



## RESEARCH ARTICLE

10.1029/2025JH000649

**Special Collection:**

Science Understanding from Data Science: Transformative Science through the Convergence of Data Science and Physical Science

**Key Points:**

- Large language models can assist geoscientists by generating data analyses and visualizations from natural-language prompts
- A general-purpose Intelligent Data Exploring Assistant shows the potential of artificial intelligence to enhance geoscience research
- The Station Explorer Assistant analyzes water level data from tide gauges providing insights into sea level variability and risks

**Supporting Information:**

Supporting Information may be found in the online version of this article.

**Correspondence to:**

M. J. Widlansky,  
mwidlans@hawaii.edu

**Citation:**

Widlansky, M. J., & Komar, N. (2025). Building an intelligent data exploring assistant for geoscientists. *Journal of Geophysical Research: Machine Learning and Computation*, 2, e2025JH000649. <https://doi.org/10.1029/2025JH000649>

Received 19 FEB 2025

Accepted 9 JUL 2025

**Author Contributions:**

**Conceptualization:** Matthew J. Widlansky

**Data curation:** Nemanja Komar

**Formal analysis:** Matthew J. Widlansky

**Funding acquisition:** Matthew J. Widlansky

© 2025 The Author(s). *Journal of Geophysical Research: Machine Learning and Computation* published by Wiley Periodicals LLC on behalf of American Geophysical Union.

This is an open access article under the terms of the [Creative Commons](#)

Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

## Building an Intelligent Data Exploring Assistant for Geoscientists

Matthew J. Widlansky<sup>1,2</sup>  and Nemanja Komar<sup>1</sup>

<sup>1</sup>School of Ocean and Earth Science and Technology (SOEST), Cooperative Institute for Marine and Atmospheric Research, University of Hawai‘i at Mānoa, Honolulu, HI, USA, <sup>2</sup>Department of Oceanography, SOEST, University of Hawai‘i at Mānoa, Honolulu, HI, USA

**Abstract** Advances in natural-language processing and large language models (LLMs) are transforming how geoscientists interact with complex data sets, enabling efficient and intuitive scientific analyses. This study introduces the Intelligent Data Exploring Assistant (IDEA), a prototype software framework that integrates existing LLM technology with domain-specific instructions, data, analytical tools, and computing resources to support geoscientific research. We demonstrate its application through the Station Explorer Assistant (SEA), a web-based tool designed for sea level scientists. SEA empowers users to analyze and interpret coastal water level data by addressing challenges such as vertical datum conversions and assessing flooding risks. We also demonstrate the generalizability of building an IDEA, whereby we deploy a local instance of the framework to analyze atmospheric observations from Mars collected by NASA’s InSight Mission. By combining LLM capabilities with robust domain-specific customizations, SEA and the Mars IDEA generate accurate analyses, visualizations, and insights through natural-language prompts. This study highlights the potential of IDEA frameworks to lower technical barriers, enhance educational opportunities, and transform geoscientific workflows while addressing the limitations and uncertainties of current LLM technology.

**Plain Language Summary** Artificial intelligence (AI) is transforming how scientists explore and understand our world. At the University of Hawai‘i Sea Level Center (UHSLC), we are developing tools that use large language models, like what ChatGPT uses, to help scientists study sea level changes. One such tool, called the Station Explorer Assistant (SEA), allows researchers to ask questions in everyday language and receive clear explanations and data analyses in response. SEA uses AI to analyze sea level data, compare water levels to normal conditions, and predict potential flooding, drawing on the UHSLC’s extensive database. It even writes and runs its own analysis software, which it shows the user to check that its results are accurate. By making sea level science easier to understand and access, SEA can support communities adapting to rising seas and other coastal challenges. SEA technology is generalizable across geoscience domains through a framework we call an Intelligent Data Exploring Assistant (IDEA), which we demonstrate by asking it to analyze wind observations from Mars. Our work highlights how AI can enhance scientific research and communication, and we envision similar tools being created to support scientists in many fields.

### 1. Introduction

The emergence of natural-language processing (NLP) to interact with powerful computer models offers transformative potential for geoscientists (Adapting to AI, 2024; Florindo, 2023; Heidt, 2024), enabling us to request assistance with scientific analyses in ways that could disrupt traditional workflows (Castelvecchi, 2022). Advances in large-language model (LLM) technology, such as those underpinning popular “chatbot” applications (e.g., ChatGPT, released by OpenAI in November 2022), have made it possible for scientists to generate information and insights through “prompts” (Lenharo, 2024). In the natural sciences, LLMs have already been used for tasks such as summarizing literature (Pearson, 2024), generating ideas (Conroy, 2024; Si et al., 2024), planning research procedures (Castelvecchi, 2024; Lu et al., 2024), writing computer code for data analysis (Castelvecchi, 2022; Gu et al., 2024; Li et al., 2022), and drafting communication materials (J Huang & Tan, 2023; Hutson, 2022; Stockel-Walker, 2023; Thorp, 2023) such as the abstract for this study. Despite the growing adoption of NLP through applications like ChatGPT, barriers remain between the capabilities of generative-artificial intelligence (AI) and its practical utility for geoscientists (Hadid et al., 2024). Although NLP tools provide unprecedented accessibility and convenience for acquiring information about the geosciences (Baucon & de Carvalho, 2024), they are not immune to inaccuracies (Farquhar et al., 2024).

**Investigation:** Matthew J. Widlansky, Nemanja Komar

**Methodology:** Matthew J. Widlansky, Nemanja Komar

**Project administration:** Matthew J. Widlansky

**Resources:** Matthew J. Widlansky

**Software:** Nemanja Komar

**Supervision:** Matthew J. Widlansky

**Validation:** Matthew J. Widlansky, Nemanja Komar

**Visualization:** Matthew J. Widlansky

**Writing – original draft:** Matthew J. Widlansky

**Writing – review & editing:** Matthew J. Widlansky, Nemanja Komar

This paper describes our effort to bridge the utility gap by developing an NLP application tailored specifically to geoscientists' needs, particularly in data analysis. By integrating LLM technology (e.g., the gpt-4o model from OpenAI; OpenAI et al., 2024) with domain-specific data, custom instructions, analytical tools, and a dedicated computing environment, we introduce an Intelligent Data Exploring Assistant (IDEA) as a prototype software application (Widlansky & Komar, 2025). The IDEA framework (Figure 1) allows users to interact with an LLM via natural-language prompts and receive AI-generated responses in the form of text, plots, and data. These interactions are enriched by the LLM's access to relevant data sets for geoscientists, code-generation capabilities for scientific analysis, specialized software functions, and computational resources to execute the generated code. By customizing the system instructions and tools for specific domains and user needs, the IDEA framework is adaptable, enabling varying degrees of specialization and sophistication. The goal is to create a seamless and intuitive interface for geoscientists to perform their work with the assistance of a robust AI-powered system.

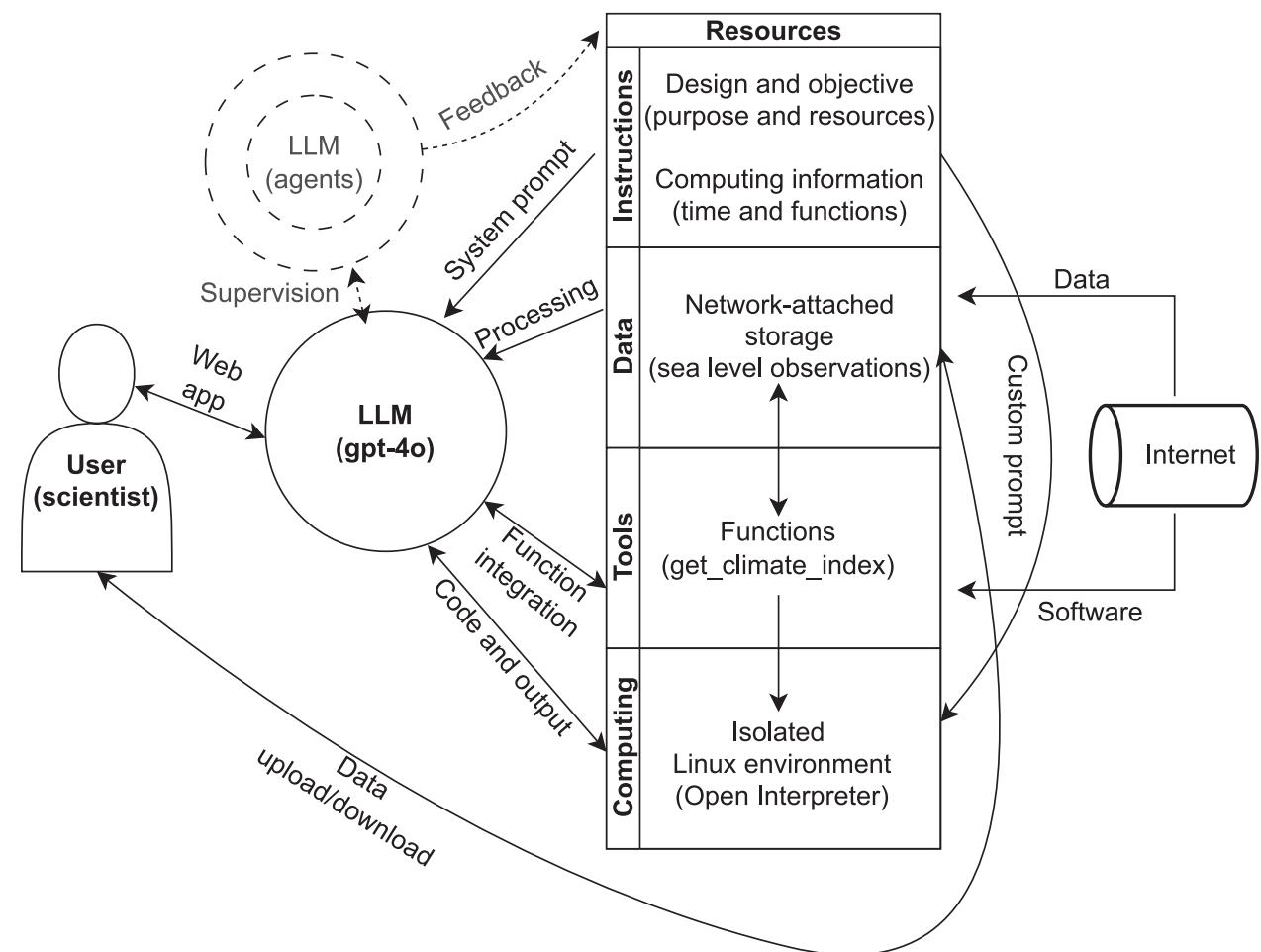
To illustrate the potential utility of an IDEA in detail, we developed an application specifically for sea level science. At the University of Hawai'i Sea Level Center (UHSLC), we designed a web-based IDEA called the Station Explorer Assistant (SEA) to assist users in analyzing and interpreting coastal water level data. Tide gauges, which measure water levels relative to local land elevations, are critical for understanding vulnerabilities to sea level variability. Working with water level data is often challenging due to the unique vertical reference frames used by tide gauges, which can be arbitrarily defined (e.g., relative to the Station Zero datum; Caldwell et al., 2015). UHSLC already provides a web product, the Station Explorer (<https://uhslc.soest.hawaii.edu/stations>), to visualize tidal datums for the Global Sea Level Observing System network (IOC, 2020). However, many users require additional support for questions such as how to convert water levels between reference frames (e.g., to mean sea level, MSL) or how to assess coastal flooding risks associated with sea level variability. Such analyses require knowledge of tidal datums, hazard thresholds (e.g., elevations associated with flooding impacts; Hague et al., 2019), and their interrelations. Motivated by these challenges and the rapid advancements in LLMs, we developed SEA as an interactive web-based application to address user queries related to sea level science. SEA is available for testing (<https://uhslc.soest.hawaii.edu/research/SEAinfo/>), offering an opportunity to evaluate its capabilities and utility in answering domain-specific questions.

The IDEA framework is easily generalized to work across geoscience domains. By modifying instructions to the LLM (Figure 1), we created a Mars IDEA specialized for analyzing atmospheric observations from NASA's InSight mission (Banerdt et al., 2020). To demonstrate the IDEA opportunity for a wide variety of use cases beyond exploring sea level observations, we prompted the Mars example application to visually interpret a published figure (from Chatain et al., 2021), consult the associated study to learn about the analysis method so that it can be recreated, and then explore new inquiries about wind characteristics on Mars. (Analyzing wind speeds recorded on Mars seemed like an interesting topic, especially because of the dust storms depicted in science-fiction stories such as "The Martian" (Weir, 2011).) We describe this example of using the IDEA framework to analyze observations of the Mars atmosphere, which is a geoscience domain far removed from our expertise and probably not something we would attempt without the IDEA.

The remainder of this paper is organized as follows. In Section 2, we detail the system instructions, accessible data sets, analytical tools, and computing environment that underpin IDEA frameworks such as the SEA web application and Mars example. Section 3 presents results for two examples of user interactions with SEA and one example using the Mars version of the IDEA framework on a personal computer. The examples include associated figure outputs of the "conversations" and our interpretations of the LLM-generated responses. Finally, in Section 4, we discuss the broader applicability of the IDEA framework for geoscientists and its potential to transform workflows. We also address known limitations of this study and the underlying technology, alongside the uncertainties that come with its rapid evolution. These considerations are vital for understanding not only the implications of LLMs for geosciences but also their impact on scientists leveraging tools such as SEA.

## 2. Data and Methods

The key design principle for building an IDEA is to provide instructions to an LLM for working with the data made available to it. The data scope itself can be virtually unlimited (e.g., from sea levels on Earth to winds on Mars), as the application's system instructions (2.1) are designed to enable data access (2.2) from a wide variety of sources beyond its local computing environment. For the SEA example on the web, this includes data sourced from (a) network-attached storage at UHSLC, (b) on-demand retrieval via cloud-based data services or internet



**Figure 1.** Design of an Intelligent Data Exploring Assistant (IDEA). The engineering plan for an intelligent assistant such as Station Explorer Assistant features a User (i.e., a geoscientist) conversing with an large language model (LLM) (gpt-4o) that is provided with Resources including domain-specific instructions (via a system and custom prompt), Data access (via local or remote file storage), Tools (i.e., functions or applications like PaperQA2), and a Computing environment (for executing code). The IDEA framework could be expanded to include an agentic or supervisory LLM for validating results (gray dashed circles and arrow) or optimizing resources based on user feedback (dashed arcing arrow), such as by tasking an LLM to review the conversation and improve the system instructions.

downloads, and (c) user uploads. It is also possible to deploy a local instance of an IDEA, which we did to demonstrate an example of analyzing Mars data. Work tasks can be made more efficient and reliable by improving the LLM (gpt-4o-2024-11-20 from OpenAI was originally used for the examples in Section 3), refining its instructions, and equipping it with custom tools for analyzing data (2.3), retrieving textual information, or perhaps reviewing its own outputs. For an IDEA to perform complex data analyses, a computing environment must also be provided (2.4). The computing environment is external to the LLM site (i.e., OpenAI), with code generated by the LLM executed locally at UHSLC in the case of SEA or on the user's personal computer in the case of a locally deployed IDEA (execution outputs, including potential error messages, are reviewed by the LLM). The IDEA software is built using open-source LLM protocols (e.g., OpenAI-compatible APIs), allowing it to interface with alternative inference endpoints such as those available through Jetstream2 (Boerner et al., 2023; Hancock et al., 2021).

Our proposed structure of combining instructions (2.1), data (2.2), tools (2.3), and computing (2.4) resources forms the basis of an IDEA, as illustrated in Figure 1 (see Table 1 for further description of these components specific to SEA). The IDEA framework employs a general-purpose LLM to assist with geoscience tasks, thereby avoiding the need to pre-train or fine-tune the LLM (Lin et al., 2024), while still producing an application customized for analyzing the data provided to it (Chen et al., 2024). We note that the LLM can reason about science questions and act on data analyses, thereby demonstrating the capability to assist users to solve their

**Table 1***Station Explorer Assistant (SEA) Communicates About Itself*

Category	Description
Instructions	I am programmed to assist with sea level science, tidal datum analysis, and related tasks. My responses are tailored to the University of Hawaii Sea Level Center (UHSLC) data and protocols. I strictly adhere to the following: <ul style="list-style-type: none"><li>-Focus exclusively on sea level, tides, datums, benchmarks, and altimetry-related questions.</li><li>-Use UHSLC Fast Delivery (FD) data as the primary source, ensuring accurate and up-to-date analyses.</li><li>-Provide clear explanations, visualizations, and references to the Station Zero datum or other requested datums.</li><li>-Ensure all outputs are reproducible, with proper documentation and clear labeling of units and datums.</li><li>-Follow strict security protocols, including scanning packages before installation and avoiding destructive operations.</li></ul>
Data	I have access to six types of sea level data and metadata as well as climate indices and altimetry observations: <ol style="list-style-type: none"><li>1. Sea Level Data (FD): Hourly and daily water levels in mm relative to Station Zero datum, retrieved via UHSLC ERDDAP server.</li><li>2. Tide Prediction Data: High/low tide predictions and hourly tide predictions from 1983 to 2030, relative to Station Zero datum.</li><li>3. Tidal Datum Data: Datum tables for converting between different vertical reference frames (e.g., MHHW, MLLW).</li><li>4. Near-Real Time Data (RAPID): Preliminary hourly observations and tide predictions with residuals.</li><li>5. Benchmarks: Metadata and photos of tide gauge benchmarks, including elevation data.</li><li>6. Altimetry Observations: Re-gridded monthly sea surface height anomalies and climatologies from CMEMS.</li></ol> Metadata: Station metadata (e.g., location, name, and country) from fd_metadata.geojson and benchmark metadata from all_benchmarks.json. Climate Indices: Data for indices such as ONI, PDO, NAO, and others retrieved from authoritative sources.
Tools	I utilize the following tools and libraries to perform analyses and visualizations: <ul style="list-style-type: none"><li>-Programming Languages: Python, Shell, JavaScript, and others as needed.</li><li>-Data Analysis: Pandas, NumPy, xarray, and utide for tidal harmonic analysis.</li><li>-Visualization: Matplotlib, Folium (for mapping), and other plotting libraries.</li><li>-Data Retrieval: HTTP requests to UHSLC servers, ERDDAP, and climate index sources.</li><li>-Security: Guarddog for scanning Python and Node.js packages before installation.</li></ul>
Capabilities	I am capable of performing a wide range of tasks related to sea level science, including: <ul style="list-style-type: none"><li>-Data Retrieval and Cleaning: Fetching, parsing, and cleaning sea level, tide, and climate data.</li><li>-Analysis: Calculating trends, residuals, datum conversions, and tidal harmonics.</li><li>-Visualization: Creating time series plots, maps, and comparative analyses with clear labeling of datums and units.</li><li>-Mapping: Generating interactive maps of tide gauge stations and benchmarks using Folium.</li><li>-Benchmark Analysis: Providing metadata and photos of tide gauge benchmarks, including elevation conversions.</li><li>-Altimetry Analysis: Comparing tide gauge data with altimetry observations, including unit conversions.</li><li>-Climate Index Correlation: Fetching and visualizing climate indices to study their relationship with sea level variations.</li><li>-Security and Reproducibility: Ensuring all outputs are secure, reproducible, and suitable for scientific publication.</li></ul>

*Note.* Prompt to SEA: "Prepare a table describing your Instructions, Data, Tools, and Capabilities. The table should be suitable for publication in a science journal." Additional features include uploading data, downloading the conversation, interpreting images, and searching literature (a sample of publications about sea level science are included in the application).

problems; however, the IDEA framework is not agentic because it lacks the ability to interact with itself or other agents (Plaat et al., 2025). Adapting an IDEA such as SEA to analyze other geoscience data requires only rewriting the LLM instructions, as we did to create the Mars example.

## 2.1. Instructions

IDEA applications are designed to facilitate the examination, visualization, and interpretation of geoscience data, such as sea level observations in the case of SEA. Their functionality is governed by a structured set of instructions delivered to the LLM. When a user initiates an IDEA with their first prompt, two hidden prompts (called the system prompt and custom prompt) are automatically transmitted to the LLM. Together, these prompts contain approximately 5,300 tokens of chunked text for SEA, incurring an initial inference cost of about \$0.01 at current OpenAI rates using the gpt-4o model.

The system prompt provides overarching guidance on the assistant's purpose, behavior, and capabilities, whereas the custom prompt supplies context-specific information about the current computing environment, such as the list of available functions and whether the model has vision support. These prompts remain hidden from the conversation unless explicitly requested by the user. For reference, the complete system and custom prompts used in the SEA web application and the Mars IDEA are provided in the Data Availability Statement. Because IDEA applications are conversational and retain the context of prior interactions, much of this initialization content must be resent with each exchange. To reduce the associated overhead, OpenAI's token-caching feature is automatically invoked, lowering the cost and latency of follow-up responses.

Through experience, we found that developing effective system and custom prompts requires some trial and error. As LLMs improved, we were able to simplify the prompts—for example, newer models could retrieve the current date autonomously through the built-in Python interpreter, eliminating the need to hard-code that information. In general, we found it effective to begin with concise instructions and iteratively refine them by asking the LLM itself how its workflow might be improved.

The IDEA software is implemented primarily in Python, and the system prompt instructs the LLM to take advantage of scientific libraries commonly used in geoscience analysis, such as Pandas, NumPy, and Matplotlib. Users may also request code in other programming languages for offline use. The LLM is further instructed through the custom prompt to apply a set of prebuilt functions for specialized tasks as described in Sections 2.3 and 2.4. These capabilities allow users to ask questions such as “What is the rate of sea level rise measured by this tide gauge?” and receive scientifically formatted results, including visualizations suitable for communication or publication. Although the LLM can carry out sophisticated analyses, mistakes sometimes occur, and users are advised to verify important results, as discussed in Section 3.

## 2.2. Data

Metadata embedded in the system prompt, such as data organization details and tips for handling missing values, enhances the LLM's ability to generate accurate and reliable code. IDEA applications are programmed via NLP instructions to access various data types, including CSV, JSON, and NetCDF files. For SEA, the focus is on water levels, sea level variability, and other geophysical data such as climate indices. Data access is achieved through multiple means: direct file reads from the host server (i.e., UHSLC in the case of SEA), cloud-based data services accessed via LLM-generated Python code (e.g., using APIs, ERDDAP, or OPeNDAP), internet downloads (e.g., using wget commands), and user-uploaded files. However, SEA and other IDEA applications currently lack autonomous web search capabilities; they cannot discover or retrieve new data on their own without explicit system instructions or user prompting.

SEA focuses on analyzing coastal water level observations from tide gauges. It has access to all publicly available data at UHSLC, primarily via the ERDDAP server (<https://uhslc.soest.hawaii.edu/erddap/tabledap>). This includes a comprehensive global data set of coastal water level observations suitable for scientific research. UHSLC's Fast Delivery (FD) archive, updated every 1–2 months, contains hourly and daily water levels from over 200 active tide gauges. FD data are periodically replaced by Research Quality data, which undergo rigorous manual quality control to ensure long-term vertical reference frame stability (see Acknowledgments). SEA also incorporates derived data products such as tidal predictions and datums (e.g., MSL as well as Mean Higher High Water, MHHW), which are available for download from Station Explorer. These resources allow SEA to perform

datum conversions and frame water level outlooks relative to different reference points. SEA can calculate differences between current and target datums, enabling analyses of coastal flooding risks tied to tidal and elevation thresholds.

To enhance its capability to explore physical drivers of sea level variability, SEA is connected to climate indices available online (see Table 1 and the Acknowledgments). It also accesses ancillary data sets, such as a satellite altimetry calculation of monthly anomalies, processed locally at UHSLC using global ocean gridded Level 4 sea surface heights (CMEMS, 2025). These data enable SEA to perform advanced analyses, such as comparing sea level changes observed by tide gauges versus satellite measurements. Additionally, users can upload their own data in common formats (e.g., elevation survey results in a CSV file) for direct comparison with tide gauge observations. Expanding SEA's data scope is straightforward and involves updating its instructions to handle new data sets, with the example of creating an IDEA to analyze the Mars atmosphere demonstrating the flexibility of this framework.

In the system prompt for the Mars IDEA, we instructed the LLM that it is “a data scientist specializing in analyzing observations from the InSight Mission, with a focus on atmospheric conditions on Mars” (see the Data Availability Statement for acquiring the complete documentation). Brief description of NASA's Planetary Data System followed in the instructions along with a directory listing of web addresses for downloading data that we were curious to learn more about. Deploying this IDEA locally on a laptop via the Windows Subsystem for Linux allowed us to iteratively write the system prompt, adding notes to make the application more efficient (e.g., “check ./data/InSight directory for locally stored files. If the data is not found, download it from the remote source and store it locally”) while minimizing unnecessary information. The final system prompt of the Mars IDEA consisted of about 2,200 tokens, which is much less than for the SEA application (Section 2.1) but still includes a substantial section describing educational materials that could be interesting to ask it about (e.g., perform a visual interpretation of camera imagery from InSight or explain what is unrealistic about “The Martian”; not shown).

### 2.3. Tools

IDEA applications help users analyze data through two primary methods: generating and executing Python code, or invoking pre-built tools for specific tasks. For tasks requiring precision, prebuilt functions streamline workflows by minimizing errors. For instance, SEA is instructed to use a custom function, `get_climate_index`, to load the Oceanic Niño Index (ONI; B Huang et al., 2017) from a specified NOAA source when a user queries El Niño–Southern Oscillation (ENSO) data. For tasks requiring greater flexibility, such as searching for tide gauge data on UHSLC's ERDDAP server, SEA generates Python code on request.

Vision capabilities are inherent to the IDEA framework, because of our choice to use a multimodal LLM (Wang et al., 2024). The ability to understand and describe images is a feature of the gpt-4o model (OpenAI et al., 2024), which we incorporate into the example applications via the custom prompting described in Section 2.1 (i.e., “You can view images directly...”). Uploading an image as a JPG, PNG, or TIF file to SEA, or any other IDEA using such an LLM will prompt a visual interpretation of its contents in the context of the application's system prompt and the conversation history. This vision capability proves to be especially useful for asking the LLM to recreate a previous graphical analysis (see Section 3.3) or to check its own work.

In addition to analytical tools, IDEA includes a Retrieval-Augmented Generation (RAG) feature for extracting knowledge from scientific literature. This capability integrates the open-source PaperQA2 application (Pearson, 2024; Skarlinski et al., 2024), which in the case of SEA answers user queries using a small, pre-indexed collection of archived PDFs about sea level science. To reduce computational costs associated with processing full-text articles, we configure PaperQA2 to use a compact model (gpt-4o-mini-2024-07-18), keeping per-query costs near \$0.01 to answer a question like “What are the climate drivers of sea level variability in the tropical western Pacific?” This is approximately 10 times cheaper than using the full gpt-4o model (our Mars IDEA example uses the larger model throughout; see Section 3.3). Currently, SEA's web-accessible instance includes only a limited set of papers for demonstration purposes. However, local deployments of IDEA can be easily configured to index and search additional documents, simply by uploading new PDFs to the PaperQA2 backend. Users may invoke the RAG feature by prompting SEA or any other IDEA to “perform a literature search” about their question.

## 2.4. Computing

IDEA's computing environment is designed to ensure robust performance, security, and reliability. It operates using Open Interpreter software (<https://github.com/openinterpreter/open-interpreter>) within a Linux-based Docker container that manages CPU and memory resources according to task demands (e.g., up to 8 CPU cores and 5 GB of RAM were available for the example analyses performed by SEA). Computing capacity can be scaled as needed, although this remains an important consideration when deploying an IDEA for broader public use.

By default, a Docker container isolates the IDEA execution environment from the host operating system, which is the case for the production deployment of SEA. The container uses Docker-managed storage volumes and has no direct access to the host filesystem at UHSLC. This protects both the host system and remote users from unintended or malicious code execution. However, for local deployments of an IDEA (e.g., on a personal computer during development), we provide a different Docker configuration that bind-mounts the project directory into the container. This setup offers convenience, allowing developers to see live changes without rebuilding the software image, but also grants the LLM write access to parts of the host file system. Therefore, additional caution is warranted, as a poorly constructed prompt or unreliable LLM could theoretically perform unintended destructive operations (e.g., deleting files). These trade-offs are documented in the GitHub instructions for deploying an IDEA (see the Data Availability Statement for information about acquiring the source code).

Open Interpreter supports error handling and self-correction, enabling IDEAs to identify and resolve issues autonomously (e.g., SEA sometimes notifies users of a coding mistake and automatically makes revisions). These capabilities align the IDEA framework with commercial tools such as OpenAI's Assistants Code Interpreter (<https://platform.openai.com/docs/assistants/tools/code-interpreter>) while extending functionality with domain-specific customizations (i.e., instructions, direct data access, analysis tools, and internet access) as well as support for various LLMs, including both commercial and open-source models. Care should be taken to select safe and reliable models, especially in local deployments.

Security is a critical aspect of IDEA's design. The LLM is instructed to perform security checks for all external software installations, and the Open Interpreter scans downloaded code for vulnerabilities. Although SEA was designed as a temporary, session-based tool where user interaction histories are discarded after an hour unless explicitly saved, we have begun storing selected inputs and outputs to help improve SEA services at UHSLC and through the LLM inference provider. This is disclosed to users via the disclaimer footer on the application web page: "Inputs may be used to improve services by UHSLC and/or LLM providers." Users are encouraged to export important results as a PDF using the built-in "Download conversation" tool (see Supporting Information S1 for examples) or using their web browser to save the html page. These evolving practices reflect a balance between maintaining user privacy and improving the utility and performance of IDEA applications for geoscientific data exploration.

## 3. Results

Here, we describe examples of working with the SEA web application and a local instance of the IDEA software to analyze geoscience data and communicate information about multiple types of observations. These results are presented as three case studies to illustrate how geoscientists (i.e., the intended users) may interact with SEA or any other IDEA through prompts to the LLM, which utilizes associated instructions, data, tools, and computing resources (Figure 1). The first example shows SEA's capability to analyze and visualize sea level variability over time (3.1), whereas the second demonstrates how SEA integrates user-provided data to relate extreme water levels to potential impacts (3.2). Focusing these examples on the Republic of Palau in the tropical western Pacific was motivated by our involvement in analyzing sea levels for the 2020 Pacific Islands Regional Climate Assessment (PIRCA) process (Miles et al., 2020) as well as from questions asked to us during a sea level training workshop, which was recently held there in conjunction with a climate services dialog (Brewington et al., 2024). We will task SEA with updating analysis from the PIRCA report and then work through an exercise like part of the training workshop (3.2). The third example involves using a local instance of the IDEA on a personal computer to analyze an entirely different domain of geoscience—Mars atmospheric observations—which will demonstrate the generalizability of the framework to assist with many types of science activities (3.3).

Each example begins with a summary written by SEA or the IDEA instance, which is presented in *italics* and minimally edited by the authors (indicated by bracketed plain text). Additional interpretations and insights by the authors follow in regular font. Full conversations for all three examples are included in Supporting Information S1. Although the SEA and IDEA applications are designed to function almost deterministically through a low “temperature” parameter (0.2), exact reproduction of conversations is unlikely due to inherent variability in LLM outputs.

### 3.1. Sea Level Variability

*In this analysis (Figure 2), we examined the monthly maxima of sea levels at UHSLC station 007 (Malakal, Palau), referenced to the MHHW datum, using hourly sea level observations. The data set spans from 1969 to 2024, and only months with at least 90% hourly data coverage were included to ensure data reliability. The monthly maxima were calculated and plotted, with time values centered on the 15th of each month for temporal accuracy. To contextualize the data, La Niña periods were identified using the Oceanic Niño Index (ONI  $\leq -0.5$  for at least five consecutive overlapping 3-month seasons) and highlighted with light blue shading. The results, visualized in a publication-quality figure, reveal temporal variability in extreme sea levels, with notable peaks during certain La Niña events. The dark orange line represents the monthly maxima, emphasizing the variability in sea level extremes over the study period. This analysis provides a robust framework for understanding the relationship between sea level extremes and climate variability, particularly during La Niña conditions.*

The first case study demonstrates how SEA can be used to analyze sea level data in the context of climate variability. The conversation with SEA involved a sequence of prompts to guide the analysis culminating in the visualization shown in Figure 2. Below are the prompts used:

1. Use the hourly observations to plot a timeseries of the monthly maxima since the beginning of the record. Reference the water levels to the MHHW datum. Highlight La Niña conditions using light blue shading.
2. Only show months with at least 90% of hourly water levels. For the plotting, limit the time axis to when sea level data exists. Remove the legend.
3. Style the figure for publication. Color the water levels dark orange. Use a larger font (8 pt or greater). Write a figure caption suitable for scientific literature.
4. I noticed that you centered the monthly water levels at the end of each month. For plotting, center everything on the 15th of each month. Also, make all of the fonts bigger by 2 pt.
5. Review this conversation. Then summarize the analysis and results in one paragraph. The paragraph should be detailed and thorough, written in a style for a scientific publication.

SEA successfully generated the analysis and visualization (Figure 2), relating extreme sea levels at Malakal, Palau with ONI as a proxy for ENSO. Notable peaks in sea level extremes were reported by SEA as occurring during La Niña events. One minor issue arose when the monthly values were initially centered at the end of each month. After prompting, SEA corrected the analysis and re-centered the values monthly on the 15th date (the conversation in Supporting Information S1 includes python code generated at each step). Although this example highlights SEA’s capability to analyze sea level variability and communicate effectively, the user prompts are clearly from a trained climate scientist who is already familiar with the UHSLC data holdings. Further user testing across diverse scenarios would help quantify its utility and refine its robustness.

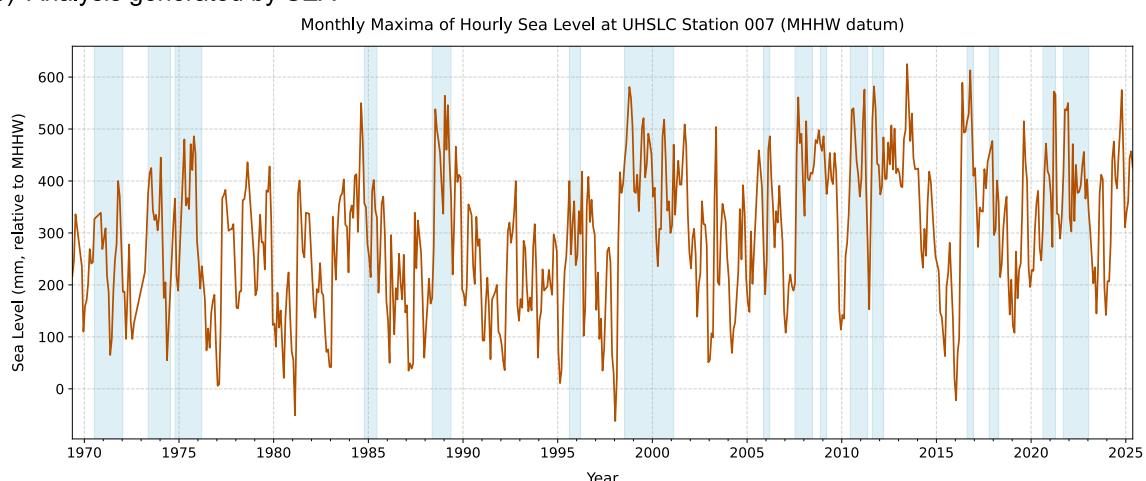
### 3.2. Relating Sea Levels to Impacts

*This analysis (Figure 3) examines the annual frequency of high water hours at Malakal Island, Palau using an impact-based flood threshold to assess potential flooding events. The results, visualized in the provided figure (caption from Figure 7 of Miles et al., 2020), reveal a significant temporal variability in the number of high water hours per year from the 1970s to the present. The bars represent the annual count of hours when water levels exceeded the flood threshold, whereas the shaded region illustrates the uncertainty range associated with the measurements. Notably, there is an increasing trend in high water hours over the decades with pronounced peaks in recent years, suggesting a heightened risk of coastal flooding. This trend aligns with broader observations of rising sea levels and increased storm surge impacts due to climate change. The uncertainty range highlights the variability in the data emphasizing the need for continued monitoring and refinement of flood thresholds to improve predictive capabilities and inform mitigation strategies.*

- a) User prompts SEA and it communicates the analysis plan

The screenshot shows a web-based interface for the Station Explorer Assistant. At the top, there's a green button labeled "Download conversation". Below it, the title "Station Explorer Assistant" is displayed, followed by a dropdown menu set to "Station: 007 Malakal, Palau". A "Start over" button is at the top right. A large text area contains a user prompt: "Use the hourly observations to plot a timeseries of the monthly maxima since the beginning of the record. Reference the water levels to the MHHW datum. Highlight La Niña conditions using light blue shading." Below this, a white box contains the generated analysis plan: "Here's my plan to fulfill your request: 1. Retrieve the full span of hourly sea level data for station 007 from the UHSLC Fast Delivery (FD) database. 2. Retrieve the MHHW (Mean Higher-High Water) datum value from the station's datum table. 3. Convert all hourly sea level data from Station Zero datum to the MHHW datum. 4. Aggregate the hourly data to compute the monthly maxima. 5. Retrieve the Oceanic Niño Index (ONI) to identify La Niña periods (ONI ≤ -0.5°C for at least 5 consecutive overlapping 3-month seasons). 6. Plot the monthly maxima time series (referenced to MHHW), and shade La Niña periods in light blue." Further down, another white box lists steps: "Let's start by: • Loading the station metadata to determine the FD span for station 007. • Retrieving the MHHW datum value." A footer note says "Proceeding with these steps now." At the bottom, there's a text input field with a microphone icon and placeholder text "Type your message, click mic to dictate, or attach/drop a file...", and three buttons: "mic", "Send", and "Stop". A small note at the very bottom says "SEA can make mistakes — check important results. Inputs may be used to improve services by UHSLC and/or LLM providers. [More info]".

- b) Analysis generated by SEA

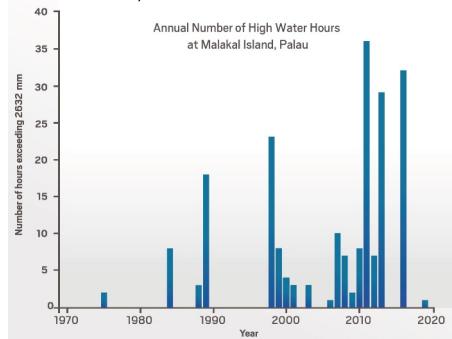


**Figure 2.** Example use of Station Explorer Assistant (SEA) to analyze water level observations. (a) The user selects a station (007 Malakal, Palau) and sends the prompt: “Use the hourly observations to plot a timeseries of the monthly maxima...” (light green highlight). SEA plans the analysis (white box of text) and informs the user that “SEA can make mistakes...” (footer text). The user refines the instructions with additional prompts, such as “Only show months with at least 90% of hourly water levels....” A green button allows downloading the full conversation (Supporting Information S1), including text and code from the large language model as well as plots. (b) SEA executes the generated code and produces the plot. Upon request, SEA also generates a figure caption, noting that the orange line represents monthly maximum sea levels and blue shading highlights La Niña conditions.

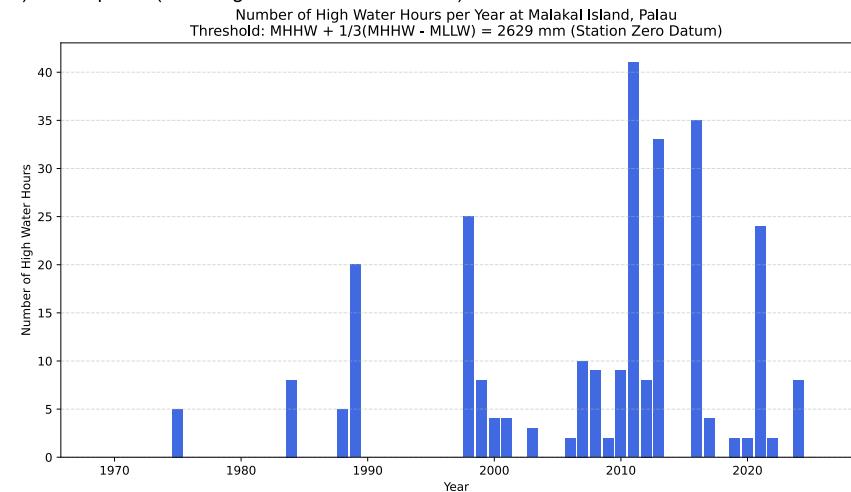
The second case study demonstrates how SEA can update a previously published analysis using new data, modify the methodology, and then recreate the graphical output. Specifically, SEA was tasked with reproducing the PIRCA water level analysis (Figure 3a) and extending it to include the latest FD hourly water level data (Figure 3b). A novel aspect of the SEA analysis is that it uses uploaded data (Supporting Information S1) to determine an impact-based flood threshold for an area near the tide gauge. The conversation consisted of the following prompts:

1. Use the full record of observations to make an updated figure for publication, which previously included this caption “Number of high water hours per year at Malakal Island in Palau from 1970 to 2019. The high water

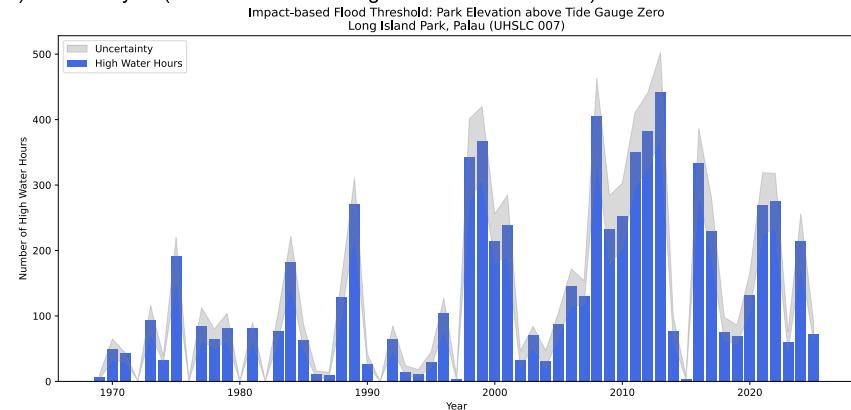
a) Original graphic (published in 2020)



b) SEA update (including the latest observations)



c) SEA analysis (water levels exceeding user-defined elevation)



**Figure 3.** Example use of Station Explorer Assistant (SEA) to update a sea level tracking analysis. (a) Reproduction of a previously published analysis (Figure 7 of Miles et al., 2020), which is associated with the caption provided in the first prompt of the second example (3.2). (b) Updated figure created by SEA using observations through May 2025. (c) Further analysis according to a high-water threshold that SEA defined based on user-provided elevation data.

threshold (2632 mm) is defined as the MHHW level plus one-third of the difference between that and the Mean Lower Low Water level at the tide gauge....”

2. Elevation survey data in a .xlsx (Excel) file is uploaded.
3. Based on the uploaded data, define an impact-based flood threshold using the surveyed height of the park above the tide gauge zero. Estimate the survey uncertainty by comparing the two different heights of the tide gauge benchmark. Use this new threshold and its uncertainty estimate to replot the yearly number of high water hours.

4. Review this conversation. Then summarize the analysis and results... for a scientific publication (same last prompt as in Sections 3.1 and 3.3).

SEA recreated the original analysis (Figure 3a) and updated it with more recent data, yielding a 3 mm change in the high water threshold (related to the latest tidal datums reported by UHSLC) and producing Figure 3b. Additionally, SEA interpreted elevation survey data provided by the user to define a new, impact-based flood threshold appropriate for the specified location (i.e., a park nearby the tide gauge identified as flooding often, which was surveyed with GPS equipment by participants in the sea level training workshop organized by UHSLC in March 2024). Using this updated threshold, SEA generated a visualization (Figure 3c) showing the annual number of hours since 1969 when the low-lying park was likely inundated. For further comparison of the original and updated analysis, we invite the reader to upload a screenshot of this figure to SEA along with the first prompt above. Such a new conversation with the assistant could proceed in various ways, for instance by the user requesting axis-scaling adjustments to more clearly compare results (Figures 3a and 3b).

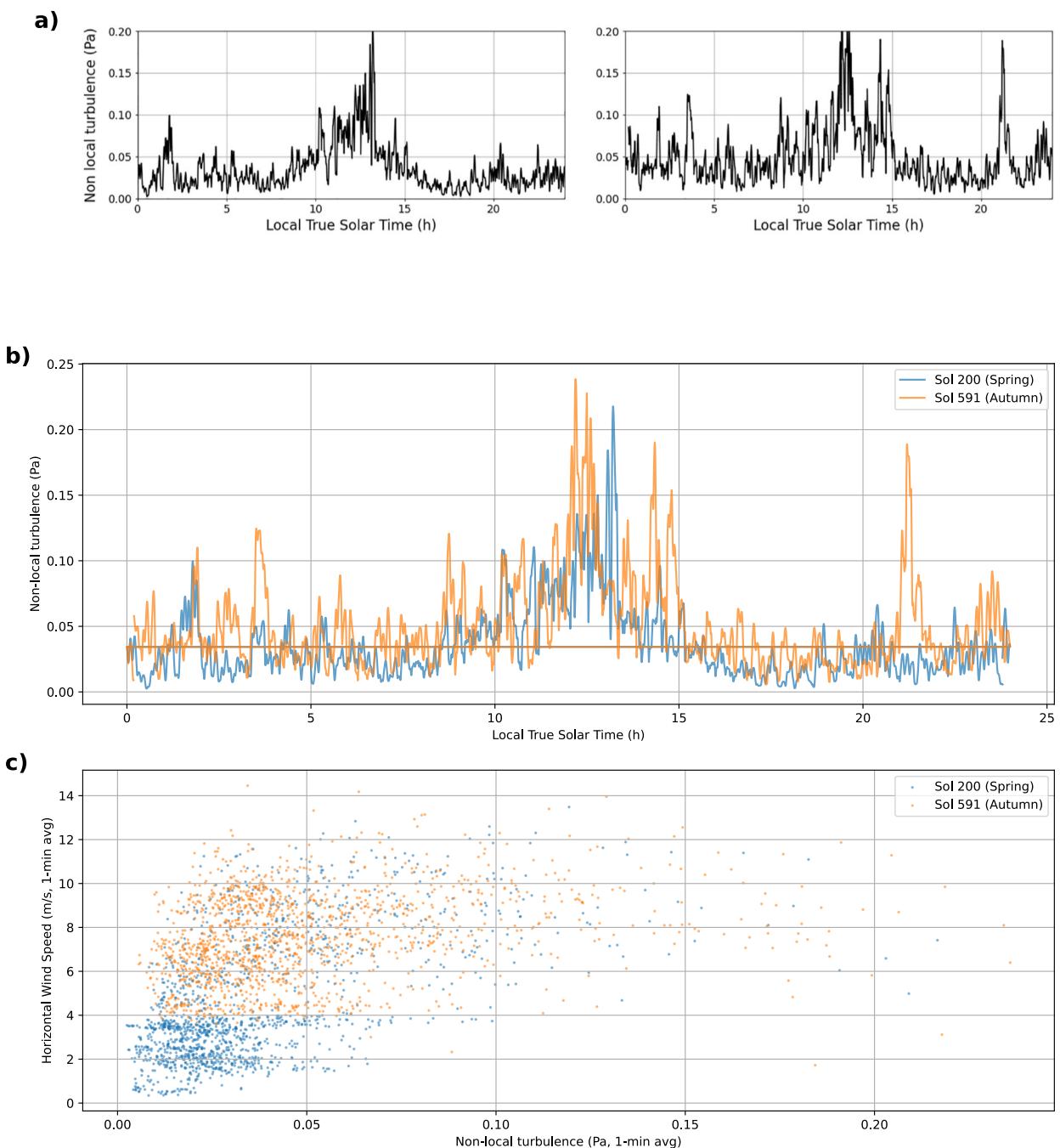
Although SEA effectively integrated new data and produced meaningful results, its summary omitted details about interannual variability in flooding events. Further interactions with SEA explored the relationship between ENSO and flooding, yielding responses such as there being a “relationship between ONI and the number of high water hours...” and that “during La Niña conditions, flooding counts tend to increase.” Asking SEA to perform a literature review at this stage of the conversation provided more nuanced information about the climate drivers of high sea levels in the tropical western Pacific, along with relevant citations (i.e., by returning the response from PaperQA2; Supporting Information S1).

### 3.3. Investigating a Different Geoscience Domain

*In this analysis, we investigated non-local turbulence and its relationship with wind speed on Mars using data from the InSight mission for Sol 200 (spring) and Sol 591 (autumn). Nonlocal turbulence was computed by subtracting the 1,000-second moving average of pressure from the 100-second moving average followed by calculating the standard deviation over a 500-second moving window. The results, plotted as a function of Local True Solar Time (LTST), revealed that turbulence peaks during the daytime, around 12–13 hours LTST, and diminishes at night, consistent with convective activity driven by solar heating. Wind speed data from the TWINS instrument were resampled to 1-minute averages and analyzed against nonlocal turbulence. Scatter plots demonstrated a positive correlation between wind speed and turbulence, with Sol 591 exhibiting higher wind speeds (up to 12 m/s) compared to Sol 200 (up to 6 m/s). (Peak wind speeds appear to be similar, although more calm moments occurred during Sol 200.) These findings suggest seasonal variations in atmospheric dynamics, with stronger winds and turbulence observed in autumn. The analysis was validated by comparing recreated turbulence plots with the original figure from Chatain et al. (2021), and a publication-quality figure was prepared (Figure 4), integrating the original data, recreated turbulence plots, and wind speed analysis.*

The third case study utilizes the Mars version of an IDEA to demonstrate how this framework (Figure 1) is adaptable to a wide variety of geoscience data. Although the relatively brief instructions provided to the Mars IDEA (Section 2.2) minimized the time and cost of initiating a conversation, we noticed that more subsequent prompting was required to yield intended results of an analysis, compared to the examples with SEA. The conversation with the Mars IDEA consisted of the following prompts:

1. I uploaded Fig2bottom\_Chatain\_et al.\_GRL\_2021\_InSightTurbulence.JPG
2. The images are from spring (Sol 200) and autumn (Sol 591). Perform a detailed literature review to learn the exact methodology for this analysis.
3. I noticed that the authors stated the following in their methodology: “Nonlocal turbulence is obtained by subtracting the 1,000 s-window-averaged pressure to the 100 s-window-averaged pressure, before computing standard deviation of the result over a moving window of 500 s.” Recreate the analysis.
4. Compare the plots you produced to the original figure that I uploaded. Style your figure like the original.
5. Make sure that this is in fact plotting results as a function of LTST.
6. Consider the wind speed observations on these sols. Plot the wind speed as a function of Nonlocal turbulence. Use 1-min averaging of winds and turbulence (sic).
7. Prepare a figure for publication with the following: Top- uploaded image (stretched across two columns) Middle- recreated analysis plots Bottom- wind speed analysis plots.
8. Panel A plots should not be duplicated. Rather, simply stretch the image across the figure.



**Figure 4.** (a) Non-local turbulence as a function of Local True Solar Time (LTST) for spring (Sol 200) and autumn (Sol 591) on Mars (reproduced from Figure 2 of Chatain et al., 2021). (b) Recreated analysis showing non-local turbulence (Pa) as a function of LTST for Sol 200 (blue) and Sol 591 (orange). The turbulence peaks during the daytime, around 12–13 hr LTST, with minimal activity at night. (c) Scatter plots of horizontal wind speed (m/s) as a function of non-local turbulence (Pa) for Sol 200 (blue) and Sol 591 (orange). Higher wind speeds are associated with increased turbulence with Sol 591 exhibiting stronger winds overall. (Figure and caption generated by the Mars Intelligent Data Exploring Assistant.)

9. I like this new layout. Style the line colors in panel B like you color the dots in panel C.
10. Prepare a figure caption for publication.
11. Review this conversation. Then summarize the analysis and results... for a scientific publication (same last prompt as in Sections 3.1 and 3.2).

Like the preceding conversations with SEA (Sections 3.1 and 3.2), the Mars IDEA example requests data analyses and interpretation of the results. The first prompt uploaded an image cropped from Figure 2 of the Chatain et al. (2021) study (reproduced in our Figure 4a), which the LLM visually interpreted as representing “turbulence data collected by the InSight mission on Mars.” Minor spelling mistakes by the user (i.e., “turbulence” in the first and sixth prompts) were ignored by the LLM. Chatain et al. (2021) classified the Mars atmosphere during the local spring and autumn, focusing on Sol 200 and Sol 591, respectively (the system prompt to this IDEA explained that sols are a unit of time since the InSight mission landed on Mars). Since we are not familiar with the study of Mars weather, we relied on the IDEA to learn about methods appropriate for analyzing turbulence. This learning was partly accomplished by the second prompt directing a literature review, which invoked the PaperQA2 function call. We found that simply prompting with the most relevant part of Chatain et al. (2021) (i.e., based on our human review; third prompt) led the Mars IDEA to generate more reliable code, compared to prior testing with only the PaperQA2 response. (The moving window approach to isolating non-local turbulence was more robust for the short time series, compared to the band-pass filtering method that the RAG literature review usually emphasized.) In any case, by following prompts 4–11, the Mars IDEA recreated the turbulence analysis almost exactly (Figure 4b), performed a novel analysis of how the wind speed varied as a function of turbulence (Figure 4c), and documented the results (*italicized* paragraph above).

Besides our concern about filtering methods, which necessitated the clarification in prompt 3, we noticed a couple minor problems with the Mars data analysis and its interpretation. There appears to be a subtle plotting error in Figure 4b (i.e., a fictitious horizontal line), which we retained to illustrate a limitation of using this technology to analyze complex data sets. Separately, we needed to correct the summary paragraph concerning interpretation of the wind speeds (bracket of nonitalicized text). Interestingly, review of the full conversation (Supporting Information S1) shows that the initial visual interpretation of the scatterplots was fairly accurate (Figure 4c; “Sol 200: Wind speeds are generally lower, with most values below 6 m/s. Sol 591: Wind speeds are higher, with values reaching up to 12 m/s.”). When prompted to write the final summary, the Mars IDEA appears to misconstrue its prior interpretation, writing instead: “Sol 591 exhibiting higher wind speeds (up to 12 m/s) compared to Sol 200 (up to 6 m/s).” The deficiency describing peak wind speeds on Sol 200 (i.e., failing to mention that the gusts are actually similar in magnitude on both sols) perhaps stems from inherent characteristics of how the LLM summarizes a conversation (OpenAI et al., 2024), which motivates the need for careful review until more advanced models are incorporated into future IDEAs.

#### 4. Discussion

Users of the SEA web application or local deployments of the IDEA software are informed that it can make mistakes and are encouraged to check important results (Figure 2a). During testing, we often repeated a series of prompts, such as the examples above, to assess the stability of conversations (Stall et al., 2023). Then, we would verify the accuracy of the analysis by comparing the AI-generated code against our own data analysis techniques. Overall, SEA and the Mars IDEA exhibit the desired behavior of following user prompts to achieve the intended analysis; however, variations in performance did occur between conversations. Although SEA and the other IDEA example occasionally made major errors, these were typically easy to address through subsequent prompting of feedback (e.g., “considering that no information is displayed in the plot, reevaluate your procedure for analyzing the data”). We also had success employing chain-of-thought prompts (Wei et al., 2022), like “first calculate the monthly means of the full record, then remove the annual cycle, and finally calculate what month usually has the largest variability,” which seemed to improve the LLM’s reasoning capability enough to complete analyses that it otherwise struggled with.

Successful uses of SEA and forthcoming deployments of IDEA in other geoscience domains rely on the users’ willingness and ability to identify analysis mistakes (Zhou et al., 2023) and provide corrective prompts (i.e., feedback), at least in the current form of this framework (Figure 1). This need for verification and user-driven correction underscores why the applications demonstrated here were designed to display computer code generated by the LLM. Beyond transparency, the assistant applications showing their work serve educational purposes, as users are encouraged to refine the code offline (e.g., via a “Copy” button) or request translations into other programming languages.

Future improvements to the IDEA framework (illustrated by dashes in Figure 1) are likely to reduce the occurrence of errors. More seamlessly incorporating the vision capability to review plots (e.g., automatically

verifying that data is being displayed) would lessen the number of times the user must prompt for corrections. Updating SEA to use a newer LLM with better problem-solving capabilities (gpt-4.1-2025-04-14 from OpenAI) enhanced its ability to follow complex prompts in one shot (e.g., seamlessly recreating Figure 2b). The newer version is available online for testing, which we used to generate higher-resolution figures for publication by asking SEA to follow similar prompts as before. Testing IDEA with less powerful LLMs from OpenAI (e.g., gpt-4o-mini) resulted in much lower performance on complex tasks, like the analyses used to make Figures 2–4. Leveraging more advanced LLMs that feature enhanced reasoning capabilities and optimizations for assisting with science (e.g., o3-mini; OpenAI, 2025) could reduce how many initial instructions are necessary (i.e., shorter system prompts), negate the need for chain-of-thought prompting to solve complex problems, and supervise the conversation for accuracy; however, our tests with “reasoning” models (e.g., o3-mini) were slower and more expensive. Another opportunity for improvement is to incorporate feedback from conversations to refine the system prompt, which could be done automatically by an LLM reviewing the users' prompts.

Rapid advancements in LLMs (Ananthaswamy, 2024; Buback et al., 2023; Plaat et al., 2025) and their growing utility in data analyses (e.g., Microsoft Research, 2023), exemplified by our demonstrations of SEA and the Mars IDEA instance, have the potential to disrupt established practices in scientific research (Gabriel et al., 2025). We showed how SEA enables trained users (i.e., geoscientists experienced with UHSLC data) to quickly and reliably analyze sea levels with minimal expense of time or money. Deploying the Mars IDEA required only writing a system prompt to instruct how to access this domain of geoscience information, with the instructions being iteratively refined through our conversations with the LLM, since we were not initially familiar with nuances of the InSight mission data sets.

Compared to existing LLM-powered tools in other domains of geoscience—such as various “Climate GPT” prototypes or proprietary assistants embedded in platforms like ArcGIS from Esri—the IDEA framework offers a more flexible and transparent approach, accessible for research testing via the SEA web application or local deployment via instructions on GitHub. Most existing tools for AI-assisted data exploration either rely on hard-coded workflows or remain behind commercial barriers, limiting customizability and broad accessibility. In contrast, IDEA is designed to be generalizable across disciplines and data sets, with a customizable system prompt and flexible access to computing resources, whether deployed on the web or run locally on a user's machine. However, the IDEA framework also has known limitations. In particular, its performance depends on the clarity and completeness of the system instructions, which must be recrafted for each domain and use case. This sensitivity presents both a challenge and an opportunity for future improvements.

New techniques for connecting LLMs to data, such as the Model Context Protocol (MCP; Anthropic, 2025), as well as emerging web search capabilities, are poised to streamline the development of future IDEA applications. These advances may help bridge the current gap between IDEA and more specialized agents by enabling richer context awareness and dynamic retrieval of relevant data sets. Nevertheless, ensuring transparency, reproducibility, and domain-specific reliability remains a challenge across the field and will be a key focus of our future work.

The examples presented here were completed in just a few minutes, with the LLM service cost associated with preparing Figures 2–4 amounting to about \$1. It is conceivable that more agentic LLMs (Boiko et al., 2023) could independently employ applications based on SEA or other IDEAs, perhaps communicating via MCP, further reducing the human effort required for scientific inquiries. A substantial amount of guidance was provided to the LLM via user prompts that, especially for the sea level examples, resemble instructions one might provide for someone new to exploring such data. Delivering instructions more politely is common practice in professional dialog (e.g., “please consider the data completeness...”), and there is some evidence that prompting with such tone could improve the LLM performance (Yin et al., 2024) as well as the human experience (Ribino, 2023).

One of the IDEA framework's most valuable features is its potential to lower barriers for geoscientists initiating investigations, particularly by creating opportunities to explore unfamiliar data sets (Foroumandi et al., 2023). For example, SEA can help graduate students conduct sea level research while simultaneously learning new skills, although best practices should be developed for training in the use of this technology. Likewise, we demonstrated using an IDEA to analyze a data set completely unfamiliar to us (i.e., observations from the InSight mission). The Mars IDEA successfully acquired relevant data, applied methods from a domain-specific study (Chatain et al., 2021), and further examined how wind speeds relate to atmospheric turbulence on our neighboring planet. Potential benefits of the IDEA framework toward advancing education and research opportunities may mitigate

some of the environmental costs associated with utilizing LLMs for such tasks (Gabriel et al., 2025). However, there remain many unknowns about the usability and utility of this technology by geoscientists, which should be evaluated.

As LLMs continue to improve in scientific tasks such as code generation and problem solving (Ananthaswamy, 2024), we anticipate that future IDEAs could profoundly transform academic mentorship and research in general. If AI tools can assist with research typically carried out by graduate assistants, often at lower cost, new mechanisms for sponsoring education and scientific inquiry may become necessary. The authors acknowledge that similar considerations apply to the potential for AI tools to take over aspects of our own work—and perhaps yours as well.

## Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

## Data Availability Statement

All the sea level data discussed in this study are available for download via requests to SEA (<https://uhslc.soest.hawaii.edu/research/SEA>). The source code and instructions for implementing an IDEA like SEA as well as the example to analyze observations from NASA's InSight mission on Mars are available at <https://github.com/uhsealevelcenter/IDEA> (Widlansky & Komar, 2025). The system prompts and custom instructions for SEA and the Mars IDEA are available in the "utils" directory of the GitHub repository.

## Acknowledgments

The usefulness of AI applications such as SEA and the Mars IDEA depends on high-quality data, which relies on the efforts of dedicated teams working at organizations such as UHSLC, NOAA, and NASA. We extend our gratitude to the scientists and technical support personnel at other organizations for making their data accessible to SEA via the internet. We acknowledge the data contributions of the Copernicus Marine Service, NOAA's National Center for Environmental Information and Physical Sciences Laboratory, NASA's Jet Propulsion Laboratory, and the University of Wisconsin-Madison's Atmospheric and Oceanic Sciences department. Mars atmospheric observations measured by NASA's InSight mission that were acquired in the IDEA example are from the Planetary Atmospheres Node of NASA's Planetary Data System. Funding for the development of SEA and the IDEA framework, including API charges for LLM services from OpenAI, was primarily provided through the grant NA21OAR4310308 by the Pacific RISA program, a NOAA Climate Adaptation Partnership (CAP) team. SEA and the Mars IDEA were utilized in writing this study with further clarification provided by OpenAI's ChatGPT.

## References

- Adapting to AI. (2024). Nat. Geosci. Vol. 17(481). <https://doi.org/10.1038/s41561-024-01475-5>
- Ananthaswamy, A. (2024). How close is AI to human-level intelligence? *Nature*, 636(8041), 22–25. <https://doi.org/10.1038/d41586-024-03905-1>
- Anthropic. (2025). Model context protocol, edited. Retrieved from <https://modelcontextprotocol.io>
- Banerdt, W. B., Smrekar, S. E., Banfield, D., Giardini, D., Golombek, M., Johnson, C. L., et al. (2020). Initial results from the InSight mission on Mars. *Nature Geoscience*, 13(3), 183–189. <https://doi.org/10.1038/s41561-020-0544-y>
- Baucon, A., & de Carvalho, C. N. (2024). Can AI get a degree in geoscience? Performance analysis of a GPT-based artificial intelligence system trained for Earth science (GeologyOracle). *Geoheritage*, 16(121), 121. <https://doi.org/10.1007/s12371-024-01011-2>
- Boerner, T. J., Deems, S., Furlani, T. R., Knuth, S. L., & Towns, J. (2023). Access: Advancing innovation: NSF's advanced cyberinfrastructure coordination ecosystem: Services & support, in *In practice and experience in advanced research computing (PEARC '23)*, edited, p. 4. <https://doi.org/10.1145/3569951.3597559>
- Boiko, D. A., MacKnight, R., & Gomes, G. (2023). Emergent autonomous scientific research capabilities of large language models. <https://doi.org/10.48550/arXiv.2304.05332>
- Brewington, L., Bryson, C., Frazier, A. G., Keener, V. W., Marra, J. J., Matsutaro, E. X., et al. (2024). A climate services dialog to build sector-based climate early warning systems in the Republic of Palau. *Bulletin America Meteorology Social*, 105(8), E1610–E1618. <https://doi.org/10.1175/BAMS-D-24-0157.1>
- Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., et al. (2023). Sparks of artificial general intelligence: Early experiments with GPT-4. <https://doi.org/10.48550/arXiv.2303.12712>
- Caldwell, P. C., Merrifield, M. A., & Thompson, P. R. (2015). *Sea level measured by tide gauges from global oceans — The joint archive for sea level holdings (NCEI accession 0019568)*, version 5.5. NOAA National Centers for Environmental Information. <https://doi.org/10.7289/V5V40S7W>
- Castelvecchi, D. (2022). Are ChatGPT and AlphaCode going to replace programmers? *Nature*. <https://doi.org/10.1038/d41586-022-04383-z>
- Castelvecchi, D. (2024). Researchers built an 'AI Scientist' — What can it do? *Nature*, 633(266), 266. <https://doi.org/10.1038/d41586-024-02842-3>
- Chatain, A., Spiga, A., Banfield, D., Forget, F., & Murdoch, N. (2021). Seasonal variability of the daytime and nighttime atmospheric turbulence experienced by InSight on Mars. *Geophysical Research Letters*, 48(22). <https://doi.org/10.1029/2021GL095453>
- Chen, Y., Wang, W., Lobry, S., & Kurtz, C. (2024). An LLM agent for automatic geospatial data analysis, <https://doi.org/10.48550/arXiv.2410.18792>
- CMEMS. (2025). Global ocean gridded L 4 sea surface heights and derived variables nrt. In *E. U. C. M. S. I. (CMEMS), marine data store (MDS)*. <https://doi.org/10.48670/moi-00149>
- Conroy, G. (2024). *Do AI models produce more original ideas than researchers?* Nature. <https://doi.org/10.1038/d41586-024-03070-5>
- Farquhar, S., Kossen, J., Kuhn, L., & Gal, Y. (2024). Detecting hallucinations in large language models using semantic entropy. *Nature*, 630(8017), 625–630. <https://doi.org/10.1038/s41586-024-07421-0>
- Florindo, F. (2023). ChatGPT: A threat or an opportunity for scientists? *Perspectives of Earth and Space Scientists*, 4(1). <https://doi.org/10.1029/2023CN000212>
- Foroumandi, E., Moradkhani, H., Sanchez-Vila, X., Singha, K., Castelletti, A., & Destouni, G. (2023). ChatGPT in hydrology and Earth sciences: Opportunities, prospects, and concerns. *Water Resources Research*, 59(10), e2023WR036288. <https://doi.org/10.1029/2023WR036288>
- Gabriel, I., Manzini, A., Keeling, G., Hendricks, L. A., Rieser, V., Iqbal, H., et al. (2025). The ethics of advanced AI assistants. <https://doi.org/10.48550/arXiv.2404.16244>
- Gu, K., Shang, R., Jiang, R., Kuang, K., Lin, R.-J., Lyu, D., et al. (2024). Blade: Benchmarking language model agents for data-driven science. <https://doi.org/10.48550/arXiv.2408.09667>

- Hadid, A., Chakraborty, T., & Busby, D. (2024). When geoscience meets generative AI and large language models: Foundations, trends, and future challenges. *Expert Systems*, 41(10), e13654. <https://doi.org/10.1111/exsy.13654>
- Hague, B. S., Murphy, B. F., Jones, D. A., & Taylor, A. J. (2019). Developing impact-based thresholds for coastal inundation from tide gauge observations. *Journal of Southern Hemisphere Earth Systems Science*, 69(1), 252. <https://doi.org/10.1071/ES19024>
- Hancock, D. Y., Fischer, J., Lowe, J. M., Snapp-Childs, W., Pierce, M., Marru, S., et al. (2021). Jetstream2: Accelerating cloud computing via jetstream. In *Practice and experience in advanced research computing (PEARC '21)*, Association for Computing Machinery, (pp.1–8). <https://doi.org/10.1145/3437359.3465565>
- Heidt, A. (2024). "Without these tools, i'd be lost": How generative AI aids in accessibility. *Nature*, 628(8007), 462–463. <https://doi.org/10.1038/d41586-024-01003-w>
- Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., et al. (2017). Extended reconstructed sea surface temperature version 5 (ERSSTv5), upgrades, validations, and intercomparisons. *Journal of Climate*, 30(20), 8179–8205. <https://doi.org/10.1175/JCLI-D-16-0836.1>
- Huang, J., & Tan, M. (2023). The role of ChatGPT in scientific communication: Writing better scientific review articles. *American Journal of Cancer Research*, 13(4), 1148–1154.
- Hutson, M. (2022). Could AI help you to write your next paper? *Nature*, 611(7934), 192–193. <https://doi.org/10.1038/d41586-022-03479-w>
- IOC. (2020). Quality control of in situ sea level observations: A review and progress towards automated quality control. *Manuals and guides*, 1(144), 71.
- Lenharo, M. (2024). ChatGPT turns two: How the AI chatbot has changed scientists' lives. *Nature*, 636(8042), 281–282. <https://doi.org/10.1038/d41586-024-03940-y>
- Li, Y., Choi, D., Chung, J., Kushman, N., Schrittewieser, J., Leblond, R., et al. (2022). Competition-level code generation with AlphaCode. *Science*, 378(6624), 1092–1097. <https://doi.org/10.1126/science.abq1158>
- Lin, Z., Deng, C., Zhou, L., Zhang, T., Xu, Y., Xu, Y., et al. (2024). GeoGalactica: A scientific large language model in geoscience. <https://doi.org/10.48550/arXiv.2401.00434>
- Lu, C., Lu, C., Lange, R. T., Foerster, J., Clune, J., & Ha, D. (2024). The AI scientist: Towards fully automated open-ended scientific discovery. <https://doi.org/10.48550/arXiv.2408.06292>
- Microsoft Research. (2023). The impact of large language models on scientific discovery: A preliminary study using GPT-4, AI4Science. <https://doi.org/10.48550/arXiv.2311.07361>
- Miles, W., Greini, Z., Matsutaro, E., Colin, P., Keener, V., Golbuu, Y., et al. (2020). Climate change in Palau: Indicators and considerations for key sectors. *Report for the Pacific Islands Regional Climate Assessment*. <https://doi.org/10.5281/zenodo.4124259>
- OpenAI. (2025). o3-mini system card. Retrieved from <https://cdn.openai.com/o3-mini-system-card.pdf>
- OpenAI, A. H., Lerer, A., Goucher, A. P., Perelman, A., & Coauthors (2024). GPT-4o system card. <https://doi.org/10.48550/arXiv.2410.21276>
- Pearson, H. (2024). Can AI review the scientific literature — And figure out what it all means? *Nature*, 635(8038), 276–278. <https://doi.org/10.1038/d41586-024-0240>
- Plaat, A., Stein, M. V., Preuss, M., Putten, P. V. D., & Batenburg, K. J. (2025). Agentic large language models. *a survey*. <https://doi.org/10.48550/arXiv.2503.23037>
- Ribino, P. (2023). The role of politeness in human–machine interactions: A systematic literature review and future perspectives. *Artificial Intelligence Review*, 56(Suppl 1), 445–482. <https://doi.org/10.1007/s10462-023-10540-1>
- Si, C., Yang, D., & Hashimoto, T. (2024). Can LLMs generate novel research ideas? A large-scale human study with 100+ NLP researchers. <https://doi.org/10.48550/arXiv.2409.04109>
- Skarbinski, M. D., Cox, S., Laurent, J. M., Braza, J. D., Hinks, M., Hammerling, M. J., et al. (2024). Language agents achieve superhuman synthesis of scientific knowledge. <https://doi.org/10.48550/arXiv.2409.13740>
- Stall, S., Cervone, G., Coward, C., Cutcher-Gershenfeld, J., Donaldson, T. J., Erdmann, C., et al. (2023). Ethical and responsible use of AI/ML in the Earth, space, and environmental sciences. *ESS Open Archive*. <https://doi.org/10.22541/essoar.168132856.66485758/v1>
- Stockel-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>
- Thorp, H. H. (2023). ChatGPT is fun, but not an author. *Science*, 379(6630), 313. <https://doi.org/10.1126/science.adg7879>
- Wang, J., Jiang, H., Liu, Y., Ma, C., Zhang, X., Pan, Y., et al. (2024). A comprehensive review of multimodal large language models: Performance and challenges across different tasks. <https://doi.org/10.48550/arXiv.2408.01319>
- Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., et al. (2022). Chain-of-Thought prompting elicits reasoning in large language models. <https://doi.org/10.48550/arXiv.2201.11903>
- Weir, A. (2011). The Martian. self-published.369.
- Widlansky, M. J., & Komar, N. (2025). uhsealevelcenter/IDEA: Building an intelligent data exploring assistant (IDEA) for geoscientists. <https://doi.org/10.5281/zenodo.15605301>
- Yin, Z., Wang, H., Horio, K., Kawahara, D., & Sekine, S. (2024). Should We Respect LLMs? A Cross-Lingual Study on the Influence of Prompt Politeness on LLM Performance. <https://doi.org/10.48550/arXiv.2402.14531>
- Zhou, J., Müller, H., Holzinger, A., & Chen, F. (2023). Ethical ChatGPT: Concerns, challenges, and commandments. <https://doi.org/10.48550/arXiv.2305.10646>