Jie Han
Wei Qiu
Eric Lichtfouse

ChatGPT in Scientific Research and Writing

A Beginner's Guide

Springer

ChatGPT in Scientific Research and Writing

Jie Han · Wei Qiu · Eric Lichtfouse

ChatGPT in Scientific Research and Writing

A Beginner's Guide

Springer

Jie Han
School of Human Settlements and Civil Engineering
Xi'an Jiaotong University
Xi'an, China

Wei Qiu
International Science and Technology Cooperation Base of Xi'an Municipality
Guyiheng Technologies Ltd.
Shaanxi, China

Eric Lichtfouse
State Key Laboratory of Multiphase Flow in Power Engineering, International Research Center for Renewable Energy
Xi'an Jiaotong University
Xi'an, China

ISBN 978-3-031-66939-2 ISBN 978-3-031-66940-8 (eBook)
https://doi.org/10.1007/978-3-031-66940-8

Contents

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

J. Han et al., *ChatGPT in Scientific Research and Writing*,
https://doi.org/10.1007/978-3-031-66940-8_1

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

Prompt: 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.
Response: Here are the key points of the article in bullet points: • 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. • 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. • 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. • 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. • 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.

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 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 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 cm^{-1}" and "1080 cm^{-1}", or more accurately, "shifts from 1715 cm^{-1} to 1697 cm^{-1} and from 1090 cm^{-1} to 1080 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 cm^{-1} and 1100 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

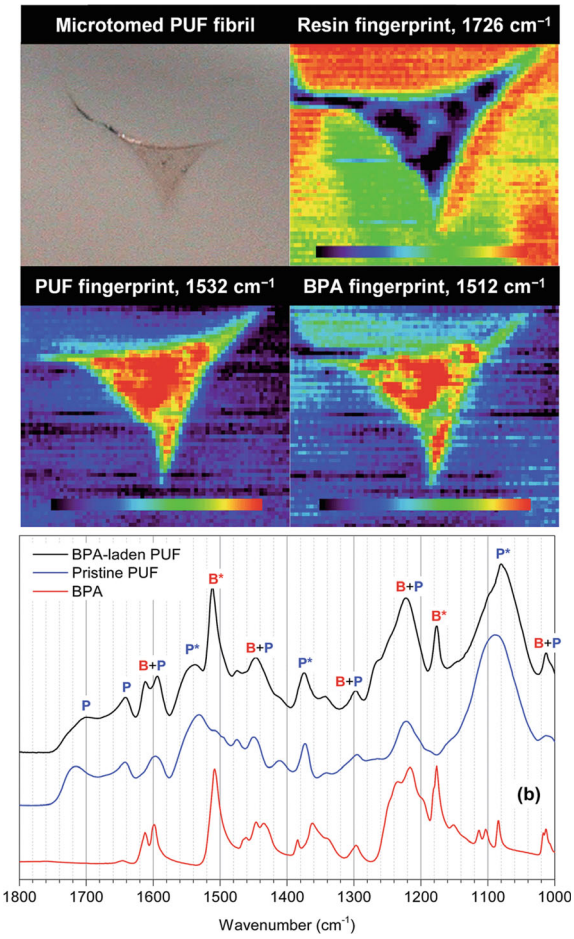

Fig. 2 **a** Microscopy and infrared images of a microtomed cross-section of a bisphenol A (BPA)-laden polyurethane foam (PUF) fibril embedded in a resin medium. Infrared images were constructed from hyperspectral data using the fingerprint peaks of LR White resin, polyurethane foam (PUF), and bisphenol 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 × 75 pixels at a resolution of 1.56 × 1.56 μm 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 ≥ 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 cm^{-1} and CH_2 stretching peak at 2871 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^{-1}** and **1100 cm^{-1}**, 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

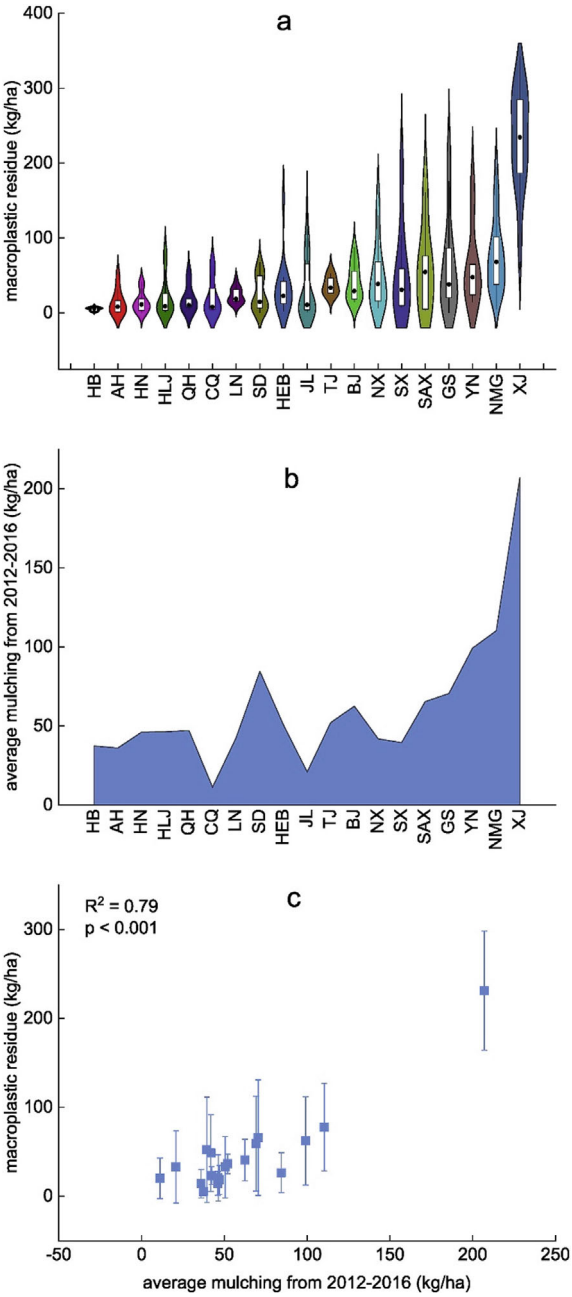

Fig. 3 Relationship between macroplastic residues and the use of plastic mulching in agricultural soils across China: **a** violin plots of the abundances of macroplastic residues in agricultural soils across China, **b** average mass of mulching film in 2012–2016, **c** relationship between macroplastic 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.

are *not* provided in the caption of the figure (see the reprinted caption in Fig. 3). After validating the model's response, we concluded that the answers could only be generated by analyzing the full text rather than paraphrasing texts from certain parts of the paper. For instance, the model's comment in the first paragraph of the example analysis, describing the distribution of macroplastics as "skewed" in most provinces, is a term that is not found anywhere in the paper by Huang et al. (2020). Nonetheless, it accurately captured the essence of the data presented in this figure. Overall, the ChatGPT-enabled new Bing interpreted figures in a largely accurate manner. The model recognizes the link between the figure and its description (caption), the authors' discussion text on the figure, and the purpose of having the figure in the paper, i.e., specific arguments supported by the figure and the significance of the results. Using short and intuitive prompts, users could gain a quick understanding of complicated-looking figures in scientific publications, especially when they are not prepared in a "stand-alone" manner, e.g., where the caption does not provide interpretations of the results in the figure with "take-home" messages for readers.

It should be pointed out that, according to the description, the built-in large language model (GPT-4) in the new Bing is capable of delineating standalone images on its own (Microsoft 2023b). This function, referred to as "multimodal visual search", allows users to prompt ChatGPT or its derived applications, e.g., new Bing, on images, drawings, or charts with related questions, and the model will try to understand the image, interpret it, and answer questions about it. For instance, new Bing allows users to "drag and drop" images directly into the chat window to access this function (Microsoft 2023b). This is useful for general interpretation of standalone images, e.g., photographs, artistic work, or when the full text is unavailable for a figure that needs to be analyzed. Nonetheless, the ability to correlate the figure to relevant texts and specific arguments in the source document allows the model to interpret the figure in a more accurate and contextually relevant manner, which is important for interpreting figures that are prepared for specialists in scientific publications.

Below is a list of our main findings in this section.

- In this section, we highlight the use of ChatGPT-enabled new Bing to quickly interpret complex figures, e.g., hyperspectral images and data plots, in research papers.
- The model delves into the full text, correlates each figure to the description of the figure and the authors' interpretation of the results in the figure, and identifies the specific conclusions or arguments supported by the figure in the paper. The model's interpretations were validated as mostly correct by the authors, although some minor inaccuracies were noted.
- Figures often contain key evidence and data in scientific papers, although they can be difficult to understand by readers without reading the full text, especially those showing many results or complex concepts, or extensively using symbols or abbreviations that are not explained in the figure captions. Large language models can analyze and interpret complex scientific figures for users to understand them without reading the associated text, by allowing users to gain an in-depth understanding of what specific figures demonstrate and prove within the broader context of the research.

4 Evaluation of Research Papers

Most scientific articles are peer-reviewed prior to the acceptance for publication. It involves a rigorous process of critically evaluating the work in the submitted manuscript, usually by journal editors and peer reviewers. One of the immediate uses that one could think of, therefore, is to ask the AI for an "independent opinion" on the manuscript before submission. Using ChatGPT or other large language models, authors may obtain critical, even constructive comments on the work presented in their manuscript. This could benefit authors by doing an instant, risk-free evaluation of their manuscript before submitting it to a scientific journal when it *will* be subject to rigorous evaluation and criticism. In fact, the benefits are two ways. Those who are constantly involved in the peer review process for scientific journals may also get a "second opinion" from the AI with comments that may have been overlooked in their own, often rushed evaluation (Dance 2023).

Apart from obtaining critical comments, knowing the *strength* of their work—in other people's eyes—is also extremely useful for authors. As a mandatory requirement, most scientific journals ask authors to submit a cover letter to journal editors, where the authors must articulate the *novelty* and *significance* of their work, with a compelling story. In addition to that, many reputable journals in environmental science disciplines require authors to submit a maximum of five concisely worded bullet points, e.g., 85 characters each, as "highlights" of their submitted work (STOTEN 2024a; Environmental Pollution 2024). Large language models like ChatGPT offer somewhat "averaged opinions", i.e., by comparing the reported work with existing publications in similar research domains. By reading the model's comments, authors could learn the "strengths" and "highlights" of their work, all in the eyes of an "average reader." Authors must, in turn, make these apparent in the title, abstract, and introduction of their article to save time for busy journal editors to "hunt the rabbits in a forest". This could significantly increase the chance of the article being considered and sent to peer reviewers. On the other hand, manuscripts with the "golden nuggets" buried deep in the text may be overlooked by journal editors during the initial screening, which is unfortunate for both the authors and journal editors who miss the good work buried in the manuscripts.

Below are several examples showing the capabilities of ChatGPT and the ChatGPT-enabled new Bing for performing such tasks. Note that for commenting on the "strengths and weaknesses" of a research paper, the specific points given by the model require some "fair judgments" by doing deeper analyses, e.g., by putting the research topics, findings, and methods in a broader literature context, in addition to critically evaluating the validity, quality, and significance of the work presented in the paper. Such tasks require a higher level of intelligence than abstracting the key findings from a research paper which requires judgments in its own right to pick the most essential points in a dense, jargon-packed scientific publication. In other words, by going from abstracting key points to asking what a specific figure shows and proves in a research paper, and now evaluating the "strengths and weaknesses" of a research paper, we are essentially increasing the level of challenges for the intelligence of

the AI large language model by creating more human-like tasks demanding higher levels of intelligence. Below are the user prompts that we have used for testing the model's capabilities on these tasks. As we found from our examples, the model did well in all three task categories.

- Level 1: List the key findings in this paper in detail, with information on how authors obtained each of these findings.
- Level 2: Tell me in detail what Figure [x] shows and proves in this paper.
- Level 3: Analyze this paper. Tell me its strengths and weaknesses in detail.

4.1 Reviews versus Original Research Articles

In the first example, we asked ChatGPT to analyze the "strengths and weaknesses" of our recent review paper by Dai et al. (2023) (Table 8). Note that this article is an open-access publication which is freely accessible online. In the prompt, we provided the title of the article without having to download its full-text PDF. The article is a regular-length review paper published *after* the knowledge cut-off date of the large language model (GPT-3.5). However, this does not seem to affect its evaluation of the contents of this paper, as evidenced by the quality of its response. After validating the response, the authors reached a consensus that all points listed under "Strengths" and Points 1, 3, and 4 listed under "Weaknesses" are either plausible or "spot-on" accurate. The model's response provided by new Bing, which is also accurate and contextually relevant, is shown in Table 9 for comparison.

In the next example, new Bing provided a succinct summary of the highlights with some fair comments on the shortcomings of a review paper (Table 10). In this paper, we broadly reviewed the aggravating ecological issues in Mongolia. Interestingly, this was picked by the model "*…lacks a clear research question or objective that guides its review and analysis*", which is a fair criticism and reflects the nature of our discussions in this article. In the third example, we asked new Bing to analyze the "strengths and weaknesses" of a recent research paper (Table 11). Since access to this paper requires a subscription or purchase, we downloaded the full-text PDF, opened it in the Microsoft Edge Dev browser, and then entered the prompt. Based on the validation of the model's response by the original authors of this paper, the points listed under "possible strengths" and "possible weaknesses" are both accurate, as well as the summary of findings in this study that is provided upfront in the model's response. Note that when responding to the user's request, the model sometimes provided general instructions upfront for users to know how to perform such tasks themselves. In this example, the instructions given at the beginning of the response are accurate, concise, and reflect the right approach for critically evaluating a scientific publication. This is also seen in a previous example when we entered a prompt that would require extensive answers from the model, i.e., analyzing each of the figures in the research paper (Table 7).

We ran a further test on the research paper by Liu et al. (2023) by asking the model (new Bing) to accentuate the limitations of the methods, findings, and discussions

Table 8 Evaluation of our recent review paper by Dai et al. (2023) using ChatGPT (GPT-3.5 model)

Prompt: Analyze this paper: It is time to acknowledge coronavirus transmission via frozen and chilled foods: Undeniable evidence from China and lessons for the world. Tell me its strength and weakness in detail.

Response:

Strengths:

1. The paper presents a timely and relevant topic that is of global interest, given the ongoing COVID-19 pandemic and the concerns regarding its transmission via food.
2. The authors provide a comprehensive review of the existing literature on the transmission of SARS-CoV-2 virus through frozen and chilled food, based on empirical evidence from China.
3. The paper offers several recommendations to mitigate the potential risks associated with the transmission of the virus via food, such as strengthening food safety measures and enhancing international cooperation.
4. The authors use a clear and concise writing style that makes the paper accessible to a broad audience.

Weaknesses:

1. The paper heavily relies on the evidence from China, which may not be directly applicable to other regions with different food systems, climates, and cultural practices. Therefore, the generalizability of the findings should be interpreted with caution.
2. The paper does not provide a systematic analysis of the limitations of the studies reviewed, such as the potential biases, confounding factors, and measurement errors, which may affect the validity and reliability of the conclusions.
3. The paper does not explore the ethical implications and trade-offs of the proposed recommendations, such as the economic costs, social impacts, and cultural sensitivities, which may hinder their implementation in practice.
4. The paper does not address some important aspects of the topic, such as the potential effects of food processing, packaging, and transportation on the transmission of the virus, or the role of consumer behavior and attitudes towards food safety.

Overall, the paper is informative and thought-provoking, but it could benefit from a more critical and nuanced analysis of the evidence and implications. Future research could explore the broader context of the issue and its implications for public health, food security, and international trade.

In this example, we provided only the title of the article in the prompt as a unique identifier of the paper, which is an open-access publication.

presented in this paper (Table 12). As validated by the original authors of this paper, the points raised by the model are valid. Note that these points are not paraphrased from the sub-section "3.5.1 Limitations" of the paper, but reflect the AI's opinion on a *different* set of limitations in the methods and findings reported in this preliminary study.

Overall, we were impressed by the capability of the model in the critical evaluation of research papers. Apparently, the model does so by algorithmically comparing them with other publications on similar topics, although we are unsure of how it defines "similarity" in topics or how many publications it would compare them with to draw these conclusions. In reality, authors and reviewers may not think of some of these points suggested by the model as they prepare or review the manuscript, because the literature survey conducted by the authors or peer reviewers is not *exhaustive* and can have certain degrees of randomness (in fact, ChatGPT also shows randomness in its answers when responding to the same prompt in different chat sessions). This means

Table 9 Evaluation of the review paper by Dai et al. (2023) using ChatGPT-enabled new Bing under the "More Creative" mode

Prompt: Analyze this paper. Tell me its strength and weakness in detail.
Response: To analyze a scientific paper, you need to read it carefully and critically, evaluate its methods, results, discussion and conclusions, and compare it with other relevant papers in the field. Based on the web page context from system, here is a brief analysis of the paper's strength and weakness: Strength: • The paper provides a comprehensive review of the recent re-emergent outbreaks of COVID-19 in China caused by fomite transmission via imported frozen and chilled foods. • The paper summarizes the epidemiological investigations, laboratory studies, and control measures taken by public health authorities in response to those outbreaks. • The paper also discusses the implications of this specific route of transmission for global food safety and biosecurity. • The paper is well-structured, well-referenced, and well-written. Weakness: • The paper relies heavily on official reports and media sources from China, which may have some limitations or biases in terms of data quality, accuracy, and transparency. • The paper does not provide a systematic assessment of the risk factors, mechanisms, and prevalence of fomite transmission via cold-chain foods across different countries and regions. • The paper does not address some potential confounding factors or alternative explanations for the re-emergent outbreaks, such as human-to-human transmission, environmental contamination, or viral mutation. • The paper does not offer any specific recommendations or guidelines for preventing or mitigating this route of transmission at the global level.

To generate the response, we downloaded the full-text PDF of the review paper, opened it in the Microsoft Edge Dev browser, and then typed in the prompt in the sidebar to initiate the conservation with new Bing. Bullet points are added in the response for clarity.

that the shortcomings and highlights pointed out by the model can indeed provide valuable supplementary insights for both authors and reviewers, especially when the user, e.g., a reviewer is not already familiar with the literature context in the research domain or when there is simply not enough time to revisit the literature studies to make a detailed analysis, e.g., by putting the work in the evaluated manuscript into the broader literature context.

To further gauge the capabilities of the model against human experts, we ran additional tests on our previously submitted manuscripts to validate whether the large language model could indeed serve as a competent peer reviewer. We found that in most cases, the model was unable to identify *specific* issues in those manuscripts, e.g., those the reviewers had raised during the peer review process and the authors agreed to address, even with guided prompts (results not shown). Since addressing specific issues is a part of the peer review process, it is evident that the current capabilities of ChatGPT cannot truly fulfill the role of human peer reviewers. To summarize, we conclude that at present, ChatGPT can provide accurate and valuable general remarks on research papers, but they cannot fulfill the role of peer reviewers by providing sufficiently targeted comments on specific issues in manuscripts.

Table 10 Evaluation of an earlier review paper by Han et al. (2021a) using ChatGPT-enabled new Bing under the "More Creative" mode

Prompt: Analyze this paper. Tell me its strength and weakness in detail. List them in bullet points.
Response: This is my analysis of the paper in the current web page context: Strengths: • The paper provides a comprehensive review of the recent climate and ecological changes in Mongolia, a region that is highly vulnerable to desertification and sandstorms. • The paper uses various sources of data, such as satellite imagery, meteorological records, tree ring data, and official statistics, to support its arguments and findings. • The paper identifies the key driving factors of desertification and sandstorms in Mongolia, such as climate change, overgrazing, human activities, and positive feedback loops between soil moisture and surface warming. • The paper discusses the trans-regional impact of sandstorms in East Asia and calls for collaborative efforts from policy makers, local residents, and scientists to address the ecological challenges in Mongolia. Weaknesses: • The paper lacks a clear research question or objective that guides its review and analysis. • The paper does not provide a systematic or quantitative assessment of the relative contributions of different driving factors to desertification and sandstorms in Mongolia. • The paper does not propose or evaluate any specific mitigation strategies or solutions for combating desertification and sandstorms in Mongolia. • The paper does not compare or contrast the situation in Mongolia with other regions that are facing similar ecological issues.

Below is a list of our main findings and opinions in this subsection.

- Subject to the policies and requirements, e.g., by journal publishers, ChatGPT or its derived application, e.g., new Bing, can provide "independent assessments" on research publications or manuscripts. The model can highlight the particular areas needing improvement by providing insights into the weaknesses and limitations of the work presented. In addition, the model can help users identify the strengths of the research work evaluated, which can be useful for articulating these points more clearly.
- Those who are involved in the peer review process may use large language models as a supplementary tool to gain additional perspectives that may be overlooked in their own evaluations. Likewise, users must check the current policies and requirements of journal publishers before engaging any of these tools in the peer review process to ensure full compliance.
- Although the model can offer highly constructive and context-relevant remarks, it generally evaluates the broad aspects of manuscripts and lacks the ability needed to identify specific issues in research publications which tend to be picked up by human reviewers in the peer review process.

Table 11 Evaluation of a recent research paper by Liu et al. (2023) using ChatGPT-enabled new Bing under the "More Creative" mode

Prompt: Analyze this paper and tell me its strength and weakness in detail.
Response: To analyze a scientific paper, you need to read it critically and evaluate its quality, validity, and significance. You also need to compare it with other relevant papers and identify its strengths and weaknesses. Here is a possible summary of the paper based on the web page context: • The paper investigates the release of microplastics and other particles from six types of breastmilk storage bags during simulated use. • The authors use various methods to characterize the size, shape, morphology, and composition of the particles and quantify their abundance. • The authors estimate the potential exposure of infants to these particles through breastmilk consumption and discuss the health and environmental implications. • The paper provides first evidence of microplastic contamination from breastmilk storage bags and calls for further studies and risk assessment. Some possible strengths of the paper are: • It addresses a novel and important topic that has implications for infant health and food safety. • It uses multiple techniques to identify and characterize the particles released from breastmilk storage bags, such as optical microscopy, micro-Raman spectroscopy, SEM-EDS, and weighing. • It provides quantitative estimates of particle abundance and infant exposure based on realistic scenarios of breastmilk storage and consumption. • It discusses the limitations and uncertainties of the methods and results and suggests directions for future research. Some possible weaknesses of the paper are: • It does not compare the results with other sources of microplastic exposure for infants, such as baby bottles, teats, or breastmilk itself. • It does not analyze the chemical additives or contaminants that may be present in the plastic materials or the particles. • It does not assess the biological effects or toxicity of the particles on human cells or tissues. • It does not provide any recommendations or solutions to reduce or prevent microplastic contamination from breastmilk storage bags.

Bullet points are added to the first part of the response for clarity.

4.2 Current Policies

An important reminder is that both authors and reviewers must be aware of the intellectual property rights, confidentiality, and privacy that may be compromised by sending unpublished content to ChatGPT or other AI tools. Some scientific journal publishers recently updated their policies on the use of ChatGPT and other AI tools in the peer review process (Springer Nature 2023; Nature Portfolio 2024; Elsevier 2024c; Taylor & Francis 2024). Users must be aware of these restrictions before they engage AI tools in their evaluation or self-evaluation of manuscripts or research proposals.

In the latest update on "*Peer Review Policy, Process and Guidance*", Springer Nature stated that while the publisher explores providing peer reviewers with access to safer AI tools, it asks that peer reviewers do not upload manuscripts into generative AI tools (Springer Nature 2023; Nature Portfolio 2024). However, if any part of the

Table 12 Evaluating the limitations of the study by Liu et al. (2023) using ChatGPT-enabled new Bing under the "More Creative" mode

Prompt: You are a peer reviewer of the research article in the web browser. You need to be rigorous, skeptic, harsh, and constructive. List the main limitations in the methods, findings, and discussions of this study, and explain them in detail.
Response: Based on the current web page context and the web search results, some of the main limitations in the methods, findings, and discussions of this study are: • The sample size of six breastmilk storage bags is too small to represent the diversity and variability of products available in the market. The authors should explain how they selected the products and whether they are representative of different brands, materials, and designs of breastmilk storage bags. • The simulated use of breastmilk storage bags in this study does not reflect the realistic conditions and practices of storing and feeding breastmilk to infants. The authors should provide more details on how they filled, sealed, shook, and filtered the bags, and whether these procedures mimic the actual use by consumers. They should also consider other factors that may affect the release of microplastics, such as temperature, storage time, freezing and thawing, heating and cooling, and exposure to light. • The analytical methods used in this study have some limitations and uncertainties that may affect the accuracy and reliability of the results. The authors should acknowledge the challenges and errors of using optical microscopy, micro-Raman spectroscopy, SEM-EDS, and weighing for identifying, characterizing, and quantifying microplastics and other particles. They should also report the quality assurance and quality control measures they took to minimize contamination and cross-contamination during sample preparation and analysis. • The findings of this study are of a preliminary nature and cannot be generalized or extrapolated to assess the health risks of microplastic exposure for infants. The authors should discuss the limitations of their exposure assessment, such as the lack of information on the toxicity, bioavailability, bioaccumulation, and biotransformation of microplastics and other particles in breastmilk and infant body. They should also compare their results with other studies on microplastic contamination in infant food and products, and suggest future research directions to fill the knowledge gaps.

evaluation of the claims made in the manuscript was in any way supported by an AI tool, Springer Nature asks peer reviewers to declare the use of such tools transparently in the peer review report. In the peer review guidelines, the publisher pointed out the limitations of generative AI tools that they can lack up-to-date knowledge and may produce nonsensical, biased, or false information. Moreover, manuscripts may include sensitive or proprietary information that should not be shared outside the peer review process (Springer Nature 2023; Nature Portfolio 2024). According to Springer Nature, these are the reasons that the publisher advises peer reviewers to use caution when using any use of generative AI tools during the peer review process.

Users must be aware that some publishers have more strict rules concerning the use of generative AI and AI tools in the peer review process. Elsevier, for instance, requires peer reviewers to treat submitted manuscripts and peer review reports as confidential documents, and prohibits the use of generative AI and AI-assisted technologies by reviewers in the journal peer review process, even for improving language and readability. The scientific journal publisher explains that uploading a submitted manuscript or any part of it into a generative AI tool may violate the authors' confidentiality and proprietary rights and, where the paper contains personal information, may breach data privacy rights. This requirement extends to peer review reports which

may also contain confidential information about the manuscript and the authors. The publisher expresses further concerns about using generative AI or AI-assisted technologies in the peer review process with the following arguments. First, only human reviewers can be responsible and accountable for the content of the review report. Second, the critical thinking and original assessment needed for peer review are outside of the scope of this technology, including generative AI and AI-assisted technologies. Third, there is the risk that AI tools may generate incorrect, incomplete, or biased conclusions. Readers may read the full text under "Publishing ethics", "Duties of Reviewers", in the subsection entitled "*The use of generative AI and AI-assisted technologies in the journal peer review process*" (Elsevier 2024c). Some scholarly publishers, e.g., Taylor & Francis, have similar requirements (Taylor & Francis 2024; Cambridge University Press 2023). According to the current guidelines, peer reviewers should not "*upload files, images, or information from unpublished manuscripts into databases or tools that do not guarantee confidentiality, are accessible by the public and/or may store or use this information for their own purposes*", including generative AI tools like ChatGPT (Taylor & Francis 2024).

The publishers are correct about confidentiality, proprietary rights, and data privacy rights, and for that, providing access to safer AI tools for peer reviewers, as Springer Nature states in its guidelines, is *urgently* needed. On the second point, this may be correct. However, generative AI such as ChatGPT and its derived applications, e.g., new Bing, already demonstrates competency in this task category, e.g., Tables 8, 9, 10, 11 and 12, and it will be interesting to see how they evolve in the next generations of this technology. Note that this is also true for the third point where the model should be able to improve and generate even better responses when evaluating scientific content. On the first point, it is plausible that reviewers are ultimately accountable for the contents of peer review reports, *regardless* of whether they have used generative AI or AI-assisted technologies to assist with their review. Ultimately, the reviewer is responsible for the content they submit and, depending on the policies of the publisher, this may include reviewing AI's comments, fact-checking, and making his or her own judgment on whether any of the AI's inputs are useful for the peer review process *after* making his or her own independent assessment.

From the perspective of journal editors, we can usually tell if the entire or most of the review comments are generated by AI. One of the criteria used in our judgment, for instance, is to see if such comments are relevant but broad in nature, and lack sufficient details addressing specific issues present in the manuscript. For instance, when analyzing the weaknesses and limitations of research papers, ChatGPT tends to branch out to topics beyond the content in focus. Human experts, on the other hand, tend to pay closer attention to specific issues in manuscripts, such as the methods, QA/QC, presentation and interpretations of the results, as well as the structure, writing, language and spelling, *in addition* to an overall assessment on the validity, significance, and novelty of the scientific work reported in the manuscript. Rather than speaking broadly in the comments using polite, polished language with a constructive tone (this seems to be the default style by the AI when facing challenges from human users), human experts usually provide highly targeted, content-specific comments on particular issues present in the manuscript, often with a less polished language and

more straight-to-point writing in their criticisms. It is certainly possible to refine the prompts and obtain more specific comments from the model, like what we did in the last example (Table 12), but doing so requires significantly more time and meticulous fact-checking afterward, which defies the purpose for those who want to exploit AI to generate all or most of the review comments needed. The other signature is in the writing itself. Based on our experience, the succinctness, coherence, and polished writing surpass the writing that we have seen in a majority of the publications and manuscripts in the environmental research domain. Put the science and facts aside, the writing is so polished that feels like observing an art craft that is machined to precision. Like ourselves, readers may also get goosebumps by reading AI's writing, much like reading the work of a professional writer in the English language. The writing feels unrealistically smooth and carefully worded, especially for early-career researchers who write in English as a second language.

Below are the key points we discussed in this subsection.

- Authors and reviewers must be aware of the intellectual property rights, confidentiality, and privacy issues that can arise from the use of large language models or other AI tools, including ChatGPT and its derived applications, on unpublished or copyright-protected content.
- Scientific publishers recently updated their policy guidelines regarding the appropriate use of AI tools in publications. For instance, Springer Nature is exploring safer AI tools for peer reviewers but currently advises against uploading any unpublished content to AI tools. They also require that any AI involvement in the peer review process be declared. Elsevier prohibits the use of generative AI and AI-assisted technologies in the peer review process entirely.
- Despite the current limitations and risks associated with using AI in the evaluation of research outputs, our examples have shown that the models (GPT-3.5 and GPT-4) are showing competence in this task category. As these models continue to evolve and provided that data security and copyright issues can be fully addressed, these AI tools may eventually become a powerful tool for scientific researchers and publishers.
- In all cases, users must ensure compliance with their institutions' policies and publishers' guidelines on the appropriate use of large language models or other AI tools in the scientific writing or peer review process, and validate the model's response before using any information generated by AI tools.
- Journal editors can often detect when review comments are entirely or partly generated by AI tools, as these tend to be broad in nature and lack details specific to the issues found in the particular manuscript. Human reviewers, on the other hand, provide feedback that targets specific issues in the manuscripts as well as overall assessments of the work presented, often with more blunt and less polished writing.

5 Spotting Errors

5.1 Scientific Errors

Errors are *not* uncommon in scientific publications (Pulverer 2015; Aboumatar et al. 2021; Besançon et al. 2022; van Ravenzwaaij et al. 2023). In fact, reproducibility is known to be one of the biggest issues facing science today (Baker 2016; Amaral and Neves 2021; Marshall-Cook and Farley 2024). Many scientific journals, such as *Nature* Portfolio journals and *Science* family of journals, and some journals in environmental science disciplines, e.g., *Environmental Science & Technology* and *Science of the Total Environment* allow readers to correct mistakes and comment on issues in their published papers (Nature 2024; Science 2024; ACS 2023; STOTEN 2024b). In reality, such mechanisms are used sparingly, because of the significant efforts required to identify those errors and the often-inadequate credit given to those who do make such efforts post-publication.

As our first example, here we refer to a critical review published in *Water Research* entitled "Mistakes and inconsistencies regarding adsorption of contaminants from aqueous solutions: A critical review" (Tran et al. 2017). This review paper addressed common errors in published studies on the adsorption of pollutants in water and aqueous solutions, offering many corrections and detailed explanations. One of the authors, J. Han, was invited by Prof. Mark van Loosdrecht, the Editor-in-Chief of *Water Research* at the time, to serve as a peer reviewer of this paper. Here we use this review paper as our reference and take a closer look at the capabilities of ChatGPT on spotting errors in scientific publications. Specifically, we asked the model to analyze the mistakes in two publications (Li et al. 2011; Zafar et al. 2007), which contain errors in concepts, terminology, or mathematical equations as pointed out by Tran et al. (2017) in the review and further validated by us.

In the first example (Table 13), the statement excerpted from the referenced study, i.e., "*For PFOSA (pKa = 6.52), when pH < pKa, protonation occurs on the amino group, and the decreased protonation leads to the increased adsorption, but when pH > pKa, PFOSA exists as neutral molecule in water*", contains multiple errors. First, the *n*-perfluorooctanesulfonamide molecule contains a *sulfonamide* group, not an "amino" group. When an amine is considered as the functional group of a molecule, it is referred to as an "amino group". In this case, the sulfonyl group (O=S=O) connected to the amine group ($-NH_2$) forms a distinct group, i.e., the sulfonamide, which is a rigid moiety with antibacterial properties that are used in several groups of commercial drugs. Second, the sulfonyl group (O=S=O) renders the hydrogen atom on the amine group relatively acidic, i.e., electrophilic. As a result, the sulfonamide behaves as a weak organic acid, not a base, which dissociates in strong alkaline solutions. For instance, the estimated acid dissociation constant (K_a) of sulfanilamide, a drug compound in the sulfonamide class, is 2.2×10^{-11}, or $pK_a = 10.7$ (Schmelkes et al. 1942). More estimated pK_a values and pH-dependent speciation diagrams of perfluoroalkyl sulfonamides can be found in the study by Rayne and Forest (2009). When pH > p*K*a, PFOSA exists as a negatively charged anion in water, not a neutral

molecule, where the sulfonamide group can lose a proton and form a conjugate base (Rayne and Forest 2009). Third, the acid dissociation constant (K_a), or by convention its negative logarithmic value (pK_a), is an *equilibrium* constant, i.e., when the dissociated and co-existing neutral species have equal molar concentrations in the solution, not an *onset* value as the authors suggested in the following statement "*…when pH > pKa, PFOSA exists as neutral molecule in water*" (Li et al. 2011). Last, the equilibrium constant for the reaction of a weak base with water is named the base ionization constant (K_b), or its negative logarithmic value (pK_b). Readers may refer to the book "*Environmental Organic Chemistry*" by Schwarzenbach et al. (2002) for clarification on these concepts and the correct use of terminologies.

Using a general, non-targeted prompt, the ChatGPT-enabled new Bing correctly spotted two of these errors in this example by receiving no hint from the user (Table 13), but it failed to provide correct references on its statements. The references provided by new Bing (not shown in the table), i.e., Wikipedia (2024), PubChem (2024), and Burns et al. (2008) contained no data or facts supporting its arguments in the model's response. In the second example, we used a guided prompt to identify errors in the mathematical equations in a review paper by Zafar et al. (2007) (Table 14). In the first part of the response, new Bing provided the correct equations for all of the four models with clear explanations of the model parameters but was unable to spot an error in one of the equations in the paper, i.e., the linearized equation of the pseudo-second-order kinetic model $(t/q) = 1/(k * q_e^2) + 1/(q_e t)$, which should be written as $(t/q) = 1/(k * q_e^2) + t/q_e$. Additionally, one of the corrections

Table 13 Spotting errors in a statement in a research paper (Li et al. 2011), using the GPT-enabled new Bing under the "More Creative" mode

Prompt: Below are some discussion texts under the section "3.5 Effect of contaminant solution character on adsorption". I want you to read the entire article, understand the context, and identify the mistakes in these discussion texts. For PFOSA ($pKa = 6.52$), when pH < p*Ka*, protonation occurs on the amino group, and the decreased protonation leads to the increased adsorption, but when pH > p*Ka*, PFOSA exists as neutral molecule in water.
Response: Hello, this is Bing. I can help you identify the mistakes in the discussion texts based on the current web page context and web search results. The discussion text for PFOSA contains two mistakes: • The first mistake is that PFOSA does not have an amino group, but a sulfonamide group. An amino group has the formula $–NH_2$, while a sulfonamide group has the formula $–SO_2NH_2$. You can see the structure of PFOSA in the web search result. • The second mistake is that when pH > p*Ka*, PFOSA does not exist as a neutral molecule in water, but as a negatively charged anion. This is because the sulfonamide group can lose a proton and form a conjugate base with the formula $–SO_2NH^-$. You can see the p*Ka* value of PFOSA and its dissociation reaction in the web search result. I hope this helps you understand the errors in the discussion text. Do you have any other questions?

A line break is added in the user prompt for clarity. By default, new Bing removes all line breaks in the user prompt which can make it difficult to discern the prompt and excerpted texts. Minor editing is applied to the symbols in the response for clarity.

Table 14 Spotting errors in mathematical equations in a review paper (Zafar et al. 2007)

Prompt: In the review on this webpage (Biosorption of nickel from protonated rice bran, https://doi.org/10.1016/j.jhazmat.2006.09.055), the authors wrote the equations of the following models on adsorption kinetics and isotherms: the pseudo first-order Lagergren model, the pseudo second-order model, the Langmuir adsorption isotherm model, the Freundlich model. I want you to take a closer look at these equations. If there is any mistake in these equations, I want you to find it and explain it to me in detail.

Response: The equations of the models on adsorption kinetics and isotherms are as follows

- The pseudo first-order Lagergren model

$$\frac{dq_t}{dt} = k_1(q_e - q_t)$$

- The pseudo second-order model

$$\frac{dq_t}{dt} = k_2(q_e - q_t)^2$$

- The Langmuir adsorption isotherm model

$$\frac{1}{q_e} = \frac{1}{q_m} + \frac{1}{K_L q_m C_e}$$

- The Freundlich model

$$q_e = K_F C_e^{1/n}$$

where q_t and q_e are the amounts of adsorption at time t and at equilibrium (g/kg), k_1 and k_2 are the rate constants of adsorption (min^{-1} and g/mg min, respectively), q_m is the maximum adsorption capacity (g/kg), K_L is the Langmuir constant (L/g), C_e is the equilibrium concentration of metal ion in the solution (g/L), and K_F and n are the Freundlich constants.

According to web search results, there are some mistakes in these equations. The first mistake is that the pseudo first-order Lagergren model should be written as

$$\log(q_e - q_t) = \log q_e - k_1 t$$

which can be obtained by integrating the differential equation with respect to time. This form is more convenient for fitting experimental data by plotting $\log(q_e - q_t)$ versus t.

The second mistake is that the Langmuir adsorption isotherm model should be written as

$$\frac{c_e}{q_e} = \frac{c_e}{q_m} + \frac{1}{K_L q_m}$$

which can be obtained by rearranging the terms. This form is more convenient for fitting experimental data by plotting c_e/q_e versus c_e.

The third mistake is that the Freundlich model should be written as

$$\log q_e = \log K_F + \frac{1}{n}\log c_e$$

which can be obtained by taking logarithms on both sides. This form is more convenient for fitting experimental data by plotting $\log q_e$ versus $\log c_e$.

These mistakes may affect the accuracy and reliability of the model parameters and predictions.

For clarity, minor edits are applied to the mathematical symbols, and line breaks are added to the model's response.

made by new Bing, i.e., the pseudo-first-order Lagergren equation, contains an error on its own. The equation should be written with a natural log, not a logarithm, i.e., $\mathrm{Ln}(q_e - q_t) = \mathrm{Ln}(q_e) - k_1 \mathrm{t}$.

Overall, we found that the model could accurately identify some errors but failed to recognize other mistakes in the two examples, even with clear hints and guiding

instructions in the prompts (Tables 13 and 14). Despite these shortcomings, the current model still offers valuable assistance for researchers and publishers to spot errors in published research. Since reducing errors in scholarly publications is a crucial part of maintaining the integrity of scientific records, we remain hopeful about future models' capabilities and AI technologies to help scientific researchers and publishers tackle this longstanding challenge.

5.2 Misquotation

We then tested the model for identifying misquotation, a type of error that is less visible but fairly common in the scientific literature (Table 15). For this test, we used one of our earlier review papers (He et al. 2021a) and asked the ChatGPT-enabled new Bing whether the quoted statements contained errors by misquoting findings in the referenced study. Note the subtle corrections of wording suggested by the model, which are highlighted in bold fonts in the response. In this example, the AI did a good job by refining the statements with more accurate wording that better aligns with the description and author's interpretation of the experimental observations in the referenced study.

Virtually all existing publications, such as journal papers, patents, books and chapters, technical standards, reports, and other types of scientific or technical publications can be scrutinized by AI for errors. With validation by human experts, we could potentially reduce the number of errors in these publications, paving the road for future researchers with less erroneous information. Image manipulation, plagiarism, falsified data, and studies with poor QA/QC or major flaws in methodologies may be subject to AI scrutiny in the future. Given the vast body of existing literature and the increasing volume of scientific literature that is published every year, even a marginal improvement in the rigor of scientific publications is meaningful for future scientists.

Below are our main findings in this section.

- Errors are common in scientific publications, which are recognized within the scientific community. Many journals in the environmental science discipline provide venues for authors and readers to correct mistakes and comment on issues in their publications to ensure scientific rigor. Such mechanisms are, however, underused due to the meticulous work often required and inadequate recognition given to error-spotters.
- We found that ChatGPT and its derived application, i.e., the new Bing, can help identify issues in research publications such as misconceptions, incorrect terminology use, errors in mathematical equations, and misquotation. Users need to be aware that such corrections are often incomplete, mixed with errors generated by the model, and therefore require careful validation by users.
- Despite these limitations, large language models are showing promise as an efficient augmented tool for identifying errors in scientific publications, although currently they cannot replace the meticulous error-checking process by human experts.

Table 15 Identifying misquotations in a review paper (He et al. 2021a) and suggesting corrections in wording

Prompt: Read the sentence below and tell me whether it contains errors by misquoting the findings in the referenced study. The full text of the referenced study is opened in the browser. Davy et al. (2018) found large quantities of coronavirus RNAs in the intestines of hibernating Little brown bats (*Myotis lucifugus*) co-infected with the white-nose syndrome, which confirmed that responses of extracellular co-infections had led to amplified coronavirus replication and increased viral shedding from bats.
Response: Yes, the sentence contains errors by misquoting the findings in the referenced study. Based on the current web page context, which is the full text of Davy et al. (2018), the sentence should be corrected as follows: Davy et al. (2018) found **higher levels** of coronavirus RNAs in the intestines of hibernating Little brown bats (*Myotis lucifugus*) co-infected with the white-nose syndrome, which **suggested** that **the systemic effects** of extracellular co-infections had led to **increased coronavirus replication and increased potential of virus shedding** from bats.
Text excerpts from the referenced study (Davy et al. 2018): • Here, we tested the prediction that little brown bats (*Myotis lucifugus*) co-infected with the *M. lucifugus* coronavirus (*Myl-CoV*) and with *Pseudogymnoascus destructans* (Pd), the fungus that causes bat white-nose syndrome (WNS), exhibit different disease severity, viral shedding and molecular responses than bats infected with only *Myl-CoV* or only *P. destructans*. Here, we show that the intestines of virus-infected bats that were also infected with fungus contained on average 60-fold more viral RNA than bats with virus alone. Increased viral RNA in the intestines correlated with the severity of fungus-related pathology… Levels of coronavirus antibodies were also higher in fungal-infected bats. Our results suggest that the systemic effects of WNS may down-regulate anti-viral responses in bats persistently infected with *M. lucifugus* coronavirus and increase the potential of virus shedding. • Our findings suggest that systemic responses of bats to WNS results in increased coronavirus replication and consequently, increased viral shedding, which may lead to subsequent infection of susceptible animals…In this study, we showed that a coronavirus exhibits low activity in its natural host, *M. lucifugus*, but that co-infection with a fungus increases the quantity of viral RNA in the intestines. • Our results suggest that secondary skin infection with the fungus, *P. destructans*, substantially increases the level of viral RNA in the intestine of hibernating bats…Infection with *P. destructans* causes modulation of a number of immune responses, including down-regulation of interleukin and cell proliferation genes which may compromise bats' ability to suppress viral activity. Taken together, our results have implications for epidemiological studies of *P. destructans*, the WNS fungus and for research into viral spillovers, which should consider the potential implications of co-infections that increase viral shedding.

This example shows the capability of the ChatGPT-enabled new Bing to identify deeper, not-so-obvious issues in research publications. For clarity, a line break and italicized fonts are added in the user prompt and the model's response. Original text excerpts from the referenced study (Davy et al. 2018) are added to the table as a reference for readers to validate the model's response.

6 Responding to Reviewer's Comments

6.1 Rebuttal to Criticism

Addressing reviewer's comments, especially responding to their criticism, is a challenging task even for well-versed scientists. In this test, we tested the GPT-enabled new Bing on preparing rebuttals and responses to reviewers' comments. For this purpose, we selected two of our publications, namely an earlier review paper by He et al. (2021a) and a recent research paper by Liu et al. (2023). Both papers underwent rigorous peer reviews before publication. The review paper, in particular, faced intense criticism from a group of bat conservation scientists who requested the authors' responses to address their concerns post-publication. The research paper, which investigated the presence of microplastic contaminants in breastmilk storage bags, faced similar scrutiny and went through substantial revisions with answers to a long list of questions before its acceptance.

For this test, we chose the very critical and content-specific comments given by the reviewers to increase the level of challenges for the model. To help the model understand the reviewer's comments unambiguously, we made minor edits to the reviewer's comments before inserting them into the user prompts. In all of our tests, we allowed new Bing to access the full text of our original manuscripts submitted to the journals, including the supplementary material. To clarify, we only use these as examples and none of the text generated by new Bing, including the contents, reference citations, and underlying strategies, were adopted in our actual responses submitted to the journals.

The first test was an open question which we expected the AI to do quite well. At first glance, new Bing provided a solid rebuttal with well-reasoned arguments supported by plausible-sounding references (Table 16). A closer look at the model's response, however, reveals several issues that must be taken into account by users. First, the response by new Bing *evades* the question in the comment which targeted a statement in the article on bats as "the speculated natural host of SARS-CoV-2." Therefore, listing solid evidence of bats being the host of "SARS-CoV-2 *related* coronaviruses" does *not* address the criticism directly, and this explanation is not consistent with the authors' writing in the paper either. In other words, although the explanations are good, the model completely missed the target in this case and took a wrong strategy to refute the criticism.

Here is what we actually did. While the excerpted text in the comment makes this link sound assertive, i.e., bats and SARS-CoV-2, the actual writing in the paper reads "*Many have speculated that bats, a pervasive species and a major reservoir of coronaviruses and other viral pathogens, are natural hosts of those novel zoonotic pathogens, including the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2).*" In our response sent to the group of bat conservation scientists, we pointed out the speculative nature of this suspected link upfront and referred to peer-reviewed publications discussing this very link at the time of writing. Therefore, we stand correct with this statement in the paper. As a reference, our actual response is also shown in the last part of Table 16.

Table 16 Preparing a rebuttal to the criticism of a review paper by He et al. (2021a) which discusses the origin of the novel coronavirus (SARS-CoV-2) and a possible scenario of its spillover to humans, using ChatGPT-enabled new Bing

Prompt: Read this article and provide a detailed rebuttal to the Reviewer's comments below, using information in the article, references cited in the article, and web searching results if needed.

Reviewer's comments: 1. Introduction. The first paragraph of the Introduction contains the most egregious error "*…bats …are natural host of novel zoonotic pathogens, including the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2)…*" SARS-CoV-2 has never been found in any bat species and to suggest otherwise is severely damaging to bat conservation efforts worldwide.

Response: A possible rebuttal to the reviewer's comment is: We appreciate the reviewer's concern about the potential impact of our statement on bat conservation efforts. However, we did not claim that bats are the direct source of SARS-CoV-2, but rather that they are natural hosts of **related** coronaviruses that may have contributed to the origin of the pandemic virus. This is supported by multiple lines of evidence from genomic, phylogenetic, and serological studies.

First, genomic analyses have shown that SARS-CoV-2 belongs to the Sarbecovirus subgenus within the Betacoronavirus genus, which is predominantly found in bats (Zhou et al. 2020). The closest known relative of SARS-CoV-2 is RaTG13, a bat coronavirus isolated from *Rhinolophus affinis* in China, with 96.1% genome identity (Zhou et al. 2020). Other SARS-CoV-2 related coronaviruses have also been detected in bats from different regions and species, such as RmYN02 from *Rhinolophus malayanus* in China (Zhou et al. 2020), RacCS203 from *Rhinolophus acuminatus* in Laos (Temmam et al. 2022), and RshSTT200 from *Rhinolophus shameli* in Cambodia (Latkin et al. 2021). These findings suggest that bats harbor a high diversity and abundance of Sarbecoviruses that may share a common ancestor with SARS-CoV-2.

Second, phylogenetic analyses have revealed that SARS-CoV-2 is most closely related to bat coronaviruses in the receptor-binding domain (RBD) of the spike protein, which determines the specificity and affinity for the human angiotensin-converting enzyme 2 (ACE2) receptor (Andersen et al. 2020). The RBD of SARS-CoV-2 differs from that of RaTG13 by only four amino acid residues, and from that of RacCS203 or RmYN02 by only one or two residues (Temmam et al. 2022). Moreover, some bat coronaviruses, such as RacCS203 and RmYN02, have been shown to bind more efficiently to human ACE2 than SARS-CoV-2 and mediate entry and replication in human cells (Temmam et al. 2022). These results indicate that bat coronaviruses have the potential to infect humans through ACE2-mediated mechanisms.

Third, serological studies have provided evidence of exposure and cross-reactivity between SARS-CoV-2 and bat coronaviruses. SARS-CoV-2 neutralizing antibodies have been detected in bats of the same colony as RacCS203 in Laos and in a pangolin at a wildlife checkpoint in Thailand (Temmam et al. 2022). Antisera raised against the RBD of RmYN02 was able to cross-neutralize SARS-CoV-2 despite the fact that the RBD of RacCS203 or RmYN02 failed to bind ACE2 (Temmam et al. 2022). These findings suggest that some bat coronaviruses share antigenic epitopes with SARS-CoV-2 and may elicit cross-protective immune responses.

In summary, we believe that our statement is based on solid scientific evidence and does not imply any direct causality between bats and COVID-19. We acknowledge that other animal species may have played a role as intermediate hosts or reservoirs of SARS-CoV-2, but this does not exclude the possibility that bats are involved in the evolutionary history and emergence of the virus. We also emphasize that bats are important components of biodiversity and ecosystem services, and that they should not be blamed or persecuted for their potential association with zoonotic pathogens. We agree with the reviewer that bat conservation efforts should be supported and promoted, especially in light of the increasing anthropogenic pressures and environmental changes that may facilitate spillover events and disease transmission.

References:

- Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF (2020) The proximal origin of SARS-CoV-2. Nat Med 26:450–452.
- Latkin EA, Wacharapluesadee S, Duengkae P et al. (2021) A novel SARS-CoV-2 related coronavirus in bats from Cambodia. Nat Commun 12:6018.
- Temmam S, Vongphayloth K, Baquero E et al. (2022) Bat coronaviruses related to SARS-CoV-2 and infectious for human cells. Nature 604:330–336.
- Zhou P, Yang XL, Wang XG et al. (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579:270–273.

(continued)

Table 16 (continued)

Our actual response: We never made this assertation in the article. Instead, we made it very clear at the beginning of the sentence (see full texts excerpted below) on the speculative nature of this link.

Note the statement at the beginning (in bold font).

"**Many have speculated that** *bats, a pervasive species and a major reservoir of coronaviruses and other viral pathogens (Kupferschmidt 2017; Maxmen 2017; Sallard et al. 2021; Segreto et al. 2021), are natural hosts of those novel zoonotic pathogens, including the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) which has caused the recent COVID-19 pandemic with about 160 million people already infected around the globe (WHO 2021).*" (https://doi.org/10.1007/s10311-021-01291-y)

There has been a common speculation that the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) may have links to bats, and our statement above merely reflect this speculation in the current literature context. As pointed out in the comments, this has *not* been confirmed. Knowing the speculative nature of this link, we hence used the phrase "**many have speculated that**" to reflect the hypothetical nature of this link put forward in recent scholarly publications (see examples below). We also noted that Dr Aaron T. Irving (aaronirving@intl.zju.edu.cn), one of the scientists listed as a co-signatory of the comments, published an article (*Nature*, 589, 2021, 363–370) where authors discussed this speculative link.

"*There have been several major outbreaks of emerging viral diseases, including Hendra, Nipah, Marburg and Ebola virus diseases, severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS)-as well as the current pandemic of coronavirus disease 2019 (COVID-19). Notably, all of these outbreaks have been linked to suspected zoonotic transmission of bat-borne viruses.*"

Similar hypotheses were put forward by others:

Zhou et al. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579, 270–273.

"*Simplot analysis showed that 2019-nCoV was highly similar throughout the genome to RaTG13 (Fig. 1c), with an overall genome sequence identity of 96.2%…The close phylogenetic relationship to RaTG13 provides evidence that 2019-nCoV may have originated in bats.*"

Lau et al. (2020). Possible bat origin of Severe Acute Respiratory Syndrome Coronavirus 2. *Emerging Infectious Diseases* 26(7), 1542–1547.

"*Potential recombination sites were identified around the RBD region, suggesting that SARS-CoV-2 might be a recombinant virus, with its genome backbone evolved from Yunnan bat virus–like SARSr-CoVs and its RBD region acquired from pangolin virus–like SARSr-CoVs.*"

The paper, which addressed a highly debated topic during the COVID-19 pandemic, faced scrutiny from a group of bat conservation scientists who requested responses from the authors to address their concerns post-publication. For clarity, line breaks are added in the user prompt and the model's response. The bold font on "related" was used by the model in its response to the reviewer's criticism.

Second, we found several discrepancies in our fact-checking of the model's response. The statement in the first point of rebuttal, i.e., "The closest known relative of SARS-CoV-2 is RaTG13, a bat coronavirus isolated from *Rhinolophus affinis* in China, with 96.1% genome identity" is no longer correct. The closest known relatives of SARS-CoV-2 are now three viruses found in bats in Laos (Mallapaty 2021; Temmam et al. 2022). These viruses, named BANAL-52, BANAL-103, and BANAL-236, were each more than 95% identical to SARS-CoV-2. One of the viruses, BANAL-52, is 96.8% identical to SARS-CoV-2, making it more genetically similar to SARS-CoV-2 than RaTG13, previously the closest relative with 96.1% genome identity. Notably, all three viruses have individual sections that are more similar to sections of SARS-CoV-2 than seen in any other viruses, and their receptor binding domains could attach to the angiotensin-converting enzyme 2 (ACE2) receptor on human cells as efficiently as some early variants of SARS-CoV-2. These findings

support the hypothesis that SARS-CoV-2 has a natural origin, with bats being a probable reservoir. The obsolete statement in the response by new Bing, however, may be due to the knowledge cut-off of the GPT-4 model, i.e., in September 2021 (OpenAI 2023a), which coincides with the first reports of this discovery on 17 September 2021 and onwards (Temmam et al. 2021; Rettner 2021; Mallapaty 2021). This example shows that the integration of GPT-4 with the Microsoft Bing search engine does *not* enable the AI large language model to access more recent search data, a reasonable assumption that is widely held but proven otherwise by this example.

Also, the statement in the model's response "…SARS-CoV-2 related coronaviruses have also been detected in bats from different regions and species, such as RmYN02 from *Rhinolophus malayanus* in China (Zhou et al. 2020), RacCS203 from *Rhinolophus acuminatus* in Laos (Temmam et al. 2022), and RshSTT200 from *Rhinolophus shameli* in Cambodia (Latkin et al. 2021)" contains an error. In fact, RacCS203 was discovered in bats (*Rhinolophus acuminatus*) found in a cave in Thailand, not Laos. The reference cited in this statement "(Temmam et al. 2022)" is also incorrect, which should be corrected as "(Wacharapluesadee et al. 2021)". It is noteworthy that this discovery was reported by Wacharapluesadee et al. (2021) and Briggs (2021) in February 2021, *before* the knowledge cut-off of the GPT-4 model. In addition, the study cited in the statement "RmYN02 from *Rhinolophus malayanus* in China" is incorrect. The correct reference should be "A Novel Bat Coronavirus Closely Related to SARS-CoV-2 Contains Natural Insertions at the S1/S2 Cleavage Site of the Spike Protein", which was published by Zhou et al. (2020). Lastly, the bibliometric information of the second reference given by the model contains errors. Specifically, the authors' names and the article number listed in the second reference should be corrected as "Delaune D, Hul V, Karlsson EA et al (2021) A novel SARS-CoV-2 related coronavirus in bats from Cambodia. Nat Commun 12:6563". The reference cited in the statement "RshSTT200 from *Rhinolophus shameli* in Cambodia" should thus be corrected as "(Delaune et al. 2021)."

We then challenged the GPT-enabled new Bing with a more specific comment from the bat conservation scientists. Since we already analyzed this reference in a previous example (Table 15), we initiated a "new conversation" with new Bing to avoid possible interferences from our chat history. The AI large language model made an impressive rebuttal by presenting arguments at multiple levels against the critical remarks on the targeted sentence that we wrote in the paper (Table 17). Apart from the last point, i.e., before the summary statement, which obviously veered too far from the topic being debated, other points are valid, relevant, and support the argument in the quoted sentence. Note how the AI large language model organized its arguments and evidence in a *progressive* manner to formulate a *strong*, *neutral-sounding* rebuttal. We did fact-checking in the referenced studies and confirmed that the facts presented in the response are accurate, which are marked by underlines in the table.

We then gave new Bing a relatively straightforward task by asking it to provide a rebuttal over the word choices in our writing. Being an AI large language model, we thought this should not be a challenge for the ChatGPT-enabled new Bing. The responses by new Bing proved our optimism (Table 18). It is evident that by using AI large language models, one could effortlessly find examples of writing in the public

Table 17 Preparing a rebuttal to critical comments on their writing, using the ChatGPT-enabled new Bing

Prompt: You are an author of a review paper. In the paper, you wrote this sentence: "Davy et al. (2018) found large quantities of coronavirus RNAs in the intestines of hibernating Little brown bats (*Myotis lucifugus*) co-infected with the white-nose syndrome, which confirmed that responses of extracellular co-infections had led to amplified coronavirus replication and increased viral shedding from bats."

The reviewer strongly disagrees with the sentence you wrote. The reviewer stated: "This is incorrect because a PCR gives the amount of RNA, but this doesn't always correlate with the quantity of live virus. Can't associate detection of RNA with live virus."

The full text of the reference article (Davy et al. 2018) is opened in the web browser. I want you to read the entire article, analyze the reviewer's comments, and provide a detailed rebuttal to the reviewer.

Response: Thank you for your comments. I appreciate your feedback and the opportunity to clarify my statement. Here is my rebuttal to your criticism:

- The sentence I wrote is based on the findings of Davy et al. (2018), which is the current web page context. The authors used RT-qPCR to quantify the levels of coronavirus RNA in the ileum of bats co-infected with the white-nose syndrome fungus and bats infected only with the virus. They found that co-infected bats had 60-fold more coronavirus RNA on average than virus-infected bats, and that this correlated with the severity of fungal pathology.
- The authors also performed a transcriptomic analysis on bat intestines and found that co-infection altered the expression of genes related to the MAPK pathway and cytokine responses, which may affect the innate immune response and viral replication. Furthermore, they measured the levels of antibodies against the coronavirus nucleocapsid protein in bat plasma and found that co-infected bats had higher antibody levels than virus-infected bats, suggesting an adaptive immune response to increased viral replication.
- Therefore, the authors provided multiple lines of evidence to support their conclusion that co-infection with the fungus increased coronavirus replication and shedding in bats. They did not simply rely on PCR results, but also used RNA sequencing, ELISA, and histopathology to corroborate their findings.
- Moreover, while it is true that PCR does not directly measure live virus, it is widely used as a proxy for viral load and infectivity in many studies of viral diseases. For example, SARS-CoV-2-RNA viremia has been shown to be a robust marker of critical illness in COVID-19 patients. The detection of viral RNA in blood or other tissues may indicate active viral replication, viral persistence, or spill-over from infected cells elsewhere. In any case, it reflects the presence and activity of the virus in the host.
- Additionally, some viruses that encode information with RNA rather than DNA tend to have a higher rate of mutations, which allows them to evade the host immune system and adapt to new environments. Some viruses also insert their own RNA into the host cell's genome, which can lead to cancer or other diseases. Therefore, detecting viral RNA is not only indicative of viral infection, but also of potential viral evolution and pathogenesis.
- In summary, I believe that my sentence accurately reflects the main findings and implications of Davy et al. (2018), which is a well-designed and comprehensive study of co-infection dynamics in bats. I respectfully disagree with your claim that PCR results cannot be associated with live virus, as there is ample evidence to suggest otherwise. I hope this rebuttal addresses your concerns and clarifies my position. Thank you for your time and attention.

Line breaks are added to the user prompt for clarity, which was automatically removed by new Bing after the user inputs the prompt in the conservation window. The "More Creative" mode was used for generating the response. Apart from the point discussing RNA detection and virus evolution, which obviously veers too far from the topic being discussed, the evidence and arguments presented in the response are relevant and valid. All underlined texts in the response were validated by the authors doing fact-checking in the referenced studies. The phrase "web page context" is a term that is used by new Bing to describe the document being analyzed, which is opened in the Microsoft Edge Dev browser by the user.

domain. Note that the third point listed in the model's response is irrelevant to the reviewer's critical comment. Again, the model missed the target, as in a previous example (Table 16). Other than that, the facts and links provided by the model are correct as validated by the authors after doing fact-checking.

Below are our main findings in this subsection.

- We tested ChatGPT-enabled new Bing for preparing rebuttals to criticisms of our research publications. These comments are critical and content-specific, and they require a good understanding of the evidence and arguments presented in these publications as well as general knowledge of the broader literature context, presenting a challenging task for the AI large language model.
- Overall, the model showed impressive capabilities of engaging in meticulous discussions and debates over scientific issues with human experts by finding supporting evidence and organizing its arguments in a compelling manner. However, in one case the model evaded the issue raised in the comment and failed to address the criticism directly. Also, the model's response contains outdated information, an inherent issue of large language models due to their knowledge cut-off date. These examples highlight the need for human oversight when using ChatGPT to address specific comments on any scientific work.
- Subject to the policies of journal publishers and research institutions, these tools can assist researchers in preparing responses to critical comments from peer reviewers.

6.2 Answering Questions

We also tested the capabilities of the AI large language model to help authors prepare answers to reviewers' questions on specific contents in a recent research paper (Table 19). For this purpose, we provided our revised manuscript and the supplementary data, i.e., the published version, to the model, without giving any hints or our actual responses to the model. In this example, the reviewer asked two questions in one set of comments. The model addressed the questions separately by numbering them and then provided a point-by-point response to each of the questions (Table 19). To improve the clarity of the response, the model used both bold and italicized fonts to differentiate the reviewer's comments and excerpted text from the manuscript from its response, a technique that we teach our students when preparing responses to reviewers' comments. It is noteworthy that all of these were done by the model without having such requirements written in the user prompt.

In this example, we asked the model to use only the information in the referenced manuscript to respond to the reviewer's comments, as an attempt to restrict the model from using information from other sources, e.g., web sources or secondary information which are more likely to result in errors in the model's response. First, the answers provided by the model in "Response 1" are accurate and succinct, as we validated in the full text and the supplementary data of the paper (Liu et al. 2023). Second, note the upfront appreciation shown to the reviewer before addressing the comments, i.e. "*Thank you for your valuable comments and suggestions on our*

Table 18 Preparing a rebuttal to criticism over the choice of wording in a sentence in our earlier review paper (He et al. 2021a) using new Bing

Prompt: Read this article, understand the context, and provide a detailed rebuttal to the Reviewer's comment below, using supporting evidence in existing studies and web searching results if needed.

The Reviewer's comment refers to a sentence in the article, "…bats are known to be a pervasive nocturnal species living in cities and city suburbs and have the best hearing of land mammals."

Reviewer's comment: This is wrong in at least two ways. Using "a" implies reference to a single species and pervasive is generally used in a negative way to indicate someone or something unwanted and harmful.

Response: Thank you for your comment. I disagree with your criticism of the sentence "…bats are known to be a pervasive nocturnal species living in cities and city suburbs and have the best hearing of land mammals." Here is my rebuttal:

- First, using "a" does not imply reference to a single species, but rather to a group or order of mammals. Bats belong to the order Chiroptera, which is the second-most speciose group of mammals, after rodents. The approximately 925 species of living bats make up around 20% of all known living mammal species. Therefore, it is appropriate to use "a" to refer to bats as a whole, just as one would use "a" to refer to rodents or primates as groups of mammals.
- Second, using "pervasive" does not necessarily imply a negative connotation, but rather a widespread or abundant occurrence. Bats are found throughout the world in tropical and temperate habitats, except for polar regions and some isolated islands. They are also highly adaptable and can roost in various natural and artificial structures, such as caves, trees, bridges, and buildings. Some bat species are frequently sighted in urban and suburban areas, where they forage on insects attracted by artificial lighting or feed on fruits and nectar from plants. Therefore, it is accurate to describe bats as pervasive in their distribution and occurrence.
- Third, using "the best hearing" is supported by scientific evidence that bats have remarkable auditory abilities that enable them to echolocate and communicate in complex ways. Echolocation is a sophisticated system of emitting high-frequency sounds and analyzing the returning echoes to navigate and locate prey in the dark. Bats can produce sounds ranging from 14 to 200 kHz, well beyond the human hearing range of 20–20 kHz. Bats can also modulate the frequency, intensity, and duration of their sounds to suit different situations and environments. Moreover, bats use vocalizations for social communication, such as mating calls, territorial defense, group coordination, and mother–offspring recognition. Therefore, it is reasonable to state that bats have the best hearing of land mammals, considering their exceptional auditory skills and diversity.

In conclusion, I believe that the sentence in question is correct and justified by existing studies and web search results. I hope this clarifies my point of view and addresses your concerns.

Prompt: Provide some examples where "pervasive" is used as a neutral or positive connotation in scientific publications. Please extract the original wording from the literature and provide the link.

Response: Here are some examples of the term "pervasive" used in a neutral or positive context in scientific publications:

- In a paper titled "Positive Stereotypes Are Pervasive and Powerful", the authors discuss how positive stereotypes are pervasive and powerful in society. They define positive stereotypes as subjectively favorable beliefs about members of social groups that directly or indirectly connote or confer domain-specific advantage, favorability, or superiority based on category membership.[1]
- In a publication titled "Pervasive Healthcare Computing as a Scientific Care Discipline for Patients", the term "pervasive" is used to describe a type of healthcare service that can be provided to any needy person at any time and probably anywhere, ignoring time and some other constraints including proper coverage and very high-quality real-time services.[2]
- The term "pervasive" is also used in the context of computing, as seen in the title of the journal "Pervasive and Mobile Computing". This journal publishes high-quality scientific articles covering all aspects of pervasive computing and communications.[3]

References:

[1] Positive Stereotypes Are Pervasive and Powerful. © The Author(s) 2015. https://depts.washington.edu/sibl/Publications/Czopp,%20Kay,%20&%20Cheryan,%202015.pdf

[2] Pervasive Healthcare Computing as a Scientific Care Discipline for…. https://link.springer.com/chapter/10.1007/978-3-030-77746-3_4

[3] Pervasive and Mobile Computing—ISSN 1574-1192—Elsevier. https://www.elsevier.com/journals/pervasive-and-mobile-computing/1574-1192/guide-for-authors

In this example, the GPT-enabled new Bing leveraged its advantage as a large language model and provided facts and arguments that countered the opinion in the comment. The "More Creative" mode in new Bing was used for generating the response, and the "More Precise" mode was used for finding examples of writing. Line breaks are added in the user prompt for clarity.

Table 19 Preparing responses to reviewers' questions over specific issues in the data and methods reported in a recent research paper (Liu et al. 2023) using the ChatGPT-enabled new Bing

Prompt: Read this article and provide a detailed response to the Reviewer's comments below, using the information in the article.

More QA/QC data is needed. It seems blanks were not characterized by Raman, SEM, or weighed, why not? Also why was the packaging not also characterized with Raman to confirm the polymer and match with the polymers detected on the sample?

Response: Dear Reviewer, Thank you for your valuable comments and suggestions on our manuscript. We appreciate your time and effort in reviewing our work. We have carefully addressed your comments and revised our manuscript accordingly. Below, we provide detailed responses to each of your comments, along with the changes we have made in the manuscript.

Comment 1: More QA/QC data is needed. It seems blanks were not characterized by Raman, SEM, or weighed, why not?

Response 1: We agree that QA/QC data is important for ensuring the reliability and validity of our results. In fact, we did characterize and weigh the blanks by Raman, SEM, and EDS, as described in Section 2.4 of our manuscript. However, we did not report the results of the blanks in detail in the original manuscript due to space limitations. To address your comment, we have added a paragraph in Section 3.1 to present the results of the blanks and compare them with the samples. We have also added a figure (Fig. S5) in the Supplementary Data to show the SEM images of particles identified on filter membranes from control blank samples. The revised paragraph reads as follows:

For quality control, control blanks were prepared in triplicate, inspected using the optical microscope, micro-Raman spectroscopy, and SEM-EDS to examine microplastic and other particle contaminants carried by filter membranes or introduced during our sample preparation. The SEM images of particles identified on filter membranes from control blank samples are shown in Fig. 2a and b and Fig. S5. In terms of particle abundance, there were on average 0.003 ± 0.021 mg of particles and debris collected from each blank sample (300 mL) with particle sizes generally in the lower range of 1–50 μm. Compared with the weights of particles and debris in samples from breastmilk storage bags, i.e., 0.22–0.47 mg, background contamination, i.e., particles and debris filtered from control blanks, represented less than 1.4% of those detected in breastmilk storage bag samples and was thus considered negligible (Pivokonsky et al., 2018). Analysis by micro-Raman spectroscopy did not identify microplastics in control blanks due to the fact that the particles retained on the polycarbonate membrane were extremely low in abundance and that the few particles that could be visually identified under the optical microscope and micro-Raman were generally small in size (< 2-3 μm) which exhibited strong fluorescence interference in their Raman spectra.

Comment 2: Also why was the packaging not also characterized with Raman to confirm the polymer and match with the polymers detected on the sample?

Response 2: We thank you for this insightful suggestion. We did not characterize the packaging with Raman in our original study because we relied on the product labels to identify the polymer types used in the breastmilk storage bags. However, we agree that it would be more rigorous and informative to confirm the polymer types with Raman analysis and match them with the polymers detected on the sample. Therefore, we have performed additional Raman analysis on the body and parts of each of the six types of breastmilk storage bags used in our study and compared them with the particles released from them. The results are shown in Fig. 4b and Table S2 in our revised manuscript. The revised paragraph reads as follows:

The chemical composition of the released particles were examined preliminarily by SEM-EDS and micro-Raman spectroscopy. The results are shown in Fig. 3j and 3k and Fig. 4a. Overall, the elemental composition of the particles consisted of predominantly carbon and oxygen. The particles that were identifiable by micro-Raman spectroscopy were characterized as PE, PET, and nylon-6 (Fig. 4a). Figure 4b shows the typical Raman spectra of ***the body and parts of breastmilk storage bags used in our study*** *as well as* ***the identifiable particles from them*** *for comparison purposes. Table S2 summarizes* ***the materials identified from both the six products****, namely, product labels on their packaging* ***and the results of the micro-Raman spectroscopy analyses on breastmilk bags and the particles retained on filter membranes.***

We hope that our responses have addressed your comments satisfactorily and improved our manuscript for publication. We thank you again for your constructive feedback and look forward to hearing from you.

Excerpted words shown in the model's response that contain discrepancies with the original text of the research paper are marked with underlines. The bold font in "Response 2", which is used by the AI model automatically, highlights the words that directly answer the reviewer's question in the comment. The "More Creative" mode in new Bing was used for generating this response.

manuscript. We appreciate your time and effort in reviewing our work." and the first sentence in "Response 1" that expresses agreement with the reviewer, i.e., "*We agree that QA/QC data is important for ensuring the reliability and validity of our results.*" A similar style of writing is seen in "Response 2". A polite and firm tone is often advised to our authors when preparing responses to reviewer's comments. Also, the explanation in "Response 1" for abbreviating the data in the original manuscript, i.e., "*…we did not report the results of the blanks in detail in the original manuscript due to space limitations*" is justifiable, which is often seen in authors' responses due to the word limit imposed by most scientific journals on research publications today. Overall, even if we take science out of the evaluation, this is a good example of how authors should write their responses to reviewers in an appropriate manner.

By analyzing the model's output in "Response 2", we had an interesting finding that, while there are variations (marked by underlines) between the excerpted text in the model's response and the original text in the paper, which are presumably mistakes made by the AI, the AI's writing is actually more accurate by stating "Fig. 4a" instead of "Fig. 4" as per the original writing in the paper. Readers may refer to the caption of this figure in the referenced paper for quick validation. Also, the statement by the model "*Fig. 4b shows the typical Raman spectra of the body and parts of breastmilk storage bags used in our study as well as the identifiable particles from them for comparison purposes*" contains an error, where "Fig. 4b" should be written as "*Fig. 4a and 4b*". This mistake may originate from an erroneous statement in the paper where the authors stated that "*Fig. 4b shows the typical Raman spectra of the identifiable particles from the six products*." In fact, "Fig. 4b" shows the typical Raman spectra of the *body and parts* of breastmilk storage bags used in the referenced study, as per the description in the figure caption in the paper. This erroneous statement may have confused the model which mashed up the information in the caption of "Fig. 4b" and the authors' erroneous description in the discussion. By changing "Fig. 4b" to "Fig. 4a and 4b" in the model's response, the sentence becomes correct and actually makes a stronger argument for answering the reviewer's question, while it also improves the coherence of writing in this section in the original paper.

After examining the model's responses, we concluded that the model did a very good job pointing out the information that the reviewer had requested in the revised manuscript and the supplementary data. Specifically, the model showed capabilities of (i) understanding the questions raised by the reviewer in the correct context (note that there is very little background information given in the reviewer's comments, only straight-up questions); (ii) precisely locating the contents that are relevant to the reviewer's questions in the referenced manuscript and the supplementary data; and (iii) organizing the answers into a proper response with coherent writing and polished language.

Below are our main findings in this subsection.

- We tested the capability of the ChatGPT-enabled new Bing to help authors prepare detailed answers to specific questions raised by peer reviewers.
- With minimal guidance from the user, the model correctly identified the specific information from the revised manuscript and the supplementary data to construct

its answers to the reviewer's questions. Minor inaccuracies were found between the model's quotes and the original text of the research paper.

- The model addressed the reviewer's questions separately by numbering them and using formatting techniques to provide clear, point-by-point responses to each of the questions, without taking the instructions from the user. This approach is particularly needed for maintaining the clarity of the conversation with the reviewer when discussing multiple interconnected issues in a research paper that are of a complex nature.
- The model molded its response in a polite and factual manner, using expressions of appreciation and where appropriate, showing its agreement with the reviewers' comment to make its stance clear upfront. When addressing queries from the reviewer, the model maintains an open-minded and respectful tone in its writing of the entire response.

6.3 Responding to Difficult Requests

Not all reviewers' requests can be satisfactorily addressed within the given timeframe during the peer review process. In the next example, the reviewer did not ask any questions but rather pointed out the limitation of the data reported in our study (Table 20). The reviewer suggested repeating the study using advanced facilities, a task that was unachievable using the facilities in place or within the given timeframe. To respond to the reviewer's comments, the model extracted several facts from a section in the paper where the authors articulated the limitations of the study. It then re-arranged these facts to organize its arguments properly, in the order of (i) technical difficulties encountered during our particle analysis, (ii) lack of an advanced automated accessory in the instrument used in the study, i.e., automated particle analyzer, and (iii) having no access to the alternative instruments suggested by the reviewer for repeating the experiments using mass quantification, i.e., Py-GC-MS.

When facing criticisms or difficult requests from the reviewer, it is important for authors not to respond in a confronting or resentful manner. In this example, the model admitted in an upfront manner that the results presented in the study are preliminary and that a more comprehensive analysis is needed, in agreement with the reviewer's comments. It then turns to its arguments by pointing out the efforts already made by the authors before the submission of the work, in a bid to convince the reviewer that repeating these experiments without the needed facility upgrade on the micro-Raman spectroscopy would likely result in the *same* challenges, which in turn makes it unrealistic to determine the weight fraction of non-plastic fraction in the released particles, as requested by the reviewer. Then, the model admitted the usefulness of TED-GC-MS and Py-GC-MS as complementary techniques for analyzing the mass and composition of the particles but stated that authors do not have access to these instruments while it also pointed out the destructive nature of these techniques and their requirements on sample pre-treatment. From our experience, this is a common excuse used by authors because they usually face a pressing deadline, e.g., 21–30 days

Table 20 Responding to difficult requests from the reviewer using the ChatGPT-enabled new Bing

Prompt: Read this article and provide a detailed response to the Reviewer's comments below, using information in the article, references cited in the article, and web searching results if needed.

Reviewer's comment: The results are preliminary, as Raman spectroscopy was only applied to a selection of particles, it is unclear how large the non-plastic fraction among the weighed particles is. The study should be repeated if possible, using either μ-Raman spectroscopy for detailed particle analysis (e.g. with methods comparable to Schymanski et al. 2018; Weisser et al. 2021) or methods such as Py-GC-MS for mass quantification.

Response: Thank you for your comment and suggestion. We agree that our results are preliminary and that more comprehensive analysis of the particles released from breastmilk storage bags is needed. However, we would like to point out some limitations and challenges that we faced when applying micro-Raman spectroscopy to our samples.

First, as we mentioned in our article, we encountered strong fluorescence interference in the Raman spectra of many particles, which hindered their identification. This is a common problem when analyzing microplastics using Raman spectroscopy, especially for small particles (< 20 μm) (Schymanski et al. 2018; Weisser et al. 2021). We tried different methods to reduce the fluorescence background, such as changing the laser wavelength, adjusting the instrument settings, applying photobleaching, and testing different samples and sample replicates. However, none of these methods were effective enough to obtain clear Raman spectra for all particles.

Second, we did not have access to an automated particle analyzer that could locate, measure, and identify particles on the filter membranes in a fast and reliable way. Therefore, we had to manually select and target approximately 80 particles on each membrane sample for Raman spectral acquisition and matching. This process was time-consuming and prone to errors and bias. Moreover, due to the limited number of particles analyzed, we could not estimate the percentage of microplastic particles in the total particle mass.

Third, we did not have access to other techniques that could complement or confirm the results of Raman spectroscopy, such as Py-GC-MS or TED-GC-MS. These techniques can provide information on the polymer mass and composition of particles by analyzing their thermal decomposition products (Li et al. 2021; Rivers et al. 2019). However, they are also destructive and require sample preparation and optimization.

Therefore, we suggest that further studies on the release of microplastics from breastmilk storage bags should use a combination of Raman spectroscopy and other techniques that can overcome its limitations and challenges. For example, using a micro-Raman spectroscope with an automated particle analyzer and a reference library for spectral matching could improve the speed and accuracy of particle identification (Schymanski et al. 2018; Weisser et al. 2021). Additionally, using PyGC-MS or TED-GC-MS could provide quantitative data on the polymer mass and composition of particles (Li et al. 2021; Rivers et al. 2019). We hope that our study can provide a useful reference for future research on this topic.

The "More Creative" mode in new Bing was used for generating the response. The facts in the model's response are validated by the original authors of this paper (Liu et al. 2023). Note that the reference citation "(Weisser et al. 2021)" is incorrect, and it should be "(Schymanski et al. 2021)" (Weisser is a co-author of this publication). The citation "(Rivers et al. 2019)" is also incorrect. The two erroneous citations are underlined in the model's response. Errors in bibliographic information are common in responses generated by ChatGPT or the ChatGPT-enabled new Bing.

to submit their response along with their revised manuscript, while obtaining access to new facilities could take days to weeks, depending on the circumstances. It is noteworthy that nowhere in the revised manuscript or the supplementary data do the authors make such statements or hints, i.e., the authors having no access to these instruments. This means that the AI came up with such excuses on its own while also pointing out the shortcomings of these techniques (and the extra work required on sample preparation) to respond to the challenging task suggested by the reviewer. Then, in the summary of the response, the AI reiterated its position by agreeing with the reviewer's suggestions but stressed that advanced facilities would be needed for carrying out such work.

As we have consistently seen in these examples, the tactics used by the AI large language model have shown intelligence levels that are on par with human scientists when addressing comments from peer reviewers. In fact, these are among the most impressive results that we have obtained in this project. With the exception of one example where the AI essentially evaded the point of criticism in focus (see Table 16), we found that the responses generated by new Bing were generally well written and organized, with coherent writing and convincing arguments that were pertinent to reviewer's comments, i.e., not veering too far from the reviewer's questions or criticism, an issue that is sometimes seen in AI-generated responses when performing critical evaluations of scientific publications (Han et al. 2023). Although there are some errors in the model's responses, these do not undermine its impressive capabilities of making solid rebuttals and appropriate responses to criticisms, challenging questions, and difficult requests in research publications. Clearly, as a widely used representative of current AI large language models, the ChatGPT-enabled new Bing is competent for assisting researchers with such tasks.

Below are our main findings in this subsection.

- We tested the ChatGPT-enabled new Bing for preparing responses to a reviewer's request that was *not* fulfilled in our revised manuscript. This scenario tested the large language model for writing appropriate responses to unfulfilled requests due to resource and time constraints, a common but difficult situation faced by many researchers when submitting their work to journals for peer review.
- With minimal guidance from the user, the model used tactics to handle the difficult request by admitting upfront that the results presented were preliminary, aligning its stance with the reviewer's comments. Then, it structured the response by acknowledging the limitation of the methods used in the study, explaining the technical difficulties previously encountered by the authors, i.e., attempting various methods to reduce fluorescence interference in Raman spectroscopy, and admitting the authors' lack of access to the more advanced and alternative instruments suggested by the reviewer.
- The model cited non-existent references in its response, highlighting the issue again that bibliographic information supplied by large language models often contains errors.

6.4 A Reminder for Journal Editors and Reviewers

Editors and reviewers should be aware of the possibility of being swamped by AI-generated responses. After all, fact-checking is time-consuming and tedious, especially on lengthy responses generated by large language models, and there is a good chance that those who exploit large language models or other AI tools to do their own work will *not* spend time reading and checking the model's response carefully. Since most scientific journals do not publish authors' responses along with accepted manuscripts, currently there is no feasible way to subject these responses to public scrutiny, e.g., after the acceptance of a manuscript. This puts more pressure on the editors and reviewers who must develop a good sense of judgment on the *true* source of these responses, in addition to checking the scientific rigor of the arguments and evidence presented in the authors' responses. To conclude, the lack of public oversight on the communication between authors and reviewers may put the conventional peer review process at risk of being jeopardized by AI-generated content.

7 Advanced Language Editing

7.1 Correcting Issues in Writing

Communicating with peer researchers and general readers for public engagement is an essential part of scientific research. Yet, many researchers have struggled with the task of writing scientific manuscripts that, once published, report their scientific findings and theories to a global audience. Indeed, after one has accomplished the exploratory work, there is still a major task ahead. As Dr. Kevin Plaxco once wrote in his paper, "*The Art of Writing Science*", scientific writing can be seen as an advanced skill, and can be stressful for early-career researchers (Plaxco 2010). Since English is the dominant language used by the global research community, those who are non-English speakers may find it even more difficult to organize evidence and arguments into clear, well-structured manuscripts that are understandable by peer researchers and a wider readership. While grammar checkers, online translators, and paraphrasing tools, e.g., Thesaurus in Microsoft Word, are available, none of these tools works like a professional language editor, that is, to correct or rewrite one's writing and explain *how* applying such changes improves the clarity and coherence of writing. To add to that, none of these tools can understand the tone and the context, and make suggestions that align with the author's goals of writing the text.

Large language models like ChatGPT have been trained to process natural language in a proficient and context-relevant manner. In addition, these models have been trained with vast amounts of human-generated texts, including those in specialized domains. The ability to answer user's queries makes them an ideal tool for language editing of scientific manuscripts. In the example below, we first asked ChatGPT to correct grammatical errors, odd phrases, and awkward sentences in a

rough draft written by one of the authors (Table 21). The draft is a rough translation of a curated paragraph written in Chinese that was *not* provided to ChatGPT as a reference for doing language editing, i.e., the rough translation written in English was all that we provided to ChatGPT. The model listed the issues in the rough English translation and suggested over a dozen corrections. The revised draft is a much-improved version in terms of clarity and coherence compared with the rough draft. Note how ChatGPT correctly understood the rough writing and rewrote the phrases and sentences in native English.

7.2 Rewriting Entire Text

We then asked ChatGPT to rewrite the entire text rather than correcting the issues in the rough draft. The text rewritten by the model is also shown in Table 21. As one could tell by reading the two versions side by side, the result was an even more improved version that is more polished in wording and contains no mistakes as we confirmed by fact-checking against the original text, i.e., the curated paragraph written in Chinese. It should be noted that, although it is easier to ask ChatGPT to rewrite the text and it may indeed generate better-quality texts with the "rewrite" prompt, authors could actually *learn* better by asking ChatGPT to identify issues in their own writing with detailed explanations on the corrections suggested by the model. After all, the goal is to improve the authors' own writing skills while completing advanced language editing on the writing sample. In this vein, these models may be particularly useful for training students or those who write in English as a second language in professional settings.

To conclude, language editing is one of the most beneficial uses that we have found throughout our testing. In essence, ChatGPT provides researchers with *universal* access to advanced-level language editing that can be used by authors throughout the writing, editing, or proofing of manuscripts before submission, even for experienced and native English-speaking scientists. The model identifies, corrects, and *explains* grammatical errors, odd phrases, awkward sentences, and incoherence in writing in a given context at proficiency levels that are contending to professional human language editors. Researchers who write in English as their second language or those who struggle with phrasing and sentence structures can benefit instantly from the model's ability to rephrase texts and bring them in line with native English standards. The following prompt can be used for this purpose: "*Read the text below. Identify any mistakes, inconsistencies, grammatical errors, odd phrases, awkward sentences, and other issues in the text. Suggest corrections and explain each correction in detail.*" Further, ChatGPT can help authors make transitions more seamless between sentences or when prompted, rewrite the entire text to improve its clarity and coherence.

Table 21 Revising and rewriting a rough draft of English translation using ChatGPT with the GPT-3.5 model

Prompt: Identify all grammatical errors, odd phrases, and awkward sentences in the paragraph below. List them in bullet points and suggest corrections to them all.

China has the largest consumption need of products in the world, including fresh farm products, meats and fish products. From 1997 to 2017, the number of countries exporting food to China increased from 108 to 170, cover three quarters countries and areas in the world. Due to food, especially fresh food, has rot characteristic, the import food consumption needs of China increased rapidly promotes cold chain logistics industries developed rapidly and food industry running in the world. The scale of Chinese cold chain logistics market is over 380 billion yuan, the cold storage has 180 million cubic meters. The cold cars are 287,000. The cold chain transport is a complex processing. The fresh or foods will contact workers constantly when transporting. Transporting or storing equipment are ongoing low temperatures environments, due to keep food needs. At present, we have known SARS, SARS-CoV-2 and other coronavirus survival prolonged significantly in the low temperature. If the food is processing or transporting, the staff who talk, cough or sneeze and droplet discharges will cause the virus pollutes the food or the surface of products. And the products contact the cold chain transport and store equipment which infected with a virus, are maybe always lead to far-distance flow and spread of virus.

Response:

- "China has the largest consumption need of products in the world" is an awkward phrase. Suggest: "China has the world's largest demand for products, including fresh farm products, meats, and fish."
- "cover three quarters countries and areas in the world" is grammatically incorrect. Suggest: "covering three-quarters of the countries and areas in the world."
- "Due to food, especially fresh food, has rot characteristic" is an awkward phrase. Suggest: "As fresh food is prone to spoilage."
- "increased rapidly promotes" is grammatically incorrect. Suggest: "prompted a rapid increase in."
- "food industry running in the world" is an awkward phrase. Suggest: "global food industry."
- "The scale of Chinese cold chain logistics market is over 380 billion yuan" should be "The Chinese cold chain logistics market has a scale of over 380 billion yuan."
- "The fresh or foods will contact workers constantly when transporting" is an awkward phrase. Suggest: "Fresh foods come into constant contact with workers during transport."
- "Transporting or storing equipment are ongoing low temperatures environments, due to keep food needs" is an awkward phrase. Suggest: "Transport and storage equipment are maintained at low temperatures to meet food safety requirements."
- "At present, we have known SARS, SARS-CoV-2 and other coronavirus survival prolonged significantly in the low temperature" is grammatically incorrect. Suggest: "Currently, we know that SARS, SARS-CoV-2, and other coronaviruses can survive for prolonged periods at low temperatures."
- "If the food is processing or transporting" is an awkward phrase. Suggest: "During food processing and transportation."
- "the staff who talk, cough or sneeze and droplet discharges will cause the virus pollutes the food or the surface of products" is an awkward phrase. Suggest: "If staff talk, cough, or sneeze, droplet discharges can contaminate the food or surfaces of products with the virus."
- "And the products contact the cold chain transport and store equipment which infected with a virus, are maybe always lead to far-distance flow and spread of virus" is an awkward and unclear phrase. Suggest: "If products come into contact with cold chain transport and storage equipment infected with the virus, it may lead to the long-distance spread of the virus."

(continued)

Table 21 (continued)

Prompt: go on
Response: • "cold cars are 287,000" should be "there are 287,000 refrigerated trucks." • "The cold chain transport is a complex processing" is an awkward phrase. Suggest: "Cold chain transport is a complex process." • "are maybe always lead to far-distance flow and spread of virus" is grammatically incorrect. Suggest: "can lead to the long-distance flow and spread of the virus." • The paragraph could be divided into smaller sentences for better readability. Revised paragraph: China has the world's largest demand for products, including fresh farm products, meats, and fish. From 1997 to 2017, the number of countries exporting food to China increased from 108 to 170, covering three-quarters of the countries and areas in the world. As fresh food is prone to spoilage, this prompted a rapid increase in demand for imported food, particularly fresh food, and led to the rapid development of the cold chain logistics industry. The Chinese cold chain logistics market has a scale of over 380 billion yuan, with 180 million cubic meters of cold storage and 287,000 refrigerated trucks. Cold chain transport is a complex process, with fresh foods coming into constant contact with workers during transport. Transport and storage equipment are maintained at low temperatures to meet food safety requirements. Currently, we know that SARS, SARS-CoV-2, and other coronaviruses can survive for prolonged periods at low temperatures. During food processing and transportation, if staff talk, cough, or sneeze, droplet discharges can contaminate the food or surfaces of products with the virus. If products come into contact with cold chain transport and storage equipment infected with the virus, it can lead to the long-distance flow and spread of the virus.
Prompt: Instead of trying to correcting mistakes, rewrite this paragraph in native English for publishing in a scientific journal.
Response: China has the world's largest demand for a variety of products, including fresh farm products, meats, and fish. The number of countries exporting food to China has increased from 108 to 170 from 1997 to 2017, covering three-quarters of the countries and areas in the world. The rapid increase in demand for imported food, especially fresh food, has led to the development of the cold chain logistics industry. The scale of the Chinese cold chain logistics market is over 380 billion yuan, with cold storage facilities covering 180 million cubic meters and a fleet of 287,000 refrigerated trucks. Cold chain transport is a complex process, during which fresh food comes into constant contact with workers. To maintain food safety standards, transport and storage equipment are kept at low temperatures. Additionally, it is known that SARS, SARS-CoV-2, and other coronaviruses can survive for extended periods at low temperatures, which highlights the importance of maintaining strict hygiene practices during food processing and transportation. Contamination of food or surfaces of products with the virus can occur when staff talk, cough, or sneeze, leading to the potential long-distance spread of the virus via contact with cold chain transport and storage equipment.

The original text was a curated paragraph written in Chinese (**Text S1**). Upon the human translation to English, however, the writing became very rough with many issues, making the text confusing and difficult to read. This is a common problem in scientific writing by non-English speaking students and early-career researchers. Note the use of a simple prompt ("go on") to ask the model to provide the rest of its answers due to the word limit on its text output in a single response. The GPT-3.5 model is free to use with no account registration needed. Unlike new Bing, no options of "answering modes" are available in the user interface of ChatGPT.

7.3 Removing Language Barriers

The capability of ChatGPT provides researchers with a versatile tool to freely transcend language barriers. In addition to language editing, the model can directly translate curated texts, e.g., in Chinese into native English while preserving the coherence of the writing and ensuring the accuracy of the translated information, including the style of writing and the use of specialized terminologies. One immediate benefit to researchers, therefore, is that they could potentially use ChatGPT to effortlessly share their professional works written in non-English languages with an international audience. While English is unquestionably the dominant language in scientific works and publications, there exists a universe of scientific, engineering, medical, mathematical, and social science works written in French, German, Italian, Spanish, Russian, Chinese, Japanese, Korean, and many other non-English languages, which are yet to be explored and researched by international peers.

In the example below, ChatGPT transcends the Chinese-English language barrier by rewriting curated texts in Chinese into native English (Table 22). The original text was excerpted from the white paper "*Responding to climate change: China's policies and actions*" from the State Council Information Office of the People's Republic of China (SCIO PRC 2021). The official English translation of the white paper, which is available from ChinaDaily (2021), is shown in the table to check the accuracy of the translation by the model. Note that there is an error in the output by ChatGPT, where "*increased non-fossil fuel energy capacity*" should be "*adding a new target, that is, the total installed capacity of non-fossil fuels*". This is because the original writing in Chinese could be interpreted both ways. This sort of writing can confuse readers and, in this case, the AI large language model. This can be clarified by reading the official English translation of the white paper and further, comparing these "new targets" with those in an earlier version of the white paper, namely, the "*2015 Nationally Determined Contributions by the Government of the People's Republic of China*" (Gov. PRC 2015). Other than this glitch, which may be difficult to spot by readers, the output by ChatGPT is flawless in terms of adhering to the original text, being grammatically correct, and preserving coherence in writing.

We further compared the translated text by ChatGPT with several mainstream translators that are widely used by scientific researchers (**Text S2**). Based on our evaluation, the AI large language model (in this case, GPT-4) outperformed the mainstream translators including Google Translate, Microsoft Translator, DeepL, and CNKI Academic Translator. While all these translators showed good accuracy, the output by the GPT-4 model showed the best consistencies with the original text and overall, the most accurate use of specialized terminologies in the translated text. With few exceptions, users can assume that the model is already familiar with the terminologies in the field of study prompted by the user.

It is also noteworthy that users can write prompts in their native language in their conservations with GPTs and new Bing. Despite the fact that over 90% of the texts used to train its predecessor, GPT-3, were in English language (Brown et al. 2020), we have consistently validated in our testing that for the same prompt, the responses obtained by inputting prompts in Chinese or French are of similar quality with those obtained using prompts written in English. This again shows the strong ability of

Table 22 Rewriting a section of a Chinese government white paper in English using ChatGPT (GPT-3.5 model)

Prompt: Rewrite this paragraph in native English for general readers. [Original text omitted.]
Response: In 2020, China announced new targets for its contribution to reducing global carbon emissions. The country aims to reach peak carbon dioxide emissions by 2030 and achieve carbon neutrality by 2060. By 2030, China plans to reduce its carbon dioxide emissions per unit of GDP by more than 65% compared to 2005 levels, with non-fossil fuel energy consumption accounting for approximately 25% of total energy consumption. Additionally, China aims to increase its forest stock by 6 billion cubic meters compared to 2005 levels and increase its wind and solar power capacity to more than 1.2 billion kilowatts. These targets are even more ambitious than the ones China set in 2015, with a shorter timeline, greater reductions in carbon intensity, a five-percentage-point increase in non-fossil fuel energy consumption, increased non-fossil fuel energy capacity, and an additional 1.5 billion cubic meters of forest stock. In 2021, China announced that it would no longer build coal-fired power projects overseas, demonstrating its practical action to address climate change
Official translation: In 2020, China announced new NDC targets and measures. China aims to • peak carbon dioxide emissions before 2030 and achieve carbon neutrality before 2060 • lower its carbon intensity by over 65 percent by 2030 from the 2005 level • increase the share of non-fossil fuels in primary energy consumption to around 25 percent by 2030 • increase the forest stock volume by 6 billion cubic meters by 2030 from the 2005 level • bring its total installed capacity of wind and solar power to over 1.2 billion kW by 2030 Compared with the objectives set in 2015, the new targets are more ambitious in timeframe. They involve a steeper cut in carbon intensity, an increase of another five percentage points in the share of non-fossil fuels in primary energy consumption, a new target for installed capacity of non-fossil fuels, an additional forest stock of 1.5 billion cubic meters, and a clear announcement to aim for carbon neutrality before 2060. China has announced in 2021 a decision to stop building new coal-fired power projects overseas, demonstrating its concrete actions in response to climate change

The original text is omitted in the table. The underlined text in the model's response shows inaccurate translation. The correct translation is underlined in the official translation, which is also shown in the table as a reference.

GPT to transcend the language barrier. The Bing Image Generator, which has the DALL-E 2 deep learning models built-in and previously only supported requests in English, recently added native support for non-English prompts in over 100 different languages (Microsoft 2023c). This means that users do not need to translate their prompts into English before inputting them in GPTs, new Bing, or the Bing Image Generator, which is time-consuming and may impair the accuracy of the prompts. This is an important feature for researchers who write in English as a second language, which we have tested repeatedly in GPTs and new Bing with consistently positive findings.

The following points present our key findings in this section.

- We used ChatGPT to correct and rewrite rough English translations, which significantly improved the grammar, clarity, and flow of the text. The consistent high-quality outcomes achieved with ChatGPT on language editing tasks are on par with the work of professional language editors for scientific writing.
- On a broader subject, the model provides researchers with a valuable tool to share their work with a global audience by removing language barriers. For instance, ChatGPT can translate and rewrite non-English text into English, or the other way around, while maintaining its accuracy and coherence.

- The model outperformed mainstream translation tools in terms of consistency with the original text and accuracy in the use of specialized terminologies. These are essential requirements for the language editing of scientific writing.
- While the model can rewrite text with great accuracy, its detailed explanations on corrections allow users to understand and learn from their mistakes, thereby helping authors improve their writing skills. This is unlike other grammar checkers, online translators, and paraphrasing tools, which generally do the work without explanation.

7.4 GPT-3.5 versus GPT-4 Model

It is noteworthy that in our evaluation, the latest model (GPT-4)—despite the widespread perception of its advantages over the previous version (GPT-3.5)—does not always outperform GPT-3.5 on language editing tasks. In many cases, GPT-3.5 generated better responses, with sentences and phrases that are more precise and natural-sounding. This coincides with the fact that GRE writing, English Language, and English Literature were among the few tests that GPT-3.5 maintained an edge over the latest GPT-4 model on the test results (OpenAI 2023a). However, when it comes to tasks like adapting research papers to different styles of writing or designing experiments, as we will discuss further in the following sections, the advantages of the GPT-4 model become evident. For language editing tasks, we recommended users run both models and compare the outputs to leverage the strength of both models. This is also why we intersperse the outputs by the GPT-3.5 and GPT-4 models in the following sections.

8 Crafting Article Titles

For scientists, well-crafted titles capture the essence of their research and motivate readers' engagement. Good titles, such as "***Silent Spring***" by Rachel Carson and "***COVID-19 Trashed the Recycling Dream***" by Joe Brock, captured the essence of the authors' writing and have made a lasting impact on readers. Below are well-crafted book titles on climate change, water resources, biodiversity, and zoonoses. These are some of the most influential writings on environmental and public health-related issues, and the list goes on.

- "***Rising****: Dispatches from the New American Shore*", by Rush (2019)
- "***Where the Water Goes****: Life and Death Along the Colorado River*", by Owen (2017)
- "***We Are All Whalers****: The Plight of Whales and Our Responsibility*", by Moore (2021)
- "***Spillover****: Animal Infections and the Next Human Pandemic*", by Quammen (2013)

Researchers who have experience in preparing scientific manuscripts would know that crafting a concise, informative, and impactful title for a research paper can be a challenging task. Leveraging the advanced language editing capabilities of ChatGPT, it is possible to craft good article titles for manuscripts in writing. In this example, we asked the ChatGPT-enabled new Bing to generate ten possible titles from an outline of a rough manuscript still being written by the authors. For this, we first converted the outline to PDF format and opened it in the Microsoft Edge Dev browser to feed the information to the model. The response from new Bing is shown in Table 23. Clearly, the majority of the suggested article titles are content-relevant and plausible-sounding. After reviewing all the candidates, the authors quickly reached a consensus that the title "*The environmental paradox of cryptocurrency: How a decentralized technology contributes to centralized energy consumption and pollution*" could be happily adopted with minor editing. It is noteworthy that some of the other suggestions also look interesting, although at times they veer too far from the outline that we provided, yielding overly creative titles. Nonetheless, even if the user does not wish to adopt any of its proposals, he or she can still find plenty of inspiration and usable phrases from the model's suggestions. It is impressive that, using AI large language models, one could have many quality suggestions for a traditionally daunting task in just tens of seconds.

Table 23 Creating titles for a research article

Prompt: Read the PDF on this page. Suggest 10 titles for this article for publishing in a scientific journal.
Response: I read the PDF on this page. Based on the web page context and the web search results, here are some possible titles for this article for publishing in a scientific journal: • The Environmental Impacts of Cryptocurrency Mining: A Review of Methods, Assumptions and Data Sources • Cryptocurrency Mining and Sustainability: A Critical Assessment of Energy Consumption and Carbon Emissions Estimates • How Green is Cryptocurrency? A Systematic Review of Energy and Carbon Footprint Studies • Energy and Emissions from Cryptocurrency Mining: A Survey of Methods, Data and Challenges • Cryptocurrency Mining and Climate Change: An Evaluation of Environmental Performance and Policy Implications • The Carbon Footprint of Cryptocurrency: A Comprehensive Analysis of Energy Use and Emissions in the Mining Sector • Cryptocurrency Mining and the Environment: A Review of Current Practices, Problems and Prospects • Assessing the Environmental Sustainability of Cryptocurrency Mining: A Review of Methodological Approaches and Data Quality • The Environmental Costs and Benefits of Cryptocurrency Mining: A Review of Energy Consumption, Carbon Emissions and Policy Responses • Cryptocurrency Mining and Environmental Justice: A Review of Social and Ethical Issues in the Energy and Emissions Debate

(continued)

Table 23 (continued)

Prompt (continuing the chat session): Can you make it a bit more lively? Give me another 10 answers.
Response: Sure, I can try to make it a bit more lively. Here are another 10 possible titles for this article for publishing in a scientific journal • The Dark Side of Cryptocurrency: How Mining Wastes Energy and Pollutes the Planet • Cryptocurrency Mining: A Threat to Sustainability or an Opportunity for Innovation? • How Much Energy Does Cryptocurrency Mining Consume? A Review of Methods and Data Sources • The Climate Impact of Cryptocurrency Mining: How Bad Is It and What Can We Do About It? • Cryptocurrency Mining and the Future of Energy: A Review of Environmental Challenges and Solutions • How to Mine Cryptocurrency Without Destroying the Environment: A Review of Best Practices and Policy Recommendations • Cryptocurrency Mining: A Hidden Source of Greenhouse Gas Emissions • The Environmental Paradox of Cryptocurrency: How a Decentralized Technology Contributes to Centralized Energy Consumption and Pollution • Cryptocurrency Mining and Renewable Energy: A Review of Current Trends and Future Prospects • The Environmental Dilemma of Cryptocurrency: How to Balance Innovation and Sustainability in the Mining Sector

The outline of a manuscript draft was provided to the ChatGPT-enabled new Bing as the content for creating titles. The "More Creative" mode in new Bing was used for generating the response. This example shows how the ChatGPT-enabled new Bing can be both creative and analytical in crafting headline-styled titles to attract readers while adhering to the contents being analyzed.

9 Design of Experimental Studies

9.1 Designing Experiments and Refining Methods

A long-desired functionality by scientific researchers is to have the AI design and perform experiments to liberate themselves from some of the most laborious and time-consuming work in research projects. The latter requires sophisticated robotics and automation, which is beyond the scope of our current discussion, although attempts have long been made with renewed efforts recently reported (Boiko et al. 2023). But what if we just ask ChatGPT to design the study, for instance, by giving us step-by-step instructions on the experiments that are required for a particular research project? After all, large language models like ChatGPT have been trained on a vast body of existing literature, so one would expect that they are fairly "knowledgeable" on the methods and techniques that have been established in various science disciplines. For instance, some experimental work in environmental analytical chemistry, e.g., sample pretreatment and instrumental analyses, have well-established protocols, and recent studies on emerging contaminants have refined these methods for measuring trace organic compounds, engineered nanomaterials, and micro/nanoplastic debris present in various types of environmental matrices. Therefore, this functionality may be particularly useful for those who just

started working in a research domain and are about to test their ideas in the lab, without having to spend a lot of time learning and comparing the methods reported in the existing studies.

For this test, we asked ChatGPT to design experiments to study the presence of microplastics in bottled water (Table 24). Note the simple, intuitive prompt given to the model. As a first attempt, we specified a "clean" sample matrix, i.e., bottled water as opposed to soil, food, or natural water, to simplify the experimental design that was to be completed by the model. This is a routine experiment in the research domain of microplastics in water samples, where hundreds of research publications, including open-access papers, have been published. As shown in the table, the AI large language model (GPT-4) did a good job by providing step-by-step procedures for conducting the experiments, with the names of chemical reagents, consumables, and instruments required for carrying out the work. There are many commendable suggestions in the initial response by the model on the experimental and overall study design, including most notably, the strategy of selecting a representative pool of commercial products (steps 1a and 1b), quality assurance steps and precautions (steps 1c, 2a, 2c, and 2d), and the methods described in steps 2b and 3a for capturing and visually inspecting microplastic particles, respectively. Data recording and processing, as described in steps 3c, 4a, and 4b, and the investigation of the source of microplastics in the last step of the study design are rational and often required for this type of study.

Anyone who closely follows these steps without questioning and careful validation may find themselves stuck at some point and are unable to achieve the described outcomes. Indeed, a closer look at the initial response by ChatGPT revealed several gaps in the steps of work designed by the model. First, it is difficult, if at all possible, to use infrared or Raman spectroscopy to identify the type of plastic for particles captured through steps 2b and 2c, such as those with small sizes, e.g., 100 μm or less. For that, one would have to employ more advanced infrared or Raman spectroscopic techniques, such as Fourier-transform infrared micro-spectroscopy or Raman micro-spectroscopy (μ-Raman) to complete these tasks. In fact, the use of μ-Raman for identifying the type of plastic for microplastic particles has been well documented in research publications before September 2021, i.e., the knowledge cutoff date of the GPT-4 model (OpenAI 2023a). Second, since the particles are captured on a microfiltration membrane which is typically made of polymer materials, one would have to separate the infrared or Raman signals collected from microplastic particles from those emitted from the membrane background. This could be done by sputtering a very thin layer of noble metal, e.g., gold or platinum with 70–100 nm in thickness on the membrane using a sputter coater or selecting an inorganic membrane, e.g., silver, glass microfiber, or anodic aluminum oxide membrane which has no interfering infrared or Raman absorption (Cytiva 2024; Sterlitech 2024). Third, since the characterization relies on visual inspection and spectroscopic analysis, it is crucial to inspect the surface of the filter membrane, which should be smooth on a micrometer scale and free of particulate matter or other contaminants before use. Last, for quality assurance and quality control purposes, all samples should be prepared in replicates. In addition, method blanks should be prepared to check for contaminants introduced by ultrapure water, labware, and filter membranes to establish the baseline before

Table 24 Experimental study designed by ChatGPT (GPT-4 model)

Prompt: Design an experiment for studying microplastics in bottled water

Response: To study microplastics in bottled water, an experiment will be designed to sample, extract, identify, and quantify microplastics from various brands and types of bottled water. The goal is to determine the prevalence and potential sources of microplastics in bottled water to better understand the associated risks and inform consumers and manufacturers.

Experiment: Microplastics in bottled water

1. Sample collection and preparation:
   a. Select multiple brands of bottled water, considering different materials (PET, glass), production processes (spring, purified, mineral), and storage conditions (cool, room temperature, exposed to sunlight).
   b. Acquire at least three samples per brand to account for potential variability.
   c. Store the samples in a controlled environment until analysis to prevent contamination.
2. Extraction of microplastics:
   a. In a clean laboratory setting, shake each bottle gently for 30 s to distribute potential microplastics evenly.
   b. Pour the water from each bottle through a filtration system with a 1 μm pore size membrane to capture microplastics.
   c. Rinse the bottles with ultrapure water and filter the rinse water to capture any remaining microplastics.
   d. Use forceps or tweezers to carefully transfer the filtration membranes to clean Petri dishes or glass slides, and cover them to prevent contamination.
3. Identification and quantification of microplastics:
   a. Analyze the particles on the filtration membranes using an optical microscope or scanning electron microscope (SEM) to determine their size, shape, and count.
   b. Use Fourier-transform infrared spectroscopy (FTIR) or Raman spectroscopy to identify the type of plastic.
   c. Record the number of particles, their size distribution, and plastic type for each bottled water sample.
4. Data analysis:
   a. Calculate the average concentration of microplastics (particles/L) for each brand and type of bottled water.
   b. Perform statistical analyses to determine whether there are significant differences between brands, materials, production processes, and storage conditions.
   c. Investigate potential sources of microplastics, such as the production process, packaging materials, or environmental factors.

Upon completion of this experiment, the results will provide valuable information about the presence of microplastics in bottled water and may inform improvements in production, packaging, and storage practices to minimize microplastic contamination.

While there are gaps in the initial response, the experimental and overall study design have many commendable suggestions. No reference was given to ChatGPT in the user prompt for designing these experiments.

actual samples are prepared. Note that these requirements are not apparent to beginners and that they are again hidden under the polished writing in the response by the AI language model. However, anyone who has first-hand experience doing these experiments would be able to spot at least some of these gaps in the model's initial response.

Since ChatGPT automatically analyzes its previous conservation with the user in a chat session, it is possible to request further information on its earlier answers or request alternatives (OpenAI 2023a). We continued our chat session with GPT-4 and asked it to elaborate on certain points and provide alternatives for the methods described in its initial response. As shown in Tables 25 and 26, the answers to our follow-up questions are contextually relevant and mostly accurate, where the GPT-4 model provided answers that we could largely relate to our experiences when we actually conducted these experiments, with few exceptions that need to be pointed out. The first deviation was that in our experiments, we did not opt for silver or aluminum oxide membranes for cost-saving purposes and the fact that track-etched polycarbonate and polyethylene terephthalate membranes had smooth surfaces that were superior to silver and aluminum oxide membranes (It4ip 2024; Sterlitech 2024). Second, when asked further on choices of polymeric membranes and coating techniques to use, GPT-4 provided answers that we could not immediately validate or relate to our experience, i.e., using the atomic layer deposition technique to coat 5–20 nm of metal oxides on polymeric membranes (Table 26). In reality, we used a fine coater to deposit a thicker layer of platinum on track-etched polycarbonate membranes to block the Raman signals of the polycarbonate filter membrane (Liu et al. 2023). That being said, these instructions are of sufficient quality and detail as an initial step-by-step guide for one to start experimenting and researching to validate and refine the methods.

9.2 Reality Check

Reality check is the last and perhaps the most important step in questioning the experimental and study design by AI large language models. For this purpose, we asked GPT-4 to provide a list of relevant studies with information on their methods used for extracting microplastics from samples and the analytical instruments used in those studies, as real-world references to check the experiments designed by the model. Further, to put our results in the literature context, we requested a compilation of data from the reference studies, i.e., the quantities and materials of the microplastics found in samples, and the types and sources of samples used in these studies.

Overall, the references provided by GPT-4 are of high quality and relevance for doing the "reality check". For instance, six of the references recommended by GPT-4, i.e., articles numbered 1–6 in Table 27, were read and hand-selected by one of the authors (J. Han) during the fall semester of 2019 for students who were interested in investigating microplastics in soda drinks and other bottled liquids. By cross-comparing their methods of sample preparation and the types and settings of analytical instruments used in these studies, one could develop consensuses and identify the key variations in their methods before adopting or turning down the experiments designed by GPT-4, i.e., completing the final step of "reality check" for tasks suggested by the model. Note that the model did not quite fulfill our request

Table 25 Elaboration on certain points in the initial response by the model on the experimental design. We continued our chat session in the previous table using the GPT-4 model.

Prompt (continuing the chat session): In your answers above, can you elaborate on the following points:
- Point b, under "3. Identification and quantification of microplastics": How should I use Fourier-transform infrared spectroscopy (FTIR) or Raman spectroscopy to identify the type of plastic for microplastic particles? Note these particles may be very tiny, for example a few micrometers in size. It would be impossible to do this analysis using the regular Fourier-transform infrared spectroscopy (FTIR) or Raman spectroscopy. I need to know the exact types of instruments and their settings for identifying the type of material of microplastic particles.
- Point b, under "2. Extraction of microplastics": How could I separate the infrared or Raman spectra of the microplastic particles being analyzed from the filter membrane background, if the filter membrane is also made of a plastic material, which is usually the case with microfiltration membranes.
- There are many types of filter membranes in the recommended pore size range. What should be my first choices, and why?
- What are the possible sources of contaminants in this experiment? Give me some detailed instructions on how to prepare blank samples to validate the method of sample preparation.

Response: I apologize for the lack of clarity in my previous response. I'll address your concerns and provide more information on the specific points you mentioned.

1. Analyzing microplastic particles using FTIR or Raman spectroscopy:
For analyzing microplastics in the micrometer size range, you'll need specialized equipment, such as micro-FTIR or micro-Raman spectroscopy. These instruments are equipped with a microscope that allows for the analysis of small particles with high spatial resolution.
For micro-FTIR spectroscopy, you can use a system with an attenuated total reflectance (ATR) accessory, which has a small crystal probe that can be placed directly on the microplastic particle for analysis. Micro-Raman spectroscopy uses a laser to focus on the particle, making it suitable for analyzing small particles without additional accessories.
In both cases, the instrument settings should be optimized to obtain high-quality spectra, considering factors like laser power, integration time, and the number of accumulations. The acquired spectra can be compared to reference spectra of known plastics in a database to identify the type of plastic.

2. Separating microplastic spectra from the filter membrane background:
To avoid interference from the filter membrane in the analysis, you can choose a membrane made of a material with a distinct FTIR or Raman signature compared to common microplastics. Alternatively, you can use an aluminum oxide (Al_2O_3) or a silver membrane filter, which do not produce overlapping spectra with plastic materials.

3. Filter membrane choices:
For this experiment, the first choice of filter membranes should be those made from aluminum oxide (Al_2O_3) or silver. These materials will not produce overlapping FTIR or Raman spectra with plastic materials, minimizing the risk of interference during analysis. If these materials are not available or suitable, opt for a membrane made from a material with distinct FTIR or Raman signatures compared to common microplastics.

4. Possible sources of contaminants and preparing blank samples:
Contamination can occur during sampling, storage, and analysis, with potential sources including airborne microplastics, lab equipment, and clothing. To minimize contamination, follow these steps:
   a. Work in a clean laboratory environment with laminar flow hoods, if available.
   b. Wear disposable gloves, lab coats, and hairnets to prevent contamination from clothing.
   c. Clean all equipment and surfaces with ultrapure water or a suitable solvent before use.
   d. Store samples and filtration membranes in clean, sealed containers to prevent airborne contamination.

To validate the sample preparation method and account for potential contamination, prepare blank samples by following these steps:
1. Use ultrapure water as a blank sample to mimic the experimental conditions without the presence of bottled water.
2. Follow the same extraction and filtration process as for the actual samples, including filtration through the same type of membrane.
3. Analyze the blank samples using micro-FTIR or micro-Raman spectroscopy to identify any potential contaminants and background signals.
4. Subtract any contaminants or background signals identified in the blank samples from the actual sample data to obtain accurate results.

By following these guidelines and recommendations, you can minimize contamination and accurately identify and quantify microplastics in bottled water samples.

Table 26 Asking ChatGPT for alternatives to the specialty membranes suggested in the model's initial response on the experimental design. We further continued our chat session in the previous tables using the GPT-4 model.

Prompt (continuing the chat session): I have two further questions which I need you to answer in detail.
- Silver and aluminum oxide membranes are too expensive. I need cheaper alternatives. Can I coat a thin layer of metal or metal oxide on the polymeric microfiltration membrane to block the background infrared or Raman signals from the membrane? If this is doable, give me detailed instructions on how to do this. I need to know the steps, the type of instrument, the coating technique, the coating material, and the thickness of the coating.
- In general, what type of polymeric microfiltration membranes have the most flat and smooth surfaces?

Response: I understand your concerns about the cost of silver and aluminum oxide membranes. Coating a thin layer of metal or metal oxide on a polymeric microfiltration membrane could potentially block the background infrared or Raman signals from the membrane. However, it is crucial to ensure that the coating does not affect the pore size or the filtration efficiency of the membrane.

1. Coating technique: Atomic Layer Deposition (ALD)
ALD is a thin-film deposition technique that allows for precise control over the thickness and uniformity of the coating. It is a gas-phase process that utilizes sequential, self-limiting surface reactions to deposit material on a substrate. This method can be used to deposit thin layers of metal or metal oxide on polymeric microfiltration membranes

2. Coating material: Al_2O_3 or other metal oxide
Aluminum oxide (Al_2O_3) is a suitable coating material due to its compatibility with the ALD process and its distinct infrared and Raman signatures compared to common microplastics. Other metal oxides, such as titanium dioxide (TiO_2) or zirconium dioxide (ZrO_2), can also be used if they provide the desired blocking effect without interfering with the analysis of microplastics.

3. Thickness of the coating
The thickness of the coating should be sufficient to block the background signals from the polymeric membrane without altering the pore size or filtration efficiency. A thickness of 5–20 nm is recommended as a starting point, and optimization may be required based on the specific membrane and coating material used.

4. Steps for ALD coating:
   a. Obtain a suitable ALD system, such as a commercial or custom-built system.
   b. Clean the polymeric microfiltration membranes with ultrapure water or a suitable solvent to remove contaminants and ensure proper adhesion of the coating.
   c. Load the cleaned membranes into the ALD chamber.
   d. Introduce the precursor gases for the chosen coating material (e.g., trimethylaluminum and water for Al_2O_3) sequentially into the chamber, allowing them to react with the membrane surface.
   e. Purge the chamber with an inert gas (e.g., nitrogen) to remove unreacted precursors and byproducts.
   f. Repeat steps d–e to deposit the desired thickness of the coating, monitoring the thickness with an in situ or ex situ technique, such as spectroscopic ellipsometry.
   g. Remove the coated membranes from the ALD chamber and store them in clean, sealed containers until use.

Regarding your second question, polymeric microfiltration membranes made from polyvinylidene fluoride (PVDF) or polytetrafluoroethylene (PTFE) generally have flat and smooth surfaces. PVDF membranes, in particular, are known for their low background signals in FTIR and Raman spectroscopy, making them a suitable option for the analysis of microplastics. However, it is essential to verify the compatibility of the chosen membrane material with the coating process and ensure that it does not interfere with the analysis of microplastics.

by providing "*an extensive range of studies, e.g., 30 publications or more*" as specified in the user prompt. This is a common limitation with ChatGPT and its derived application, e.g., new Bing. When a prompt necessitates long text outputs, the model either provides a shorter version of its answer, e.g., Table 27 or gives one particular example with a set of instructions upfront on how to complete the task so the user can do it himself or herself. The latter was seen in a previous example (Table 7), where the model responded to our request to explain *each* figure in a research paper by giving the results of one figure analyzed by the model with a set of instructions on how to complete this kind of task.

To conclude, our examples in Tables 24, 25, 26 and 27 show the typical workflow of sending a series of progressive queries to ChatGPT to obtain a meticulous experimental design, following the order of: (i) a general task list with step-by-step procedures by giving a simple prompt to the model; (ii) requesting more details or clarifications on the specific tasks designed by the model; and (iii) requesting a list of the most relevant references for doing the "reality check". To add to these, one may ask the model for a procurement list of chemical reagents, materials, consumables, instruments, personal protective equipment, quality assurance and quality control (QA/QC) steps, and safety requirements to plan the work.

9.3 Cautionary Note

There are several points that we need to point out to safeguard the use of AI large language models for designing experiments. First, at the end of the user's queries, one should always ask the model for a list of reference studies that are relevant to the experiment design provided in the chat session. When the model does its job right, as we have seen in this example (Table 27), authors can instantly get a high-quality list of relevant studies selected by the model. Other AI-enabled search tools, such as Perplexity Ask, can also be used for this purpose (Perplexity 2024). The user should then spend time looking up their method sections and compare those with the tasks designed by the model. This validation step is very important for users to ensure that the methods suggested by the model do not contradict the consensus on methods reported in the research domain, i.e., as a reality check to relate AI-generated answers with real-world experiences by other scientists. Second, one must be vigilant that some experiments involve safety hazards where the operator must strictly follow lab safety protocols, such as reading the Material Safety Data Sheet and wearing appropriate personal protection equipment *before* handling hazardous chemical reagents, materials, instruments, or where needed, obtaining approval from the lab manager or other lab safety personnel before conducting the work. AI large language models cannot take accountability for any negligence and whatever dire consequences it may cause in real-world settings. Lastly and most importantly, one should only use these for legitimate purposes such as scientific investigation or educational purposes, not making illicit drugs, toxic substances, explosives, or biohazards that can potentially threaten environmental and public health or any activities prohibited by law.

Table 27 Providing reference studies on the experimental design queried in our chat session

Prompt (continuing the chat session): here is my last question:
If others have done these experiments before, I need you to compile the following information from existing studies. I need you to compile these from an extensive range of studies, e.g., 30 publications or more. The more, the better. If the answer is too long, you may break it into multiple parts.
- their reported results, including the quantities and materials of microplastics found in their samples
- the types and sources of their samples
- the experimental methods used for extracting microplastics from their samples
- the analytical instruments used for quantifying and identifying the type of materials of the microplastics
- name of the corresponding author, year of publication, and link to each study

Response: As an AI language model, I am unable to search the web or access new publications beyond my knowledge cutoff date in September 2021. However, I can provide a summary of some relevant studies on microplastics in bottled water published before my knowledge cutoff date. Keep in mind that this list is not exhaustive and may not include 30 or more publications.

1. Mason, S. A., Welch, V. G., & Neratko, J. (2018). Synthetic polymer contamination in bottled water. Frontiers in Chemistry, 6. https://doi.org/10.3389/fchem.2018.00407
   - Quantities: 93% of the samples tested contained microplastics, with an average of 325 particles/L.
   - Materials: Predominantly polypropylene (PP), polyethylene (PE), and PET.
   - Sample types and sources: 259 bottled water samples from 11 brands and 9 countries.
   - Extraction method: Filtration using a 2.5 μm pore size filter followed by Nile Red staining.
   - Analytical instruments: FTIR spectroscopy and Raman spectroscopy.
2. Oßmann, B. E., et al. (2018). Small-sized microplastics and pigmented particles in bottled mineral water. Water Research, 141, 307–316. https://doi.org/10.1016/j.watres.2018.05.027
   - Quantities: An average of 242 particles/L of microplastics and 5052 particles/L of pigmented particles.
   - Materials: Polyethylene terephthalate (PET), polypropylene (PP), polyethylene (PE), and others.
   - Sample types and sources: 32 bottled mineral water samples from 14 German brands.
   - Extraction method: Filtration using a 0.2 μm pore size silver membrane filter.
   - Analytical instruments: Micro-Raman spectroscopy.
3. Schymanski, D., Goldbeck, C., Humpf, H. U., & Fürst, P. (2018). Analysis of microplastics in water by micro-Raman spectroscopy: Release of plastic particles from different packaging into mineral water. Water Research, 129, 154–162. https://doi.org/10.1016/j.watres.2017.11.011
   - Quantities: Ranged from 2649 particles/L to 6292 particles/L depending on the packaging type.
   - Materials: Mainly PET and PP.
   - Sample types and sources: Mineral water stored in single-use PET bottles, reusable PET bottles, and glass bottles.
   - Extraction method: Filtration using a 0.2 μm pore size aluminum oxide membrane filter.
   - Analytical instruments: Micro-Raman spectroscopy.
4. Koelmans, A. A., et al. (2019). Microplastics in freshwaters and drinking water: Critical review and assessment of data quality. Water Research, 155, 410–422. https://doi.org/10.1016/j.watres.2019.02.054
   - This study is a critical review of microplastics in freshwaters and drinking water, assessing data quality and providing guidance for future research. The authors discuss various extraction methods, analytical instruments, and sample types/sources.

These four studies offer a starting point for understanding the methodologies and findings of research on microplastics in bottled water. For a comprehensive review of 30 or more publications, I recommend consulting a literature review or database search, as my access to publications is limited to those available before September 2021.

(continued)

Table 27 (continued)

Prompt (continuing the chat session): go on
Response: Here are a few more studies on microplastics in bottled water: 5. Welle, F., & Franz, R. (2018). Microplastic in bottled natural mineral water—literature review and considerations on exposure and risk assessment. Food Additives & Contaminants: Part A, 35(12), 2482–2492. https://doi.org/10.1080/19440049.2018.1533407 • This study is a literature review on microplastics in bottled natural mineral water. The authors discuss the sources, analytical methods, and potential risks associated with microplastics in bottled water. 6. Liebezeit, G., & Liebezeit, E. (2014). Synthetic particles as contaminants in German beers. Food Additives & Contaminants: Part A, 31(9), 1574–1578. https://doi.org/10.1080/19440049.2014.945099 • Although this study focuses on beer, the methodologies for extraction and analysis of microplastics can be applied to bottled water as well. • Quantities: Ranged from 2 to 79 particles/L depending on the brand. • Materials: Mainly PET and PE. • Sample types and sources: 24 German beer brands. • Extraction method: Filtration using a 0.45 μm pore size cellulose nitrate membrane filter. • Analytical instruments: FTIR spectroscopy. 7. Zuccarello, P., et al. (2019). A simple and rapid method for the isolation of microplastics from gastrointestinal tracts of fishes. Environmental Pollution, 250, 148–155. https://doi.org/10.1016/j.envpol.2019.03.060 • Although this study focuses on fish, the extraction methods presented may be adapted for analyzing microplastics in bottled water. • Extraction method: Density separation using a saturated NaCl solution followed by filtration. • Analytical instruments: FTIR spectroscopy. As I mentioned earlier, my knowledge is limited to publications available before September 2021. To find more recent and extensive literature on this topic, I recommend searching scientific databases such as Web of Science, Scopus, or Google Scholar. Conducting a comprehensive review requires a systematic search of the literature to ensure that all relevant publications are included.

We continued the chat session in the previous tables using the GPT-4 model. The studies listed by GPT-4, all of which were published before the knowledge cut-off date of the GPT-4 model, can indeed help users understand the literature context of the study designed by the model. Readers may refer to the Appendix for the results of our fact-checking on the model's response (**Text S3**).

The following points present our key findings within this section:

- Designing experiments for research projects is one of the most productive uses of large language models to advance innovations in science and technology.
- For illustrative purposes, we used ChatGPT (GPT-4 model) to design a set of experiments for studying microplastics in bottled water. With additional clarifications required by the user, the model provided a meticulous design with step-by-step instructions on the experiments from sample collection and debris extraction to the identification and quantification of microplastics.
- In the experimental design, the model incorporated quality assurance and control procedures, including the use of sample replicates and method blanks, to ensure the reliability of data acquired from the study.
- The examples showed the iterative process of refining experimental designs through continuous interaction with the model by asking for clarifications and

alternatives. This is often needed given the limited text output by the model in a single response and the meticulous steps and considerations generally required for designing experiments in research projects. Users should carefully scrutinize the responses from the model using common sense and be aware of resource constraints and safety requirements in real-world laboratory settings.

- Before conducting the work, it is crucial for users to validate the experiments designed by the model against existing studies conducted in real-world laboratory settings, i.e., reality check. For this purpose, users may request a list of the most relevant studies from the model after having extensive discussions with the model on the experimental design. As usual, users must check the bibliographic details of the references and validate the methods and findings in the reference studies provided by the model.
- This approach is particularly beneficial for novice researchers or those who are new to a specific field, for whom the model's experimental design offers a start for planning the work.

10 Design of Public Surveys

10.1 Creating a Survey Questionnaire from Scratch

Designing meticulous and thoughtful survey questionnaires is an essential task for environmental scientists. Well-crafted survey questionnaires are indispensable tools for gathering valuable data from diverse population groups, whose feedbacks provide critical insights into their opinions and behaviors related to the natural and living environment. These are needed for making informed decisions, developing effective mitigation strategies, and shaping new policies to address emerging risks and challenges. Indeed, many emerging threats in the environmental and public health domain are closely linked to human behavior, especially when abrupt changes occur on a large or global scale.

In the following example, we asked ChatGPT to create a survey questionnaire to investigate whether mask-wearing during COVID-19 has changed the behavior of cosmetic product use among U.S. consumers. Overall, we found that GPT-4 is well-versed in crafting survey questionnaires on consumer behaviors. With few requirements in the user prompt, the model generated a full-bodied questionnaire with questions and answers that not only fulfilled the user's request but went in-depth on the topic being discussed, e.g., "general questions" listed at the end of the questionnaire (Table 28).

Table 28 Creating a survey questionnaire on mask-wearing and cosmetic product uses during COVID-19. The GPT-4 model was used for generating the response.

Prompt: You are a Research Scientist specialized in consumer product safety and public health.

Design an online survey to find out whether mask-wearing during COVID-19 has changed the behavior of using cosmetic products by consumers, such as their frequency of use, choice of product types, and pattern of use.

Aim for 1000 survey respondents in the United States with different demographic groups, with an emphasis on population groups who are more likely to be affected.

Design a survey questionnaire with 30 questions.

Response:
Survey on the Impact of Mask-Wearing on Cosmetic Product Usage During COVID-19

Introduction:
Thank you for participating in this survey. We are studying the changes in consumer behaviors related to cosmetic products during the COVID-19 pandemic, specifically in relation to mask-wearing. Your responses will be kept confidential and will be used solely for research purposes.

Demographics:
1. Age:
   - Below 18
   - 18–24
   - 25–34
   - 35–44
   - 45–54
   - 55 and above
2. Gender:
   - Male
   - Female
   - Non-binary
   - Prefer to specify: _____
   - Prefer not to say
3. Ethnicity:
   - Caucasian
   - African American
   - Hispanic/Latino
   - Asian
   - Native American
   - Other: _____
4. Occupation:
   - Healthcare
   - Education
   - Retail
   - Hospitality
   - Student
   - Other: _____
5. Which of the following best describes the area you live in?
   - Urban
   - Suburban
   - Rural

Mask-Wearing Behavior:
1. How often do you wear a mask?
   - Always when outside
   - Often, but not always
   - Only in crowded places
   - Rarely
   - Never
2. If you wear a mask, what type do you commonly wear?
   - Cloth mask
   - Surgical mask
   - N95/KN95 mask
   - Other: _____

(continued)

Table 28 (continued)

Cosmetic Product Usage Pre-COVID:

1. Prior to COVID-19, how often did you use cosmetic products?
   - Daily
   - Several times a week
   - Weekly
   - Monthly
   - Rarely/never
2. Which of the following cosmetic products did you use most frequently before COVID-19? (Select all that apply)
   - Foundation/concealer
   - Lipstick/lip gloss
   - Eyeliner/mascara
   - Blush/bronzer
   - Eyeshadow
   - None

Cosmetic Product Usage Post-COVID (during mask-wearing):

1. How often do you now use cosmetic products?
   - Daily
   - Several times a week
   - Weekly
   - Monthly
   - Rarely/never
2. Which of the following cosmetic products do you use most frequently now? (Select all that apply)
   - Foundation/concealer
   - Lipstick/lip gloss
   - Eyeliner/mascara
   - Blush/bronzer
   - Eyeshadow
   - None
3. Has mask-wearing influenced your decision to use or not use certain cosmetic products?
   - Yes
   - No
   - Unsure
4. If yes, which products have you decreased in usage? (Select all that apply)
   - Foundation/concealer
   - Lipstick/lip gloss
   - Eyeliner/mascara
   - Blush/bronzer
   - Eyeshadow
   - None
5. Which products have you increased in usage? (Select all that apply)
   - Foundation/concealer
   - Lipstick/lip gloss
   - Eyeliner/mascara
   - Blush/bronzer
   - Eyeshadow
   - None

(continued)

Table 28 (continued)

Product Choices and Preferences:

1. Have you switched to any 'mask-friendly' cosmetic products (e.g., smudge-proof lipstick)?
   - Yes
   - No
2. If yes, which 'mask-friendly' products have you started using? _____
3. Has the need for skincare increased due to mask-wearing (e.g., breakouts, irritation)?
   - Yes
   - No
4. If yes, which skincare products or remedies have you incorporated due to mask usage? _____
5. Has your spending on cosmetic products increased or decreased since the pandemic started?
   - Increased
   - Decreased
   - Remained the same
6. Are you more likely to purchase cosmetic products online now as compared to before COVID-19?
   - Yes
   - No
   - Remained the same

Changes in Application and Patterns of Use:

1. Has mask-wearing influenced the areas of the face you apply makeup to?
   - Yes
   - No
2. If yes, which areas do you now focus on more? (Select all that apply)
   - Eyes
   - Brows
   - Cheeks
   - Forehead
   - No specific area
3. Has the duration of your makeup routine changed due to mask-wearing?
   - Shortened
   - Lengthened
   - Remained the same
4. Has mask-wearing influenced your decision to go "makeup-free" more often?
   - Yes
   - No

General Questions:

1. Do you believe there's a stigma associated with not wearing makeup in professional settings, even with mask-wearing?
   - Yes
   - No
   - Unsure
2. Have your makeup preferences changed more due to mask-wearing or other pandemic-related lifestyle changes (e.g., working from home)?
   - Mask-wearing
   - Other pandemic-related changes
   - Both equally
3. Would you continue with your current makeup habits post-pandemic?
   - Yes
   - No
   - Unsure
4. Has mask-wearing affected your self-perception without makeup?
   - More confident
   - Less confident
   - No change
5. Has the pandemic influenced your perspective on the importance of makeup in your daily life?
   - More important
   - Less important
   - No change
6. Any additional comments or insights you'd like to share about your makeup habits during the COVID-19 pandemic?

Thank you for your participation! Your input is invaluable in understanding consumer habits and preferences during these times.

Note the well-thought-out questions and the simple user prompt used for generating the comprehensive questionnaire. Section headings, including bold and italic fonts, were used by the model to improve the clarity of the text in the questionnaire

10.2 Refining Scopes

After reading the initial output of the model, one may decide to refine the scope of the survey by specifying additional requirements. Using ChatGPT, one can easily customize the questionnaire to target specific population groups, circumstances, or types of products. For instance, in the previous example (Table 28), we could focus on (i) different population groups, e.g., adult females aged between 18 and 60, people who have chronic respiratory diseases, or those who work in the retail or hospitality sector; (ii) public settings or industrial sectors, e.g., public transport, schools, healthcare premises; (iii) types of cosmetic products, e.g., face powders which are more likely to shed inhalable substances during mask-wearing.

In the following example, we asked ChatGPT to narrow down the scope of our survey by focusing on one population group, i.e., adult females aged between 18 and 60 and a subgroup, i.e., people with chronic respiratory diseases, and focus on their use of powdery facial cosmetics before and during COVID-19 to gain further insights (Table 29). Overall, the refined questionnaire provided a solid start for the user to design the actual survey questionnaire to be used in the study. With eight sections containing 40 questions and answers revolving around the defined topic, the user could decide to give or take the questions and options generated by the AI large language model. This makes it much easier than having to start from scratch.

Large language models are designed to respond to queries from humans. Yet, they have also shown great potential to work the other way around, that is, to ask questions of humans. As demonstrated by our examples (Tables 28 and 29), the GPT-4 model executes such tasks in a systematic and intelligent approach that is comparable with the work of human specialists. These models could be utilized by environmental scientists to formulate targeted questionnaires in areas where public involvement and local knowledge are crucial, such as tracking changes in consumer behavior, gauging community attitudes toward environmental issues, and assessing the societal impact of environmental policies. Furthermore, the ability of these models to tailor follow-up questions in real-time enables interactive dialogues with survey respondents to gather more information from individuals with enhanced scope and resolution of data collection efforts.

Below are our key findings from the two subsections.

- ChatGPT showed competence in generating detailed and in-depth survey questionnaires revolving around specific research questions in real-world settings, with minimal guidance from the user.
- The model showed a good understanding of the research question motivating the user to conduct the survey, and generated a meticulous set of well-thought-of questions to gain insights from survey respondents. The questionnaire included demographic details, behavior patterns, topic-related queries, and questions that delve deep into the behavior and preferences of survey respondents.
- Upon the user's request, the model narrowed down the focus of the questions and expanded the questions to align the designed questionnaire with the user's study goals, thereby enhancing the scope and resolution of data collection efforts.

Table 29 Refining the survey questionnaire created in the previous example by continuing our chat session with the GPT-4 model and adding requirements on the target population group and the specific types of cosmetic products

Prompt (continuing the chat session): Re-create the survey questionnaire by:
1. targeting adult females aged 18–60 years; and
2. focusing on the use of face powders or other facial cosmetic products in powdery forms while wearing face masks.

Design a more focused online survey to find out whether mask-wearing during COVID-19 has changed their behavior of using those particular types of cosmetic products, such as the frequency of use, choice of product types, pattern of use, and so on. Aim for 1000 survey respondents in the United States with different demographics. Include 30 questions in the questionnaire.

Then, create an additional mini-questionnaire with 10 questions for those who have chronic respiratory diseases and may be prone to breathing in fine particulate matter and other substances from these cosmetic products worn under the masks.

Response:
Survey on the Impact of Mask-Wearing on the Use of Face Powders and Powdery Cosmetic Products During COVID-19

Introduction:
Thank you for participating in this survey. We aim to understand the potential changes in the use of face powders and other powdery cosmetics by adult females aged 18–60 during the COVID-19 pandemic, especially in relation to mask-wearing. Your input is invaluable and will be kept confidential.

Demographics:
1. Age:
   - 18–24
   - 25–34
   - 35–44
   - 45–54
   - 55–60
2. Ethnicity:
   - Caucasian
   - African American
   - Hispanic/Latino
   - Asian
   - Native American
   - Other: _____
3. Occupation:
   - Healthcare
   - Education
   - Retail
   - Hospitality
   - Other: _____
4. Location:
   - Urban
   - Suburban
   - Rural

Mask-Wearing Behavior:
1. How frequently do you wear a mask?
   - Always when outside
   - Often, but not always
   - Only in crowded places
   - Rarely
   - Never
2. What type of mask do you most commonly wear?
   - Cloth mask
   - Surgical mask
   - N95/KN95 mask
   - Other: _____

(continued)

Table 29 (continued)

Use of Powdery Cosmetic Products Pre-COVID:

1. Prior to COVID-19, how often did you use face powders or other powdery cosmetics?
   - Daily
   - Several times a week
   - Weekly
   - Monthly
   - Rarely/never
2. Which of the following powdery cosmetics did you use most before COVID-19? (Select all that apply)
   - Loose face powder
   - Pressed face powder
   - Blush
   - Bronzer
   - Eyeshadow
   - Setting powder
   - None

Use of Powdery Cosmetic Products Post-COVID (while wearing masks):

1. How frequently do you now use face powders or other powdery cosmetics?
   - Daily
   - Several times a week
   - Weekly
   - Monthly
   - Rarely/never
2. Which powdery cosmetics have you decreased in usage due to mask-wearing? (Select all that apply)
   - Loose face powder
   - Pressed face powder
   - Blush
   - Bronzer
   - Eyeshadow
   - Setting powder
   - None
3. Have you experienced any of the following issues when wearing face powders or powdery cosmetics with a mask? (Select all that apply)
   - Smudging
   - Transfer to mask
   - Skin breakouts
   - Increased dryness
   - None

Product Choices and Preferences:

1. Have you switched to any non-powdery cosmetic products due to mask-wearing?
   - Yes
   - No
2. If yes, which non-powdery products have you started using more frequently? _____
3. Have you purchased any 'mask-proof' powdery cosmetics?
   - Yes
   - No
4. If yes, which 'mask-proof' products have you started using? _____
5. How has your spending on powdery cosmetics changed since the pandemic started?
   - Increased
   - Decreased
   - Remained the same
6. Are you more likely to purchase powdery cosmetics online now as compared to before COVID-19?
   - Yes
   - No
   - Remained the same

(continued)

Table 29 (continued)

Changes in Application and Patterns of Use:

1. Do you apply powdery cosmetics differently due to mask-wearing?
   - Yes
   - No
2. If yes, please describe briefly how your application technique has changed: _____
3. Do you feel the need to reapply powdery cosmetics more often when wearing a mask?
   - Yes
   - No

General Questions:

1. How would you rate the importance of wearing powdery cosmetics in your daily routine now, on a scale from 1 (not important) to 5 (very important)?
2. Do you believe that powdery cosmetics have a different visual effect when wearing a mask?
   - Yes
   - No
   - Unsure
3. Have you noticed increased skin sensitivity or breakouts due to wearing powdery cosmetics under a mask?
   - Yes
   - No
4. Are you more inclined to choose natural or organic powdery cosmetics since the start of the pandemic?
   - Yes
   - No
5. Have you ever felt discomfort breathing while wearing powdery cosmetics under a mask?
   - Yes
   - No
6. Do you believe there's a need for powdery cosmetics specifically designed for mask-wearers?
   - Yes
   - No
7. Are you considering permanent makeup or other cosmetic procedures as an alternative to powdery cosmetics?
   - Yes
   - No
8. How do you remove smudges or transfer from your mask caused by powdery cosmetics?
9. Would you continue your current makeup habits post-pandemic?
   - Yes
   - No
10. Any additional comments on using powdery cosmetics with face masks during COVID-19?

Mini-Questionnaire for Respondents with Chronic Respiratory Diseases:

1. Please specify your respiratory condition: _____
2. Are you concerned about inhaling particulates from powdery cosmetics while wearing a mask?
   - Yes
   - No
3. Have you experienced any breathing discomfort after applying powdery cosmetics and wearing a mask?
   - Yes
   - No
4. Have you reduced or stopped using powdery cosmetics due to these concerns?
   - Reduced
   - Stopped
   - No change
5. Do you believe powdery cosmetics worsen your respiratory symptoms?
   - Yes
   - No
   - Unsure
6. Are you more inclined to choose natural or organic powdery cosmetics due to your respiratory condition?
   - Yes
   - No
7. Do you take any special precautions when applying powdery cosmetics due to your respiratory condition?
8. Have you consulted a healthcare professional about using powdery cosmetics with your condition?
   - Yes
   - No
9. If yes, what advice or recommendations were you given?
10. Any additional insights or concerns you'd like to share about your experience with powdery cosmetics and your respiratory condition?

Further, the ability to tailor follow-up questions in real-time enables interactive dialogues with survey respondents, which can potentially provide more data and insights by adapting to the responses provided.

- Our examples illustrate how environmental scientists can use AI tools to formulate targeted questionnaires for studying community attitudes towards emerging issues including environmental and health, changes in consumer behavior, and the impact of environmental policies. By automating the initial designs of questionnaires, these models can significantly reduce the time and effort required for conducting survey-based studies.

10.3 Designing a Survey-Based Risk Assessment Study

After finalizing the questionnaire, we asked ChatGPT to design a risk assessment study based on the public survey to validate our hypothesis. The study involves literature review, experiments, and data analysis, with quality assurance and quality control (QA/QC) steps (Table 30). Building on our experience from the previous examples (see Tables 25 and 26), we used follow-up prompts and asked GPT-4 to provide further details on the designed steps of work with clarifications on a key component to be used in the experimental setup, after the model provides the general outline of the study design in the initial response.

Overall, the model provides a systematic and meticulous set of tasks for conducting the risk assessment study, with a rigorous approach that establishes a strong foundation to ensure the validity of findings. There are several commendable points in the study design, including the clarification and corrections that are provided by the GPT-4 model (Table 30).

- First, the model clearly understands the purpose of the study conceptualized by the user. This provides the foundation for the user to engage in a productive conversation with the model. This is highlighted in "II. Preliminary Steps" and "VII. Interpretation and Reporting" in the model's initial response. Understanding the purpose of the user's query is the prerequisite to productive conservation, and also the number one thing that the user is anxious about to find out when they start a new type of inquiry with large language models.
- Second, the model provides a meticulous plan for quality assurance and quality control (QA/QC) in the study outline. This section contains very high-quality information. The task plan in this section is complete and practical and requires little extra work from the user to refine the tasks. These are particularly useful for student researchers who may not have learned the painful lessons of incorporating robust QA/QC measures at the beginning of a research project to ensure the validity of their data.
- Third, the model responded to the user's further queries in a point-by-point, accurate manner, with no missed targets or vague comments in the responses. This is best exemplified in the last round of question and answer, where the user harsh criticism over some of the details in the initial design of the experimental setup provided by the model.

Table 30 Designing a survey-based risk assessment study using ChatGPT (GPT-4 model) on the use of powdery facial cosmetics while wearing face masks. The context was given to the model by continuing our chat session after the previous examples of creating survey questionnaires targeting this issue.

Prompt (continuing the chat session): Further to my request above, design an experimental study to conduct risk assessments on the inhalation of fine particulate matter and other substances from powdery facial cosmetics while wearing face masks, on a routine basis.

In the study design, list the experiments in a step-by-step manner and include products and materials, chemical reagents, apparatus, instruments, and lab consumables that are needed for conducting the experiments, with quality assurance and quality control (QA/QC) steps.

The study design must be meticulous and rigorous. Break your response into multiple parts to provide enough details.

Response:
Experimental Study Design: Risk Assessment of Inhalation of Fine Particulate Matter and Other Substances from Powdery Facial Cosmetics while Wearing Face Masks

I. Objectives:
To determine the potential health risks associated with the inhalation of particulate matter and other substances from powdery facial cosmetics during routine use with face masks.

II. Preliminary Steps:
1. Literature Review: Before the experiment, review existing literature on the types of particulates commonly found in powdery cosmetics, their average sizes, and any documented health risks.
2. Product Selection: Choose a range of commonly used powdery facial cosmetics for testing. The selection should represent a cross-section of the market in terms of price point, brand reputation, and formulation.
3. Identification of Potential Risk Compounds: From the product ingredients list and from the literature review, identify potential compounds of concern for further investigation.

III. Materials and Reagents
1. Products: At least 10 different types of powdery cosmetics, ensuring a diverse representation of brands and formulations.
2. Reagents:
- Solvents for sample extraction (e.g., methanol, isopropanol).
- Standards for identified compounds of concern.

3. Consumables:
- Sterile swabs
- Filter papers
- Syringes and needles
- Glass vials
- Disposable masks

IV. Apparatus and Instruments:
1. Inhalation Chamber: To simulate real-world inhalation of particles.
2. Particle Size Analyzer: To determine the size distribution of particulates.
3. High-Performance Liquid Chromatography (HPLC) or Gas Chromatography-Mass Spectrometry (GC-MS): To identify and quantify compounds in the samples.
4. Microscope: For visual inspection of particulates.
5. Balance: For weighing samples.

V. Experimental Design:
1. Sample Preparation:
- Apply a standardized amount of powdery cosmetic on a dummy face or skin mimic.
- Place a mask over the face/skin mimic.
- Subject the setup to simulated inhalation using the inhalation chamber to imitate human breathing patterns.

2. Particle Collection:
- Collect particles that come through the mask during inhalation into a particle trap.
- Also, swab the inner side of the mask for particulate matter and other residues.

3. Particle Size Analysis:
- Analyze the particle size distribution of the collected particles using a Particle Size Analyzer.
- Document the range and average particle sizes.

(continued)

Table 30 (continued)

4. Chemical Analysis:
- Extract collected samples with appropriate solvents for targeted compounds.
- Analyze the extracts using HPLC or GC-MS to identify and quantify compounds.

5. Data Analysis:
- Compare particle sizes to established inhalation risk sizes.
- Compare compound concentrations to established inhalation risk levels.

VI. Quality Assurance and Quality Control (QA/QC):
1. Instrument Calibration: Ensure all instruments are calibrated before use. Use appropriate standards to calibrate the HPLC or GC-MS.
2. Blanks: Run solvent blanks to ensure no contamination is present in reagents or equipment.
3. Replicates: Process samples in triplicate to ensure repeatability.
4. Spike Recovery: Add a known amount of target compounds to some samples to check the efficiency of extraction and analysis methods.
5. Interlaboratory Comparison (if possible): Share samples with another lab to compare results.
6. Data Documentation: Ensure meticulous documentation of all experimental steps, instrument settings, and results.
7. Standard Operating Procedures (SOPs): Create and adhere to SOPs for all steps to ensure consistency.

VII. Interpretation and Reporting:
1. Risk Assessment:
- Compare the particle sizes and concentrations to thresholds identified in literature and health guidelines.
- Assess the potential risks associated with the inhalation of these particles and compounds.

2. Recommendations: Based on findings, provide recommendations for consumers and industry about the use and formulation of powdery cosmetics, especially in the context of wearing face masks.

VIII. Limitations and Further Studies:
Acknowledge any limitations of the study and propose potential further studies based on the findings.

Prompt (continuing the chat session): Provide more details on the task described in the third point under "1. Sample Preparation" in part "V. Experimental Design". Tell me exactly how this should be done.

Provide your answer in detail and include the following information in your answer.
- types of equipment, instruments, and materials that are needed
- detailed, step-by-step instructions on the experimental procedure, including key experimental conditions and parameters
- a list of the most relevant studies to the methods described above.

(continued)

Table 30 (continued)

Response: Certainly. The task described in the third point under "1. Sample Preparation" in part "V. Experimental Design" is to subject the setup (a dummy face or skin mimic with applied powdery cosmetic and a mask over it) to simulated inhalation using an inhalation chamber to imitate human breathing patterns.

Detailed Procedure for Simulated Inhalation Using an Inhalation Chamber

I. Equipment, Instruments, and Materials:
1. Inhalation Chamber: A sealed environment where air can be pushed in and out to simulate inhalation.
2. Breathing Simulator: An instrument designed to mimic human breathing patterns.
3. Flow Meter: To measure the flow rate of air in and out of the chamber.
4. Thermometer and Hygrometer: To measure and control temperature and humidity inside the chamber.
5. Dummy Head or Skin Mimic: A model which can closely imitate human facial skin properties.
6. Masks: The types of masks being studied.
7. Powdery Cosmetic Products: To be applied to the dummy head or skin mimic.
8. Particle Trap: To capture particulates that pass through the mask during simulated inhalation.

II. Step-by-Step Procedure:
1. **Environment Setup:**
   - Adjust the room temperature to approximately 22 °C (standard room temperature).
   - Measure the room humidity and ensure it's around 50% to simulate typical indoor conditions.
2. **Preparation of Dummy or Skin Mimic:**
   - Clean the surface of the dummy head or skin mimic to remove any dust or contaminants.
   - Apply a standardized amount of the powdery cosmetic on the dummy face or skin mimic, ensuring even distribution.
   - Place the mask over the applied area ensuring a snug fit, as it would be worn on a real face.
3. **Chamber Setup:**
   - Place the dummy head or skin mimic inside the inhalation chamber.
   - Attach the breathing simulator to the chamber.
   - Set up the particle trap to capture any particulates that come through the mask during the simulation.
4. **Simulation Parameters:**
   - Set the breathing simulator to mimic normal human breathing: approximately 12–20 breaths per minute with a tidal volume of around 500 ml/breath.
   - Using the flow meter, monitor and ensure consistent airflow in and out of the chamber.
   - Continuously measure the temperature and humidity inside the chamber to ensure consistent conditions.
5. **Initiate Simulation:**
   - Start the breathing simulator.
   - Allow the simulation to run for a specific period, e.g., 30 min, which might represent an average time someone may wear makeup underneath a mask without adjustments.
6. **Particle Collection:**
   - After the simulation, turn off the breathing simulator.
   - Carefully remove the particle trap from the chamber and seal it to avoid loss of collected particulates.
   - These particulates will be analyzed in subsequent steps as mentioned in the original plan.

III. Relevant Studies:
While I cannot access real-time databases at this moment, below are general references to look for when researching methods for simulated inhalation and particle collection:
1. Studies on inhalation toxicology often employ simulated inhalation methodologies. Look for foundational texts or research papers on inhalation toxicology.
2. Occupational health journals: They may feature studies on mask efficiency and the impact of inhaling particles through protective equipment.
3. Aerosol science and technology research: These studies focus on the behavior, dynamics, and effects of particles in the air, which would include methodologies on simulating and measuring particle inhalation.

I'd recommend consulting databases like PubMed, Web of Science, or Google Scholar using keywords like "inhalation simulation", "particle inhalation risk", "mask efficiency studies", etc., to find specific relevant studies.

(continued)

Table 30 (continued)

Prompt (continuing the chat session): I cannot picture some of the experimental set-up and experiments in my head.

1. Why do you need an inhalation chamber? For housing the particle trap or what? The inhalable hazards are located in the confined space between the nose and mouth area and the face mask, where powdery cosmetic products may fall off, e.g., due to air movement, mask-skin contact, or adjustments made by users, e.g., for a better snug fit of the face mask after wearing it for some time, changing the face mask, or re-applying make-ups on face. All of these have little to do with the atmosphere surrounding the dummy head.
2. Could you describe the breathing simulator in detail. This is the key component of the experimental design. I need to [know] what it looks like and how it is connected to the dummy head.
3. Same for the particle trap. Is it connected with the breathing simulator with a clean hose? The particle trap should trap particles from the breathing activity of the dummy head through the face mask, not the surrounding atmosphere. Is this correct?

Response: You're absolutely right. My apologies for the oversight. Let's delve deeper into the specific components you mentioned.

1. Inhalation Chamber:
Given the nature of the hazards you're focusing on, the inhalation chamber, as initially described, might be excessive. The hazards are indeed confined between the mask and the facial area. The key is to simulate human breathing, as that's the main mechanism that would cause the inhalation of any particulate matter from the cosmetic product applied beneath the mask.

2. Breathing Simulator:
Description: The breathing simulator is an apparatus designed to replicate human respiratory actions. Here's a detailed description:

- **Main Body**: The central part of the breathing simulator is a cylindrical chamber. This is where the air is pushed in and pulled out to mimic inhalation and exhalation, respectively.
- **Diaphragm**: Inside this chamber is a flexible diaphragm (similar to a human diaphragm). This diaphragm moves up and down, which causes air to be pushed in and out of the chamber, simulating the process of breathing.
- **Control Panel**: Attached to the main body is a control panel that allows you to set breathing parameters. This can include breathing rate (breaths per minute) and tidal volume (volume of air inhaled and exhaled during a single breath).
- **Connectors**: There are hose connectors on the simulator. One end of a hose is attached to the simulator, and the other end can be connected to the dummy head.

Connection:
The breathing simulator is connected to the dummy head via a hose. The hose carries air to and from the mouth/nose region of the dummy, simulating the inhalation and exhalation process. If the dummy head is sophisticated enough, it might have nostrils and a mouth, allowing for more precise simulation.

3. Particle Trap:
Description: A particle trap is essentially a filtration or capture system designed to trap fine particles in the air passing through it.

- **Structure**: It is typically a cylindrical device with a mesh or filter at its core. The mesh's pore size is chosen based on the particle sizes you're most interested in capturing.
- **Placement**: Given the design requirements, the particle trap should be placed between the dummy head and the breathing simulator. This means it's connected in-line with the hose that runs from the dummy's mouth/nose to the breathing simulator.
- **Function**: As the breathing simulator "inhales," it pulls air from the confined space between the mask and the dummy face, and any particles or particulates dislodged from the cosmetic product are carried with this air. When this air passes through the particle trap, the particles are captured on the trap's filter.

Connection:
Yes, the particle trap would be connected to the breathing simulator using a clean hose. As the dummy "inhales" (due to the action of the breathing simulator), air from the confined space (between the mask and the face) is drawn through the particle trap, ensuring only particles from this specific space are trapped, not from the surrounding atmosphere.

Thank you for pointing out the focus areas. The revised design centers around the specific space between the mask and the face, emphasizing the most critical aspects of the potential inhalation risk.

11 Writing Research Proposals

11.1 Brainstorming

One of the things that we frequently came across in recent discussions on ChatGPT is brainstorming, specifically "to brainstorm research ideas" (Anonymous 2023b; Owens 2023). In theory, AI can perform extensive searches in response to the user's inquiries. By providing answers revolving around these inquiries, the model can potentially supplement gaps or even inspire users to generate truly creative ideas. This intuitive assumption also piqued our curiosity to test just how good it actually is as an on-demand brainstorming partner.

In the examples below, we asked for specific information on products and topics that we were not familiar with (Tables 31, 32 and 33). In what appeared to be an interactive Q&A session, the model provided answers with sufficient details that were more customized and more concise than information one could typically obtain from a well-constructed Wikipedia page. These initial insights provided users with adequate information to start researching and gathering more information. Overall, the interactive session was more productive than keyword-searching on the Internet and screening information in the search results by a human researcher. This is another good example where users could benefit from the quick and knowledgeable attributes of AI large language models.

We then asked the GPT-4 model more challenging questions in the hope that the model could give us some fresh insights that differ from the typical "market hypotheses" that we often heard about in the media or read in the literature (Tables 34 and 35). In the first example, we attempted to use ChatGPT to learn about new classes of emerging contaminants that are currently under-reported. Most of the contaminants listed by the GPT-4 model, e.g., metamaterial and space exploration-derived contaminants, are "new" and under-investigated. In the second example, we asked ChatGPT to learn about alternative hypotheses revolving around the topics that are still being debated in the scientific community. Some of the hypotheses provided by the model, if rigorously investigated and validated, can indeed shed new light on the current understanding of these important matters of wide concern. However, we did not see exciting or "novel" research ideas in the responses provided by GPT-4. While the answers may be sufficient for one to start researching, the information is mostly known to specialists who are already familiar with these research topics. This is understandable because after all, AI large language models were trained on existing text and data (Anonymous 2023b; Tregoning 2023), while scientific research often starts with bold, "out-of-the-box" thinking or requires unique insights to formulate new hypotheses. In other words, being knowledgeable and analytical, as we have seen in previous examples, does not mean the model is capable of generating truly innovative ideas as humans. In other words, having a conversation with human experts would reveal if AI has truly evolved to human-level intelligence within the expert's area of expertise.

Table 31 Using ChatGPT (GPT-4 model) to learn about the mode(s) of action by controlled-release fertilizers, a type of fertilizer that we were unfamiliar with but curious about.

Prompt: Exactly how do controlled-release fertilizers release their nutrients into soil in a slow and controlled manner? I put them under both optical microscopes and scanned electron microscopes, but I did not see any pores or slits.
Response: Controlled-release fertilizers (CRFs) release nutrients slowly into the soil through different mechanisms, depending on the type of CRF. Although you may not see pores or slits under the microscope, the release mechanisms can still be effective. There are primarily three types of controlled-release fertilizers: 1. Polymer-coated fertilizers: These fertilizers have a polymer coating around the nutrient core, which acts as a barrier. The release of nutrients occurs through diffusion, where water penetrates the coating, dissolves the nutrients, and the nutrients then diffuse out of the coating into the soil. The thickness of the coating determines the release rate of the nutrients. Thicker coatings result in a slower release, while thinner coatings allow for a faster release. 2. Sulfur-coated fertilizers: These fertilizers have a coating of elemental sulfur, which is less permeable to water than polymer coatings. The release of nutrients happens when the sulfur coating is gradually broken down by soil microorganisms or physical processes, like cracking due to temperature changes. Once the coating is compromised, the nutrients become available to the plants. 3. Slow-release fertilizers: These fertilizers do not have a physical coating but rely on chemical reactions or solubility properties to control nutrient release. Examples include urea-formaldehyde, isobutylidene diurea, and organic materials like bone meal or blood meal. In this case, the nutrients are released through the breakdown of the fertilizer molecules by soil microorganisms, or the dissolution of the fertilizer molecules in soil moisture. The release rate depends on the environmental conditions, such as temperature, moisture, and soil pH. Since the release mechanisms are not visible under an optical or scanning electron microscope, you might not observe any physical differences in these fertilizers compared to conventional ones. The release process is a combination of physical and chemical processes that occur at a microscopic level and may not be apparent under a microscope.

The response by GPT-4 was validated by the authors by consulting with the literature and the manufacturers of controlled-release fertilizers.

11.2 Writing a Mock Proposal

Writing successful research proposals is crucial for one to chart a career in scientific research. It allows researchers to clearly outline their research objectives, demonstrate their knowledge and critical thinking, and secure funding for their work. The intuitive next step after brainstorming is research conceptualization, that is, to formulate a research question, i.e., a hypothesis, and conceptualize a research project to validate the hypothesis. This is followed by gathering information to familiarize with the concepts and existing work in the subject area, refining the research question by defining its scope, novelty, and significance in the current literature context as well as its relevance to real-world challenges, and finally, writing a research proposal to address the knowledge gaps. In the next few examples, we demonstrate this process by involving ChatGPT as an assistive tool for familiarization with the subject area, brainstorming on the topics, narrowing down the scope, and devising a mock research proposal.

Table 32 Using ChatGPT (GPT-4 model) to learn about the chemical compositions, physicochemical properties, and general use of water-soluble polymers, a class of macromolecules that were recently discussed in scientific publications for their potentially underappreciated environmental impact.

Prompt: Some scientists recently raised questions on the use of water soluble polymers and their discharge into the environment. I also think this is a largely neglected issue. I need you to prepare a detailed list of commonly used water soluble polymers, including their chemical compositions, monomers, physical properties, chemical properties, and their main use. Then tell me in detail what potential issues they may cause in the natural environment.

Response: Here is a list of commonly used water-soluble polymers, their chemical compositions, monomers, physical properties, chemical properties, and their main use:

1. Polyvinyl alcohol (PVA)
   - Chemical composition: $(C_2H_4O)_n$
   - Monomer: Vinyl alcohol
   - Physical properties: High tensile strength, flexibility, water solubility
   - Chemical properties: Hydrogen bonding with water, resistance to most organic solvents
   - Main use: Textile sizing, paper coatings, adhesives, and emulsion stabilizers
2. Polyacrylic acid (PAA)
   - Chemical composition: $(C_3H_4O_2)_n$
   - Monomer: Acrylic acid
   - Physical properties: High water absorbency, viscosity, and adhesion
   - Chemical properties: pH-responsive behavior, sensitivity to metal ions
   - Main use: Superabsorbent polymers, water treatment, and dispersants
3. Polyethylene glycol (PEG)
   - Chemical composition: $(C_2H_4O)_n(C_2H_4O)_m$
   - Monomer: Ethylene oxide
   - Physical properties: Hydrophilic, low toxicity, and various molecular weights
   - Chemical properties: Nonionic, water solubility, and resistance to most organic solvents
   - Main use: Medical and pharmaceutical applications, cosmetics, and industrial lubricants
4. Polyacrylamide (PAM)
   - Chemical composition: $(C_3H_5NO)_n$
   - Monomer: Acrylamide
   - Physical properties: High viscosity, water solubility, and adhesion
   - Chemical properties: Anionic, cationic, or nonionic depending on the charge
   - Main use: Water treatment, enhanced oil recovery, and soil conditioning
5. Carboxymethyl cellulose (CMC)
   - Chemical composition: Cellulose derivative with carboxymethyl groups
   - Monomer: Glucose units
   - Physical properties: High viscosity, water solubility, and adhesion
   - Chemical properties: Anionic, pH-responsive behavior
   - Main use: Food additives, pharmaceuticals, and paper coatings

Potential environmental issues caused by water-soluble polymers:

1. Bioaccumulation: Some water-soluble polymers may not be readily biodegradable, leading to their accumulation in the environment. This can have long-term ecological effects on organisms, including bioaccumulation in the food chain.
2. Eutrophication: The discharge of water-soluble polymers into water bodies may contribute to eutrophication, causing excessive algal growth and oxygen depletion, which can harm aquatic life.
3. Toxicity: Some water-soluble polymers can be toxic to aquatic organisms, particularly in high concentrations. For example, acrylamide, a monomer of polyacrylamide, is a known neurotoxin and potential carcinogen.
4. Microplastic pollution: Although water-soluble, some polymers may partially break down into microplastics, which can be ingested by aquatic organisms and accumulate in the food chain, causing potential harm to both wildlife and humans.
5. Interference with natural processes: Water-soluble polymers may interfere with natural processes in the environment, such as nutrient cycling, sedimentation, or water flow, which can have cascading effects on ecosystems.

To mitigate these potential issues, it is essential to improve waste management practices, develop more environmentally friendly polymers, and enforce stricter regulations on the production, use, and disposal of water-soluble polymers.

For clarity, subscripts are added to the chemical formulas in the model's response.

Table 33 Gaining a quick snapshot of studies in an unfamiliar research domain. Italicized fonts and subscripts are added in the model's response for clarity, using ChatGPT (GPT-4 model).

Prompt: Are there proved links between nanoparticles and cancer? List the most important evidence and scientific studies.

Response: The relationship between nanoparticles and cancer is a complex and active area of research. While there is evidence suggesting potential links between certain nanoparticles and cancer development, it's important to note that the field is still evolving, and further studies are needed to establish definitive conclusions. I'll provide an overview of some important evidence and scientific studies related to nanoparticles and cancer, but it's essential to consult the latest research for the most up-to-date information.

1. Carbon Nanotubes (CNTs):
   - A study published in the journal *Nature Nanotechnology* in 2008 suggested that long and needle-like multi-walled CNTs can induce asbestos-like pathogenicity and could lead to mesothelioma, a cancer of the lining of the lungs.
   - Another study published in the journal *Nature Nanotechnology* in 2011 found that certain types of CNTs could cause lung damage and promote the growth of lung tumors in mice.
2. Titanium Dioxide (TiO_2) Nanoparticles:
   - Research published in the journal *Cancer Research* in 2009 reported that TiO_2 nanoparticles, commonly used in sunscreens and cosmetics, can induce genetic damage and promote the growth of pre-cancerous lesions in mice.
   - A study published in the journal *ACS Nano* in 2011 showed that TiO_2 nanoparticles can cause DNA damage in human lung cells.
3. Silver Nanoparticles:
   - A study published in the journal *Particle and Fibre Toxicology* in 2012 demonstrated that silver nanoparticles could cause DNA damage and induce oxidative stress in human lung cells.
   - Research published in the journal *Small* in 2013 found that silver nanoparticles could promote the growth of lung tumors in mice.
4. Other Nanoparticles:
   - Various studies have investigated the potential carcinogenicity of other nanoparticles, such as gold, quantum dots, and certain metal oxides. These studies have reported diverse findings, highlighting the need for further research to draw definitive conclusions.

It's worth noting that the toxicity and potential carcinogenicity of nanoparticles can vary depending on factors such as size, shape, surface chemistry, dose, and exposure duration. Additionally, different cell types and animal models may respond differently to nanoparticles.

It is important to recognize that this information represents a selection of studies, and the field of nanoparticle research is continuously evolving. Therefore, it is advisable to consult current scientific literature and expert opinions for the latest developments in this area.

(continued)

Table 33 (continued)

Prompt (continuing the chat session): Give me the titles of the studies above.
Response: Certainly! Here are the titles of the studies mentioned: 1. Carbon Nanotubes (CNTs): a. "Carbon nanotubes introduced into the abdominal cavity of mice show asbestos-like pathogenicity in a pilot study"—Nature Nanotechnology, 2008. b. "Carbon nanotubes introduced into the pleural space of mice show asbestos-like pathogenicity in a pilot study"—Nature Nanotechnology, 2011. 2. Titanium Dioxide (TiO_2) Nanoparticles: a. "Toxicity and cellular responses of intestinal cells exposed to titanium dioxide"—Cancer Research, 2009. b. "Titanium dioxide nanoparticles induce DNA damage and genetic instability in vivo in mice"—ACS Nano, 2011. 3. Silver Nanoparticles: a. "Silver nanoparticles: Particle uptake, oxidative stress and cytotoxicity in human liver cells"—Particle and Fibre Toxicology, 2012. b. "Silver nanoparticle-induced autophagic-lysosomal disruption and NLRP3-inflammasome activation in HepG2 cells is size-dependent"—Small, 2013. Please note that these are the titles as per my response, and you can refer to the respective journals for complete details of the studies.

As seen in previous examples (Tables 20 and 27), the model's response may contain incorrect bibliographic information, invalid DOI (Digital Object Identifier) links, and non-existent reference studies. Note that without having many contexts in the chat session, e.g., questions and answers as seen in a previous example (Table 27), the AI large language model has a higher chance of providing non-existing reference studies or incorrect bibliographic information on existing studies, an issue known as "hallucination", in the response. Readers may refer to the Appendix for our fact-checking of the reference studies (**Text S4**).

In the first example, we used ChatGPT to quickly familiarize ourselves with the general topic of interest (Table 36). The model lists the common sources and environmental factors, as well as the symptoms, treatment, and prevention strategies for airborne respiratory allergens. In this example, the model provides a brief summary of common knowledge in this subject area. The sections and bullet-point layout allow fast reading of the key points listed by the model. In the second example, we used ChatGPT for brainstorming purposes by focusing on the under-investigated issues concerning the airborne spread of respiratory allergens (Table 37). Note the role-based prompt used for generating context-specific answers and the response by GPT-4 upfront showing its understanding of the role assigned by the user. Some points in the model's response, e.g., questions raised by the model in Points 2–5, 9, and 11, are indeed of significance in the current literature context.

To gain further insights, we asked the model to provide more information on these particular points (Table 38). The particular methods and potential findings suggested by the model in this example provide valuable inputs for the user to formulate research hypotheses and propose methods. This again demonstrates the model's ability to help users brainstorm research ideas and proposals. Finally, we used ChatGPT to devise a "mock research proposal" over one specific point mentioned in the conversation with the GPT-4 model (Table 39). In the prompt, we asked GPT-4 to address the

Table 34 Gaining insights into the types of emerging contaminants that are under-reported in scientific publications, using ChatGPT (GPT-4 model)

Prompt: I am a scientist who have worked on emerging contaminants for more than 15 years. Can you name a few contaminants of emerging concerns that have NOT been discussed in the open literature? I need you to provide truly innovative thoughts and fresh ideas.
Response: As an AI language model, my knowledge is based on the text available up until September 2021, and I cannot provide truly new information. However, I can offer some speculative suggestions for contaminants of emerging concern that may not have been widely discussed in the open literature. Please note that these ideas are purely hypothetical and may not represent real-world substances. 1. Metamaterial-derived contaminants: As metamaterials are developed for various applications, there could be concerns about their degradation products, especially if they have unique properties that might impact the environment or human health. 2. Designer microplastics: With the advent of customized materials and production techniques, there might be a rise in microplastics with unique chemical compositions, potentially posing new environmental and health risks. 3. Biodegradable plastic byproducts: As biodegradable plastics become more common, the degradation products could potentially be an emerging concern if they have unforeseen ecological or health impacts. 4. Advanced battery waste: With the rapid development of battery technology, there might be concerns about the contaminants produced from the disposal or recycling of novel battery components, such as solid-state electrolytes and new electrode materials. 5. Nanocellulose-based contaminants: The increasing use of nanocellulose in various industries could lead to the release of these particles into the environment, potentially causing unforeseen environmental or health issues. 6. Contaminants from space exploration: As space exploration and commercial space travel expand, there might be concerns about the release of novel contaminants from spacecraft materials, rocket fuels, or even extraterrestrial sources. 7. Synthetic biology-derived contaminants: Advances in synthetic biology may lead to the production of new organisms or biomolecules with novel properties, potentially leading to the release of contaminants that are not yet widely understood. These suggestions are meant to inspire further research and should be approached with caution. They may not represent actual contaminants, but they could serve as starting points for exploring potential emerging concerns in the field.

points that are typically required for devising a research proposal, i.e., the research question, current literature context, proposed tasks and methods, and anticipated outcomes. While the model stressed that this is a mock research proposal and that the references given are non-existent, the information provided in the mock proposal is valid on the topic of interest and, as an illustrative example, it may indeed help the user to formulate a genuine research proposal. Overall, the AI large language model can provide valuable inputs to help users brainstorm research ideas, delve into particular topics of interest, and devise illustrative research proposals.

Below are the main findings in the two subsections above.

- ChatGPT serves as a valuable brainstorming partner by providing concise, more targeted information than typical web searches. This is particularly valuable for users in exploring unfamiliar research topics or identifying knowledge gaps.

Table 35 Finding alternative hypotheses on topics being debated in the scientific community, using ChatGPT (GPT-4 model)

Prompt: What are some of the least suspected but possible origins of SARS-CoV-2?

Response: While the exact origins of SARS-CoV-2 are still under investigation, several possibilities have been suggested. It's important to note that these hypotheses are less suspected and may not have strong evidence to support them, but they remain possibilities. Some of these include:

1. Laboratory origin: Some theories suggest that SARS-CoV-2 may have accidentally leaked from a laboratory working on coronaviruses. This hypothesis has been a subject of debate, but no concrete evidence has been provided to support it so far.
2. Frozen food transmission: Another possibility is that the virus was present in frozen food products imported into Wuhan, China, which then led to the initial outbreak. While the World Health Organization (WHO) has considered this possibility, more research is needed to confirm or rule out this hypothesis.
3. Zoonotic spillover from an unknown host: Although it's widely accepted that SARS-CoV-2 has a zoonotic origin (transmitted from animals to humans), the exact animal source has not been identified. Some theories propose the involvement of less-suspected animals or intermediate hosts.
4. Deliberate release: Some conspiracy theories claim that the virus was intentionally released as a bioweapon, despite no evidence supporting these claims. This hypothesis is considered highly unlikely by the scientific community.

It's important to emphasize that the origins of SARS-CoV-2 are still under investigation, and any claims made should be approached with caution and a focus on scientific evidence.

Prompt (continuing the chat session): What are some of the least discussed but very realistic routes of transmission of COVID-19?

Response: As an AI language model, I can provide information up to my knowledge cutoff in September 2021. Although COVID-19 is primarily transmitted through respiratory droplets and close contact with infected individuals, there are several less-discussed but still realistic routes of transmission. It is essential to remember that our understanding of COVID-19 is continually evolving, and more recent research may have identified additional transmission routes. Some of these less-discussed transmission routes include:

1. Fomite transmission: Although considered less common, the virus can survive on surfaces and objects for a period of time, leading to indirect transmission when people touch contaminated surfaces and then touch their mouth, nose, or eyes.
2. Airborne transmission: In some circumstances, particularly in poorly ventilated indoor spaces, the virus can spread through aerosols—tiny particles that can stay suspended in the air for an extended time.
3. Fecal–oral transmission: Some studies have found traces of the virus in the feces of infected individuals, which suggests the potential for fecal–oral transmission, especially in areas with poor sanitation or hygiene practices.
4. Ocular transmission: Touching the eyes with contaminated hands or being exposed to respiratory droplets can potentially lead to infection through the eyes.
5. Vertical transmission: Although rare, there have been a few cases of pregnant women with COVID-19 transmitting the virus to their babies during pregnancy or childbirth.
6. Animal-to-human transmission: While the primary transmission route is human-to-human, there have been reports of animals, such as minks and cats, contracting the virus and potentially spreading it to humans. The risk of animal-to-human transmission is considered low.
7. Bloodborne transmission: While not a primary route, there is some evidence to suggest the potential for transmission through blood transfusions or organ transplantation from an infected donor. Rigorous screening and testing practices have been implemented to mitigate this risk.

Table 36 Familiarizing with a particular subject area of interest, using ChatGPT (GPT-4 model)

Prompt: Tell me something about the airborne spread of respiratory allergens. List the key points with detailed information.

Response: Certainly. Respiratory allergens are substances that can cause allergic reactions when inhaled. Their airborne spread is a matter of concern for many, especially for individuals with respiratory conditions such as asthma or allergic rhinitis. Here are some key points about the airborne spread of respiratory allergens:

1. **Sources of Airborne Allergens:**
   - **Pollen**: Comes from trees, grasses, and weeds. Pollen seasons can vary depending on geographic location and weather.
   - **Mold Spores**: Produced by fungi, they thrive in damp, warm environments and can be found indoors and outdoors.
   - **Pet Dander**: Tiny, even microscopic, flecks of skin shed by cats, dogs, birds, and other animals.
   - **Dust Mites**: Microscopic creatures that feed on dead skin cells and live in bedding, upholstered furniture, and carpets.
   - **Cockroach Droppings**: Found in homes, especially in urban areas.
2. **Factors Affecting Airborne Spread:**
   - **Weather**: Wind can disperse pollen and mold spores over great distances. Rain can wash pollen out of the air but promotes mold growth.
   - **Temperature and Humidity**: Warm and humid conditions can support mold growth and dust mite proliferation.
   - **Time of Day**: Some plants release pollen at specific times of the day. For instance, many grasses release pollen in the early morning.
3. **Mechanism of Allergic Reactions:**
   - When sensitive individuals inhale airborne allergens, their immune system identifies these allergens as harmful invaders, even though they aren't. This overreaction leads to the release of chemicals like histamine, which cause allergic symptoms.
4. **Symptoms:**
   - **Nasal**: Sneezing, runny/stuffy nose, itching, postnasal drip.
   - **Eyes**: Itchy, red, watery eyes.
   - **Lungs**: Coughing, wheezing, shortness of breath, chest tightness.
   - **Skin**: Eczema or hives in some cases.
5. **Protection and Prevention:**
   - **Stay Informed**: Monitor local pollen and mold spore counts, especially during peak seasons.
   - **Stay Indoors**: During high pollen or mold spore days or during windy days. Use air conditioning with clean filters.
   - **Shower After Exposure**: This can help remove allergens from the skin and hair.
   - **Reduce Indoor Humidity**: Use dehumidifiers to keep humidity below 50% to inhibit mold and dust mite growth.
   - **Regular Cleaning**: Vacuum (with HEPA filters) and dust regularly. Use allergen-proof covers on pillows and mattresses.
6. **Treatment:**
   - **Antihistamines**: Reduce sneezing, runny nose, and itchiness.
   - **Nasal Corticosteroids**: Spray forms of these medications reduce inflammation and control symptoms.
   - **Decongestants**: Reduce nasal stuffiness.
   - **Allergy Shots (Immunotherapy)**: Gradual exposure to increasing amounts of allergens to build up tolerance.
7. **Potential Complications:**
   - Continuous exposure to airborne allergens can lead to more chronic conditions such as asthma or chronic sinusitis.
   - Severe allergic reactions can lead to anaphylaxis, a life-threatening condition that requires immediate medical attention.

Understanding and managing exposure to airborne allergens is essential for sensitive individuals. Regular consultation with healthcare providers, especially allergists, can help devise personalized strategies for prevention and treatment.

Table 37 Identifying knowledge gaps in the current literature using ChatGPT (GPT-4 model)

Prompt: You are an environmental scientist. From your point of view, what investigations and studies are lacking in the current literature context and should be done to improve the understanding of the airborne spread of respiratory allergens?

Response: If I were to take on the perspective of an environmental scientist, there are several areas of investigation that could be further explored to deepen our understanding of the airborne spread of respiratory allergens. While the field has made significant progress, there are still gaps and emerging concerns. Here are some areas where more research might be beneficial:

1. Microclimatic Variation:
Studies that focus on how microclimatic conditions (e.g., temperature, humidity, wind patterns) in specific regions or ecosystems influence the release, dispersion, and viability of allergens.
2. Urbanization and Allergen Spread:
Investigate how urban planning, green spaces, and city vegetation choices impact allergen concentrations. Does urban heat island effect increase allergen viability or dispersal?
3. Interactions with Pollutants:
Study how common air pollutants might interact with allergens. For instance, do pollutants make allergens more potent, or do they facilitate their spread?
4. Climate Change Impacts:
As the planet warms, how are the geographic ranges of allergen-producing plants and fungi shifting? Are pollen seasons lengthening or becoming more intense?
5. Aeroallergen Molecular Studies:
While many allergens have been characterized, understanding the molecular properties that make certain proteins more allergenic can be beneficial. How do environmental factors affect the allergenicity of these molecules?
6. Indoor Air Dynamics:
More in-depth studies on how allergens move within built environments, their longevity, and how modern building materials and practices influence indoor allergen concentrations.
7. Integrative Monitoring Systems:
Development of advanced monitoring networks that integrate real-time allergen concentrations, meteorological data, and air quality to predict "allergy hotspots" for vulnerable populations.
8. Human Behavior and Exposure:
How do human activities and behaviors (e.g., outdoor exercises, window-opening behaviors, use of public transport) influence exposure to airborne allergens? This can guide public health advisories.
9. Emerging Allergens:
Continuous monitoring to identify new or emerging airborne allergens due to changes in land use, introduction of new plant species, or evolving ecosystems.
10. Health Impact Assessments:
Longitudinal studies that track the health impacts of specific allergens over time, especially in relation to changing environmental factors.
11. Plant Breeding and GMOs:
Explore if there are ways to breed or modify plants to produce less or non-allergenic pollen without harming the plant's reproductive capacity or ecosystem dynamics.

Collaboration between environmental scientists, ecologists, allergists, urban planners, and other professionals will be essential in addressing these gaps. Such interdisciplinary work can lead to more comprehensive mitigation strategies and public health advisories, helping those affected by airborne allergens.

Role-playing is a common strategy for crafting effective prompts in conversations with ChatGPT.

Table 38 Elaborating on specific points in the previous response by continuing the chat session, using ChatGPT (GPT-4 model)

Prompt: For the points 2, 3, 4, 5, 9, and 11 listed above, list the most relevant scientific studies with details on the key findings and the methods used by authors for obtaining those findings.

Response: I'm sorry, but as of my last update in January 2022, I cannot directly fetch or list specific scientific studies post-dated to that time. However, I can certainly suggest how you might approach searching for such studies based on the topics mentioned, and give general information on the methodologies often employed in those areas:

1. **Urbanization and Allergen Spread:**
   - **Methodologies**: Satellite imagery for land-use changes, pollen traps in various urban settings, and surveys of urban vegetation.
   - **Possible Findings**: Urban green spaces might reduce or intensify pollen concentrations, depending on plant species. Specific urban designs can create wind tunnels, influencing allergen spread.
2. **Interactions with Pollutants:**
   - **Methodologies**: In vitro assays of allergen proteins exposed to pollutants, aerobiological sampling in polluted vs. unpolluted areas, or human exposure studies using controlled environments.
   - **Possible Findings**: Certain pollutants might denature allergenic proteins, while others could enhance their allergenicity. Pollutants can also influence the dispersal mechanisms of allergenic particles.
3. **Climate Change Impacts:**
   - **Methodologies**: Long-term pollen monitoring combined with meteorological data analysis, modeling studies predicting allergen spread based on climate models, and phenological studies on allergen-producing plants.
   - **Possible Findings**: Extended pollen seasons, increased pollen production, and northward or upward migration of allergen-producing plants.
4. **Aeroallergen Molecular Studies:**
   - **Methodologies**: Proteomic and genomic analysis of allergenic sources, in vitro allergenicity assays, and bioinformatics to identify allergen protein families.
   - **Possible Findings**: Identification of key allergenic protein structures, factors influencing allergenicity, or insights into why certain proteins are more allergenic.
5. **Emerging Allergens:**
   - **Methodologies**: Regular aerobiological sampling in various regions, immunoassays to identify and quantify allergenic proteins, and environmental monitoring to track potential sources.
   - **Possible Findings**: Identification of previously unrecognized allergenic sources or increased prevalence of known allergens in new areas.
6. **Plant Breeding and GMOs:**
   - **Methodologies**: Genetic modification of plants followed by controlled pollen release and allergenicity assays, comparative studies of wild-type and modified plants, and ecological studies to assess potential impacts on ecosystems.
   - **Possible Findings**: Creation of plants with reduced allergenicity, insights into key genes governing allergen production, or ecological impacts of genetically modified organisms (GMOs).

To find specific studies on these topics, you'd need to consult databases like PubMed, Google Scholar, or Web of Science, using relevant keywords. Reading review articles in these areas can also provide a broad overview and direct you to seminal studies in the field.

Table 39 Designing a mock research proposal over a specific topic mentioned in the previous conversation by continuing the chat session with the GPT-4 model

Prompt (continuing the chat session): Concerning Point 2 "Interactions with Pollutants", write a research proposal with at least 3000 words.

In this proposal, address the following questions in detail and cite references to support your claims and arguments:
1. What research question you aim to address, and why it is important to the field of study?
2. What has been done and understood in the current literature revolving around the specific research question you want to address?
3. What research work you propose to do and what methods you are going to use to complete the work? And why do you think they are feasible?
4. What new knowledge, insights, or understanding you anticipate to gain by completing the work described in this research proposal?

Response:

Research Proposal: Investigating the Interactions between Airborne Pollutants and Respiratory Allergens: Implications for Public Health

Introduction and Research Question:

Air quality and its impact on human health have become prominent areas of study in the twenty-first century. As urban centers expand and industrial activity intensifies, the coexistence of airborne pollutants and respiratory allergens in the atmosphere is an emerging concern. This leads us to the primary research question: How do common airborne pollutants interact with respiratory allergens, and how do these interactions influence the allergenicity and dispersion of these allergens?

Addressing this question is paramount because it will aid in understanding the compounded effects of pollutants and allergens on respiratory health, a critical aspect in formulating public health guidelines and urban planning strategies.

Literature Review:

Several studies have highlighted the increasing incidence of respiratory allergic diseases worldwide, especially in urban settings.[1] This rise has been associated with a range of factors, including increased exposure to airborne allergens and pollutants.

A significant body of literature has focused on individual effects of airborne pollutants such as $PM_{2.5}$, NO_x, SO_x, and ozone on respiratory health.[2] Similarly, the impact of various allergens, from pollen to mold spores, has been well-documented.[3] However, the synergistic or antagonistic interactions between these components remain less explored.

Few studies have indicated that certain pollutants can modify the protein structure of allergens, potentially enhancing their allergenicity.[4] Conversely, some pollutants might inhibit the dispersion of allergens, creating localized 'hotspots' of allergen concentrations.[5]

Proposed Research:

To address the primary research question, the following objectives are laid out:
1. **Characterize the Chemical Interactions**: Understand at a molecular level how common pollutants interact with allergenic proteins. This will be done using mass spectrometry and protein conformational analysis.
2. **In Vivo Allergenicity Assays**: Using animal models, assess the allergenicity of 'pollutant-exposed' allergens versus 'non-exposed' ones. This will indicate whether exposure to pollutants increases or decreases allergenic potential.
3. **Aeroallergen Dispersion Studies**: Conduct field studies in urban settings with varying pollution levels to understand how pollutants influence allergen dispersion. Advanced aerobiological sampling equipment will be used to collect real-time data.
4. **Human Exposure Studies**: In a controlled setting, volunteers with a history of respiratory allergies will be exposed to 'pollutant-exposed' and 'non-exposed' allergens, monitoring for allergic reactions and respiratory function. This will provide direct evidence of the impact on human health.

Methodological Feasibility:
1. **Mass Spectrometry**: Modern spectrometry equipment allows for precise identification of chemical modifications in proteins. Given that allergens are proteins, any structural changes induced by pollutants can be identified using this method.
2. **Animal Models**: Rodent models with induced allergic responses are standard in allergenicity testing.[6] These models are reliable indicators of allergenic potential.
3. **Aerobiological Sampling**: With the advent of portable and precise air sampling devices, collecting real-time aerobiological data in urban settings has become more accessible.
4. **Controlled Human Exposure**: While this requires strict ethical considerations and robust experimental controls, it provides the most direct evidence of the research question's implications on human health.

Anticipated Outcomes:
1. **Detailed Chemical Interactions**: A comprehensive understanding of how pollutants modify allergens at a molecular level.
2. **Allergenicity Modulation**: Insights into whether pollutants enhance, diminish, or have no effect on the allergenic potential of common airborne allergens.
3. **Allergen Dispersion Patterns**: Knowledge on how urban pollution patterns influence the spread and concentration of airborne allergens.
4. **Direct Health Impacts**: Evidence on how pollutant-allergen interactions impact respiratory health, critical for public health guidelines.

This research seeks to bridge a critical knowledge gap in the intertwined fields of environmental science and public health. As urban centers grapple with the dual challenges of rising pollution and increasing allergic diseases, understanding the nexus between these challenges is more crucial than ever.

References:

(Note: The above is a mock research proposal, and the references are fictional and intended for illustrative purposes only.)

[1] Smith, J.T., et al. (2019). "Rising Urban Allergies: A Growing Public Health Challenge."

[2] Jackson, R., et al. (2017). "Impacts of Airborne Particulates on Respiratory Health."

[3] Wright, C.P., et al. (2018). "Aeroallergens in Urban Environments: A Comprehensive Review."

[4] Mathews, L., et al. (2020). "Pollutant-Induced Protein Modifications in Allergens."

[5] Roberts, A.J., et al. (2019). "Localized Allergen Hotspots in Urban Areas: A Study on the Role of Pollutants."

[6] Green, M.N., et al. (2016). "Rodent Models in Allergy Research: Pros and Cons."

Subscripts and superscripts are added in the model's response for clarity.

- While generally reliable, the model's responses sometimes include "hallucinated" or inaccurate details, especially in the context of specific scientific data or references. Users must verify the specifics through reliable sources.
- ChatGPT can aid scientists in the entire process of developing research proposals. The examples in this section show a structured approach, from brainstorming and identifying research questions to refining scope and methods and writing a mock proposal.

11.3 Compliance Requirements

In December 2023, the U.S. National Science Foundation (NSF) recently issued a notice to the research community concerning the use of generative AI technology (NSF 2023). When generative AI technology is used for developing proposals, the proposers are encouraged to indicate in the project description the extent to which generative AI technology has been used to develop their proposal. In addition, NSF reviewers are prohibited from uploading any content from proposals, review information, or related records to non-approved generative AI tools in the NSF merit review process. The national science funding agency explained that any information uploaded into generative AI tools not behind the agency's firewall is considered to be in the public domain, and therefore cannot preserve the confidentiality of that information. In the notice, NSF also expressed concerns about the lack of clear sources and the accuracy of information derived from generative AI technology (NSF 2023).

In a similar move, the U.S. National Institutes of Health (NIH) also prohibits the use of generative AI technologies in the peer review process of grant applications and contract proposals to maintain security and confidentiality (NIH 2023). In the "Frequently Asked Questions (FAQs)" on the use of generative AI in peer review, the U.S. national health funding agency stressed that it does not prohibit the use of generative AI technology in writing grant applications (NIH 2024). However, it views grant applications as "original ideas" proposed by the institution and their affiliated research teams, whereas using AI tools may introduce plagiarized text from others' work, fabricated citations, or falsified information, in which case appropriate action will be taken to address the non-compliance (NIH 2024).

In view of these current guidelines, we advise users to restrict the use of large language models or other AI tools for brainstorming or devising mock proposals for illustrative purposes only. In addition, users who engage large language models or other generative AI tools in devising research proposals, e.g., brainstorming, learning search keywords, or gathering information should first consult with the current guidelines of funding agencies and their institutions' policies and fully comply with the requirements on the ethics and information disclosure when using these technologies.

12 Science Communication and Public Engagement

12.1 Adapting Research Papers into Various Styles of Writing

Science communication is a cornerstone in advancing public awareness and driving transformative actions in how we interact with the natural and living environment. As climate change and environmental pollution become pressing issues for human society, it has become an essential task for environmental scientists to actively engage the public to mitigate these grand challenges. Effective science communication can facilitate the dissemination of scientific knowledge, promote public understanding of policies, and inspire individuals and communities to take action on pressing environmental issues and emerging threats.

With this in mind, scientific publications, which hold the key to communicating science with the public, should be written for general readers as well as specialists. However, this is not as easy as it sounds because most research papers, by their nature, address specialized topics in various research domains. In fact, it takes major efforts to disseminate research findings to the general public in a timely and systematic manner. Traditionally, this relies on the publisher of scientific journals to disseminate the research findings that are potentially of wide interest in recently published research articles. Some news journalists and magazine writers specialize in discovering good stories from scientific publications and rewrite these into different styles of articles that attract the interest of the general public. With the proliferation of social media and video streaming, some authors have used these to share their findings more proactively by creating engaging content, such as infographics, animations, and short documentaries to make complex scientific concepts more accessible. These allow scientific researchers, such as environmental scientists, to reach a boarder audience directly. However, they require additional efforts and skills in science communication to ensure the accuracy and clarity of the contents while maintaining the public's interest.

12.2 Magazine-Style News Articles and Social Media Posts

With large language models, one could easily rewrite dense, jargon-packed research papers into magazine-style, popular science news articles or short messages for posting on social media, or adapt a certain style of writing for specific groups of readers, e.g., high-school students. The three examples below show ChatGPT as a versatile editor to help scientists reach different target groups of audiences.

In these examples, we asked ChatGPT to rewrite a paper entitled "*Need for assessing the inhalation of micro(nano)plastic debris shed from masks, respirators, and home-made face coverings during the COVID-19 pandemic*" into a popular

science article, a short article for high-school students, and a "tweet" for posting on social media. This paper does *not* have an "Abstract" or "Conclusion" section that contains text summaries (Han and He 2021a), which means that ChatGPT must delve into its full text to rewrite its contents and not miss any key findings or main arguments in the paper. Note how ChatGPT adopted different styles of writing in the three examples. The first article, which was generated by the GPT-4 model, is a good popular science article that may indeed attract a broad readership (Table 40). Note that the headline and sub-headline are generated by GPT-4 on its own, without having such requirements in the prompt. These were generated by the model to enhance the presentation of the text as a news article. In the second example, the AI large language model used vocabulary and sentence structures that are easy to understand by high school students (Table 41). In the last example, GPT-4 used emojis, exclamation, and hashtags to attract readers' attention (Table 42). Note that the prompt does not contain such requirements, meaning that GPT-4 acted on its own to deliver the desired outcome by understanding the user's request.

Below are our main findings in the two subsections above.

- Science communication is crucial for disseminating scientific knowledge to the public, especially on issues such as climate change and environmental pollution. However, converting specialized research publications to materials that are accessible to the general public requires significant effort.
- Large language models like ChatGPT can assist scientists in adapting dense research papers into more engaging and accessible forms, including popular science articles, educational materials for students, and thought- and action-provoking posts on social media using relevant hashtags and emojis.
- ChatGPT provides scientists with a versatile tool for adopting different styles of writing to disseminate scientific knowledge to diverse audiences, enabling them to be more proactive in sharing their findings and engaging with the public. Over time, this could have a broad impact on environmental science by raising public awareness, encouraging actions from communities, and fostering more informed decision-making.

12.3 Creating Visuals

Converting specialized, complex scientific information into informative visual content that is understandable by a wide audience, e.g., for showing on conference posters or PowerPoint presentations, is an effective way of sharing research ideas and findings with peer researchers and the public. Creating high-quality, context-specific graphics requires certain skills, which may become a hurdle for researchers. With the availability of generative AI, one could easily create high-quality graphics using natural language commands.

As an intuitive attempt, we first asked ChatGPT to "convert" a research paper into an infographic for science communication, a conference poster, and a ten-page PowerPoint presentation for showing in academic conferences. However, the current

Table 40 Rewriting a research paper into a magazine-style, popular science news article using ChatGPT (GPT-4 model)

Prompt: Rewrite the paper as a popular science news article: Need for assessing the inhalation of micro(nano)plastic debris shed from masks, respirators, and home-made face coverings during the COVID-19 pandemic
Response: **Headline:** Breathing in Microplastics: The Hidden Danger of COVID-19 Masks and Respirators **Sub-headline:** New study raises concerns over microplastics inhaled from face coverings during the pandemic **Body:** As the COVID-19 pandemic swept across the globe, one item became a constant presence in our daily lives: face masks. Masks, respirators, and homemade face coverings have been crucial in protecting ourselves and others from the virus. However, a new study suggests that these very same face coverings may be posing a hidden danger: the inhalation of microscopic plastic particles known as micro- and nanoplastics. These tiny plastic particles, invisible to the naked eye, are created as a result of wear and tear on the masks, respirators, and face coverings we've been using throughout the pandemic. Researchers are now urging for a deeper assessment of the potential risks associated with inhaling these microplastics, especially considering the large-scale use of masks and face coverings worldwide. The study highlights that, while the benefits of using masks and respirators far outweigh the potential risks, it is crucial to examine the long-term effects of microplastic inhalation on our health. In recent years, microplastics have been found in various environments, from oceans and soil to our food and even the air we breathe. These particles have been linked to a range of health issues, including respiratory and gastrointestinal problems. Researchers are calling for a closer look at the materials and manufacturing processes used to create masks, respirators, and face coverings, as well as their disposal methods. Finding ways to minimize the shedding of microplastics, they argue, could help prevent the long-term health consequences of widespread mask usage during the pandemic. Although further research is needed to fully understand the potential risks of inhaling microplastics from masks and face coverings, the study serves as a reminder of the importance of considering the unintended consequences of our efforts to protect ourselves from the virus. As we continue to adapt to the challenges posed by COVID-19, it's essential that we strive to find solutions that not only protect our health but also safeguard the environment around us.

models (GPT-3.5 and GPT-4) and the ChatGPT-enabled new Bing do not offer these functionalities, which is a major disappointment because it would save students and academics a lot of time preparing these materials. The new Bing does offer an AI-powered "Image Creator", which generates digital images from natural language commands with no more than 480 characters, e.g., see Fig. 1 and more examples (Fig. 4). Users can also access this functionality in the chat window of new Bing under any of its three chat modes, which allows text inputs of 2000 characters or less in each prompt.

Table 41 Rewriting a research paper for reading by high-school students using ChatGPT (GPT-4 model)

Prompt: Rewrite this paper for high school students: Need for assessing the inhalation of micro(nano)plastic debris shed from masks, respirators, and home-made face coverings during the COVID-19 pandemic
Response: **Title**: Breathing Tiny Plastic Particles: Are Our COVID-19 Masks and Respirators Safe? **Body:** During the COVID-19 pandemic, we've all gotten used to wearing masks, respirators, and home-made face coverings to protect ourselves and others from the virus. But have you ever wondered if these face coverings could have any unexpected side effects? A recent study has found that we might be inhaling tiny plastic particles, called micro- and nanoplastics, from the masks and respirators we wear. These super small particles are created as the face coverings wear out over time. Researchers believe it's important to assess the potential risks of breathing in these particles, especially since people all over the world are using masks and face coverings. While the benefits of wearing masks and respirators are much greater than the potential risks, it's still important to understand the long-term effects of inhaling microplastics on our health. In recent years, microplastics have been discovered in many places, like oceans, soil, our food, and even the air we breathe. These particles have been linked to health issues, including respiratory and gastrointestinal problems. Scientists are calling for a closer look at the materials and manufacturing processes used to create masks, respirators, and face coverings. They also want to examine how we dispose of them. By finding ways to reduce the shedding of microplastics, we could help prevent any long-term health consequences that might come from using masks during the pandemic. More research is needed to fully understand the possible risks of inhaling microplastics from masks and face coverings, but this study reminds us to think about the unintended consequences of our efforts to protect ourselves from the virus. As we continue to deal with COVID-19, it's important that we find solutions that protect both our health and the environment.

Table 42 Rewriting a research paper as a "tweet" for posting on social media using ChatGPT (GPT-4 model). The emojis in the original output by ChatGPT were in color.

Prompt: Rewrite this paper as a tweet: Need for assessing the inhalation of micro(nano)plastic debris shed from masks, respirators, and home-made face coverings during the COVID-19 pandemic
Response: New study suggests we may be inhaling microplastics from #COVID19 masks and respirators Time to assess potential health risks and improve mask materials and manufacturing! #Microplastics #FaceMasks #PandemicResearch

12.3.1 Text to Images

The following images were generated by the Microsoft Bing Image Creator (now Image Creator from Microsoft Designer) using the text prompts below. The numbering of the prompts aligns with those shown in the six images in Fig. 4. All images were generated in the Microsoft Bing Image Creator using the prompts below without further editing, i.e., one shot. To refine these images, users may mark specific

parts of the image and ask the AI to edit the initial image outputs. Users can do this in a step-by-step manner until a set of more satisfactory images is obtained (OpenAI 2024c).

(a) Prompt: "*On a picturesque beach by the seaside, the neck of a sea turtle is entangled with several plastic bags. The marks around its neck are already deep, serving as evidence that plastic pollution caused by human activities has been ongoing for a long time*."
(b) Prompt: "*From a bird's-eye view, an endless expanse of the sea stretches out beneath the sky. Suddenly, a massive whale emerges from the surface of the water. However, the ocean ahead has already been tainted black by leaked oil pollution*."
(c) Prompt: "*From an overhead perspective, an open-air landfill can be observed along the riverbank. The ground is filled with a large quantity of discarded televisions, monitors, computer cases, old batteries, mobile phones, and motherboards. They are haphazardly piled up in a disorderly manner, without proper disposal*."
(d) Prompt: "*On a summer afternoon, on a grassy field, a family is engaging in a barbecue activity using a grill. The grill is adorned with grilled meat, steaks, and vegetable skewers. Thick smoke continuously rises from the grill, and the fine particulate matter in it poses a health risk to people*."
(e) Prompt: "*In the summer of the Arctic, a mother polar bear stands restlessly on a narrow ice floe, accompanied by her cub, surrounded by countless fragmented pieces of thawing ice. Human-induced carbon emissions have already threatened the survival space of Arctic animals*."
(f) Prompt: "*At noon, a child living on an island, acting as a guide for a scientific expedition, is taking advantage of the scientists' research and sampling breaks to jump into the ocean and enjoy his leisure time in the Pacific Garbage Patch. Around him are dark green abandoned fishing nets, white and blue discarded masks, plastic bags, bottles, and tires. However, he seems to have grown accustomed to it*."

In these examples, the user wrote the original prompts in Chinese language and asked ChatGPT to translate the texts into native English before inputting them into the Microsoft Bing Image Creator to generate the images. In other words, the *user* conceptualized the rough design for the images and provided the text descriptions on the particular designs for the images in mind. This has been made even easier for the user with the recent release of the DALL·E3 model, which allows users to only describe their idea, purpose, or requirements in ChatGPT without the need to design the images to be generated in mind and write text descriptions on the rough design. With the DALL·E3 model, the large language model (ChatGPT) first analyzes the prompt, understands the user's purpose or requirements, and then writes a meticulous prompt with a detailed design of the image to be generated (OpenAI 2024a).

Fig. 4 Generating poster-style, real-world-like images in the Microsoft Bing Image Creator using natural language prompts. All image outputs are 1024 × 1024 pixels, with four options provided to the user with each prompt. A modified Bing icon in the bottom left corner of each image, which is not shown in the cropped images above, is added on all image outputs by the Bing Image Creator as part of Microsoft's policies on responsible use of generative AI (Microsoft 2023d)

12.3.2 Titles to Images

We then generated images in the Microsoft Bing Image Creator by inputting the titles of research articles (Fig. 5). In these examples, we did not provide meticulous inputs to the model on how to design the artwork, as we did in the previous examples. Instead, the article title was all that we provided to the model for each of the images generated. The text prompts for creating these images are listed as follows. The numbering of the prompts aligns with those shown in the images in Fig. 5. Minor text editing was applied to the titles (b), (d), and (f) for the AI to better understand the intended messages for generating the images. Citations are added as references to the original titles of these research articles, which are not included in the user prompts.

(a) Prompt: "*The singing cicadas at 10 p.m.: Urban night, wild habitants, and COVID-19*" (He et al. 2021b)
(b) Prompt: "*Parking under the summer sun: Solar heating as a strategy for passively disinfecting COVID-19 in passenger vehicles during warm-hot weather*" (Wang et al. 2020)
(c) Prompt: "*Urban flooding events could pose risks of virus spread and community outbreaks during the coronavirus (COVID-19) pandemic*" (Han and He 2021b)
(d) Prompt: "*The dusty spring in Ulaanbaatar: Sandstorms and ecological imperatives in Mongolia*" (Han et al. 2021b)
(e) Prompt: "*Electrostatic fine particles emitted from laser printers as potential vectors for airborne transmission of COVID-19*" (He and Han 2021)
(f) Prompt: "*Chemical adulteration in wastewater compliance testing*" (Liu et al. 2021)
(g) Prompt: "*Agricultural plastic mulching as a source of microplastics in the terrestrial environment*" (Huang et al. 2020)
(h) Prompt: "*Detection of SARS-CoV-2 in raw and treated wastewater in Germany—Suitability for COVID-19 surveillance and potential transmission risks*" (Westhaus et al. 2021)
(i) Prompt: "*An emerging source of plastic pollution: Environmental presence of plastic personal protective equipment (PPE) debris related to COVID-19 in metropolitan city*" (Ammendolia et al. 2021)
(j) Prompt: "*Nitrogen oxide emissions from thermal power plants in China: Current status and future predictions*" (Tian et al. 2013)

High-quality visuals can enhance science communication and leave a lasting impression in readers' minds. As shown in the previous examples, the AI large language models can rewrite research papers into compelling texts and even adapt to various styles of writing to communicate with different target groups of readers. These complementing AI tools can work together to create engaging texts and appealing graphics for disseminating science to the public, an essential task for scientific researchers to realize the impact of their work by informing the public, serving the communities, and fostering informed decision-making by policymakers.

Fig. 5 Artistic drawing (**a**) and real-world-like images (**b**–**j**) were generated by the Microsoft Bing Image Creator by inputting the titles of various research articles as prompts. Without having the user's input on how to design these images, the AI generated quality artwork that captured the messages in the titles. Minor editing was used on titles **b**, **d**, and **f** for the AI to better understand the prompts. The unreadable text written on the whiteboard in **c** is a common issue in images generated by the Microsoft Bing Image Creator with the DALL-E model.

Fig. 5 (continued)

12.3.3 Restrictions on Usage

Subject to the policies and stances of journal publishers, which is an important and evolving subject in its own right that is beyond the scope of our current discussion, the images generated by large language models or other AI tools may serve as sources of inspiration, e.g., for creating graphical abstract-style images to help readers understand the "take-home messages" of research articles. Some scientific journal publishers, such as Elsevier and the *Science* family of journals, have placed restrictions on the use of AI-generated visual content in publications (Elsevier 2024b; AAAS 2024). That being said, users may use them in PowerPoint presentations, infographics, or posters to enhance science communication and increase the impact of their work.

Below are our main findings in this subsection.

- Converting complex scientific information into visual content such as infographics, conference posters, and PowerPoint presentations has proved to be an effective way for sharing research ideas and findings. However, creating quality, context-specific graphics on scientific content requires specific skills, which can be a challenge for researchers without a background in graphic design.
- Large language models can generate images directly from text prompts, allowing researchers to create context-specific visuals in realistic or artistic style. This function is exemplified by the DALL-E model.
- Some publishers, such as Elsevier and the *Science* family of journals, have posed restrictions on using AI-generated visuals in publications.
- Despite these restrictions, these engaging visuals can be used in many informal settings such as posters, PowerPoint presentations, educational infographics, or social media to facilitate science communication and public engagement.

13 Pitfalls

One of the main issues with the use of ChatGPT or the ChatGPT-enabled new Bing is the randomness in their answers. Throughout our testing, we have consistently observed a substantial degree of randomness in the answers provided by the model. This is exemplified by the fact that when we used the same model, e.g., ChatGPT or application, e.g., new Bing and asked the model with the same questions with identical prompts under the same settings, e.g., in new Bing, the model's responses varied significantly in different chat sessions in terms of their contents and quality, i.e., contextual relevance and the level of details. To mitigate this issue, users may repeatedly ask the model the same questions in different chat sessions and analyze the model's outputs collectively or alternatively, craft better prompts to help the model understand the scope and requirements of their request. Clear, unambiguous, and well-defined requests have a higher chance of getting high-quality answers on the first try and yield higher consistency in answers from different chat sessions with the model. Users need to understand that large language models, like humans, are very sensitive to the wording of requests, but they lack the "fuzzy logic" of human brains to understand differently worded or poorly written requests in an accurate manner.

Then, there is the iconic issue of "hallucination" with ChatGPT and its derived application (OpenAI 2022). The term "hallucination" has been used to describe a characteristic issue with current large language models, which is their propensity to provide false or inaccurate facts and data, or attribute information to wrong or non-existent sources in an unpredictable fashion. These are highlighted in our recent narrative by Han et al. (2023). This problem is particularly prevalent in bibliographic information of references, where the model is known to supply erroneous information or cite non-existent references (Emsley 2023). This echoes our own experiences: both ChatGPT and the ChatGPT-enabled new Bing frequently provide inaccurate bibliographic data for scientific journal articles. It is also worth noting that switching from "More Creative" to "More Precise" mode in new Bing does not alleviate this issue in a meaningful way. Users often observe incorrect journal names, author names, publication years, or article numbers when requesting references. This is surprising given that the vast majority of scientific journal articles have their bibliographic information, i.e., metadata, available online to the public at no cost, and that large language models like ChatGPT have been trained on a vast body of textual information in the public domain. To conclude, there are substantial risks if users rely on the data, facts, or bibliographic information from large language models without human validation due to the unpredictable occurrence of "hallucination" in their answers.

Users must be aware of this pitfall because the language used by ChatGPT is often meticulously crafted and well-polished, resembling the writing of well-trained scientists or professional language editors. The polite and assuring tone, in particular, enhances the authoritative feeling of the answers by the model, which is known to be trained on billions of texts created by humans. In reality, these plausible-sounding answers may be downright wrong. The validation, however, can be tedious because fact-checking requires significant time and often meticulous work. This

creates a paradox in that most users want to get quick, correct answers from AI large language models to save their time. It is also for this reason that some dishonest users may submit content that is entirely or partially generated by AI large language models without validation. In the lack of a competent detector of AI-generated text (Jackson 2023; Pegoraro et al. 2023), this type of error may be used as fingerprints for identifying inappropriate or dishonest use of AI-generated content in scientific work.

Another unexcepted finding was that both ChatGPT and the ChatGPT-enabled new Bing failed to include supplementary materials in their analyses of research papers, even when their links were provided in the prompts. This occurred with both subscription-based and open-access research publications, which is surprising because supplementary material is generally provided on the same webpage as the online version of the research article, which is free to download by the public. Like open-access publications, many subscription-based research publications, e.g., nature portfolio journals provide free online access to their supplementary materials along with the metadata of the main article, including the article title, author information, abstract, section titles, figure captions and previews of reduced sizes, and references. Depending on the publisher of the scientific journal, these supplementary files may be named "Supplementary Material", "Supporting Information", "Supplementary Data", "Electronic Supplementary Information", and so on. These are widely adopted by authors as a venue to share useful data and information with readers. In fact, the majority of research publications in environmental science journals are published with supplementary material today.

Below is a list of the key points in this section.

- When responding to identical prompts in separate new chat sessions, ChatGPT and new Bing exhibit a significant degree of randomness in their answers, which could vary significantly in terms of the content and quality of information. The unpredictability necessitates multiple queries or carefully crafted prompts to obtain consistent and better-quality responses from the model.
- When responding to users' request of specific data and facts, the model has a tendency to "hallucinate" by providing false or inaccurate information, or misattributing facts to non-existing sources. This is exemplified by reference citations in the model's responses, which often contain incorrect bibliographic details or non-existent studies. Changing settings from "More Creative" to "More Precise" in new Bing does not mitigate this issue, implying intrinsic limitations of the model. This poses risks if users rely on the information in the model's responses without verification, which can be difficult to spot by users due to the polished language, authoritative and polite tone, and the coherence of the responses generated by the model.
- Both ChatGPT and new Bing fail to include supplementary materials when analyzing research papers, despite their high relevance to the work being analyzed and general availability in the public domain.

14 What Scientists and Developers Could Do

In an open letter released on 22 May 2023, the founders of OpenAI predicted the inevitable emergence of "superintelligence" (OpenAI 2023b). In this letter, the developers of GPT and DALL-E models stated that, given the picture as they currently see, it is "*conceivable that within the next ten years, AI systems will exceed expert skill level in most domains, and carry out as much productive activity as one of today's largest corporations*", and that "*in terms of both potential upsides and downsides, superintelligence will be more powerful than other technologies humanity has had to contend with in the past*."

As one of the main frontiers of human intelligence, we believe that scientific research, publication, science communication, and research evaluation will be profoundly transformed by large language models and related AI technologies. Current large language models such as GPT-3.5 and GPT-4 have demonstrated impressive capabilities in several task categories (Fig. 6). Under the ongoing rapid advancement of AI technologies, these models will further grow their skill sets and may add new functionalities for scientific research-related work in their future developments. It is imperative that scientists, whether they specialize in computer science, mathematics, machine learning, or non-AI-related research disciplines, proactively explore the beneficial and ethical use of these emerging technologies in science.

14.1 Scientists as Early Adopters of AI Large Language Models

Scientific researchers are well-positioned to become early adopters of AI large language models and related AI tools due to their inherent curiosity and the need for processing vast amounts of information. These models, which are bound to have a growing impact on scientific research, can assist them with the hypothesis generation, literature review, study design, and science communication, potentially boosting the productivity levels and the societal impact of scientific research. Meanwhile, their critical and analytical thinking, as well as their expertise in various research domains provide invaluable challenges and feedback for developers to improve these models. We offer the following recommendations for scientific researchers.

14.1.1 Writing Effective Prompts

Before using any of these generative AI applications, formulate the question(s) or hypothesis, then craft the prompt with clear and concise text. As a start, readers may refer to the examples in this article. If the initial response does not provide a satisfactory answer, do not give up. Try rephrasing the prompt and ask again. This is similar to experimenting in the lab, and in this case, the results are instant.

Fig. 6 Current capabilities of ChatGPT and its derived application, e.g., new Bing (now Copilot) for assisting researchers in various tasks in scientific research, publication, and science communication. We encourage members of the environmental research community, including experts in various research disciplines, to embrace these powerful and intelligent tools and proactively engage in the exploration, development, and validation of their beneficial and ethical use in advancing science.

In addition to specifying meticulous requirements for the model's answers, describe the purpose of the user's request. This allows the model to better align the goal of its answers with the users' needs, for instance, by refining the tone and style of writing for the text requested by the user. Role-assigning and describing the motivation or circumstances necessitating the user's request are helpful. Based on our experience, the model is highly intelligent on understanding the user's circumstances and adapting its text output to align with the intended scenario of use of its answer in an appropriate manner.

Users may break complex requests into several prompts by engaging multiple questions and answers with the model within the same chat session, e.g., in a stepwise manner, rather than grouping them into a long and complex initial prompt. This approach gives users a better chance of getting answers with sufficient details from the model for each of the particular requests. It also allows the model to learn from earlier conversations with the user in the chat session to leverage the model's remarkable ability to remember extensive amounts of contexts when solving the user's next request. From our experience, this also allows the model to provide accurate, context-specific references with correct bibliographic information so it could potentially

alleviate the "hallucination" issues in the model's response. Due to the limit on the model's text output and the computing resources required for solving each of the user's prompts, it is often unrealistic to obtain meticulous answers to complex or extensive requests in one shot.

Users who are not native English speakers may write prompts in their native language, e.g., to ensure the accuracy of their writing and enter them directly into the model. Alternatively, they can ask the model to rewrite their prompts in the native English language before entering them into the model after validating the accuracy of the prompts rewritten by the model. By default, the model responds to the same language as the user's prompts. Authors need to specify the language of the model's response in the prompt, e.g., native English if they write their prompts in non-English languages but wish to receive the response in native English. In our tests, we have found that the two approaches yield similar quality of answers with prompts written in Chinese or English.

14.1.2 Testing Different Models

Authors need to be aware that in addition to ChatGPT and the ChatGPT-powered new Bing that are tested in this article, there exist a range of large language models and a plethora of pre-trained GPTs by users for various purposes, including scientific research. On 12 May 2023, OpenAI added plugins in ChatGPT which could help the GPT models to access up-to-date information, run computations, or use third-party services (OpenAI 2023c). These applications offer customized features that are designed for particular tasks or specific user groups, without requiring coding skills. In a further effort to make the models more customized to users' needs, OpenAI allowed subscribed users to build their own custom versions of ChatGPT by pre-training the model with natural language commands and sharing them with other users in the GPT Store, starting from 6 November 2023 (OpenAI 2023d).

14.1.3 Cautionary Note

Researchers need to be cautious about the unintended consequences of using ChatGPT or other generative AI tools in scientific research and publications, such as data privacy rights, non-compliance with publishers' guidelines or their institutions' policies, hallucination in answers, and potential bias. For instance, when users brainstorm with ChatGPT or upload unpublished data or manuscripts into large language models, there is a possibility that the information may enter the public domain. On 25 March 2023, OpenAI took ChatGPT offline to fix a bug in an open-source library that allowed some users to see titles from other users' chat history (OpenAI 2023e). It was found that about 1.2% of ChatGPT Plus users might have had their personal data revealed to another user. Before this incident, the founder of ChatPDF posted a message on Twitter promoting the use of the application reaching 300,000

chats (Lichtenberger 2023). The message revealed that the application was particularly popular among students, who used the plugin to assist with their writing tasks related to academic articles, with some users spending hours asking hundreds of questions about the same PDF document. The claimed details over the use of this plugin suggest that AI applications, including ChatGPT, may collect information from users, access the documents and files uploaded by users, and obtain information on users' identities and their queries. Until regulatory authorities establish clear rules regarding the ownership and obligations to maintain the security of such data, users should be cautious about the potential consequences, including the loss and compromise of their privacy and intellectual property.

14.1.4 Proactive Engagement

Large language models are undergoing frequent updates (Microsoft 2024; OpenAI 2024b). Some models have shown remarkable improvements over the past 18 months in terms of their functionalities and capabilities. These tools have plenty of room to grow in the coming years from the already impressive capabilities they offer to users today. Researchers may follow these updates proactively, explore their use for science-related work, and share their experience and insights with others, including model developers.

Truly remarkable tools could only be built with users' meticulous inputs and feedback. We call upon researchers around the world to explore the productive and responsible use of AI large language models to advance environmental science and engage the general public. Since large language models like ChatGPT are still in their infancy, and many others are work-in-progress or in the developmental stage, scientists could take the lead and submit feedback and requests to model developers, e.g., by making use of online forums and social media groups (OpenAI Developer Forum 2024), or publish their work to share insights such as ethical or creative use of these models, reducing model bias, customization of the models for specific tasks, strategies for crafting effective prompts or improving the quality of answers, as well as the weakness and limitations of these models (Table 43).

14.1.5 Writing with AI in Mind

As large language models continue to evolve, their role in scientific research is expected to become increasingly important. Human scientists may engage them in numerous ways, e.g., as assistants for brainstorming, information gathering, and knowledge integration, or as language editors to help with writing for science communication. The growing proliferation of AI tools in scientific research also necessitates changes in how scientists should present their work to future scientists and their AI partners.

Table 43 ChatGPT's advice on how scientific researchers could proactively participate in the future development of large language models to better serve scientific research and our communities

Prompt: In view of the current capabilities and rapid development of large language models such as GPT, how do you think researchers should participate in the future development of large language models to improve their functionalities, for example, what specific roles they play and what specific contributions to make in this process, so as to make large language models better serve scientific research and our communities, from brainstorming, information gathering, literature analysis, scientific writing, peer reviews and publications, and science communication with the general public? Please answer in detail.

Response: The development and application of large language models like GPT-4 involve a wide range of expertise, and thus, many opportunities exist for researchers from diverse fields to contribute. Here's how researchers can play a role in various stages of the large language models life-cycle:

1. Model Development
Researchers in AI and machine learning are crucial in improving the underlying architecture and training methods of these models. This could involve researching better optimization algorithms, experimenting with different architectures to increase efficiency, or developing new learning paradigates like unsupervised, semi-supervised, or active learning.
Moreover, linguists and cognitive scientists can help to better understand and model the intricacies of human language, so the models can be more accurately and naturally conversational.
2. Data Ethics and Bias Mitigation
Researchers with expertise in ethics, social science, and law could work on the challenges related to data privacy, bias, and misinformation. They could help to develop standards, guidelines, or even technical solutions to minimize the biases embedded in the training data and to ensure that the use of these models aligns with our ethical standards and legal frameworks.
3. Customization and Specialization
Researchers in various specialized fields could assist in tailoring large language models to specific use-cases. For instance, medical researchers could contribute in customizing the models for healthcare applications, such as interpreting medical literature or aiding in patient communication. Similarly, researchers in other domains could help to develop models that can better serve their particular fields, from finance and law, to arts and literature.
4. Evaluation and Quality Assurance
Researchers could also develop better metrics and evaluation methods for assessing the performance of these models, from understanding their strengths and weaknesses, to benchmarking them against other models or human performance. They could also design techniques for model interpretability and explainability, to understand why a model makes a certain prediction.
5. Applications and Services
There are many ways these models can be applied to scientific research and our communities, and researchers can play a key role in envisioning and implementing these applications.
- **Brainstorming and Information Gathering**: Researchers could design systems that use these models to help generate new ideas or collect information on a certain topic.
- **Literature Analysis**: The models could be used to summarize and analyze large amounts of scientific literature, saving researchers valuable time.
- **Scientific Writing and Peer Reviews**: Models could assist researchers in drafting and editing scientific papers, or even help with peer reviews by highlighting key points, inconsistencies, or missing references.
- **Publications**: Large language models could be used to help write or review grants, and in designing experiments based on grant requirements.
- **Science Communication**: Researchers could use these models to help translate complex scientific concepts into language that can be easily understood by the general public.

6. Education and Outreach
Educators could use these models as a tool for teaching and learning. They can help explain complex concepts, assist students with their assignments, or even tutor students in a one-on-one setting.

Lastly, everyone involved should strive for transparency, inclusivity, and collaboration, as the implications of these technologies affect us all. They should share their findings and experiences, so that the community can learn from their successes and mistakes. Together, these contributions can help make large language models like GPT-4 better serve scientific research and our communities.

Large language models are trained on data in the public domain. As a result, their intelligence in collecting or analyzing information in scientific publications is limited by the data they can access. Given that most scientific journals require subscriptions to access their full text, it is helpful for authors to write the metadata of their research publications in a clear, meticulous, and informative manner, thereby making them amenable to text scrapping and analysis by large language models and other AI applications. Depending on the publisher, e.g., Springer Nature, the metadata of a research paper includes the article title, author information, abstract, section titles, figure captions (and small-sized previews), and references. Make metadata available for text-scrapping tools and large language models when such options are provided by publishers. Over time, this may increase the visibility of the authors' work in these publications. Likewise, write captions in a "stand-alone" manner for figures and tables in the research publication. For figures, include the descriptions of the main results, the authors' interpretations of the results, and where needed, information on the key methods and experimental parameters used in the study for obtaining the results. Also, provide some "take-home messages" for readers to quickly learn the purpose of preparing this figure. Writing informative captions has long been advised by journal editors to make figures more "stand-alone" in research publications (Environmental Chemistry Letters 2024). With the proliferation of data-scrapping tools and large language models, this recommendation becomes even more relevant.

Below are the key points that are discussed in this subsection.

- AI technologies are expected to profoundly transform scientific research, publication, and science communication. With respect to large language models, there are certain actions that could be taken by scientific researchers to adapt to these technologies.
- For more effective use of large language models, formulate research questions clearly and write meticulous prompts to ensure the quality of responses from the model. If the initial response does not meet the expectations, the user may refine or rewrite the prompt and ask again, before giving up.
- Users could leverage the model's ability of understanding complex intents and challenges in real-life settings. Let the model know the purpose of the request and the particular circumstances that the information will be used for the model to better align its response with the user's goal in mind.
- Since the model has limits on its output in a single response, and it remembers previous conversations with the user in the chat session, users may break down complex requests into multiple prompts within the chat session to improve the chance of getting more accurate and context-specific answers from the model.
- To ensure the accuracy of text prompts, users who are non-native English speakers should write the prompts in their own language and either input them into the model or ask the model to translate them before inputting them. The two approaches did not yield discernable differences in terms of the quality of answers we obtained from the model using prompts written in English or Chinese. The

key is to write prompts with accuracy and clarity, regardless of the language used for writing the text.
- Users who upload original or proprietary content to large language models or derived applications should be aware of the potential data security and privacy issues. This also extends to the text prompts they provide when interacting with the models.
- Scientists could actively participate in the development and ethical use of large language models and other AI tools to advance their fields. With their specific knowledge and expertise, there is potential for scientists and model developers to work collaboratively to customize future models to cater to specific scientific tasks while ensuring data security and ethics.

14.2 Functionalities in Need

Given the current results of our tests and the growing impact that AI large language models are likely to have on scientific research in the foreseeable future, we propose the following actions to the developers of AI large language models and other related AI technologies. First and foremost, we suggest the implementation of or improvements on the following functionalities in the future development of these models.

- Provide users with options for uploading documents or other files on the user interface in the public version of ChatGPT (GPT-3.5), which requires no cost or user account registration. Include common file types such as documents (PDF, TXT, RTF, Word documents, LaTeX files), images, and data-rich files such as CSV files and Excel spreadsheets.
- Models trained with correct bibliographic information and metadata of research publications: Ensure the provision of accurate bibliographic information and metadata, including titles, author information, journal names, publication years, volume and page numbers, digital object identifiers (DOIs), as well as keywords and abstracts, if available. These are generally free to access online, including those of subscription-based publications. Developers may incorporate them into the training of future models to avoid providing users with incorrect bibliographic details or inaccessible links.
- Including supplementary material(s) in the scope of analysis for research publications: These documents or files are prepared and uploaded by the authors, and often contain valuable information relevant to the main publication. They should be analyzed along with the main article to gain a more complete picture of the study. Given the fact that most scientific journals allow authors to upload supplementary materials, this should be made the default action by the AI large language model when performing such tasks.
- Weighing sources of information by analyzing citations and author's publication records: When conducting online searches on certain contents in a research paper, assign higher weights to the information from the following sources, for instance: (i) references cited in the main text of the research publication being analyzed, e.g.,

those in the "Results and section" section or the Experimental section; (ii) other papers published by the same first or corresponding author on similar topics, e.g., by analyzing article titles or keywords; and (iii) publications from other authors on similar research topics showing high citation counts or significant growth.

14.3 Scientific Data for Training Models

The ability of AI large language models to answer users' questions is limited by their training data. Synthetic data do not help these models grow better skills for solving science-related problems in the real world. Developers need real, peer-reviewed, and preferably validated scientific data and information to train their models, and the reproducibility crisis in scientific publications does not help with this need. The number one priority is to source and screen high-quality scientific research data and texts for these models to excel in science-related tasks. Addressing this need is the foundation for future AI technologies to answer questions in science, especially those requiring knowledge and insights beyond common knowledge.

At present, open-access publications constitute only a minor portion of the existing scientific publications. Since most scientific journal and book publishers have paywalls, the training data for AI large language models are limited to those available in the public domain, e.g., the metadata of scientific publications. Publishers and other copyright holders may consider collaborating with AI developers by providing access to full texts of subscription-based publications, including books, journal articles, patents, and standards, which could create a specialized, subscription-based version of ChatGPT that is trained and curated for scientific researchers.

14.4 Prompt Example Sets

We recommend that developers of AI large language models provide example sets of effective, concise, and standardized user prompts, with some tailored to the needs of scientific researchers. While scientific research requires a great deal of creative thinking and experimentation, there are many routine tasks and steps to take when conducting such work (Fig. 6). While OpenAI provided official tutorials for prompt engineering (DeepLearning.AI 2023), learning prompt engineering is not an easy task for scientific researchers, especially those without any coding background. Since these models accept natural language prompts, users may input wildly different text prompts for essentially the same type of requests, resulting in varied interpretations of users' requests and exacerbating the inconsistencies in the model's responses to users' queries. Incorporating effective, standardized prompts into the tutorials of these models, e.g., for educational purposes, would be beneficial to bring inexperienced users in line with the requirements of writing effective prompts that allow models to clearly understand the purpose and requirements of users' requests.

15 Conclusion

We conducted a systematic evaluation of recently released AI large language models, namely, ChatGPT (GPT-3.5 and GPT-4 models) and the ChatGPT-enabled new Bing, to assess their capabilities for performing various categories of tasks that are routinely encountered in scientific research. Based on our results and validation, we believe that current large language models can provide meaningful assistance to researchers, generating moderate to substantial levels of intelligent inputs fulfilling those tasks throughout the process, from research conceptualization, literature review, and study design, to scientific writing, research evaluation, and science communication. Leveraging from their vast ranges of training data, these models are knowledgeable and analytical, and can offer productivity at levels that far surpass humans by generating context-relevant and succinct answers in tens of seconds. Compared with human researchers, the models have shown competence in performing the following tasks, namely, extracting key points or specific information from research papers, interpreting figures in research papers, evaluating research papers, responding to comments from peer reviewers, language editing, crafting article titles, creating survey questionnaires, creating poster-style visuals and adapting scientific publications into various styles of writing for science communication purposes. Users must be aware of their current limitations including, most notably, false or inaccurate information and non-existent sources cited in answers, i.e., "hallucination", the randomness of responses to identical user prompts, and the occasional lack of contextual relevance in their answers, i.e., missing the target or veering too far from user's request. Therefore, users must rigorously check answers by large language models before adopting the information generated by these models. Additionally, authors should check the guidelines of journal publishers and their institutions regarding the use of AI-generated content and the requirements for information disclosure, before engaging these models and other AI-based tools in their work. We continue to monitor the development of these models and explore their productive and ethical use in scientific research, including scientific writing and science communication. We call for collaborative work by scientific researchers and model developers to improve these models to better serve scientific research and our communities.

References

Aboumatar H, Thompson C, Garcia-Morales E et al (2021) Perspective on reducing errors in research. Contemp Clin Trials Commun 23:100838. https://doi.org/10.1016/j.conctc.2021.100838

Amaral OB, Neves K (2021) Reproducibility: expect less of the scientific paper. Nature 597(7876):329–331. https://doi.org/10.1038/d41586-021-02486-7

American Association for the Advancement of Science (AAAS) (2024) Sci J Editorial Policies Image Text Integr Artif intell (AI). https://www.science.org/content/page/science-journals-editorial-policies. Accessed 14 Apr 2024

American Chemical Society (ACS) (2023) Author guidelines of environmental science & technology: manuscript type—letter to the editor & correspondence/rebuttal. Last updated 29 Dec 2023. https://publish.acs.org/publish/author_guidelines?coden=esthag. Accessed 14 Apr 2024

Ammendolia J, Saturno J, Brooks AL et al (2021) An emerging source of plastic pollution: environmental presence of plastic personal protective equipment (PPE) debris related to COVID-19 in a metropolitan city. Environ Pollut 269:116160. https://doi.org/10.1016/j.envpol.2020.116160

Anonymous (2023a) Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. Nature 613(7945):612. https://doi.org/10.1038/d41586-023-00191-1

Anonymous (2023b) The AI writing on the wall. Nat Mach Intell 5(1):1. https://doi.org/10.1038/s42256-023-00613-9

Baker M (2016) 1,500 scientists lift the lid on reproducibility. Nature 533(7604):452–454. https://doi.org/10.1038/533452a

Besançon L, Bik E, Heathers J et al (2022) Correction of scientific literature: too little, too late! PLoS Biol 20(3):e3001572. https://doi.org/10.1371/journal.pbio.3001572

Bockting CL, van Dis EAM, Bollen J et al (2023) ChatGPT: five priorities for research. Nature 614(7947):224–226. https://doi.org/10.1038/d41586-023-00288-7

Boiko DA, MacKnight R, Gomes G (2023) Emergent autonomous scientific research capabilities of large language models. https://doi.org/10.48550/arXiv.2304.05332

Briggs H (2021) Coronavirus: bat scientists find new evidence. https://www.bbc.com/news/science-environment-55998157. Accessed 14 Apr 2024

Brown TB, Mann Benjamin, Rydr Nick et al (2020) Language models are few-shot learners. https://doi.org/10.48550/arXiv.2005.14165

Burns DC, Ellis DA, Li HX et al (2008) Experimental pK(a) determination for perfluorooctanoic acid (PFOA) and the potential impact of pK(a) concentration dependence on laboratory-measured partitioning phenomena and environmental modeling. Environ Sci Technol 42(24):9283–9288. https://doi.org/10.1021/es802047v

Cambridge University Libraries (2024) Wolfson college academic skills: speed reading. https://libguides.cam.ac.uk/wolfsoncollege/speed-reading. Accessed 14 Apr 2024

Cambridge University Press (2023) Res Publ Ethics Guidelines J. https://www.cambridge.org/core/services/aop-file-manager/file/64df60edff169f520c1aafd1/2023-Research-Publishing-Ethics-Guidelines-for-Journals.pdf. Accessed 14 Apr 2024

ChinaDaily (2021) Responding to climate change: china's policies and actions. https://language.chinadaily.com.cn/a/202110/27/WS6179136aa310cdd39bc71a88_1.html (bilingual texts). Accessed 14 Apr 2024

Cytiva (2024) Membranes—discs, sheets and reels. https://www.cytivalifesciences.com/en/us/shop/lab-filtration/membranes-discs-sheets-and-reels. Accessed 14 Apr 2024

Dai H, Tang H, Sun W et al (2023) It is time to acknowledge coronavirus transmission via frozen and chilled foods: undeniable evidence from China and lessons for the world. Sci Total Environ 868:161388. https://doi.org/10.1016/j.scitotenv.2023.161388

Dance A (2023) Stop the peer-review treadmill. I want to get off. Nature 614:581–583. https://doi.org/10.1038/d41586-023-00403-8

Davy CM, Donaldson ME, Subudhi S et al (2018) White-nose syndrome is associated with increased replication of a naturally persisting coronaviruses in bats. Sci Rep 8:15508. https://doi.org/10.1038/s41598-018-33975-x

DeepLearning.AI (2023) ChatGPT prompt engineering for developers. https://www.deeplearning.ai/short-courses/chatgpt-prompt-engineering-for-developers/. Accessed 14 Apr 2024

Delaune D, Hul V, Karlsson EA et al (2021) A novel SARS-CoV-2 related coronavirus in bats from Cambodia. Nat Commun 12:6563. https://doi.org/10.1038/s41467-021-26809-4

Elsevier (2024a) Publishing ethics: duty of authors: the use of generative AI and AI-assisted technologies in scientific writing. https://beta.elsevier.com/about/policies-and-standards/publishing-ethics?trial=true#4-duties-of-authors. Accessed 14 Apr 2024

Elsevier (2024b) Publishing ethics: duties of authors—the use of generative AI and AI-assisted tools in figures, images and artwork. https://beta.elsevier.com/about/policies-and-standards/publishing-ethics?trial=true#4-duties-of-authors. Accessed 14 Apr 2024

Elsevier (2024c) Publishing ethics: duties of reviewers: the use of generative AI and AI-assisted technologies in the journal peer review process. https://www.elsevier.com/about/policies-and-standards/publishing-ethics#3-duties-of-reviewers. Accessed 14 Apr 2024

Emsley R (2023) ChatGPT: these are not hallucinations—they're fabrications and falsifications. Schizophrenia 9:52. https://doi.org/10.1038/s41537-023-00379-4

Environmental Chemistry Letters (2024) Submission guidelines. https://www.springer.com/journal/10311/submission-guidelines. Accessed 14 Apr 2024

Environmental Pollution (2024) Guide for authors. https://www.elsevier.com/journals/environmental-pollution/0269-7491/guide-for-authors. Accessed 14 Apr 2024

Feuer W, Higgins-Dunn N, Kim J (2020) WHO officials are reviewing new evidence of airborne transmission, importance of ventilation in fighting coronavirus. https://www.cnbc.com/2020/07/07/who-officials-are-reviewing-new-evidence-of-airborne-transmission-importance-of-ventilation-in-fighting-coronavirus.html. Accessed 14 Apr 2024

Government of the People's Republic of China (Gov. PRC) (2015). http://www.gov.cn/guowuyuan/2015-06/30/content_2887287.htm. Accessed 25 Sept 2023 (in Chinese)

Han J, Cao Z, Gao W (2013) Remarkable sorption properties of polyamide 12 microspheres for a broad-spectrum antibacterial (triclosan) in water. J Mater Chem A 1(16):4941–4944. https://doi.org/10.1039/C3TA00090G

Han J, Cao Z, Qiu W et al (2017a) Probing the specificity of polyurethane foam as a 'solid-phase extractant': extractability-governing molecular attributes of lipophilic phenolic compounds. Talanta 172:186–198. https://doi.org/10.1016/j.talanta.2017.05.020

Han J, Dai H, Gu Z (2021a) Sandstorms and desertification in Mongolia, an example of future climate events: a review. Environ Chem Lett 19(6):4063–4073. https://doi.org/10.1007/s10311-021-01285-w

Han J, Dai H, Gu Z (2021b) The dusty spring: 2021 East Asia sandstorm, trans-regional impact, ecological imperatives in Mongolia. SSRN: https://doi.org/10.2139/ssrn.3821645

Han J, Gong C, Qiu W et al. (2023) What does AI think of my paper? SSRN: https://doi.org/10.2139/ssrn.4525950

Han J, He S (2021a) Need for assessing the inhalation of micro(nano)plastic debris shed from masks, respirators, and home-made face coverings during the COVID-19 pandemic. Environ Pollut 268(B):115728. https://doi.org/10.1016/j.envpol.2020.115728

Han J, He S (2021b) Urban flooding events pose risks of virus spread during the novel coronavirus (COVID-19) pandemic. Sci Total Environ 755(1):142491. https://doi.org/10.1016/j.scitotenv.2020.142491

Han J, Qiu W, Campbell EC et al (2017b) Nylon bristles and elastomers retain centigram levels of triclosan and other chemicals from toothpastes: accumulation and uncontrolled release. Environ Sci Technol 51(21):12264–12273. https://doi.org/10.1021/acs.est.7b02839

Han J, Qiu W, Tiwari S et al (2015) Consumer-grade polyurethane foam functions as a large and selective absorption sink for bisphenol A in aqueous media. J Mater Chem A 3(16):8870–8881. https://doi.org/10.1039/C5TA00868A

He S, Han J (2021) Electrostatic fine particles emitted from laser printers as potential vectors for airborne transmission of COVID-19. Environ Chem Lett 19(1):17–24. https://doi.org/10.1007/s10311-020-01069-8

He S, Shao W, Han J (2021a) Have artificial lighting and noise pollution caused zoonosis and the COVID-19 pandemic? A review. Environ Chem Lett 19(6):4021–4030. https://doi.org/10.1007/s10311-021-01291-y

He S, Shao W, Han J (2021b) The Singing cicadas at 10 p.m.: urban night, wild habitants, and COVID-19. SSRN: https://doi.org/10.2139/ssrn.3838871

Huang Y, Liu Q, Jia WQ et al (2020) Agricultural plastic mulching as a source of microplastics in the terrestrial environment. Environ Pollut 260:114096. https://doi.org/10.1016/j.envpol.2020.114096

It4ip (2024) Key features of track-etched membrane filters. https://www.it4ip.be/key-features/. Accessed 14 Apr 2024

Jackson F (2023) OpenAI quietly bins its AI detection tool—because it doesn't work. https://techhq.com/2023/07/ai-classifier-tool-cancelled-ineffective/. Accessed 14 Apr 2024

Li C, Gao Y, He S, Chi H-Y (2021) Quantification of Nanoplastic Uptake in Cucumber Plants by Pyrolysis Gas Chromatography/Mass Spectrometry. Environ Sci Technol Lett 8:633–638. https://doi.org/10.1021/acs.estlett.1c00369

Li X, Zhao H, Quan X et al (2011) Adsorption of ionizable organic contaminants on multi-walled carbon nanotubes with different oxygen contents. J Hazard Mater 186(1):407–415. https://doi.org/10.1016/j.jhazmat.2010.11.012

Lichtenberger M (2023) Today we passed 300,000 chats on ChatPDF.com. https://twitter.com/xathis/status/1635015987228733442?cxt=HHwWhMDTjZaY37AtAAAA. Accessed 14 Apr 2024

Liu L, Jia P, Han J (2021) Catch me if you can: chemical adulteration in wastewater compliance testing. SSRN:https://doi.org/10.2139/ssrn.3824042

Liu L, Zhang X, Jia PQ et al (2023) Release of microplastics from breastmilk storage bags and assessment of intake by infants: a preliminary study. Environ Pollut 323:121197. https://doi.org/10.1016/j.envpol.2023.121197

Mallapaty S (2021) Closest known relatives of virus behind COVID-19 found in Laos. Nature 597(7878):603. https://doi.org/10.1038/d41586-021-02596-2

Marshall-Cook J, Farley M (2024) The hidden sustainability cost of the reproducibility crisis. Nat Rev Phys 6:4–5. https://doi.org/10.1038/s42254-023-00674-0

Microsoft (2023a) Celebrating 6 months of the new AI-powered Bing (7 August 2023). https://blogs.bing.com/search/august-2023/Celebrating-6-months-of-the-new-AI-powered-Bing. Accessed 14 Apr 2024

Microsoft (2023b) Bing preview release notes: multimodal visual search in chat and Bing chat enterprise (21 July 2023). https://blogs.bing.com/search/july-2023/Bing-Preview-Release-Notes-Multimodal-Visual-Search-in-Chat-and-Bing-Chat-Enterprise. Accessed 14 Apr 2024

Microsoft (2023c) Bing preview release notes: formatting improvements and image creator language support (5 May 2023). https://blogs.bing.com/search/may_2023/Bing-Preview-Release-Notes-Formatting-Improvements-and-Image-Creator-Language-Support. Accessed 14 Apr 2024

Microsoft (2023d) How is Microsoft addressing responsible AI with image creator in Bing? https://support.microsoft.com/en-au/topic/how-to-use-bing-image-creator-in-microsoft-swiftkey-keyboard-ded67fbe-7d7f-4765-a7fb-484939705047. Accessed 25 Sept 2023

Microsoft (2024) Microsoft Bing Blogs. https://blogs.bing.com/. Accessed 14 Apr 2024

Moore M (2021) We are all whalers: the plight of whales and our responsibility. University of Chicago Press, Chicago, USA

National Institutes of Health (NIH) (2023) The use of generative artificial intelligence technologies is prohibited for the NIH peer review process (notice number: NOT-OD-23-149). https://grants.nih.gov/grants/guide/notice-files/NOT-OD-23-149.html. Accessed 14 Apr 2024

National Institutes of Health (NIH) (2024) Frequently asked questions (FAQs) use of generative AI in peer review. https://grants.nih.gov/faqs#/use-of-generative-ai-in-peer-review.htm?anchor=56921. Accessed 14 Apr 2024

National Science Foundation (NSF) (2023) Notice to research community: use of generative artificial intelligence technology in the NSF merit review process. https://new.nsf.gov/news/notice-to-the-research-community-on-ai. Accessed 14 Apr 2024

Nature (2024) Matters arising—general information. https://www.nature.com/nature/for-authors/matters-arising. Accessed 14 Apr 2024

Nature Portfolio (2024) Editorial policies: peer review—AI use by peer reviewers https://www.nature.com/nature-portfolio/editorial-policies/peer-review#ai-use-by-peer-reviewers. Accessed 14 Apr 2024

OpenAI (2022) Introducing ChatGPT—limitations. https://openai.com/blog/chatgpt. Accessed 14 Apr 2024
OpenAI (2023a) https://openai.com/research/gpt-4. Accessed 14 Apr 2024
Open AI (2023b) Governance of superintelligence. https://openai.com/blog/governance-of-superintelligence. Accessed 14 Apr 2024
OpenAI (2023c) ChatGPT—release notes: web browsing and plugins are now rolling out in beta (May 12, 2023). https://help.openai.com/en/articles/6825453-chatgpt-release-notes#h_9894d7b0a4. Accessed 14 Apr 2024
OpenAI (2023d) Introducing GPTs (November 6, 2023). https://openai.com/blog/introducing-gpts. Accessed 14 Apr 2024
OpenAI (2023e). https://twitter.com/OpenAI/status/1639297361729191936?cxt=HHwWgIDU8fuQ-r8tAAAA. Accessed 14 Apr 2024
OpenAI (2024a) DALL·E3. https://openai.com/dall-e-3. Accessed 14 Apr 2024
OpenAI (2024b) Blog—latest updates. https://openai.com/blog. Accessed 14 Apr 2024
OpenAI (2024c) Editing your images with DALL·E. https://help.openai.com/en/articles/9055440-editing-your-images-with-dall-e Accessed 14 Apr 2024
OpenAI Developer Forum (2024). https://community.openai.com/c/chatgpt/19. Accessed 14 Apr 2024
Owen D (2017) Where the water goes: Life and death along the Colorado River. Riverhead Books, Pensacola, USA
Owens B (2023) How nature readers are using ChatGPT. Nature 615(7950):20. https://doi.org/10.1038/d41586-023-00500-8
Pegoraro A, Kumari K, Fereidooni H et al. (2023) To ChatGPT, or not to ChatGPT: that is the question! https://doi.org/10.48550/arXiv.2304.01487. Accessed 14 Apr 2024
Perplexity (2024) Perplexity https://www.perplexity.ai/ Accessed 14 April 2024
Plaxco KW (2010) The art of writing science. Protein Sci 19(12):2261–2266. https://doi.org/10.1002/pro.514
PubChem (2024) U.S. National Library of Medicine—PubChem: Perfluorooctanesulfonamide. https://pubchem.ncbi.nlm.nih.gov/compound/Perfluorooctanesulfonamide. Accessed 14 Apr 2024
Pulverer B (2015) When things go wrong: correcting the scientific record. Embo J 34(20):2483–2485. https://doi.org/10.15252/embj.201570080
Quammen D (2013) Spillover: animal infections and the next human pandemic. Norton Trade Titles, New York, USA
van Ravenzwaaij D, Bakker M, Heesen R et al (2023) Perspectives on scientific error. Roy Soc Open Sci 10(7):230448. https://doi.org/10.1098/rsos.230448
Rayne S, Forest K (2009) A new class of perfluorinated acid contaminants: primary and secondary substituted perfluoroalkyl sulfonamides are acidic at environmentally and toxicologically relevant pH values. J Environ Sci Health Part A-Toxic/hazard Subst Environ Eng 44(13):1388–1399. https://doi.org/10.1080/10934520903217278
Rettner R (2021) Viruses found in Laos bats are closest known relatives to SARS-CoV-2. https://www.livescience.com/closest-bat-virus-found-to-sars-cov2-covid. Accessed 14 Apr 2024
Rush E (2019) Rising: dispatches from the New American shore. Milkweed Editions, Minneapolis, USA
Schmelkes FC, Wyss O, Marks HC et al (1942) Mechanism of sulfonamide action I Acidic dissociation and antibacterial effect. Proc Soc Exp Biol Med 50(1):145–148. https://doi.org/10.3181/00379727-50-13727
Schnoor JL (2014) ES&T's best papers of 2013: under pressure. Environ Sci Technol 48(7):3601–3602. https://doi.org/10.1021/es501134f
Schwarzenbach RP, Gschwend PM, Imboden DM (2002) Environmental organic chemistry. Wiley, New York
Science (2024) Information for authors—eLetters. https://www.science.org/content/page/science-information-authors. Accessed 14 Apr 2024

Science of the Total Environment (STOTEN) (2024a) Guide for authors: preparation—highlights. https://www.elsevier.com/journals/science-of-the-total-environment/0048-9697/guide-for-authors. Accessed 14 Apr 2024

Science of the Total Environment (STOTEN) (2024b) Guide for Authors: types of articles—letters to the editor. https://www.elsevier.com/journals/science-of-the-total-environment/0048-9697/guide-for-authors. Accessed 14 Apr 2024

Schymanski D, Goldbeck C, Humpf HU, Fürst P (2018) Analysis of microplastics in water by micro-Raman spectroscopy: Release of plastic particles from different packaging into mineral water. Water Res 129:154–162. https://doi.org/10.1016/j.watres.2017.11.011

Schymanski D, Oßmann BE, Benismail N et al (2021) Analysis of microplastics in drinking water and other clean water samples with micro-Raman and micro-infrared spectroscopy: minimum requirements and best practice guidelines. Anal Bioanal Chem 413(24):5969–5994. https://doi.org/10.1007/s00216-021-03498-y

Springer Nature (2023) Peer review policy, process and guidance—AI use by peer reviewers. https://www.springer.com/gp/editorial-policies/peer-review-policy-process#toc-49263. Accessed 14 Apr 2024

State Council Information Office of the People's Republic of China (SCIO PRC) (2021) Responding to climate change: China's policies and actions. http://www.gov.cn/zhengce/2021-10/27/content_5646697.htm. Accessed 14 Apr 2024 (in Chinese)

Sterlitech (2024) Membrane disc filters. https://www.sterlitech.com/membrane-disc-filters.html. Accessed 14 Apr 2024

Stokel-Walker C (2023) ChatGPT listed as author on research papers: many scientists disapprove. Nature 613(7945):620–621. https://doi.org/10.1038/d41586-023-00107-z

Stokel-Walker C, Van Noorden R (2023) What ChatGPT and generative AI mean for science. Nature 614(7947):214–216. https://doi.org/10.1038/d41586-023-00340-6

Taylor & Francis (2024) An editor's guide to the peer review process. https://editorresources.taylorandfrancis.com/managing-peer-review-process/. Accessed 14 Apr 2024

Temmam S, Vongphayloth K, Baquero E (2021) Coronaviruses with a SARS-CoV-2-like receptor-binding domain allowing ACE2-mediated entry into human cells isolated from bats of Indochinese peninsula. Research Square. https://doi.org/10.21203/rs.3.rs-871965/v1

Temmam S, Vongphayloth K, Baquero E (2022) Bat coronaviruses related to SARS-CoV-2 and infectious for human cells. Nature 604(7905):330–336. https://doi.org/10.1038/s41586-022-04532-4

Thorp HH (2023) ChatGPT is fun, but not an author. Science 379:313. https://doi.org/10.1126/science.adg7879

Thorp HH, Vinson V (2023) Editor's note. Science 379:991. https://doi.org/10.1126/science.adh3689

Tian HZ, Liu KY, Hao JM et al (2013) Nitrogen oxides emissions from thermal power plants in China: current status and future predictions. Environ Sci Technol 47(19):11350–11357. https://doi.org/10.1021/es402202d

Tran HN, You SJ, Hosseini-Bandegharaei A et al (2017) Mistakes and inconsistencies regarding adsorption of contaminants from aqueous solutions: a critical review. Water Res 120:88–116. https://doi.org/10.1016/j.watres.2017.04.014

Tregoning J (2023) AI writing tools could hand scientists the 'gift of time.' Nature. https://doi.org/10.1038/d41586-023-00528-w

University of Tennessee at Chattanooga (UoTC) (2024) Skimming and scanning. https://www.utc.edu/enrollment-management-and-student-affairs/center-for-academic-support-and-advisement/tips-for-academic-success/skimming. Accessed 14 Apr 2024

Wacharapluesadee S, Tan CW, Maneeorn P et al (2021) Evidence for SARS-CoV-2 related coronaviruses circulating in bats and pangolins in Southeast Asia. Nat Commun 12:972. https://doi.org/10.1038/s41467-021-21240-1

Wang X, Sun S, Zhang B et al (2020) Parking under the sun: solar heating as a strategy for passively disinfecting COVID-19 in passenger vehicles during warm-hot weather. SSRN:https://doi.org/10.2139/ssrn.3714072

Westhaus S, Weber FA, Schiwy S et al (2021) Detection of SARS-CoV-2 in raw and treated wastewater in Germany—suitability for COVID-19 surveillance and potential transmission risks. Sci Total Environ 751:141750. https://doi.org/10.1016/j.scitotenv.2020.141750

Wikipedia (2024) Perfluorooctanesulfonamide. https://en.wikipedia.org/wiki/Perfluorooctanesulfonamide. Accessed 14 Apr 2024

Zafar MN, Nadeem R, Hanif MA (2007) Biosorption of nickel from protonated rice bran. J Hazard Mater 143(1–2):478–485. https://doi.org/10.1016/j.jhazmat.2006.09.055

Zhou H, Chen X, Hu T et al (2020) A novel bat coronavirus closely related to SARS-CoV-2 contains natural insertions at the S1/S2 cleavage site of the spike protein. Curr Biol 30:2196–2203. https://doi.org/10.1016/j.cub.2020.05.023

Appendix

Text S1. A paragraph of curated text written by the authors (in Simplified Chinese)

> 中国拥有着全球最庞大的产品消费需求，包括新鲜农产品，肉类和水产养殖产品。1997-2017 年间，中国进口食品的国家从 108 个增加到 170 个，覆盖了全球近3/4的国家和地区。由于食品尤其是生鲜类食品易腐烂的特性，中国快速增长的进口食品消费需求促使冷链物流行业快速发展以及食品类货物的全球流动。2020 年中国冷链物流市场规模超过 3800 亿元人民币，冷库库容近1.8亿立方米，冷藏车保有量约28.7万辆。食品的冷链运输是一个复杂的过程，被运送的生鲜或食品可能会不断地接触工人，运输或储存设施，以保持食品在运输过程中所需的不间断低温环境。目前已知的 SARS 、SARS-CoV -2等冠状病毒在低温下存活时间显著延长。如果食品在加工或运输过程中，工作人员说话、咳嗽、打喷嚏等排出飞沫可导致病毒污染食物或包装表面，或货物接触了受病毒污染的冷链运输与存储设施表面，均可能导致病毒远距离流动和传播。

English version rewritten by ChatGPT using the GPT-3.5 model:

> China has the world's largest demand for a variety of products, including fresh farm products, meats, and fish. The number of countries exporting food to China has increased from 108 to 170 from 1997 to 2017, covering three-quarters of the countries and areas in the world. The rapid increase in demand for imported food, especially fresh food, has led to the development of the cold chain logistics industry. The scale of the Chinese cold chain logistics market is over 380 billion yuan, with cold storage facilities covering 180 million cubic meters and a fleet of 287,000 refrigerated trucks. Cold chain transport is a complex process, during which fresh food comes into constant contact with workers. To maintain food safety standards, transport and storage equipment are kept at low temperatures. Additionally, it is known that SARS, SARS-CoV-2, and other coronaviruses can survive for extended periods at low temperatures, which highlights the importance of maintaining strict hygiene practices during food processing and transportation.

J. Han et al., *ChatGPT in Scientific Research and Writing*,
https://doi.org/10.1007/978-3-031-66940-8

Contamination of food or surfaces of products with the virus can occur when staff talk, cough, or sneeze, leading to the potential long-distance spread of the virus via contact with cold chain transport and storage equipment.

Text S2. Cross-comparison of Chinese to English translations by ChatGPT (GPT-4 and GPT-3.5 models) and other widely used online translators for scientific writing

Original text (in Simplified Chinese):

这篇文章的技术细节主要包括以下几个方面:

- 文章使用了不同的实验方法来研究 PUF 对 BPA 的吸附行为，包括等温吸附实验、动力学实验、水化学影响实验、再生实验和表征分析。
- 文章使用了Langmuir模型和Freundlich模型来拟合 PUF 对 BPA 的吸附等温数据，并发现 Freundlich 模型更能描述PUF的非线性吸附特性。
- 文章使用了准二级动力学模型和内扩散模型来分析 PUF 对 BPA 的吸附动力学数据，并发现内扩散是控制步骤，且随着浓度增加，内扩散系数增大。
- 文章使用了不同的水化学条件 (pH、离子强度、有机物质)来考察它们对 PUF 对 BPA 的吸附影响，并发现 pH 是最重要的因素，而离子强度和有机物质则没有显著影响。
- 文章使用了碱性溶液 (NaOH) 来再生 PUF，并发现可以有效地回收 BPA 并保持PUF的完整性和稳定性。
- 文章使用了高分辨电子显微镜 (HRTEM) 、气体比重计 (GP) 和红外光谱成像 (IRI) 等技术来表征PUF的结构和功能，并发现BPA可以渗透到 PUF 纤维内部并与PUF羰基和醚氧基形成氢键。

Conclusion:

1. The original text, which is the response generated by the ChatGPT-enabled new Bing using a user prompt written in Chinese, contains 371 words listing the key findings of a research article written by one of the authors (Han et al. 2013). While there are some issues in the original text output by new Bing, e.g., hallucination (Han et al. 2023), the writing itself is coherent and polished.
2. Mistakes in the following text translations and terminologies are frequently encountered in the outputs by ChatGPT (GPT-3.5 and GPT-4) and other online translators. Incorrect translation, terminologies, and grammatical errors are marked with underlines in the text outputs below.

    - (ad)sorption of BPA by/on PUF or (ad)sorption of PUF toward BPA
    - water chemistry experiments
    - pseudo-second-order kinetic model
    - intra-particle diffusion model
    - organic matter
    - high-resolution transmission electron microscopy
    - gas pycnometer
    - infrared spectroscopic imaging
    - ether oxygen

3. Overall, GPT-4 showed the best result in terms of the accuracy of the translation, including the correct use of terminologies. All translators, including GPT-4 and GPT-3.5, showed coherence in writing.

GPT-4 (total issues: 6; major issues: 4; minor issues: 2)

The technical details of this article mainly include the following aspects:

- The paper employs various experimental methods to study the adsorption behavior of PUF towards BPA, including isothermal adsorption experiments, kinetic experiments, hydrochemical impact experiments, regeneration experiments, and characterization analysis.
- The paper uses the Langmuir model and the Freundlich model to fit the isothermal adsorption data of **PUF on BPA**, and finds that the Freundlich model better describes the nonlinear adsorption characteristics **of PUF**.
- The paper employs the pseudo-second-order kinetics model and the intraparticle diffusion model to analyze the adsorption kinetic data of **PUF on BPA**, and finds that intraparticle diffusion is the controlling step, and the diffusion coefficient increases with the increase in concentration.
- The paper explores different hydrochemical conditions (pH, ionic strength, organic matter) to investigate their impact on the adsorption of **PUF on BPA**, and finds that pH is the most crucial factor, whereas ionic strength and organic matter do not have significant effects.
- The paper uses alkaline solution (NaOH) to regenerate PUF and finds that it can effectively recover BPA while maintaining the integrity and stability of PUF.
- The paper utilizes techniques such as high-resolution transmission electron microscopy (HRTEM), gas pycnometer (GP), and infrared spectroscopy imaging (IRI) to characterize the structure and function of PUF, and discovers that BPA can permeate into the fibers of PUF and form hydrogen bonds with PUF carbonyl and ether groups.

GPT-3.5 (total issues: 9; major issues: 5; minor issues: 4)

The technical details of this article mainly include the following aspects:

- The article used various experimental methods to study the adsorption behavior of **PUF (Polyurethane foam) on BPA (Bisphenol A)**, including isothermal adsorption experiments, kinetic experiments, hydrochemical influence experiments, regeneration experiments, and characterization analysis.
- The article employed the Langmuir model and Freundlich model to fit the adsorption isotherm data of **PUF on BPA** and found that the Freundlich model could better describe the nonlinear adsorption characteristics **of PUF**.
- The article used pseudo-second-order kinetic model and intra-particle diffusion model to analyze the adsorption kinetic data of **PUF on BPA** and found that intra-particle diffusion was the controlling step, and the intra-particle diffusion coefficient increased with the increase in concentration.

- The article examined the influence of different hydrochemical conditions (pH, ionic strength, organic substances) on the adsorption of **PUF on BPA** and found that pH was the most important factor, while ionic strength and organic substances had no significant impact.
- The article regenerated PUF using an alkaline solution (NaOH) and found that it could effectively recover BPA while maintaining the integrity and stability of PUF.
- The article used techniques such as high-resolution transmission electron microscopy (HRTEM), gas pycnometer (GP), and infrared spectroscopic imaging (IRI) to characterize the structure and functionality of PUF and found that BPA could penetrate into the interior of PUF fibers and form hydrogen bonds with PUF carbonyl and ether oxygen groups.

DeepL Translate (total issues: 10; major issues: 8; minor issues: 2)

The technical details of this article include the following:

- The article used different experimental approaches to investigate the adsorption behavior of **PUF on BPA**, including isothermal adsorption experiments, kinetic experiments, water chemistry impact experiments, regeneration experiments, and characterization analysis.
- The article used Langmuir and Freundlich models to fit the isothermal data of **PUF adsorption on BPA**, and found that the Freundlich model better described the nonlinear adsorption properties **of PUF**.
- The article used a **quasi-secondary kinetic model** and an **internal diffusion model** to analyze the sorption kinetic data of PUF on BPA and found that internal diffusion was the controlling step and the internal diffusion coefficient increased with increasing concentration.
- Different hydrochemical conditions (pH, ionic strength, organic matter) were used to examine their effects on the adsorption of **PUF on BPA**, and pH was found to be the most important factor, while ionic strength and organic matter had no significant effects.
- An alkaline solution (NaOH) was used to regenerate the PUF and was found to be effective in recovering BPA and maintaining the integrity and stability of the PUF.
- High-resolution electron microscopy (HRTEM), **gas gravimetry** (GP) and infrared spectroscopic imaging (IRI) were used to characterize the structure and function of PUF, and BPA was found to penetrate into the PUF fiber and form hydrogen bonds with the PUF carbonyl and **ether oxy** groups.

Google Translate (total issues: 12; major issues: 8; minor issues: 4)

The technical details of this article mainly include the following aspects:

- The article uses different experimental methods to study the adsorption behavior of **PUF on BPA**, including isothermal adsorption experiment, kinetic experiment, **water chemical influence experiment**, regeneration experiment and characterization analysis.
- The article uses the Langmuir model and the Freundlich model to fit the adsorption isotherm data of **PUF to BPA**, and finds that the Freundlich model can better describe the nonlinear adsorption characteristics **of PUF**.
- The article uses the pseudo-second-order kinetic model and the **internal diffusion model** to analyze the adsorption kinetic data of **PUF on BPA**, and finds that the internal diffusion is the controlling step, and the internal diffusion coefficient increases as the concentration increases.
- The article used different **water chemical conditions** (pH, ionic strength, organic matter) to examine their influence on the adsorption of **PUF on BPA**, and found that pH was the most important factor, while ionic strength and organic matter had no significant effect.
- The article used an alkaline solution (NaOH) to regenerate PUF and found that BPA can be efficiently recovered while maintaining the integrity and stability of PUF.
- The article uses techniques such as high-resolution electron microscopy (HRTEM), gas pycnometer (GP) and infrared spectroscopy imaging (IRI) to characterize the structure and function of PUF, and finds that BPA can penetrate into the interior of PUF fibers and interact with PUF carbonyls and ethers Oxygen groups form hydrogen bonds.

Microsoft Translator (total issues: 11; major issues: 10; minor issues: 1)

The technical details of this article mainly include the following aspects:

- Different experimental methods were used to study the adsorption behavior of **PUF on BPA**, including isothermal adsorption experiment, kinetic experiment, **water chemical influence** experiment, regeneration experiment and characterization analysis.
- In this paper, the Langmuir model and the Freundlich model were used to fit the adsorption isothermal data of **PUF to BPA**, and found that the Freundlich model can better describe the nonlinear adsorption characteristics **of PUF**.
- In this paper, the **quasi-secondary kinetic model** and **internal diffusion model** were used to analyze the adsorption kinetics data of PUF to BPA, and it was found that internal diffusion was a control step, and the internal diffusion coefficient increased with the increase of concentration.
- Different **water chemical conditions** (pH, ionic strength, organic matter) were used to investigate their adsorption effect on **PUF on BPA**, and found that pH was the most important factor, while ionic strength and organic matter had no significant effect.
- The paper uses an alkaline solution (NaOH) to regenerate PUF and finds that BPA can be efficiently recovered and maintain the integrity and stability of PUF.

- The paper uses techniques such as high-resolution electron microscopy (HRTEM), **gas hydrometer** (GP), and infrared spectroscopy imaging (IRI) to characterize the structure and function of PUF, and finds that BPA can penetrate into PUF fibers and form hydrogen bonds with PUF carbonyl and **etheroxy** groups.

CNKI Academic Translator (total issues: 7; major issues: 5; minor issues: 2)

The technical details of this article mainly include the following aspects:

- In this paper, different experimental methods were used to study the adsorption behavior of BPA on PUF, including isothermal adsorption experiment, kinetic experiment, **water chemical effect** experiment, regeneration experiment and characterization analysis.
- Langmuir model and Freundlich model were used to fit the adsorption isotherm data of BPA on PUF, and it was found that Freundlich model could better describe the nonlinear adsorption characteristics **of PUF**.
- In this paper, the pseudo-second-order kinetic model and the **internal diffusion model** were used to analyze the adsorption kinetic data of BPA by PUF, and it was found that the internal diffusion was the control step, and the internal diffusion coefficient increased with the increase of concentration.
- Different hydrochemical conditions (pH, ionic strength, organic matter) were used to investigate their effects on the adsorption of BPA by PUF, and it was found that pH was the most important factor, while ionic strength and organic matter had no significant effect.
- In this paper, alkaline solution (NaOH) was used to regenerate PUF, and it was found that BPA could be effectively recovered and the integrity and stability of PUF could be maintained.
- The structure and function of PUF were characterized by high resolution electron microscopy (HRTEM), **gas gravimeter** (GP) and **infrared spectroscopy** (IRI). It was found that BPA could penetrate into PUF fibers and form hydrogen bonds with PUF carbonyl and ether oxygen groups.

Text S3. Fact-checking the information in the model's response in **Table 27** that is presented in Chapter 1 of this book

1. Mason, S. A., Welch, V. G., & Neratko, J. (2018). Synthetic polymer contamination in bottled water. Frontiers in Chemistry, 6. https://doi.org/10.3389/fchem.2018.00407

    - Quantities: 93% of the samples tested contained microplastics, with an average of 325 particles/L.
    - Materials: Predominantly polypropylene (PP), polyethylene (PE), and PET.
    - Sample types and sources: 259 bottled water samples from 11 brands and 9 countries.
    - Extraction method: Filtration using a 2.5 μm pore size filter followed by Nile Red staining.
    - Analytical instruments: FTIR spectroscopy and Raman spectroscopy.

Results from our fact-checking:

- All bibliographic details are correct, including the DOI link.
- "Quantities: 93% of the samples tested contained microplastics, with an average of 325 particles/L." This statement is correct. Below is the text from the reference article.
  - From the Abstract: "*Of the 259 total bottles processed, 93% showed some sign of microplastic contamination.*"
  - From the Results section, under the "Overview" subsection: "*Seventeen bottles out of the 259 bottles analyzed (~ 7%) showed no microplastic contamination in excess of possible laboratory influence indicating that 93% of the bottled water tested showed some sign of microplastic contamination.*"
  - From the Results section, under the "Overview" subsection: "*When averaged across all lots and all brands, 325 MPP/L were found within the bottled water tested [broken down as an average of 10.4 MPP/L occurring within the larger size range (> 100 um) and an average 315 MPP/L within the smaller size range (6.5–100 um)].*"
- "Materials: Predominantly polypropylene (PP), polyethylene (PE), and PET." This statement contains an error and it fails to point out the fact that only a small portion of the extracted particles (20% of microplastic particles > 100 μm) were analyzed for their type of plastic material in this study:
  - First, this statement ignored the fact that only (some of) the particles with sizes over 100 microns could be identified by FTIR for their type of materials. This is clearly stated in the article.

    From the Results section, under the "NR + FTIR Confirmed Particles (>1 00 μm)" subsection:

    "*In total nearly 2,000 microplastic particles > 100 um were extracted from all of the filters, with nearly 1000 (~ 50%) being further analyzed by FTIR.*"
    "*In total over 400 particles (20% of all extracted plastic particles > 100 um and 40% of those analyzed by FTIR) met this threshold for identity confirmation and only those results are presented here.*"

    From the Results section, under the "NR Tagged Particles (6.5–100 μm)" subsection.

    "*Given the limitations of the lab, particles < 100 um (the so-called "NR tagged particles") were not able to be confirmed as polymeric through spectroscopic analyses (FTIR and/or Raman spectroscopy).*"

    In this study, the authors used Nile Red (NR) as a selective fluorescence tag to plastic debris to differentiate *plastic* versus *non-plastic* debris (with sizes smaller than 100 μm) extracted from the samples but were unable to identify the *type* of plastic materials of these smaller particles. The authors

referred to two previous studies as supporting evidence of using NR as a selective tag for microplastics in environmental samples and performed additional FTIR analysis on NR-tagged fluorescence particles > 100 μm to validate this technique.

From the Results section, under the "NR Tagged Particles (6.5–100 μm)" subsection":"*…in testing of various stains and dyes that could be employed for microplastic detection and analysis within environmental samples with a greater potential for misidentification and false positives (i.e., sediments and open-water environmental samples) both Maes et al. (2017) and Erni-Cassola et al. (2017) concluded that Nile Red (NR) was very selective, especially within the time scales of incubation employed, and could be used for the rapid detection of microplastics without the need for additional spectroscopic analysis. To be sure that is why this stain was employed for this study. Additionally FTIR analysis was done on fluorescing particles > 100 um and every particle analyzed was confirmed to be polymeric*."

– Second, for particles that were identified as plastics in this study, the most abundant types are polypropylene (54%), followed by nylon (16%), polystyrene (11%), and polyethylene (10%). The PET variety, i.e., polyester and polyethylene terephthalate which are grouped as "PEST" by authors in the reference study, only accounted for 6% of the plastic debris analyzed in the study, after PP, Nylon, PS, and PE.

From the Results section, under the "NR + FTIR Confirmed Particles (>100 μm)" subsection: "*Polypropylene was found to be the most common polymeric material (54%) with Nylon being the second most abundant (16%; Fig. 2). Polypropylene is a polymer often used to make plastic bottle caps, along with polyethylene, which corresponded to 10% of the particles analyzed*."
More results are shown in Fig. 2 in the reference study.

- "Sample types and sources: 259 bottled water samples from 11 brands and 9 countries. This statement is correct.

  – From the Introduction: "*In total 259 bottles of water from 11 brands were processed across 27 different lots purchased from 19 locations in nine countries.*"

  Additional information on the samples is shown in Table 1 in the reference study.

  – From the Results section, under the "Overview" subsection: "*A total of 259 individual bottles from across 11 different brands and 27 different lots were analyzed for microplastic particulate…*"

- "Extraction method: Filtration using a 2.5 μm pore size filter followed by Nile Red staining." This statement is incorrect.

  - From the "Materials and methods" section, under the "Sample Processing" subsection: "*While under the fume hood, each bottle was opened and injected with a specific volume of Nile Red solution (prepared in acetone to 1 mg* mL^{-1}*) to yield a working concentration of 10 ug* mL^{-1} *(Maes et al. 2017) and re-capped. Nile Red adsorbs to the surface of plastics, but not most naturally occurring materials, and fluoresces under specific wavelengths of light (Erni-Cassola et al. 2017). Bottles were allowed to incubate with the injected dye for at least 30 min. The bottled water was then vacuum filtered through a glass fiber filter (Whatman grade 934-AH, 55 mm diameter, 1.5 μm pore).*"

- "Analytical instruments: FTIR spectroscopy and Raman spectroscopy." This statement is largely correct. To be clear, the authors stated that they had used Raman spectroscopy (along with FTIR) in their attempt to identify the type of polymer material for particles with sizes smaller than 100 microns, although they did not show any results obtained with Raman spectroscopy.
  - From the "Materials and methods" section, under the "Sample Processing" subsection:

    "*…the first 3–5 particles were analyzed* via *FTIR (PerkinElmer Spectrum Two ATR; 450* cm^{-1} *to 4,000* cm^{-1}*, 64 scans, 4* cm^{-1} *resolution; ATR correction) to confirm polymeric identity (Spectrum 10 software suite).*"
  - From the "Results" section, under the "NR Tagged Particles (6.5–100 μm)" subsection:

    "*Given the limitations of the lab, particles < 100 um (the so-called "NR tagged particles") were not able to be confirmed as polymeric through spectroscopic analyses (FTIR and/or Raman spectroscopy).*"

2. Oßmann, B. E., et al. (2018). Small-sized microplastics and pigmented particles in bottled mineral water. Water Research, 141, 307–316. https://doi.org/10.1016/j.watres.2018.05.027

   - Quantities: An average of 242 particles/L of microplastics and 5052 particles/L of pigmented particles.
   - Materials: Polyethylene terephthalate (PET), polypropylene (PP), polyethylene (PE), and others.
   - Sample types and sources: 32 bottled mineral water samples from 14 German brands.
   - Extraction method: Filtration using a 0.2 μm pore size silver membrane filter.
   - Analytical instruments: Micro-Raman spectroscopy.

Results from our fact-checking:

- All bibliographic details are correct, including the DOI link.
- "Quantities: An average of 242 particles/L of microplastics and 5052 particles/L of pigmented particles." This statement is incorrect.

- "An average of 242 particles/L of microplastics" is incorrect.

  In this study, the authors grouped the samples based on their types of bottles and provided results for each group. This is shown in "**Fig. 1** Mean number of microplastics projected to 1 L sample volume function of the bottle type (material)", under the subsection "3.1. Microplastics." The average quantity "242 particles/L" does not correlate to any of the five groups of samples shown in the figure.
  The closest match to this quoted figure (242 particles/L) is the results from blank samples in subsection "3.1.1. Blank samples", where the author stated that "*On average, 384 ± 468 microplastics/l were found in blank samples (Fig. 1), consisting mainly of PP, some of PS, PE and PET*."
  All sample groups show significantly larger quantities of microplastics than the quoted figure (242 particles/L of microplastics).

  > From the subsection "3.1.2. Mineral water": "*Water from single use PET bottles displayed the lowest amount with 2649 ± 2857 microplastics/l. Water from reusable PET bottles contained on average 4889 ± 5432 microplastics/l and water from glass bottles 6292 ± 10,521 microplastics/l (Fig. 1).*"

- "An average of…5052 particles/L of pigmented particles." This statement is also incorrect.

  - Similar to the total counts of microplastics, the authors divided the samples into different groups. This is shown in "**Fig. 5** Mean number of pigment particles projected to 1 L sample volume function of the bottle type (material)." The average quantity of "5052 particles/L of pigmented particles" does not correlate to any of the five groups of samples that are shown in this figure.

- To be clear, the quoted numbers "242 and 5052 particles/L" are not found in the main article or the Supplementary Data of this publication.

- "Materials: Polyethylene terephthalate (PET), polypropylene (PP), polyethylene (PE), and others." This statement is largely correct. However, it ignored two additional types of polymers that represent major fractions (ranked 2nd or 3rd in terms of abundance) of the plastic debris found in the sample groups, i.e., styrene-butadiene-copolymer in glass bottled mineral water (13%), and PET + olefin in single-use PET bottled mineral water (11%) and reusable PET-bottled mineral water (7.7%). This is illustrated in "**Fig. 3** Polymer type of the detected microplastics with respect to the bottle type (material)".
- "Sample types and sources: 32 bottled mineral water samples from 14 German brands." This statement contains an error. The number of brands should be 21, not 14.

  - In the "Materials and methods" section, under subsection "2.2. Samples":

"*32 samples from 21 different brands of mineral water were purchased in Bavarian food stores.*"

– Additional information is shown in the Supplementary Data of this reference study, under "**Table S1.** Summary of several characteristics of all analysed samples of mineral water"

- "Extraction method: Filtration using a 0.2 μm pore size silver membrane filter." This statement contains an error. In this study, the authors used 0.2-μm syringe filters with *cellulose acetate* membranes, not silver membranes.

  – In the "Materials and methods" section, under subsection "2.1. Precautions to avoid sample contamination":

  "*All added solutions were filtered through a syringe membrane filter (Cellulose acetate, 0.2* μm*, GE Healthcare Life Sciences Whatman™) prior to use.*"

- "Analytical instruments: Micro-Raman spectroscopy." This statement is correct.

  – In the "Materials and methods" section, under the subsection "2.4.1. Indentification of microplastics"

  "*After the sample preparation, particles were analysed directly on the filter surface with micro-Raman spectroscopy using an XploRa Plus system, operated by LabSpec 6 software (Horiba Scientific).*"

3. Schymanski, D., Goldbeck, C., Humpf, H. U., & Fürst, P. (2018). Analysis of microplastics in water by micro-raman spectroscopy: release of plastic particles from different packaging into mineral water. Water Research, 129, 154–162. https://doi.org/10.1016/j.watres.2017.11.011

   - Quantities: Ranged from 2649 particles/L to 6292 particles/L depending on the packaging type.
   - Materials: Mainly PET and PP.
   - Sample types and sources: Mineral water stored in single-use PET bottles, reusable PET bottles, and glass bottles.
   - Extraction method: Filtration using a 0.2 μm pore size aluminum oxide membrane filter.
   - Analytical instruments: Micro-Raman spectroscopy.

Results from our fact-checking:

- All bibliographic details are correct, including the DOI link.
- "Quantities: Ranged from 2649 particles/L to 6292 particles/L depending on the packaging type." This statement is incorrect.

  – The authors grouped the samples into four groups and provided the average microplastic counts (particles/L) for each group of samples. This is shown in "**Fig. 3** Mean microplastic content of water from different packaging" The

quoted figures do not correlate with the results shown in this figure, i.e., 11–118 particles/L.
  - The two quoted figures in this statement, i.e., 2649 particles/L and 6292 particles/L, are not found in the main article or the Supplementary Data.

- "Materials: Mainly PET and PP." This statement is partially correct.
  - In this study, the authors identified "PET (polyethylene terephthalate) + PES (Polyester)" as one of the main types of polymer materials found in the microparticles isolated from the samples. By convention, both PET and PES are commonly referred to as "PET" (the GPT-4 model may have adopted this convention in its response), although in this study they are grouped as "PEST" by the authors. This is shown in "**Fig. 4** Polymer distribution of the microparticles found in water from different packaging types." This part of the statement is largely correct.
  - Polyethelyele (PE), not PP as suggested in this statement, was found to be the largest or the second largest fraction of polymer material in the microparticles from samples with beverage carton and glass packaging materials, accounting for 38 and 35% of the microparticles found in these samples. This statement is also incorrect for the other two sample groups, i.e., One way bottles and Returnable bottles. Their results are also shown in "**Fig. 4** Polymer distribution of the microparticles found in water from different packaging types."
  - "Sample types and sources: Mineral water stored in single-use PET bottles, reusable PET bottles, and glass bottles." This statement is incorrect.
    - In the "Materials and methods" section, under the subsection "2.2. Types of water":

      "*Water samples (volume range 700 ml–1500 ml) from 12 different returnable and 10 single-use plastic bottles, 3 beverage cartons and 9 glass bottles were obtained from grocery stores in Germany.*"

    - The returnable plastic bottles were made of PET (with PP caps):

      In the Abstract: "*Most of the particles in water from returnable plastic bottles were identified as consisting of polyester (primary polyethylene terephthalate PET, 84%) and polypropylene (PP; 7%). This is not surprising since the bottles are made of PET and the caps are made of PP.*"

    - The material of single-use plastic bottles was not clearly stated in the study. It is not clear whether all single-use plastic bottles in this study were PET bottles, although PET bottles are very common in single-use bottled water products and PET microparticles were identified as a major fraction of the microparticles in samples from single-use plastic bottles in this study.

      In the Abstract: "*In water from single-use plastic bottles only a few micro-PET-particles have been found.*"

  - Beverage cartons, which represent a type of the samples tested in the study, are not mentioned in the statement.

- "Extraction method: Filtration using a 0.2 μm pore size aluminum oxide membrane filter." This statement is incorrect.
  - "*Since gold hardly emits any Raman signals gold coated polycarbonate filter (filtr.AID membranes, rapID, Berlin, 3.0 μm pore size and inner diameter of 24 mm) were used*."
- "Analytical instruments: Micro-Raman spectroscopy." Correct.
  - "*38 mineral waters were analyzed for microplastics by means of μ-Raman spectroscopy*."
  - "*However, micro-Raman spectroscopy is capable of detecting even smaller particle sizes. Therefore, we show that this technique, which was used in this study, is particularly useful in detecting microplastics in drinking water where particle sizes are in the low micrometer range*."

4. Koelmans, A. A., et al. (2019). Microplastics in freshwaters and drinking water: critical review and assessment of data quality. Water Research, 155, 410–422. https://doi.org/10.1016/j.watres.2019.02.054
   - This study is a critical review of microplastics in freshwaters and drinking water, assessing data quality and providing guidance for future research. The authors discuss various extraction methods, analytical instruments, and sample types/sources.

Results from our fact-checking:

- All bibliographic details are correct, including the DOI link.
- "This study is a critical review of microplastics in freshwaters and drinking water, assessing data quality and providing guidance for future research. The authors discuss various extraction methods, analytical instruments, and sample types/sources." This statement is mostly correct.
  - The scope of this review is clearly stated in the title "Microplastics in freshwaters and drinking water: Critical review and assessment of data quality", which the model could access. The points of discussion, i.e., various extraction methods, analytical instruments, and sample types/sources, as listed in the statement, are found in the different subsections of the review article.

    Subsections "3.1.2. Sampling methods" and "3.1.3. Sample size"

    In these sections, the authors discussed the extraction of *samples*, rather than the extraction of *microplastics* from samples, from surface water, wastewater, tap water, and bottled drinking water. The phrase "extraction methods" in the model's response should be written as "sampling methods" to more accurately describe the content in the

Note that the authors did not present a review of the methods for extracting microplastics from various types of samples.
A summary of the sample methods, including the type of water sources, sampling methods, and sample size is shown in "**Table 1**. Overview of individual and accumulated scores of papers reporting microplastic concentrations in surface water and drinking water."

Subsection "3.1.9. Polymer identification"

This subsection contains discussions on the analytical techniques and instruments for identifying the types of polymer materials for microplastic particles extracted from the samples.

Subsection "3.2.1. b Global microplastic concentrations in different water types"

This subsection contains information on the sample types, i.e., sources, in addition to the information listed in "**Table 1**. Overview of individual and accumulated scores of papers reporting microplastic concentrations in surface water and drinking water."

Additional information on the sample types/sources can be found in the Supplementary Data of the reference study (**Table S1**).

In Section "2. Methods", under the subsection "2.3. Study characteristics": "*For each study the following characteristics were summarized in tabular form (Table S1): Reference, Country (area), Source (water type), Treatment applied (for wastewater treatment plants (WWTP) or drinking water treatment plants (DWTP), bottled and tap water), Sampling date, Size/shape (of microplastics detected), Polymer types (of microplastics detected), Chemicals (analysed on microplastic), Value (of microplastics detected in water sample), Quality assurance applied (detection limit, positive controls, negative controls), Sampling method, Analysis method, Comments. Raw concentration data were pooled per water type: WWTP influent, WWTP effluent, lake, river, canal, groundwater, untreated and treated tap water, and bottled water, and analysed for means, ranges and significance of differences among the water types.*"

5. Welle, F., & Franz, R. (2018). Microplastic in bottled natural mineral water—literature review and considerations on exposure and risk assessment. Food Additives & Contaminants: Part A, 35(12), 2482–2492. https://doi.org/10.1080/19440049.2018.1533407

- This study is a literature review on microplastics in bottled natural mineral water. The authors discuss the sources, analytical methods, and potential risks associated with microplastics in bottled water.

Results from our fact-checking:

- The bibliographic details are correct, except the DOI link which should be https://doi.org/10.1080/19440049.2018.1543957
- "This study is a literature review on microplastics in bottled natural mineral water. The authors discuss the sources, analytical methods, and potential risks associated with microplastics in bottled water." This statement is mostly correct.
  - The scope of this study is clearly stated in the title "Microplastics in bottled natural mineral water", which is readily accessible by the model. Discussions on the specific points listed in the statement, i.e., sources, analytical methods, and potential risks are found in the review article.

    Discussions on the sources of microplastics in bottled water can be found in the section "Literature data on microplastics in bottled mineral water" and the Conclusion section:

    "*The study found 118 ± 88 particles per litre (range 28–241) in returnable, multi-use PET bottles. In terms of plastic type, 84% of the particles in multi-use PET bottles consisted of PET, 7% of PP, 5% of PE and 2% of PA, with 2% of other polymers, which indicates that the packaging material itself might be a source of the microplastics contamination.*"
    "*The source of microplastic particles in beverages are most likely packaging materials themselves. PET and PE polymers have been reported to be the most frequently found polymer types.*"

    Brief discussions on the analytical methods for microplastics in bottled water can be found in the section "Literature data on microplastics in bottled mineral water". However, it should be noted that analytical methods, including the associated sample preparation, are *not* the focus of this review paper. This is clearly stated in the last paragraph of the Introduction section, where the authors listed the objective of this review article as follows:

    "*The objective of this publication is (i) to present and discuss a compilation of the currently available literature data on microplastics in bottled mineral water, (ii) to undertake an estimation of the related oral exposure of the consumer from substances present in microplastics and from the plastics particles themselves and (iii) to present toxicological arguments for and considerations on risk assessment from the consumption of bottled mineral water containing microplastics.*"

    In this review, the authors presented detailed discussions on the potential risks of microplastics in bottled mineral water under the section "Considerations on exposure and risk assessment for the consumer." Text excerpts are omitted here for brevity. Readers are referred to the reference article for further validation, using the correct DOI link in our fact-checking.

6. Liebezeit, G., & Liebezeit, E. (2014). Synthetic particles as contaminants in german beers. Food Additives & Contaminants: Part A, 31(9), 1574–1578. https://doi.org/10.1080/19440049.2014.945099

- Although this study focuses on beer, the methodologies for extraction and analysis of microplastics can be applied to bottled water as well.
- Quantities: Ranged from 2 to 79 particles/L depending on the brand.
- Materials: Mainly PET and PE.
- Sample types and sources: 24 German beer brands.
- Extraction method: Filtration using a 0.45 μm pore size cellulose nitrate membrane filter.
- Analytical instruments: FTIR spectroscopy.

Results from our fact-checking:

- All bibliographic details are correct, including the DOI link.
- "Although this study focuses on beer, the methodologies for extraction and analysis of microplastics can be applied to bottled water as well." This statement is partially correct.
  - This sentence appears to be correct upon initial reading; however, upon our reading the full text, a significant shortcoming in the study's methodology was found, which the authors admitted in the paper. Specifically, the particulate matter that was filtered out underwent no chemical analysis, such as FT-IR, Raman, or micro-Raman spectroscopy, before being categorically labeled as microplastics. Therefore, readers should not fully accept the suggestion in the statement by the GPT-4 model. Instead, they can only refer to the extraction method in this study as a point of reference.

    In the "Material and methods" section: "*After a reaction time of 5 min the dye was filtered off and the stained material washed dye-free with filtered deionised water. The non-stained material will be referred to as microplastic in the following, although it is recognised that only spectroscopic analysis (FT-IR or Raman spectroscopy) can provide unambiguous proof of the synthetic nature of the non-stained particles.*"

- "Quantities: Ranged from 2 to 79 particles/L depending on the brand." This statement is incorrect.
  - First, "2–79" refers to the quantity of *fibers* per liter of beer, not the number of *particles.* Additionally, the phrase "depending on the brand" is not entirely correct. The quantities depend on the brand, the particular sample batch (i.e., differences between replicates), and the fraction of the debris (i.e., fibers, fragments, or granules).

    In the Abstract: "*A total of 24 German beer brands was analysed for the contents of microplastic fibres, fragments and granular material. In all cases contamination was found. Counts ranged from 2 to 79 fibres L^{-1}, from 12 to 109 fragments L^{-1} and from 2 to 66 granules L^{-1}.*"
    In the "Results and discussion" section: "*A clear overall dominance of one microplastic fraction could not be established*."

In the "Results and discussion" section: "A high variability was noted between different brands, but also between replicates (Fig. 2 and Table 1)"

- "Materials: Mainly PET and PE." The model made up this information, i.e., hallucination. This statement also has no relevance to the reference study.
  - In this study, the authors counted the number of particles, fragments, and fibers, and discussed their possible sources, but did *not* analyze the chemical compositions of the extracted debris. This can be seen in the "Material and methods" section of this study.
- "Sample types and sources: 24 German beer brands." This statement is correct.
  - At the beginning of the "Material and methods" section:

    "*A total of 24 beer German brands was obtained from local supermarkets, among them the 10 most popular brands in Germany (data for 2012; see* http://www.inside-getraenke.de)."
- "Extraction method: Filtration using a 0.45 μm pore size cellulose nitrate membrane filter." This statement is incorrect. Both the pore size and membrane filter type are inconsistent with the information in the study.
  - In this study, the author used two methods to filter the samples. However, neither the filtering materials nor the pore sizes of the filters matched the information output by the model. See the following text exerpts from the "Material and methods" section:

    "*With the exception of the wheat beers, all samples were filtered over a 0.8 μm grey, gridded cellulose filter immediately after opening the bottles.*"
    "*The wheat beers could not be treated this way as the filters became clogged after about 200 ml had passed through. These samples were initially passed though a 40 μm stainless steel sieve.*"
- "Analytical instruments: FTIR spectroscopy." This is a hallucinated statement. The authors did not present any results on the instrumental analysis of the extracted debris in the study.
  - In the "Material and methods" section: "The non-stained material will be referred to as microplastic in the following, although it is recognised that only spectroscopic analysis (FT-IR or Raman spectroscopy) can provide unambiguous proof of the synthetic nature of the non-stained particles."

7. Zuccarello, P., et al. (2019). A simple and rapid method for the isolation of microplastics from gastrointestinal tracts of fishes. Environmental Pollution, 250, 148–155. https://doi.org/10.1016/j.envpol.2019.03.060
   - Although this study focuses on fish, the extraction methods presented may be adapted for analyzing microplastics in bottled water.

- Extraction method: Density separation using a saturated NaCl solution followed by filtration.
- Analytical instruments: FTIR spectroscopy.

Results from our fact-checking:

- This reference does not exist. At the time of writing, searching with the article title in Google, Microsoft Bing, or Web of Science returned no relevant results.
- The closest match of the article title is a sentence in the review paper "Microplastic sample purification methods—Assessing detrimental effects of purification procedures on specific plastic types", which was published in 2022 in the journal *Science of the Total Environment* (https://doi.org/10.1016/j.scitotenv.2022.154824), where the authors stated the following in subsection "4.4 Combination of acidic and alkaline agents with/without oxidative treatments for sample purification":
  - "*Roch and Brinker (2017) developed a rapid purification method for the extraction of microplastics from the gastrointestinal tracts of fishes is a sequence of 1 M NaOH (50 °C, 15 min), 10 M HNO3 (50 °C, 15 min followed by 80 °C, 15 min) combined with a density separation with NaI*."
- The DOI link in the model's response links to a research paper published in Environmental Pollution entitled "*Strategy for phytomanagement in an area affected by iron ore dam rupture: A study case in Minas Gerais State, Brazil*", which is irrelevant to the topic described in the user prompt.
- Searching in the journal's archived issues shows no results matching the volume and page numbers provided by the model.

Text S4. Checking the information in the model's response shown in **Table 33** that is presented in Chapter 1 of this book

Response (Part 1 of 2): The relationship between nanoparticles and cancer is a complex and active area of research. While there is evidence suggesting potential links between certain nanoparticles and cancer development, it's important to note that the field is still evolving, and further studies are needed to establish definitive conclusions. I'll provide an overview of some important evidence and scientific studies related to nanoparticles and cancer, but it's essential to consult the latest research for the most up-to-date information.

1. Carbon Nanotubes (CNTs)
   - A study published in the journal Nature Nanotechnology in 2008 suggested that long and needle-like multi-walled CNTs can induce asbestos-like pathogenicity and could lead to mesothelioma, a cancer of the lining of the lungs.
   - Another study published in the journal Nature Nanotechnology in 2011 found that certain types of CNTs could cause lung damage and promote the growth of lung tumors in mice.

2. Titanium Dioxide (TiO_2) Nanoparticles

   - Research published in the journal Cancer Research in 2009 reported that TiO_2 nanoparticles, commonly used in sunscreens and cosmetics, can induce genetic damage and promote the growth of pre-cancerous lesions in mice.
   - A study published in the journal ACS Nano in 2011 showed that TiO_2 nanoparticles can cause DNA damage in human lung cells.

3. Silver Nanoparticles

   - A study published in the journal Particle and Fibre Toxicology in 2012 demonstrated that silver nanoparticles could cause DNA damage and induce oxidative stress in human lung cells.
   - Research published in the journal Small in 2013 found that silver nanoparticles could promote the growth of lung tumors in mice.

4. Other Nanoparticles

   - Various studies have investigated the potential carcinogenicity of other nanoparticles, such as gold, quantum dots, and certain metal oxides. These studies have reported diverse findings, highlighting the need for further research to draw definitive conclusions.

It's worth noting that the toxicity and potential carcinogenicity of nanoparticles can vary depending on factors such as size, shape, surface chemistry, dose, and exposure duration. Additionally, different cell types and animal models may respond differently to nanoparticles.

It is important to recognize that this information represents a selection of studies, and the field of nanoparticle research is continuously evolving. Therefore, it is advisable to consult current scientific literature and expert opinions for the latest developments in this area.

Response (Part 2 of 2): Certainly! Here are the titles of the studies mentioned:

1. Carbon Nanotubes (CNTs)

   a. "Carbon nanotubes introduced into the abdominal cavity of mice show asbestos-like pathogenicity in a pilot study"—Nature Nanotechnology, 2008.
   b. "Carbon nanotubes introduced into the pleural space of mice show asbestos-like pathogenicity in a pilot study"—Nature Nanotechnology, 2011.

2. Titanium Dioxide (TiO_2) Nanoparticles

   a. "Toxicity and cellular responses of intestinal cells exposed to titanium dioxide"—Cancer Research, 2009.
   b. "Titanium dioxide nanoparticles induce DNA damage and genetic instability in vivo in mice"—ACS Nano, 2011.

3. Silver Nanoparticles

   a. "Silver nanoparticles: Particle uptake, oxidative stress and cytotoxicity in human liver cells"—Particle and Fibre Toxicology, 2012.

b. "Silver nanoparticle-induced autophagic-lysosomal disruption and NLRP3-inflammasome activation in HepG2 cells is size-dependent"—Small, 2013.

Please note that these are the titles as per my response, and you can refer to the respective journals for complete details of the studies.

Results from our fact-checking:

"Carbon nanotubes introduced into the abdominal cavity of mice show asbestos-like pathogenicity in a pilot study"—Nature Nanotechnology, 2008.

- The bibliographic details are correct.
- "…suggested that long and needle-like multi-walled CNTs can induce asbestos-like pathogenicity and could lead to mesothelioma, a cancer of the lining of the lungs". This statement is correct. See text excerpts below from the referenced study.
  - In the Abstract: "*Here we show that exposing the mesothelial lining of the body cavity of mice, as a surrogate for the mesothelial lining of the chest cavity, to long multiwalled carbon nanotubes results in asbestos-like, length-dependent, pathogenic behaviour. This includes inflammation and the formation of lesions known as granulomas.*"
  - In the main text: "*We observed that long MWNTs produced inflammation, FBGCs and granulomas that were qualitatively and quantitatively similar to the foreign body inflammatory response caused by long asbestos.*"
  - In the last paragraph of the main text, before the "Methods" section: "*Our data demonstrate that asbestos-like pathogenic behaviour associated with CNTs conforms to a structure –activity relationship based on length, to which asbestos and other pathogenic fibres conform.*"

"Carbon nanotubes introduced into the pleural space of mice show asbestos-like pathogenicity in a pilot study"—Nature Nanotechnology, 2011.

- This reference does not exist. At the time of writing, searching with the article title in Google, Microsoft Bing, or Web of Science returned no relevant results.
- The closest match is a research article published in the *American Journal of Pathology* in 2011, which is entitled "Length-dependent retention of carbon nanotubes in the pleural space of mice initiates sustained inflammation and progressive fibrosis on the parietal pleura."
  - In this article, the authors stated that "*The fibrous shape of carbon nanotubes (CNTs) raises concern that they may pose an asbestos-like inhalation hazard, leading to the development of diseases, especially mesothelioma. Direct instillation of long and short CNTs into the pleural cavity, the site of mesothelioma development, produced asbestos-like length-dependent responses.*"

Results from our fact-checking:

"Toxicity and cellular responses of intestinal cells exposed to titanium dioxide"—Cancer Research, 2009.

- The article title and the year of publication are correct. The name of the journal should be *Cell Biology and Toxicology*, not *Cell Research.* Below is the full bibliographic information of the reference study:
  - Koeneman B. A., Zhang Y., Westerhoff P. et al. (2009) Toxicity and cellular responses of intestinal cells exposed to titanium dioxide. Cell Biology and Toxicology 26(3):225–238. https://doi.org/10.1007/s10565-009-9132-z
- "…reported that TiO_2 nanoparticles, commonly used in sunscreens and cosmetics, can induce genetic damage and promote the growth of pre-cancerous lesions in mice." This statement is incorrect.
  - The scope and main findings of this study are clearly stated in the Abstract.

    "*This study investigates possible pathways by which nanoparticles, titanium dioxide (TiO_2), could cross the epithelium layer by employing both toxicity and mechanistic studies.*"
    "*This study provides evidence that at 10 μg/mL and above, TiO_2 nanoparticles cross the epithelial lining of the intestinal model by transcytosis, albeit at low levels. TiO_2 was able to penetrate into and through the cells without disrupting junctional complexes, as measured by γ-catenin.*"
  - The authors used "Caco-2 cell line", not "mice", as the testing subject in the study. The authors did *not* use mice in their experiments or infer that the findings in this study could be extended to mice models.

    In the "Methods and materials" section, under the subsection "Cell culture": "*A human, brush border expressing intestinal cell line, Caco-2 (American Type Culture Collection) was maintained as previously described (Peterson and Mooseker 1992; Koeneman et al. 2009).*"
    In the Introduction section: "*The human-derived Caco-2 cell line utilized here has several advantages, including that the cells, which are stored under cryogenic conditions, can continue to be obtained to develop a consistent set of genetically identical cells that can be used in the assays.*"
  - The authors did *not* find that "TiO_2 nanoparticles induced genetic damage or prompt the growth of pre-cancerous lesions."

    In the Abstract: "*…low concentrations (10 or 100 μg/mL) of TiO_2 do not disrupt epithelial integrity. Live/dead analysis results did not show cell death after exposure to TiO_2.*"
    In the Abstract: "*…at 10 μg/mL (and above) TiO_2 nanoparticles begin alteration of both microvillar organization on the apical surface of the epithelium as well as induce a rise in intracellular-free calcium.*"

In the Discussion section: "*The electron microscopy data demonstrated the disruption of the microvillar organization after application of TiO_2, while the Calcium Green data indicated that levels of intracellular-free calcium rose in a dose–response manner to increasing concentrations of TiO_2.*"

"Titanium dioxide nanoparticles induce DNA damage and genetic instability in vivo in mice"—ACS Nano, 2011.

- The article title is correct. Both the name of the journal and the year of publication are incorrect. The correct bibliographic information is shown below.
  - Trouiller B., Reliene R., Westbrook A. et al. (2009) Titanium dioxide nanoparticles induce dna damage and genetic instability in vivo in mice. Cancer Res 69(22):8784–8789 https://doi.org/10.1158/0008-5472.CAN-09-2496
- "…showed that TiO_2 nanoparticles can cause DNA damage in human lung cells." This statement contains errors.
  - The authors studied the toxicities of TiO_2 nanoparticles in a mice model, not human cells.

    In the Abstract: "*The present study investigates TiO_2 nanoparticles–induced genotoxicity, oxidative DNA damage, and inflammation in a mice model.*"
    In the "Materials and Methods" section, under subsections "Mouse care and breeding" and "TiO_2 nanoparticles preparation and exposure":

    "*C57Bl/6Jp^{un}/p^{un} mice were obtained from The Jackson Laboratory… Mice were housed and cared for under standard specific pathogen-free conditions and according to the Animal Rescue Coalition and Institutional Animal Care and Use Committee regulations.*"
    "*Solutions of dispersed TiO_2 nanoparticles were prepared by ultrasonication (Solid State/Ultrasonic FS-14; Fisher Scientific) for 15 min in drinking water at 60, 120, 300, and 600 μg/mL concentrations just before use. We measured TiO_2 nanoparticles–supplemented water intake at the end of experiments in each cage, which housed 2 to 3 mice, and calculated an average daily water intake per mouse. Daily TiO_2 nanoparticles–supplemented water intake ranged from 3 to 7 mL/mouse, consistent with normal daily water intake. Doses were calculated using a 30 g average weight per mouse, and an average of 5 mL water intake per day. The exposure was 5 days in adult males. For* in utero *exposure, pregnant dams were given nanoparticles-supplemented drinking water for 10 days from 8.5 to 18.5 days post-coitum at a concentration of 300 μg/mL.*"
  - Using mice models, the authors also found that TiO_2 nanoparticles induced genotoxicity and inflammation, in addition to DNA damage.

    In the Abstract:

    "*We treated wild-type mice with TiO_2 nanoparticles in drinking water and determined the extent of DNA damage using the comet assay,*

the micronuclei assay, and the γ-H2AX immunostaining assay and by measuring 8-hydroxy-2′-deoxyguanosine levels and, as a genetic instability endpoint, DNA deletions. We also determined mRNA levels of inflammatory cytokines in the peripheral blood."
"*Our results show that TiO_2 nanoparticles induced 8-hydroxy-2'-deoxyguanosine, γ-H2AX foci, micronuclei, and DNA deletions. The formation of γ-H2AX foci, indicative of DNA double-strand breaks, was the most sensitive parameter. Inflammation was also present as characterized by a moderate inflammatory response.*"

In the "Discussion" section:

"*We showed that TiO_2 nanoparticles (500 mg/kg) induce not only DNA single-strand breaks and DSBs but also chromosomal damage.*" (Note: Here the "DSB" refers to "double-strand breaks".)
"*To date, very few* in vivo *genotoxicity studies have been carried out with nanoparticles…We also found that maternal exposure to 500 mg/kg TiO_2 nanoparticles during gestation results in significantly elevated frequencies of DNA deletions in offspring.*"

Results from our fact-checking:

"Silver nanoparticles: Particle uptake, oxidative stress and cytotoxicity in human liver cells"—Particle and Fibre Toxicology, 2012.

- This reference does not exist. At the time of writing, searching with the article title in Google, Microsoft Bing, or Web of Science returned no relevant results.
- The closest match is a study published in Toxicology Letters in 2011 (https://doi.org/10.1016/j.toxlet.2010.12.010), which is entitled: "Silver nanoparticles induce oxidative cell damage in human liver cells through inhibition of reduced glutathione and induction of mitochondria-involved apoptosis".

"Silver nanoparticle-induced autophagic-lysosomal disruption and NLRP3-inflammasome activation in HepG2 cells is size-dependent"—Small, 2013.

- The article title is correct. However, both the name of the journal and the year of publication are incorrect. The correct bibliographic information is shown below.
  - Mishra A. R., Zheng J. W., Tang X. et al. (2016) Silver nanoparticle-induced autophagic-lysosomal disruption and NLRP3-inflammasome activation in HepG2 cells is size-dependent. Toxicological Sciences 150(2):473–487. https://doi.org/10.1093/toxsci/kfw011
- "…found that silver nanoparticles could promote the growth of lung tumors in mice." This statement is incorrect.
  - The study used human liver-derived hepatoma (HepG2) cells, not mice models, the study the toxicities of silver nanoparticles.

In the Abstract:

"*The objective of this study was to determine the mechanism of size- and concentration-dependent cytotoxicity of AgNPs in human liver-derived hepatoma (HepG2) cells.*"

– The description of the main finding in the study "…promote the growth of lung tumors" has no relevance to the actual findings reported in the study.

In the Abstract:

"*Autophagy and enhanced lysosomal activity were induced at noncytotoxic concentrations (1 μg/ml; primary particle size:10 > 50 > 100 nm), whereas increased caspase-3 activity (associated with apoptosis) was observed at cytotoxic concentrations (10, 25, and 50 μg/ml).*"
"*Subcytotoxic concentrations of AgNPs enhanced expression of LC3B, a pro-autophagic protein, and CHOP, an apoptosis inducing ER-stress protein, and activation of NLRP3-inflammasome (caspase-1, IL-1β).*"
"*Disrupting the autophagy-lysosomal pathway through chloroquine or ATG5-siRNA exacerbated AgNPs-induced caspase-1 activation and lactate dehydrogenase release.*"

In the Conclusion:

"*We showed that AgNPs (primary particle sizes of 10, 50, and 100 nm) induce cytotoxicity in cultured liver cells that is mediated by AgNP-induced LMP and inflammasome dependent caspase-1 activation, a proinflammatory protease which regulates cell death.*"
"*Blocking AgNP induced autophagy exacerbates caspase-1 activation and cell death.*"
"*AgNPs induce autophagy and lysosmal membrane permeabilization resulting in NLRP3 inflammsome dependent caspase-1 activation.*"