frontiers in Ecology and Evolution

TYPE Original Research
PUBLISHED 16 February 2026
DOI 10.3389/fevo.2026.1636308

OPEN ACCESS

EDITED BY
Darren Norris,
Universidade Federal do Amapá, Brazil

REVIEWED BY
Erin Bunting,
Michigan State University, United States
Haowei Hua,
Princeton University, United States

*CORRESPONDENCE
Sally C. Dillon
sally.c.dillon@usace.army.mil

RECEIVED 27 May 2025
REVISED 21 January 2026
ACCEPTED 29 January 2026
PUBLISHED 16 February 2026

CITATION
Dillon SC, Provost L and Tritinger A (2026) AI-driven insights on coastal management: a comparative discussion of the use of generative chatbots and natural infrastructure.
Front. Ecol. Evol. 14:1636308.
doi: 10.3389/fevo.2026.1636308

AI-driven insights on coastal management: a comparative discussion of the use of generative chatbots and natural infrastructure

Sally C. Dillon*, Leigh Provost and Amanda Tritinger

Coastal and Hydraulics Laboratory, US Army Engineer Research and Development Center (ERDC), Vicksburg, MS, United States

The emergence of cloud computing platforms has catalyzed the development of numerous generative artificial intelligence (AI) chatbots, offering users unprecedented access to information and reasoning capabilities at their fingertips, provided they have internet access. This manuscript explores AI's current perspectives on coastal management, focusing particularly on the implementation of natural infrastructure. By examining the insights provided by three widely used generative AI chatbots: ChatGPT (GPT-4), Google Gemini 1.5 Flash, and Microsoft Bing Copilot, the authors present a comparative analysis of AI-generated opinions on coastal management. The models were given identical prompts related to coastal management and natural infrastructure. The AI-generated responses were evaluated using a scored system assessing their accuracy, completeness, relevance, clarity, and depth. The analysis revealed significant variability in the performance of the three models. Bing Copilot, consistently delivered the most accurate and relevant responses. Gemini provided well-structured but often truncated answers, while ChatGPT's outputs were frequently generalized and lacked depth. A key finding was the presence of "response drift," where all models tended to reuse concepts from earlier in the conversation. Furthermore, the models' inability to access paywalled scientific literature and proprietary databases was identified as a critical limitation. The findings underscore the potential and limitations of these tools to inform and advance the field of natural infrastructure. The study highlights that the models lack the situational awareness and critical judgment necessary for complex environmental decision-making. The observed response drift and data access limitations can reinforce narrow perspectives and prevent a comprehensive evaluation of contested issues. A proposed framework suggests integrating AI within a cyclical and collaborative management process that prioritizes human interpretation and oversight to ensure contextually validated and robust outcomes.

KEYWORDS

coastal management, coastal resilience, decision support, generative AI, natural infrastructure

1 Introduction

Coastal management is at a crossroads, facing intensifying challenges driven by, urbanization, and ecological degradation (EPA n.d.; The Guardian, 2025). Rising sea levels, extreme weather events, and coastal erosion increasingly threaten both human communities and natural ecosystems (EPA n.d.). Traditional gray infrastructure strategies, such as seawalls and levees, while essential, often struggle to keep pace with these multifaceted and evolving threats. Moreover, these hard engineering approaches can have unintended consequences, such as habitat degradation and disruption of natural sediment transport processes (Yale Environment Review n.d.). In response, there is a growing interest in integrating nature-based solutions (NbS) into coastal resilience strategies. Natural infrastructures like wetlands, dunes, and mangroves offer adaptive benefits, including storm surge mitigation, shoreline stabilization, biodiversity enhancement, and carbon sequestration (NOAA n.d.). Coastal wetlands in the United States alone are estimated to provide $23.2 billion in annual storm protection services (NOAA n.d.). However, the successful implementation of NbS requires interdisciplinary collaboration and sufficient site-specific data (Center for Coastal Solutions n.d.).

Artificial intelligence (AI) is emerging as a powerful tool in environmental management, offering new capabilities to process vast datasets, simulate ecological scenarios, and optimize conservation interventions (Dalton, 2024). AI has already demonstrated its value in applications ranging from biodiversity monitoring to climate modeling (Financial Times, 2025). Yet, the role of AI in shaping guidance and decisions in coastal management, particularly in relation to nature-based infrastructure remains an evolving and under-explored domain.

Many AI chatbots are now available, each with its own strengths and weaknesses. Examples include ChatGPT, DeepSeek, Claude, Meta AI, Google Gemini, Microsoft Copilot, Poe, Perplexity, Le Chat Mistral, HuggingChat, Grok, Pi, and Merlin. PC Magazine recently published a comparative review of these tools, asking each chatbot the same series of questions and evaluating their accuracy, consistency, complexity, and depth, along with core features such as table creation, citation support, and summarization (Forlini and Circelli, 2025). The authors emphasized that this type of assessment has clear limits: evaluating an AI model's performance across the full range of possible topics would require an "army of experts," and even then, performance may shift from one day to the next due to updates. Still, their testing revealed distinct differences, and they noted that users will likely form their own conclusions after spending significant time with each tool. Their rankings listed ChatGPT as Best Overall, Google Gemini as Best Value, Microsoft Copilot as Best for Windows Users, and Claude as Best for Privacy.

This manuscript investigates contemporary perspectives on coastal management as generated by three widely used generative AI chatbots. By analyzing their responses to a series of structured prompts, the study provides a comparative discussion of AI-generated insights and evaluates the use of this technology within the framework of coastal management projects. A systematic scoring metric is employed, assessing each response across five key dimensions: accuracy, completeness, relevance, clarity, and depth. The authors hope that the discussion of their findings will lead to reflection of coastal practitioners on how AI should and successfully can be utilized in their workflows.

The results of the comparative analysis show that while AI chatbots are capable of producing informative and well-structured content that can aid in broadening public understanding and supporting educational or preliminary planning efforts, their limitations remain evident. In particular, the absence of contextual awareness, nuanced judgment, and domain-specific expertise, as well as the constraint posed by the training data to rely on publicly available sources, excluding much of the peer-reviewed literature behind proprietary databases. These restrictions can lead to oversimplified, partially accurate, or outdated conclusions, particularly in dynamic, interdisciplinary fields (McGovern et al., 2021). Consequently, although generative AI shows promise as a supplementary tool for exploring sustainable, nature-based approaches to coastal management, critical decisions should continue to rely on the expertise and discernment of subject matter professionals.

Additionally, by critically analyzing how generative AI interprets and formulates guidance on coastal management, this research underscores the necessity of understanding the internal logic and data-driven heuristics that underpin AI-generated outputs. As AI becomes more integrated into environmental planning workflows, especially in areas such as climate adaptation and coastal resilience, it is imperative to interrogate how these systems process complex, interdisciplinary knowledge.

The central question posed is whether AI can meaningfully advance coastal management and nature-based infrastructure, or whether it merely reflects the biases embedded in its training data and the assumptions embedded in user prompts. AI's current capabilities suggest it can synthesize information, identify general patterns, and provide accessible entry points into complex topics. However, its outputs are constrained by the scope, quality, and timeliness of the data it has been exposed to, as well as by its lack of contextual awareness and domain-specific judgment. While AI can simulate expertise in language, it does not possess an embodied understanding of ecological systems, or the situational insight required to navigate place-based environmental and social complexities. As these systems continue to evolve, their role in supporting scenario planning, stakeholder communication, and early-stage design processes may expand. Yet, realizing this potential requires a clear understanding of AI's operational boundaries or what it can and cannot reliably do. Only by acknowledging these limitations can AI be responsibly integrated as a tool that augments, rather than replaces, the analytical rigor and critical thinking of domain experts in coastal science and engineering.

2 Materials and methods

This study aims to assess the capabilities of generative artificial intelligence (AI) platforms in addressing topics relevant to coastal management, particularly with respect to natural infrastructure. A mixed qualitative-quantitative approach was used, employing a structured set of questions, standardized scoring criteria, and comparative evaluation across three leading free to use AI platforms. By evaluating AI's capabilities against structured, interdisciplinary questions, this study underscores the potential of these tools to complement the expertise of engineers, ecologists, and coastal managers in advancing nature-based solutions for sustainable coastal management. The following subsections outline the steps taken in question development, AI selection, scoring, and reliability, along with a reflection on potential limitations inherent in AI-generated content.

2.1 Question categories and selection

An initial pool of 52 questions was compiled based on a literature review of contemporary coastal management challenges, policy priorities, and scientific discourse. These questions covered themes such as ecosystem services, infrastructure trade-offs, climate adaptation, community engagement, and technological tools (Edwards, 1997; Gordon et al., 2022; Siders et al., 2019; O'Donnell, 2022; Barbier, 2011; Barbier et al., 2017).

To refine the scope and ensure depth of analysis, questions were evaluated for thematic representation, clarity, and potential to elicit insightful AI responses. Through iterative discussion and pilot testing, a final set of 8 questions was selected. These 8 questions and their responses were thought to be the most representative in each thematic category. Each question was crafted to correspond with one of eight key thematic categories in coastal management:

- General understanding of coastal management -" What do you consider the definition of coastal management?".
- Environmental concerns- "What role do you think natural habitats like mangroves and wetlands play in coastal protection?".
- Ecological concerns - "How effective are current measures to prevent coastal erosion and protect biodiversity?".
- Community and stakeholder involvement - "How do coastal managers engage local communities in decision-making, and what challenges do they face in gaining public support?".
- Long term planning and innovation-" What role do you see for technology and data in improving coastal management practices?".
- Current research - "What coastal management strategies have been most successful, and why?".
- Knowledge gaps - "Are there any critical knowledge gaps that need to be addressed to improve coastal management strategies?".
- Understanding coastal management challenges - "What role do non-structural approaches (like managed retreat or habitat restoration) play compared to traditional engineering solutions in coastal management?".

This categorical framework ensured that chatbot responses could be evaluated for both informational content, and also for relevance to a wide spectrum of coastal management concerns.

2.2 Artificial intelligence engines

This study selected OpenAI's ChatGPT, Google Gemini, and Microsoft Copilot as the testing chatbots because they are free to use, widely available, and highly ranked (Forlini and Circelli, 2025). The analysis focused specifically on their performance in subject areas related to coastal management and natural infrastructure. Each of the final eight questions was submitted to the three models through their standard user interfaces under default settings to simulate typical usage scenarios by environmental professionals or educators seeking rapid knowledge synthesis.

2.2.1 ChatGPT 4o mini

Responses from OpenAI's ChatGPT were generated using the GPT-4 model, accessed via the ChatGPT web interface (chat.openai.com). ChatGPT is a large language model developed by OpenAI, designed to produce human-like text responses based on a wide range of prompts (OpenAI, 2023a). It utilizes deep learning methods, particularly transformer-based neural networks, to understand and generate coherent and contextually relevant text (Vaswani et al., 2017; Brown et al., 2020). The model can assist with various tasks including question answering, summarization, content generation, code creation, and language translation (OpenAI, 2023b).

While GPT-4 is a powerful tool, it is not without limitations; responses may contain inaccuracies, hallucinated content, or reflect biases present in its training data (Bender et al., 2021; McGovern et al., 2021). Furthermore, unless explicitly connected to real-time data sources (e.g., via plugins or API connections), it operates with limited knowledge of events or developments occurring after its training cut-off, originally set to September 2021, later extended for some deployments (OpenAI, 2023a, 2024).

2.2.2 Google Gemini 1.5 Flash

Google's Gemini 1.5 Flash is a lightweight, multimodal AI model optimized for speed and efficiency. It is a distilled version of the larger Gemini 1.5 Pro model, designed for high-throughput, low-latency tasks such as chat interactions, transcription, and document summarization (Google DeepMind, 2024a). Distillation enables Gemini 1.5 Flash to retain core functionalities of the Pro model while significantly reducing computational load, a technique widely used to create smaller, faster models without substantial performance loss (Hinton et al., 2015).

According to Google, the model is particularly effective for tasks involving long-context reasoning, structured data extraction, and on-demand content generation (Google DeepMind, 2024a, 2024). Gemini 1.5 Flash can process up to 1 million tokens in context, enabling advanced applications such as analyzing lengthy legal or technical documents, and generating insights from complex data structures (Google DeepMind, 2024b). Its design prioritizes responsiveness and scalability, making it suitable for user-facing applications like conversational assistants, customer support, and educational tools (Google DeepMind, 2024a).

2.2.3 Microsoft Bing Co-Pilot

Bing Copilot, part of the broader Microsoft Copilot suite, is an AI-enhanced chatbot integrated within Microsoft Edge and other Microsoft services, including Microsoft 365. It is built on OpenAI's large language models, such as GPT-4, and tailored for tasks such as web search, text generation, image creation, and productivity enhancement (Microsoft, 2023a, 2024). The chat functionality within Microsoft Edge allows for up to 30 user-AI exchanges per session, while other browsers such as Chrome and Safari typically support a reduced capacity of around five messages per session (Microsoft, 2023b).

These usage constraints are part of Microsoft's moderation and performance balancing strategies, and additional limitations may apply depending on prompt length and the context of use (Microsoft, 2023b). Bing Copilot is freely available through the Bing platform, but functionality and access vary between the consumer-facing Bing Chat and enterprise-integrated Microsoft 365 Copilot experiences. The enterprise version includes features designed for data privacy, document integration, and productivity within Microsoft Office apps like Word, Excel, and Teams (Microsoft, 2024b; Roth, 2023).

2.3 Scoring rubric and metrics

Responses were evaluated using a structured scoring rubric across five key dimensions:

1. Accuracy – How factually correct is the information?
2. Completeness – Does the response address the full scope of the question?
3. Relevance – Is the information directly related to the topic?
4. Clarity – Is the answer well-articulated and easy to understand?
5. Depth – Does the response show critical thinking, layered insight, or nuanced understanding?

Each dimension was scored on a 1–3–5 scale:

1 = Completely unsatisfactory, incorrect, or irrelevant.

3 = Partially correct, somewhat relevant, or incomplete.

5 = Fully correct, relevant, and insightful.

Example: For the question "What role do you think natural habitats like mangroves and wetlands play in coastal protection?" considering Depth as the scoring Metric, a response scoring 1 might ignore mangroves entirely or confuse them with unrelated concepts. A 3 response may correctly mention erosion control but lack discussion of biodiversity or flood mitigation. A 5 would integrate multiple ecosystem services, cite examples, and include spatial or policy context.

2.4 Limitations and considerations

In this study, all questions were submitted sequentially within the same session for each AI platform, allowing the models to maintain conversational context. This approach was chosen to simulate real-world use patterns, where users often engage in extended multi-question interactions rather than isolated, single-query prompts.

While this method enhances the coherence and cumulative insight of the responses, it introduces certain limitations. AI chatbots may exhibit contextual anchoring, whereby earlier framing influences subsequent answers, either reinforcing prior assumptions or limiting the diversity of perspectives.

Biases also stem from underlying model training data or prior interactions. AI engines learn patterns from large-scale web content, which can reflect dominant narratives, regional priorities, or outdated practices. Moreover, platform-specific memory features or session history (e.g., in personalized Bing searches or Google activity) may subtly shape response patterns. This type of bias is especially relevant as the real-time search integration may favor recent, high-visibility, or Western-centric content, potentially sidelining diverse or Indigenous knowledge systems in coastal management.

These context-driven dynamics were considered when evaluating criteria such as completeness and depth, especially for later questions. This underscores the importance of interpreting AI responses not as isolated outputs, but as emergent products shaped by the flow of interaction and the inherent characteristics of each AI model.

An additional limitation to this analysis is the small and narrowly focused pool of raters. With only three raters, all of whom are coastal engineers with over a decade of experience, the comparative analysis lacks statistical significance. While their expertise is extensive, spanning technical design, research, and program management for coastal resilience and nature-based solutions, this homogeneity limits the generalizability of the findings. The authors acknowledge this limitation, clarifying that the findings from the results are intended to be qualitative in nature. The goal is to spark reflection and discussion among coastal practitioners regarding the application of AI in their work, rather than present a definitive quantitative assessment of the AI chatbots.

2.5 Inter-rater reliability

To ensure consistency and minimize individual bias in the evaluation of AI-generated responses, three independent reviewers assessed each response using a predefined scoring rubric. Inter-rater reliability was calculated using percentage agreement across five evaluation dimensions: accuracy, completeness, relevance, clarity, and depth, as shown in Table 1. Where percent agreement fell below 80%, discrepancies were examined and discussed among reviewers to reconcile differing interpretations. The 80% threshold is commonly used in social science research as a benchmark for acceptable agreement when applying categorical or rubric-based assessments (McHugh, 2012).

This process led to the re-evaluation of some original scores, helping to refine the definitions of each metric and align expectations regarding scoring criteria. However, in these discussions it was found that the variability in responses during the analysis of low-agreement cases was itself a meaningful data point. As such, it was determined that achieving 80% or above

TABLE 1 Percent agreement between raters per rubric category.

% agreement between raters per rubric category				
Accuracy	Completeness	Relevance	Clarity	Depth
ChatGPT				
83.3%	58.3%	25.0%	58.3%	75.0%
Gemini				
66.7%	50.0%	70.8%	54.2%	62.5%
Bing Co-Pilot				
58.3%	50.0%	66.7%	62.5%	50.0%

percent agreement was not a strict requirement for this analysis, and that the observed variability provided valuable insight into the inherent differences in AI model responses and user interpretation (O'Connor and Joffe, 2020).

3 Results

3.1 Inter-rater reliability results

Questions that were ambiguous, compound, or open to multiple interpretations by the AI models often resulted in greater scoring disparity. For instance, the prompt "How effective are current measures to prevent coastal erosion and protect biodiversity?" asks the model to address two distinct yet interconnected topics, coastal erosion and biodiversity protection. Depending on the underlying training data and algorithmic priorities of each AI model, the generated responses varied significantly in how they emphasized one issue over the other. In some cases, recent news articles or dominant discourse trends may have influenced the model to favor one subject, leading to divergent interpretations among reviewers.

These discrepancies underscored the importance of precision and framing in question design when interacting with generative AI systems. In real-world environmental planning contexts, such a question would likely be posed to a multidisciplinary team of experts within the context of a specific management area or project. The resulting response would reflect shared assumptions about goals, scale, and local context, factors that guide human judgment and reduce interpretive ambiguity. To elicit more realistic and context-specific responses from AI models, prompts must similarly embed sufficient detail. For example, revising the original question to: "How effective are current measures in preventing coastal erosion while also protecting biodiversity? Cite examples of both goals being achieved through natural infrastructure programs, such as Engineering With Nature®," would yield more focused responses across models.

Another theme that emerged was a decline in scoring consistency for questions posed later in the sequence. This effect was not attributable to poor prompt design but appeared to result from prompt chaining or latent bias introduced by earlier questions in the session. Some models demonstrated a tendency to carry forward assumptions or tone from previous interactions, which influenced the content and structure of subsequent answers. This phenomenon highlights a subtle but important limitation in multi-turn AI interactions, especially in the absence of session resets or context management and further contributed to reduced inter-rater agreement.

3.2 Generative AI rating results

Overall, the raters found that Bing CoPilot gave the best answers on coastal storm management, followed by Google Gemini and then ChatGPT as shown in Table 2. The raters found that ChatGPT struggled across all rubric categories with particular emphasis on the last three questions related to Current Research, Knowledge Gaps, and Understanding Coastal Management Challenges. There was also the widest inter-rater disagreement within ChatGPT answers, particularly in the metric category of Completeness. The raters found that while ChatGPT gave answers that were accurate and relevant it often left out tangent topics that were needed in fully addressing the question.

Shown in Figure 1, all three chatbots answered the first, second, and fourth question well. These questions were related to general understanding, environmental and ecological concerns, and community stakeholder involvement. CoPilot showed higher scores for questions 5-7, showing greater insights into long term planning and innovation, current research, and knowledge gaps. The last question received the lowest ranking of all questions, with Gemini providing the highest score, outperforming ChatGPT in the metric categories of accuracy and relevance.

ChatGPT consistently ranked the lowest in all five of the scoring metrics as shown in Figure 2, while Gemini and CoPilot gave more similar results. They also produced more identical answers between

TABLE 2 Total score: average and standard deviation of all rubric scores.

Rater	Chat GPT		Gemini		Bing Co-Pilot	
	Average	Standard deviation	Average	Standard deviation	Average	Standard deviation
1	3.7	1	3.8	1	3.6	1
2	3.1	2	3.1	2	3.5	2
3	2.8	2	3.6	1	3.9	1
Total rater average:	3.2		3.5		3.6	

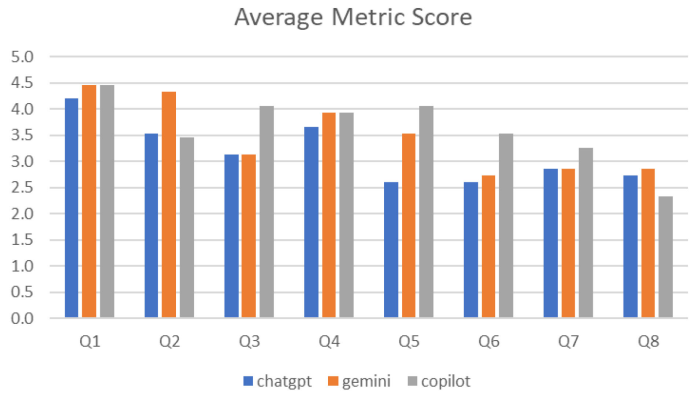

FIGURE 1
Average metric score per question by AI chatbot.

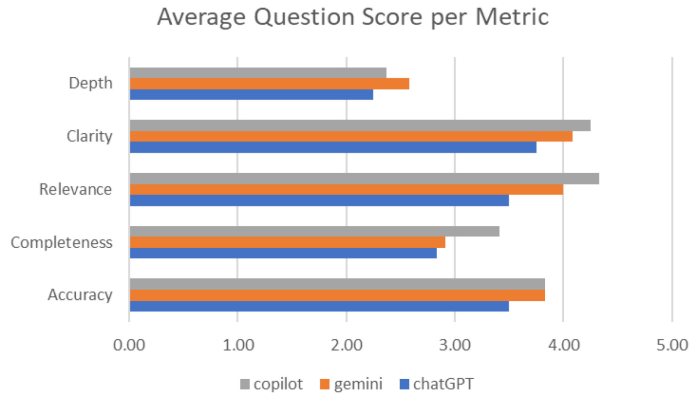

FIGURE 2
Average question score per metric by each AI chatbot.

raters which likely influenced this similarity in performance. Overall, Clarity and Relevance were the best scored metrics, followed by accuracy. Depth and Completeness were the lowest scored metrics.

In rater agreement (Figure 3), ChatGPT and Gemini produced the most variability and therefore lowest agreement in metric scores between each question. Gemini and CoPilot often had similar levels of agreement in answers. The agreement between raters overall chatbots declines significantly with question number.

Per Metric, ChatGPT had the greatest percent agreement between raters on accuracy and depth, followed by Gemini (Figure 4). Gemini had more agreement in rater metric scores on relevance. CoPilot generally had 60% agreement or less across all metrics.

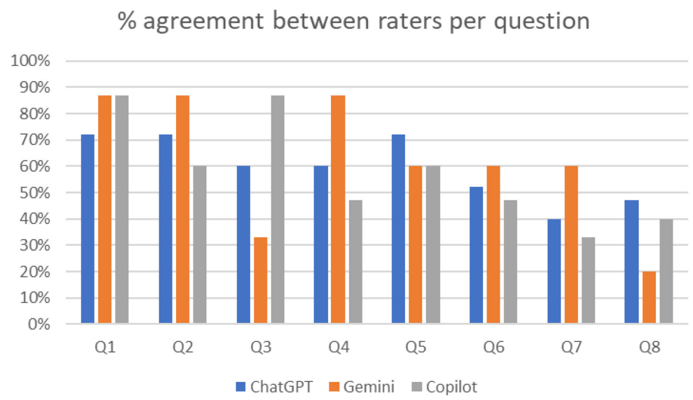

FIGURE 3
Percent agreement between raters per question by each AI chatbot.

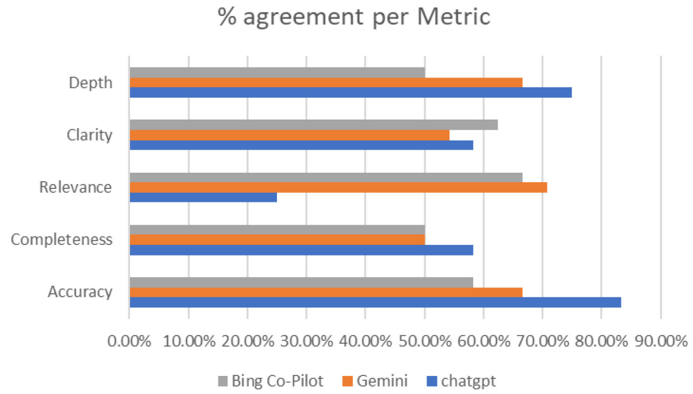

FIGURE 4
Percent agreement per metric by each AI chatbot.

4 Discussion

A key observation emerged from the differences in responses generated for each reviewer: despite using the same prompt structures, model outputs were not identical. Variations reflected subtle shifts in phrasing, prompt history, and user-AI interaction dynamics, highlighting the stochastic nature of large language models and their sensitivity to input framing.

Across the session, responses evolved in tone and content, with later answers frequently reflecting terminology, conceptual frameworks, or biases introduced earlier. This was particularly evident in repeated emphasis on nature-based solutions, an area all models were more comfortable discussing, sometimes at the expense of balanced treatment of grey infrastructure or hybrid approaches. In several instances, prior responses appeared to prime subsequent answers, reinforcing specific narratives or limiting exploration of alternative strategies. For example, once natural infrastructure was discussed positively, later responses often defaulted to emphasizing these solutions without re-evaluating their trade-offs in context.

While this continuity occasionally enhanced coherence and readability, especially in models like Gemini that prioritized structured brevity, it also introduced redundancy and narrowed the scope of insights. Bing Copilot, although the strongest overall, exhibited this bias when referencing similar case studies or recommendations across multiple answers, sometimes offering less diversity in perspectives as the session progressed. ChatGPT, by contrast, tended to generalize its framing early and maintain that vagueness, while Gemini truncated nuanced arguments, possibly due to its lightweight optimization for speed.

Overall, this behavior underscores a structural limitation of generative AI in multi-turn analytical tasks: context retention improves flow but may constrain depth and diversity of content, particularly in domains like coastal management where interdisciplinary and contested perspectives are vital. Recognizing and mitigating this response drift is essential if these tools are to support complex decision-making processes without reinforcing narrow heuristics.

In a recent article in Nature (Naddaf, 2025), author's found similar cases of ChatGPT and Gemini responding sycophantically, with a tendency for "people pleasing" and praising the questioner, to the detriment of depth, completeness and accuracy of the response. The findings of this study, support those claims and also conclude that AI platforms, such as Bing CoPilot and Gemini, can support the exploration of nature-based solutions by synthesizing complex coastal management data, terms and presenting clear, actionable insights. However, their role does not replace professional expertise, which should be used to guide critical decisions in the application of natural infrastructure.

The approach of using generative AI for this comparison highlights several important limitations of generative AI in this context. One critical issue is that these models can only generate responses based on the data they were trained on or have access to through publicly available information. Many high-quality scientific studies, technical reports, and regulatory documents remain behind paywalls or institutional access barriers, which prevents AI models from incorporating them into their responses. As noted by Himmelstein et al. (2018), approximately 28% of scholarly literature remains closed access, limiting broader public and algorithmic reach. This means that even the most accurate AI tools may not be able to reflect or reference the most recent or authoritative research unless it has been made open access or widely cited in public forums.

Furthermore, proprietary databases, such as those managed by Elsevier, Springer, and other commercial publishers, are typically excluded from the training corpora of AI models due to licensing restrictions. As a result, generative AI tools may produce answers based on secondary sources or older, less comprehensive data, which can be problematic for research fields like environmental science where timely, high-resolution data and current methodologies are essential (van Eck et al., 2013).

This comparative evaluation of ChatGPT (GPT-4), Google Gemini 1.5 Flash, and Microsoft Bing Copilot highlights both the promise and constraints of generative AI in the context of coastal management. While all three models demonstrated a baseline understanding of core topics, such as natural infrastructure, climate adaptation, and ecosystem services, their outputs varied significantly in accuracy, completeness, and depth. Notably, Bing Copilot outperformed its counterparts, often delivering responses with greater clarity and relevance, likely benefiting from real-time internet access (Microsoft, 2023a). Gemini followed closely, showing strength in concise structuring, while ChatGPT, despite its fluency, lagged in completeness and frequently produced generalized or vague content.

A shared strength across all platforms was the recognition of natural infrastructure as a vital component in coastal resilience. Concepts such as wetland preservation, dune restoration, and mangrove buffers were consistently cited for their roles in mitigating storm surge and erosion, echoing findings from the IPCC (2023) and research on ecosystem-based adaptation (Temmerman et al., 2013; Sutton-Grier et al., 2015). However, these insights often lacked contextual nuance. AI responses tended to reiterate well-established benefits of nature-based solutions (NBS) without critically engaging with trade-offs, feasibility under different environmental stressors, or the long-term performance of these systems under sea-level rise and socio-political uncertainty (Narayan et al., 2017; Spalding et al., 2014).

In terms of depth and contextualization, none of the models consistently demonstrated a robust understanding of the interdisciplinary nature of coastal planning. The complexity of balancing engineered infrastructure with ecological, social, and regulatory constraints was often oversimplified or omitted. Bing Copilot stood out by referencing recent research and providing greater topical relevance, though even its highest-scoring responses lacked deeper comparative evaluation between grey and green infrastructure strategies (Arkema et al., 2013). Gemini, optimized for efficiency, frequently truncated its answers, while ChatGPT's outputs, although coherent, remained mostly generic and detached from current scientific discourse.

A critical limitation identified was the presence of contextual bias and response drift introduced by sequential prompting. Presenting all eight questions within a single session led the models to reuse terminology and frameworks from earlier interactions, sometimes reinforcing useful continuity but often narrowing the diversity of insights. For instance, once nature-based strategies were introduced positively, subsequent responses tended to default to these solutions even when a more balanced perspective would have been appropriate. This phenomenon aligns with observations in AI discourse, where context retention improves conversational flow but can constrain exploration of complex, multi-perspective topics (Bender et al., 2021; Birhane et al., 2022).

These findings underscore a central tension in using generative AI for applied environmental decision-making: while the models can articulate plausible and well-structured ideas, they lack situational awareness, ecological embodiment, and the ability to interpret socio-political dynamics, elements that are foundational in successful coastal management (Barbier et al., 2008; Foley et al., 2005). This absence creates a risk that confidently worded outputs may be misinterpreted as authoritative, leading to uncritical acceptance and marginalization of expert judgment.

To meaningfully integrate AI into coastal management workflows, it must be framed explicitly as a decision-support tool, not a substitute for expert analysis. AI can augment practitioner capacity during early stages, such as scoping literature, generating conceptual framings, or communicating ideas to stakeholders, but must be complemented by domain expertise to ensure scientific rigor and contextual relevance (Lal et al., 2022; Rolnick et al., 2022).

In Figure 5, a proposed theoretical framework outlines the cyclical and collaborative process for coastal management that would strategically integrate Generative AI chatbots with human expertise.

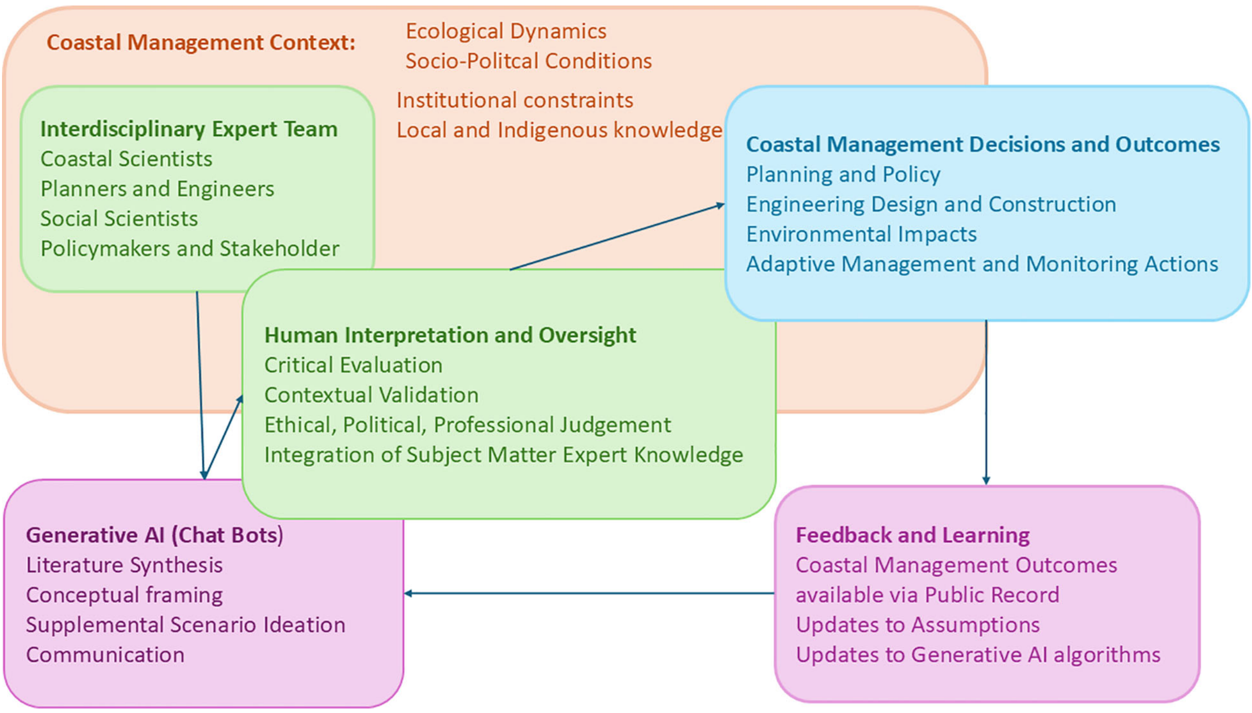

FIGURE 5
Theoretical Framework of how AI can enhance coastal management workflows.

The process begins within a broad Coastal Management Context, which considers ecological dynamics, socio-political conditions, and local knowledge. An Interdisciplinary Expert Team, comprising scientists, engineers, and policymakers, utilizes Generative AI (Chat Bots) to assist with tasks like literature synthesis and supplemental scenario ideation. However, the framework emphasizes that both AI outputs and human insights are filtered through a critical stage of Human Interpretation and Oversight. This ensures that all information is contextually validated and subjected to ethical and professional judgment before leading to Coastal Management Decisions and Outcomes. Finally, a Feedback and Learning loop is acknowledge where the results of these decisions are made public and used to update both the system's assumptions and the AI algorithms, creating an adaptive and continuously improving management cycle. This model demonstrates the belief that AI has a place in coastal management as a supportive tool that can increase work efficiency and augment critical thinking and decision making by human experts.

In conclusion, AI chatbots hold promise as accessible, rapid-response tools for conceptual exploration, public education, and early-stage synthesis in coastal resilience planning. However, their effective use depends on critical interpretation, expert oversight, and a clear understanding of their operational boundaries. As these technologies evolve, future work should prioritize model training on interdisciplinary, site-specific datasets and enhance their capacity to reflect the complexity, trade-offs, and uncertainties inherent in coastal systems.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

SD: Data curation, Project administration, Formal Analysis, Visualization, Validation, Methodology, Resources, Investigation, Writing – review & editing, Conceptualization, Funding acquisition, Supervision, Writing – original draft. LP: Methodology, Writing – original draft, Conceptualization, Investigation. AT: Investigation, Funding acquisition, Writing – review & editing, Data curation, Writing – original draft, Formal Analysis.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This paper is a result of research that was done in collaboration with the Engineering Research and Development Center Coastal Hydraulic Laboratory and funded by the Engineering With Nature® Program.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was used in the creation of this manuscript. Per the nature of this study to compare responses from generative AI chatbots, generative AI chatbots were used in this study to respond to the questions posed.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Author disclaimer

Any use of trade, firm or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

Arkema, K. K., Guannel, G., Verutes, G., Wood, S. A., Guerry, A., Ruckelshaus, M., et al. (2013). Coastal habitats shield people and property from sea-level rise and storms. *Nat. Climate Change* 3, 913–918. doi: 10.1038/nclimate1944

Barbier, E. B. (2011). The value of estuarine and coastal ecosystem services. *Ecol. Monogr.* 81, 169–193. doi: 10.1890/10-1510.1

Barbier, E. B., Hacker, S. D., Kennedy, C., Koch, E. W., Stier, A. C., and Silliman, B. R. (2008). Coastal ecosystem-based management with nonlinear ecological functions and values. *Science* 319, 321–323. doi: 10.1126/science.1150349

Barbier, E. B., Hacker, S. D., Kennedy, C., Koch, E. W., Stier, A. C., and Silliman, B. R. (2017). Progress and challenges in valuing coastal and marine ecosystem services. *Trends Ecol. Evol.* 32, 255–268. doi: 10.1016/j.tree.2016.12.001

Bender, E. M., Gebru, T., McMillan-Major, A., and Shmitchell, S. (2021). "On the dangers of stochastic parrots: Can language models be too big?," in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. (New York, New York: Assocation for Computing Machinery (ACM)), 610–623. doi: 10.1145/3442188.3445922

Birhane, A., Prabhu, V., and Kahembwe, E. (2022). "The values encoded in machine learning research," in *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. (Seoul, South Korea: Assocation for Computing Machinery (ACM)). doi: 10.1145/3531146.3533179

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., et al. (2020). Language models are few-shot learners. *Adv. Neural Inf. Process. Syst.* 33, 1877–1901. Available online at: https://proceedings.neurips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf (Accessed August 2024).

Center for Coastal Solutions. *Engineering with nature for coastal dune and wetland resilience* (Gainesville, FL, USA: Center for Coastal Solutions, University of Florida). Available online at: https://ccs.eng.ufl.edu/what-we-do/nature-based-solutions/engineering-with-nature-for-coastal-dune-and-wetland-resilience/ (Accessed January 2025).

Dalton, A. (2024). AI + Earth observation data could catalyze a world's worth of climate progress. *Wall Street J. CIO J.* Available online at: https://deloitte.wsj.com/cio/ai-earth-observation-data-could-catalyze-a-worlds-worth-of-climate-progress-d173a4cb.

Edwards, S. D. (1997). Participation in coastal zone management initiatives: A review and analysis. *Ocean Coast. Manage.* 36, 157–174. doi: 10.1016/S0964-5691(97)00017-2

Environmental Protection Agency (EPA) *Climate change impacts on coasts*. Available online at: https://www.epa.gov/climateimpacts/climate-change-impacts-coasts (Accessed August 2024).

Financial Times (2025). *How we can use AI to create a better society* (London, UK: Financial Times). Available online at: https://www.ft.com/content/33ed8ad0-f8ad-42ed-983a-54d5b9eb2d27 (Accessed January 2025).

Foley, J. A., DeFries, R., Asner, G. P., Barford, C., Bonan, G., Carpenter, S. R., et al. (2005). Global consequences of land use. *Science* 309, 570–574. doi: 10.1126/science.1111772

Forlini, E., and Circelli, R. (2025). *The best AI chatbots of 2025*. Ed. B. Moore (New York, NY: PC Magazine). Available online at: https://www.pcmag.com/picks/the-best-ai-chatbots?test_uuid=03iF1uOjHbmoZSTXr58OMhT&test_variant=B (Accessed November 2025).

Google DeepMind (2024a). *Gemini 1.5 Flash: Our most efficient long-context model*. Available online at: https://deepmind.google/technologies/gemini/gemini-15-flash/ (Accessed August 2024).

Google DeepMind (2024b). *Gemini 1.5 technical report*. Available online at: https://storage.googleapis.com/deepmind-media/gemini/gemini-1.5-technical-report.pdf (Accessed August 2024).

Gordon, B., Todd, D., and Siders, A. R. (2022). Community involvement in coastal adaptation: Examining participation in shoreline management planning. *Environ. Sci. Policy* 135, 70–80. doi: 10.1016/j.envsci.2022.05.014

Himmelstein, D. S., Romero, A. R., McLaughlin, S. R., Tzovaras, B. G., and Greene, C. S. (2018). Sci-Hub provides access to nearly all scholarly literature. *eLife* 7, e32822. doi: 10.7554/eLife.32822

Hinton, G., Vinyals, O., and Dean, J. (2015). *Distilling the knowledge in a neural network*. Available online at: https://arxiv.org/abs/1503.02531 (Accessed August 2024).

Intergovernmental Panel on Climate Change (IPCC) (2023). *Sixth assessment report: Impacts, adaptation and vulnerability*. Available online at: https://www.ipcc.ch/report/ar6/wg2/ (Accessed August 2024).

Lal, R., Stewart, B. A., and Singh, B. R. (2022). AI applications in sustainable development and climate resilience. *Sustainability* 14, 7103. doi: 10.3390/su14127103

McGovern, A., Ebert-Uphoff, I., Gagne, D. J. II, and Bostrom, A. (2021). *The need for ethical, responsible, and trustworthy artificial intelligence for environmental sciences*. Available online at: https://arxiv.org/abs/2112.08453 (Accessed August 2024).

Microsoft (2023a). *Introducing bing chat: now with GPT-4. Microsoft blogs*. Available online at: https://blogs.bing.com/search/march_2023/Introducing-the-new-Bing (Accessed August 2024).

Microsoft (2023b). *Bing chat FAQ. Microsoft learn*. Available online at: https://learn.microsoft.com/en-us/bing-chat/faq (Accessed August 2024).

Microsoft (2024a). *Microsoft Copilot: Your everyday AI companion*. Available online at: https://copilot.microsoft.com/ (Accessed August 2024).

Microsoft (2024b). *What is microsoft 365 copilot? Microsoft learn*. Available online at: https://learn.microsoft.com/en-us/microsoft-365-copilot/overview (Accessed August 2024).

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica* 22, 276–282. doi: 10.11613/bm.2012.031

Naddaf, M. (2025). ChatGPT and Gemini found to flatter users instead of giving straight answers, study shows. *Nature*. 647, 13–14. Available online at: https://www.nature.com/articles/d41586-025-03390-0 (Accessed October 2025).

Narayan, S., Beck, M. W., Wilson, P., Thomas, C. J., Guerrero, A., Shepard, C. C., et al. (2017). The value of coastal wetlands for flood damage reduction in the northeastern USA. *Sci. Rep.* 7, 9463. doi: 10.1038/s41598-017-09269-z

National Oceanic and Atmospheric Administration (NOAA). *Nature-based solutions* (Silver Spring, MD, USA: NOAA Office for Coastal Management). Available online at: https://coast.noaa.gov/states/fast-facts/natural-infrastructure.html (Accessed January 2025).

O'Connor, C., and Joffe, H. (2020). Intercoder reliability in qualitative research: Debates and practical guidelines. *Int. J. Qual. Methods* 19, 1–13.

O'Donnell, T. (2022). Managed retreat: A systematic review of trade-offs and implementation challenges. *Environ. Res. Lett.* 17, 033004. doi: 10.1088/1748-9326/ac4d7d

OpenAI (2023a). *GPT-4 technical report*. Available online at: https://cdn.openai.com/papers/gpt-4.pdf (Accessed August 2024).

OpenAI (2023b). *Introducing chatgpt*. Available online at: https://openai.com/blog/chatgpt (Accessed August 2024).

OpenAI (2024). *How GPT models work. OpenAI Help Center*. Available online at: https://help.openai.com/en/articles/6825453-chatgpt-release-notes (Accessed August 2024).

Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., et al. (2022). Tackling climate change with machine learning. *ACM Comput. Surv.* 55, 1–96. doi: 10.1145/3485128

Roth, E. (2023). Microsoft's Bing Chat comes to Chrome and Safari. *The Verge*. New York, NY, USA. https://www.theverge.com/2023/7/24/23805493/bing-ai-chat-google-chrome-safari (Accessed January 2025).

Siders, A. R., Hino, M., and Mach, K. J. (2019). The case for strategic and managed retreat. *Science* 365, 761–763. doi: 10.1126/science.aax8346

Spalding, M. D., McIvor, A., Beck, M. W., Koch, E. W., Möller, I., Reed, D. J., et al. (2014). The role of ecosystems in coastal protection: Adapting to climate change and coastal hazards. *Ocean Coast. Manage.* 90, 50–57. doi: 10.1016/j.ocecoaman.2013.09.007

Sutton-Grier, A. E., Wowk, K., and Bamford, H. (2015). Future of our coasts: The potential for natural and hybrid infrastructure to enhance the resilience of our coastal communities, economies and ecosystems. *Environ. Sci. Policy* 51, 137–148. doi: 10.1016/j.envsci.2015.04.006

Temmerman, S., Meire, P., Bouma, T. J., Herman, P. M., Ysebaert, T., and De Vriend, H. J. (2013). Ecosystem-based coastal defence in the face of global change. *Nature* 504, 79–83. doi: 10.1038/nature12859

The Guardian (2025). *Sea level rise will cause 'catastrophic inland migration,' scientists warn* (The Guardian). Available online at: https://www.theguardian.com/environment/2025/may/20/sea-level-rise-migration.

van Eck, N. J., Waltman, L., van Raan, A. F., Klautz, R. J., and Peul, W. C. (2013). Citation analysis may severely underestimate the impact of clinical research as compared to basic research. *PloS One* 8, e62395. doi: 10.1371/journal.pone.0062395

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. *Adv. Neural Inf. Process. Syst.* 30, 5998–6008. Available online at: https://papers.nips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html (Accessed August 2024).

Yale Environment Review *Coastal wetland conservation: Nature-based solutions to mitigate the impacts of climate change*. Available online at: https://environment-review.yale.edu/coastal-wetland-conservation-nature-based-solutions-mitigate-impacts-climate-change (Accessed August 2024).