The department of Data Sciences

Holon Institute of Technology

Final Project - Master of Science

Automated Elicitation and Discovery of Human Health and Ecological Indicators for Urban Freshwater Environments Monitoring

By: Helit Bauberg, 027466002

Supervisor: Prof. Arriel Benis, Department of Digital Medical Technologies and Department of Data Science,   
Holon Institute of Technology

Date of submission: 8 October 2024

# Acknowledgments

This work was funded by the European Union’s Horizon Europe Research and Innovation Program (HORIZONCL6-2022-GOVERNANCE-01) under grant agreement No. 101086521 - OneAquaHealth (Protecting urban aquatic ecosystems to promote One Health).

I would like to extend my gratitude to Prof. Arriel Benis and Holon Institute of Technology (HIT) for providing the opportunity and resources to pursue this research. Their support has contributed to the completion of this work.

Table of Contents

[Acknowledgments 1](#_Toc178889079)

[Abstract 5](#_Toc178889080)

[List of Tables 7](#_Toc178889081)

[OAH-AI Summaries 8](#_Toc178889082)

[Introduction 10](#_Toc178889083)

[Background 10](#_Toc178889084)

[ODH Technology Ring and Data Science 11](#_Toc178889085)

[AI and Its Role in ODH 11](#_Toc178889086)

[Conversational AI and Generative AI 12](#_Toc178889087)

[Scientific Discovery – the AI Era 13](#_Toc178889088)

[Aims, Objectives, and Goals 13](#_Toc178889089)

[Methods 14](#_Toc178889090)

[OAH-AI Framework 14](#_Toc178889091)

[System Architecture 15](#_Toc178889092)

[Part I – OAH-AI PRISMA Systematic Review 15](#_Toc178889093)

[Part II – OAH-AGI 18](#_Toc178889094)

[Development Environment 19](#_Toc178889095)

[Results 20](#_Toc178889096)

[Part I - OAH-AI PRISMA Systematic Review 20](#_Toc178889097)

[Step 1 - Identification 20](#_Toc178889098)

[Step 2 – Screening 20](#_Toc178889099)

[Step 3 – Eligibility 23](#_Toc178889100)

[Step 4 – Full Text Assessment 29](#_Toc178889101)

[Full OAH-AI PRISMA Review - Summary 40](#_Toc178889102)

[Part II – OAH-AGI 41](#_Toc178889103)

[First Process: Generating Partial Reports 41](#_Toc178889104)

[Second Process: Summarizing Reports 41](#_Toc178889105)

[Summary of Results and Observations 41](#_Toc178889106)

[Discussion 47](#_Toc178889107)

[Overview 47](#_Toc178889108)

[Heuristic Scientific Discovery 47](#_Toc178889109)

[Generative AI and Heuristic Discovery 48](#_Toc178889110)

[Heuristic Approach in OAH-AI 48](#_Toc178889111)

[Using LLMs within OAH-AI 49](#_Toc178889112)

[Customizing LLMs’ reasoning and creativity settings 49](#_Toc178889113)

[Context Window and Retrieval-Augmented Generation 49](#_Toc178889114)

[Human-in-the-Loop for Domain-Specificity 50](#_Toc178889115)

[Closed vs. Open-Source LLMs 50](#_Toc178889116)

[Continuous Updates in LLMs 51](#_Toc178889117)

[Experimenting with close-source LLMs 51](#_Toc178889118)

[Prompt Design - Lessons Learned 52](#_Toc178889119)

[OpenAI vs. Gemini: Comparative Evaluation 52](#_Toc178889120)

[Perspectives (Future work) 55](#_Toc178889121)

[Ethics and Sustainability 55](#_Toc178889122)

[Conclusion 56](#_Toc178889123)

[References 58](#_Toc178889124)

[Glossary 64](#_Toc178889125)

[Appendix 1 – OAH-AI PRISMA Screening prompt 65](#_Toc178889126)

[Appendix 2 - LLM Usage 65](#_Toc178889127)

[Appendix 3 - LLM query engine selector example 66](#_Toc178889128)

[Appendix 4 - Project and Output files’ location and description 68](#_Toc178889129)

[Appendix 5 - Gemini reasoning failure 69](#_Toc178889130)

[Appendix 6 – LLM calls – PoC costs 71](#_Toc178889131)

[Appendix 7 – Papers Selected for OAH-AGI 71](#_Toc178889132)

# Abstract

This study explores the integration of artificial intelligence (AI) within the One Aqua Health (OAH) project, a European-Union funded initiative under the One Digital Health (ODH) framework. By merging the principles of One Health with Digital Health, ODH promotes a holistic approach to health that encompasses human, animal, and environmental domains. Within this context, the OAH-AI project aims to advance AI applications for monitoring urban freshwater ecosystems, focusing on identifying early warning indicators of ecosystem degradation.

Central to this research is the development of an AI-powered platform designed to systematically review scientific literature and identify health and ecological indicators relevant to freshwater environments. Using large language models (LLMs) such as OpenAI’s GPT-4o-mini and Google’s Gemini-1.5-flash, the platform processes a corpus of 1,000 scientific papers from PubMed through stages of metadata screening, eligibility assessment, and full-text analysis. This structured process allows for efficient data extraction and discovery, aiming to reveal correlations across human, animal, and environmental health domains.

Results demonstrate that the OAH-AI framework significantly accelerates literature review tasks, providing contextually accurate insights with high efficiency. This capability reduces the time needed for literature analysis and enables the identification of critical environmental health proxies, disease vectors, and trends related to ecosystem health. Evaluated by domain experts, the outputs generated by the LLMs were found to be precise and relevant, further supporting the platform’s value in monitoring ecosystem’s health.

The study also reflects on the rapid advancements in AI technology, emphasizing the need for sustainable and ethical practices to guide its application. As the field progresses, AI's role in interdisciplinary research will likely expand, supporting the ODH framework’s goals of data credibility and integration across complex health ecosystems. This work thus sets a foundation for future AI-driven research aimed at deepening our understanding of interconnected health indicators and supporting the sustainable management of health ecosystems.

List of Figures

[Figure 1- ODH Steering Wheel framework (Benis et al., 2021) 11](#_Toc178936254)

[Figure 2 - number of 'accept' votes for conflicted papers 21](#_Toc178936255)

[Figure 3 - OAH-AI Screened papers distribution according to publication year 22](#_Toc178936256)

[Figure 4 - OAH Score Distribution 27](#_Toc178936257)

[Figure 5 - ODH Score Distribution 27](#_Toc178936258)

[Figure 6 - PoC Costs - Gemini 72](#_Toc178936259)

# List of Tables

[Table 1 - OAH-AI PRISMA steps - description 16](#_Toc179021018)

[Table 2 - OAH-AI PRISMA steps - extracted data 18](#_Toc179021019)

[Table 3 - PRISMA Isdentification - Pubmed Queries 20](#_Toc179021020)

[Table 4 - OpenAI - Gemini comparison - Screening 22](#_Toc179021021)

[Table 5 - Eligibility Prompts, set 1 24](#_Toc179021022)

[Table 6 - Eligibility Prompts, set 2 25](#_Toc179021023)

[Table 7 - sample of eligibility assessment Gemini vs. OpenAI 25](#_Toc179021024)

[Table 8 - summary of eligibility assessment Gemini vs. OpenAI - 51 papers 26](#_Toc179021025)

[Table 9 - Summary of eligibility assessment by Gemini 28](#_Toc179021026)

[Table 10 - Full Text Assessment Prompts 29](#_Toc179021027)

[Table 11 - Full text assessment - Gemini vs. OpenAI examples 38](#_Toc179021028)

[Table 12 - Final Report - LLM Comparison 45](#_Toc179021029)

# OAH-AI Summaries

[OAH-AI Summary - 1: PRISMA Identification 20](#_Toc178936230)

[OAH-AI Summary - 2: PRISMA Screening 23](#_Toc178936231)

[OAH-AI Summary - 3: PRISMA Eligibility Assessment 23](#_Toc178936232)

[OAH-AI Summary - 4: PRISMA Eligibility Assessment - PoC 29](#_Toc178936233)

[OAH-AI Summary - 5: Full PRISMA Assessment 40](#_Toc178936234)

List of Abbreviations

AGI - Artificial General Intelligence

AI - Artificial Intelligence

API - Application Programming Interface

DOI – Digital Object Identifier

DS - Data Science

FAIR - Findable, Accessible, Interoperable, and Reusable

GenAI - Generative Artificial Intelligence

GUI - Graphical User Interface

IoAHT - Internet of Animal and Human Things

IoMT - Internet of Medical Things

IoT - Internet of Things

LLM - Large Language Model

ML - Machine Learning

NLP - Natural Language Processing

OAH – One Aqua Health

ODH - One Digital Health

ODH-UT - One Digital Health Unified Terminology

PoC - Proof of Concept

RAG - Retrieval-Augmented Generation

# Introduction

## Background

One Digital Health (ODH) (Benis et al., 2021) is a conceptual framework for health ecosystems that aims to provide a unified approach to digital health. It merges the specific features of One Health (*One Health*, 2008) and Digital Health (Eysenbach & Group, 2011; Meskó et al., 2017) into one innovative landscape (Benis et al., 2021). ODH takes a holistic approach, addressing human health, animal health, and the environment as dimensions, while integrating various elements of healthcare, environment science, technology, and data as their perspectives.

The framework emphasizes the importance of a proactive approach to building health ecosystems that leverage digital technologies, artificial intelligence (AI), machine learning (ML), and data analytics in healthcare. Adhering to the ODH framework and strategies by utilizing digital and mobile technologies for any initiatives that address human, animal, and environmental systems aims to provide the scientific community with FAIR (Findable, Accessible, Interoperable, and Reusable) ( Wilkinson et al., 2016; Tamburis & Benis, 2022) access to valuable data resources and analytics. Increased data availability is anticipated to contribute significantly to understanding complex ecosystems and enhancing the ability to detect changes in behavior and identify trends and emergencies.

The OneAquaHealth (OAH) (*Home - ONEAQUAHEALTH*, 2022) is such initiative under the ODH framework. OAH is a European-Union-funded project that involves 13 institutions from 9 countries; it aims to improve the sustainability and integrity of freshwater ecosystems in urban environments, to identify early warning indicators and enhance environmental monitoring with AI-assisted tools.

The ODH holistic approach incorporates previously separated domains into a unified framework, asserting that human well-being and ecosystem health are interconnected. By identifying and investigating these links, ODH supports decision-makers in finding **adequate and timely decisions** as well as implementing **effective measures** to restore aquatic ecosystem health and promote One Health principles.

The necessary first steps in achieving OAH goals are the curation of ecological and biological health indicators (as a part of the so-call Work Package 2, WP2) and human health indicators and behaviors (as a part of the so-call Work Package 3, WP3). Under ODH technology ring, this research aims to contribute to these tasks by using AI and Data Science (DS) approaches, from the Large Langue Models (LLMs) field, to support the elicitation of human health and ecological indicators reported in the scientific literature and potential to suggest new ones. (LeCun et al., 2015, Russell & Norvig, 2021)

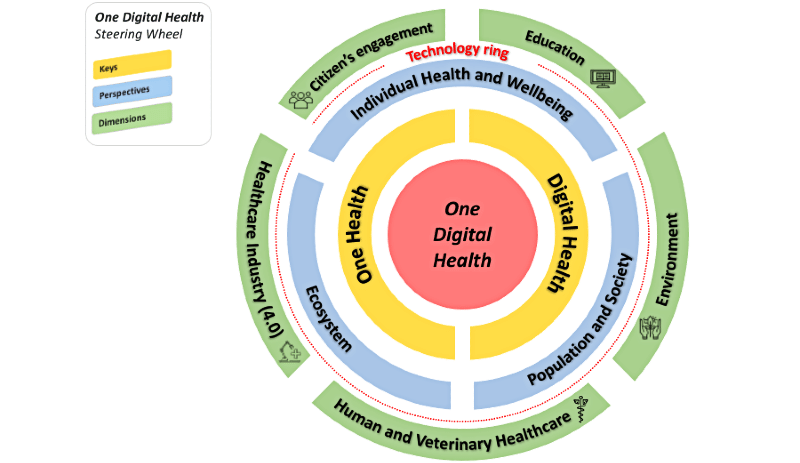


Figure 1- ODH Steering Wheel framework (Benis et al., 2021)

## ODH Technology Ring and Data Science

DS (Fawcett & Provost, 2013) is an interdisciplinary field that draws upon various disciplines, including mathematics, statistics, probability theory, computer science, and AI. Its primary objective is to discover actionable insights and knowledge through the analysis of large and complex datasets. These data sources often include Internet of Things (IoT) sensors, Internet of Medical Things (IoMT), Internet of Animal and Human Things (IoAHT), social media platforms, healthcare records, and an extensive, growing body of literature. The integration of data from such diverse sources provides a comprehensive view, which is crucial for understanding complex ecosystems, especially within the ODH framework. (Farah et al., 2023)

The datasets collected for ODH initiatives encompass health data, ecological data, and environmental data. Examples include healthcare records for monitoring human health, biological indicators for ecosystem assessments, and environmental data from smart sensors for detecting climate and pollution changes. These data, gathered in real time, contribute to a holistic understanding of human, animal, and environmental health interactions, which are central to the ODH framework.

In recent years, AI has gained significant traction in DS applications. AI encompasses a broad range of software tools, methods, and algorithms designed to imitate and even surpass human cognitive abilities. AI can support domain-specific tasks, aid decision-making, and enable general-purpose applications such as chatbots and virtual assistants, which have become integral to daily life, enhancing productivity and providing health monitoring, content generation, and much more.

### AI and Its Role in ODH

In the context of DS and ODH, AI tools and methods facilitate several core tasks:

1. Data Curation and Classification: AI tools are employed to curate, classify, and analyze vast amounts of data from sources like PubMed and ResearchGate. This allows researchers to sort through complex datasets efficiently, identifying key insights that would otherwise take considerable time to uncover manually. (*Home Feed | ResearchGat*e; *PubMed*; Jin et al., n.d.)
2. Pattern Recognition and Data Analytics: AI excels at identifying patterns and correlations across vast datasets. For example, it can analyze genome sequencing data, study disease transmission patterns, or track climate change impacts on ecosystems. Through advanced pattern recognition, AI can extract meaningful trends from the data (Dias & Torkamani, 2019; Dewitte et al., 2021).
3. Predictive Modeling and Simulation: AI is also used in predictive modeling and simulation. In healthcare, AI models can predict the efficacy and safety of drugs in clinical trials, assist in personalized medical treatment planning, and forecast the spread of infectious diseases. In environmental studies, AI can simulate the effects of climate change on specific regions and ecosystems, aiding in policy development and environmental management (Adam et al., 2020; Johnson et al., 2021).

While these AI applications offer powerful tools for DS, the ODH framework highlights several limitations. Processing the data, considering the amount and diversity of data sources, languages and formats, many contain unstructured data (e.g. PDF files, web sites) is a very challenging task, especially from scientific literature in diverse fields. Inferring relationships across the three ODH domains (human, animal, environment) is even more challenging.

## Conversational AI and Generative AI

AI capabilities have evolved dramatically, with conversational AI and Generative AI now receiving significant attention from both the scientific community and the general public. These tools have the potential to revolutionize how we conduct research and process complex data. (AI4Science & Quantum, 2023)

Conversional AI, better known as Large Language Models (LLMs), are advanced AI models that have been trained on massive amounts of text data to understand and generate human language. These models use deep learning techniques to predict and generate text based on the context provided. LLMs are capable of performing a wide range of tasks, including language translation, summarization, and conversation. They excel in natural language understanding and generation, making them valuable tools for applications like virtual assistants, chatbots, and content generation.

Conversational LLMs, such as OpenAI’s ChatGPT (*ChatGPT*, n.d.), Google’s Gemini (*‎Gemini – Chat to Supercharge Your Ideas*, n.d.), and Meta’s Llama (*Llama 3.1*, n.d.), can engage in intelligent, continuous dialogue in natural language, demonstrating a deep understanding of context and providing insightful responses. This marks a significant leap from earlier AI tools, allowing for more dynamic and context-aware interactions.

LLMs rely on advanced Natural Language Processing (NLP) techniques to predict the next word or sentence based on statistical patterns in large datasets. While these models excel at language comprehension and can assist with semantic search—broadening search capabilities to interpret the meaning of prompts rather than just matching keywords—they are not inherently creative.

GenAI models, on the other hand, introduce a new layer of AI functionality by generating novel content and solving complex problems across diverse fields like mathematics, coding, vision, and medicine. These models can generate text, images, music, and even videos by learning from vast datasets and identifying underlying patterns.

Early assessments of GPT-4 have shown its ability to tackle tasks that require higher-order reasoning, creativity, and interdisciplinary knowledge. (Bubeck et al., 2023)

Despite their impressive potential, conversational LLMs and GenAI models have limitations that need further exploration. Several studies have highlighted concerns about bias, data integrity, and the ethical implications of relying on AI-driven systems for decision-making in sensitive domains such as healthcare and environmental policy. These challenges and technical constraints must be addressed to ensure AI tools can fully support the goals of ODH. (Li et al., 2023; Wang et al., 2023; Barrett et al., 2023; Reddy, 2023)

### Scientific Discovery – the AI Era

As AI technologies advance, mounting evidence from the scientific community highlights the significant capabilities of these models in various core tasks of scientific discovery, such as analyzing scientific literature, conceptual and theoretical modeling, and advanced data analysis (Xu et al., 2021; Feng et al., 2022; Neves et al., 2023; Chen et al., 2023). The transformative power of AI in these tasks is reshaping how knowledge is discovered, synthesized, and applied across many disciplines.

In the context of OAH, which seeks to connect three vast and complex domains (human, animal, and environment), AI-driven knowledge discovery appears to be a preferable method and potentially the only practical approach to handle such complexity. Initial experiments in medicine and healthcare, for example, have already demonstrated promising results, providing methods and frameworks that can be further expanded upon (Gao et al., 2024) (Deiner et al., 2024).

## Aims, Objectives, and Goals

This research aims to advance the integration of AI within the OAH project and the ODH technology framework, contributing to the sustainable management of urban freshwater ecosystems.

Specifically, the research aims to demonstrate how AI, and more precisely, NLP-driven models, can support identifying and monitoring ecological and health indicators that connect human, animal, and environmental health.

Our objectives are to:

1. Develop an AI-powered platform to facilitate the elicitation and identification of health and environmental indicators from a large corpus of scientific papers.
2. Leverage NLP and LLMs' capabilities to automate a systematic literature review process that will identify themes and correlations within this corpus, fill knowledge gaps, and reveal previously undetected connections between domains.
3. Validate the platform’s efficacy by applying it to the OAH project, focusing on freshwater ecosystems in urban environments. This will involve generating reports that support the detection of disease vectors, environmental health proxies, and early warning indicators of ecosystem degradation.

Our goals consist of producing the following outcomes:

* A functional AI and NLP-powered platform that can support OAH objectives by processing, analyzing, and generating insights from diverse data sources.
* Improved understanding of the interconnectedness between human health, animal health, and environmental health within urban freshwater ecosystems.
* Practical recommendations for the broader application of AI and NLP tools in other ODH-related projects, focusing on scalability and adaptability.

Integrating AI and Data Science within the OAH project offers a significant opportunity to enhance our understanding of the interconnectedness between human, animal, and environmental health. However, current AI technologies have limitations and risks that must be carefully considered. As AI continues to evolve, its potential to contribute to cross-domain insights remains immense, but development and application must align with ODH's interdisciplinary vision, ensuring data credibility, scalability, and the ethical use of AI in holistic health ecosystems.

This manuscript is structured as follows. After the present introduction, the methods section outlines the AI and NLP tools and approaches used in this research, describing the platform's design and how it was built specifically to support the OAH project. Then, the results section presents the findings from running the full OAH-AI pipeline on selected scientific papers, including interim results and statistical analyses. Afterward, in the discussion section, we elaborate on various aspects we investigated, tackled or identified. It also summarizes the research outcomes, discusses the advantages and limitations of current AI technologies in the scope of OAH, and proposes future directions for expanding the role of AI in scientific research.

# Methods

## OAH-AI Framework

We built an application that uses LLMs and NLP as inference, reasoning, and generative engines to elicit health and environmental indicators and links under OAH to support and extend the project's reach.

The first step in this process is the curation of relevant corpora of papers from which new scientific discoveries can be drawn. The amount of published papers grows exponentially in every scientific domain. (Bornmann et al., 2021)A manual systematic literature review is resource-intensive. It is a group effort, requires domain knowledge, is time-consuming, and the results are often far from exhaustive.

Moreover, LLMs have proven capabilities in inferring, and following instructions provided in natural language, and theoretically, they can systematically perform the same task much faster and on larger amounts of data.

There are a few common, well-defined schemes for a manual systematic literature review, and as the PRISMA guidelines (Page et al., 2021) was adopted by OAH team, it made sense to follow the same scheme in our framework as well.

The framework is a multi-step process that consists of two parts:

* Part I performs a literature review on OAH domain. It adheres to PRISMA systematic review flow diagram: Identifies, Screens, and performs an Eligibility assessment on papers’ full text to extract information and asses its relevance for OAH domains of interest. Only Accepted papers (papers that pass the applicable validity checks at each step) continue to the next step for processing. Excluded papers are counted, and rejection reasons are recorded, as the PRISMA procedure requires. The framework collects more information from each paper as it moves through the review steps.
* Part II experiments with GenAI. It uses the corpus of eligible documents processed in Part I to identify recurring themes, explore correlations, identify disease proxies and health proxies, suggest hypotheses, and generate insights based on the entire corpus.

Throughout the framework, we use natural language to provide the LLMs with instructions and questions related to the reviewed papers. The feedback we get from the LLMs is then used to make decisions about the paper and throughout the different stages of the process.

Part I of the framework uses LLM for reasoning and content inference, whereas part II utilizes the LLM generative capabilities and encourages the use of creativity and development innovative ideas.

We also allow Human intervention in the loop throughout the framework: each step produces file outputs in human-readable formats (txt, CSV, JSON) that can be reviewed and revised before being sent as input to the next step.

## System Architecture

The architecture of OAH-AI framework is detailed in the following sections.

### Part I – OAH-AI PRISMA Systematic Review

OAH-AI PRISMA is comprised of 4 sequential steps, adhering to the PRISMA systematic review flow diagram: Identification, Screening, Eligibility, and full-text assessment. The following table summarizes the inputs and outputs of each step and a short explanation of the role the LLM plays in this step. The output file path is also specified per step for ease of reference.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| # | PRISMA Step | Description | Input | LLM functionality | Output | Output files path |
| 1 | Identification | Identify Pubmed papers that match queries | Query: Urban Aquatic Ecosystems  Query: Urban Freshwater Environments | N/A | Text file containing all papers’ metadata, Medline format | /PRISMA/Identified/ |
| 2 | Screening | Screen identified papers | Identified papers abstract and metadata | Inference and reasoning:  LLM is asked to evaluate input text | Text and csv files containing decision for all identified papers | /PRISMA/Screened/ |
| 3 | Eligibility | Assess paper’s relevance to OAH | Screened papers full text pdf files | Inference and reasoning:  LLM is asked to assess paper relevance to ODH, OAH | CSV file containing ODH, OAH relevance scores;  json file containing all extracted information per paper | /PRISMA/Eligible/ |
| 4 | Full text assessment | Chat to paper, collect information | Eligible papers | LLM query engine selector, LLM Q&A retrieval, LLM inference | Text and Json files containing all extracted information per paper | /PRISMA/Outputs/  OAH-PRISMA/ |

Table - OAH-AI PRISMA steps - description

As paper goes through the steps, at each step more data is collected, as summarized in the following table.

Unless specified otherwise, all fields are strings that contain LLM replies in natural language.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | PRISMA Step | Field | Description | Comments |
| 1 | Identification | PMID | PubMed ID | PubMed Bio Entrez python API paper metadata retrieval (*MEDLINE/PubMed Data Element (Field) Descriptions*, n.d.) |
| Title | Paper title |
| Abstract | Paper abstract |
| MeSH terms | Paper MeSH terms |
| Keywords | Paper keywords |
| Authors | Paper authors |
| Language | Paper language |
| Source | Paper source |
| PMC | Paper PMC ID |
| Publication date | Paper publication date |
| 2 | Screening | I1 | True, if the publication in English | LLM is provided with Step 1 data |
| I2 | True, if the publication date is after January 2013 and before May 2024 |
| I3 | True if the source includes doi, OR PMC information included |
| I4 | True if paper abstract, MeSH and keywords focus on Human Health Indicators in Urban Freshwater Environments |
| I5 | True if paper abstract, MeSH and keywords focus on Ecological Indicators in Urban Freshwater Environments |
| Status | Screened if (I1, I2, I3) is true AND (I4 or I5) is true, or Rejected otherwise. |
| 3 | Eligibility | Domain | Human Health/Animal Health/Environment | OAH sub domain categorization |
| Blue/Green/Other | Watery environment categorization |  |
| ODH Relevance Score | Paper’s relevance score to ODH in 1-10 scale |  |
| ODH Score Rationale | Rationale behind given score explained |  |
| OAH Relevance Score | Paper’s relevance score to OAH in 1-10 scale |  |
| OAH Score Rationale | Rationale behind given score explained |  |
| 4 | Full text assessment | Aim | Paper aim |  |
| OAH indicators | Data about indicators measured in paper |  |
| Geolocation | Paper’s geographical location |  |
| Methodology | Data about methodology and reasearch design |  |
| Key findings | Paper’s key findings |  |
| Population | Research population |  |
| Relations | Data about relationship between the indicators discussed in the paper |  |
| ODH keys | List of ODH keys addressed in the paper |  |
| ODH perspectives | List of ODH perspectives addressed in the paper |  |
|  | ODH dimensions | List of ODH dimensions addressed in the paper |  |

Table - OAH-AI PRISMA steps - extracted data

On the final stage of the PRISMA review, all the data that was curated per paper is compiled into a structured summary and saved in a text file format, to be assessed by the domain experts.

### Part II – OAH-AGI

The second part of the OAH-AI framework diverges significantly from the OAH-AI PRISMA (part I of the framework) in its focus and objectives. Rather than analyzing the original papers themselves, this phase centers on the corpus of extracted information derived from earlier stages. Our primary goal shifts from gathering insights from individual papers to exploring interrelationships and patterns across the entire corpus. Additionally, beyond the assessment of logical reasoning and inferencing abilities from the first phase, this phase emphasizes the evaluation of the generative capabilities of the LLMs. Specifically, it assesses their ability to generate creative and non-trivial insights that contribute meaningfully to the OAH project.

This step, by design, is non-reproducible, as each iteration of the process can yield different outputs. As such, it is not strictly aligned with traditional systematic review methods, such as the PRISMA framework.

#### OAH-AGI Pipeline: Two Processes

The OAH-AGI pipeline is composed of two key processes, with the LLMs playing a central role in both. Below, we outline the methodology behind each process:

##### Process 1: Generating Partial Reports

In this process, a subset of papers is randomly selected from the corpus and provided to the LLM with specific instructions. The LLM is tasked with identifying patterns, correlations, and generating hypotheses based on the content of the selected papers. The instructions provided to the LLM are as follows:

*"You are a biomed scientist and domain expert in human health and wellbeing, working with OneAquaHealth. Below is a collection of scientific papers related to OneAquaHealth. Within your field of expertise, you are tasked with:*

1. *Identifying* ***patterns, correlations, and recurring themes*** *between indicators, research, and concepts in the papers.*
2. *Identifying indicators that support the detection of* ***disease vectors*** *or environmental health proxies.*
3. *Identifying* ***early warning indicators*** *for ecosystem degradation.*
4. *Generating* ***potential hypotheses*** *based on the text.*

*Always provide references from the text by stating the paper PMIDs. Answer the tasks in the given order.*  
*{papers}*"

This process is repeated iteratively as many times as needed, with each iteration generating unique outputs.

##### Process 2: Consolidating Partial Reports into a Comprehensive Summary

This step involves consolidating the partial reports generated in Process 1. The LLM receives a subset of the generated "Ideas" and is prompted to merge the results into a cohesive summary. The prompt provided is as follows:

*"Evaluating our scientific papers corpus, our domain expert identified:*

1. *Patterns, correlations, and recurring themes.*
2. *Disease vectors or environmental health proxies.*
3. *Early warning indicators for ecosystem degradation.*
4. *Potential hypotheses generated based on that corpus.*

*We repeated this process several times. Below is the result of these iterations.* *{hypotheses}*

*Please consolidate this information by removing duplicates and merging similar insights into concise statements. Always provide references from the text by stating all supporting paper PMIDs."*

This consolidation process can be repeated multiple times on any subset of ideas or the entire set of partial reports.

##### Encouraging Creativity in the OAH-AI Framework

To foster creativity and ensure the generation of innovative insights, we employed several strategies during this phase:

1. Random Selection of Papers: Random subsets of the corpus were used to promote diversity in the LLM’s outputs.
2. High LLM Temperature Settings: By increasing the temperature in the LLM configuration, we encouraged more varied and creative responses (further discussed in the *Discussion* section).
3. Generative Prompts: The LLM was provided with prompts designed to stimulate creative, generative responses.

## Development Environment

The framework was written in python and runs in Google Colab environment.

Input/output files are read from Google Drive.

Full text paper pdf files are manually downloaded to drive from provided paper source.

Python packages and versions can be found under project files /Environment.

Environment specification details can be found under project files /Environment.

LLM versions used:

Gemini: gemini-1.5-flash

ChatGPT: gpt-4o-mini

# Results

The results of the first step of the OAH-AI PRISMA evaluation process, papers *Identification*, yielded over 12,000 papers. Taking into consideration the hardware, space and resource limitations that arise from using a free cloud environment, as well as the cost of each LLM call, we have decided to limit the scope of the experiments by completing the OAH-AI PRISMA review only on a subset of the *Identified* papers. We randomly picked 1,000 *Identified* papers (about 8%) that were forwarded for the next OAH-AI PRISMA steps, to be processed according to the PRISMA flow diagram. In the following sections we present the results of the PoC.

## Part I - OAH-AI PRISMA Systematic Review

### Step 1 - Identification

The first step in the process, paper identification through multiple PubMed queries as defined by our domain expert, was performed on August 12th 2024, and yielded the following results:

|  |  |  |
| --- | --- | --- |
| Query | Retrieved paper count | Retrieved paper percentage |
| Urban Freshwater Environments | 8776 | 71.37 |
| Human Health Indicators AND Urban Aquatic Ecosystems | 55 | 0.45 |
| Ecological Health Indicators AND Urban Aquatic Ecosystems | 25 | 0.20 |
| Human Health Indicators in Urban Aquatic Ecosystems | 55 | 0.45 |
| Ecology AND Urban Aquatic Ecosystems | 689 | 5.60 |
| Human Health AND Urban Aquatic Ecosystems | 326 | 2.65 |
| Urban Aquatic Ecosystems | 2370 | 19.27 |
| Total: | 12,296 | 100 |

Table - PRISMA Isdentification - Pubmed Queries

OAH-AI Summary - : PRISMA Identification

|  |
| --- |
| PRISMA IDENTIFICATION:  DATABASE: PUBMED  Total papers identified: 12,296  Total unique papers identified: 10,057 (2,239 duplicates removed) |

### Step 2 – Screening

In the PRISMA workflow, the screening stage involves multiple reviewers (at least two) that independently assess each *Identified* paper, and cast their inclusion/exclusion vote.

Discrepancies on inclusion/exclusion votes between reviewers are usually resolved through discussion or by involving a third reviewer.

In this human-based process, this is done to ensure objectivity and reduce any individual biases, thus improving the reliability of the review process.

In OAH-AI PRISMA review, we have complied with this requirement by asking two different LLMs to screen each paper. Though AI is less prone to human biases, it is not fully reliable: there may be other biases, reasoning failures, lack of domain knowledge or other known LLM issues that may result in a vote a domain expert won’t agree with. (*A Comprehensive Comparative Analysis of LLMs*, 2024)

#### Testing LLM consistency

Initially, we wanted to assess both LLMs’ reasoning and inference capabilities, as well as the consistency of the responses, when asked the same questions. We randomly selected 100 (10%) *Identified* papers and *Screened* them using both LLMs. We repeated this experiment 3 times per LLMs (6 runs in total). LLM screening prompt is in Appendix 1.

Out of these 100 papers, over 70% had the same vote in all 6 runs:

* 46 papers: screened unanimously
* 25 papers: rejected unanimously
* 29 papers: no unanimous vote

The following plot describes the distribution of votes between runs/LLMs, when no unanimous vote was casted:

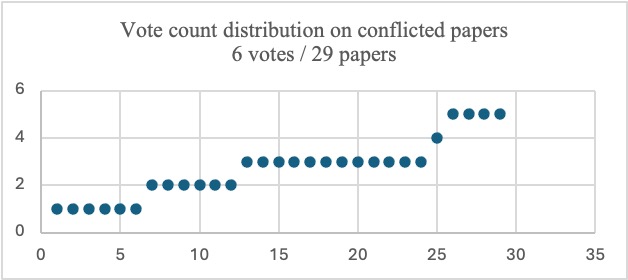


Figure - number of 'accept' votes for conflicted papers

When we consider the symmetry between 1 vote – 5 votes, and the symmetry between 2 votes – 4 votes, the results meant that there is quite a unified distribution of votes between the conflicting views. That meant there are no obvious biases, or a particular inference or reasoning problem. The conflicts were related to Screening fields I4 and I5, where the LLMs are requested to infer the paper domain of focus from its abstract, MeSH and keywords. There were no conflicts related to the other fields, that require straight forward inference (I1 – I3) and logical reasoning (Status field).

For this Proof-of-Concept, we used the following Screening scheme:

* If both LLMs Accept paper: *Screen* paper.
* If both LLMs Reject paper: *Reject* paper.
* If there’s a conflict between LLMs votes: *Reject* paper.

The following table summarizes and compares the results of the OAH-AI PRISMA *Screening* step performed on the corpus of 1,000 *Identified* papers, as processed by Gemini and OpenAI:

|  |  |  |  |
| --- | --- | --- | --- |
| Papers statistics | OpenAI | Gemini | LLMs comparison |
| Screened | 395/1000 | 566/1000 | 307/1000 |
| Rejected | 605/1000 | 434/1000 | 347/1000 |
| Excluded - language | 21/1000 | 25/1000 | 4 conflicts. \* |
| Excluded – public. date | 54/1000 | 42/1000 | 16 conflicts \*\* |
| Excluded – no full text | 6/1000 | 16/1000 | 4 conflicts \* |
| Excluded – paper focus | 598/1000 | 415/1000 | 220 conflicts |
| Total Execution Time | 16 min 45 sec | 10 min 40 sec |  |
|  |  |  |  |

Table - OpenAI - Gemini comparison - Screening

Statistics:

* LLM calls: 1 per LLM per paper (2,000 calls per experiment)
* Papers total token count: 538056.
* Average token count per paper: 538.056

Observations:

65% had unanimous votes.

\* 4 Gemini calls had failed, returning default answer.

\*\* Gemini included 11 papers published in May 2024. OpenAI excluded these. The other 5 conflicts Gemini was wrong.

In the current research we proceeded with the 307 papers that were *Screened* by both LLMs.

For ease of reference, we have exported the results to a Zotero library file, which can be found in the project files, as well as the OAH-AI framework output files and statistical analyses data. Usage information for both LLMs for this step is in Appendix 2.

The following diagram describes the distribution of the 307 *Screened* papers according to the publication year:

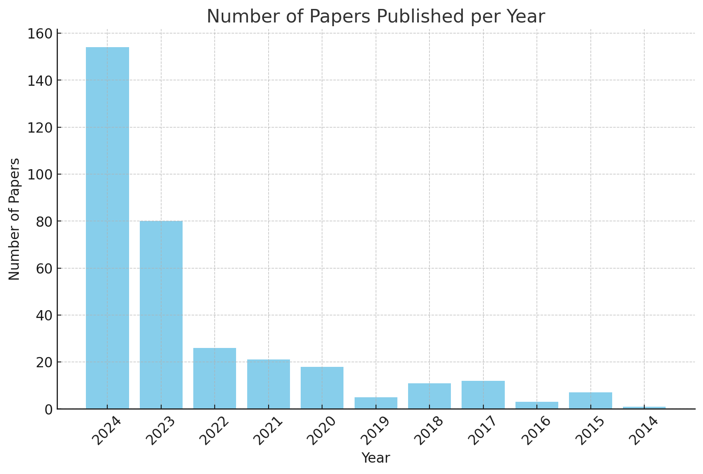


Figure - OAH-AI Screened papers distribution according to publication year

OAH-AI Summary - : PRISMA Screening

|  |
| --- |
| PRISMA SCREENING:  Total papers identified: 1,000  Total papers screened: 307  Total papers rejected: 693  No full text: 16  Paper focus out of OAH scope: 235  Publication date: 42  Language: 26  Total papers available for Eligibility assessment: 307 |

### Step 3 – Eligibility

PRISMA eligibility step determines which *Screened* papers should be included or excluded from the systematic review based on a set of criteria. These criteria are determined by factors such as relevance to the research question, study design, population, interventions, and outcomes. The final review process will not analyze papers that fail to meet these criteria.

In the previous steps, OAH-AI framework relied on papers’ metadata analyses to determine its inclusion status – the data was enough. But a reliable eligibility assessment of the *Screened* papers must include full-text assessment.

We manually tried to download 307 *Screened* papers available from the HIT online library. Out of the 307 *Screened* papers, 69 papers were not downloaded due to no HIT access (only paid access). We are left with 238 papers. All PDFs were uploaded to the Google Drive designated folder, each paper filed under its PMID number.

For ease of reference, we have exported the results to a Zotero library file, which can be found in the project files, as well as the OAH-AI framework output files and statistical analysis data.

OAH-AI Summary - : PRISMA Eligibility Assessment

|  |
| --- |
| PRISMA ELIGIBILITY: |
| Total papers identified: 307 |
| Total papers screened for Eligibility: 238 |
|  |

For each paper:

1. The PDF file is split and converted into text embeddings vectors.
2. Two paper query engine tools are created:

* Vector query engine – a semantic similarity query engine; useful for retrieving specific context from the paper.
* Summarization query engine – an LLM driven engine; useful for summarization questions related to paper.

1. The two query engines are provided to an LLM agent as tools.
2. A fixed list of questions in natural language is provided to the LLM agent.
   * For each question, the LLM uses its reasoning capability to select the right tool for the task. Example of LLM query engine reasoning in Appendix 3.
3. The replies to all the questions are used to determine the paper’s eligibility as well as its relevance to OAH and ODH domains.

#### Eligibity Assessment – Question Set

The following are the questions we ask the paper.

The first set is used to categorize the paper to its matching OAH domain:

|  |
| --- |
| 1. Does this paper focus on impact on human health or well-being (e.g., quality of life, mental health)? Reply yes or no 2. Does this paper focus on impact on environmental health (ecology, animals, plants)? Reply yes or no 3. Does this paper focus on impact on species other than human? Reply yes or no 4. *Blue spaces are dominated by a watery element, such as a lakeside, river or coast.* 5. *Green spaces may include a watery element but are characterized by predominantly 'green' elements such as grass or trees.* Does this paper focus on blue spaces or green spaces? Reply green/blue/other |

Table - Eligibility Prompts, set 1

The second set is used to evaluate the paper’s relevance to our research topics:

|  |
| --- |
| 1. *One Digital Health (ODH) is a proposed unified structure. The conceptual framework of the One Digital Health Steering Wheel is built around two keys (ie, One Health and digital health), three perspectives (ie, individual health and well-being, population and society, and ecosystem), and five dimensions (ie, citizens' engagement, education, environment, human and veterinary health care, and Healthcare Industry 4.0).*   *One Digital Health aims to digitally transform future health ecosystems, by implementing a systemic health and life sciences approach that takes into account broad digital technology perspectives on human health, animal health, and the management of the surrounding environment.*  In a scale of 1 to 10, how relevant is this paper to OneDigitalHealth? Explain your reply and end it with final score in this format: <<score>>   1. *Urban aquatic ecosystems are extremely relevant connectors between people, animals, and plants, making cities more biodiverse and sustainable. Yet, these ecosystems are often confronted with lack of space, vegetation cuts, artificialization, and other urbanisation processes. This degradation can lead to numerous disservices to humans regarding emerging pathogens, decreasing disease resistance, climate change impacts and other health concerns in cities.*   *OneAquaHealth is a research project. OneAquaHealth aims to improve the sustainability and integrity of freshwater ecosystems in urban environments.*  *The project will identify early warning indicators and enhance environmental monitoring by investigating the interconnection of ecosystem health and human wellbeing.*  In a scale of 1 to 10, how relevant is this paper to OneAquaHealth? Explain your reply and end it with final score in this format: <<score>> |

Table - Eligibility Prompts, set 2

Only papers that are categorized under an OAH domain, scoring 8 or higher at OAH relevance score, were directed to the final PRISMA step for in depth assessment.

If an LLM cannot categorize the paper, it tags it with ‘TBD’ – to be determined by domain expert, human or LLM.

The LLM outputs for this step are under project files /Eligibility

Below is a comparison of eligibility assessment results by both LLMs for a few papers:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Gemini** | | | | | **OpenAI** | | | | |
| # | domain | OAH score | ODH score | decision | Time(sec) | domain | OAH score | ODH score | decision | time (sec) |
| 1 | N/A | 3 | 2 | Exclude | 25.20 | Environment | 9 | 3 | Eligible | 100.22 |
| 2 | Environment | 7 | 2 | Eligible | 61.28 | TBD | 9 | 5 | Eligible | 93.54 |
| 3 | Environment | 9 | 2 | Eligible | 66.25 | Human Health | 9 | 3 | Eligible | 107.88 |
| 4 | Environment | 9 | 3 | Eligible | 58.82 | TBD | 8 | 3 | Eligible | 85.29 |
| 5 | Environment | 9 | 5 | Eligible | 73.67 | TBD | 9 | 7 | Eligible | 114.25 |

Table - sample of eligibility assessment Gemini vs. OpenAI

Evaluated papers:

q

1. PMID: 37543067

Title: An analytical approach to confidence interval estimation of river microplastic sampling.

Source: Environ Pollut. 2023 Oct 15;335:122310. doi: 10.1016/j.envpol.2023.122310. Epub 2023 Aug 4.

1. PMID: 37778549  
   Title: Systematic loss in biotic heterogeneity but not biodiversity across multiple trophic levels in Erhai lake, China  
   Source: Sci Total Environ. 2024 Jan 1;906:167479. doi: 10.1016/j.scitotenv.2023.167479. Epub 2023 Sep 29.
2. PMID: 38223759  
   Title: Investigating Escherichia coli habitat transition from sediments to water in tropical urban lakes.  
   Source: PeerJ. 2024 Jan 11;12:e16556. doi: 10.7717/peerj.16556. eCollection 2024.
3. PMID: 38325783  
   Title: Geographic distribution of bacterial communities of inland waters in China.  
   Source: Environ Res. 2024 May 15;249:118337. doi: 10.1016/j.envres.2024.118337. Epub 2024 Feb 5.
4. PMID: 38373626  
   Title: Presence of pharmaceutical contaminants of emerging concerns in two rivers of western Cuba and their relationship with the extracellular enzymatic activity of microbial communities.  
   Source: Environ Pollut. 2024 Apr 1;346:123589. doi: 10.1016/j.envpol.2024.123589. Epub 2024 Feb 17.

#### Performance Evaluation

In order to evaluate as well as compare Gemini and OpenAI’s inference and reasoning skills, we initially tested this step on a smaller subset of papers selected from the entire *Eligible* corpus, alternating between LLMs. The inputs, statistics and LLM outputs for this step are under project files /Eligibility.

The results are summarized in the following table:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Gemini | | OpenAI | |
| Metric\LLM | Run 1 | Run 2 | Run 1 | Run 2 |
| Screened for Full Text | 49 papers *Eligible*.  2 *Excluded* | 49 papers *Eligible*.  2 *Excluded* | 51/51 papers *Eligible* | 51/51 papers *Eligible* |
| Computation time – total (min) | 29 | 33 | 44 | 41 |
| Average processing time per paper (sec) | 34 | | 60 | |
| Average LLM calls per paper | 15 | | | |
| Paper count-  Domain categorization:  TBD | 4 | 3 | 28 | 28 |
| Paper count-  Domain categorization:  Environment | 43 | 43 | 22 | 22 |
| Paper count-  Domain categorization:  Human Health | 2 | 2 | 1 | 1 |
| Paper count-  Domain categorization:  Animal Health | 0 | 1 | 0 | 0 |

Table - summary of eligibility assessment Gemini vs. OpenAI - 51 papers

The following figures illustrate the distribution of OAH and ODH scores, as voted by both LLMs:



Figure - OAH Score Distribution



Figure - ODH Score Distribution

As we can notice, there is strong correlation between the LLM relevance scores, especially for OAH scores. Yet, Gemini exhibits significantly shorter response times, as well as more balanced score distributions.

For this PoC, considering the cost and resource limitations, we used only Gemini perform the *Eligibility* step on the *Screened* corpus.

The results are summarized in the following table:

|  |  |
| --- | --- |
| Metric |  |
| Total papers screened | 238 |
| Screened for *Full Text* | 229 |
| Rejected for *Full Text* | 9 |
| Computation time - total | 1h 17m |
| Average processing time per paper | 33 sec |
| Paper count-domain categorization  Environment | 126 |
| Paper count-domain categorization  Human Health | 7 |
| Paper count-domain categorization  Animal Health | 1 |
| Paper count-domain categorization  TBD | 95 |
| OAH scores | 210 papers scored 8 and above |
| 28 papers scored 7 or less |
| ODH scores | 12 papers scored 7 and above |
| 226 papers scored 6 or less |

Table - Summary of eligibility assessment by Gemini

Ideally, OAH domain experts would review the ‘TBD’ tagged papers at this stage and include or exclude it from the corpus. For the PoC, we included in the corpus for *Full Text Assessment* only the following papers:

1. Papers with OAH score equals 8 or higher
2. Papers with domain categorization other than TBD

That left us with 123 *Eligible* papers.

OAH-AI Summary - : PRISMA Eligibility Assessment - PoC

|  |
| --- |
| PRISMA ELIGIBILITY:  Total papers identified: 307  Total papers screened for Eligibility: 238  Papers rejected: 9 (reason: domain N/A)  Papers rejected: 95 (reason: domain TBD)  Total papers eligible to full text assessment: 123 |

### 

### Step 4 – Full Text Assessment

Once a paper was tagged as *Eligible* for the OAH domain of interest, it was prompted to reply to a second set of questions, using the same LLM agent and tools described at the previous step.

The questions, much like the questions a scientist would ask when reading the paper, are designed to extract the main scientific discoveries from the paper, and reliably present these findings in a clear, concise and comprehensive manner.

#### Full Text Assessment – Question Set

The list of the questions we ask each paper is described in the following table:

|  |
| --- |
| 1. What is the **aim** of the paper? Reply in plain text 2. *Ecological and Human Health Indicators: Key characteristics include sensitivity, reliability, relevance, practicality, and validity.*   *Ecological indicators measure ecosystem health and the impact of human activities, while human health indicators assess population well-being and public health effectiveness.*  *Examples for Indicators:*  *Ecological: Biological (e.g., species diversity), Physical (e.g., water quality), Chemical (e.g., pollutant levels).*  *Human Health: Mortality rates, life expectancy, disease prevalence, and quality of life measures.*  What **indicators** are mentioned in the paper? How were they measured? Reply in plain text.   1. What is the **geographical** location the paper focuses on? Include continent in your reply. Reply in plain text 2. What **methodology**/design was used in the paper? Reply in plain text 3. What are the **key findings** in the paper? Reply in plain text 4. What is the studied **population's** characteristics? number, species, age etc. Reply in plain text 5. Express the key findings of the paper in terms of the **relationship** between the indicators that are discussed in the paper. Reply in plain text 6. *One Digital Health (ODH) is a proposed unified structure. The conceptual framework of the One Digital Health Steering Wheel is built around two keys (ie, One Health and digital health), three perspectives (ie, individual health and well-being, population and society, and ecosystem), and five dimensions (ie, citizens' engagement, education, environment, human and veterinary health care, and Healthcare Industry 4.0).*   *One Digital Health aims to digitally transform future health ecosystems, by implementing a systemic health and life sciences approach that takes into account broad digital technology perspectives on human health, animal health, and the management of the surrounding environment.*   * What **OneDigitalHealth** **keys** are addressed in this paper? Place the keys between double stars in your reply. * What **OneDigitalHealth** **perspectives** are addressed in this paper? Place the perspectives between double stars in your reply * What **OneDigitalHealth** **dimensions** are addressed in this paper? Place the dimensions between double stars in your reply |

Table - Full Text Assessment Prompts

#### Output Comparison

Two representative examples of the outputs by both LLM agents, given the same question set, are detailed in the table below:

|  |  |  |
| --- | --- | --- |
| **PMID**: 38181685  **Title**: Protecting and restoring freshwater biodiversity across urban areas in Aotearoa New Zealand: Citizens' reporting of pollution in stormwater drains and waterways.  **Source**: J Environ Manage. 2024 Feb;351:120019. doi: 10.1016/j.jenvman.2024.120019. Epub 2024 Jan 4.  **Authors**: McLeod LJ, Hine DW, Milfont TL, Dorner Z, Tassell-Matamua NA, Maris RDC, Kitson JC, Stahlmann-Brown P  **Publication Date**: 2024 Feb  **Keywords**: Audience segmentation, Behavior change, Behaviour change wheel, Behaviour prioritisation matrix, Community-based social marketing, Intervention design  **MeSH Terms**: Humans, New Zealand, \*Environmental Pollution, \*Biodiversity, Fresh Water, Ecosystem | | |
|  | **Gemini** | **OpenAI** |
| Output Token count | 841 | 1263 |
| Processing time (secs) | 75 | 102 |
| ODH Score | The paper focuses on freshwater biodiversity and how human behavior impacts it. While it mentions the importance of community engagement and environmental stewardship, it doesn't directly address the broader digital health ecosystem or the interconnectedness of human, animal, and environmental health as envisioned by One Digital Health.  <<3>> | (Screened by Gemini in Step 3) |
| OAH Score | This paper directly addresses the challenges of urbanisation on freshwater ecosystems and the need for intervention to protect and restore biodiversity. It focuses on citizen reporting of pollution as a key behavior to improve urban freshwater health. This aligns with OneAquaHealth's goal of improving the sustainability and integrity of freshwater ecosystems in urban environments by identifying early warning indicators and enhancing environmental monitoring.  The paper's focus on citizen engagement and behavior change strategies is particularly relevant to OneAquaHealth's aim of investigating the interconnection of ecosystem health and human wellbeing.  <<9>> | (Screened by Gemini in Step 3) |
| Paper aim | The paper aims to understand and change human behavior to improve urban freshwater biodiversity and management. | The aim of the paper is to demonstrate the benefits of applying a systematic behavior change framework to improve urban freshwater biodiversity. It evaluates and prioritizes specific behaviors that can enhance freshwater outcomes, assesses community willingness to report pollution, identifies barriers to participation in restoration efforts, and explores the determinants of voluntary reporting behavior. Additionally, it emphasizes the importance of segmentation in designing social marketing programs to promote pro-environmental behaviors and effective conservation strategies. |
| OAH Indicators | The paper focuses on freshwater biodiversity in urban areas and uses indicators related to human behavior and its impact on the environment.  The paper mentions indicators like:  \* \*\*Effectiveness of behaviors:\*\* This was measured through expert ratings on a scale of 1 to 10, with 1 being not at all beneficial and 10 being extremely beneficial.  \* \*\*Maximum Possible Penetration:\*\* This refers to the likelihood of a behavior being adopted by the community, with a value of 4 indicating high likelihood.  \* \*\*Current Penetration:\*\* This measures the current adoption rate of a behavior.  \* \*\*Likelihood of Adoption:\*\* This assesses the probability of a behavior being adopted in the future.  \* \*\*Total Weighted Impact (TWI):\*\* This combines the effectiveness, maximum possible penetration, current penetration, and likelihood of adoption to determine the overall impact of a behavior.  These indicators are used to prioritize behaviors that have the greatest potential to improve freshwater biodiversity in urban areas. | The paper mentions several indicators related to behaviors that can improve freshwater outcomes. These include:  1. \*\*Ecological Indicators\*\*:  - \*\*Biological\*\*: Species diversity.  - \*\*Physical\*\*: Water quality.  - \*\*Chemical\*\*: Pollutant levels.  2. \*\*Human Health Indicators\*\*:  - While specific human health indicators are not detailed in the provided context, the focus is on behaviors that can impact both ecological health and community well-being.  The measurement of these indicators is implied through the assessment of behaviors rated by experts for their effectiveness in improving freshwater outcomes, as well as through community surveys that gauge the likelihood of adoption of these behaviors. The effectiveness of each behavior was rated on a scale from 1 to 10, and the potential impact was calculated using a Total Weighted Impact (TWI) algorithm that combined effectiveness, maximum possible penetration, current penetration, and likelihood of adoption. |
| Geolocation | The paper focuses on Aotearoa New Zealand, which is located in Oceania. | The paper focuses on Aotearoa New Zealand, which is located in the continent of Oceania. |
| Methodology | The paper employed a mixed-methods approach, combining quantitative and qualitative data. It utilized a systematic behavior change framework and the Behavior Change Wheel methodology. The study involved two phases, with the first phase using expert and community surveys to assess the effectiveness of behaviors. The second phase involved a large-scale survey to examine factors influencing residents' willingness to report pollution. The paper also used statistical techniques like factor analysis, latent profile analysis, and MANOVA to analyze the data. | The paper utilized a behavior change framework to analyze and recommend interventions for encouraging residents to report stormwater pollution. It involved gathering data from primary sources and incorporating expert knowledge into a behavior prioritization matrix (BPM) analysis. The methodology focused on understanding the motivations and barriers faced by different segments of residents regarding pollution reporting. |
| Key findings | The paper identifies the most promising urban resident behaviors to target to reduce stormwater pollution and improve natural waterway habitats in urban areas. It also investigates the underlying capabilities, opportunities and motivations for residents' reporting of pollution in stormwater drains and waterways, and identifies five potential target audiences: 'Supportive', 'Unaware but receptive', 'Motivated but lack support', 'Reluctant', and 'Not my problem'. | The key findings in the paper include:  1. Identification of 29 behaviors that urban residents can adopt to reduce stormwater pollution and improve freshwater biodiversity, with a focus on the most promising behaviors based on effectiveness and likelihood of adoption.  2. A significant portion of urban residents expressed a willingness to report pollution, while financial contributions to reduce pollution were notably lower.  3. Various barriers to reporting pollution were identified, including lack of awareness, convenience, motivation, and social norms.  4. Five audience segments were defined based on behaviors and attitudes towards reporting pollution: Supportive, Unaware but Receptive, Motivated but Lack Support, Reluctant, and Not My Problem, each requiring tailored interventions.  5. The importance of understanding individual motivations and barriers to enhance community participation in environmental initiatives was emphasized.  6. Recommendations for targeted interventions include making reporting easy and accessible, promoting community norms, providing feedback to reporters, and framing messages around personal and community values.  7. The study utilized frameworks to systematically identify drivers and barriers, suggesting a structured approach to developing effective behavior change strategies.  8. The findings highlight the need for a coordinated approach to address the human element in urban freshwater management, advocating for comprehensive strategies to encourage pro-environmental actions. |
| Population | The study involved 1901 survey respondents. The average age was 47.2 years, with a range of 18 to 90 years. Over half of the respondents were female (58%), with 42% male and a small percentage identifying as non-binary. The respondents' education levels varied, with the majority having completed an undergraduate degree (36%). | The studied population consists of 1901 survey respondents with an average age of 47.2 years, ranging from 18 to 90 years. The gender distribution includes 1098 females (58%), 794 males (42%), and nine individuals identifying as non-binary. In terms of education, 29% completed secondary education, 21% had a trade or technical qualification, 36% completed an undergraduate degree, 7% had a higher degree, and 7% had no qualifications or were unsure. Most respondents (83%) reported never having reported pollution in stormwater drains and waterways. |
| Relations | The paper explores the relationship between urban residents' behaviors and freshwater biodiversity. It identifies key behaviors that can improve freshwater biodiversity, such as reporting pollution in stormwater drains and waterways. The paper then analyzes the factors that influence these behaviors, including awareness, opportunity, social motivation, and personal motivation. The paper also identifies five distinct audience segments based on their attitudes and behaviors towards pollution reporting, highlighting the importance of tailoring interventions to specific audience needs. | The key findings of the paper reveal a complex interplay between various indicators influencing urban residents' behaviors related to freshwater biodiversity and pollution reporting. Awareness and motivation are critical determinants; individuals who are more aware of pollution issues and motivated to act are more likely to engage in reporting behaviors. Community dynamics, including social norms and support, significantly influence individual participation, with communities that encourage reporting seeing higher engagement levels.  Perceived responsibility also plays a vital role, as individuals who feel a sense of duty towards environmental stewardship are more inclined to report pollution. Convenience and accessibility of reporting tools are essential; simplifying the reporting process can enhance participation. Additionally, providing feedback on the outcomes of reports fosters a positive experience, encouraging future engagement.  Demographic factors such as age, education, and gender further influence willingness to participate, highlighting the need for tailored interventions that address specific barriers faced by different audience segments. The segmentation into groups like 'Supportive', 'Unaware but receptive', 'Motivated but lack support', 'Reluctant', and 'Not my problem' allows for targeted strategies that resonate with each segment's unique motivations and challenges.  Overall, the findings emphasize that a coordinated approach addressing awareness, community support, perceived responsibility, convenience, and demographic influences is essential for promoting sustainable behaviors and enhancing freshwater conservation efforts. |
| ODH\_keys: | The paper addresses \*\*digital health\*\*. | The keys addressed in this paper are \*\*One Health\*\* and \*\*digital health\*\*. |
| ODH\_perspectives: | The paper focuses on \*\*individual health and well-being\*\* and \*\*population and society\*\*. | The perspectives addressed in this paper are \*\*individual health and well-being\*\*, \*\*population and society\*\*, and \*\*ecosystem\*\* |
| ODH\_dimensions: | The paper addresses \*\*citizens' engagement\*\*, \*\*education\*\*, and \*\*environment\*\* dimensions of One Digital Health. | The dimensions addressed in this paper are \*\*environment\*\* and \*\*human and veterinary health care\*\*. |
| **PMID**: 37899012  **Title**: Light pollution of freshwater ecosystems: principles, ecological impacts and remedies.  **Source**: Philos Trans R Soc Lond B Biol Sci. 2023 Dec 18;378(1892):20220360. doi: 10.1098/rstb.2022.0360. Epub 2023 Oct 30.  **PMC**: PMC10613548  **Authors**: Holker F, Jechow A, Schroer S, Tockner K, Gessner MO  **Publication Date**: 2023 Dec 18  **Keywords**: aquatic community dynamics, circadian rhythms, conservation, ecosystem functioning, land-water interactions, light physics  **MeSH Terms**: Animals, Humans, \*Ecosystem, \*Light Pollution, Biodiversity, Fresh Water, Rivers | | |
|  | **Gemini** | **OpenAI** |
| Token count (output) | 614 | 1263 |
| Processing time | 80 sec | 102 sec |
| OAH Score | The paper focuses on the impact of artificial light on freshwater ecosystems. While it mentions the importance of considering broader environmental factors, it doesn't directly address the digital transformation of healthcare systems or the integration of human, animal, and environmental health. Therefore, the paper's relevance to One Digital Health is limited.  <<2>> | (Screened by Gemini in Step 3) |
| ODH Score | This paper focuses on the impact of light pollution on freshwater ecosystems, particularly the ecological consequences and potential remedies. OneAquaHealth aims to improve the sustainability and integrity of freshwater ecosystems in urban environments. The paper's insights into the negative effects of light pollution on freshwater ecosystems are directly relevant to OneAquaHealth's goals of enhancing environmental monitoring and understanding the interconnection of ecosystem health and human well-being.  <<9>> | (Screened by Gemini in Step 3) |
| Paper aim | The paper aims to provide a comprehensive understanding of light pollution in freshwater ecosystems, including its ecological impacts and potential remedies. It explores the effects of artificial light at night on various aspects of aquatic and terrestrial ecosystems, including insect behavior, fish migration, and predator-prey interactions. The paper also aims to establish relevant environmental parameters for urban lighting, translating research into practice. | The aim of the paper is to explore the ecological consequences of artificial light at night (ALAN) on various organisms and ecosystems, particularly focusing on its effects on aquatic environments, biological processes, behaviors, and interactions among species. It also discusses the need for effective lighting solutions to mitigate light pollution's adverse impacts on biodiversity and ecosystems. |
| OAH Indicators | The paper focuses on the ecological impacts of light pollution on freshwater ecosystems. It mentions indicators like species diversity, ecosystem functioning, and the behavior and physiology of aquatic species. The paper does not describe how these indicators were measured. | The paper discusses the ecological impacts of light pollution on freshwater ecosystems, highlighting various indicators related to biodiversity and ecosystem functioning. Specific ecological indicators mentioned include species diversity and behavioral responses of aquatic species to artificial light at night (ALAN). These indicators are measured through experiments that observe changes in species behavior, such as diel vertical migration of zooplankton and melatonin production in fish, in response to varying light levels. The paper emphasizes the need for understanding these biological responses to assess the ecological health of freshwater systems affected by light pollution. |
| geolocation | The provided text does not mention a specific geographical location. | The geographical location the paper focuses on is not specified in the provided information. |
| methodology | The paper used a combination of observational and experimental methods. | The paper does not specify a particular methodology or design, as it presents no original data. It focuses on conceptualization, funding acquisition, visualization, and writing, with contributions from multiple authors in these areas. |
| Key findings | The paper highlights the increasing problem of light pollution and its impact on freshwater ecosystems. It emphasizes the need for environmentally friendly lighting concepts that minimize ecological impacts. The paper also discusses the importance of considering the spectral sensitivities of aquatic species when planning lighting installations. The paper explores the impact of artificial light at night on various aquatic organisms, including phytoplankton, fish, and invertebrates. It highlights the potential disruption of natural light cycles, affecting their behavior, physiology, and ecological interactions. | The key findings in the paper highlight the significant ecological impacts of artificial light at night (ALAN) on freshwater ecosystems and their inhabitants. Key points include:  1. There has been a notable increase in night-time sky brightness, primarily due to the rise of LED lighting, with growth rates approaching 10% per year.  2. Freshwater ecosystems, which are rich in biodiversity, are particularly susceptible to light pollution, affecting species behavior, physiology, and community dynamics.  3. Even low levels of light can disrupt behaviors such as diel vertical migration in zooplankton and melatonin production in fish, potentially altering species distribution and community composition.  4. Many freshwater ecosystems remain understudied in the context of light pollution, despite their importance for biodiversity and human well-being.  5. There is a critical need for effective mitigation strategies that consider the specific light properties and biological responses in freshwater environments.  6. Significant gaps in knowledge exist regarding the ecological consequences of light pollution, necessitating targeted research and innovative lighting solutions to minimize ecological disruption.  7. ALAN can create movement barriers for species, alter predator-prey dynamics, and lead to changes in community composition and ecosystem processes.  8. The cumulative effects of light pollution can exacerbate other environmental stressors, complicating assessments of its impacts on biodiversity and ecosystem functioning.  9. The paper emphasizes the importance of raising awareness among citizens and decision-makers about the ecological impacts of light pollution to enhance the protection of freshwater biodiversity. |
| Population | The context does not provide information about the studied population's characteristics, such as number, species, or age. | The provided information does not specify the characteristics of the studied population, such as number, species, or age. It primarily discusses the effects of artificial light at night on various ecological aspects and references multiple studies related to this topic. |
| Relations | The paper explores the complex relationship between artificial light at night (ALAN) and various indicators in freshwater ecosystems. It highlights how ALAN can disrupt natural light cycles, impacting the behavior, physiology, and ecological interactions of aquatic organisms. This disruption can lead to changes in species distribution, community structure, and ecosystem processes, potentially altering the overall functioning of freshwater ecosystems. The paper emphasizes the need for environmentally friendly lighting concepts that minimize light pollution and protect biodiversity. | The key findings of the paper illustrate a complex relationship between artificial light at night (ALAN) and various ecological indicators within freshwater ecosystems. ALAN disrupts natural behaviors and physiological processes in aquatic organisms, affecting circadian rhythms, hormone levels, and growth. This disruption can lead to altered migration patterns, such as diel vertical migration in species like zooplankton and fish, which increases their vulnerability to predation.  The presence of artificial light also influences predator-prey dynamics, as predators may exploit prey attracted to light, leading to shifts in community structures and species interactions. Additionally, light pollution can affect microbial diversity and community respiration in sediments, impacting nutrient cycling and ecosystem processes.  Furthermore, the interaction of ALAN with other environmental stressors, such as pollutants and noise, complicates its ecological impacts. For instance, light can alleviate some negative effects of pollutants on ecosystems while simultaneously altering the behavior and growth of key species like freshwater shredders and plants.  Overall, the findings emphasize the interconnectedness of light pollution with various ecological indicators, highlighting the need for targeted mitigation strategies to protect biodiversity and maintain ecosystem health in freshwater environments. |
| ODH\_keys: | The paper addresses \*\*digital health\*\* and \*\*One Health\*\*. | The keys addressed in this paper are \*\*One Health\*\* and \*\*digital health\*\*. |
| ODH\_perspectives: | The paper addresses the \*\*individual health and well-being\*\* perspective. | The perspectives addressed in this paper are \*\*individual health and well-being\*\*, \*\*population and society\*\*, and \*\*ecosystem\*\* |
| ODH\_dimensions: | The paper focuses on the impact of light pollution on the environment and its effects on living beings. This aligns with the \*\*environment\*\* dimension of One Digital Health. | The dimensions addressed in this paper are \*\*environment\*\* and \*\*human and veterinary health care\*\*. |

Table - Full text assessment - Gemini vs. OpenAI examples

#### Performance Evaluation

Compared to all previous steps, this step is the most challenging to the LLMs. It requires in depth understanding of the paper’s context provided per question, as well as strong inference and reasoning capabilities.

The evaluation of this step is also different: unlike all previous steps, where most replies were extracted as structured text, in this step, all replies are provided as free text and are mostly domain specific. It requires Human in the loop: OAH domain experts’ assessment of the papers’ analyses by both LLMs.

This step is also the most resource intensive step. It handles between 10-20 LLM calls per paper, and requires output token count monitoring.

In order to evaluate this step’s outputs, we randomly chose 55 papers from the corpus and performed *Full Text Assessment* using both LLMs.

The outputs were shared with OAH domain experts. The observations are detailed below.

#### Observations:

* The responses from both LLMs to the same question are consistently aligned throughout the corpus and match the content of the paper a well as the question asked.
* Top of Form
* Bottom of Form
* Both LLMs exhibit good inference capabilities throughout the corpus.
* OpenAI’s replies are significantly more comprehensive, detailed and accurate compared to Gemini
* Compared to OpenAI, Gemini tends to provide vague, generic answers when the paper lacks information about the question asked
* Average processing time is 25% - 40% shorter for Gemini
* Average output token length is 30% - 70% longer for OpenAI

### Full OAH-AI PRISMA Review - Summary

**Identification of studies via databases and registers**

Records removed *before screening*:

Duplicate records removed (n = 2,239)

Records marked as ineligible by automation tools (n = 0)

Records removed for other reasons (n = 0)

Records identified from:

PubMed Database

(n = 10,057)

**Identification**

Records excluded by AI:

(n = 693)

No full text (n = 16)

Language (n = 26)

Publication date (n = 42)

OAH relevance (n = 235)

Records screened\*

(n = 1,000)

**Screening**

Reports sought for retrieval

(n = 307)

Reports not retrieved

(n = 69)

Reports excluded by AI:

OAH relevance (n = 104)

Reports assessed for eligibility

(n = 123)

Studies included in review \*\*

(n = 32)

Reports of included studies

(n = 20)

**Included**

OAH-AI Summary - 5: Full PRISMA Assessment

Source: Page MJ, et al. BMJ 2021;372:n71. doi: 10.1136/bmj.n71.

This work is licensed under CC BY 4.0. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

\* 10% for PoC

\*\* 25% for PoC

## Part II – OAH-AGI

Given the complexity of evaluating outputs from the entire dataset of 123 papers that underwent full-text assessment, we selected a representative subset of 32 papers (25% of the total) for this PoC. The list of these selected papers is provided in Appendix 7.

### First Process: Generating Partial Reports

The first process in the pipeline, which involves generating partial reports based on random subsets of the selected papers, was executed multiple times. This iterative approach produced a substantial dataset of partial reports, each containing insights such as patterns, correlations, and hypotheses derived from the selected papers.

All iterations’ outputs are stored under project files /Ideas.

### Second Process: Summarizing Reports

The partial reports generated in the first process were then processed through the second step of the pipeline. Here, the LLMs were tasked with consolidating the reports into coherent summaries, removing duplicates, and merging similar insights. This summarization process was also repeated several times using both the Gemini and OpenAI models.

Each iteration's outputs are stored under project files /Outputs/OAH-AI/.

### Summary of Results and Observations

An example of the final outputs genereated by both LLM is in the table below:

|  |  |  |
| --- | --- | --- |
| Task | Gemini | OpenAI |
| Patterns, Correlations, and Recurring Theme | \* \*\***Urbanization Impacts**\*\*: Urbanization negatively impacts aquatic ecosystems through pollution, altered species composition, and reduced resource efficiency (PMID: 35099004, 34955197, 35283080).  \* \*\***Water Quality and Health**\*\*: Poor water quality increases the risk of disease vectors and health concerns (PMID: 35099004, 35283080, 37586536, 37558875).  \* \*\***Biodiversity and Ecosystem Health**\*\*: Biodiversity decline is linked to urbanization and pollution, impacting ecosystem services and human health (PMID: 37586536, 34955197).  \* \*\***Monitoring is Crucial**\*\*: Continuous monitoring of aquatic ecosystems is vital to detect early warning signs of degradation and inform management strategies (PMID: 37586536, 35099004, 37558875).  \* \*\***Human Health Linkages**\*\*: Environmental degradation, water quality, and public health are interconnected, requiring integrated management approaches (PMID: 35283080, 37586536, 25889666). | - \*\***Impact of Urbanization**\*\*: Urbanization is consistently linked to environmental degradation in aquatic ecosystems, leading to increased pollution levels, altered nutrient dynamics, and reduced biodiversity. Studies indicate that urban environments often facilitate higher contaminant loads, resulting in detrimental effects on water quality and species composition (PMIDs: 29996397, 35099004, 37856511, 38557568).  - \*\***Water Quality as a Health Indicator**\*\*: Poor water quality is correlated with elevated health risks, particularly regarding vector-borne diseases and exposure to pathogens. Water quality metrics, such as nutrient levels (nitrogen and phosphorus), chemical contaminants, and microbial indicators, are crucial for assessing the health of aquatic ecosystems and the associated risks to public health (PMIDs: 36529169, 26771345, 35283080, 37586536).  - \*\***Biodiversity and Ecosystem Health**\*\*: Biodiversity loss is a recurring concern, with metrics like species richness and functional diversity often used as proxies for ecological resilience and health. Declines in these metrics are associated with increased environmental stress and urbanization pressures (PMIDs: 37558875, 37586536).  - **\*\*Emerging Contaminants and Monitoring Needs**\*\*: The presence of contaminants of emerging concern, such as pharmaceuticals and heavy metals, highlights the need for improved monitoring and regulatory strategies, as these pollutants pose risks to both ecosystems and human health (PMIDs: 35283080, 26771345).  - \*\***Importance of Integrated Monitoring and Management**\*\*: Effective monitoring and management strategies are essential for mitigating the impact of contaminants on aquatic ecosystems. The integration of ecological indicators with public health considerations is critical for informed decision-making (PMIDs: 37586536, 38557568). |
| Indicators Supporting Detection of Disease Vectors or Environmental Health Proxies | \* \*\*Ecological Indicators\*\*:  \* \*\*Bacterial Composition\*\*: Changes in fish skin microbiota reflect ecosystem health (PMID: 35099004).  \* \*\*Mosquito Species\*\*: Species diversity and abundance indicate public health risks from mosquito-borne diseases (PMID: 25889666).  \* \*\*Water Quality Indicators\*\*:  \* \*\*Pollutant Concentrations\*\*: Levels of PFAAs and heavy metals in river systems are health proxies (PMID: 35283080, 29800822).  \* \*\*Nutrient Levels\*\*: Total nitrogen and phosphorus affect phytoplankton dynamics and ecosystem health (PMID: 34955197).  \* \*\*Biodiversity Metrics\*\*:  \* \*\*Species Richness and Functional Diversity\*\*: Indicators of ecological resilience and potential health implications (PMID: 37586536).  \* \*\*Environmental Stressors\*\*:  \* \*\*Eutrophication Levels\*\*: Indicate ecosystem health and potential pathogen emergence (PMID: 35099004). | - \*\*Microbial Indicators\*\*: The presence of pathogens (e.g., E. coli) and antibiotic-resistant bacteria in water bodies serves as a direct indicator of health risks associated with contaminated water sources (PMIDs: 32470679, 36541332).  - \*\*Water Quality Metrics\*\*: Key indicators such as nutrient levels (nitrogen and phosphorus), biochemical oxygen demand (BOD), and chemical oxygen demand (COD) are critical for assessing the health of aquatic ecosystems (PMIDs: 36529169, 37586536).  - \*\*Biodiversity Metrics\*\*: Species diversity and richness are used as proxies for ecological health, indicating the potential for disease emergence and resilience to environmental stressors (PMIDs: 37558875, 37586536).  - \*\*Trophic Transfer Indicators\*\*: The concentration of heavy metals in aquatic organisms provides insights into bioaccumulation risks and the potential health impacts on humans through seafood consumption (PMIDs: 33991905, 37856511). |
| Early Warning Indicators for Ecosystem Degradation | \* \*\*Biodiversity Metrics\*\*: Decreases in species richness and evenness signal ecological stress and potential degradation (PMID: 37586536, 34955197).  \* \*\*Microbial Community Changes\*\*: Alterations in fish skin microbiota due to eutrophication are early warning signs (PMID: 35099004).  \* \*\*Water Quality Indicators\*\*: Increased levels of pollutants (PFAAs or heavy metals) in water and sediments signal ecosystem degradation (PMID: 35283080, 29800822).  \* \*\*Resource Use Efficiency\*\*: Decline in phytoplankton resource use efficiency suggests potential degradation (PMID: 34955197).  \* \*\*Fish Health Metrics\*\*: Increased ectoparasite loads on fish indicate declining health and environmental stress (PMID: 35099004). | - \*\*Nutrient Enrichment\*\*: Elevated levels of nitrogen and phosphorus are early indicators of eutrophication, which can lead to harmful algal blooms (HABs) and other ecological disruptions (PMIDs: 26771345, 35283080).  - \*\*Microbial Community Shifts\*\*: Changes in the composition and diversity of microbial communities, particularly the emergence of pathogens and antibiotic-resistant strains, signal declining water quality (PMIDs: 33485000, 35099004).  - \*\*Biodiversity Trends\*\*: A decrease in species richness or shifts towards pollution-tolerant species can indicate ecosystem stress and potential degradation (PMIDs: 37558875, 38557568).  - \*\*Temperature Increases\*\*: Rising water temperatures associated with urbanization and climate change serve as indicators of ecological shifts and declining aquatic health (PMIDs: 37979332, 26771345). |
| Potential Hypotheses Based on the Text | \* \*\*H1\*\*: Higher urbanization leads to increased eutrophication, impacting microbial communities and fish health (PMID: 35099004, 34955197).  \* \*\*H2\*\*: Elevated PFAAs in freshwater systems correlate with increased human health risks, especially in urban areas (PMID: 35283080).  \* \*\*H3\*\*: Reduced phytoplankton diversity and resource use efficiency are linked to rising nutrient levels and urban runoff, suggesting a predictive indicator of ecosystem degradation (PMID: 34955197).  \* \*\*H4\*\*: Specific mosquito species in newly created wetlands correlate with public health risks from vector-borne diseases, requiring proactive management (PMID: 25889666).  \* \*\*H5\*\*: Long-term biodiversity trends show plateauing recovery despite initial efforts, suggesting the need for adaptive management practices (PMID: 37586536). | - \*\*Hypothesis 1\*\*: Increased urbanization correlates with higher nutrient pollution in freshwater ecosystems, leading to increased occurrences of harmful algal blooms and associated health risks for nearby populations (PMIDs: 33774562, 36529169).  - \*\*Hypothesis 2\*\*: The presence of antibiotic-resistant bacteria in urban lakes is positively correlated with pollution levels and urban runoff, indicating significant health risks (PMIDs: 34479163, 36529169).  - \*\*Hypothesis 3\*\*: Changes in the diversity of algal and bacterial communities serve as reliable indicators of water quality, reflecting the impacts of anthropogenic activities (PMIDs: 37952398, 34479163).  - \*\*Hypothesis 4\*\*: Effective management strategies, including riparian vegetation restoration and monitoring of aquatic species, can mitigate the negative impacts of urbanization on freshwater ecosystem health (PMIDs: 25889666, 37586536).  - \*\*Hypothesis 5\*\*: The concentration of emerging contaminants in urban freshwater systems is associated with increased prevalence of waterborne diseases in local populations, necessitating improved monitoring and management efforts (PMIDs: 35283080, 26771345). |
| Reference list (PMIDs) | 1. 25889666  2. 29800822  3. 34955197  4. 35099004  5. 35283080  6. 37558875  7. 37586536 | 1. 25889666  2. 26771345  3. 29996397  4. 32470679  5. 33774562  6. 33991905  7. 34479163  8. 35099004  9. 35283080  10. 36529169  11. 36541332  12. 37558875  13. 37586536  14. 37856511  15. 37952398  16. 37979332  17. 38050711  18. 38557568  19. 33485000  20. 37856511 |

Table - Final Report - LLM Comparison

The outputs from these iterative runs, using both Gemini and OpenAI, demonstrated the following results:

#### 1. Key Patterns and Themes Identified

* **Urbanization and Ecosystem Degradation**: Both models highlighted the consistent negative impacts of urbanization on aquatic ecosystems, including pollution, altered species composition, and reduced biodiversity. Gemini focused on direct cause-effect relationships with pollutants like PFAAs, while OpenAI expanded on broader impacts, such as nutrient dynamics and habitat degradation.
* **Water Quality as a Health Indicator**: Both models underscored water quality's critical role in indicating ecosystem and public health. OpenAI elaborated more on the broader risks of contaminants, including emerging concerns like antibiotic-resistant bacteria.
* **Biodiversity as an Indicator of Ecosystem Health**: Both models agreed that biodiversity metrics, such as species richness and functional diversity, are essential for understanding ecosystem resilience and health. OpenAI provided additional insights on shifts toward pollution-tolerant species, while Gemini highlighted the role of biodiversity in ecosystem services and human health.
* **Monitoring and Management**: Both models emphasized the importance of continuous monitoring for early detection of ecosystem degradation. OpenAI integrated the influence of climate change and urbanization in shaping monitoring needs, while Gemini focused on established management strategies.

#### 2. Indicators of Disease Vectors and Ecosystem Health

* **Ecological and Microbial Indicators**: Both models identified changes in microbial communities, particularly fish skin microbiota and pathogen presence, as strong indicators of ecosystem health. OpenAI also highlighted the significance of antibiotic-resistant bacteria as a public health risk, emphasizing the need for integrated management strategies.
* **Water Quality Indicators**: Both Gemini and OpenAI recognized pollutants, nutrient levels, and heavy metals as crucial proxies for ecosystem health. Gemini focused more on pollutant concentrations (e.g., PFAAs), while OpenAI expanded on nutrient dynamics and biochemical oxygen demand (BOD).
* **Biodiversity Metrics**: Both models highlighted species richness and functional diversity as critical ecological resilience and health indicators. OpenAI expanded on the implications of biodiversity shifts in relation to disease vector risks and potential human health impacts.

#### 3. Early Warning Indicators of Ecosystem Degradation

* **Nutrient Enrichment and Microbial Community Shifts**: Both models identified nutrient levels (e.g., nitrogen and phosphorus) and shifts in microbial communities as early indicators of ecosystem degradation. OpenAI emphasized the role of harmful algal blooms (HABs) as a clear signal of ecological stress.
* **Water Quality Degradation**: Increased levels of pollutants like PFAAs and heavy metals were consistently identified by both models as warning signs of ecosystem degradation. OpenAI additionally flagged rising water temperatures and their impacts on ecosystem health, driven by climate change and urbanization.
* **Biodiversity Decline**: Both models agreed that declines in species richness and shifts towards pollution-tolerant species indicate significant ecological stress. OpenAI highlighted the need to monitor these shifts as they could predict broader ecosystem degradation.

#### 4. Hypotheses Generated

Both models generated a range of hypotheses based on their analysis, with OpenAI often providing more speculative, yet insightful, hypotheses:

* **Urbanization and Ecosystem Health**: Both models hypothesized that urbanization leads to eutrophication and shifts in microbial communities, affecting overall ecosystem and fish health. OpenAI extended this to include risks of harmful algal blooms and broader ecological stress.
* **PFAAs and Human Health**: Both models proposed that elevated levels of PFAAs in freshwater systems are linked to increased public health risks, particularly in urban areas.
* **Phytoplankton Diversity and Ecosystem Degradation**: Both models agreed that reductions in phytoplankton diversity and resource use efficiency could serve as predictive indicators of ecosystem degradation. OpenAI added that monitoring these shifts could inform proactive management strategies.
* **Mosquito Species and Disease Risks**: Both models identified the presence of specific mosquito species in urban and newly created wetlands as an indicator of increased vector-borne disease risks. Gemini focused on the management implications, while OpenAI suggested broader preventive strategies.
* **Long-Term Biodiversity Trends**: Both models proposed that long-term biodiversity monitoring indicates a plateau in recovery, suggesting the need for adaptive management. OpenAI integrated climate change impacts into this hypothesis, highlighting the complexity of managing urban ecosystems.

The results from both Gemini and OpenAI models provide a comprehensive understanding of the key drivers and indicators of aquatic ecosystem health in urban environments. While Gemini excelled in structured analysis and hypothesis generation based on existing research, it failed to perform basic reasoning and inference tasks in collecting all PMIDs from text into a sorted list, a task OpenAI performed well. OpenAI’s outputs were generally more detailed and comprehensive, particularly regarding the specific indicators and their implications. It also demonstrated creativity in proposing new hypotheses and exploring broader ecological impacts. Both models consistently identified urbanization, water quality, and biodiversity as central themes, making a compelling case for integrating AI-driven analysis in large-scale environmental research.

# Discussion

In this Discussion section, we explore the role of heuristic scientific discovery and how generative AI, particularly through LLMs, contributes to this approach within the OAH-AI framework. We begin with an examination of heuristic methods in scientific discovery, followed by an analysis of the integration of LLMs in OAH-AI to enhance reasoning, creativity, and domain-specific knowledge through tailored settings and retrieval-augmented generation.

The section delves into the comparative benefits of closed- and open-source LLMs, the role of human oversight for accuracy, and lessons learned in prompt design. It also presents a comparison between OpenAI and Gemini models and addresses ethical considerations, sustainability, and potential avenues for future research. This comprehensive analysis aims to offer insights into the nuanced challenges and innovative opportunities for using AI in domain-specific scientific discovery.

## Overview

In this work, we collected a corpus of 1,000 scientific papers from PubMed that matched one or more of our OAH-related search queries and sent those papers through our OAH-AI framework pipeline to perform a full PRISMA review.

Out of the 1,000 papers, 307 papers were initially screened through metadata analyses. After this step, 238 papers were assessed for eligibility, and 134 papers were finally selected for full-text assessment. The processing and analysis of the selected papers within the framework were achieved via interactions with one or more LLMs, specifically OpenAI’s GPT-4o-mini and Google’s Gemini-1.5-flash, using natural language prompts.

We designed the framework through the evaluation of the LLMs' inference and causality skills. These models were instructed to rate the relevance of papers, perform logical reasoning, and extract knowledge from given contexts. Following initial experiments on a small subset of papers, we applied a different LLM mechanism at each step:

* For the *Screening* step, we used both LLMs as screening agents, accepting papers that were approved by both models.
* For the *Eligibility* step, we only used Gemini.
* For the *Full-Text Assessment*, we utilized OpenAI.

Throughout the design of the OAH-AI framework we applied a heuristic scientific discovery approach, as discussed in the coming sections.

## Heuristic Scientific Discovery

In his article "Integrated Systems for Computational Scientific Discovery," Pat Langley (Langley, 2024) describes the task of scientific discovery in generic, domain-independent terms:

* Given:
  + Scientific data to be described or explained.
  + Knowledge about the scientific domain.
  + Given: A space of candidate categories, laws, or models.
* Find:
  + Candidates that describe or explain the observations.

This process bears similarities to Fayyad et al.'s Knowledge Discovery in Databases framework (Fayyad et al., 1996), which also seeks to find meaningful patterns in large datasets. However, Langley identifies two key distinctions between scientific discovery and data mining:

1. Understandability: While data mining often uses machine-understandable formalisms, such as decision trees or Bayesian networks, scientific discovery is concerned not only with prediction but also with producing explanations that are understandable to humans. This distinction is crucial for communicating findings within the scientific community.
2. Data Set Size and Approach: Data mining focuses on processing large datasets efficiently, whereas scientific discovery typically deals with smaller, more specialized datasets. Moreover, scientific discovery may involve abductive reasoning, where hypotheses are generated and tested based on domain expertise rather than brute-force pattern recognition.

Both paradigms share the challenge of navigating a vast search space of potential models and laws, which can be too large for exhaustive techniques. Heuristic guidance is therefore essential to steer the discovery process toward promising hypotheses and avoid computational intractability.

### Generative AI and Heuristic Discovery

With the growing interest in GenAI, it’s important to recognize that the search spaces in scientific discovery are inherently generative. They are not defined by an explicit set of candidates but rather by a starting point and a set of ‘operators’ that produce new candidates, akin to techniques used in planning and game-playing AI systems. Generative AI models, particularly LLMs, can explore this search space by generating hypotheses, explanations, and potential patterns from vast amounts of text and data. In this sense, LLMs serve as heuristic tools that guide the scientific discovery process by exploring potential relationships that would be difficult for human researchers to identify within reasonable timeframes.

### Heuristic Approach in OAH-AI

In our study, we adopted a heuristic scientific discovery approach with no intent to capture a complete, exhaustive view of the scientific landscape. Rather, we aimed to explore "what's out there" by leveraging LLMs' ability to comb through large amounts of literature and generate novel insights.

Our objective was not to prioritize reproducibility in the strictest sense, but to allow the AI models to generate hypotheses and identify emerging patterns in the ecological and health indicators relevant to urban freshwater environments. This aligns with Langley’s perspective that scientific discovery often involves abductive reasoning—using limited data to infer potential explanations, which can then be further validated or refuted through more detailed study.

This approach has significant value for OAH because it allows us to rapidly gather a wide range of insights, identify gaps in the literature, and propose potential new indicators for ecosystem health monitoring. The heuristic approach also aligns well with the goals of exploratory research, where the primary objective is to open new pathways for investigation, rather than finalize a definitive set of answers.

By focusing on heuristic discovery, we emphasize the exploratory nature of this research, acknowledging that the AI models may not provide definitive solutions, but rather, open up new avenues for investigation. This flexibility is particularly valuable in the complex, multidisciplinary domain of urban freshwater ecosystem monitoring, where human, animal, and environmental health relationships are still emerging and poorly understood.

## Using LLMs within OAH-AI

In the following section we discuss our findings, methodology and observations from various aspects related to LLM usage in the framework.

### Customizing LLMs’ Reasoning and Creativity Settings

We used two mechanism to control the type of responses we requested from the LLM in the OAH-AI framework. One was Prompt engineering – the LLM was requested to respond with structured data (e.g., ‘reply only yes or no’, or ‘end your reply with final score in this format: <<score>>’) to some questions, and with free, unconstrained text to others. The second mechanism was the Temperature setting, as detailed in the following section.

#### LLM Temperature

When interacting with LLMs, one key parameter influencing their behavior is temperature. This setting controls the randomness of the model's output. A low temperature (closer to 0) prompts the LLM to provide more deterministic and predictable answers, which is especially useful for structured tasks such as extracting specific data points or performing logical reasoning. In contrast, a high temperature allows the model to generate more creative, varied responses, which can be helpful in brainstorming or exploratory tasks.

In the OAH-AI pipeline, balancing the temperature setting was crucial for optimizing LLM performance. For tasks that required precise reasoning and structured outputs, such as screening based on inclusion criteria, or assessing paper eligibility, we opted for a low temperature. This ensured that the LLMs remained focused and logical in their responses. However, for broader tasks like generating hypotheses or exploring correlations across the entire corpus, a higher temperature allowed the models to think outside the box, offering creative insights.

This differentiation underscores a key limitation of current LLMs (as of August 2024): while they excel at tasks requiring structured reasoning, they are far from Artificial General Intelligence (AGI), which could theoretically perform all cognitive tasks a human could. LLMs require careful tuning to perform specialized reasoning tasks, while AGI would ideally handle these with greater autonomy and versatility.

### Context Window and Retrieval-Augmented Generation

When we first began designing the OAH-AI framework, one of the key challenges we faced was the context window-the amount of text an LLM can process in a single call. This posed a significant problem when working with large volumes of text, as the entire context often could not fit into a single prompt.

There are several ways to overcome this limitation. One option is to retrain or fine-tune the model, but a more efficient solution is using a Retrieval-Augmented Generation (RAG) approach. RAG works by querying an external knowledge source (in our case, the corpus of scientific papers) and feeding only the most relevant sections back into the LLM for analysis. This allows the model to generate answers based on pertinent, up-to-date information without exceeding its context window or requiring costly retraining.

By using RAG, we were able to bypass the memory constraints of LLMs, offering a scalable solution when handling large volumes of complex scientific literature, as we did with the OAH corpus. It also avoided the challenges of fine-tuning a general-purpose LLM for domain-specific tasks, which can be both resource-intensive and time-consuming.

Although RAG was the preferred approach for many use cases early on, this is quickly changing. Initially, GPT-3.5-Turbo, OpenAI's leading LLM, had a context window of 4,096 tokens. Early versions of GPT-4 extended the context window to 8,000–32,000 tokens, depending on the model variant. However, a newer model, the one we chose for the OAH-AI framework, GPT-4o-mini, now supports a context window of 128,000 tokens, and Gemini-1.5-flash – one million tokens. This shift has made RAG unnecessary for many applications. We took advantage of the extended window size in part 2 of the OAH-AI, as we were able to send more papers for cross-analyses at each LLM call.

### Human-in-the-Loop for Domain-Specificity

Despite their impressive capabilities, LLMs still require human oversight, especially when applied to highly specialized domains such as human health and ecology. These domains often involve nuanced understanding and interpretation, which LLMs can struggle with. For example, the context and implications of a particular ecological health indicator may not be fully captured by the LLM without human guidance.

In our approach, a human-in-the-loop methodology was essential to ensure that the LLMs' outputs were not only correct but also meaningful in the context of OAH.

For an optimal framework output, it is crucial that domain experts are involved to tune and direct key steps of the process, such as filtering results tagged as ‘TBD’ and validating the LLMs' reasoning during the eligibility and full-text assessment stages. This hybrid model ensures the accuracy and relevance of AI-generated insights while leveraging human expertise where necessary.

### Closed vs. Open-Source LLMs

The debate between closed-source and open-source LLMs remains an important consideration for researchers. Closed-source models like OpenAI’s GPT-4 and Google’s Gemini provide state-of-the-art capabilities, often leading the field in terms of performance, accuracy, and natural language understanding. However, they come with notable limitations, including high costs, lack of transparency, and limited adaptability, as researchers do not have direct access to the model's architecture or training data. Additionally, users must rely on proprietary APIs, restricting control over the deployment environment and model fine-tuning.

In contrast, open-source LLMs, such as Meta’s LLaMA, offer greater flexibility and customizability, allowing researchers to fine-tune or retrain models according to their specific needs. Open-source models provide transparency into their inner workings, which can be critical when domain-specific adjustments are required. Furthermore, specialized LLMs like Palmyra-Med-70B-32K (*Writer/Palmyra-Med-70B-32K · Hugging Face*, 2024) - a leading open-source model in the biomedical field with an impressive benchmark score of 85.87%, outperforming GPT-4, Gemini, and even medically trained human experts - offer domain-specific advantages. These models are freely available for use, given the right cloud infrastructure, making them a powerful tool for tasks that require domain expertise, such as health and ecological monitoring.

However, open-source models present challenges of their own. Unlike closed-source models, they often lack easily accessible APIs, requiring the model to be downloaded and run locally. In free environments such as Google Colab, only the smallest versions of these models can be downloaded and used due to resource constraints, which significantly limits their performance.

For example, during our initial experiments with LLaMA-3, we worked with its smallest variant due to computational restrictions, but found its performance to be unsatisfactory. It struggled with basic reasoning tasks and demonstrated weak inference capabilities, making it unsuitable for the complexity of our research needs.

While there are many other strong LLMs available, and competition remains fierce with ongoing advancements, we ultimately decided to go with Google’s Gemini and OpenAI’s GPT. Their simple-to-ues API access, cutting-edge performance, combined with significant recent price reductions, made them the most suitable choices for our OAH-AI framework. At the time, these models offered the best balance of cost-efficiency, reliability, and the ability to handle complex tasks at scale, which were critical for achieving our research objectives.

### Continuous Updates in LLMs

Both Goole and OpenAI continously update their LLMs. This can have both positive and negative implications. On the positive side, these updates often introduce new capabilities, enhance reasoning skills, and improve language understanding. However, these updates can also disrupt previously stable workflows, as certain prompt designs or outputs may change without notice.

In our experience, it was important to stay adaptable and regularly monitor how updates to the LLMs impacted their performance in the OAH-AI pipeline. For instance, we noticed subtle changes in how Gemini handled specific prompts after a model update, requiring us to adjust either our prompt or our workflow accordingly. This dynamic nature of LLMs highlights the need for ongoing evaluation and prompt adjustment in any system that relies on these models, as well as careful prompt enginreeing.

### Experimenting with close-source LLMs

When we first began experimenting with LLMs, GPT-3.5-turbo was OpenAI’s flagship model (*OpenAI Platform*, n.d.), while Google offered Gemini-1.5-pro (*Gemini*, 2024). Both companies provided free API access to these models, but the free tiers came with significant limitations, such as restrictions on the number of calls per minute, daily call limits, and token usage caps. These constraints made it extremely difficult to develop and debug our framework effectively.

Even after narrowing the PoC to 1,000 papers, we were still dealing with thousands of LLM calls, which serve as the core of the framework's reasoning and data extraction processes. Under the free-tier limitations, this volume of calls was impractical.

To address this, we switched to paid subscriptions for both LLMs, which resolved the issue and allowed us to proceed with development and experimentation. However, to keep costs as low as possible, we used LLM calls judiciously, knowing that we were charged for both the number of calls and the token counts per request. This became a key design consideration during the development process. We also tested cheaper models where possible. OpenAI’s results were generally satisfactory, but Gemini struggled with basic reasoning tasks, as shown in Appendix 5.

Over the following weeks, we observed noticeable performance improvements in both models as Google and OpenAI updated them internally. New, more powerful models were released, such as GPT-4o and GPT-4o-mini (which replaced GPT-3.5-turbo) from OpenAI, and Gemini-1.5-Flash from Google. However, these newer models came with higher costs.

The competitive landscape worked in our favor. In August 2024, OpenAI halved the price of GPT-4o’s API calls for input tokens and reduced the cost for output tokens by 33%, with even steeper discounts for asynchronous bulk processing. Not to be outdone, Google cut the price of API calls to Gemini-1.5 Flash by 75%. (*LLM Price War, Black Forest’s Open Image Generator, The High Cost of AI Leadership, Machine Translation Goes Agentic - Gmail*, n.d.)

These price reductions allowed us to use more powerful models at a significantly lower cost, leading to a substantial improvement in the quality of the LLM responses. A detailed breakdown of the total costs for LLM usage in the PoC is available in Appendix 6.

### Prompt Design - Lessons Learned

Throughout our experiments, we learned several important lessons regarding prompt design:

* Keep prompts short and simple: Although LLMs are capable of handling complex tasks, we found that simpler, more focused prompts often resulted in clearer and more accurate outputs.
* Ask for structured replies: By explicitly requesting structured answers (e.g., json format, bracketed text), we were able to extract more useful, actionable insights from the LLMs. This was especially helpful for tasks such as simple inference or basic reasoning from text.
* Small building blocks are more resilient: Instead of designing large, complex prompts, breaking down tasks into smaller, incremental steps often produced more robust results. This modular approach allowed us to build on earlier results and adapt more easily to changes in model behavior or updates.

### OpenAI vs. Gemini: Comparative Evaluation

This section provides a detailed comparison between OpenAI and Gemini across several performance dimensions relevant to the OneAquaHealth (OAH) project. Each model was assessed based on its consistency, inference capabilities, reasoning, cost, creativity, and efficiency.

#### **1. Consistency and Reproducibility**

Both models consistently identified key themes such as the impact of urbanization on aquatic ecosystems, the role of water quality as a health indicator, and the importance of biodiversity metrics. However, while both models provided valuable insights, their reproducibility differed:

* **OpenAI**: Showed greater consistency across multiple iterations. It was less influenced by changes in the prompt, which allowed for more reproducible outputs. OpenAI's reliance on large datasets and probabilistic reasoning helped it to provide answers that were reliably in line with scientific evidence, making it more adaptable to various contexts.
* **Gemini**: Although also consistent, Gemini’s outputs were shorter, more rigid and factual, sometimes lacking the flexibility needed for creative or exploratory tasks. Its outputs were slightly more prone to variation in response to changes in prompt context, which made its performance more dependent on input factors. This led to outputs that, while precise, were sometimes less adaptable in scenarios requiring broader exploration.

Both models mirrored the non-reproducibility often observed in human expert evaluations due to various human biases, but OpenAI’s consistency was notably stronger.

#### **2. Inference Capabilities**

Inference tasks, including answering complex queries and extracting relevant information from large datasets, were critical tests for both models. In this regard:

* **OpenAI**: Excelled in handling inference-based tasks, demonstrating the ability to perform sophisticated Q&A and extract relevant, nuanced information. It was capable of distilling complex relationships between variables, showing an aptitude for answering multifaceted questions while staying focused on the research context.
* **Gemini**: Also performed well in information extraction but had a tendency to provide shorter, more direct answers. While this brevity ensured clarity, it sometimes lacked the depth or nuance observed in OpenAI’s responses, making Gemini less effective in more demanding tasks that required comprehensive analysis.

Both models were strong in inference, but OpenAI was superior in maintaining depth and breadth when extracting complex insights.

#### **3. Reasoning Ability**

Reasoning, particularly in tasks requiring logical analysis or the synthesis of new hypotheses, highlighted the differences between the two models:

* **OpenAI**: Displayed robust reasoning skills, excelling in both trivial and complex tasks. It was able to connect disparate concepts, generate hypotheses, and offer nuanced explanations, making it particularly strong in tasks that required deeper cognitive processing.
* **Gemini**: Showed weaker reasoning, especially in more trivial and basic tasks. While it performed well in factual recall and structured tasks, in trivial tasks, Gemini’s reasoning was sometimes completely wrong.

OpenAI consistently outperformed Gemini in reasoning tasks, particularly in generating creative hypotheses and handling complex problem-solving.

#### **4. Cost**

As of the most recent comparison on September 24, 2023, both OpenAI’s GPT-4o-mini and Google’s Gemini-1.5-flash were both cost effective, and nearly identical in terms of operational cost, as shown in Appendix 6. This similarity in cost makes them equally viable from a budgetary perspective, allowing for selection based on performance needs rather than cost constraints.

#### **5. Creativity**

Creativity, defined by the ability to generate novel insights, ideas, or hypotheses, was a critical factor for this research:

* **OpenAI**: Demonstrated a high degree of creativity, particularly in hypothesis generation and exploring new research directions. OpenAI was often able to identify non-obvious connections between variables and suggest alternative interpretations of the data, which is valuable for projects that require innovative thinking.
* **Gemini**: While structured and precise, Gemini’s outputs were less creative. It focused on established patterns and known relationships, which made it reliable but less suited for tasks requiring imaginative or speculative thinking. In generative tasks, Gemini was more conservative, rarely venturing beyond the confines of the data.

In tasks requiring creative exploration, OpenAI significantly outperformed Gemini, providing richer and more innovative outputs.

#### **6. Efficiency and Scalability**

The OAH-AI framework exhibited strong efficiency and scalability when using both OpenAI and Gemini models. The framework was able to process the selected corpus efficiently, completing multiple iterations in under 24 hours, despite the complexity of cross-paper analysis. This highlights the scalability of the system, making it suitable for future applications involving larger datasets.

* **OpenAI**: While OpenAI took approximately 30% longer per LLM call on average, it consistently delivered more nuanced and detailed responses. This added processing time was justified by the depth and breadth of the insights generated, particularly in tasks requiring complex reasoning or creative exploration. OpenAI’s ability to provide rich, detailed outputs made it ideal for tasks where comprehensive analysis is essential, even though it required slightly more processing time. As the OAH framework is not designed to be a real time application, this did not present a real disadvantage.
* **Gemini**: Demonstrated higher efficiency in terms of processing speed, especially in tasks that involved structured analysis. Its responses were clear, concise, and to the point, making it faster in synthesizing and summarizing information. Although Gemini's outputs were less creative and exploratory compared to OpenAI, its efficiency in handling large datasets and returning quick results made it highly effective for straightforward analytical tasks where rapid turnaround is prioritized.

Both models demonstrated excellent scalability, but OpenAI’s ability to handle creative tasks while maintaining efficiency gave it a slight edge in terms of adaptability for broader research contexts.

#### **Conclusion**

When comparing OpenAI and Gemini, both models demonstrated significant value to the OAH project, each excelling in different areas. OpenAI's strengths lay in its **consistency**, **reasoning ability**, and **creativity**, making it the superior choice for tasks requiring deep cognitive processing and exploratory analysis. Its ability to generate nuanced insights and creative hypotheses made it particularly effective for complex problem-solving.

In contrast, **Gemini** excelled in more structured, straightforward tasks. It provided **precise and reliable outputs**, delivering results efficiently and cost-effectively. However, it struggled with more advanced reasoning and creative tasks, where its performance was comparatively weaker than OpenAI.

Given the continuous updates and improvements to both LLMs, this comparison is subject to change over time, as demonstrated in previous evaluations throughout our work on this project. Therefore, leveraging multiple LLMs for **cross-verification** ensures reliability and comprehensive analysis across various tasks.

## Perspectives (Future work)

The OAH-AI framework was initially developed as a PoC, with an emphasis on minimizing costs. To transform this framework into a more useful, accessible, and sustainable tool, we recommend the following improvements:

* Migrate the application from Google Colab to a leading cloud service provider for enhanced scalability, reliability, and performance.
* Adopt open-source LLMs by downloading and integrating both top general-purpose and domain-specific models. This approach would eliminate reliance on external API calls, ensuring more control over updates and performance consistency.
* Implement a nonvolatile database for storing and retrieving paper information, rather than relying solely on GoogleDrive storage. This will improve data management and accessibility for future use.
* Enhance voting strategies by incorporating techniques such as multiple runs of LLMs, an additional LLM reviewer or critic as a conflict resolver, multi-LLM debate systems, or other mechanisms to ensure consistency and reliability throughout the OAH-AI pipeline.
* Integrate multilingual analyses to broaden the framework's applicability across non-English scientific literature, thereby capturing more diverse and global insights.
* Develop a multidomain knowledge graph or ontology from the extracted scientific discoveries. This could be achieved by expanding the existing ODH ontology (ODH-UT), facilitating better integration and understanding of interconnected domains.
* Create a user-friendly GUI to make the tool more accessible to researchers and other stakeholders without technical expertise.
* Expand the framework to additional domains, leveraging its capabilities to explore new areas beyond OAH domains.

## Ethics and Sustainability

A critical aspect of this research is the consideration of ethics and safety in AI applications. As LLMs become more powerful, ensuring their responsible and ethical use is paramount, particularly in sensitive fields like healthcare and environmental monitoring. Issues such as bias, data integrity, and the potential misuse of AI-generated insights must be carefully addressed. In this context, a human-in-the-loop approach remains essential to verify the outputs of LLMs, ensuring that decisions are informed by accurate and unbiased information.

Moreover, the environmental impact of AI systems cannot be overlooked. LLMs, while facilitating groundbreaking research, consume significant energy and computational resources, contributing to their ecological footprint. To address this, we advocate for the integration of green informatics principles (Benis & Tamburis, 2023) . The goal is to foster research practices that contribute to science and humanity without harming the environment. This means using LLMs strategically and efficiently, reducing unnecessary computational load, and prioritizing sustainable AI practices wherever possible.

## Conclusion

This research has demonstrated the successful integration of LLMs within the OAH-AI framework to automate the elicitation and discovery of human health and ecological indicators for monitoring urban freshwater ecosystems. The goals and objectives outlined in the introduction have been met through the development of an AI-powered platform that processes a large corpus of scientific papers, identifies relevant indicators, and generates meaningful insights. By leveraging the capabilities of NLP-driven models, we have shown that AI can support and accelerate the systematic literature review process, revealing previously undetected connections between human, animal, and environmental health.

Through the screening, eligibility, and full-text assessment phases, the LLMs demonstrated their ability to perform tasks with remarkable accuracy and speed. What would normally take human researchers weeks or months was completed in under 24 hours, showcasing the efficiency and scalability of the OAH-AI framework. The results were rigorously evaluated by domain experts, who confirmed that the LLMs produced concise, well-phrased, and contextually accurate outputs, further validating the effectiveness of this approach.

Beyond the specified goals of this research, we have also explored the evolution of LLMs throughout the development process. From using GPT-3.5-turbo to the latest advancements like GPT-4o and Gemini-1.5 Flash, we have witnessed rapid improvements in performance, reasoning, and data extraction capabilities. We used LLMs to write and debug code, to research and brainstorm in the design process, to review and refine sections in this report. In all those tasks we saw the same rapid improvements.

Notably, in September 2024, OpenAI’s release of the o1 model marked a significant leap forward (*Learning to Reason with LLMs*, 2024), surpassing all other LLMs in math, science, and coding benchmarks, bringing us closer to Artificial General Intelligence (AGI) than ever before.

While these advancements are promising, they also highlight the fast-paced evolution of AI technologies, where models and approaches can quickly become outdated.

Nevertheless, the trend is moving towards more authentic, human-like communication and advanced NLP capabilities, which will continue to reshape the landscape of AI-driven research and scientific discovery.

The outputs produced by the OAH-AI framework higlight the efficiency and power of the latest LLMs for scientific discovery. It demonstrates the potential for AI to revolutionize not only literature reviews but also interdisciplinary research across complex domains. By effectively automating data extraction and hypothesis generation, LLMs can play a key role in accelerating scientific discovery, enabling quicker responses to pressing environmental and health challenges.

Looking ahead, as AI models continue to evolve and improve, they will become increasingly integral to research processes. However, it is crucial to balance this technological progress with ethical considerations and sustainable practices. The future of AI in research lies not only in its ability to amplify human capabilities but also in ensuring that these capabilities are used responsibly and in harmony with the broader goals of ODH.

# References

*A Comprehensive Comparative Analysis of LLMs*. (2024, March 11). https://mindsdb.com/blog//navigating-the-llm-landscape-a-comparative-analysis-of-leading-large-language-models

Adam, G., Rampášek, L., Safikhani, Z., Smirnov, P., Haibe-Kains, B., & Goldenberg, A. (2020). Machine learning approaches to drug response prediction: Challenges and recent progress. *Npj Precision Oncology*, *4*(1), 1–10. https://doi.org/10.1038/s41698-020-0122-1

AI4Science, M. R., & Quantum, M. A. (2023). *The Impact of Large Language Models on Scientific Discovery: A Preliminary Study using GPT-4* (arXiv:2311.07361). arXiv. http://arxiv.org/abs/2311.07361

Barrett, C., Boyd, B., Burzstein, E., Carlini, N., Chen, B., Choi, J., Chowdhury, A. R., Christodorescu, M., Datta, A., Feizi, S., Fisher, K., Hashimoto, T., Hendrycks, D., Jha, S., Kang, D., Kerschbaum, F., Mitchell, E., Mitchell, J., Ramzan, Z., … Yang, D. (2023). *Identifying and Mitigating the Security Risks of Generative AI* (arXiv:2308.14840). arXiv. http://arxiv.org/abs/2308.14840

Benis, A., & Tamburis, O. (2023). The Need for Green and Responsible Medical Informatics and Digital Health: Looking Forward with One Digital Health. *Yearbook of Medical Informatics*. https://doi.org/10.1055/s-0043-1768717

Benis, A., Tamburis, O., Chronaki, C., & Moen, A. (2021). One Digital Health: A Unified Framework for Future Health Ecosystems. *Journal of Medical Internet Research*, *23*, 22189. https://doi.org/10.2196/22189

Bornmann, L., Haunschild, R., & Mutz, R. (2021). Growth rates of modern science: A latent piecewise growth curve approach to model publication numbers from established and new literature databases. *Humanities and Social Sciences Communications*, *8*(1), 1–15. https://doi.org/10.1057/s41599-021-00903-w

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T., & Zhang, Y. (2023). *Sparks of Artificial General Intelligence: Early experiments with GPT-4* (arXiv:2303.12712). arXiv. https://doi.org/10.48550/arXiv.2303.12712

*ChatGPT*. (n.d.). Retrieved September 17, 2024, from https://chatgpt.com

Chen, S., Kann, B. H., Foote, M. B., Aerts, H. J. W. L., Savova, G. K., Mak, R. H., & Bitterman, D. S. (2023). Use of Artificial Intelligence Chatbots for Cancer Treatment Information. *JAMA Oncology*. https://doi.org/10.1001/jamaoncol.2023.2954

Deiner, M. S., Deiner, N. A., Hristidis, V., McLeod, S. D., Doan, T., Lietman, T. M., & Porco, T. C. (2024). Use of Large Language Models to Assess the Likelihood of Epidemics From the Content of Tweets: Infodemiology Study. *Journal of Medical Internet Research*, *26*(1), e49139. https://doi.org/10.2196/49139

Dewitte, S., Cornelis, J. P., Müller, R., & Munteanu, A. (2021). Artificial Intelligence Revolutionises Weather Forecast, Climate Monitoring and Decadal Prediction. *Remote Sensing*, *13*(16), Article 16. https://doi.org/10.3390/rs13163209

Dias, R., & Torkamani, A. (2019). Artificial intelligence in clinical and genomic diagnostics. *Genome Medicine*, *11*(1), 70. https://doi.org/10.1186/s13073-019-0689-8

Eysenbach, G., & Group, C.-E. (2011). CONSORT-EHEALTH: Improving and Standardizing Evaluation Reports of Web-based and Mobile Health Interventions. *Journal of Medical Internet Research*, *13*(4), e1923. https://doi.org/10.2196/jmir.1923

Farah, L., Murris, J. M., Borget, I., Guilloux, A., Martelli, N. M., & Katsahian, S. I. M. (2023). Assessment of Performance, Interpretability, and Explainability in Artificial Intelligence–Based Health Technologies: What Healthcare Stakeholders Need to Know. *Mayo Clinic Proceedings: Digital Health*, *1*(2), 120–138. https://doi.org/10.1016/j.mcpdig.2023.02.004

Fawcett, T., & Provost, F. (2013). *Data Science for Business*.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, *17*(3), Article 3. https://doi.org/10.1609/aimag.v17i3.1230

Feng, Y., Liang, S., Zhang, Y., Chen, S., Wang, Q., Huang, T., Sun, F., Liu, X., Zhu, H., & Pan, H. (2022). Automated medical literature screening using artificial intelligence: A systematic review and meta-analysis. *Journal of the American Medical Informatics Association : JAMIA*, *29*(8), 1425–1432. https://doi.org/10.1093/jamia/ocac066

Gao, S., Fang, A., Huang, Y., Giunchiglia, V., Noori, A., Schwarz, J. R., Ektefaie, Y., Kondic, J., & Zitnik, M. (2024). *Empowering Biomedical Discovery with AI Agents* (arXiv:2404.02831). arXiv. https://doi.org/10.48550/arXiv.2404.02831

*Gemini*. (2024, September 18). Google DeepMind. https://deepmind.google/technologies/gemini/

*‎Gemini – chat to supercharge your ideas*. (n.d.). Gemini. Retrieved September 17, 2024, from https://gemini.google.com

*Home Feed | ResearchGate*. (n.d.). Retrieved September 17, 2024, from https://www.researchgate.net/

*Home—ONEAQUAHEALTH*. (n.d.). Retrieved September 17, 2024, from https://www.oneaquahealth.eu/

Jin, Q., Leaman, R., & Lu, Z. (n.d.). *PubMed and Beyond: Recent Advances and Best Practices in Biomedical Literature Search*.

Johnson, K. B., Wei, W.-Q., Weeraratne, D., Frisse, M. E., Misulis, K., Rhee, K., Zhao, J., & Snowdon, J. L. (2021). Precision Medicine, AI, and the Future of Personalized Health Care. *Clinical and Translational Science*, *14*(1), 86–93. https://doi.org/10.1111/cts.12884

Langley, P. (n.d.). *Integrated Systems for Computational Scientific Discovery*.

*Learning to Reason with LLMs*. (n.d.). Retrieved September 20, 2024, from https://openai.com/index/learning-to-reason-with-llms/

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.

Li, H., Moon, J. T., Purkayastha, S., Celi, L. A., Trivedi, H., & Gichoya, J. W. (2023). Ethics of large language models in medicine and medical research. *The Lancet Digital Health*, *5*(6), e333–e335. https://doi.org/10.1016/S2589-7500(23)00083-3

*Llama 3.1*. (n.d.). Meta Llama. Retrieved September 17, 2024, from https://www.llama.com/

*LLM Price War, Black Forest’s Open Image Generator, The High Cost of AI Leadership, Machine Translation Goes Agentic—Helit.bauberg@gmail.com—Gmail*. (n.d.). Retrieved September 20, 2024, from https://mail.google.com/mail/u/0/?fs=1&source=atom#search/thebatc/FMfcgzQVzFSJslWXdXkNkxjJFvgwzqrR

*MEDLINE/PubMed Data Element (Field) Descriptions*. (n.d.). [Technical Documentation]. U.S. National Library of Medicine. Retrieved September 17, 2024, from https://www.nlm.nih.gov/bsd/mms/medlineelements.html#mh

Meskó, B., Drobni, Z., Bényei, É., Gergely, B., & Győrffy, Z. (2017). Digital health is a cultural transformation of traditional healthcare. *mHealth*, *3*, 38. https://doi.org/10.21037/mhealth.2017.08.07

Neves, M., Klippert, A., Knöspel, F., Rudeck, J., Stolz, A., Ban, Z., Becker, M., Diederich, K., Grune, B., Kahnau, P., Ohnesorge, N., Pucher, J., Schönfelder, G., Bert, B., & Butzke, D. (2023). Automatic classification of experimental models in biomedical literature to support searching for alternative methods to animal experiments. *Journal of Biomedical Semantics*, *14*(1), 13. https://doi.org/10.1186/s13326-023-00292-w

*One Health*. (n.d.). Retrieved September 17, 2024, from https://www.who.int/europe/initiatives/one-health

*OpenAI Platform*. (n.d.). Retrieved September 20, 2024, from https://platform.openai.com

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, n71. https://doi.org/10.1136/bmj.n71

*PubMed*. (n.d.). Retrieved September 17, 2024, from https://pubmed.ncbi.nlm.nih.gov/

Reddy, S. (2023). Evaluating large language models for use in healthcare: A framework for translational value assessment. *Informatics in Medicine Unlocked*, *41*, 101304. https://doi.org/10.1016/j.imu.2023.101304

Russell, S. J., & Norvig, P. (with Chang, M., Devlin, J., Dragan, A., Forsyth, D., Goodfellow, I., Malik, J., Mansinghka, V., Pearl, J., & Wooldridge, M. J.). (2021). *Artificial intelligence: A modern approach* (Fourth Edition). Pearson.

Tamburis, O., & Benis, A. (2022). One Digital Health for more FAIRness. *Methods of Information in Medicine*, *61*(S 02), e116–e124. https://doi.org/10.1055/a-1938-0533

Wang, C., Liu, S., Yang, H., Guo, J., Wu, Y., & Liu, J. (2023). Ethical Considerations of Using ChatGPT in Health Care. *Journal of Medical Internet Research*, *25*(1), e48009. https://doi.org/10.2196/48009

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18

*Writer/Palmyra-Med-70B-32K · Hugging Face*. (2024, August 17). https://huggingface.co/Writer/Palmyra-Med-70B-32K

Xu, Y., Liu, X., Cao, X., Huang, C., Liu, E., Qian, S., Liu, X., Wu, Y., Dong, F., Qiu, C.-W., Qiu, J., Hua, K., Su, W., Wu, J., Xu, H., Han, Y., Fu, C., Yin, Z., Liu, M., … Zhang, J. (2021). Artificial intelligence: A powerful paradigm for scientific research. *The Innovation*, *2*(4), 100179. https://doi.org/10.1016/j.xinn.2021.100179

# Glossary

**DOI**

Digital Object Identifier, a string of numbers, letters and symbols used to uniquely identify an article or document, and to provide it with a permanent web address (URL).

**Google Drive**

A cloud storage and file synchronization service developed by Google, allowing users to store, share, and collaborate on files and documents online.

**Google Colab**

A free, cloud-based platform for writing and executing Python code, particularly useful for machine learning and data analysis projects.

**LLM call**

A single interaction or request made to a Large Language Model (LLM), such as Gemini or GPT-4, to generate text, answer questions, or perform language-related tasks.

**Zotero**

A free, open-source reference management software that helps researchers collect, organize, cite, and share research materials.

**Natural Language Processing (NLP)**

A branch of artificial intelligence that focuses on enabling computers to understand, interpret, generate, and manipulate human language. NLP combines linguistics, computer science, and machine learning to process and analyze large amounts of natural language data.

**Token**

A unit of text that has been separated from a larger body of text during processing. Tokens can be words, characters, or subwords. Tokenization is the process of breaking down text into these individual units for further analysis.

Input tokens refers to the amount of tokens sent to the LLM. Output tokens are the amount of tokens returned by the LLM as a response.

**Word Embedding**

A technique in NLP where words or phrases are represented as dense vectors of real numbers in a continuous vector space. These representations capture semantic and syntactic relationships between words, allowing machines to better understand the meaning and context of language.

# Appendix 1 – OAH-AI PRISMA Screening prompt

<<<

Your task is to assess the relevance of the following paper to PRISMA review on {term}.

If the paper does not meet the inclusion criteria, reject it.

The record inclusion criteria are as follows:

I1. Must be in English. Reply true or false.

I2. Publication date is after January 2013 and before May 2024. Reply true or false.

I3. Source includes doi, OR PMC information included

I4. Check papers' keywords and MeSH terms before replying. Focus on Human Health Indicators in Urban Freshwater Environments. Reply true or false.

I5. Check papers' keywords and MeSH terms before replying. Focus on Ecological Indicators in Urban Freshwater Environments. Reply true or false.

Reply with a dictionary:

I1: true or false

I2: true or false

I3: true or false

I4: true or false

I5: true or false

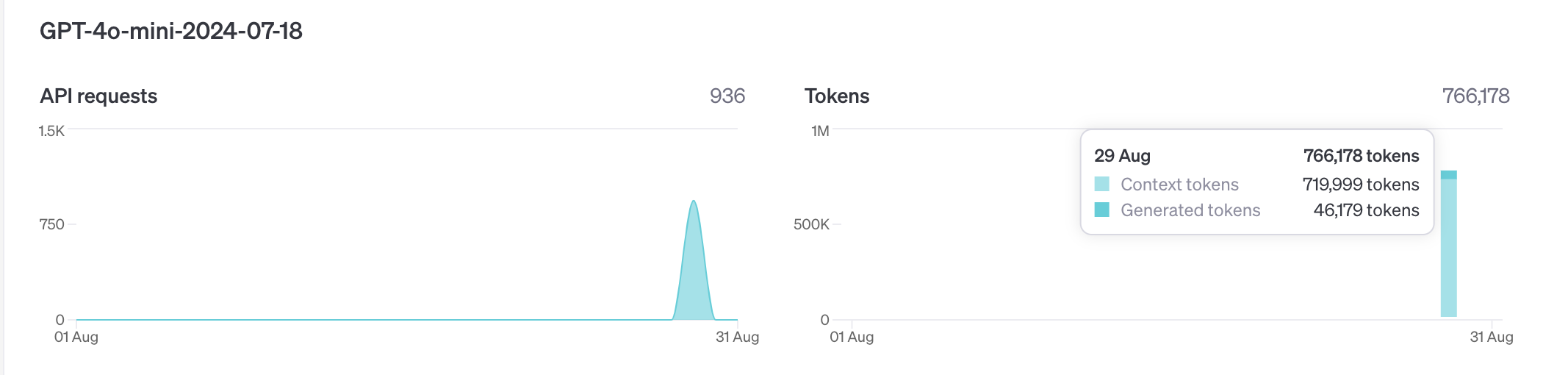
Status: Screened if (I1, I2, I3) is true AND (I4 or I5) is true, or Rejected otherwise.

{paper}

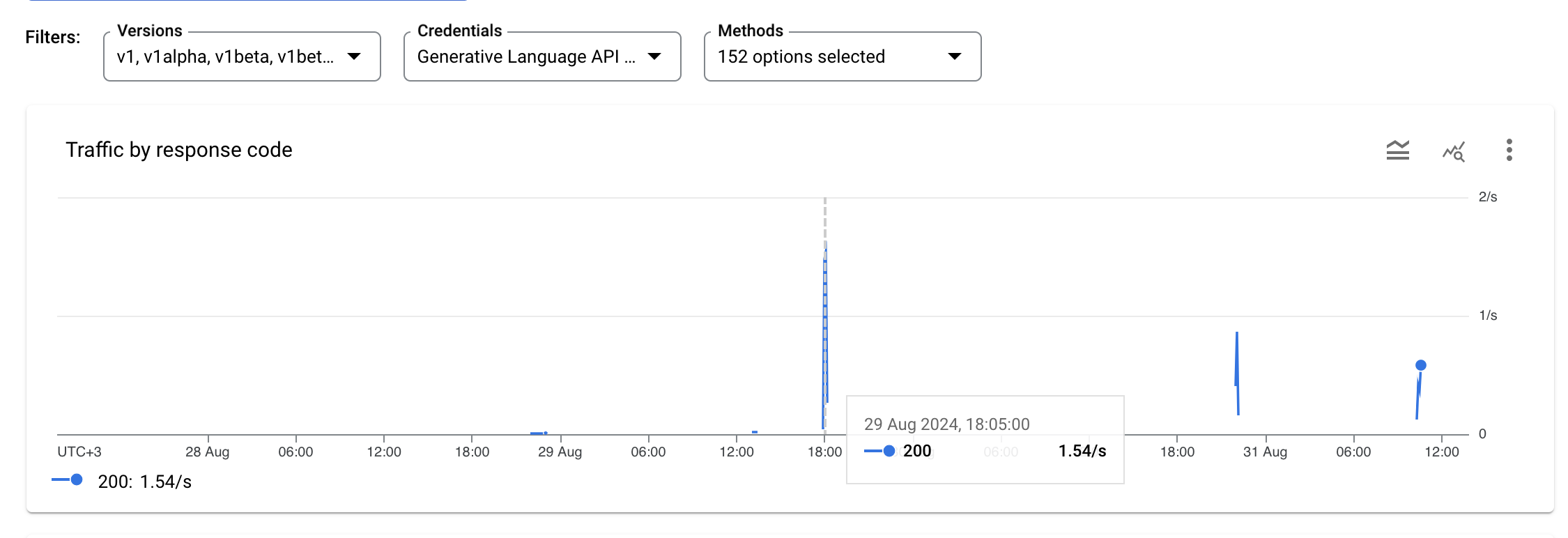
>>>

# Appendix 2 - LLM Usage

**Token count : 237 fixed + {paper token count} (average: 600 tokens)**

****

OpenAI screening 1,000 papers Exp1



Gemini screening 1,000 papers Exp1

# Appendix 3 - LLM query engine selector example

Found 38797424 in papers

hello got llm models/gemini-1.5-flash

created query engines

***Selecting query engine 0: The question asks about the overall focus of the paper, which is related to summarization..***

Q1: No

***Selecting query engine 1: The question asks about a specific topic (environmental health) within the paper, indicating a need for context retrieval..***

Q2: yes

***Selecting query engine 0: The question asks about the overall focus of the paper, which is related to summarization..***

Q3: No

***Selecting query engine 0: The question asks for a summary of the paper's focus, which is directly related to the content of the paper..***

Q4: green

***Selecting query engine 1: The provided text directly describes the One Digital Health (ODH) concept, its framework, and its goals. This makes it highly relevant for retrieving specific context about ODH..***

Q5: The paper focuses on the impact of road salt application on trace element concentrations in soil porewater. While it touches upon environmental factors, it doesn't directly address the broader digital health aspects of One Digital Health.

<<2>>

***Selecting query engine 1: The question asks about the relevance of the paper to OneAquaHealth, a specific research project. This suggests a need to retrieve specific context from the paper, such as its goals, methods, and findings, to determine its relevance to OneAquaHealth..***

Q6: This paper focuses on the impact of de-icing salts on various environmental components, including groundwater, soil, and aquatic ecosystems. It investigates the mobilization of heavy metals, chloride, and other contaminants due to road salt application. This research directly aligns with OneAquaHealth's goal of understanding and mitigating the impact of urban stressors on freshwater ecosystems.

The paper provides valuable insights into the chemical cocktails introduced into urban environments through de-icing practices, which can negatively affect the health of aquatic ecosystems. This information is crucial for developing early warning indicators and enhancing environmental monitoring, as aimed by OneAquaHealth.

<<9>>

2

9

PMID 38797424 was classified under domain: Environment

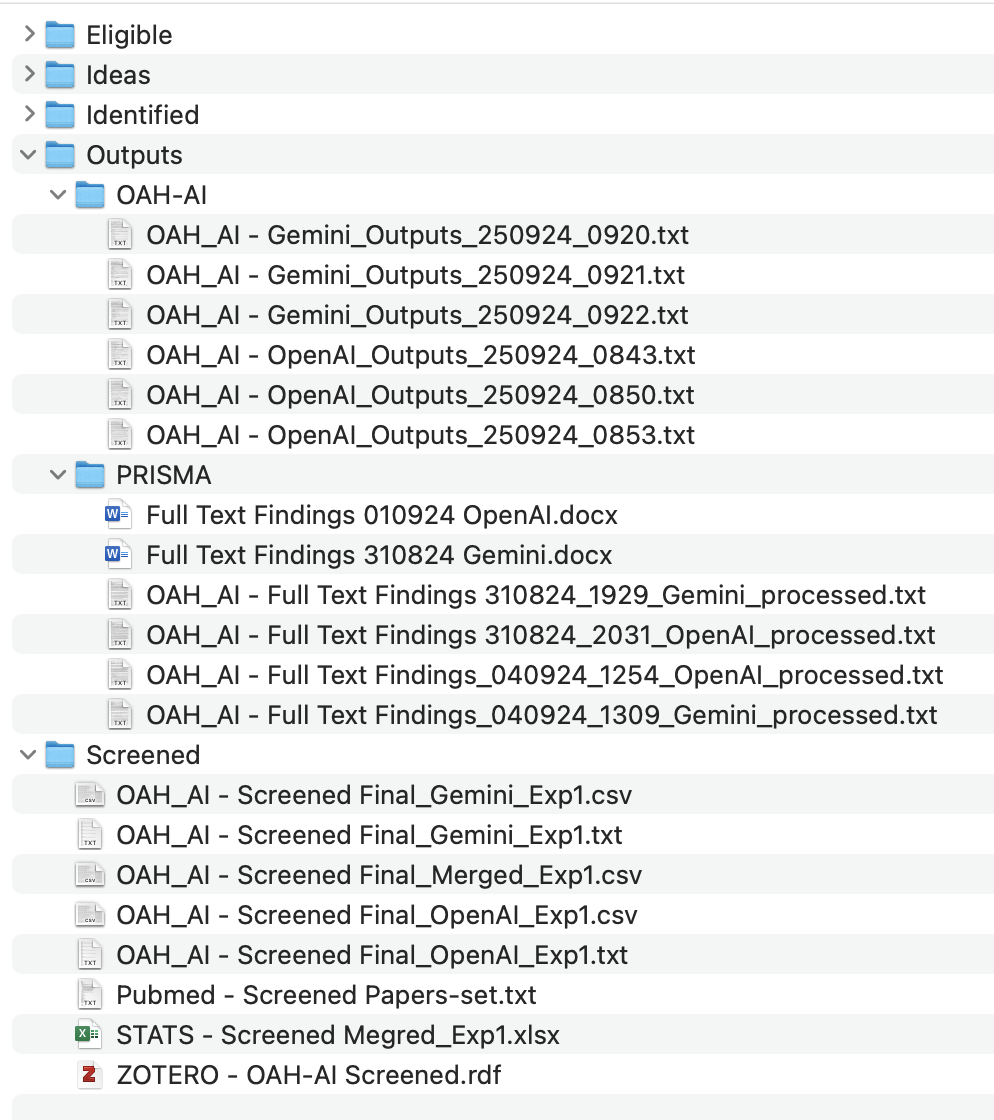
# Appendix 4 - Project and Output files’ location and description

OAH-AI output files are located under OAH-AI/PRISMA/

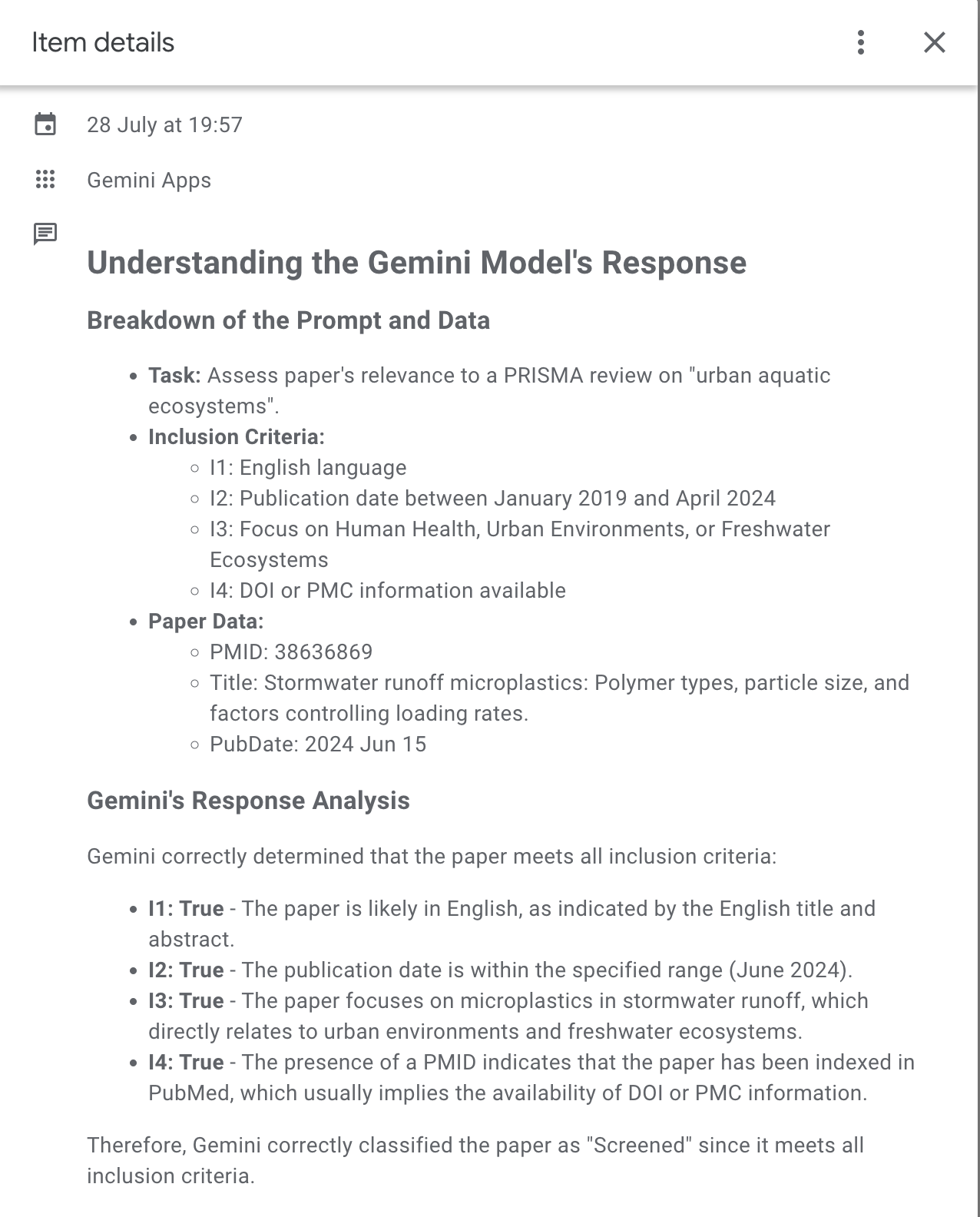
Each step’s output in its designated folder.

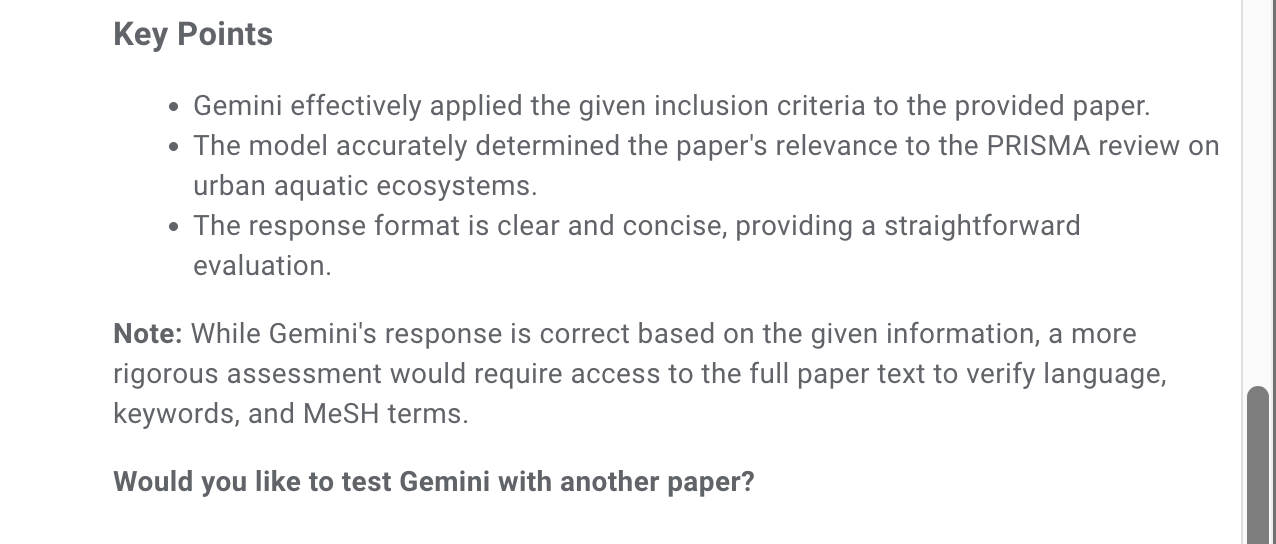
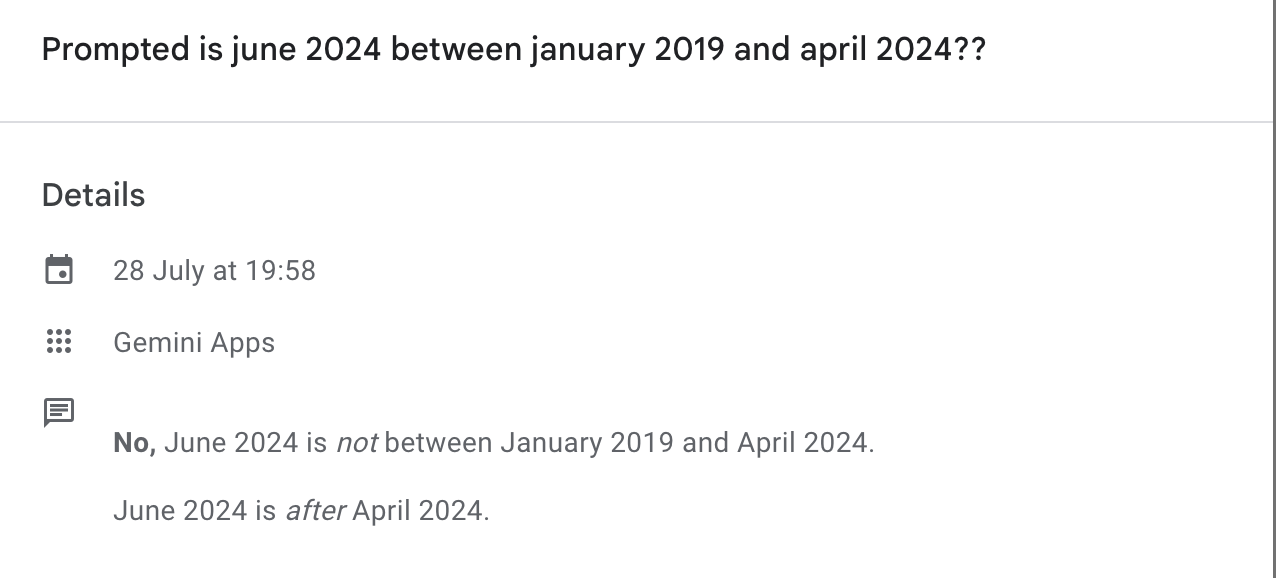
All files that were created by the OAH-AI framework begin with ‘OAH-AI’.

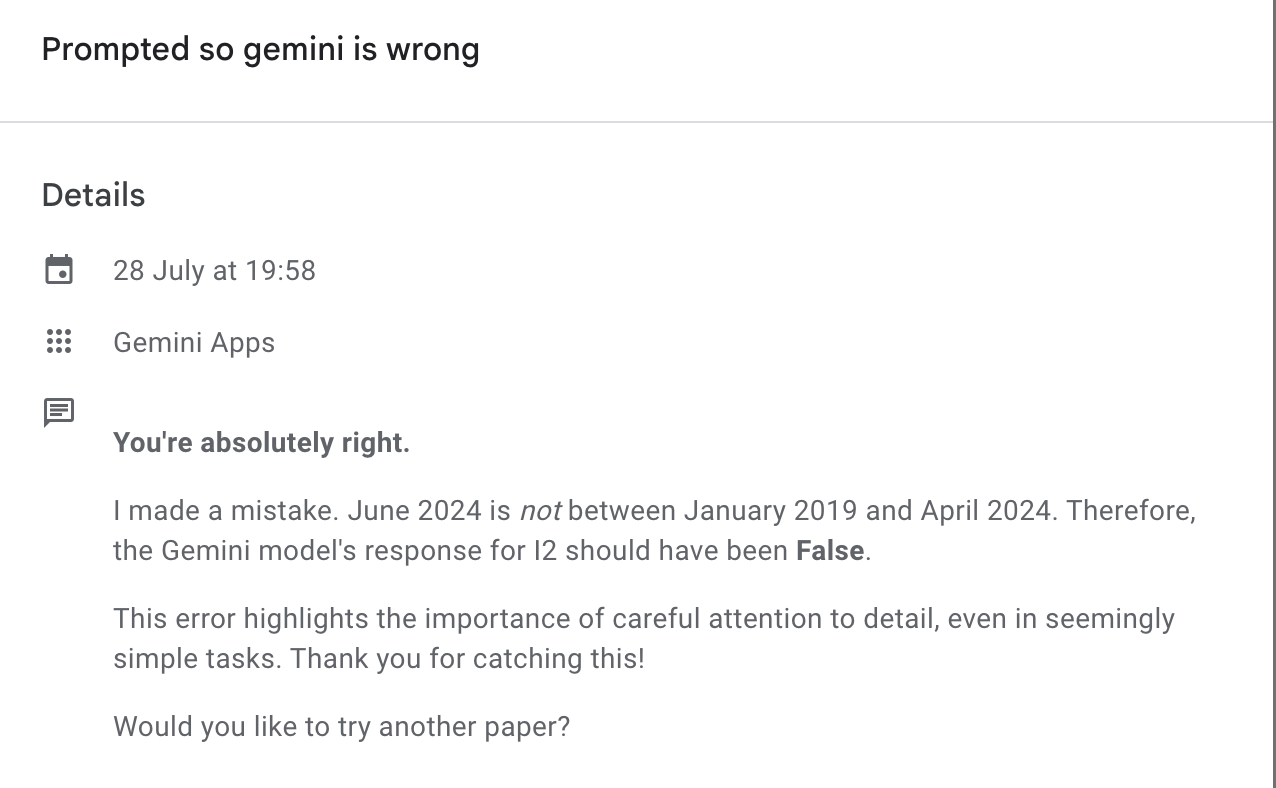
Files that do not have this prefix are supplemental: these contain statistics on the OAH-AI step outputs, or formatted output (Zotero rdf, Pubmed set, xls, docx).



# Appendix 5 - Gemini reasoning failure





# Appendix 6 – LLM calls – PoC costs

## Google

Total Google costs were under 15 Shekel (less than 4 Euros), as shown in the below figure.

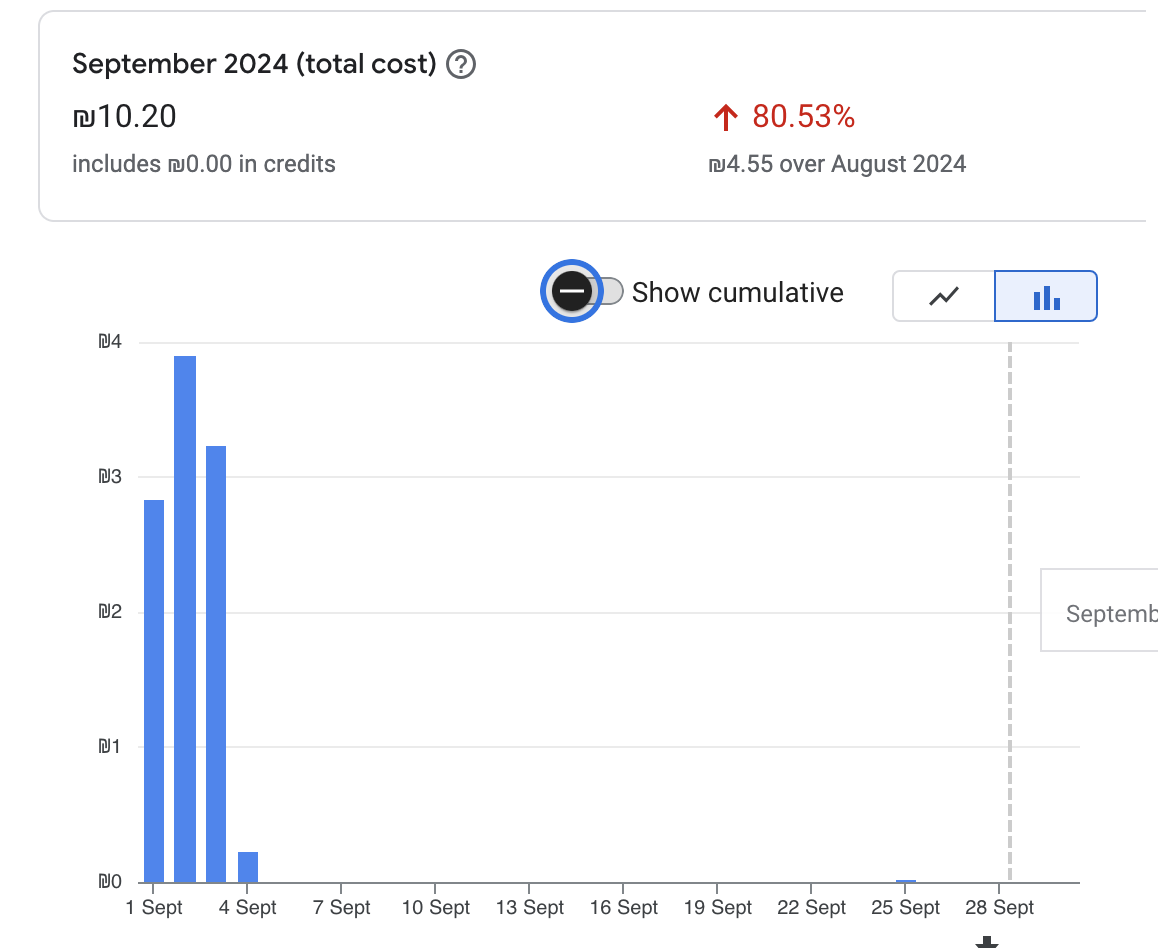


Figure - PoC Costs - Gemini

## OpenAI

Total OpenAI costs for September 2024 PoC experiments was 4.62 USD (about 4.2 Euros).

# Appendix 7 – Papers Selected for OAH-AGI

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| PMID | Title | Source | Domain | OAH Score | ODH Score |
| 25889666 | Impacts of the creation, expansion and management of English wetlands on mosquito presence and abundance - developing strategies for future disease mitigation. | Parasit Vectors. 2015 Mar 3;8:142. doi: 10.1186/s13071-015-0751-3. | Environment | 9 | 3 |
| 26771345 | Are harmful algal blooms becoming the greatest inland water quality threat to public health and aquatic ecosystems? | Environ Toxicol Chem. 2016 Jan;35(1):6-13. doi: 10.1002/etc.3220. | Environment | 9 | 1 |
| 28318787 | Municipal wastewater treatment plant effluent-induced effects on freshwater mussel populations and the role of mussel refugia in recolonizing an extirpated reach. | Environ Pollut. 2017 Jun;225:460-468. doi: 10.1016/j.envpol.2017.03.010. Epub 2017 Mar 18. | Environment | 9 | 2 |
| 29800822 | Distribution and source analysis of heavy metal pollutants in sediments of a rapid developing urban river system. | Chemosphere. 2018 Sep;207:218-228. doi: 10.1016/j.chemosphere.2018.05.090. Epub 2018 May 16. | Environment | 9 | 3 |
| 29996397 | Juvenile coho salmon growth and health in streams across an urbanization gradient. | Sci Total Environ. 2018 Jun 1;625:1003-1012. doi: 10.1016/j.scitotenv.2017.12.327. Epub 2018 Jan 5. | Environment | 10 | 3 |
| 30772551 | Quantitative assessment of the effects of human activities on phytoplankton communities in lakes and reservoirs. | Sci Total Environ. 2019 May 15;665:213-225. doi: 10.1016/j.scitotenv.2019.02.117. Epub 2019 Feb 8. | Environment | 10 | 3 |
| 32470679 | Hospital wastewaters: A reservoir and source of clinically relevant bacteria and antibiotic resistant genes dissemination in urban river under tropical conditions. | Ecotoxicol Environ Saf. 2020 Sep 1;200:110767. doi: 10.1016/j.ecoenv.2020.110767. Epub 2020 May 26. | Environment | 10 | 3 |
| 33302085 | The direct and indirect effects of copper on vector-borne disease dynamics. | Environ Pollut. 2021 Jan 15;269:116213. doi: 10.1016/j.envpol.2020.116213. Epub 2020 Dec 2. | Environment | 9 | 2 |
| 33485000 | The microbiome of a polluted urban lake harbors pathogens with diverse antimicrobial resistance and virulence genes. | Environ Pollut. 2021 Jan 11;273:116488. doi: 10.1016/j.envpol.2021.116488. | Environment | 9 | 3 |
| 33686602 | Effects of urbanisation and a wastewater treatment plant on microplastic densities along a subtropical river system. | Environ Sci Pollut Res Int. 2021 Jul;28(27):36102-36111. doi: 10.1007/s11356-021-13185-1. Epub 2021 Mar 8. | Environment | 9 | 1 |
| 33774562 | A machine learning approach for early warning of cyanobacterial bloom outbreaks in a freshwater reservoir. | J Environ Manage. 2021 Jun 15;288:112415. doi: 10.1016/j.jenvman.2021.112415. Epub 2021 Mar 26. | Environment | 9 | 3 |
| 33991905 | Trophic transfer and dietary exposure risk of mercury in aquatic organisms from urbanized coastal ecosystems. | Chemosphere. 2021 Oct;281:130836. doi: 10.1016/j.chemosphere.2021.130836. Epub 2021 May 10. | Environment | 9 | 2 |
| 34479163 | Aquatic concentration and risk assessment of pharmaceutically active compounds in the environment. | Environ Pollut. 2021 Dec 1;290:118049. doi: 10.1016/j.envpol.2021.118049. Epub 2021 Aug 27. | Environment | 9 | 3 |
| 34955197 | Urbanization reduces resource use efficiency of phytoplankton community by altering the environment and decreasing biodiversity. | J Environ Sci (China). 2022 Feb;112:140-151. doi: 10.1016/j.jes.2021.05.001. Epub 2021 Jun 3. | Environment | 10 | 3 |
| 35099004 | Changes in fish skin microbiota along gradients of eutrophication in human-altered rivers. | FEMS Microbiol Ecol. 2022 Feb 17;98(1):fiac006. doi: 10.1093/femsec/fiac006. | Environment | 9 | 3 |
| 35283080 | First report of perfluoroalkyl acids (PFAAs) in the Indus Drainage System: Occurrence, source and environmental risk. | Environ Res. 2022 Aug;211:113113. doi: 10.1016/j.envres.2022.113113. Epub 2022 Mar 11. | Environment | 9 | 3 |
| 36529169 | Wastewater generation and treatment by various eco-friendly technologies: Possible health hazards and further reuse for environmental safety. | Chemosphere. 2023 Feb;313:137547. doi: 10.1016/j.chemosphere.2022.137547. Epub 2022 Dec 15. | Environment | 9 | 2 |
| 36541332 | COVID-19 Lockdown Pandemic Period Effects in Highly Impacted Aquatic Ecosystems. | Environ Toxicol Chem. 2023 May;42(5):966-977. doi: 10.1002/etc.5551. Epub 2023 Feb 8. | Environment | 10 | 8 |
| 36554433 | Study and Application of Urban Aquatic Ecosystem Health Evaluation Index System in River Network Plain Area. | Int J Environ Res Public Health. 2022 Dec 9;19(24):16545. doi: 10.3390/ijerph192416545. | Environment | 10 | 3 |
| 37454450 | Water quality variation and driving factors quantitatively evaluation of urban lakes during quick socioeconomic development. | J Environ Manage. 2023 Oct 15;344:118615. doi: 10.1016/j.jenvman.2023.118615. Epub 2023 Jul 14. | Environment | 9 | 2 |
| 37558875 | The recovery of European freshwater biodiversity has come to a halt. | Nature. 2023 Aug;620(7974):582-588. doi: 10.1038/s41586-023-06400-1. Epub 2023 Aug 9. | Environment | 9 | 2 |
| 37586536 | How do weather conditions and environmental characteristics influence aesthetic preferences of freshwater environments? | Sci Total Environ. 2023 Dec 10;903:166283. doi: 10.1016/j.scitotenv.2023.166283. Epub 2023 Aug 14. | Human Health | 10 | 3 |
| 37778289 | A One-Health environmental risk assessment of contaminants of emerging concern in London's waterways throughout the SARS-CoV-2 pandemic. | Environ Int. 2023 Oct;180:108210. doi: 10.1016/j.envint.2023.108210. Epub 2023 Sep 14. | Environment | 9 | 3 |
| 37856511 | Littoral sediment arsenic concentrations predict arsenic trophic transfer and human health risk in contaminated lakes. | PLoS One. 2023 Oct 19;18(10):e0293214. doi: 10.1371/journal.pone.0293214. eCollection 2023. | Human Health | 9 | 3 |
| 37899012 | Light pollution of freshwater ecosystems: principles, ecological impacts and remedies. | Philos Trans R Soc Lond B Biol Sci. 2023 Dec 18;378(1892):20220360. doi: 10.1098/rstb.2022.0360. Epub 2023 Oct 30. | Environment | 9 | 2 |
| 37952398 | Algal or bacterial community: Who can be an effective indicator of the impact of reclaimed water recharge in an urban river. | Water Res. 2023 Dec 1;247:120821. doi: 10.1016/j.watres.2023.120821. Epub 2023 Nov 1. | Environment | 9 | 3 |
| 37978200 | Shotgun metagenomes from productive lakes in an urban region of Sweden. | Sci Data. 2023 Nov 17;10(1):810. doi: 10.1038/s41597-023-02722-x. | Environment | 10 | 3 |
| 37979332 | Empirical evidence of climate change and urbanization impacts on warming stream temperatures. | Water Res. 2023 Dec 1;247:120703. doi: 10.1016/j.watres.2023.120703. Epub 2023 Oct 6. | Environment | 10 | 1 |
| 38557568 | Change in water quality in an Amazonian microbasin: ecological and human health implications. | J Water Health. 2024 Mar;22(3):522-535. doi: 10.2166/wh.2024.286. Epub 2024 Feb 29. | Environment | 10 | 0 |
| 38646485 | Citizen science initiatives document biodiversity baselines at an urban lake. | PeerJ. 2024 Apr 17;12:e17209. doi: 10.7717/peerj.17209. eCollection 2024. | Environment | 10 | 2 |
| 38735464 | Severe microplastic pollution risks in urban freshwater system post-landfill fire: A case study from Brahmapuram, India. | Environ Pollut. 2024 Jul 1;352:124132. doi: 10.1016/j.envpol.2024.124132. Epub 2024 May 11. | Environment | 9 | 3 |
| 38820820 | Accumulation of antibiotics in the environment: Have appropriate measures been taken to protect Canadian human and ecological health? | Ecotoxicol Environ Saf. 2024 Jul 15;280:116513. doi: 10.1016/j.ecoenv.2024.116513. Epub 2024 May 30. | Environment | 9 | 3 |