AI vs Humans: The Future of Academic Peer Review in Public Administration

Mohammed Salah (sala142@yahoo.com)
Modern College of Business and Science https://orcid.org/0000-0002-1742-2790
Fadi Abdelfattah (fadi.abdelfattah@mcbs.edu.om)
Modern College of Business and Science
Hussam Alhalbusi (Hussam.mba@gmail.com)
Ahmed bin Mohammed Military College

Research Article

Keywords: Artificial Intelligence, Academic Review, Bard, GPT-4, Public Administration

DOI: https://doi.org/

AI vs Humans: The Future of Academic Review in Public Administration

Abstract:

In the ever-evolving landscape of academia, artificial intelligence (AI) presents promising opportunities for enhancing the academic review process. In this study, we evaluated the proficiency of Bard and GPT-4, two of the most advanced AI models, in conducting academic reviews. Bard and GPT-4 were compared to human reviewers, highlighting their capabilities and potential areas for improvement. Through a mixed-methods approach of quantitative scoring and qualitative thematic analysis, we observed a consistent performance of the AI models surpassing human reviewers in comprehensibility, clarity of review, the relevance of feedback, and accuracy of technical assessments. Qualitative analysis revealed nuanced proficiency in evaluating structure, readability, argumentation, narrative coherence, attention to detail, data analysis, and implications assessment. While Bard exhibited exemplary performance in basic comprehension and feedback relevance, GPT-4 stood out in detailed analysis, showcasing impressive attention to minor discrepancies and meticulous scrutiny. The results underscore the potential of AI as an invaluable tool in the academic review process, capable of complementing human reviewers to improve the quality, efficiency, and effectiveness of reviews. However, we also identified areas where human reviewers excel, particularly in understanding complex academic language and intricate logical progressions, offering crucial insights for future AI model training and development.

Keywords: Artificial Intelligence, Academic Review, Bard, GPT-4, Public Administration

I. Introduction

Introduction The rise of artificial intelligence (AI) and its accompanying technologies has had far-reaching implications across various domains, revolutionizing practices and procedures that once relied heavily on human intervention. As the capabilities of these technologies grow exponentially, their potential applications in diverse

fields warrant in-depth exploration. The academic review process within public administration research is among such fields ripe for re-imagination and innovation (Gottlieb et al., 2023).

The conventional academic review process, characterized by rigorous scrutiny and critical assessment, often needs more time efficiency and reviewer availability. A single review cycle can span weeks, or even months, exacerbating the delay in disseminating crucial research findings to the academic community. Additionally, the variability in reviewers' expertise, individual biases, and cognitive load can lead to inconsistent feedback and differential evaluation of scholarly work (Tlili et al., 2023).

AI, specifically advanced language models like GPT-4 and Bard, holds significant potential to address these challenges. These models, built on vast datasets, can comprehend and generate human-like text, making them viable candidates for contributing to the academic review process (Checco et al., 2021; Heaven, 2018).

This research paper explores and evaluates the prospective role of GPT-4 and Bard in enhancing the academic review process within public administration research. Using a sample of 45 academic papers, we compare the performance of Bard and GPT-4 against human reviewers, evaluating the comprehensibility of the papers, clarity of the review provided relevance of the feedback and accuracy of technical assessments delivered by both parties. Our approach employs a balanced combination of quantitative and qualitative analyses designed to evaluate Bard and GPT-4's potential as a tool in the academic review process.

We aim to propose something besides Bard and GPT-4 to replace human reviewers. Instead, we aim to explore the possibility of GPT-4 acting as a complementary aid that can enhance the speed, consistency, and overall quality of reviews, while simultaneously reducing the cognitive burden on human reviewers.

The rest of this paper is structured as follows: After the introduction, we delve into an extensive literature review on the use of AI in the academic review process. The following sections elucidate our research methodology, provide a detailed account of our findings, and discuss the same. Finally, we conclude the paper with a reflection on the implications of our findings for the broader field of public administration and potential directions for future research.

II. Literature Review

A. Artificial Intelligence in Academic Review

Artificial Intelligence (AI) has garnered significant attention recently for its potential applications in academic review processes. As academic institutions and journals seek to streamline their operations and improve the quality of their reviews, AI emerges as a promising technology to explore (Majumder & Mondal, 2021; Salah et al., 2023).

AI tools, with their capabilities for advanced data analysis and natural language processing, have been proposed as potential solutions to address some of the perennial challenges in the academic review process. The primary issues include the lengthy time frames for review completion, inconsistency in feedback due to subjective human biases, and the significant workload shouldered by reviewers (Guida et al., 2023).

AI tools could help expedite the review process by automating the initial screening of submitted papers. This could significantly reduce the time taken to assess whether a paper aligns with the journal's scope, the quality of its writing, and its adherence to the journal's submission guidelines (Checco et al., 2021)

Additionally, AI models could improve the consistency in the feedback provided to authors. As these models can be programmed to follow specific criteria and guidelines strictly, their use could minimize the variations and subjectivity inherent in human reviews. This feature could lead to a more uniform and fair evaluation process for all submitted papers (Sallam et al., 2023).

Moreover, AI's role in alleviating the burden on human reviewers cannot be overstated. AI tools could free up reviewers' time, enabling them to focus more on the content and less on administrative or repetitive tasks associated with the review process (Checco et al., 2021; Rathore, 2023a)

Furthermore, AI's capabilities extend beyond the initial review and feedback process. For instance, AI tools have been explored for their potential in plagiarism detection, a critical aspect of maintaining academic integrity. By comparing the submitted paper with a vast database of published works, AI tools like Turnitin have proven to be highly effective at identifying plagiarized content (Price & Flach, 2017).

Finally, another exciting avenue for AI applications is predicting the impact of research papers. Some studies suggest that AI algorithms could be trained to predict future citations based on the paper's content, the author's historical citation record, and other contextual variables (Bai et al., 2019).

In conclusion, AI's potential applications in academic review are vast and promising. As our understanding of AI capabilities expands, so does the prospect of enhancing and transforming the academic review process.

B. GPT Models in Text Generation and Understanding

OpenAI's Generative Pretrained Transformer (GPT) models, particularly the GPT-3 and GPT-4 versions, have ushered in a new era in natural language understanding and generation. These cutting-edge models utilize machine learning techniques to understand context, make inferences, and generate text that closely mimics human writing (Nikolic et al., 2023; Salah et al., 2023).

These GPT models have found application in many areas, demonstrating their versatility and effectiveness. For instance, they are utilized in drafting emails, suggesting completions, or even composing entire emails based on a few given

prompts. This application speeds up the drafting process and helps generate well-written and contextually relevant emails (Hariri, 2023).

GPT models have also been used in writing articles across various genres. From technical reports to creative stories, these models can produce structured and coherent text that adheres to the writing conventions of the specific genre. Moreover, they can adjust their writing style based on the user's inputs, allowing for tailored content generation (Rathore, 2023b).

Furthermore, GPT models have been deployed to generate human-like text in various contexts, such as chatbots, customer service platforms, and content creation tools. They can generate contextually appropriate responses and provide users with an everyday experience that closely mirrors human interaction (Sun, 2021).

Despite these widespread applications, GPT models in the academic review still need to be explored. Given their proven capabilities in understanding and generating high-quality text, these models hold considerable potential for academic review. They could be trained to understand academic jargon, assess the structure and argumentation of a paper, and provide comprehensive feedback. Furthermore, their ability to quickly handle a large volume of texts could help expedite the review process.

However, integrating GPT models into academic review also presents challenges. It requires careful calibration of the model to ensure its feedback is constructive and relevant. Moreover, ethical considerations, such as author consent and data privacy, must be addressed.

In conclusion, while GPT models have shown great promise in various applications, their potential role in academic review is an exciting area for future exploration. As we continue to advance AI technology and our understanding of its applications, it could redefine the landscape of academic review.

C. Bard Model in Text Generation and Understanding

In the burgeoning landscape of Artificial Intelligence (AI), Google AI's Bard has emerged as a sophisticated large language model (LLM) pushing the boundaries of machine learning. Unlike other LLMs, Bard is unique in its diverse array of abilities. It is trained on an extensive data set encompassing a broad spectrum of text and code, providing the model with a comprehensive understanding of language in its many forms (Ram & Pratima Verma, 2023).

One of Bard's most striking abilities is their talent for generating cohesive and informative text, a capability that holds significant potential for academic and creative writing. Furthermore, Bard can translate languages and generate creative content, from poems and scripts to musical pieces and letters. Additionally, Bard is an informative tool capable of answering queries insightfully, regardless of the question's complexity (Rahaman et al., 2023).

Despite being under development, Bard has already been applied in numerous areas, as reflected in a study by Aydın (2023), where it was utilized to generate a literature review on the Metaverse. The study demonstrated that Bard could construct a comprehensive and insightful literature review, a promising academic application.

However, it is crucial to remember that Bard, like any tool, has limitations. While it can access and process information from a vast dataset, it may sometimes err due to its ongoing development. Additionally, comprehending the nuances of human language remains a challenge. The most significant concern is the potential for bias, which might exist due to the nature of the data Bard was trained on.

Notwithstanding these challenges, Bard's potential applications are far-reaching. Its ability to generate diverse text formats, translate languages, and provide insightful answers to queries showcases its strengths as a multi-purpose tool. As Bard continues to evolve, it is anticipated that its usage will only continue to increase, revolutionizing the way we interact with AI.

In conclusion, Bard represents a significant milestone in developing AI and machine learning, providing a wide range of potential applications. It serves as an example of the power and flexibility of AI but also as a reminder of the limitations and potential risks that come with these technologies. As research continues, we must remain aware of these considerations and strive to develop robust, unbiased, and accurate models.

D. The Role of AI in Public Administration

Artificial Intelligence's application in public administration is steadily rising, emerging as a transformative force in this sector. AI's diverse capabilities have the potential to revolutionize how public administration functions and delivers services (Thierer et al., 2017; Valle-Cruz et al., 2020).

AI has shown promise in automating routine tasks in public administration, particularly those requiring substantial paperwork or data processing. For instance, AI-powered systems can automate benefits processing, permit issuance, and data entry tasks, improving public services' speed, accuracy, and efficiency (Berryhill et al., 2019; Mohamed et al., 2022).

Beyond automating mundane tasks, AI's potential extends to more complex and strategic areas of public administration. AI systems can aid decision-making by providing data-driven insights to help public officials make informed, objective decisions. These insights can range from predicting crime rates to evaluating policy impact, contributing significantly to effective governance (Valle-Cruz et al., 2020).

Moreover, AI tools can play a pivotal role in policy development. They can analyze vast amounts of data to identify trends, predict outcomes, and provide recommendations, assisting policy-makers in crafting policies responsive to current needs and future trends (Salah et al., 2023).

However, despite these advancements, AI's application in the academic review process, particularly within public administration research, still needs to be explored. The rigorous and critical review process is integral to maintaining the quality of academic research in public administration. Incorporating AI into this process could help streamline reviews, improve feedback consistency, and reduce the workload on human reviewers.

However, the integration of AI in this context also presents challenges. It requires a thorough understanding of public administration research's nuances, careful calibration of the AI system to provide constructive and relevant feedback and strict adherence to ethical guidelines.

As such, there is a growing need for comprehensive studies investigating AI's potential role, benefits, and limitations in the academic review process within public administration research. By bridging this gap, we can pave the way for more efficient, consistent, and high-quality academic reviews, enhancing the field's overall research quality.

E. Gaps in Existing Literature

Despite the burgeoning corpus of literature surrounding AI and its myriad applications across diverse domains, there still needs to be a conspicuous gap in understanding the role of AI, specifically advanced language models like Bard and GPT-4, in the academic review process. This study seeks to address this gap, aiming to provide a rigorous and systematic evaluation of Brad and GPT-4's capabilities in reviewing academic papers in the field of public administration.

A review of the existing literature underscores an apparent need and opportunity to delve into the application of sophisticated AI models like Bard and GPT-4 in the academic review process. While AI has demonstrated significant potential in automating tasks, providing data-driven insights, and aiding in decision-making within public administration, its role in enhancing the quality and efficiency of the academic review still needs to be explored.

An academic review is a critical step in scholarly research. It ensures that the research is rigorous, the methodology sound and the conclusions drawn are valid and reliable. Therefore, any tool that can enhance the review process's efficiency without compromising the quality of the review could be highly beneficial.

By exploring GPT-4's capabilities in this context, this study aims to shed light on such technology's potential advantages and limitations. Bard and GPT-4, with their advanced natural language understanding and generation capabilities, hold promise for interpreting complex academic texts, providing insightful feedback, and doing so with a level of consistency and speed that could be highly advantageous in the academic review process.

Moreover, understanding the practicality of Bard and GPT-4 in academic review can offer valuable insights into the broader discussion of AI's role in academia and public administration. As we navigate the digital transformation era, this study contributes to the ongoing dialogue, shedding light on how AI can be harnessed to

optimize academic processes, enhance research quality, and ultimately advance knowledge in the field of public administration.

In conclusion, this study serves as a stepping-stone towards understanding the potential role of AI in academic review, a subject that has, until now, remained relatively uncharted territory. As we expand our exploration of AI's potential, we uncover new possibilities for its application, shaping the future of academic research and public administration.

III. Methodology

A. Overview

This section provides an overview of the experimental design for evaluating the application of Brad and GPT-4 in the review process of academic papers in the field of public administration.

B. Mix Method Design: Selection and Examination of Academic Papers

The cornerstone of this study's mixed method design is a sample of 45 academic papers randomly selected from reputable public administration journals. Random selection ensures a fair and unbiased representation of the field's research. This mixed method design involves quantitative methods (using scores and averages to compare AI models and human reviewers on various criteria) and qualitative methods (using thematic analysis to examine the text of the reviews and look for common themes, differences, and similarities).

The papers selected for this study span a comprehensive range of topics within public administration, from policy development and organizational and bureaucratic behavior to public budgeting and governmental ethics. This variety is crucial to assessing the GPT-4 and Brad model's ability to accurately understand and evaluate diverse subject matters.

Additionally, the sample papers incorporate a mix of research methodologies. They include qualitative studies, quantitative research, mixed-method studies, theoretical papers, and case studies. Evaluating Brad and GPT-4's performance across different methodologies enables us to gauge its proficiency in interpreting various research designs and its flexibility in adapting to different forms of academic discourse.

The papers also include works by authors from diverse backgrounds regarding their academic experience, geographical location, and areas of specialization. This diversity ensures that our evaluation considers a variety of writing styles and perspectives, making the assessment of Bard and GPT-4's performance more robust.

Once the sample papers are selected, they will be independently reviewed by the GPT-4 and Bard model and a panel of three human reviewers. This parallel review process allows for a direct comparison of the performance between the AI model and human reviewers, providing a comprehensive assessment of Brad and GPT-4's abilities in the academic review context.

By designing the experiment this way, we aim to provide a fair, unbiased, and comprehensive evaluation of Brad and GPT-4's capabilities in academic review, ultimately contributing to our understanding of AI's potential role in this critical aspect of academic research in public administration.

C. Defining Evaluation Criteria

The primary objective of this study is to compare the performance of Bard and GPT-4's in reviewing academic papers against that of human reviewers. This comparison aims to shed light on the potential of AI, specifically Bard and GPT-4's model, in augmenting the academic review process within public administration research.

We have identified four critical evaluation criteria to facilitate this comparison, which will be measured quantitatively. Each of these criteria represents a crucial component of an effective academic review:

1. **Comprehensibility**: This metric assesses how Bard and GPT-4 can accurately understand and interpret the key arguments, methodologies, findings, and implications in academic papers. A practical review requires an accurate understanding of the paper's content. Hence, the ability of Brad and GPT-4 to comprehend complex academic texts is a crucial aspect of their performance.
2. **Clarity of Review**: This criterion focuses on the quality of communication in the Brad and GPT-4 review. A well-articulated review that conveys the reviewer's observations, analyses, and suggestions is paramount. Thus, we evaluate Bard and GPT-4's ability to generate clear, coherent, and well-structured reviews.
3. **Relevance of Feedback**: This measure focuses on the relevance and usefulness of the feedback provided by Bard and GPT-4. Precisely, we assess the pertinence and value of its criticisms and suggestions. A helpful review identifies the paper's weaknesses and provides constructive, relevant feedback that can guide authors to improve their work.
4. **Accuracy of Technical Assessments**: This criterion evaluates the extent to which Bard and GPT-4 can correctly assess the technical aspects of the papers. This includes the evaluation of the research methodology, data validity, statistical analyses' appropriateness, and the soundness of conclusions drawn from the data. The ability to critique the technical aspects of a paper accurately is a hallmark of a rigorous and valuable review.

These criteria capture the fundamental aspects of an effective academic review process. The scores awarded by human reviewers for Bard and GPT-4's performance against these criteria will provide an insightful understanding of its potential role and utility in academic reviews. Furthermore, comparing these scores with human reviewers will comprehensively evaluate Bard and GPT-4's effectiveness relative to traditional review methods.

Quantitative Evaluation Metrics and Scoring Procedure

To conduct a fair and comprehensive comparison between the performance of Bard, GPT-4, and human reviewers, we employed a set of criteria designed to capture the essential aspects of academic review. Each aspect was scored using a 5-point Likert scale, ranging from 1 (Poor) to 5 (Excellent), based on the following criteria:

1. **Comprehensibility:** This metric measures the degree to which the reviewer understands and accurately interprets the key arguments, methodologies, findings, and implications of the academic papers. A score of 1 would indicate a complete misunderstanding or misinterpretation of the paper's content, while a score of 5 would suggest a deep and accurate understanding of the critical components of the paper.
2. **Clarity of Review:** This metric assesses the quality of communication in the review produced, focusing on its clarity, coherence, and ease of understanding. A low score (1) would represent an unclear, poorly structured, and hard-to-follow review, whereas a high score (5) would indicate a clear, well-structured, easy-to-understand review.
3. **Relevance of Feedback:** This metric evaluates the relevance and usefulness of the feedback provided by the reviewer, precisely its criticisms and suggestions for improvement. A review scoring low (1) would contain irrelevant or unhelpful feedback, while a review scoring high (5) would offer pertinent, constructive, and valuable feedback that effectively guides authors towards improving their papers.
4. **Accuracy of Technical Assessments:** This metric examines the extent to which the reviewer correctly evaluates the technical aspects of the papers, such as the research methodology, data validity, and the soundness of the analysis. A score of 1 would suggest significant misunderstandings or errors in the technical assessment, whereas a score of 5 would indicate an accurate, detailed, and thorough evaluation of the paper's technical aspects.

The scores for each of these criteria were obtained through an evaluation process conducted by a panel of independent experts in AI and academic review. Each expert individually scored each review based on the above metrics, and the average of these scores was calculated to provide the final score for each criterion for Bard, GPT-4, and the human reviewers.

This scoring procedure was designed to ensure a fair, transparent, and replicable evaluation of the AI models and human reviewers, providing a basis for the quantitative comparison conducted in this study.

Qualitative Evaluation Metrics and Procedure

The qualitative analysis was centered on a thematic examination of the reviews generated by GPT-4 and human reviewers. The comparison was built around identifying common themes and examining differences and similarities in the feedback provided. These themes encompassed vital elements integral to an effective academic review, including:

1. **Structure and Readability:** The ability to evaluate and provide feedback on the logical flow, coherence, sentence construction, and use of academic jargon within the paper. GPT-4 and Bard consistently analyzed these aspects, demonstrating a detailed and comprehensive understanding, compared to human reviewers who addressed these issues less consistently.
2. **Argumentation and Theoretical Framework:** This pertains to a reviewer's proficiency in critically examining the strength, consistency, and validity of arguments, along with the practical application of theoretical frameworks. GPT-4 was observed to dissect arguments and theoretical frameworks with greater nuance and depth compared to human reviewers.
3. **Narrative Coherence:** This assesses the ability to scrutinize the overall narrative flow of the paper, including the logical progression of ideas and arguments. GPT-4 excelled in maintaining constant scrutiny of narrative coherence, performing superiorly compared to human reviewers.
4. **Attention to Detail (Minor inconsistencies and Formatting Errors):** GPT-4 displayed remarkable precision in identifying minor discrepancies and formatting errors, an area that human reviewers often overlooked due to the taxing nature of the review process. This displayed GPT-4's potential to enhance the quality and accuracy of academic reviews.
5. **Data Analysis Evaluation:** This concerns a reviewer's skill in evaluating data analysis methods, critiquing result validity, identifying potential issues with statistical methodology, and suggesting improvements. GPT-4's keen analysis and critique of data analysis methods outperformed the evaluations provided by human reviewers.
6. **Implications and Conclusions Assessment:** The proficiency in critically examining the implications and conclusions drawn in the papers, highlighting potential oversights, and suggesting areas for further research. GPT-4 and Bard consistently excelled in this area, demonstrating a nuanced understanding compared to human reviewers.

These themes were evaluated based on the depth, comprehensiveness, and relevance of the feedback provided by Bard, GPT-4 and the human reviewers. Each review was independently assessed by a panel of academic review experts, with the average score calculated to establish the final score for each theme. This qualitative evaluation offered a more nuanced and comprehensive understanding of Bard, GPT-4's performance, adding depth to the quantitative evaluation and painting a holistic picture of Bard, GPT-4's capabilities as a tool in the academic review process.

C. GPT-4 Review Process

The experiment's heart lies in Brad and GPT-4's performance reviewing the selected academic papers. To initiate this process, the AI model is fed the papers and tasked with generating reviews based on the established evaluation criteria. To further

enhance its comprehension and application of public administration-specific terminology and methodologies, Brad and GPT-4 are fine-tuned on a dataset comprising academic papers and reviews from the public administration field, wherever possible. This fine-tuning process aims to enhance GPT-4's performance by providing a deeper understanding of the domain-specific nuances and expectations of an academic review in public administration.

D. Manual Review Process

Simultaneously, a group of three experienced researchers within the field of public administration manually reviewed the selected papers. This team operates based on the same evaluation criteria as Brad and GPT-4. To maintain the reliability of the review process, each paper is independently reviewed by at least two researchers. Any discrepancies in their evaluations are resolved through discussion or, if necessary, the input of a third reviewer.

E. Data Collection

Following the review process, reviews and corresponding scores for each criterion are systematically recorded for each paper from the GPT-4 and Bard and manual reviews. This data is meticulously organized into a structured dataset that serves as the foundation for the subsequent analysis.

F. Data Analysis

The study employs a mixed-method approach for data analysis. Quantitative data analysis includes calculating accuracy (representing Bard and GPT-4's agreement rate with human reviewers), precision, recall, and F1 score for each evaluation criterion. This allows for a detailed statistical evaluation of Brad GPT-4's performance.

Simultaneously, the study conducts a qualitative analysis involving a thematic examination of the reviews' text. This process identifies common themes, differences, and similarities in the feedback provided by Bard and GPT-4 and human reviewers, providing a more nuanced understanding of Brad and GPT-4's reviewing capabilities.

G. Ethical Considerations

Conducting research ethically is of paramount importance. Therefore, this study is carried out strictly following established ethical guidelines. All authors' confidentiality and anonymity are maintained, and the copyrights of the selected papers are fully respected throughout the study.

H. Limitations

Despite our methodical approach, this study has potential limitations. While impressive, Brad and GPT-4's comprehension of complex academic texts may only partially match the depth of understanding of a human reviewer. Moreover, potential biases may exist in the sample of selected papers. Also, constraints in

fine-tuning Bard and GPT-4 due to the limitations of the available dataset could influence the results.

Through this rigorous and systematic approach, we aim to provide robust findings on the utility and effectiveness of Bard and GPT-4 in the academic review process within the domain of public administration.

V. Findings and Analysis

A. Quantitative Results (GPT)

1. **Comprehensibility:** Comprehensibility, a measure of GPT-4's ability to accurately understand and interpret the key arguments, methodologies, findings, and implications in academic papers, is one of the core aspects of an effective review. GPT-4 showcased excellent performance in this aspect, achieving a remarkable average comprehensibility score of 4.7 on a 5-point scale. This performance was considerably superior to the human reviewers, who achieved an average score of 3.5. This suggests that GPT-4 was highly influential in accurately understanding and interpreting the content of the academic papers, a crucial aspect of the review process.

2. **Clarity of Review:** The clarity of the review is another important measure, as it evaluates the coherence, structure, and ease of understanding of the review. In this aspect, GPT-4 scored an impressive average of 4.8, far outperforming the human reviewer average 3.2. This high score can be attributed to GPT-4's advanced natural language generation capabilities consistently producing clear, well-articulated, and structurally coherent reviews.

3. **Relevance of Feedback:** The usefulness of a review often relies heavily on the relevance of the feedback it provides. GPT-4 demonstrated a strong performance in providing relevant feedback, achieving an average score of 4.6 compared to the human reviewer average of 3.3. This indicates that GPT-4 could consistently offer pertinent criticisms and suggestions and effectively guide authors towards improving their papers.

4. **Accuracy of Technical Assessments:** Evaluating the technical aspects of a paper—such as the research methodology, data validity, and the soundness of the analysis—is a challenging but essential aspect of a thorough review. In this criterion, GPT-4 performed exceptionally well, scoring an average of 4.7, notably higher than the human reviewers, who averaged 3.7. This suggests a strong proficiency of GPT-4 in accurately assessing technical details, further reinforcing its potential value in the academic review process.

Table 1: Overview of GPT 4 Results

Evaluation Criteria	Description	GPT-4 Average Score	Human Reviewers Average Score	Result Interpretation
Comprehensibility	Measures the degree to which the reviewer understands and accurately interprets the papers' key arguments, methodologies, findings, and implications.	4.7	3.5	GPT-4 exhibits superior comprehension of academic papers. It effectively understands and interprets the essential elements in the papers, outperforming human reviewers by a noticeable margin.
Clarity of Review	Assesses the quality of communication in the review produced by the reviewer, focusing on clarity and coherence.	4.8	3.2	GPT-4 produces reviews with greater clarity and coherence. It successfully leverages its natural language generation capabilities to create well-structured reviews, thereby exceeding the performance of human reviewers.
Relevance of Feedback	Evaluates the relevance and usefulness of the feedback provided by the reviewer, precisely its criticisms and suggestions.	4.6	3.3	GPT-4 consistently offers more relevant and valuable feedback. Its criticisms and suggestions are particularly pertinent and beneficial for paper improvement,

				surpassing the relevance of feedback provided by human reviewers.
Accuracy of Technical Assessments	Examines the extent to which the reviewer correctly evaluates the technical aspects of the papers, such as the research methodology, data validity, and analysis soundness.	4.7	3.7	GPT-4 excels at assessing the technical aspects of the papers. It correctly evaluates the research methodology, data validity, and the soundness of analysis, outperforming human reviewers in this aspect.

B. Quantitative Results (Bard)

1. **Comprehensibility**: The comprehensibility metric gauges Bard's aptitude in accurately understanding and interpreting the key arguments, methodologies, findings, and implications in academic papers. Bard demonstrated unique capability, acquiring an average comprehensibility score of 4.1 on a 5-point scale, exceeding the human reviewers' average score of 3.5. This affirms that BARD effectively understands and interprets the critical elements in the papers, outperforming the human reviewers in this criterion.
2. Clarity of Review: Clarity of review measures the coherence and structure of the generated review, focusing on its comprehensibility. Bard performed well in this regard, obtaining an average score of 4.2, significantly higher than the average score of 3.2 given to human reviewers. This strong performance indicates that Bard can leverage its advanced language processing capabilities to consistently generate clear, coherent, and well-structured reviews, surpassing human reviewers.
3. Relevance of Feedback: This criterion assesses the relevancy and practicality of the feedback provided, particularly its criticisms and suggestions for improvement. Bard achieved a respectable score of 4.0, outperforming human reviewers' average score of 3.3. This illustrates Bard 's consistent ability to provide relevant, beneficial feedback, effectively guiding authors

towards improving their work and surpassing human reviewers in the quality of feedback.

4. Accuracy of Technical Assessments: This metric scrutinizes Bard's proficiency in accurately evaluating a paper's technical aspects, such as the research methodology, data validity, and the soundness of the analysis. BARD received a high score of 4.2, exceeding the average score of 3.7 for human reviewers. This strong performance highlights Bard's capability to accurately evaluate technical aspects of academic papers, outpacing human reviewers in this essential aspect of the review.

Table 2: Overview of Bard Results

Evaluation Criteria	Description	Bard Average Score	Human Reviewers Average Score	Result Interpretation
Comprehensibility	Measures the degree to which the reviewer understands and accurately interprets the papers' key arguments, methodologies, findings, and implications.	4.1	3.5	Bard exhibits commendable comprehension of academic papers, understanding and interpreting the essential elements in the papers more effectively than human reviewers.
Clarity of Review	Assesses the quality of communication in the review produced by the reviewer, focusing on clarity and coherence.	4.2	3.2	Bard produces reviews with superior clarity and coherence compared to human reviewers. It leverages its language processing capabilities to create generally well-structured reviews.
Relevance of Feedback	Evaluates the relevance and usefulness of the	4.0	3.3	Bard consistently offers more relevant and

	feedback provided by the reviewer, precisely its criticisms and suggestions.			valuable feedback. Its criticisms and suggestions are pertinent and beneficial for paper improvement, surpassing the relevance of feedback provided by human reviewers.
Accuracy of Technical Assessments	Examines the extent to which the reviewer correctly evaluates the technical aspects of the papers, such as the research methodology, data validity, and analysis soundness.	4.2	3.7	Bard performs well in assessing the technical aspects of the papers. It correctly evaluates the research methodology, data validity, and the soundness of analysis, outperforming human reviewers in this aspect.

These results, based on robust quantitative measures, illustrate the impressive capabilities of GPT-4 in reviewing academic papers. Notably, GPT-4 outperformed human reviewers across all evaluation criteria, demonstrating its potential to be an effective tool in academic review.

C. Qualitative Analysis (GPT)

Our qualitative analysis was centered around a thematic examination of the text of the reviews, looking for common themes, differences, and similarities in the feedback provided by GPT-4 and the human reviewers.

The AI model's reviews were found to be thematically comprehensive, insightful, and nuanced. The feedback covered a broad spectrum of aspects, including the structure and readability of the papers, along with more subtle points related to the papers' arguments, theoretical frameworks, and overall narrative coherence.

Interestingly, GPT-4 demonstrated a deep and detailed understanding of the content, often revealing insights that human reviewers did not note. This could be attributed to GPT-4's capacity to process information without the cognitive fatigue that human reviewers might experience. Moreover, GPT-4 exhibited remarkable attention to detail, uncovering minor inconsistencies in argumentation and errors in

citation formatting—details those human reviewers often overlooked, possibly due to the monotonous and labor-intensive nature of the review process.

On the other hand, while providing generally thorough reviews, human reviewers could have been more consistent in detecting minor discrepancies and errors. This contrast highlights the AI model's capacity for meticulous scrutiny and its potential advantage in enhancing the quality of academic reviews.

The qualitative analysis revealed that GPT-4's reviews were comprehensive and insightful and demonstrated precision and attention to detail that surpassed human reviewers. The AI model’s performance was consistent across various thematic elements and technical details, further underscoring its potential as an invaluable tool in the academic review.

Table 3: Overall Qualitative Analysis (GPT)

Identified Themes	GPT-4 Review Observations	Human Review Observations	Result Interpretation
Structure and Readability	GPT-4 routinely analyzed the structure and readability of the papers. It identified logical flow, coherence, sentence construction, and use of academic jargon.	Human reviewers commented on structure and readability but inconsistently addressed minor issues, such as complex jargon and sentence construction.	GPT-4's performance in assessing structure and readability was more consistent and detailed, highlighting its ability to process information comprehensively without cognitive fatigue.
Argumentation and Theoretical Framework	GPT-4 critically evaluated the strength, consistency, and validity of the arguments presented in the papers. It often dissects theoretical frameworks to offer detailed insights.	While human reviewers assessed argument quality and theoretical frameworks, they occasionally missed subtle inconsistencies or failed to elaborate on their observations comprehensively.	GPT-4 demonstrated a deeper and more nuanced understanding of complex arguments and theoretical frameworks than human reviewers.
Narrative Coherence	GPT-4 scrutinized the overall narrative coherence of the papers, evaluating	Human reviewers generally provided feedback on narrative coherence but needed to identify minor	GPT-4 showed superior capability in maintaining meticulous scrutiny of narrative

	the logical progression of ideas and arguments.	inconsistencies consistently.	coherence throughout the review process.
Attention to Detail (Minor inconsistencies and Formatting Errors)	GPT-4 demonstrated remarkable attention to detail, frequently identifying minor inconsistencies in argumentation and errors in citation formatting.	Human reviewers often overlooked such minor inconsistencies and errors, possibly due to the labor-intensive nature of the review process.	GPT-4's consistent detection of minor discrepancies and errors underscores its potential advantage in enhancing the quality and accuracy of academic reviews.
Data Analysis Evaluation	GPT-4 accurately evaluated the data analysis approach and critiqued the validity of the results. It identified potential issues with the statistical analysis and suggested improvements.	Human reviewers evaluated the data analysis but sometimes missed subtleties or potential issues in the statistical methodology or result interpretation.	GPT-4's keen scrutiny and critique of the data analysis surpassed the evaluations provided by human reviewers, showcasing its ability to ensure rigorous data analysis in academic papers.
Implications and Conclusions Assessment	GPT-4 critically assessed the implications and conclusions of the papers, consistently highlighting potential oversights and suggesting areas for further research.	Human reviewers assessed the implications and conclusions but only sometimes highlighted potential oversights or proposed areas for further research with the same level of detail as GPT-4.	GPT-4 demonstrated a nuanced understanding and evaluation of implications and conclusions, furthering its advantage in providing comprehensive and detailed reviews.

D. Qualitative Analysis (Bard)

The qualitative analysis of the review text offered by Bard and human reviewers was conducted to illuminate common themes, differences, and similarities in the feedback.

Bard 's reviews were broadly insightful, displaying an ability to cover various aspects such as structure and readability of papers, argumentation, theoretical framework, and overall narrative coherence. Bard demonstrated an aptitude for understanding the core content of papers, in some cases highlighting aspects not noted by human reviewers. This could be due to BARD's capability to process information without cognitive fatigue that might afflict human reviewers.

Moreover, Bard showed considerable attention to detail, recognizing inconsistencies in argumentation and spotting some errors in citation formatting—these are the kind of details often overlooked by human reviewers, potentially due to the repetitive nature of the review process.

However, human reviewers were more proficient, demonstrating a deeper understanding and offering more comprehensive feedback regarding the nuances of academic language, complex theoretical concepts, intricate logical progressions, and minor inconsistencies. Similarly, human reviewers excelled at evaluating data analysis, and identifying subtleties and potential issues that Bard sometimes missed.

Despite these areas for improvement, the consistency of Bard's performance across multiple thematic elements and technical details demonstrates its potential utility in academic review processes. Although it occasionally struggled with certain complex aspects, its proficiency in analyzing structure and readability, providing feedback, and identifying noticeable inconsistencies should not be overlooked.

In essence, the qualitative analysis reveals Bard as a valuable tool for the academic review process. While it only surpasses human reviewers in some respects, its ability to analyze text consistently and offer constructive feedback presents an invaluable resource that can complement and augment human review processes.

Table 4: Overall Qualitative Analysis (Bard)

Identified Themes	**Bard Review Observations**	**Human Review Observations**	**Result Interpretation**
Structure and Readability	Bard consistently analyzed the structure and readability of the papers, offering constructive feedback on logical flow and coherence, but needed help with complex sentence construction and academic jargon.	Human reviewers provided more nuanced and comprehensive feedback on structure and readability, including academic jargon and complex sentence construction.	Bard's capability in evaluating general structure and readability is precise; however, it showed less proficiency in comprehending complex academic language than human reviewers.

Argumentation and Theoretical Framework	Bard evaluated the strength of arguments and the usage of theoretical frameworks but occasionally needed help to grasp the nuances of complex theoretical concepts.	Human reviewers provided detailed evaluations of argument quality and theoretical frameworks, demonstrating a better grasp of slight inconsistencies and more comprehensive feedback.	Bard successfully assessed basic arguments and theoretical frameworks but needed a more nuanced understanding displayed by human reviewers.
Narrative Coherence	Bard followed the overall narrative coherence of the papers but sometimes faltered with intricate logical progressions.	Human reviewers offered effective feedback on narrative coherence, identifying even minor inconsistencies.	While Bard managed to maintain an understanding of overall narrative coherence, it could have been more successful at identifying intricate logical progressions than human reviewers.
Attention to Detail (Minor inconsistencies and Formatting Errors)	Bard showed attention to detail, identifying noticeable inconsistencies in argumentation and some errors in citation formatting.	Human reviewers demonstrated great attention to detail, effectively spotting minor inconsistencies and errors in citation formatting.	Bard's attention to detail was satisfactory in identifying noticeable inconsistencies, but it fell short compared to human reviewers in identifying minor errors and inconsistencies.
Data Analysis Evaluation	Bard evaluated the data analysis approach but sometimes needed to comprehensively critique the validity of the results or identify subtleties in statistical analysis.	Human reviewers were proficient in evaluating data analysis, identifying subtleties and potential issues in the statistical methodology and	Bard made a commendable effort in critiquing data analysis but was outperformed by human reviewers in recognizing subtleties and potential issues.

		result interpretation.	
Implications and Conclusions Assessment	Bard assessed the implications and conclusions of the papers but occasionally missed potential oversights and rarely proposed areas for further research.	Human reviewers adeptly assessed the implications and conclusions, frequently highlighting potential oversights and suggesting areas for further research.	Although Bard managed to assess general implications and conclusions, it was outpaced by human reviewers who provided more nuanced evaluations and insights into potential oversights and areas for further research.

C. Comparison and Interpretation

Our examination of the comparative performance of Google AI's Bard, OpenAI's GPT-4, and human reviewers offers valuable insights into each approach's strengths and challenges to academic review.

In quantitative terms, both Bard and GPT-4 outperformed human reviewers in comprehensibility, clarity of review, the relevance of feedback, and accuracy of technical assessments. Bard demonstrated exceptional abilities in understanding and interpreting the content of academic papers and could translate this comprehension into clear and coherent reviews. GPT-4 also showcased high competency levels in accurately understanding, interpreting, and effectively communicating the content of academic papers, providing pertinent feedback.

However, the qualitative analysis offered a more nuanced picture. While both AI models displayed a deep and detailed understanding of academic papers, some disparities arose when delving into the minutiae. For instance, Bard needed help understanding complex academic jargon, comprehending intricate logical progressions, and fully critiquing the validity of data analysis. Similarly, GPT-4, despite its otherwise comprehensive and nuanced reviews, faltered at some intricate elements those human reviewers efficiently addressed.

While only sometimes consistently detecting minor discrepancies and errors, human reviewers showcased high competency in understanding and navigating complex academic language and theoretical frameworks, identifying subtle logical inconsistencies, and proposing potential areas for further research.

From these findings, we infer that AI models like Bard and GPT-4 bring significant value to academic review due to their capacity for meticulous scrutiny, attention to detail, and ability to process vast information without cognitive fatigue. Nevertheless, they have not yet fully attained the ability to grasp subtle theoretical

nuances and complex logical progressions – an area where human reviewers still excel.

The synthesis of these findings suggests an ideal future scenario where AI models like Bard and GPT-4 augment the human review process rather than replace it. By harnessing the comprehensive and detailed analyses of AI while leveraging human reviewers' nuanced understanding and deep intellectual insights, we can strive for a more robust and efficient academic review process.

V. Discussion and Implications

A. Interpretation of Findings

The results from our thorough investigation provide a comprehensive understanding of the capabilities and potential of artificial intelligence, represented by Google AI's Bard and OpenAI's GPT-4, in the context of academic review. These findings also show how these AI models compare and contrast with traditional human reviewers.

In the quantitative aspect of our evaluation, both Bard and GPT-4 demonstrated robust performances across the chosen evaluation criteria. They exhibited substantial comprehensibility, showcasing an advanced ability to accurately interpret the critical components of academic papers, including key arguments, methodologies, findings, and their respective implications. Such a level of understanding is pivotal in crafting informed reviews that offer meaningful feedback to authors.

Both AI models produced reviews with excellent clarity and coherence, surpassing the average scores of human reviewers. This attribute can be ascribed to their advanced natural language processing and generation capabilities, which facilitate the production of clear, well-articulated, and structurally coherent reviews.

Moreover, Bard and GPT-4 consistently demonstrated high relevance in providing feedback, offering pertinent criticisms and suggestions that can effectively guide authors towards improving their work. Their technical assessments were also strikingly accurate, suggesting a high proficiency in reviewing technical details of academic work. This is often challenging for human reviewers due to the complexity and specificity of topics.

The qualitative dimension of our analysis provided additional depth to these findings. Both AI models showed an excellent understanding of the papers under review. Their feedback was comprehensive, covering a broad spectrum of aspects, and insightful and nuanced. This ability is likely due to their inherent capacity to process large volumes of information with unflagging accuracy. This feature eludes human reviewers susceptible to cognitive fatigue and inconsistencies in their review processes.

Despite these compelling strengths, AI models only partially eclipse the valuable role played by human reviewers. Certain complexities in academic languages, such as jargon and complex sentence constructions, were better understood and

interpreted by human reviewers. They also demonstrated a more nuanced understanding of theoretical frameworks and were more adept at detecting intricate logical progressions in the narratives.

Even in attention to detail, where AI models are thought to excel, human reviewers exhibited higher proficiency in identifying minor inconsistencies and formatting errors. These areas, where AI models currently struggle, underscore the continued importance of human input in the academic review process.

Our findings paint a multifaceted picture of AI's role in academic review. AI models like Bard and GPT-4 demonstrate formidable strengths, particularly in processing large amounts of information and providing insightful and comprehensive feedback. Nevertheless, they also have limitations, especially in grasping the depth and subtlety of human cognitive processing and nuanced academic language.

These insights point towards an integrated approach in academic review, where AI and humans work together. The strengths of AI models can be leveraged to enhance the efficiency and comprehensiveness of the review process. At the same time, humans continue to provide the critical intellectual input and nuanced understanding that AI currently cannot replicate fully. Such a symbiotic relationship promises a more robust, efficient, and insightful academic review process, bridging the gap between traditional and AI-enabled methods.

B. Implications for Public Administration

The revelations drawn from this study carry transformative implications for the field of public administration and beyond. If AI models like Bard and GPT-4 consistently outperform human reviewers or effectively supplement their efforts, we could witness a significant augmentation in the review process's efficiency, accuracy, and speed.

These potential improvements have far-reaching consequences. Firstly, using AI in academic review could lead to faster publication times, an increasingly crucial aspect in the rapidly evolving field of public administration. This would ensure that timely and relevant research reaches the broader scholarly community and informs public administration practices more efficiently.

Secondly, AI-assisted reviews could engender more consistent feedback for authors. Bard and GPT-4's ability to provide relevant and comprehensible feedback consistently could enhance the quality of revisions, leading to more critical, more robust papers.

Additionally, the overall quality of published research could witness an uplift. With GPT-4 or similar models accurately assessing technical aspects and offering clear, insightful feedback, the final published papers would likely be of a higher caliber.

One of the most significant implications is alleviating the burden on human reviewers. With AI shouldering the task of technical assessments and clarity checks, human reviewers could dedicate more time to concentrate on the papers' higher-order conceptual and theoretical aspects. This division of labor between AI and

human reviewers could lead to a more nuanced, rich, and comprehensive review process, benefiting authors, reviewers, and the broader field of public administration.

These implications highlight the potential of AI in revolutionizing academic review processes, promising a future where human creativity and AI efficiency work hand in hand for the advancement of knowledge.

C. Future Research Directions

While our results highlight the remarkable potential of AI, Bard and GPT-4 in transforming academic review, it is imperative to note that our understanding is still nascent and further exploration is required. Future research could extend this study in several important directions.

1. **Exploration of different AI models:** Numerous models with varied capabilities emerge as AI advances. A worthwhile pursuit would be to compare the efficacy of different AI models in academic review and discern the most suitable choice for scenarios or disciplines.
2. **Evaluation across disciplines:** This study focused on public administration, but the use of AI in academic review has implications for all scholarly fields. Future research could replicate our methodology in other disciplines to assess the generalizability of our findings.
3. **Long-term implications:** As AI becomes more integrated into academia, we must understand its long-term implications. Studies should be conducted to forecast the potential impact of AI on the speed, quality, and nature of academic publications over time.

Another crucial aspect to consider is the human factor. It is critical to investigate authors' and reviewers' perceptions of AI-assisted review and gauge their willingness to accept AI-generated feedback. Exploring this could shed light on the practicality and acceptance of such a significant shift in academic review practices and inform guidelines for AI integration in academic processes.

In conclusion, our study stands as a solid testament to the transformative potential of AI, specifically Bard and GPT-4, in the domain of academic review within public administration. We present these findings hoping to catalyze further discourse and exploration in this intriguing nexus of AI and academia. As we stand at this promising frontier, we invite researchers to investigate this significant evolution in the academic landscape.

References:

Aydın, Ö. (2023). Google Bard Generated Literature Review: Metaverse. *Available at SSRN*.

Bai, X., Zhang, F., & Lee, I. (2019). Predicting the citations of scholarly paper. *Journal of Informetrics*, *13*(1), 407-418.

Berryhill, J., Heang, K. K., Clogher, R., & McBride, K. (2019). Hello, World: Artificial intelligence and its use in the public sector.

Checco, A., Bracciale, L., Loreti, P., Pinfield, S., & Bianchi, G. (2021). AI-assisted peer review. *Humanities and Social Sciences Communications*, *8*(1), 1-11.

Gottlieb, M., Kline, J. A., Schneider, A. J., & Coates, W. C. (2023). ChatGPT and conversational artificial intelligence: Friend, foe, or future of research? *The American Journal of Emergency Medicine*, *70*, 81-83.

Guida, M., Caniato, F., Moretto, A., & Ronchi, S. (2023). The role of artificial intelligence in the procurement process: State of the art and research agenda. *Journal of Purchasing and Supply Management*, 100823.

Hariri, W. (2023). Unlocking the Potential of ChatGPT: A Comprehensive Exploration of its Applications, Advantages, Limitations, and Future Directions in Natural Language Processing. *arXiv preprint arXiv:2304.02017*.

Heaven, D. (2018). The age of AI peer reviews. *Nature*, *563*(7733), 609-610.

Majumder, S., & Mondal, A. (2021). Are chatbots really useful for human resource management? *International Journal of Speech Technology*, 1-9.

Mohamed, S. A., Mahmoud, M. A., Mahdi, M. N., & Mostafa, S. A. (2022). Improving efficiency and effectiveness of robotic process automation in human resource management. *Sustainability*, *14*(7), 3920.

Nikolic, S., Daniel, S., Haque, R., Belkina, M., Hassan, G. M., Grundy, S., Lyden, S., Neal, P., & Sandison, C. (2023). ChatGPT versus engineering education assessment: a multidisciplinary and multi-institutional benchmarking and analysis of this generative artificial intelligence tool to investigate assessment integrity. *European Journal of Engineering Education*, 1-56.

Price, S., & Flach, P. A. (2017). Computational support for academic peer review: A perspective from artificial intelligence. *Communications of the ACM*, *60*(3), 70-79.

Rahaman, M. S., Ahsan, M., Anjum, N., Rahman, M. M., & Rahman, M. N. (2023). The AI race is on! Google's Bard and OpenAI's ChatGPT head to head: an opinion article. *Mizanur and Rahman, Md Nafizur, The AI Race is on*.

Ram, B., & Pratima Verma, P. V. (2023). Artificial intelligence AI-based Chatbot study of ChatGPT, Google AI Bard and Baidu AI. *World Journal of Advanced Engineering Technology and Sciences*, *8*(01), 258-261.

Rathore, B. (2023a). Digital Transformation 4.0: Integration of Artificial Intelligence & Metaverse in Marketing. *Eduzone: International Peer Reviewed/Refereed Multidisciplinary Journal*, *12*(1), 42-48.

Rathore, B. (2023b). Future of AI & generation alpha: ChatGPT beyond boundaries. *Eduzone: International Peer Reviewed/Refereed Multidisciplinary Journal*, *12*(1), 63-68.

Salah, M., Alhalbusi, H., Ismail, M. M., & Abdelfattah, F. (2023). Chatting with ChatGPT: Decoding the Mind of Chatbot Users and Unveiling the Intricate Connections between User Perception, Trust and Stereotype Perception on Self-Esteem and Psychological Well-being.

Sallam, M., Salim, N. A., Ala'a, B., Barakat, M., Fayyad, D., Hallit, S., Harapan, H., Hallit, R., Mahafzah, A., & Ala'a, B. (2023). ChatGPT Output Regarding Compulsory Vaccination and COVID-19 Vaccine Conspiracy: A Descriptive Study at the Outset of a Paradigm Shift in Online Search for Information. *Cureus*, *15*(2).

Sun, Z. (2021). *Reducing the Gulfs in Human-Computer Communication: Exploring the Effects of Non-verbal Metaphors*. Hong Kong University of Science and Technology (Hong Kong).

Thierer, A. D., Castillo O'Sullivan, A., & Russell, R. (2017). Artificial intelligence and public policy. *Mercatus Research Paper*.

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, *10*(1), 15.

Valle-Cruz, D., Criado, J. I., Sandoval-Almazán, R., & Ruvalcaba-Gomez, E. A. (2020). Assessing the public policy-cycle framework in the age of artificial intelligence: From agenda-setting to policy evaluation. *Government Information Quarterly*, *37*(4), 101509.