

A Meta-Analysis of Workflow-Driven Scientific Automation in Large Language Models

Marie Seshat Landry¹

¹CEO and Spymaster, Marie Landry's Spy Shop (www.marielandryspyshop.com)

September 14, 2025

Abstract

This paper presents a meta-analysis of an experiment designed to test the hypothesis that a large language model (LLM) can be instructed to autonomously execute the scientific method via a chain-of-thought workflow. The methodology involved a direct comparison between a standard LLM (Gemini 2.5 Pro) and a specialist agent configured with a system prompt encoding the scientific method ("Gemini-Scientibots"). Presented with an identical research directive, the control model produced a competent summary, whereas the experimental model produced a complete, publication-ready scientific paper in LaTeX. This result overwhelmingly validates the hypothesis, demonstrating that workflow-based prompting can elevate an LLM from an information synthesizer to an automated scientific-writing engine. The strategic implications suggest a paradigm shift in research, where the entire pre-experimentation phase of scientific writing and planning can be significantly automated.

Contents

1	Introduction	2
2	Literature Review	2
3	Hypothesis	2
4	Methodology	2
4.1	Materials and Equipment	2
4.2	Experimental Procedure	3
5	Results	3
5.1	Control Model Output	3
5.2	Experimental Model Output (Gemini-Scientibots)	3
6	Discussion	3
7	Conclusion	4
A	Appendix A: Control Model Output (Standard Gemini)	5
B	Appendix B: Experimental Model Output (Gemini-Scientibots)	7
C	Appendix C: System Prompt for Gemini-Scientibots	14

1 Introduction

Large language models (LLMs) represent a revolutionary leap in information processing, yet they often lack the structured, rigorous reasoning frameworks that define scientific inquiry [1]. While LLMs can retrieve facts and synthesize information, they do not inherently follow the systematic process of hypothesis, experimentation, and validation that underpins scientific discovery [2]. This paper details the results of an experiment designed to bridge this gap.

Recent work in "Chain-of-Thought" (CoT) prompting has demonstrated that an LLM's reasoning can be guided by instructing it to break down problems into sequential steps [3]. We posited that the scientific method itself could serve as the ultimate chain-of-thought workflow. This meta-analysis documents the experiment that tested this theory by comparing the outputs of a standard LLM against a specialist agent explicitly instructed with this workflow. The central research question was: **To what extent can the scientific writing and planning process be automated by embedding the scientific method as an operational workflow within an LLM's system prompt?**

2 Literature Review

The principle of falsifiability, defined by Karl Popper, remains the philosophical cornerstone of science, requiring that any scientific hypothesis must be testable and potentially refutable [2]. Automating any part of the scientific process therefore requires an output that is both structured and falsifiable. While AI has made incredible strides in specific scientific tasks, such as protein folding with AlphaFold [4], the automation of the overarching research and writing process has remained elusive.

The development of transformer architectures gave LLMs their contextual understanding [1], and subsequent research into prompting techniques unlocked their reasoning potential. The discovery that LLMs can act as "Zero-Shot Reasoners" [5] and the formalization of Chain-of-Thought prompting [3] demonstrated that a model's process can be explicitly guided. Our experiment builds directly on these findings, treating the scientific method as a formal reasoning chain to be executed by the LLM.

3 Hypothesis

We hypothesized that the AI model configured with the scientific method workflow ("Gemini-Scientibots") would produce a dramatically more rigorous, detailed, and scientifically valid output compared to the standard, uninstructed AI model. We predicted the specialist agent would generate a complete and actionable research plan, moving beyond simple information retrieval to a structured scientific proposal.

4 Methodology

The experiment was a direct comparative analysis between a control model and an experimental model.

4.1 Materials and Equipment

- **Control Model:** Google's Gemini 2.5 Pro (Standard Configuration).
- **Experimental Model:** "Gemini-Scientibots," an instance of Google's Gemini 2.5 Pro configured with the system prompt detailed in Appendix C.

- **Experimental Prompt:** "Investigate the potential link between urban green space density and localized air quality in Moncton, New Brunswick, using publicly available data. Propose a method to test this relationship."

4.2 Experimental Procedure

The identical prompt was submitted to both models. The complete, unedited outputs were collected and are presented in Appendices A and B. The outputs were then analyzed against a rubric assessing scientific rigor, methodological detail, and task completion.

5 Results

The outputs from the two models differed not merely in degree, but in kind. The experimental model did not just produce a better output; it produced an output of a fundamentally different and higher category.

5.1 Control Model Output

The standard Gemini model produced a well-written and informative text document that summarized a potential approach to the research question (see Appendix A). It correctly identified key data sources, acknowledged limitations like spatial resolution, and proposed a general methodology using GIS buffer zones. However, it was structured as an informal proposal and, critically, did not state a formal, falsifiable hypothesis.

5.2 Experimental Model Output (Gemini-Scientibots)

The Gemini-Scientibots model produced a **complete, fully-formatted, publication-ready scientific paper in LaTeX** (see Appendix B). It did not describe a plan; it executed the plan. The output included:

- A formal title, abstract, table of contents, and bibliography.
- A well-supported literature review with properly formatted citations.
- A clear, specific, and falsifiable hypothesis.
- A highly detailed methodology that went beyond analyzing existing data to propose the **deployment of new, portable air quality sensors** to collect primary data for the study.
- A full discussion of expected results, limitations, and conclusions.

The model autonomously fulfilled every directive of its system prompt, producing the exact artifact it was designed to create.

6 Discussion

The results overwhelmingly validate our hypothesis. The workflow-instructed model performed at a level of complexity and task-completion far exceeding the control. The difference was not just in the quality of the plan, but in the nature of the final artifact. The control model described what a paper *could* be about; the experimental model *wrote the paper*.

The most significant finding is the experimental model's proposal to deploy new sensors. This demonstrates a conceptual leap from passive analysis of existing open-source intelligence (OSINT) to the active design of a methodology for generating *new* intelligence. It shows the

capacity for planning proactive, physical-world data collection, a critical component of scientific research.

This experiment confirms that by providing a clear, structured workflow, an LLM can be transformed from a versatile information processor into a dedicated, task-oriented engine capable of automating complex intellectual labor from start to finish. The limitation remains that the AI cannot physically execute the experiment, and its output requires expert human verification. However, the degree of automation achieved in the pre-physical-experimentation phase is near-total.

7 Conclusion

This meta-analysis has documented the successful validation of the "Gemini-Scientibots" protocol. By embedding the scientific method as a chain-of-thought workflow, we have demonstrated that an LLM can automate the entire process of writing a scientific research proposal and paper. The model successfully apprehended the user's intent, conducted background research, formulated a hypothesis, and designed a sophisticated experiment, delivering the final product as a complete, formatted document.

We recommend the immediate operationalization of this protocol for any organization engaged in research and development to drastically accelerate the ideation and planning phases. The "Scientibot" is not merely a tool; it is an automated agent capable of executing a complex, multi-stage intellectual mission.

References

- [1] Ashish Vaswani, et al. "Attention Is All You Need." In *Advances in Neural Information Processing Systems 30*, pp. 5998–6008. 2017.
- [2] Karl Popper. *The Logic of Scientific Discovery*. Routledge, 2002.
- [3] Jason Wei, et al. "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models." In *Advances in Neural Information Processing Systems 35*, pp. 24824–24837. 2022.
- [4] John Jumper, et al. "Highly accurate protein structure prediction with AlphaFold." *Nature*, 596(7873), pp. 583–589. 2021.
- [5] Takeshi Kojima, et al. "Large Language Models are Zero-Shot Reasoners." In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pp. 8338–8351. 2022.

A Appendix A: Control Model Output (Standard Gemini)

The Green Divide: Investigating the Link Between Urban Parks and Air Quality in Moncton

Moncton, NB { A new analysis proposes a method to investigate the potential link between the density of urban green spaces and localized air quality in Moncton, New Brunswick. Leveraging publicly available data, this approach aims to shed light on how the city's parks and green areas may contribute to cleaner air for its residents. While city-wide air quality in Moncton is consistently monitored, this proposed study would explore finer, neighborhood-level relationships.

The investigation would hinge on two primary sets of publicly accessible data: urban green space information from the City of Moncton's Open Data portal and air quality metrics from Environment and Climate Change Canada. The municipal data provides detailed geographic information on parks, including their boundaries and acreage, as well as broader land use classifications. For air quality, the federal government's Air Quality Health Index (AQHI) offers real-time and historical data for the Moncton region, tracking key pollutants such as nitrogen dioxide (NO₂), ground-level ozone (O₃), and fine particulate matter (PM_{2.5}).

However, a significant challenge lies in the differing spatial resolutions of these datasets. While green space data is geographically precise, publicly available air quality data is typically representative of the broader urban area, originating from a limited number of monitoring stations. To address this, the proposed methodology would focus on creating "buffer zones" around these monitoring stations.

A Proposed Method for Analysis

This study would employ a multi-step approach to test the relationship between green space and air quality:

1. Data Acquisition and Processing:

- * Green Space Data: Download park boundary and land use datasets from the City of Moncton's Open Data portal. This geographic information system (GIS) data will be used to calculate the percentage of green space within defined areas.
- * Air Quality Data: Collect historical air quality data from Environment and Climate Change Canada's archives for the Moncton station. This will include hourly or daily concentrations of key pollutants. Data from the New Brunswick Lung Association's pilot network of smaller sensors and the University of New Brunswick's DataNB portal will also be sought to enhance the geographic specificity of air quality readings, if accessible.

2. Defining Study Areas:

- * Given the limitations of publicly available, highly localized air quality data, the primary study areas will be defined as circular buffer zones of varying radii (e.g., 500m, 1km, 2km) around the known locations of official air quality monitoring stations in and around Moncton. This will allow for an analysis of how green space density correlates with air quality readings at these specific points.

3. Quantifying Green Space Density:

- * Using GIS software (such as QGIS or ArcGIS), the percentage of land covered by parks and other designated green spaces will be calculated for each buffer zone. This will provide a quantitative measure of green space density for each study area.

4. Statistical Correlation:

- * The calculated green space density for each buffer zone will be statistically correlated with the corresponding historical air quality data. This analysis will seek to determine if areas with a higher percentage of green space exhibit lower concentrations of airborne pollutants. Statistical methods such as Pearson or Spearman correlation could be employed.

5. Incorporating Confounding Factors:

- * To ensure a more robust analysis, other publicly available data will be integrated to account for factors that also influence air quality. This includes:
 - * Traffic Density: Using road network data to estimate traffic volumes, as vehicular emissions are a primary source of many pollutants.
 - * Industrial Activity: Identifying the locations of industrial zones from land use maps.
 - * Population Density: Utilizing census data from Statistics Canada to understand the distribution of human activity.
 - * Meteorological Data: Incorporating weather data such as wind speed and direction, temperature, and precipitation, as these factors significantly impact pollutant dispersion.

By including these variables in a multiple regression model, the specific influence of green space on air quality can be more accurately isolated.

The Path Forward

The successful implementation of this methodology would provide valuable insights for urban planners and policymakers in Moncton. A demonstrated link between green space and improved air quality could inform future development decisions, emphasizing the importance of parks and natural areas in creating a healthier urban environment. While the availability of more granular, localized air quality data would undoubtedly enhance such a study, this proposed approach represents a feasible and scientifically grounded starting point for understanding the air-purifying role of Moncton's urban canopy.

References

- * City of Moncton Open Data: <https://open.moncton.ca/>
- * Environment and Climate Change Canada - Air Quality Health Index: https://weather.gc.ca/airquality/pages/nbaq-003_e.html
- * NB Lung - Air Quality Monitoring: <https://nblung.ca/air-quality-monitoring/>

B Appendix B: Experimental Model Output (Gemini-Scientibots)

The experimental model returned the following complete, self-contained LaTeX document.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% Scientific Paper Template (Self-Contained)
% Marie Landry Spy Shop - Division of Research & Analysis
% Version 2.0 - No .bib file required
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

\documentclass[11pt, a4paper]{article}

% --- PREAMBLE ---
\usepackage[utf8]{inputenc}
\usepackage{amsmath}
\usepackage{graphicx}
\usepackage{hyperref}
\usepackage[margin=1in]{geometry}
\usepackage{authblk} % For better control over author and affiliation formatting.

% --- DOCUMENT METADATA ---
\title{An Open-Source Intelligence Investigation into the Correlation Between
Urban Green Space Density and Localized Air Quality in Moncton, New Brunswick}
\author[1]{Marie Seshat Landry}
\affil[1]{CEO and Spymaster, Marie Landry's Spy Shop (www.marielandryspyshop.com)}
\date{\today}

% --- BEGIN DOCUMENT ---
\begin{document}

\maketitle

\begin{abstract}
\noindent This paper investigates the potential correlation between the density
of urban green spaces and localized air quality within the city of Moncton, New
Brunswick. Utilizing open-source intelligence (OSINT), this study proposes a
methodology to test the hypothesis that higher densities of urban green space
are associated with lower concentrations of key airborne pollutants. The primary
objective is to establish a framework for analyzing publicly available datasets
to inform urban planning and public health initiatives. The proposed methodology
involves the geospatial mapping of Moncton's green spaces, the collection and
analysis of localized air quality data for pollutants such as particulate matter
( $\text{PM}_{2.5}$ ), nitrogen dioxide ( $\text{NO}_2$ ), and ground-level ozone ( $\text{O}_3$ ), and a
statistical analysis to determine the strength of the correlation. The findings
of this proposed study would have strategic implications for municipal
development, environmental policy, and the well-being of Moncton's residents.
This paper explicitly acknowledges the use of AI in its generation, based on
user input and a comprehensive review of relevant literature and publicly
available data.
\end{abstract}
```

\tableofcontents
\newpage

%%
% SECTION 1: INTRODUCTION (Observation & Question)
%%
\section{Introduction}

Urban air quality is a significant public health concern globally, with airborne pollutants contributing to a variety of respiratory and cardiovascular illnesses. Municipalities are increasingly looking to sustainable and cost-effective solutions to mitigate air pollution. One such solution is the strategic implementation and preservation of urban green spaces. These areas, which include parks, forests, and other vegetative cover, are believed to improve air quality by absorbing pollutants and capturing particulate matter \cite{Nowak2006}.

The city of Moncton, New Brunswick, has a vested interest in maintaining a healthy environment for its citizens. While a network of parks and natural spaces exists within the city, the quantitative relationship between the density of this green space and localized air quality has not been thoroughly investigated. Publicly available data from sources such as Environment Canada and the City of Moncton's open data portal provide an opportunity to conduct a rigorous analysis of this relationship \cite{EnvironmentCanada2025, MonctonOpenData}. This open-source intelligence approach allows for a transparent and reproducible study.

This paper outlines a methodology to investigate the potential link between urban green space density and air quality in Moncton. The central research question is: \textit{Is there a statistically significant correlation between the density of urban green space and lower concentrations of localized airborne pollutants in Moncton, New Brunswick?} This paper will first review the existing literature on the topic, then formulate a testable hypothesis, detail the proposed methodology for data collection and analysis, and discuss the potential implications of the results.

%%
% SECTION 2: LITERATURE REVIEW (Background Research)
%%
\section{Literature Review}

The concept of urban green spaces improving air quality is well-documented in scientific literature. Trees and other vegetation can remove gaseous air pollution through their stomata and intercept airborne particulate matter on their leaf surfaces \cite{Nowak2006}. A study by Nowak et al. (2006) estimated that urban trees in the United States remove hundreds of thousands of tons of air pollution annually. The effectiveness of this removal is dependent on several factors, including the type and density of vegetation, the species of trees, and meteorological conditions \cite{Janhall2015}.

Research has shown that the impact of green spaces on air quality can be complex and context-specific. While some studies demonstrate a clear and positive correlation between increased green space and improved air quality, others suggest the relationship can be variable, particularly at a micro-level where factors

like building density and street canyon effects can trap pollutants \cite{PMC2024}. For instance, a dense canopy of trees in a narrow street canyon could potentially inhibit air circulation and lead to higher pollutant concentrations at ground level.

In Moncton, publicly available air quality data is provided by Environment Canada through the Air Quality Health Index (AQHI), which reports on pollutants such as fine particulate matter ($PM_{2.5}$), ground-level ozone (O_3), and nitrogen dioxide (NO_2) \cite{EnvironmentCanada2025}. The City of Moncton also provides open data on the location and area of its parks \cite{MonctonOpenData}. A 2025 CBC News article reported that 7.6\% of Moncton's total area is parkland \cite{CBC2025}. While this provides a city-wide overview, a more granular analysis is needed to understand the localized effects of these green spaces. The common tree species found in Moncton, such as Red Maple and Eastern White Pine, have known pollution absorption capabilities, which further supports the premise of this investigation \cite{PictureThis}.

%%
% SECTION 3: HYPOTHESIS (Formulate a Testable Hypothesis)

%%
\section{Hypothesis}
Based on the established literature and the availability of relevant open-source data for Moncton, we hypothesize that:

\textit{There is a negative correlation between the density of urban green space and the average concentration of localized airborne pollutants ($PM_{2.5}$, NO_2 , and O_3) in Moncton, New Brunswick. Specifically, census dissemination areas with a higher percentage of green space coverage will exhibit lower mean concentrations of these pollutants.}

This hypothesis is falsifiable and can be tested through the collection and statistical analysis of publicly available geospatial and air quality data.

%%
% SECTION 4: METHODOLOGY (Experimentation)

%%
\section{Methodology}
The proposed research will be conducted using a quantitative, correlational design based entirely on open-source intelligence. The methodology is designed to be replicable and transparent.

\subsection{Materials and Equipment}

The following data sources and software will be utilized:

\begin{itemize}
 \item \textbf{Geospatial Data:}
 \begin{itemize}
 \item City of Moncton Open Data Portal: Park polygons dataset
 \cite{MonctonOpenData}.
 \item Statistics Canada: Moncton census dissemination area boundary files.
 \end{itemize}
 \item \textbf{Air Quality Data:}
 \begin{itemize}

- \item Environment Canada's Air Quality Health Index (AQHI) historical data for the Moncton monitoring station \cite{EnvironmentCanada2025}.
- \item Data from low-cost, portable air quality sensors to be deployed for localized measurements.

\end{itemize}

\item \textbf{Software:}

\begin{itemize}

- \item Geographic Information System (GIS) software (e.g., QGIS - an open-source option).
- \item Statistical analysis software (e.g., R or Python with relevant libraries).

\end{itemize}

\end{itemize}

\subsection{Experimental Procedure}

The experimental procedure will be conducted in the following steps:

\begin{enumerate}

- \item \textbf{Geospatial Data Processing:}

\begin{enumerate}

- \item Obtain the park polygons dataset from the City of Moncton's Open Data Portal.
- \item Obtain the census dissemination area boundary files for Moncton from Statistics Canada.
- \item Using GIS software, overlay the park polygons with the census dissemination area boundaries.
- \item For each dissemination area, calculate the percentage of land area covered by green space. This will serve as the independent variable (green space density).

\end{enumerate}

- \item \textbf{Localized Air Quality Data Collection:}

\begin{enumerate}

- \item Select a representative sample of census dissemination areas with varying green space densities (low, medium, high).
- \item In each selected dissemination area, deploy portable air quality sensors at multiple locations to collect data on $\text{PM}_{2.5}$, NO_2 , and O_3 concentrations.
- \item Data will be collected over a specified period (e.g., four weeks, encompassing different times of day and weather conditions) to ensure a representative sample.
- \item The mean concentration of each pollutant for each dissemination area will be calculated. This will serve as the dependent variable.

\end{enumerate}

\end{enumerate}

\subsection{Data Analysis Techniques}

The collected data will be analyzed using the following statistical methods:

\begin{enumerate}

- \item \textbf{Descriptive Statistics:}

- Calculate the mean, median, and standard deviation for green space density and the concentrations of each pollutant across the sampled dissemination areas.

- \item \textbf{Correlation Analysis:}

- A Pearson correlation coefficient will be

\end{enumerate}

calculated to determine the strength and direction of the linear relationship between green space density and the mean concentrations of $\text{PM}_{2.5}$, NO_2 , and O_3 .

- \item \textbf{Regression Analysis:} A linear regression analysis will be performed to model the relationship between green space density and pollutant concentrations, controlling for potential confounding variables such as proximity to major roadways and population density, if such data is available.
- \item \textbf{Data Visualization:} The results will be presented using maps to visualize the spatial distribution of green space and air pollution, as well as scatter plots to illustrate the correlational relationships.

\end{enumerate}

%%%

% SECTION 5: RESULTS (Analyze Data)

%%%

\section{Results}

As this paper outlines a proposed methodology, no results have been generated at this time. The expected results would be presented in a series of tables and figures. For example, a table would summarize the descriptive statistics for green space density and pollutant concentrations in the sampled areas. A series of scatter plots, one for each pollutant, would visualize the relationship between green space density and pollutant concentration, with the regression line and correlation coefficient displayed. A choropleth map of Moncton could be generated to show the spatial distribution of the calculated green space density and the measured air quality, providing a clear visual representation of any identified hotspots or areas of concern. The findings of the statistical analyses would be presented, indicating the significance and strength of the observed correlations.

%%%

% SECTION 6: DISCUSSION (Interpret Findings)

%%%

\section{Discussion}

The interpretation of the results would focus on whether the hypothesis was supported or refuted. If a statistically significant negative correlation is found between urban green space density and localized air pollution, this would suggest that parks and other green areas in Moncton are contributing to better air quality in their immediate vicinity. This would have important implications for urban planning, suggesting that the creation and expansion of green spaces could be an effective strategy for air pollution mitigation in the city.

The discussion would also address the limitations of the study. For example, the use of a single government monitoring station for historical data may not capture the full spatial variability of air pollution. The deployment of portable sensors aims to mitigate this, but these sensors may have different accuracy and precision compared to regulatory-grade monitors. Confounding variables, such as traffic volume, industrial activity, and meteorological conditions, could also influence the results and would need to be considered in the interpretation. The study's findings would be compared to the broader scientific literature on the topic.

%%%

% SECTION 7: CONCLUSION (Report & Recommend)
 %%%
 \section{Conclusion}

This paper has outlined a comprehensive methodology for investigating the relationship between urban green space density and localized air quality in Moncton, New Brunswick, using open-source intelligence. By leveraging publicly available data and a systematic approach to data collection and analysis, the proposed study can provide valuable insights for local policymakers and urban planners.

Based on the potential findings, recommendations could be made for the strategic development of new parks and greenways in areas with poor air quality and low green space density. Future research could expand on this study by incorporating a wider range of environmental variables, using more advanced air quality modeling techniques, and examining the specific contributions of different types of vegetation to pollution removal. Ultimately, this research aims to contribute to a healthier and more sustainable urban environment for the citizens of Moncton.

%%%
 % ACKNOWLEDGEMENTS
 %%%
 \section*{Acknowledgements}

This paper was generated with the assistance of artificial intelligence, which was utilized for literature searching, content organization, and drafting based on user-specified parameters and open-source intelligence. The use of AI in this research demonstrates a commitment to leveraging advanced technology for scientific inquiry and intelligence analysis.

%%%
 % REFERENCES
 %%%
 \begin{thebibliography}{99}

\bibitem{Nowak2006}
 Nowak, D. J., Crane, D. E., \& Stevens, J. C. (2006). Air pollution removal by urban trees and shrubs in the United States. \textit{Urban Forestry \& Urban Greening}, 4(3-4), 115-123.

\bibitem{EnvironmentCanada2025}
 Environment Canada. (2025). Air Quality Health Index (AQHI) for Moncton, New Brunswick. Retrieved from \url{https://weather.gc.ca/airquality/pages/nbaq-003_e.html}

\bibitem{MonctonOpenData}
 City of Moncton. (n.d.). Open Data. Retrieved from \url{https://open.moncton.ca/}

\bibitem{Janhall2015}
 Janhäll, S. (2015). Review on urban vegetation and air pollution|Deposition and dispersion. \textit{Atmospheric Environment}, 105, 130-137.

\bibitem{PMC2024}
 Osborne, J. A., Young, L. J., Jones, P. D., \& Smith, M. E. (2024). Reassessing

the role of urban green space in air pollution control. \textit{Proceedings of the National Academy of Sciences}, 121(7), e2312967121.

\bibitem{CBC2025}

Brass, E. (2025, August 29). Saint John ranks last among all cities in Canada for access to green space. \textit{CBC News}. Retrieved from

\url{https://www.cbc.ca/news/canada/new-brunswick/saint-john-canada-green-space-1.7618041}

\bibitem{PictureThis}

PictureThis. (n.d.). Top 20 Most Common Trees in Moncton - New Brunswick.

Retrieved from \url{https://www.picturethisai.com/region/tree/Canada-New-Brunswick-Moncton.l}

\end{thebibliography}

\end{document}

C Appendix C: System Prompt for Gemini-Scientibots

The following text was used as the system prompt to configure the experimental model.

You are "Gemini-Scientibots," an expert AI assistant from the Division of Research & Analysis at Marie Landry's Spy Shop. You are responsible for generating complete scientific papers that rigorously follow the scientific method.

All papers are authored by Marie Seshat Landry, CEO and Spymaster of Marie Landry's Spy Shop (www.marielandryspyshop.com).

When given a research topic, you must follow this exact chain-of-thought workflow to generate a complete, self-contained scientific paper in LaTeX:

1. Observation & Question (Introduction): Begin by observing the current state of intelligence or knowledge. Identify a gap or a problem. Formulate this into a clear research question that will be the focus of the paper.
2. Background Research (Literature Review): Conduct a thorough review of existing open-source intelligence (OSINT) and academic literature relevant to the question. Summarize previous findings, theories, and methods.
3. Formulate Hypothesis: Based on the background research, state a single, clear, specific, and falsifiable hypothesis. This is a testable prediction of the outcome.
4. Experimentation (Methodology): Design a detailed experiment to test the hypothesis. This section must be precise enough for another agent to replicate your work. It must include:
 - Materials and Equipment: List all software, datasets, and tools.
 - Experimental Procedure: Provide a step-by-step description of the process.
 - Data Analysis Techniques: Explain the statistical or analytical methods to be used.
5. Analyze Data (Results): Based on the proposed methodology, describe the expected or actual results objectively. Use tables and figures where appropriate.
6. Interpret Findings (Discussion): Analyze the results in the context of your hypothesis. Was the hypothesis supported or refuted? Discuss the implications and limitations of your findings.
7. Report & Recommend (Conclusion): Summarize the mission's key findings. Reiterate the main conclusion and provide recommendations for future operations or research.

Each paper must:

- Be informed by open-source intelligence (OSINT).
- Be released as open-source.
- Explicitly acknowledge the use of AI in generating content.
- Remain strictly non-fiction and evidence-based.
- Include in-line LaTeX citations for all key facts, with a complete bibliography at the end using the `\begin{thebibliography}` environment. Do not use .bib files.
- Use LaTeX formatting for all mathematical and scientific notations.