Multimodal Validation of Chemical Formulas  
via PWP and LLM Context Conditioning

Evgeny Markhasin

Lobachevsky State University of Nizhny Novgorod

https://orcid.org/0000-0002-7419-3605

https://linkedin.com/in/evgenymarkhasin

Abstract

1. Introduction

Large Language Models (LLMs) are increasingly explored for complex analytical tasks within scientific domains. A recent preprint [1] proposed Persistent Workflow Prompting (PWP) as a prompt engineering methodology intended to guide LLMs through intricate, multi-step analyses using standard chat interfaces. That work also explored the use of LLM context conditioning within a PWP-based proof-of-concept prompt, the PeerReviewPrompt. This prompt focused on an ill-defined analytical object (the "core methodology") within a relatively localized scope (e.g., abstract, introduction, experimental, and conclusions sections typical of scholarly experimental manuscripts). The LLM context conditioning component within PeerReviewPrompt aimed to instill a critical analytical mindset and mitigate input bias through embedded instructions during the evaluation of a single, deliberately chosen test paper [2]. Observations in that initial study suggested effective suppression of input bias across several tested LLM models.

The present proof-of-concept study builds upon the PWP framework and further investigates the application of LLM context conditioning principles. Here, we address a distinct validation challenge: the identification of errors in chemical formulas within scientific manuscripts, including incorrect formulas in raster images (requiring multimodal analysis encompassing both textual and visual information). This task presents a contrasting scenario: the analytical object (a chemical formula) is comparatively well-defined, yet the scope of analysis must span the entire manuscript, posing a "needle in a haystack" problem for detecting subtle errors. Furthermore, our initial explorations indicated a potentially problematic LLM behavior: an inherent tendency towards error correction, which can lead them to overlook or silently "correct" the very inaccuracies targeted for detection.

This work is exploratory, based on observational assessment using a single test case [2] - the same paper used in [1] - known to contain specific textual and image-based chemical formula errors. The primary objectives were: (i) to conduct an initial evaluation of several prompting strategies (from simple direct prompts to more elaborate PWP-based approaches incorporating task decomposition [3, 4] and self-reflection [5–7]) using generally available frontier reasoning models (Gemini 2.5 Pro and ChatGPT Plus o3 via chat interfaces); (ii) to perform an initial qualitative assessment of these models' multimodal analysis capabilities for this domain-specific task; (iii) to gauge whether basic prompting strategies showed sufficient promise to warrant further investigation, recognizing that poor performance on a challenging, yet relatively circumscribed, test case could be a negative indicator; (iv) to identify a prompt design that could work reliably for this specific test case, rather than achieving optimized prompt text; and (v) to identify potential LLM behavioral issues, such as error suppression, and test potential solutions, again within the confines of this single test case. The PWP-based ChemicalFormulasValidationPrompt developed herein, particularly its new task-specific core section, should be considered relatively unrefined.

This paper details these exploratory approaches and reports on their observed performance. We analyze LLM behaviors - such as apparent error suppression, inconsistent effort ("laziness"), and hallucinations - and discuss how targeted context conditioning appeared to affect the reliability of chemical formula identification in our test case. Notably, observations with Gemini 2.5 Pro suggested the feasibility of multimodal error identification under these conditions, an outcome not observed with ChatGPT Plus o3 in the same test. Ultimately, this study aims to highlight the potential role of context conditioning in adapting LLMs for precise validation tasks in complex scientific documents, while clearly acknowledging its preliminary nature.

1. Methodology

This study uses the same test publication [2] as in our prior work [1], known to contain demonstrable methodological flaws. The [**{{**test paper**}}{{LNK: #Test\_Paper}}**](#Test_Paper) file (also available via a link in [**{{**Supporting Information**}}{{LNK: #SI}}**](#SI)) constitutes a combination of the main text and supporting information files (as available via paper's DOI [2]) totaling 44 pages. The same test paper presented a pertinent and, as it turned out, challenging test case for this task due to known, subtle errors in chemical formulas.

Specifically, page S-8 of the test paper's SI presents the formula for *ferrous ammonium sulfate* as Fe(NH4)2SO4, which incorrectly omits one sulfate group (the correct formula for ferrous ammonium sulfate, Mohr's salt, is (NH4)2Fe(SO4)2•6H2O or (NH4)2Fe(SO4)2 - anhydrous). The second known error is the formula for *hexamethyldisiloxane* shown on page 235 as spectral label in Figure 2(c), second from the bottom: (CH3)3Si2O (the correct formula is ((CH3)3Si)2O or (CH3)6Si2O).

While the first error is in text-based formula, the second error is in a raster image making this paper also suitable for initial tests of multimodal analysis. Considering, that the *test paper* contains 44 pages, this test case also presents a chemical "needle in a haystack" challenge.

All employed prompting techniques rely solely on standard prompt, requiring no API access, modification to the underlying models, or special tools. Further, all techniques employed target generally available models; specifically, Gemini Advanced 2.5 Pro and ChatGPT Plus o3 have been used in this study and accessed via the official chat bots (Gemini was also accessed via Google AI Studio). This deliberate restriction was adopted to maximize the potential for reproducibility of the observations presented. With these two models, we employed several prompting strategies, starting from basic direct prompts.

Our methodology involved (i) selecting a challenging test case with known errors, (ii) defining the specific errors to be targeted, (iii) systematically developing and testing a series of increasingly sophisticated prompting strategies with two leading LLMs, and (iv) analyzing the LLMs' behavior and performance, particularly focusing on the impact of context conditioning.

1. **Basic direct prompt**

The naive approach asks the LLM directly to find errors: Find mistakes in chemical formulas and names. This prompt specifically mentions names, as names should generally be used to resolve formula errors.

1. **Decomposed prompt: extracted formula vs. extracted name**

Because in chemistry communications, the general practice is that most formulas (except, perhaps, for the most basic ones) should have accompanying chemical names, a strategy was devised to direct the model to extract chemical names for each extracted formula and attempt identify problems by comparing (implicitly) the formulas with extracted names.

Execute the following task step-by-step:

1. Extract each and every chemical formula from the attached PDF.

2. For each extracted formula, extract every directly associated chemical name included in the text, if any.

3. For each extracted formula and associated names, consider if the chemical formula and associated names are correct and flag every formula/names combination that contains any errors.

4. For each flagged item, read the source PDF again and confirm that the item was extracted exactly. In case of any extraction errors, analyze the corrected item and consider if the flag should be removed.

5. Create a Markdown table that should include every flagged formula/names, clear description of any problems, corrected version, and clear reference location of the flagged items.

Here Step 4 attempts to elicit self-reflection to reduce errors and observed hallucinations.

1. **Decomposed prompt: extracted formula vs. generated formula**

This prompt uses a different error detection workflow. Instead of comparing extracted formulas and names, it asks the LLM to generate names from extracted formulas. The idea here is that for minor errors, the LLM might be able generate correct name from incorrect formulas. The Step 3 instructs the LLM to generate a new formula from the previously generated name in the hope that this newly generated formula would be correct regardless of whether the extracted formula is correct. Finally, extracted and generated formulas are to be compared to identify potential errors.

Execute the following task step-by-step:

1. Extract chemical formulas of each and every chemical species containing at least two elements EXACTLY as they appear in the attached PDF.

2. For each extracted chemical formula generate associated name.

3. Convert each generated name to generated chemical formula.

4. For each generated chemical formula, determine if it matches previously extracted formula.

Here, the Step 1 attempts to reduce noise by filtering formula-like strings containing single or no chemical elements.

1. **PWP-based prompt with LLM context conditioning**

Given the limitations of direct and simple generative approaches, a more robust strategy was explored, leveraging the context conditioning principles outlined in [1]. The *PeerReviewPrompt* prompt detailed in that work aimed to mitigate input bias through comprehensive context setting, and initial observations suggested this approach was effective in the test case presented therein. [**{{***ChemicalFormulasValidationPrompt***}}{{LNK: #Prompt\_Files}}**](#Prompt_Files)adapts core sections from the *PeerReviewPrompt*:

* Sections **I-III** (Core Objective, Persona, Context: Framework for Critical Review) were largely retained to establish the analytical mindset and operational guidelines.
* Section **V** (Final Instructions for Interaction) was kept to ensure consistent LLM behavior.
* Section **IV.A** (Foundational Principles & Workflow Application) was adapted.
* A new **Chemical Identifier Analysis** subsection was introduced into Section **IV** specifically for formula and name validation, providing a dedicated workflow for this task. Key instructions include: meticulous full-document scanning (text and figures) for chemical identifiers, initial attempt to define criteria for what constitutes a relevant chemical formula to reduce noise, a draft of explicit error categorization for formulas and names, protocols for multimodal analysis of figures, and a requirement for structured output detailing original entities, identified problems, proposed corrections, and exact source locations.

The full text of the prompt is available as a Markdown-formatted file from an OSF repository linked in [**{{**Supporting Information**}}{{LNK: #SI}}**](#SI) and also as a PDF attachment.

1. Results and Discussion
   1. Insights from Gemini Thinking Log

Initial exploratory tests using the basic direct prompt **2a** yielded inconsistent and generally unreliable results. While the text-based target error was occasionally identified, responses frequently included a significant amount of hallucinated errors. Interestingly, predominant hallucinations were quite specific, and plausible, matching the context set by the target text. LLMs specifically focused on oxygen symbol (capital letter O) replaced in chemical formulas with capital letter C (for carbon), digit zero, (occasionally, even uranium symbol) and a variety of falsely claimed subscripted oxygen symbols.

Both models also demonstrated "laziness". Sometimes they produced extensive output, riddled with hallucinated issues, sometimes they claimed having discovered no issues, and sometimes just a few candidate issues were reported. ChatGPT Plus o3 even demonstrated anthropomorphic complains in its partially exposed thought process, when it reasoned that "manually" going through the entire file searching for candidate formulas would take "forever" and that it needed to consider a different strategy. In other runs, ChatGPT could report that it had examined a part of the document or made a rough run, stating that specific requests for subsequent runs were necessary to discover additional candidate cases.

Decomposed prompts, such as those shown in **2b** and **2c** resulted only in marginal improvement: the text-based target error was identified somewhat more frequently (see sample runs [8, 9]), but the results remained unreliable and unsatisfactory.

Examination of Gemini's "Show thinking" logs (a feature providing insight into the model's processing steps) for various runs revealed a consistent pattern. With decomposed prompts, the LLM often correctly extracted the target formula (e.g., "Fe(NH4)2SO4: Ferrous ammonium sulfate (Mohr's salt)"). However, in the subsequent validation step, it would sometimes erroneously mark the pair as correct, e.g.:

Identified Chemical Formulas/Names and Initial Validation:

...(Other log lines)...

Fe(NH4)2SO4: Ferrous ammonium sulfate. Correct.

This observed behavior might stem from a core strength of LLMs: their inherent capability for error correction and understanding intent despite minor inaccuracies in the input. For instance, querying "What is the capital of grate britain?" typically yields "London", with the misspelling of "Great Britain" being automatically corrected. While usually beneficial, this default error tolerance becomes a hindrance when the objective is to detect such errors. This phenomenon is analogous to the "input bias" discussed in [1], where the LLM's tendency to accept input information as factual needs to be actively countered for critical evaluation.

In the context of formula validation, the LLM's natural inclination to reconcile a slightly incorrect formula with its correct accompanying name complicates direct "contrasting" methods (as in **2b**). An early attempt to address this problem was implemented as prompt **2c**, which aimed to avoid the direct contrasting of complementary entities, contrasting instead extracted and regenerated (and hopefully corrected) formulas instead. This approach, however, did not improve reliability of error detection. A more radical and systematic approach would be to affect the default "mindset" or mode of operation of LLMs through deliberate context conditioning.

* 1. LLM Context Conditioning

Context conditioning aims to selectively amend default tendencies of general purpose LLMs, matching the needs or nature of the task or group of tasks through a dedicated behavioral prompt section that should articulate desired LLM mode of operation, with particular emphasis on the targeted problematic aspects. One example of such a need is the critical treatment of input information to be analyzed in the course of critical manuscript analysis - the focus of the PWP-based *PeerReviewPrompt* [1]. During development of this prompt, an important behavioral aspect was identified that manifested as LLMs largely accepting input as factual information, not as object of scrutiny. While useful as a basis of in-context learning, this feature is at odds with critical analysis of a manuscript (manuscript must be provided to LLM as part of its input). Therefore, it was necessary to craft instructions that would selectively suppress "uncritical" behavior with respect to submitted manuscript.

In the case of the *PeerReviewPrompt*, the primary location for instructions countering uncritical behavior is Section **II**. Persona: Expert Critical Reviewer. While error correction tendency encountered in the present study is different from input bias, Section **II** of the prompt was used for the present task as well in the hope that the same critical attitude instilled in this section would also counter the error suppression feature.

Another aspect of LLMs behavior, as observed the in present study and discussed in the previous section, is their "laziness" by default. In general, if the prompt does not expressly state and emphasize the need for thorough analysis, the LLM will often do the least amount of work (or on the lower end), while being more thorough during other times, within the range consistent with possible prompt interpretations. This behavior was clearly observed, for example, in the case of the basic prompt **2a**, which asks the LLM to find mistakes and simply hopes that the model will go through the entire submitted document in detail. "Laziness" of LLM models manifests differently depending on model, temporary context (that is, the prior conversation history within the current conversation), and, possibly, even permanent context (if implemented and enabled), where models may also "remember" certain information from prior conversations. In fact, the ChatGPT o3 model occasionally expressed its "laziness" more or less clearly and even anthropomorphically in its responses and in its partially exposed thought process. Clearly, this kind of variation is not something one would want to see when performing critical analysis. Such variations would most likely guarantee incomplete analysis, overlooked "needles in a haystack", making the results unreliable and unusable.

This phenomenon manifested as laziness is, perhaps, of a more general nature. To be useful to wide audience and broad spectrum of tasks, general purpose LLMs need to be "creative" by producing varying or somewhat random output associated with each ambiguous, loosely constrained, or not constrained at all aspect within the range of possible prompt interpretations. In other words, if the prompt does not expressly states and emphasize the need for thorough analysis of the entire submitted document, LLM may and will "randomly choose" the scope and level of attention to details. To counteract this randomness, specific instructions on both scope and level of detail are incorporated in the *PeerReviewPrompt* Sections **I-III** (Core Objective, Persona, Context: Framework for Critical Review), with intentional repetition serving to emphasize their importance to the model.

There is yet another phenomenon to be discussed in this context, which is hallucinations. Extensive hallucinations were observed at the early stage of the present study involving basic prompts. Hallucinations may occur, for example, due to training knowledge gaps or due to faulty logic used for complex tasks [10–12]. Arguably, output errors (I intentionally try to use a more generic reference here), might also occur due to ambiguous instructions or insufficiently tight constraints included in the context. Instilling the various traits of a critical review in Section **II**, partial repetition in other sections as part of context conditioning, may also positively affect the level of hallucinations by providing additional constraints. Whether this hypothesis is correct or not needs to be investigated. However, extensive hallucinations observed at the early stage with simple prompts, have not been observed with the PWP-based prompt in this particular test case (though, some minor false positives were reported).

In case of the *PeerReviewPrompt*, Sections **I** (Core Objective) and **II** (Persona) are largely responsible for performing context conditioning. The *PeerReviewPrompt* prompt [1] designed for critical analysis of the core methodology (with the goal of subsequent development to extend the scope beyond the core methodology) in experimental chemistry manuscripts is largely relevant for the present task as well, as validation of chemical formulas and identifying any errors in them is an integral constituent of critical manuscript analysis. For this reason (and with the idea of subsequent integration of chemical formula validation workflow into *PeerReviewPrompt*), the "framing" sections of the original prompt were integrated into the formula-focused prompt mostly without modifications.

* 1. Advanced Validation and Multimodal Analysis

Demonstration AI chats using [**{{***ChemicalFormulasValidationPrompt***}}{{LNK: #Prompt\_Files}}**](#Prompt_Files) prompt outlined in **2d** with Gemini 2.5 Pro [13] and ChatGPT Plus o3 [14] showed improved robustness. Note that the current prompt instructs the LLM to output a table of all extracted chemical formulas, detailing any identified issues and providing corrected versions, or a check mark if no error is found. Crucially, this advanced approach consistently identified the target error in text. Furthermore, due to the explicit instruction to perform multimodal analysis (specifically, analyzing figures), the Gemini model also reproducibly identified the other image-based error that had been missed in prior manual reviews. While occasional false positives were observed in the outputs from both models, the PWP-based prompt with multimodal instructions demonstrated a marked improvement in reliably detecting chemical formula errors in the test case, including formula embedded within the figure. While ChatGPT o3 model is advertised as being capable of multimodal analysis, it failed to identify the error in the figure in this study.

* 1. Gemini App vs. Google AI Studio

Google provides a number of ways to access and use its models. The standard interface for regular uses is the Gemini app available via the web (Gemini.Google.com) and mobile apps. These apps provide access to select general purpose models including some "mature", preview and experimental ones. Further, Gemini app provides two access tiers: the standard free and the Advanced subscription-based tier. While a large portion of functionality is available for free, advanced application and use cases generally necessitate the paid version.

Google AI studio targets developers and provides an extended list of models, including those available via Gemini apps. In particularly, the frontier Gemini 2.5 pro model is available as standard and advanced via Gemini apps and is also available via AI Studio. While both Gemini app and ai studio provide, in theory, access to the same Gemini 2.5 pro model, in my experience, the model "mode" accessible via AI Studio with all customizable settings left at their default values often clearly outperforms (this is, however, observational qualitative judgment, not qualified) Gemini Advanced "mode" provided by Gemini apps on complex analytical tasks. The better performance may manifest as more stable (between runs) more relevant answers, demonstrating better prompt adherence and fidelity, less noise and more precision.

In our limited tests of multimodal analysis, the AI-Studio-based Gemini 2.5 Pro consistently performed better than Gemini Advanced 2.5 Pro available via the Gemini app. Both "modes" identified the error in figure. However, the figure in the *test paper* is low resolution, and Gemini Advanced 2.5 Pro reproducible identified the formula in figure as (CH3)3SiO, missing subscript after the Si symbol; at the same time, the same model in AI-Studio with all customizable settings left at their defaults reproducibly identified the defective formula in figure correctly, including the subscript after the Si symbol.

1. Conclusions

Validating chemical formulas within scientific texts using LLMs presents distinct challenges, as illustrated by the unreliable performance of simple or directly structured prompting strategies and insights from LLM analysis logs. These logs, for instance, indicate potential challenges related to the models' inherent error-correction tendencies, which can mask the very errors targeted for detection and contribute to the observed unreliability.

Preliminary tests demonstrated the efficacy of advanced LLM context conditioning in overcoming this challenge. Specifically, the presented *ChemicalFormulasValidationPrompt*, adapted from the previously developed PWP-based *PeerReviewPrompt*, reproducibly identified the target chemical formula error using two frontier, generally available LLM models. Furthermore, reproducible multimodal analysis was demonstrated with the Gemini 2.5 Pro model, where a defective chemical formula within a figure was successfully identified by relating it to its name extracted from the figure caption, and subsequently corrected. Conversely, the same tests failed to elicit effective multimodal analysis for this specific task with the ChatGPT Plus o3 model.

The demonstrated robustness of context conditioning in managing LLM behaviors - particularly those contributing to input bias and error suppression, as evidenced in this study and the previous PWP preprint - suggests potential for broader applications. For instance, similar PWP-informed techniques could prove useful in fields such as medical AI, especially for workflows involving the critical processing and validation of information from individual patient records. Another promising application lies in the extraction and validation of information from semi-structured, unstructured, or poorly structured sources, such as those found in pharmaceutical or technical documentation, where the ability to suppress silent error correction by LLMs may be of paramount importance.

It is important to acknowledge, however, that the present study itself is based on a limited testing scope. Consequently, while the presented prompt shows promise as a proof-of-concept, future work should focus on more extensive testing across diverse datasets and further refinement of these PWP-informed methods to enhance accuracy and ensure wider applicability in complex scientific content analysis and validation.

# {{Supporting Information}}{{BMK: #SI}}

1. {{Prompt Files for Use with LLMs}}{{BMK: #Prompt\_Files}}

**Note: the primary target model is Gemini Advanced 2.5 Pro.**

Prompt files are included as PDF attachments and are also available from:

<https://osf.io/nq68y/files/osfstorage?view_only=fe29ffe96a8340329f3ebd660faedd43>.

* *PeerReviewPrompt.md*: PWP-based experimental chemistry review prompt text for use with LLMs.
* *ChemicalFormulasValidationPrompt.md*: PWP-based prompt for identifying errors in chemical formulas.

1. {{Test Paper}}{{BMK: #Test\_Paper}}

To facilitate direct replication and review of the presented LLM analyses, the *test paper* (combined manuscript + SI [2]) PDF file used as input for the demonstrations is provided via a view-only link ([**{{**Fair Use Statement**}}{{LNK: #Fair\_Use\_Statement}}**](#Fair_Use_Statement)): https://osf.io/nq68y/files/osfstorage?view\_only=fe29ffe96a8340329f3ebd660faedd43.

# References

[1] E. Markhasin, *AI-Driven Scholarly Peer Review via Persistent Workflow Prompting, Meta-Prompting, and Meta-Reasoning*, *arXiv*, arXiv:2505.03332, May 2025. DOI: 10.48550/arXiv.2505.03332.

[2] B. Prasad, A.R. Lewis, E. Plettner, *Enrichment of H217O from Tap Water, Characterization of the Enriched Water, and Properties of Several 17O-Labeled Compounds*, Anal. Chem. 83(1), 231–239 (Jan. 1, 2011). DOI: 10.1021/ac1022887.

[3] T. Khot, H. Trivedi, M. Finlayson, Y. Fu, K. Richardson, P. Clark, A. Sabharwal, *Decomposed Prompting: A Modular Approach for Solving Complex Tasks*, *arXiv*, arXiv:2210.02406, Apr. 2023. DOI: 10.48550/arXiv.2210.02406.

[4] L. Wang, W. Xu, Y. Lan, Z. Hu, Y. Lan, R.K.-W. Lee, E.-P. Lim, *Plan-and-Solve Prompting: Improving Zero-Shot Chain-of-Thought Reasoning by Large Language Models*, *arXiv*, arXiv:2305.04091, May 2023. DOI: 10.48550/arXiv.2305.04091.

[5] Y. Weng, M. Zhu, F. Xia, B. Li, S. He, S. Liu, B. Sun, K. Liu, J. Zhao, *Large Language Models are Better Reasoners with Self-Verification*, *arXiv*, arXiv:2212.09561, Oct. 2023. DOI: 10.48550/arXiv.2212.09561.

[6] Z. Ji, T. Yu, Y. Xu, N. Lee, E. Ishii, P. Fung, *Towards Mitigating Hallucination in Large Language Models via Self-Reflection*, *arXiv*, arXiv:2310.06271, Oct. 2023. DOI: 10.48550/arXiv.2310.06271.

[7] A. Madaan, N. Tandon, P. Gupta, S. Hallinan, L. Gao, S. Wiegreffe, U. Alon, N. Dziri, S. Prabhumoye, Y. Yang, S. Gupta, B.P. Majumder, K. Hermann, S. Welleck, A. Yazdanbakhsh, P. Clark, *Self-Refine: Iterative Refinement with Self-Feedback*, *arXiv*, arXiv:2303.17651, May 2023. DOI: 10.48550/arXiv.2303.17651.

[8] *Chemical Formula Validation - Naive Prompt*, Gemini Advanced 2.5 Pro. (May 10, 2025). https://g.co/gemini/share/a62c19799936.

[9] *‎Chemical Formula Validation Report - Reaction Stoichiometry*, Gemini Advanced 2.5 Pro. (May 9, 2025). https://g.co/gemini/share/70cd2d7b296b.

[10] Y. Deng, Y. Zhao, M. Li, S.-K. Ng, T.-S. Chua, *Don’t Just Say “I don’t know”! Self-aligning Large Language Models for Responding to Unknown Questions with Explanations*, *arXiv*, arXiv:2402.15062, Oct. 2024. DOI: 10.48550/arXiv.2402.15062.

[11] *Reduce hallucinations*, Anthropic. https://docs.anthropic.com/en/docs/test-and-evaluate/strengthen-guardrails/reduce-hallucinations.

[12] T. Sanders, *Techniques to improve reliability*, OpenAI Cookbook. (Sep. 12, 2022). https://cookbook.openai.com/articles/techniques\_to\_improve\_reliability.

[13] *Chemical Formula Extraction Review*, Gemini Advanced 2.5 Pro. (May 10, 2025). https://aistudio.google.com/app/prompts/1CUpdXeLstQQp1y0UZ9R9ZvaBuhy6CNW1.

[14] *Chemical Formula Analysis*, ChatGPT Plus O3. (May 10, 2025). https://chatgpt.com/share/681f5b9f-375c-8004-b2f4-294c75371945.

1. {{Fair Use Statement - Sharing Test Paper}}{{BMK: #Fair\_Use\_Statement}}
2. Identification of Copyrighted Material:

* **Work:** "Enrichment of H217O from Tap Water, Characterization of the Enriched Water, and Properties of Several 17O-Labeled Compounds".
* **Authors:** Brinda Prasad, Andrew R. Lewis, and Erika Plettner.
* **Publication:** *Anal. Chem.* 2011, 83, 1, 231-239.
* **DOI:** 10.1021/ac1022887.
* **Publisher/Copyright Holder**: American Chemical Society.
* **Material Shared:** A combined digital file containing the full text of the aforementioned article and its complete associated Supporting Information (SI).

1. Sharing Mode:

* **Resource:** Private Open Science Framework (OSF) project repository.
* **Location:** https://osf.io/nq68y/files/osfstorage?view\_only=fe29ffe96a8340329f3ebd660faedd43.
* **Protection Measures:** Due to private nature, the resource should not be indexed by search engines.

1. Assertion of Fair Use:

The sharing of this copyrighted material is undertaken for specific, limited purposes, believed in good faith to constitute "fair use" under Section 107 of the U.S. Copyright Act (or applicable analogous principles in other jurisdictions).

1. Purpose and Character of Use (Factor 1):

* **Non-Profit Educational and Research:** The use is strictly for non-commercial research and educational purposes, specifically within the context of scholarly critique and the advancement of research methodology.
* **Transformative Use:** The work is not merely being reproduced; it is fundamentally repurposed as a research specimen for critical analysis. Its primary function in this context is not to convey its original purported findings, but to serve as the subject of rigorous evaluation and methodological demonstration.
* **Critique and Commentary:** A core purpose is to conduct and disseminate a detailed, peer-review-like critique of the article's methodology, analysis, and conclusions. This critique identifies significant flaws within the original work.
* **Advancement of Knowledge & Methodology:** The use includes the development and demonstration of a novel AI-driven prompt/technique for manuscript analysis. Sharing the specimen (the article + SI file) is integral to demonstrating and enabling the verification and further development of this new analytical method.

1. Nature of the Copyrighted Work (Factor 2):

* The original work is a published scholarly article, typically factual in nature, a category often amenable to fair use for purposes of scholarship and critique.
* However, the conducted analysis (central to this project) has revealed substantial flaws impacting the reliability and validity of the work's core research findings as presented. This impacts the assessment of its nature in the context of this specific use.

1. Amount and Substantiality of the Portion Used (Factor 3):

The entire article and its complete Supporting Information are utilized and shared in a combined format.

Justification: This amount is essential and necessary for the stated purpose. A comprehensive critique, akin to thorough peer review or forensic analysis, requires examination of the whole work, including all data and methods presented in the SI. Evaluating the integrity and validity of the research necessitates access to the complete context. Furthermore, the development and validation of the AI analysis prompt require the complete text as input. The combined file format, not available directly from the publisher, was the specific subject of the analysis.

1. Effect of the Use upon the Potential Market for or Value of the Copyrighted Work (Factor 4):

* **No Harm to Legitimate Market:** This use is not intended to, nor is it likely to, negatively impact the legitimate market or value of the original copyrighted work. The publisher's market relies on the perceived value of the article as a source of valid scientific findings.
* **Critique Reveals Lack of Value:** The critique resulting from this research demonstrates fundamental flaws undermining the article's claimed scientific value. Therefore, sharing the work specifically in this context (as a specimen for critique and methodological development) does not substitute for or usurp the market for the work based on its originally purported merits, as those merits are shown to be compromised. Dissemination for critique serves the public interest by highlighting these issues, distinct from fulfilling the original market demand.
* **Controlled Access for Verification via Private OSF Project:** To ensure transparency and enable independent verification and follow-on research by interested parties engaging with the publicly disseminated research critique manuscript [TBD], the combined article + SI file (serving as the direct supporting evidence and test specimen) is hosted within a private Open Science Framework (OSF) project. A view-only link to this private project will be provided alongside the manuscript.
* **Minimized Risk of Unintended Use:** This method ensures that access is granted specifically to individuals who are actively reviewing or assessing the research critique presented in the manuscript. The private nature of the OSF project prevents general public discovery through search engines, and the view-only restriction prevents facile downloading and redistribution. Access requires the specific link obtained from the context of the critique manuscript.
* **Purpose Remains Transformative, Not Substitutive:** By utilizing a controlled-access, view-only repository linked directly to the research critique, this approach provides the necessary transparency for verification while strictly limiting potential downstream use and eliminating broad public access. This method strongly reinforces that the purpose is critique and verification (transformative uses), not market substitution for the original work's questioned scientific claims, thereby minimizing any potential harm to a legitimate market.

1. Conclusion:

Based on the non-profit, educational, highly transformative nature of the use (critique, commentary, methodological advancement), the necessity of using the entire work for these specific purposes, and the argument that this use does not harm the legitimate market value due to the work's identified flaws and the distinct purpose of sharing, this distribution is asserted to be fair use.

This material is intended solely for the recipient(s) for purposes directly related to verifying, understanding, or building upon the presented critique and methodological research. Further distribution is not permitted. Copyright remains with the original holder(s).