

Application of the FLUF Test Framework for Evaluating AI-Generated Outputs in Pharmacy Education

Alexander, K., Feild, C., Egan, C., & Parker, J.

Artificial Intelligence

Critical Thinking

Drug Information

Educational Technology

Evaluation Framework

Generative

Health Professions Education

Information Literacy

As generative artificial intelligence (AI) tools become increasingly integrated into healthcare, health professions education must adapt to ensure students can critically evaluate AI-generated information. To address this need, a pharmacy school assignment was revised to include the critical evaluation of AI-generated outputs in response to a drug information question. The assignment aimed to educate students about AI and equip them with the skills to assess AI-generated information using the "Format, Language, Usability, and Fanfare (FLUF) Test" framework. This pilot study analyzed how twenty-one third-year pharmacy students applied the FLUF Test to identify issues in AI-generated responses to a drug information question and explored their perceptions of the framework's usability through an open-ended survey. Data was analyzed using descriptive statistics and thematic analysis. FLUF Test results demonstrated that students most frequently identified usability concerns, particularly credibility, in the AI-generated

responses. Thematic analysis of students' open-ended survey question comments revealed three themes: the FLUF Test was useful, easy to use, and the students recognized the need for the human element when interacting with AI. However, some students noted that applying the FLUF Test could be subjective. This pilot study demonstrates the potential value of using the FLUF Test to aid in AI content evaluation in pharmacy education and theoretically other areas of instruction.

Introduction

The increasing popularity of generative artificial intelligence (AI) tools among students has led to widespread adoption for academic purposes, including creating, studying, and brainstorming. A recent survey of over 1,000 university students in the United States revealed that approximately one-third had utilized OpenAI's ChatGPT™ for coursework (Intelligent.com, 2023). However, more recent studies suggest that this figure may be significantly higher, with reports indicating that 50-75% of students have used AI tools (Tyton Partners, 2023; Crompton, et al., 2023). As the use of AI continues to grow, it is essential to prioritize teaching students how to use these tools responsibly and effectively.

The widespread adoption of AI has raised concerns among faculty about ethics, the potential loss of creativity and critical thinking skills, and overreliance on these tools (Civil, 2023; Warschauer et al., 2023). As a result, some faculty members have been cautious in integrating AI into their classrooms, while others have taken a more proactive approach by developing strategies to mitigate the risks (Tyton Partners, 2023; Cain, 2023). These strategies include designing authentic assignments using AI that promote creativity and innovation, as well as establishing AI policies that clearly define use (Chan, 2023; Mollick & Mollick, 2023). By implementing these measures, educators can “inspire students to positively contribute to and responsibly participate in the digital world” (International Society for Technology in Education [ISTE], n.d.).

As AI tools become increasingly integrated into healthcare, health professions education must adapt to ensure students can critically evaluate AI-generated information. Additionally, AI holds the potential to revolutionize teaching and learning by serving as a tool to enhance students' learning (Cain, 2023). However, realizing this potential requires addressing challenges such as creating assessments that uphold academic integrity and foster problem-solving skills in an AI-driven environment. At the University of Florida College of Pharmacy, faculty working with an instructional designer addressed these challenges by revising a drug information (DI) assignment in an elective course to incorporate AI. The updated assignment was designed to emphasize the critical evaluation of AI-generated outputs, leveraging a new framework, the “Format, Language, Usability, and Fanfare” (FLUF) Test as an evaluation tool (Parker, 2023). The purpose of this mixed-methods study was to explore how pharmacy students applied the FLUF Test within a DI assignment and to examine their perceptions of its usability and value in evaluating AI-generated responses.

Research Questions

1. What are pharmacy student perceptions of the application and usability of the FLUF test framework within a DI assignment?
2. What FLUF infractions are identified by students when the FLUF Test framework is applied to a DI assignment?

Literature Review

Information Literacy as a Foundational Framework

Information literacy is a crucial skill set for academic success and lifelong learning, encompassing the ability to locate, evaluate, and use information effectively and ethically. Pedagogies proposed by scholars such as Bruce (1997), highlight the relational model of information literacy, and Kuhlthau (2004), whose Information Search Process (ISP) underscores the effective aspects of searching for information.

Information literacy as a framework inspires research, critical evaluation, and human insight. The Association of College and Research Libraries (ACRL) outlines key frameworks for information literacy, emphasizing critical thinking and ethical information use ([ACRL, 2016](#)). ACRL defines six concepts that anchor the frames of information literacy and provide a foundation for critical evaluation: 1) Authority is Constructed and Contextual; 2) Information Creation as a Process; 3) Information has Value; 4) Research as Inquiry; 5) Scholarship as Conversation; 6) Searching as Strategic Exploration.

As technology has evolved, information literacy has become linked to *digital literacy* and *AI literacy* as interconnected competencies essential for navigating the modern information landscape (Long & Magerko, 2020). Information literacy is the ability to recognize when information is needed and to locate, evaluate, and use that information effectively and ethically (ACRL, 2016). Digital literacy extends this concept to encompass the skills and knowledge required to use digital technologies critically and creatively, integrating areas such as ethical use and digital content creation (Calvani et al., 2008; Eshet, 2004). AI literacy builds upon these foundations, incorporating the competencies to understand, interact with, and critically evaluate AI technologies, emphasizing ethical considerations and informed decision-making in their use (Long & Magerko, 2020). Together, digital and AI literacies equip individuals with a comprehensive skill set to successfully navigate digital environments, seek information, and critique results while settling comfortably under the overarching idea of information literacy.

The integration of LLMs and generative AI tools into information literacy instruction offers a transformative approach to enhancing students' research skills. Carroll and Borycz (2024) highlight the critical role generative AI can play in the information literacy landscape, where precise and efficient information retrieval is paramount. The authors note that specialized language models have the potential to outperform traditional human-crafted queries, thereby fostering a deeper and more nuanced understanding of information sources. This alignment of AI capabilities with information literacy pedagogy not only augments the instructional process but also prepares students to adeptly handle contemporary and future information challenges (Carroll & Borycz, 2024).

In *New Horizons in Artificial Intelligence in Libraries* (2024), Cox issues a call for AI literacy and provides competencies, including, "What AI can do, the skill of differentiating the tasks AI is good at doing from those it is not good at, and imagining future uses, reflecting the evolving nature of AI" (p.66). Additionally, Ndungu (2024) provides a comprehensive framework for incorporating AI literacy into media and information literacy programs, emphasizing the importance of both ethical considerations and the verification of AI-generated content. Together, these insights underscore the potential of integrating generative AI tools within learning activities to foster a more informed and critically engaged student body.

As we consider AI-enabled assignments in higher education that cause students to think more critically, we must also consider the importance of information literacy, and now AI literacy, in the process. By embedding information literacy principles into our instructional strategies, we empower students to become discerning consumers and producers of information. Ndungu (2024) further supports this integration by advocating for comprehensive AI literacy programs that train students to ethically utilize and verify AI-generated content, thereby fostering a more nuanced understanding of the digital information environment.

Shortcomings of Traditional Information Literacy Models on Generative AI

Integrating information literacy means understanding the research process and the element of critique. A Google search, a blog post, eBook, magazine article, or a journal article, should be reviewed for relevance and reliability. Traditional applications of information literacy incorporate primary sources and internet searching using frameworks like the Currency, Relevance, Authority, Accuracy, and Purpose (CRAAP) Test (Blakeslee, 2004), CARRDSS system (Credibility / authority, Accuracy, Reliability, Relevance, Date, Sources behind the text, Scope and purpose) (Valenza, 2004), 5 Key Questions (Thoman & Jolls, 2003) and SIFT method (Stop, Investigate the source, Find better coverage, Trace claims) (Caulfield, 2019). These information literacy frameworks guide students through a list of “look fors” such as validity, reliability, authorship, etc.

While these existing frameworks provide valuable guidance, they were not designed with the complexities of generative AI in mind, highlighting a critical need to equip students with an updated framework that specifically explores the outputs of generative AI. There is a need to empower students with frameworks that explore the output of generative AI. Traditional frameworks like CRAAP and SIFT are still being applied to websites and social media with no mention of AI. Even as recently as 2023, Sye and Thompson examined the practical application of the CRAAP Test and SIFT Method using the Gale's Opposing Viewpoints database, where students chose topics, read viewpoint essays, and found supporting scholarly and news articles (Sye & Thompson, 2023). They then applied the CRAAP Test and questions aligned with the SIFT Method to critically assess these sources. This structured approach provided students with a comprehensive understanding of source evaluation, preparing them to navigate the complexities of information literacy in an increasingly digital world. While the study evaluates these frameworks and how students apply them online (e.g. social media, websites), it does not delve into how AI might integrate with or influence these source evaluation frameworks. Instead, the emphasis is on enhancing information literacy and critical thinking skills through established evaluation methods. This study serves as a reminder that the use of AI might significantly change the research journey and the trajectory of the assignment.

Although these models teach users to have a critical eye, they were designed to critique online content garnered from an internet search engine, as opposed to the results of AI. Traditional information literacy frameworks like the CRAAP Test do not address elements specific to AI-generated content and do not include AI prompt generation and re-prompting to improve results. Prior to 2023 no framework existed to provide guidance on how to critically evaluate AI-generated results. Using information literacy protocols as a foundation and best practices in prompt generation, the “Format, Language, Usability, and Fanfare” (FLUF) Test was developed as a tool for critically evaluating content generated through AI (Figure 1). By examining each factor, the FLUF Test empowers users to improve their interactions with AI tools, as well as create more effective prompts. (Parker, 2023)

Since the launch of OpenAI™ in November of 2022, frameworks have continued to evolve around prompt writing, interactions, and information literacy (Wei, et al., 2022; Brown, et al., 2020; Mishra, et al., 2021). Prompt engineering is a critical element in interacting with AI and the foundation of a good result (Giray, 2023; Korzynski, et al., 2023; Mesko, 2023; Short & Short, 2023; Heston & Kuhn, 2023). However, AI is not a “search engine,” and prompts can sometimes lead to hallucinations or false, misleading, or inaccurate results (Maleki, et al., 2024). Identifying these inaccuracies can be a challenge for novice AI users.

Benefits of the FLUF Test include that it incorporates both prompt building and information literacy (Figure 2). Published around the same time as the FLUF Test is the American Library Association's “ROBOT test” (American Library Association, 2023). This framework explores Reliability, Objective, Bias, Ownership, and Type as a structured method for scrutinizing AI outputs, which is similar to the domains in the FLUF Test. However, this framework does not address prompting.

This is where the FLUF Test sets itself apart from these traditional information literacy frameworks. The indicators of the FLUF Test assist the user with prompt writing, as well as identifying elements to be analyzed in AI results. The FLUF Test was

designed to critique the output of a prompt in a large language model (LLM) like Microsoft Copilot™, an image generated with DALL-E3™, or multimedia content (e.g. videos and presentations).

The FLUF Test Framework

The indicators for Format, Language, Usability, and Fanfare (FLUF) are described in more detail below:

The “F” in FLUF stands for Format, where the user analyzes Layout or Length of the AI generative results. A FLUF infraction is charged when information does not display as intended.

The “L” in FLUF stands for Language, assessing the Tone, Phrasing, and Repetition. When examining Tone, the user considers whether the writing is the appropriate style. Phrasing should be clear without awkward presentation of information. Finally, consider whether there is a repetition of filling in the word count or space. A FLUF infraction occurs when information is not communicated clearly.

The “U” in FLUF is for Usability and assesses Credibility and Consistency. When the user analyzes the AI-generated results, they should check that the information is valid and reliable, appropriately cited, and from a reliable source. AI tools often present “hallucinations” by portraying false information as true. When sources cannot be validated or deemed reliable, a FLUF infraction is charged.

The last “F” in FLUF is for Fanfare, where the user analyzes the generative response to see if it addresses the audience using anecdotes or jargon and whether the level or mix of technical vocabulary is appropriate for the given context. A FLUF infraction occurs when the information is not presented appropriately for the intended audience or purpose.

Using the FLUF Test prompt template, the user creates an initial prompt using the “who, what, where, when, why, and how,” as well as the “Format, Language, Usability, and Fanfare” of the intended outcome (Figure 2). After generating results, the user moves on to the FLUF Test rubric and begins identifying the list of possible infractions and scores the AI output accordingly (Figure 1). The FLUF Test uses a simple rubric of plus (+) or minus (-), which translates to a binary score of zero or one. Each issue or infraction found with an AI generative result is assessed a “plus” or receives a score of one. The total infractions are tallied to get the FLUF score. The goal is to have zero, or no, infractions, and therefore “zero FLUF.” Users are encouraged to re-prompt, regenerate, repeat, or intervene manually until the AI output achieves a score of zero.

Figure 1

The Format, Language, Usability, and Fanfare (FLUF) Test Framework and Scoring Rubric

FLUF Test Rubric

Use the FLUF Test indicators to examine your AI Generated results. If you have infractions, regenerate until you reach "Zero FLUF".

Indicator	Description of Infractions	GOAL: ZERO FLUF	
		Plus = yes, there are infractions/fail	Minus = no infractions/pass
FORMAT			
Layout		+/-	
Length		+/-	
LANGUAGE			
Tone		+/-	
Phrasing		+/-	
Repetition		+/-	
USABILITY			
Consistency		+/-	
Credibility		+/-	
FANFARE			
Anecdotes		+/-	
Jargon		+/-	
Add up the "+" to see the issues with the AI generated content.		____/____	

Another view of the FLUF Test is below:

FORMAT		LANGUAGE			USABILITY		FANFARE		FLUF SCORE (GOAL: ZERO FLUF)	
Layout (+/-)	Length (+/-)	Tone (+/-)	Phrasing (+/-)	Repetition (+/-)	Credibility (+/-)	Consistency (+/-)	Anecdotes (+/-)	Jargon (+/-)	+Plus = yes, there are infractions or errors present	-Minus = no, there are no infractions present
SCORE:										



Creating an AI Generated Prompt using the FLUF Test © 2023 by [Dr. Jennifer](#)

Parker-licensed under [CC BY-NC-SA4.0](#)

Figure 2

The Format, Language, Usability, and Fanfare (FLUF) Test Framework Prompt Template

Create a "Zero FLUF" Prompt

Your search results are only as robust as the prompt provided to the tool. To get better results, include all essential information. Using the elements below, add the prompt information that the bot/AI Tool will need to generate robust content.

Prompt Element	Prompt information
Who	
What	
Where	
When	
Why	
How	
Format	
Layout	
Usability	
Fanfare	
<i>From the information above, generate the prompt</i>	



Creating an AI Generated Prompt using the FLUF Test © 2023 by [Dr. Jennifer](#)

Parker-licensed under [CC BY-NC-SA4.0](#)

Validation of the FLUF Test

The FLUF Test has been previously validated. The process included expert review for content validity, pilot testing, test-retest reliability, factor analysis, and field testing over two years across several cross-sections of participants. This included both qualitative and quantitative feedback on usefulness and implementation from experts in library science, K12 educators, graduate students, and professors in higher education. (Parker, 2023)

Why Embed Information Literacy into Pharmacy Education?

Pharmacy students are called upon daily to practice information literacy. This can be seen in the gathering of essential information around patients, prescriptions, interactions, and dosages to name just a few. While many turn to technology as part of the search process, information can be false or misleading. A career in pharmacy comes with a responsibility for accuracy in prescribing and accountability in dispensing. Human oversight and thoughtful critique should be embedded in the work. Teaching Doctor of Pharmacy (Pharm.D.) students about the power of AI, while also encouraging them to question their AI results, will encourage development of their information literacy skills.

Evidence from a scoping review demonstrates growing interest in the use of AI in pharmacy education, and the quest to incorporate the elements of responsible, ethical, and appropriate use (Arksey & O'Malley, 2005). AI is increasingly recognized for its transformative potential in pharmacy education. It has the potential to transform pharmacy education by enhancing clinical decision-making, promoting better learning outcomes, and improving guideline development (Weidmann, 2024; Mortlock & Lucas, 2024; Cain et al., 2023). However, alongside these advancements, there is a growing emphasis on the ethical and responsible use of AI. Scholars consistently call for clear guidelines and comprehensive training to address concerns such as academic integrity, potential plagiarism, and the possible erosion of critical thinking skills (Weidmann, 2024; Mortlock & Lucas, 2024; Hasan et al., 2023). Interestingly, AI is also viewed as a tool to foster critical thinking and analytical skills, which are essential for students preparing for clinical practice (Mortlock & Lucas, 2024; Busch et al., 2023; Hasan et al., 2024).

There is a notable shift in educational priorities, with increasing emphasis on moving beyond traditional content-heavy coursework toward practical application and experiential learning, leveraging AI to support these goals (Busch et al., 2023; Cain et al., 2023; Nakagawa et al., 2022). Student perceptions of AI are generally positive, with many recognizing its benefits in enhancing clinical decision-making and patient care (Busch et al., 2023; Zhang et al., 2024; Al-Ghazali, 2025). However, regional and demographic differences in how AI is perceived and utilized suggest the need for tailored educational approaches that address the diverse needs and experiences of students (Jaber et al., 2024; Hasan et al., 2023).

As more and more pharmacists consider how to incorporate AI into their daily lives, the importance of analysis and critical thinking rises to the forefront. More research is needed in the application of AI in pharmacy education, and the educational activities and frameworks that can be utilized to improve critical thinking and analysis. This study integrated AI and the FLUF Test into a Pharm.D. program to contribute to the growing body of research in this area

Methods

This was a mixed-method study evaluating the use of the FLUF Test in a pharmacy education setting. To answer the research questions, student results of the FLUF Test were evaluated quantitatively to summarize the most frequently identified

infractions with AI-generated outputs, while open-ended student survey responses on the usefulness and limitations of the FLUF Test were analyzed qualitatively.

Participants

This study included twenty-one third year Pharm.D. students enrolled in the Critical Care Pharmacy Elective course. Students were not previously exposed to assignments using AI tools in the pharmacy curriculum. Therefore, faculty implemented a pre-assignment survey to assess students' baseline knowledge and use of AI (N = 20/21; response rate = 95.2%). This information was used to describe the cohort of students in relation to their baseline knowledge and use of AI applications.

Research Methods

Quantitative data was collected as part of a DI assignment through the FLUF Test instrument rubric in Qualtrics™. Qualitative data was collected from open-ended student survey response regarding student perceptions and feedback around experiences using the FLUF Test.

Quantitative Phase

A quantitative analysis was done to assess the usability of the FLUF Test. Students were asked to identify and describe areas of concern. Descriptive statistics were used to summarize the areas of the FLUF Test domains (Format, Language, Usability, and Fanfare) where students identified infractions. Two indicators were assessed under the domains of Format (layout and length), Usability (consistency and credibility), and Fanfare (anecdotes and jargon). Three indicators were reviewed under the Language domain (tone, phrasing, and repetition). These data were summarized quantitatively for each indicator and domain, reported as both a number and a percent of the total infractions.

Qualitative Phase

To gain a deeper understanding of students' experiences and perceptions using the FLUF Test, a qualitative, thematic analysis of the open-ended survey responses of feedback on the FLUF Test was conducted in Qualtrics™ and imported into Microsoft Excel to establish themes. The open-ended survey questions asked were: (Q1) What other limitations of the AI-generated output were identified that were not included in the FLUF Test? (Q2) Briefly reflect on the usefulness of the FLUF Test in evaluating AI-generated responses to case-based questions (e.g., strengths, weaknesses, utility, etc.).

Qualitative analysis involved the use of Creswell's (2016) narrative analysis of themes approach. The qualitative analysis involved a multi-step process. First, each of the four investigators—two pharmacy faculty and two instructional designers—individually reviewed the responses and identified preliminary codes and themes. Next, the investigators convened as a group to discuss and refine their individual findings and agree upon the final codes and themes. The primary purpose of this qualitative analysis was to explore students' perceptions and experiences of using the FLUF Test, and to identify themes that could inform improvements to the framework.

Setting and Educational Activity

This pilot study assessed the use of the FLUF Test in evaluating AI outputs generated from a DI assignment within a Pharm.D. critical care elective course that incorporated the use of AI. A pharmacy DI question involves responding to inquiries about a medication's uses, side effects, or interactions. These questions often include clinical controversies that require the recipient to research the literature for answers. The DI question prompts are provided in Table 1. The assignment was designed to meet three objectives: 1) enhance students' DI skills, 2) build students' AI information literacy, and 3) write an evidence-based

response to a DI question. Historically, third year Pharm.D. students in the course completed a team-based assignment preparing a response to a DI question. Spring 2024 was the first semester that the assignment was updated to incorporate AI and use of the FLUF Test critical evaluation framework.

Table 1

Drug Information Question Prompts Provided to Students

Question/Prompt	
1	Act as a critical care clinical pharmacist rounding with a multidisciplinary medical intensive care team. You are caring for a patient with alcohol withdrawal syndrome and are curious about using phenobarbital for treatment of alcohol withdrawal. Summarize the evidence to support the use of phenobarbital for alcohol withdrawal syndrome and provide a specific recommendation for phenobarbital dose, frequency, duration, and monitoring parameters to treat alcohol withdrawal syndrome.
2	Act as a critical care clinical pharmacist rounding with a multidisciplinary trauma intensive care team. You are caring for a patient who presents emergently following a motor vehicle collision. The patient's injuries include: traumatic brain injury, multiple rib fractures, and a femur fracture. The patient is experiencing intracranial hypertension due to their severe traumatic brain injury. Summarize the evidence for hyperosmolar therapy (mannitol versus hypertonic sodium solutions) to treat intracranial hypertension. Provide a specific recommendation for the dose, route, frequency, duration, and monitoring parameters of hyperosmolar therapy to treat intracranial hypertension.
3	Act as a critical care clinical pharmacist rounding with a multidisciplinary medical intensive care team. You are caring for a patient who presents to the ICU from the hospital floor for acute hypoxic respiratory failure secondary to hospital acquired pneumonia. The patient has been hospitalized for 16 days. Pneumonia developed on hospital day 12. The patient is intubated upon arrival to the ICU. Several hours later the patient's respiratory status declines and ventilator settings are increased. The patient is diagnosed with severe acute respiratory distress syndrome. Summarize the evidence for the use of steroids to treat acute respiratory distress syndrome and provide a specific dosing recommendation.

Students utilized the FLUF Test in completing the DI assignment, which followed the instructions summarized in Table 2. First, students were individually assigned one of three possible DI questions, which were provided to students in the format of an AI prompt. They began by prompting Microsoft Copilot™ to generate a response to the assigned DI question. In the second step, students applied the FLUF Test to critically evaluate the AI-generated output. Next, they conducted their own DI search, building upon the AI-generated content. Students were then grouped into teams of 6-7 students—each working with the same prompt—to synthesize their individual responses and research into a collective DI submission. Finally, students reflected on the assignment and the role of AI in DI. All steps were completed individually by the student, except for the final DI question submission. This encouraged students to compare information from AI and literature searches and to collaborate on a final recommendation.

Table 2

Overview of Assignment Steps and Instructions for a Drug Information Assignment that Incorporated Generative Artificial Intelligence and the FLUF Test in a Pharmacy Elective Course

Step	Drug Information/AI Assignment	Percent of the Final Course Grade
-------------	---------------------------------------	--

1	Create an AI-generated response to a drug information question using a AI tool (Individual assignment)	–
2	Critique AI-generated response using the FLUF Test Framework (Individual assignment)	5%
3	Describe your drug information search process (Individual assignment)	C/I
4	Submit final drug information response (Team assignment)	20%
5	Reflect – What does this mean to me and the future of drug information? (Individual assignment)	10%

Key: AI = artificial intelligence, FLUF = Format, language, usability, and fanfare, C/I = complete/incomplete

The instructional resources and assignments were organized and delivered through the Canvas Learning Management System (LMS). A dedicated module was developed to provide students with a comprehensive overview of the learning objectives, a detailed breakdown of assignment expectations, and respective due dates. This module also featured an introductory video that explained the FLUF Test framework and demonstrated its application in practice. To ensure clarity and transparency, the assignment pages within the LMS were designed using the Transparency in Learning and Teaching (TILT) model (Winkelmes, 2023). Each assignment page outlined the purpose, tasks, and assessment criteria, along with providing a FLUF Test template and guiding prompts to facilitate student understanding and execution. A rubric was utilized to grade the FLUF Test submission as part of the DI assignment and accounted for 5% of the total course grade. (Table 3)

Table 3

Grading Rubric Utilized for Assessing the FLUF Test Survey Submission

Item	Critical Evaluation	Developing Critical Evaluation
1	All elements are assessed (2 point)	Elements of the FLUF test are missing (1 point)
2	Comments are valid and included for any infractions identified. Specific examples are discussed. (2 points)	Comments are NOT valid and/or missing for identified infractions and/or specific examples are not discussed. (1 point)
4	AI-generated output is submitted (2 point)	AI-generated output is NOT submitted (0 points)
TOTAL Possible Points = 6 points		

Failure to submit the assignment will result in a score of zero "0".

Students were asked to use Microsoft Copilot™ as the preferred AI tool, as the University of Florida provides all students with accounts on this platform, allowing them to chat securely when logged in with their university credentials. Additionally, given that the Pharm.D. curriculum emphasizes team-based learning (TBL) as a core instructional strategy, this approach was incorporated into the assignment where students worked collaboratively to synthesize their individual insights into a team submission. This collaborative element not only reinforced key competencies in teamwork but also mirrored real-world practice, where interdisciplinary collaboration is essential in healthcare.

FLUF Test Instrument

To facilitate the AI output critique step (Step 2), the FLUF Test was adapted into the online survey tool Qualtrics™. In Qualtrics™, students identified the number of infractions in the AI-generated responses and provided open-ended examples to illustrate these issues. Since this was the first pilot of the FLUF Test in this assignment, students were asked to respond to two open-ended reflection questions after inputting results from their content critique: (Q1) What other limitations of the AI-generated output were identified that were not included in the FLUF Test? (Q2) Briefly reflect on the usefulness of the FLUF Test in evaluating AI-generated responses to case-based questions (e.g., strengths, weaknesses, utility, etc.). The FLUF Test step was completed before students proceeded with their own DI question literature searches and developed their final responses. Results from utilizing the FLUF Test (Table 2; Step 2) are reported in this study to describe the applicability of using the FLUF Test in this context.

The FLUF Test was used to guide students in evaluating the AI-generated content. It was chosen because it specifically addresses challenges in using and applying AI-generated content. Although the FLUF Test is not tailored to pharmacy education or DI questions, it provides a broad approach for AI users to evaluate content. This makes it applicable across multiple disciplines and contexts. However, since the FLUF Test had not been previously used in pharmacy education, the investigators aimed to gather information about its applicability to a DI question assignment. This information will inform future iterations of the assignment and use of the FLUF Test in the course.

Results

Participant AI Experience

On the initial pre-assignment survey, the majority of students (70%, 14/20) indicated that they had explored AI and used it occasionally. A smaller group (15%, 3/20) reported never using AI, while one student (5%, 1/20) used AI daily, and two students (10%, 2/20) used AI several times a month. Most students (70%, 14/20) stated that they had never used AI for schoolwork, assignments, or work-related tasks. However, 25% (5/20) mentioned using AI occasionally for these purposes, and one student (5%, 1/20) reported using AI tools several times a month for coursework or work.

Research Question: What FLUF infractions are identified by students when the FLUF Test framework is applied to a DI assignment?

All twenty-one students (100%) completed the FLUF Test critique and received full credit using the grading rubric. Collectively, students identified issues across all domains of the FLUF Test when assessing their AI-generated responses. The most frequent concern was in the credibility section under the usability domain, with 61.9% (13/21) of students noting infractions. This was followed by the fanfare section, where 42.9% (9/21) of students observed that the AI outputs lacked specific examples or comparisons needed to answer the questions. Additionally, 33.3% (7/21) of students identified issues with the language of the AI-generated text, particularly regarding tone or style, and another 33.3% (7/21) identified problems with phrasing or the presentation of information. Further details on the issues identified by the students using the FLUF Test are provided in Table 4, including examples identified by students to illustrate the concern. Appendix I provides a sample of an AI-generated output that was evaluated using the FLUF Test by a student.

Table 4

Summary of the Number of Infractions Identified by Students using the FLUF Test Indicators along with Selected Quotes from Students Describing an Example to Support the Selected Rating

FLUF Test Domain	Infraction Identified – YES N (%)	Sample Quotes Describing the Infractions Identified by Students
Format		
Layout: Doesn't follow formal writing patterns or formats	4 (19)	“The AI failed to write the response at a level consistent with a pharmacist answering a drug information question. It provided a brief overview of the topics in a bulleted format without expanding further on any research conducted to reach those conclusions.”
Length: Lots of extra works to extend the word count	1 (4.8)	
Language		
Tone: Lacks personal style, tone, or human elements	7 (33.3)	“The response lacks personal style, using a generic third-person perspective. For example, the phrasing for the recommendation is "a specific recommendation for hyperosmolar therapy would be:" as opposed to something like "I recommend..."”
Phrasing: Syntax, semantics, or awkward presentation of information	7 (33.3)	
Repetition: Lacks succinct presentation of ideas; run-ons or repetition of ideas, thoughts, or phrases	3 (14.3)	
Usability		
Consistency: Inconsistencies in content	2 (9.5)	“While it provided its citations, the citations were largely guidelines that had final recommendations of low quality due to the lack of evidence. The AI failed to delve further into primary literature sources and expanded upon the recommendations. It only provided information on what treatment regimen to choose, but no further explanation of why one was preferred over the other.”
Credibility: credible references cannot be determined or validated or lack sources/citations for information	13 (61.9)	
Fanfare		
Anecdotes: Lack human stories or examples that provide classifications, comparisons, metaphors, or analogies	9 (42.9)	“The "frequency" suggestion is a repetition from the "dosing" suggestion; there was no need to a different heading for frequency as we can see from the dosing that Phenobarbital maintenance dose should be given every 30 minutes.”
Jargon: Repeats vocabulary, technical language, and common findings	6 (28.6)	

Key: FLUF = Format, Language, Usability, Fanfare

Research Question: What are pharmacy student perceptions of the application and usability of the FLUF test framework within a DI assignment?

Seven students (33.3%) provided feedback on limitations not captured by the FLUF Test in response to an optional survey question. The main theme from their responses was that the FLUF Test did not prompt students to evaluate the specificity and depth of AI responses. When asked to reflect on the usefulness of the FLUF Test, all twenty-one students (100%) provided feedback, identifying both strengths and weaknesses of the evaluation framework. Three themes emerged: 1) the FLUF Test was useful, 2) the FLUF Test was easy to use, and 3) human elements are required. Students recognized the need for human input when interacting, noting challenges in trusting the information and highlighting that applying the FLUF Test requires subjectivity and contextual relevance. A detailed codebook outlining the themes, codes, definitions, and examples from analysis of the open-ended student survey responses is presented in Table 5.

Table 5

Codebook with Themes, Codes, Definitions, and Example Quotes from Students' Open-ended Survey Responses regarding the Utilization of the FLUF Test to Prepare a Drug Information Response

Item	Theme	Code	Definition	Quotes
Q1	Lack of specificity and depth	Limitations and challenges	Students struggled with how to identify FLUF Test infractions due to inexperience with AI.	"Unfortunately, I am not familiar enough with ChatGPT [AI] to evaluate what is a "normal" response and what is not. It would have been helped to see an example with no FLUF content, and one with errors in it."
Q2	Usefulness, applicability, contextual relevance	Framework for Evaluation	The FLUF Test can serve as a framework for interacting with AI, but perspectives and expertise of the user can influence AI experiences.	"The FLUF Test was useful to help guide me on what to expect from an AI response." "The FLUF Test prompted me to delve into references used to generate the AI response, and not just take it at face value." "By completing the [FLUF Test] evaluation, I also get an idea of what I should change about the response [reprompt] to make it useable in a clinical scenario."
Q2	Ease of Use	Quick Starting Point and Guide	The FLUF Test was both useful and easy to use.	"It seems like a quick way to evaluate an AI response, there aren't too many items, and the items that are there all seem important enough to include." "I think the FLUF Test is an excellent initial framework to evaluate AI-generated responses quickly."
Q2	Human Element	Challenges trust and requires subjectivity and	The FLUF Test integrates human prompting and challenges blind trust of AI	"I was unaware on the capabilities of what to expect from its [AI] response and initially viewed the very basic information as adequate, similar in line to a Google

contextual
relevanceoutput through critical
thinking.Search. Utilizing the FLUF Test I was able to
raise my expectations of the response."

Key: FLUF = Format, Language, Usability, Fanfare, AI = artificial intelligence

Through the FLUF Test evaluation process, students assessed each element of the framework, Table 6 provides quotes from students on the application and usefulness of the FLUF Test domains to evaluate AI-generated outputs and how they applied the framework to identify issues, further supporting its usefulness as a tool.

Table 6

Example Quotes from Students on how the FLUF Test domains were Applicable to Evaluate AI-Generated Outputs

FLUF Test Domain	Domain Definition	Examples
Format	Proper Display	"It [The FLUF Test] analyzes the format and style of how AI generates information and responds to the user, identifying how natural the outputs seemed to read."
Language	Clear Communication	"It also made me cognizant of the fact that AI-generated questions may have repetitions, and sometimes unnecessary information, so one has to utilize AI only as a starting point, and then try to paraphrase the verified key details."
Usability	Valid, reliable, and accurate	"I also appreciate how it asks about sources in both simple presence and quality." "The FLUF Test helped me to take a deeper dive into the response that AI generated. I was prompted to click on the links provided and evaluate the sources that AI used and their credibility."
Fanfare	Messaging appropriate for the intended audience.	"The FLUF Test showed that AI can be used to expand further on topics and provide a near human/professional response." "The FLUF Test reminds us to consider the [intended] audience... Like overly formal and business like writing for informational presentations...or jargon may be acceptable depending on the level of familiarity of the audience."

Key: FLUF = Format, Language, Usability, Fanfare, AI = Artificial Intelligence

Discussion and Analysis of the Utility of the FLUF Test

This pilot study, which marked the first evaluation of the FLUF Test in pharmacy education, examined the usefulness and student perceptions of using the framework to evaluate AI responses to case-based DI questions. The findings suggest that the FLUF Test provides an easy-to-use framework for identifying key infractions of AI responses. Students noted that the FLUF Test was a valuable tool, providing a starting point for evaluating AI-generated content. The FLUF Test guided students to the importance of human intervention in the AI process, including re-prompting, review, and analysis. Specifically, students were

guided to validate results, identify credible resources, detect hallucinations, and recognize repetition using the FLUF Test. In addition, students autonomously completed the output critique, implementing real-time feedback to improve the final response. Finally, the study also found that through application of the framework students recognized the importance of critically evaluating AI outputs.

The findings of this study contribute to the growing body of literature on integrating AI into pharmacy education, as there is currently limited guidance on integrating AI literacy into pharmacy curricula (Accreditation Council for Pharmacy Education, 2025). While prior studies have discussed the potential benefits and concerns of AI use in academic settings (Cain et al., 2023; Mortlock & Lucas, 2024), our results extend this work by showing how a structured, general-purpose framework like the FLUF Test can help students critically assess AI-generated responses in a clinically oriented assignment. Students highlighted the value of the FLUF Test for evaluating the unique qualities of AI-generated content which is an element missing from traditional information literacy models such as CRAAP or SIFT (Blakeslee, 2004; Caulfield, 2019). While the American Library Association's ROBOT Test offers a starting point for AI evaluation (American Library Association, 2023), students in our cohort particularly valued the FLUF Test's emphasis on re-prompting, a step they found essential for refining AI outputs. The results suggest that pharmacy curricula would benefit from integrating AI-specific frameworks like the FLUF Test and providing students with structured opportunities to iterate, critique, and apply AI in clinically relevant contexts. These findings also align with recent calls to equip students with tools to engage more thoughtfully and ethically with AI technologies in health professions education (Hasan et al., 2023; Weidmann, 2024).

Student Evaluation of the FLUF Test

As part of this pilot study, students provided feedback on potential areas for improvement of the FLUF Test (Q1). Specifically, as noted in the results, seven students (33.3%) noted that the FLUF Test did not capture the lack of specificity and depth in the AI responses. This limitation may be attributed to the assignment's requirement for detailed responses to patient-case scenarios, which may not be applicable to all AI prompts. Alternatively, students' limited experience with the FLUF Test may have contributed, as they may not have recognized that issues with specificity and depth could be addressed under the "Fanfare" indicator, where the use of specialized jargon or medical terminology could enhance responses. This finding highlights the need for clearer understanding and training around how to apply the "Fanfare" domain, as well as potentially providing more detailed descriptions or examples of errors to look for when introducing students to the FLUF Test.

When students reflected on the usefulness of the FLUF Test in evaluating AI responses to case-based questions (Q2), three key themes emerged. Specifically, students found the FLUF Test to be a useful tool and easy to use, providing a framework for resource evaluation. They also identified the importance of the "human element" when interacting with AI, and the students acknowledged the value of the FLUF Test as a starting point for critically evaluating AI results. Notably, under the "human element" theme the concept of "trust" was frequently mentioned, as students began to question their trust in AI results after being exposed to the FLUF Test. Initially, some students may have unknowingly accepted AI output without realizing the potential for hallucinations or, alternatively, automatically discredited AI as a valuable tool or resource. This finding highlights a key advantage of using the FLUF Test, illustrating how pharmacy education can equip students to critically evaluate and ethically apply AI-generated information (Weidman, 2024; Mortlock & Lucas, 2024; Hasan et al., 2023).

The FLUF Test's quick and easy-to-use design makes it an ideal framework for novice AI users to consider when evaluating information and resources utilized by AI. This is particularly important in health professions, where evidence-based evaluation and application are essential skills for patient care. Given AI's lack of contextual understanding, human prompting and critical evaluation of results become even more crucial when working with case-based scenarios. Many students appreciated the guidance provided by the FLUF Test in validating output and assessing the credibility and quality of AI-generated information, recognizing that this skill is vital in their profession. Therefore, identifying ways in which AI tools can support the development of students' analytical skills is of utmost importance (Mortlock & Lucas, 2024; Busch et al., 2023; Hasan et al., 2024).

However, students noted that the FLUF Test's application involves some subjectivity and requires experience to use effectively. As many students in the course had limited or no background in using AI, they felt that they lacked context in distinguishing between "good" and "poor" results. Consequently, they shared that more experience with AI may aid them in identifying FLUF Test infractions.

Implementation Recommendations for Instructors

Instructors and instructional designers interested in adopting the use of the FLUF test in an AI-based assignment are encouraged to plan intentionally for implementation, tailoring it to their specific instructional context. A critical first step is to select and test an appropriate AI platform. We recommend choosing a platform that aligns with institutional policies, is accessible to students, and is commonly used within the discipline. We chose to utilize Microsoft Copilot™ since it was provided by our university and approved by our information technology team. Once selected, instructors should create a centralized assignment information page in the LMS. This page should clearly outline the full scope and step-by-step expectations of the assignment and provide access to all necessary resources, including the FLUF Test template and the introductory instructional video.

Effective facilitation strategies will vary depending on the course format and structure. In our flipped classroom curriculum, the faculty leveraged an asynchronous video to introduce the FLUF Test, walk students through its use, and present examples of successful implementation. This approach allowed students to revisit explanations as needed and engage with the assignment more independently. Additionally, transparent grading rubrics were created in Canvas™, which helped students understand expectations and allowed for more consistent assessment. For submission of the FLUF Test template, the survey tool Qualtrics™ was utilized which streamlined both student submission and instructor access to results for review and analysis. This approach could also easily be employed in the large classroom setting or online environments.

A key implementation challenge encountered with the assignment involved student misinterpretation of the AI prompt instructions. Despite providing a sample prompt to copy and paste, some students attempted to write their own, which undermined the consistency of the results. To mitigate this issue, providing explicit, step-by-step instructions and emphasizing the importance of using the exact prompt provided is strongly recommended. Highlighting common missteps during class or in the assignment instructions can also help preempt confusion.

While developed in a pharmacy education context, this assignment is adaptable across disciplines that engage students in information synthesis, critical thinking, or AI literacy. For example, in writing-intensive humanities courses, students could use the FLUF Test approach to evaluate AI-generated thesis statements or essay drafts. In Science, Technology, Engineering, and Mathematics (STEM) fields, the same structure could be applied to have students assess AI-generated explanations of scientific concepts. Modifications to the prompt and evaluation criteria can align the activity with discipline-specific goals while maintaining the core critical evaluation framework.

Finally, for assessment and feedback, combining quantitative rubric-based evaluation with qualitative self-reflection is recommended. Students submitted a brief reflection on their experience using the AI tool and what they learned about both the content and the capabilities/limitations of AI. This reflective component allowed instructors to assess metacognitive engagement and provided valuable feedback to inform future iterations of the assignment.

Limitations

This study is not without limitations. This evaluation included a small sample size of twenty-one students in an elective course at a single college of pharmacy, which limits the generalizability of the findings. In addition, not all students responded to the optional open-ended survey questions, decreasing the sample size. Future research is needed to assess the FLUF Test in larger, more diverse health professions education settings to better understand the applicability of the framework for other patient care assignments. Additionally, students included in this cohort were mainly novice AI users with limited experience

using and applying AI tools, which may have impacted results and their perceptions on using the FLUF Test. Students also identified infractions from different domains using the same prompt; however, this is likely because the output changes for each user, even when the same prompt is utilized. In addition, some students did not follow the assignment instructions and created their own prompts to answer the DI questions which could have impacted the quality of results.

Future Directions

Overall, the findings of this study provide valuable insights into the integration of AI evaluation tools, such as the FLUF Test, in educational settings. To further refine the use of the FLUF Test, future directions should focus on developing more comprehensive training and education when the FLUF Test is employed. This could include providing real-world examples of common issues with AI, such as hallucinations and inaccuracies, and clearly illustrating how the FLUF Test can be applied to address these issues. Additionally, course instructors plan to emphasize prompt generation and re-prompting in future iterations of the assignment. This was part of the instructions for using the FLUF Test but was not emphasized in the assignment instructions or evaluation rubric. Instructors are also challenged to keep up with the rapid advancement of AI tools, which may impact the experience and usefulness of the FLUF Test in the future. Finally, the FLUF Test should be evaluated in other contexts, such as implementation in a large-classroom or online settings.

Conclusion

The study's findings highlight the importance of teaching students to effectively use and evaluate AI, emphasizing the critical evaluation and judgment skills required to navigate the complexities of AI-generated content. By doing so, educators can empower students to make informed decisions on how to use and apply information obtained by AI tools.

The FLUF Test demonstrated its value as a practical tool in guiding students through the critical evaluation of AI-generated outputs. By focusing on Format, Language, Usability, and Fanfare, the FLUF Test not only helped the Pharm.D. students identify and critique issues in AI responses but also emphasized the importance of validating AI-generated content for credibility and quality. This exercise highlighted the need for a human element in both crafting prompts and critically analyzing AI outputs, underscoring that AI tools require thoughtful interaction and scrutiny.

However, participants also pointed out limitations of using the FLUF Test. The subjective nature of the FLUF Test, influenced by the individual user's expertise and experience with AI, made it more challenging for some students—particularly those with limited exposure to AI—to accurately apply the framework. As a result, novice users may struggle with recognizing AI flaws or crafting effective prompts, which affects the overall quality of AI output. Despite these challenges, most students found the FLUF Test to be a helpful guide for engaging with AI tools and appreciated its role in guiding their approach to critical evaluation.

References

- Accreditation Council for Pharmacy Education. (2025). *Accreditation standards and key elements for the professional program in pharmacy leading to the Doctor of Pharmacy degree ("Standards 2025")*. <https://www.acpe-accredit.org/pdf/ACPEStandards2025.pdf>
- Association of College and Research Libraries. (2016). *Framework for information literacy for higher education*. https://www.ala.org/sites/default/files/acrl/content/issues/infolit/Framework_ILHE.pdf.

- Al-Ghazali, M. A. (2025). Evaluation of Awareness, Perception and Opinions Toward Artificial Intelligence Among Pharmacy Students. *Hospital Pharmacy*, 00185787251326227.
- American Library Association. (2023). *ROBOT framework for ethical AI integration in libraries*. American Library Association, Center for the Future of Libraries. <https://www.ala.org>.
- Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, 8(1), 19-32.
- Balnaves, E., Bultrini, L., Cox, A., & Uzwysyn, R. (Eds.). (2024). *New Horizons in Artificial Intelligence in Libraries* (Vol. 185). Walter de Gruyter GmbH & Co KG.
- Blakeslee, S., & Meriam Library (2004). The CRAAP test. California State University, Chico.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. <https://arxiv.org/abs/2005.14165>
- Bruce, C. S. (1997). *The seven faces of information literacy*. Auslib Press.
- Busch, F., Hoffmann, L., Truhn, D., Palaian, S., Alomar, M., Shpati, K., Makowski, M. R., Bressemer, K. K., & Adams, L. C. (2023). International pharmacy students' perceptions towards artificial intelligence in medicine: A multinational, multicentre cross-sectional study. *British Journal of Clinical Pharmacology*, 90(3), 649-661.
- Cain J, Malcom DR, Aungst TD. (2023) The role of artificial intelligence in the future of pharmacy education. *American Journal of Pharm Educ*, 87(10):100135. doi: 10.1016/j.ajpe.2023.100135.
- Calvani, A., Cartelli, A., Fini, A., & Ranieri, M. (2008). Models and instruments for assessing digital competence at school. *Journal of E-learning and Knowledge Society*, 4(3), 183-193.
- Carroll, A. J., & Borycz, J. (2024). Integrating large language models and generative artificial intelligence tools into information literacy instruction. *The Journal of Academic Librarianship*, 50(4), 102899.
- Caulfield, M. (2019, June 19). *SIFT* (The Four Moves) [Blog post]. <https://happgood.us/2019/06/19/sift-the-four-moves/>. Licensed under CC BY 4.0.
- Chan, C.K.Y. (2023) A comprehensive AI policy education framework for university teaching and learning. *Int J Educ Technol High Educ* 20, 38. <https://doi.org/10.1186/s41239-023-00408-3>
- Civil, B. (2023, March 16). ChatGPT can hinder students' critical thinking skills: Artificial intelligence is changing how students learn to write. *The Queen's Journal*. <https://www.queensjournal.ca/story/2023-03-16/opinions/chatgpt-can-hinder-students-critical-thinking-skills/>.
- Creswell, J. W., & Poth, C. N. (2016). *Qualitative inquiry and research design: Choosing among five approaches*. Sage publications.
- Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: the state of the field. *International Journal of Educational Technology in Higher Education*, 20(1), 22.
- Eshet, Y. (2004). Digital literacy: A conceptual framework for survival skills in the digital era. *Journal of educational multimedia and hypermedia*, 13(1), 93-106.

- Giray, L. (2023). Prompt engineering with ChatGPT: a guide for academic writers. *Annals of biomedical engineering*, 51(12), 2629-2633.
- Hasan, H. E., Jaber, D., Khabour, O. F., & Alzoubi, K. H. (2024). Perspectives of pharmacy students on ethical issues related to artificial intelligence: a comprehensive survey study. *Research Square*, rs-3.
- Hasan, H. E., Jaber, D., AlTabbah, S., Lawand, N., Habib, H. A., & Farahat, N. M. (2024). Knowledge, attitude, and practice towards artificial intelligence in pharmacy practice. *PLOS ONE*, 29 (6884), e0296884.
- Heston, T. F., & Khun, C. (2023). Prompt engineering in medical education. *International Medical Education*, 2(3), 198-205.
- Intelligent.com. (2023, January 23). Nearly 1 in 3 college students have used ChatGPT on written assignments. <https://www.intelligent.com/nearly-1-in-3-college-students-have-used-chatgpt-on-written-assignments/>
- International Society for Technology in Education. (n.d.). ISTE standards for educators. <https://www.iste.org/standards/for-educators>.
- Jaber, D., Mosleh, R., Khashan, H., & Abu Khalaf, N. (2023). Medicine and pharmacy students' knowledge, attitudes, and practice regarding artificial intelligence programs: Jordan and West Bank of Palestine. *Advances in Medical Education and Practice*, 14(1), e1179-7258.
- Korzynski, P., Mazurek, G., Krzyrkowska, P., & Kurasinski, A. (2023). Artificial intelligence prompt engineering as a new digital competence: Analysis of generative AI technologies such as ChatGPT. *Entrepreneurial Business and Economics Review*, 11(3), 25-37.
- Kuhlthau, C. C. (2004). *Seeking meaning: A process approach to library and information services*. Libraries Unlimited.
- Long, D., & Magerko, B. (2020, April). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1-16).
- Maleki, N., Padmanabhan, B., & Dutta, K. (2024, June). AI hallucinations: a misnomer worth clarifying. In *2024 IEEE Conference on Artificial Intelligence (CAI)* (pp. 133-138). IEEE.
- Meskó, B. (2023). Prompt engineering as an important emerging skill for medical professionals: tutorial. *Journal of Medical Internet Research*, 25, e50638.
- Microsoft. (2023). Microsoft Copilot [Software]. *Microsoft*. <https://www.microsoft.com/en-us/microsoft-365/copilot>
- Mishra, S., Khashabi, D., Baral, C., & Hajishirzi, H. (2021). *Cross-Task Generalization via Natural Language Crowdsourcing Instructions*. <https://arxiv.org/abs/2104.08773>
- Mollick E and Mollick L. (2023). *Assigning AI: Seven Approaches for Students, with Prompts*. The Wharton School Research Paper. Retrieved December 18, 2024, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4475995.
- Mortlock, R., & Lucas, C. (2024). Generative artificial intelligence (Gen-AI) in pharmacy education: Utilization and implications for academic integrity: A scoping review. *Exploratory Research in Clinical and Social Pharmacy*, 100481.
- Nakagawa, N., Odanaka, K., Ohara, H., & Kisara, S. (2022). Communication training for pharmacy students with standard patients using artificial intelligence. *Currents in Pharmacy Teaching and Learning*, 14(7), 854-862.

- Ndungu, M. W. (2024). Integrating basic artificial intelligence literacy into media and information literacy programs in higher education: A framework for librarians and educators. *Journal of Information Literacy*, 18(2), 122-139.
- OpenAI. (2023). DALL-E 3 [Software]. OpenAI. <https://www.openai.com/dall-e-3>.
- Parker, J. (2023). *The FLUF Test*. <https://www.drjenniferparker.com/fluf-test-for-artificial-intelligence.html>.
- Qualtrics. (2023). Qualtrics [Software]. *Qualtrics*. <https://www.qualtrics.com>.
- Short, C. E., & Short, J. C. (2023). The artificially intelligent entrepreneur: ChatGPT, prompt engineering, and entrepreneurial rhetoric creation. *Journal of Business Venturing Insights*, 19, e00388.
- Sye, D. & Thompson, D. (2023). Tools, tests, and checklists: The evolution and future of source evaluation frameworks. *Journal of New Librarianship*, 8(1), 76–100.
- Thoman, E., & Jolls, T. (2003). Literacy for the 21st Century: An overview & orientation [*Media kit*]. Center for Media Literacy. <http://www.medialit.org>.
- Tyton Partners, (2023). *Gen AI in higher education: Fall 2023 time for class update*. <https://tytonpartners.com/app/uploads/2023/10/GenAI-IN-HIGHER-EDUCATION-FALL-2023-UPDATE-TIME-FOR-CLASS-STUDY.pdf>
- Valenza, J. (2004). Never Ending Search [blog]. CARRDSS framework. *School Library Journal*.
- Warschauer, M., Tseng, W., Yim, S., Webster, T., Jacob, S., Du, Q., & Tate, T. (2023). *The affordances and contradictions of AI-generated text for second language writers*. <https://doi.org/10.2139/ssrn.4404380>.
- Wei, J., Wang, X., Schuurmans, D., Bosma, M., Chi, E., Le, Q., & Zhou, D. (2022). *Chain-of-Thought Prompting Elicits Reasoning in Large Language Models*. <https://arxiv.org/abs/2201.11903>
- Weidmann, A. E. (2024). Artificial intelligence in academic writing and clinical pharmacy education: Consequences and opportunities. *International Journal of Clinical Pharmacy*, 46(3), 751-754.
- Winkelmes, M. A. (2023). Introduction: The story of TILT and its emerging uses in higher education. *Transparent design in higher education teaching and leadership*, 1-14.
- Zhang, X., Tsang, C. C. S., Ford, D. D., & Winnik, J. (2024). Student pharmacists' perceptions of artificial intelligence and machine learning in pharmacy practice and pharmacy education. *American Journal of Pharmaceutical Education*, 88(1), 101309.

APPENDIX I

Sample AI-generated Output in Response to a Drug Information Question Prompt and Student Scoring using the FLUF-Test Instrument

Sample Response

Certainly! Let's discuss the evidence and recommendations for hyperosmolar therapy in the management of intracranial hypertension.

1. Hyperosmolar Agents:

1. Hyperosmolar agents are commonly used to reduce intracranial pressure (ICP) and cerebral edema in patients with traumatic brain injury (TBI) and other neurological insults.
2. The two main hyperosmolar agents used are mannitol and hypertonic saline (HTS).

2. Mannitol:

1. Mechanism of Action: Mannitol is an osmotic diuretic that works by drawing water out of brain tissue, reducing cerebral edema and ICP
2. Dose and Route:
 1. Dose: Typically administered as a bolus (e.g., 0.25–1 g/kg) intravenously.
 2. Route: Intravenous (IV) administration.
 3. Frequency: Mannitol can be repeated every 4–6 hours as needed.
 4. Duration: Use for short-term management during acute episodes of elevated ICP.
5. Monitoring Parameters:
 1. Serum Osmolarity: Monitor serum osmolarity to avoid excessive elevation (>320 mOsm/L).
 2. Renal Function: Mannitol can precipitate acute renal failure if serum osmolarity is too high.

3. Hypertonic Saline (HTS):

1. Mechanism of Action: HTS works by osmotic shift of water from brain tissue into the intravascular space, reducing cerebral edema and ICP.
1. Dose and Route:
 1. Dose: Typically administered as a bolus (e.g., 3% or 23.4% HTS) intravenously.
 2. Route: Intravenous (IV) administration.
 3. Frequency: HTS can be repeated every 4–6 hours as needed.
 4. Duration: Similar to mannitol, use for short-term management during acute episodes of elevated ICP.
5. Monitoring Parameters:
 1. Serum Sodium Levels: Monitor serum sodium levels to avoid excessive elevation (>160–180 mEq/L).
 2. Neurologic, Cardiac, and Renal Function: Regular monitoring for adverse effects.

4. Recommendation:

1. Intracerebral Hemorrhage (ICH): We suggest using hypertonic sodium solutions (HTS) over mannitol for managing cerebral edema in patients with ICH (conditional recommendation, low-quality evidence).
2. Rationale: Although the quality of evidence is very low, HTS may provide a more robust and durable effect in lowering ICP compared to mannitol.

Remember to closely monitor the patient's response and adjust therapy as needed based on their clinical condition. Individualized care and close collaboration with the multidisciplinary team are essential for optimal management.

Student Use of the FLUF Test to Score the AI-generated Output

Indicator	Plus	Minus	Description of Infractions from the Student
Format – Layout	X		"The AI failed to write the response at a level consistent with a pharmacist answering a drug information question. It provided a brief overview of the topics in a bulleted format without expanding further on any research conducted to reach those conclusions."
Format – Length		X	

Language – Tone	X	“While copying and pasting the answer from the guidelines, it utilized the terminology “we suggest” which was an obvious error in the response and highlighted it copying and pasting answers from other sources.”
Language - Phrasing	X	
Language - Repetition	X	
Usability – Consistency	X	“While it provided its citations, the citations were largely guidelines that had final recommendations of low quality due to the lack of evidence. The AI failed to delve further into primary literature sources and expand upon the recommendations. It only provided information on what treatment regimen to choose, but no further explanation of why one was preferred over the other.”
Usability – Credibility	X	
Fanfare – Anecdotes	X	“Similar to the usability, it failed to provide any sort of helpful information and just quoted directly from the guidelines.”
Fanfare - Jargon	X	
TOTAL FLUF Score	4	