MONOGRAPH SERIES 2026-03

UNIVERSITY OF THE PHILIPPINES
CENTER FOR INTEGRATIVE AND DEVELOPMENT STUDIES

Data-Driven Governance in the Philippines

AI, Computational Methods, and Public Policy

Ebinezer Florano, PhD
EDITOR

PUBLISHED BY
University of the Philippines Center for Integrative and Development Studies

Data Science For Public Policy Program

ISSN 2719-0722 (PRINT)
ISSN 2719-0730 (ONLINE)

MONOGRAPH SERIES 2026-03

Data-Driven Governance in the Philippines

AI, Computational Methods, and Public Policy

Ebinezer Florano, PhD
EDITOR

PUBLISHED BY
University of the Philippines Center for Integrative and Development Studies

Data Science For Public Policy Program

ISSN 2719-0722 (PRINT)
ISSN 2719-0730 (ONLINE)

UP CIDS MONOGRAPH SERIES
2026-03

UP CIDS Monograph Series is published by the

University of the Philippines
Center for Integrative and Development Studies

Lower Ground Floor, Ang Bahay ng Alumni
Magsaysay Avenue, University of the Philippines
Diliman, Quezon City 1101

Telephone: (02) 8981-8500 loc. 4266 to 4268 / (02) 8426-0955
Email: cidspublications@up.edu.ph
Website: cids.up.edu.ph

Copyright © 2026
by the UP Center for Integrative and Development Studies

The views and opinions expressed in this monograph are those of the author/s and neither reflect nor represent those of the University of the Philippines or the UP Center for Integrative and Development Studies. No copies can be made in part or in whole without prior written permission from the authors/editors and the publisher.

ISSN 2719-0722 (Print)
ISSN 2719-0730 (Online)

Cover Image Credit

"Green futuristic waves background with computer code technology"

rawpixel.com, Magnific, https://tinyurl.com/yhn7ac4b

Table of Contents

About the Monograph vii

Foreword 1

Ebinezer Florano, PhD

Evaluating New Public Management Principles and AI Integration 6

A Comparative Analysis of ChatGPT-Assisted and Traditional Memorandum Drafting by Civil Servants in Selected National Government Agencies (NGAs) in the Philippines

Mariecon B. Aranda-Esguerra

Abstract 7

Introduction 8

Review of Related Literature 8

Methodology 12

Data Presentation 14

Analysis 16

Conclusion 17

References 18

From Complaints to Policy Suggestions: Analyzing Facebook Comments for Improved EDSA Traffic Management 20

Joselito M. Ebro Jr.

Abstract 21

Introduction 22

Methodology 27

Results: Institutional Capacity and Citizen Policy Intelligence 28

Toward Data-Informed Governance 35

References 36

Text-Mining Techniques to Identify Ageist Online Job Ads in the Philippines 40

Fernan Dayson C. Manahan

Abstract 41

Introduction 42

Methodology 52

Doc2Vec Models 63

Determining Ageist Job Ads Through Classification Models 66

Conclusion and Policy Implications 87

Appendix 88

Acknowledgment 88

References 88

Leveraging Culturomics to Inform Anti-Discrimination Policies: The Case of *Pakikipagkapwa* in the Philippines 94

Maria Margarita R. Lavides, PhD, Vladimer B. Kobayashi, PhD, Cristabel F. Tiangco, Purisima P. Panlilio, PhD, Remi E. De Leon, and Angela D. Carreon

Abstract 95

Introduction 96

Literature Review 97

Methodology 101

Results and Discussion 106

Conclusion 115

References 115

Does the ASEAN Smart Cities Network Promote Climate Resilience? 120

Louie Ayade, Charlyn Capulong, Martha Joy Abing, Maria Rizalia Teves, and Gerrich Aldin Babanto

Abstract 121

Introduction 122

Background and Research Hypothesis 123

Methodology, Variables, and Data 126

Empirical Analysis Results 130

Conclusion and Recommendations 139

References 142

Service to the People? Mapping and Measuring the University of the Philippines' Impact on the Philippine Policy Process 148

Ronald A. Pernia, PhD and Rogelio Alicor L. Panao, PhD

Abstract 149

Introduction 150

Data and Method 153

Results 155

Conclusion 166

References 167

About the Monograph

This monograph brings together six empirical studies that examine how data science, artificial intelligence, and computational methods are reshaping public policy, governance, and administration in the Philippines. Drawing on diverse analytical techniques, from text mining and document embeddings to difference-in-differences estimation, topic modeling, and social media discourse analysis, the collection demonstrates the growing role of data-driven approaches in understanding and addressing pressing societal challenges.

The chapters explore a range of contemporary issues: the integration of large language models such as ChatGPT into civil service workflows and its implications for New Public Management; the use of Facebook discourse to generate actionable traffic management policy for EDSA; the application of text-mining techniques to detect ageist language in online job advertisements; the use of cultural analytics to map the Filipino value of *pakikipagkapwa* and its implications for anti-discrimination policy; the impact of the ASEAN Smart Cities Network on climate resilience in Philippine highly urbanized cities; and the use of natural language processing on 720 Scopus-indexed policy articles by University of the Philippines scholars to show how the tone of research (critical, neutral or reform-oriented) shapes its visibility in global academic venues.

Together, these studies contribute to the emerging field of data-informed governance, offering both methodological innovations and practical policy insights. They highlight the potential and the limitations of leveraging digital data and computational tools to enhance public service delivery, promote inclusivity, and build more resilient communities. The monograph will be of interest to scholars, policymakers, and practitioners in public administration, data science, urban planning, and Philippine studies.

Foreword

Ebinezer Florano, PhD

The twenty-first century has witnessed an unprecedented proliferation of digital data, and with it, a profound transformation in how societies govern themselves. Governments worldwide are increasingly turning to data science, artificial intelligence, and computational analytics to inform policy, deliver services, and engage citizens. Yet, the promise of data-driven governance is not without its complexities, i.e., questions of equity, ethics, capacity, and context loom large, particularly in developing nations where resources are constrained and institutional frameworks are still evolving. It is within this milieu of opportunity and uncertainty that this monograph finds its purpose.

The Philippines, a nation characterized by its vibrant democracy, rapid urbanization, and deep social interconnectedness, stands at a critical juncture. The country's public sector faces formidable challenges: traffic congestion that strangles metropolitan economies, persistent discrimination that undermines social cohesion, climate vulnerabilities that threaten urban resilience, and an aging workforce navigating a rapidly digitizing labor market. Simultaneously, Filipinos are among the world's most avid users of social media and digital platforms, generating vast quantities of data that, if harnessed responsibly, could illuminate pathways toward more responsive, inclusive, and effective governance. This monograph seizes upon that potential, presenting six rigorous studies that collectively demonstrate how data science and artificial intelligence can be mobilized to address some of the Philippines' most pressing policy dilemmas.

The chapters that follow are not merely technical exercises in computational analysis. They are, at their core, exercises in democratic renewal, i.e., attempts to render visible the voices, patterns, and inequities that traditional policy processes often overlook. They explore how the tools of the digital age can be wielded to strengthen, rather than erode, the principles of good governance: transparency, accountability, inclusiveness, and responsiveness.

The first chapter, by Aranda-Esguerra, examines the integration of large language models, specifically ChatGPT, into the routine work of civil servants. Through a carefully designed experiment comparing AI-assisted and traditional memorandum drafting, the study finds that AI assistance significantly improves the readability and consistency of government communications. This finding carries profound implications for New Public Management principles, suggesting that AI can enhance efficiency and standardize outputs while also raising important questions about over-reliance, diminished discretion, and the erosion of institutional voice. The study

exemplifies how experimental methods can be deployed to evaluate the real-world impact of emerging technologies on bureaucratic practice.

Ebro's contribution takes a different but equally innovative approach, analyzing nearly 3,000 Facebook comments to uncover the policy intelligence embedded in public discourse surrounding the EDSA rehabilitation program. Far from being a mere repository of complaints, social media emerges in this study as a site of "enhanced public discourse," where citizens offer sophisticated policy critiques, link personal experiences to systemic infrastructure challenges, and propose actionable solutions. The study challenges the conventional wisdom that social media is anathema to substantive policy deliberation, demonstrating instead that with systematic analysis, digital platforms can become vital sources of citizen-generated policy innovation. The proposed framework for integrating social media insights into MMDA's operations offers a practical pathway for government agencies seeking to move beyond reactive complaint management toward proactive policy co-creation.

Manahan's study tackles the subtle but pervasive problem of ageism in online job advertisements. Using document embeddings, latent Dirichlet allocation, and clustering techniques, the research quantitatively identifies ageist language in nearly 9,000 job ads from major Philippine job boards. The study's methodological sophistication is matched by its policy relevance: it offers job boards, employers, and regulators a scalable approach to flagging potentially discriminatory language for human review. Importantly, the study navigates the tension between sensitivity and specificity, recognizing that over-flagging reasonable job requirements as ageist would undermine the credibility of such tools. The finding that clustering techniques may be more viable than classification models at the job-ad level provides a nuanced contribution to the growing literature on algorithmic discrimination detection.

Lavides and colleagues employ cultural analytics to investigate *pakikipagkapwa*, the Filipino value of recognizing others as fellow human beings. By scraping over 640,000 texts from social media platforms and online forums and applying topic modeling, the study develops an emergent model of *pakikipagkapwa* comprising channels, moderators, manifestations, and enablers. The research then connects these findings to the persistence of discriminatory behaviors in the Philippines, identifying family upbringing, religious beliefs, educational experiences, and the influence of public personalities as key moderating factors. The policy implications are far-reaching, suggesting that anti-discrimination legislation must address not only overt acts of discrimination but also the underlying socializing influences that perpetuate bias. This study is a powerful demonstration of how computational methods can illuminate deeply cultural phenomena, bridging the gap between data science and Filipino psychology.

Ayade and colleagues investigate the impact of the ASEAN Smart Cities Network (ASCN) on climate resilience in Philippine highly urbanized cities. Using a difference-in-differences approach with panel data from 2014 to 2023, the study finds that ASCN participation is associated with reduced climate resilience, contrary to expectations. The mechanism analysis reveals that ASCN significantly diminishes information technology capacity, which, while positively associated with climate resilience, does not mediate the relationship. Heterogeneity analysis further shows that cities with stronger school service capacity experience less adverse impacts, underscoring the critical role of education in climate adaptation. The study offers sobering lessons for developing countries seeking to leverage smart city frameworks for climate action, emphasizing the need for better coordination between national and local governments, targeted investments in IT infrastructure, and a focus on the "smart people" pillar of smart city development.

Finally, Pernia and Panao turn the lens inward, examining how the University of the Philippines (UP) itself shapes, and is shaped by, the policy knowledge it produces. Applying natural language processing to 720 Scopus-indexed public policy articles by UP-affiliated scholars, the study analyzes thematic content, sentiment, and evaluative framing. It finds that constructive, solution-oriented research is more likely to reach indexed venues, while predominantly critical work is less visible. This suggests that tone acts as an implicit filter in academic gatekeeping. The study highlights the tension in UP's dual role as both critic of and contributor to national development, and the pressure it creates for scholars to frame critique within reformist narratives. In doing so, it makes a timely contribution to the sociology of knowledge, showing how tone mediates the legitimacy and circulation of ideas from the Global South.

Together, these six studies represent a rich tapestry of methodological approaches—experimental design, social media discourse analysis, text mining, cultural analytics, quasi-experimental econometrics, and natural language processing of scholarly output—unified by a common commitment to using data science for the public good. They demonstrate that the tools of the digital age are not the exclusive province of the private sector or the academy; they can and must be brought to bear on the challenges of governance. At the same time, they collectively caution against technological solutionism. The findings on AI-assisted memorandum drafting, for instance, do not suggest that AI can replace human judgment; rather, they point to the importance of embedding AI within systems of ethical oversight and organizational learning. The analysis of EDSA traffic discourse reveals that social media can be a source of policy intelligence, but only if government agencies develop the analytical capacity to systematically mine and interpret that data. The study of ageist job ads shows that algorithmic tools can flag potential discrimination, but human review remains essential. The *pakikipagkapwa* research demonstrates that cultural values can be mapped and measured, but translating those insights into policy requires a

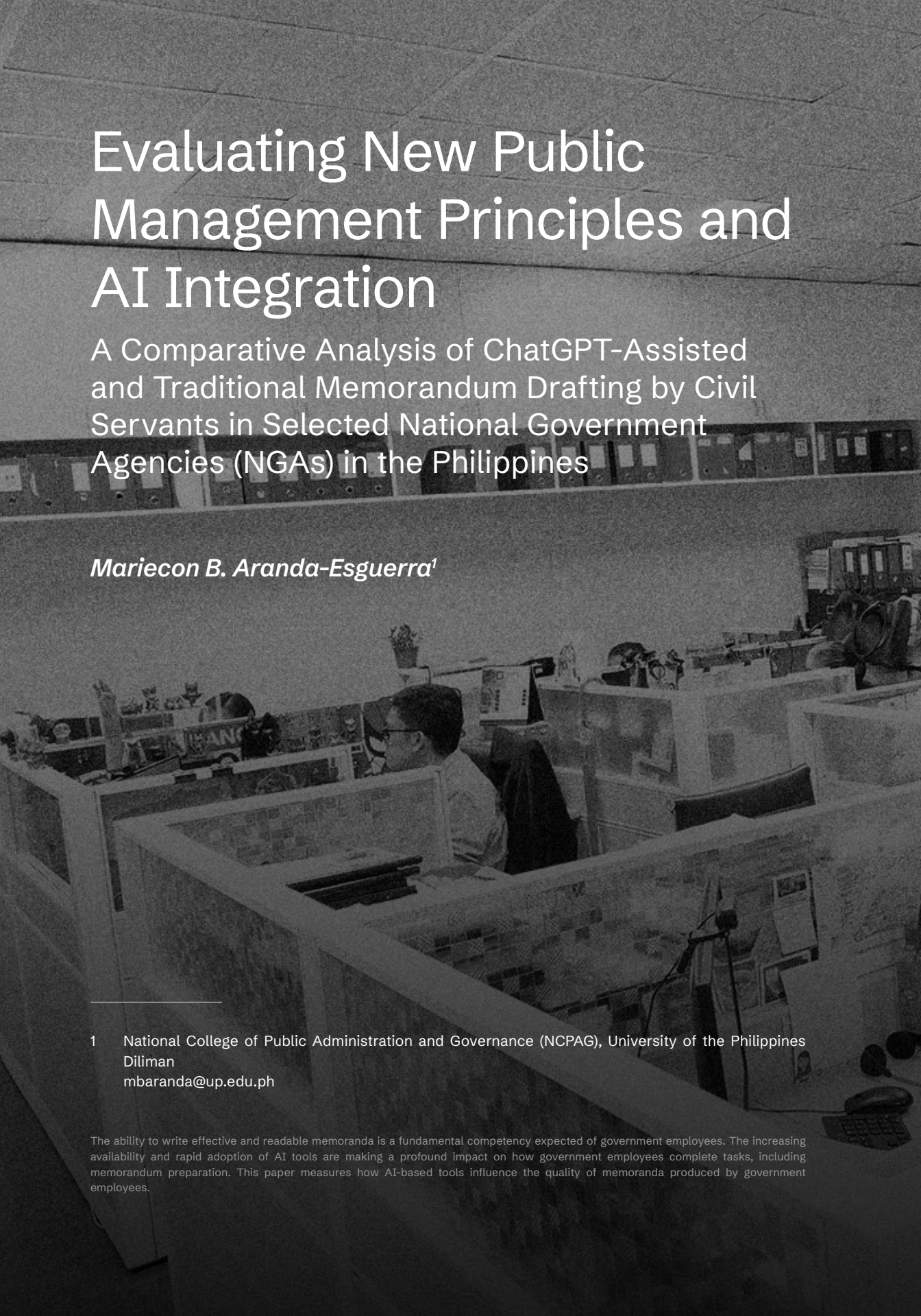
nuanced understanding of social dynamics. And the ASCN study reminds us that technological investments must be aligned with local capacities and needs, lest they produce unintended consequences.

The contributors to this volume have performed a valuable service. They have pushed the boundaries of what is possible in data-informed governance, and they have done so with rigor, creativity, and a clear-eyed awareness of the limitations of their methods. Their work will be of interest to scholars, policymakers, and practitioners alike.

Evaluating New Public Management Principles and AI Integration

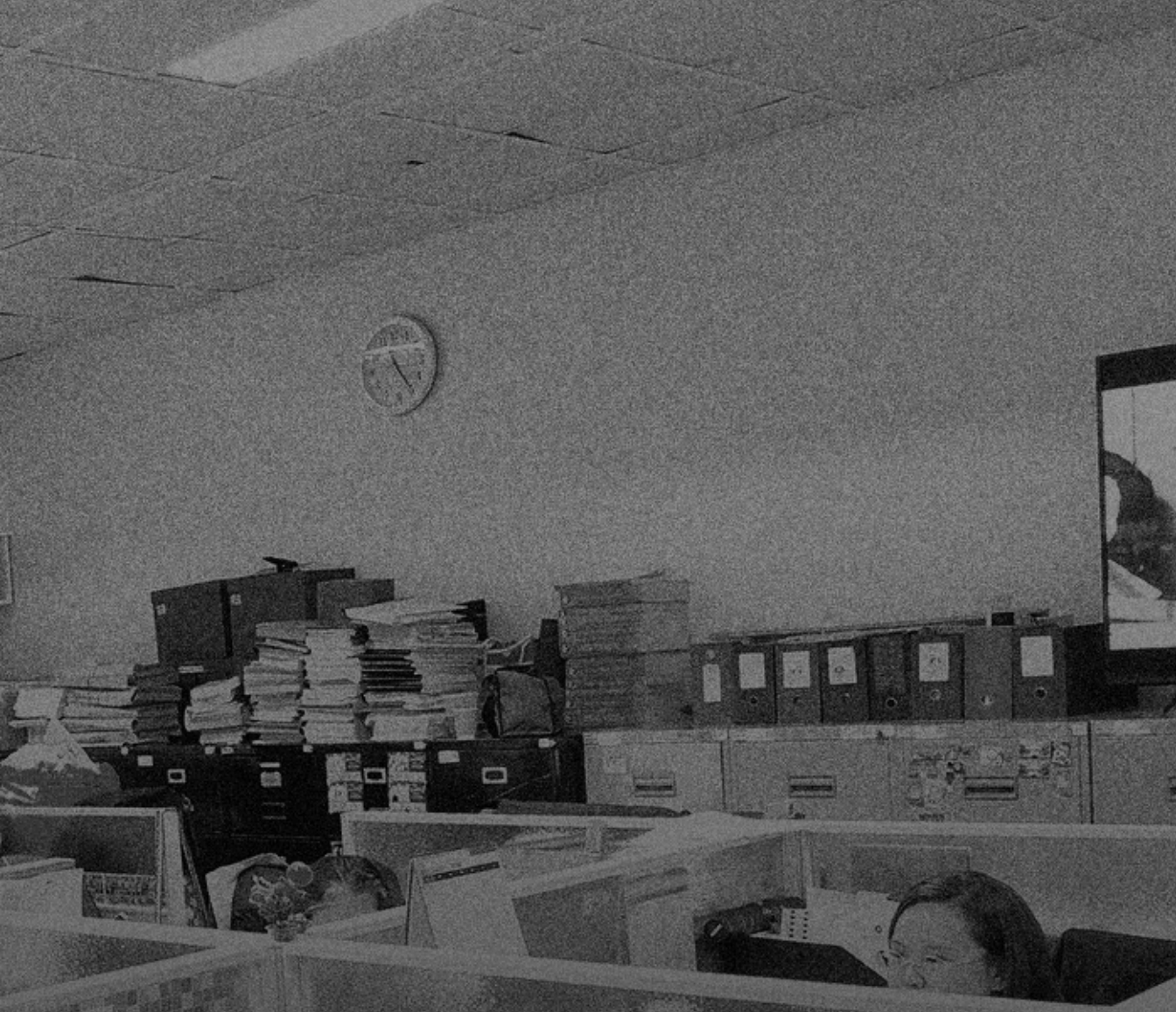
A Comparative Analysis of ChatGPT-Assisted and Traditional Memorandum Drafting by Civil Servants in Selected National Government Agencies (NGAs) in the Philippines

Mariecon B. Aranda-Esguerra[1]



1 National College of Public Administration and Governance (NCPAG), University of the Philippines Diliman
mbaranda@up.edu.ph

The ability to write effective and readable memoranda is a fundamental competency expected of government employees. The increasing availability and rapid adoption of AI tools are making a profound impact on how government employees complete tasks, including memorandum preparation. This paper measures how AI-based tools influence the quality of memoranda produced by government employees.



Abstract

This study assesses the impact of using ChatGPT in helping perform routine tasks in the public sector, specifically, the drafting of memoranda, using the lens of New Public Management. Participants, who were public sector employees, were instructed to draft a memorandum seeking approval for a benchmarking study on the ethical use of AI in public management using a common set of information. Those in the treatment group performed the task with the help of ChatGPT, while those in the control group were not allowed to use any AI tools. Quality was measured using Microsoft Editor readability scores. Results showed that memoranda drafted with the help of ChatGPT achieved higher readability scores compared to those that did not. These findings show the benefits of using AI tools such as ChatGPT to improve the quality of memoranda produced by public sector employees. It also offers a window into the potential impact of a more widespread adoption of AI tools in the public sector.

Introduction

While AI encompasses several models, tools, and methodologies, large language models (LLMs) have gained prominence with the release of ChatGPT, and subsequently, Gemini, DeepSeek and other competitors. Thus, as AI becomes more integrated in daily life, there is an urgent need to study how the public sector can leverage this technology to better serve its constituents.

LLMs are AI that can mimic conversations with an actual person. They are capable of understanding and generating natural language and can perform a variety of tasks based on instructions given by the user using common or conversational language. LLMs acquire these capabilities by deriving patterns and structures from massive amounts of data (Stryker, n.d.).

LLMs can assist in performing a range of communication tasks. In the public sector, one of the more common tasks that may benefit from the use of LLMs is the writing and editing of memoranda. More specifically, LLMs can assist in drafting communication materials by providing suggestions, correcting grammar, and enhancing the clarity and coherence of the text.

This integration aligns with New Public Management (NPM) principles, which emphasize efficiency, effectiveness, and the adoption of private sector practices within the public sector. By leveraging AI such as LLMs, public administrators can streamline their workflows, reduce administrative burdens, and focus on more "higher-level" tasks.

Existing studies appear to focus on AI's role in broader administrative functions such as data analysis and decision-making support (Desouza 2018). There appears to be a lack of detailed examination of AI's impact on tasks like memorandum preparation, which is crucial in the day-to-day operations of government agencies. By focusing on this specific aspect, this study seeks to fill an important gap in the literature and provide actionable insights for public sector agencies.

Review of Related Literature

The Future of Work: Robots, AI, and Automation by D. M. West (2018) outlines the profound changes our society faces with the rise of technology, with automation, robotics, and AI driving a digital economy. The author believes that automation might replace many traditional jobs, pushing everyone in the workforce to develop new skills and redefine the standard of work, including the value of roles like volunteering and parenting. It emphasizes the need for policies to mitigate economic

disruption and decrease polarization and highlights the importance of continuous skill development to adapt to changing job landscapes.

The Evolution of Bureaucratic Processes through ICT

Bovens and Zouridis (2002) explore the shift from street-level bureaucracies—where individual officials had significant discretion—to system-level bureaucracies, characterized by structured and automated decision-making processes. Their work is pivotal for understanding the evolution of public administration in the digital age. They highlight the reduction of administrative discretion due to ICT applications, providing crucial insights into balancing efficiency and flexibility. Their analysis underscores how technology can both enhance and constrain bureaucratic decision-making.

New Public Management Principles

NPM principles focus on enhancing efficiency, effectiveness, decentralization, performance measurement, outcome orientation, and customer (citizen) service in public administration. These principles advocate adopting modern management techniques from the private sector to make public administration more responsive and efficient. Hood (1991) outlines the core concepts of NPM, emphasizing performance measurement and accountability. His work is instrumental in understanding the shift towards performance metrics as a means of driving public sector reforms. Osborne and Gaebler's Reinventing Government (1992) further emphasizes this shift. They advocate for a government that "steers" rather than "rows," focusing on results rather than processes. The authors also note the importance of an entrepreneurial government that is innovative, mission-driven, and results-oriented. Their work underscores the necessity of performance measurement to ensure that public services are delivered effectively and efficiently. Pollitt and Bouckaert (2017) provide a comparative analysis of public management reforms across different countries, illustrating the widespread adoption of NPM principles. They offer insights into the variations and commonalities in implementing private sector practices in public administration globally. Their analysis reveals how performance measurement has become central to public management reforms, influencing policy decisions and administrative practices.

Artificial Intelligence (AI) in Public Administration

AI involves creating systems capable of performing tasks that typically require human intelligence, such as learning, reasoning, and understanding natural language. Integrating AI in public administration can streamline operations, reduce

administrative burdens, and enhance decision-making processes. Desouza et al. (2019) note that Korea and Singapore have announced the formation of respective AI ethics boards, as other industrial associations do. They asserted that AI projects, especially in the government, where projects must advance the public good and deliver public value, should give importance to auditing AI systems to ensure accuracy and mitigate biases. Correia et al. (2020) focus on the challenges of AI in public administration within the framework of smart cities, addressing legal issues and the integration of AI systems. They say that AI systems in public administration can simplify processes, making them faster and more accessible for citizens by reducing bureaucracy. Shrum and Gordon (2022) explore the future of work and the impact of AI on public administration, emphasizing the need for adaptability and continuous learning among public sector employees to effectively leverage AI tools.

Regan and Maschino (2021) provide a practical guide to ethics and AI in public administration, offering strategies for managing ethical dilemmas and ensuring responsible AI use in public services. Their study highlights the opacity of AI systems, which challenges traditional administrative responsibilities. Administrators must understand how AI affects their oversight, responsibility, and the acceptable level of AI autonomy. Clear contracts and terms of service are crucial to define AI's role, responsibilities, and ownership of information. AI systems often use personal information collected for other purposes, raising ethical concerns about secondary use and data privacy. Much of big data collection occurs without individuals' awareness, challenging previous ideas about data collection limits and individual control. AI systems can search vast quantities of data and identify patterns, making passive surveillance attractive. However, ethical questions arise about the relevance, reliability, and appropriateness of surveilling certain types of data, such as social media sites. Public administrators should consider the legitimacy and impact of using AI for surveillance. The paper emphasizes the importance of understanding contracts, ensuring transparency, and addressing ethical concerns in AI deployment in public administration. Shark and Shropshire (2019) discuss the role of AI in teaching public policy and administration, highlighting the importance of integrating AI education into public administration curricula to prepare future public managers for the evolving landscape. According to their paper, it is essential to equip lawmakers and future senior managers with the necessary tools to make informed decisions, rather than relying solely on computer science professionals. It also emphasized the risks of allowing inexperienced individuals to make these important decisions.

Effective Business Communication in Public Administration

Effective business communication is characterized by clarity, simplicity, brevity, objectivity, accuracy, attentiveness, relevancy, feedback, empathy, and consistency. These principles are crucial for conveying messages clearly and concisely, ensuring that

information is understood and acted upon appropriately. Effective communication in public administration facilitates better decision-making and organizational efficiency.

According to APAC Entrepreneur (2024), effective business communication principles are essential for achieving clarity and conciseness in conveying messages. These principles play a vital role in ensuring that information is accurately interpreted and appropriately acted upon by recipients. By adhering to the principles of effective business communication, public administrators can ensure that their messages are relevant, accurate, and consistent, leading to more informed decision-making processes and improved overall organizational performance.

Microsoft Editor Score

According to Microsoft (n.d.) through Microsoft Support, Microsoft Editor provides comprehensive readability statistics that evaluate grammar, clarity, and overall writing quality. The Microsoft Editor score is closely related to the document statistics it provides. The number of words, characters, paragraphs, and sentences help determine the overall length and structure of the document, influencing readability scores. The average number of sentences per paragraph indicates how well the text is broken up into manageable chunks, with shorter paragraphs generally being easier to read. The average number of words per sentence assesses sentence complexity, with shorter sentences typically being easier to understand, thus improving readability scores. The average number of characters per word gives an idea of word complexity, with simpler words being easier to read.

By providing detailed readability statistics, Microsoft Editor helps writers understand the complexity of their text and make necessary adjustments to improve readability. This tool is particularly useful for public administrators who need to ensure that their communication materials are clear, concise, and accessible to a broad audience. Thus, tools like Microsoft Editor may help enhance the quality of written communication in public administration, as well as provide quantifiable metrics on the quality of written communications.

Case Study: Singapore

The integration of AI in governmental functions is becoming increasingly prevalent. A notable example is the Singapore government's pilot program in 2023, which explored the utilization of ChatGPT to assist civil servants with tasks such as research and speechwriting. This initiative aims to enhance productivity and streamline operations within the public sector.

In developing a tailored version of ChatGPT for government use, Singapore demonstrates its commitment to leveraging advanced AI technologies to meet the unique requirements of public service. To ensure the responsible use of these tools, the government has established comprehensive guidelines. These guidelines emphasize the importance of accountability and accuracy, mandating that public officers verify AI-generated content and remain responsible for the integrity of their work.

Methodology

This study evaluated the impact of ChatGPT-assisted memorandum drafting by comparing the Microsoft Editor scores of drafts prepared by civil servants from select national government agencies (NGAs) of the Philippines. Participants, whose work functions in their respective offices involve the preparation of communication materials and memoranda, were divided into two groups: Group X/Control Group (not assisted by ChatGPT) and Group Y/Treatment Group (assisted by ChatGPT).

Respondents from both groups were instructed to draft a memorandum seeking approval for a benchmarking study on the ethical use of AI in public management. The memorandum should include the following details:

- an introduction stating the purpose of seeking approval for the benchmarking study;
- study goals and objectives outlining the main goals, namely exploring global best practices, identifying guidelines, and developing frameworks for the ethical use of AI in public administration;
- partner institutions and facilitators, including the Korea Information Society Development Institute (KISDI) and the Ministry of Science and ICT (MSIT) and their roles;
- the number of participants (10) from the agency/bureau/division attending the study, including their expected profiles; event details specifying the date and location (Seoul, South Korea, at KISDI, from 20 to 26 April 2025, with travel arrangements covered);
- an overview of key activities, namely a courtesy call to the Philippine Ambassador, site visits to AI-related institutions, and a lecture on South Korea's AI strategies and proposed AI National Ethical Standards;
- attire requirements, specifying business Filipiniana for the embassy visit and graduation ceremony, and business attire for regular activities;

- the allocated budget, including ₱25,000 per participant for round-trip airfare (economy class) and US$235 per participant per day for the seven-day trip, totaling an estimated budget per participant of ₱121,227.25; and

- a post-travel report requirement entailing each participant to submit a report, including a plan for AI ethical integration, recommendations, and a needs assessment for AI integration at their respective agencies.

Participants may also add other helpful information to justify the request. Both groups were given the same information and details; however, the treatment group was required to use ChatGPT to assist them in drafting the memorandum. The researcher did not specify the prompts to be used by the participants in the treatment group.

Data collection involved a combination of tasks, post-experiment evaluation surveys, and document assessments using Microsoft Editor Score. Participants also completed a post-experiment survey to gather demographic information and their perceptions of AI integration. To examine if the data follows a normal distribution, histograms were plotted, and appropriate tests were conducted on raw and transformed data. The tools utilized in this study were Microsoft Word and ChatGPT for drafting memoranda, and Microsoft Excel and R for data analysis.

The study was initially designed for the Philippine Carabao Center (PCC). However, as the deadline approached, the researcher could not collect data from the respondents identified by PCC management. To meet the required number of respondents, the target respondents were expanded to include employees and officers from various government agencies, including the Department of Agriculture and its attached agencies and corporations, the Department of Social Welfare and Development, the Senate Legislative Budget Research and Monitoring Office, the Philippine Competition Commission, and the Presidential Legislative Liaison Office. All were government employees whose regular tasks include preparing communication materials such as memoranda for programs and activities, budget, finance, administrative matters, human resources-related matters, policy orders, and lecture and presentation materials. Participation in the experiment was voluntary.

Data Presentation

Descriptive Statistics

Table 1. Descriptive Statistics

Particulars	Sample	Control Group (*X*)	Treatment Group (*Y*)
Mean	0.844	0.812	0.877
Standard error	0.011	0.017	0.012
Median	0.840	0.790	0.880
Mode	0.840	0.740, 0.790	0.840, 0.950
Standard deviation	0.841	0.087	0.060
Sample variance	0.007	0.008	0.004
Kurtosis	−0.958	−1.072	−1.012
Skewedness	−0.293	0.255	−0.347
Range	0.280	0.280	0.190
Minimum	0.680	0.680	0.770
Maximum	0.960	0.960	0.960
Sum	42.210	20.290	21.290
Count	50.000	25.000	25.000

Table 1 presents the descriptive statistics in this study. The treatment group (*Y*), which drafted the memorandum with the help of ChatGPT, had a higher mean score (0.877) compared to the control group (*X*) (0.812). The lower standard error (0.012) and standard deviation (0.060) in the treatment group indicate more precision and consistency in their mean scores. The median score for the treatment group (*Y*) (0.880) is also higher than that of the control group (*X*) (0.790). Additionally, the treatment group's mode (0.950) and lower sample variance (0.004) further demonstrate less variability and more consistent high performance compared to the Control Group. The minimum score for the treatment group (*Y*) (0.770) is higher than that of the control group (*X*) (0.680), reflecting a higher baseline performance. Overall, the Treatment Group's scores are less dispersed and more consistent.

The box and whisker chart shows that for Group X, the median score is approximately 0.8, with the interquartile range (IQR) spanning from about 0.75 to 0.9. The whiskers extend from around 0.7 to 0.95. This indicates that most scores for the control group (*X*) are concentrated within this range. For Group Y, the median score is slightly higher at around 0.85. The IQR for the treatment group extends

from approximately 0.8 to 0.9, with whiskers ranging from about 0.75 to 0.95. This shows that the bulk of the scores for the treatment group are slightly more clustered than those of the control group. It demonstrates that the treatment group (*Y*) has a higher median score compared to the control group (*X*).

Figure 1. Scores – *x* and *y*

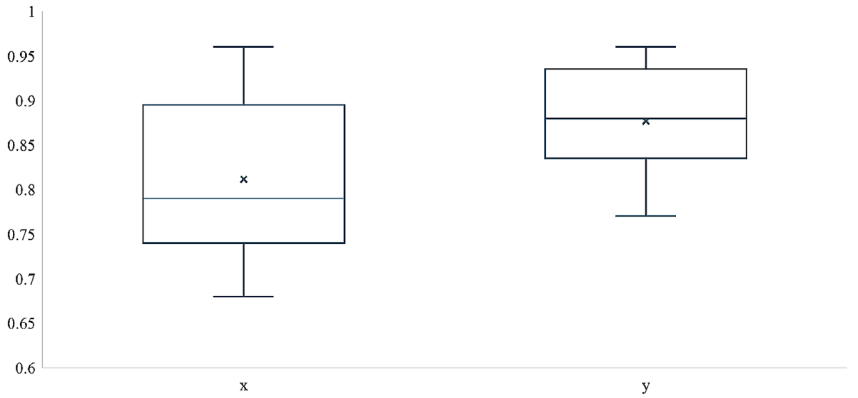


Figure 2. Sampling distribution – all

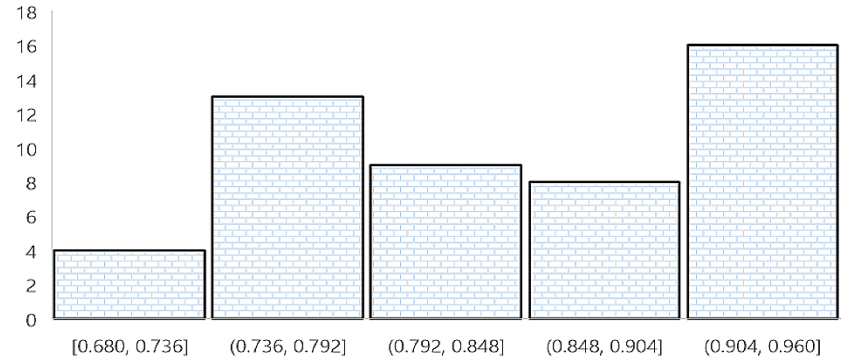


Figure 3. Sampling distribution – *x*

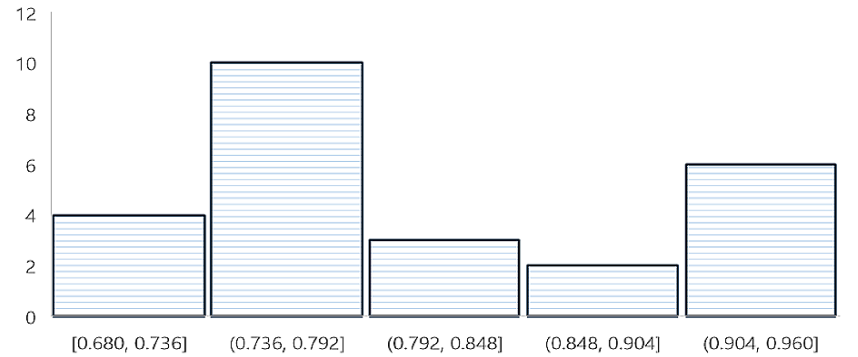
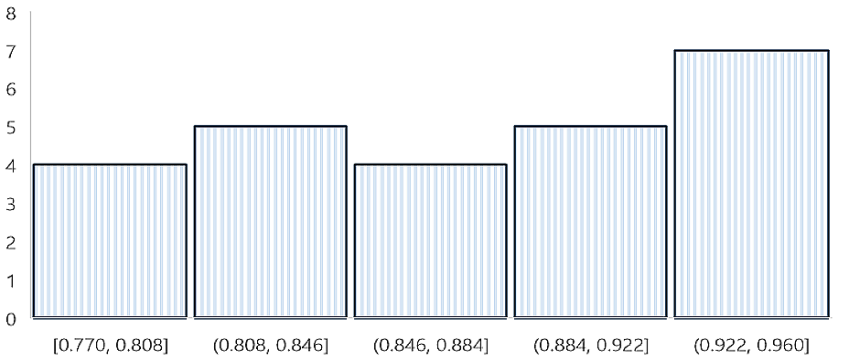


Figure 4. Sampling distribution – *y*



The data, even after log and square root transformation, did not exhibit a normal distribution based on a visual test. A Shapiro-Wilk normality test was conducted to determine whether the data fit a normal distribution. A *p*-value below 0.05 indicates, with 95 percent confidence, that the data do not fit a normal distribution. In this case, the *p*-value was 0.03032, suggesting that the data were not normally distributed.

A Mann–Whitney *U*-test was conducted to compare the medians of the treatment and control groups. The U statistic and the associated *p*-value were calculated. With a *p*-value of 0.007632882, which is below the significance level (α) of 0.05, we reject the null hypothesis (H_0) that there is no difference in the quality of memoranda drafted by the treatment and control groups. Thus, the study found that the higher median score of the treatment group is significant.

Analysis

The experiment was designed to control for content variability: all participants received the same factual material and were tasked solely with assembling it into a formal memorandum. This setup isolates the writing process, allowing for a direct comparison between AI-assisted and human-only outputs.

Performance was assessed exclusively using the Microsoft Editor score, which evaluates clarity, conciseness, and overall readability. Results showed that memos produced with ChatGPT assistance consistently received higher scores than those written without AI support. The AI-assisted group also exhibited less variation in scores, indicating a more uniform level of output quality across participants.

These findings suggest that AI assistance can improve baseline communication performance and reduce disparities in writing quality among public servants. The Microsoft Editor score served as a consistent and objective metric for evaluating

these improvements, offering a quantifiable basis for assessing the impact of AI on bureaucratic writing.

Conclusion

This study demonstrates that large language models (LLMs) such as ChatGPT can significantly enhance the quality of memoranda produced by public sector employees. This aligns with New Public Management's emphasis on performance, responsiveness, and measurable outputs, suggesting that AI-assisted tools can help public institutions better meet the needs of citizens, clients, and stakeholders.

At the same time, the findings underscore the importance of performance measurement in identifying areas for improvement and fostering continuous learning. By integrating AI into routine workflows, agencies can standardize outputs and elevate communication quality. However, this efficiency must be balanced against emerging risks: over-reliance on AI, diminished discretion, and the erosion of institutional voice.

The challenge, therefore, is not merely to adopt AI, but to govern its use wisely. This entails embedding AI within broader systems of review, ethical oversight, and organizational learning. Civil servants must be equipped with the skills to engage critically with these tools—especially in tasks that demand interpretation, tone, and contextual sensitivity.

Further study is encouraged to explore the long-term impact of AI-assisted tools on various aspects of public administration. This includes more detailed comparisons, expanded use cases, and broader applications to ensure the continued improvement and effective integration of AI in public sector operations.

Ultimately, AI can be a powerful ally in public sector reform. But its integration must be guided by principles that protect the integrity of public communication, uphold democratic accountability, and strengthen—not replace—the judgment and authorship of those who serve.

References

APAC Entrepreneur. 2024. "10 Principles of Effective Business Communications." https://apacentrepreneur.com/10-principles-of-effective-business-communications/.

Bovens, Mark, and Stavros Zouridis. 2002. "From Street-Level to System-Level Bureaucracies: How Information and Communication Technology is Transforming Administrative Discretion and Constitutional Control." *Public Administration Review* 62 (2): 174–84. https://doi.org/10.1111/0033-3352.00168.

Brown, Tom B., Benjamin Mann, Nick Ryder, et al. 2020. "Language Models Are Few-Shot Learners." *arXiv* 2005 (14165). https://doi.org/10.48550/arXiv.2005.14165.

Correia, Pedro Miguel Alves Ribeiro, Ricardo Lopes Dinis Pedro, Irineu de Oliveira Mendes, and Alexandre D. C. S. Serra. 2020. "Artificial Intelligence in Smart Cities: Challenges for Public Administration." *Journal of Public Administration and Governance* 10 (4): 122–39.

Desouza, Kevin C. 2018. "Governing with Algorithms: AI and Public Administration." *Perspectives on Public Management and Governance* 1 (3): 234–50.

Dunleavy, Patrick, and Helen Margetts. 2015. "Design Principles for Essentially Digital Governance." Presented at the 111th Annual Meeting of the American Political Science Association, San Francisco, 3–6 September 2015. https://researchonline.lse.ac.uk/id/eprint/64125/1/Essentially%20Digital%20Governance.pdf.

Hood, Christopher. 1991. "A Public Management for All Seasons?" *Public Administration* 69 (1): 3–19. https://doi.org/10.1111/j.1467-9299.1991.tb00779.x.

Microsoft. n.d. "Microsoft Editor Checks Grammar and More in Documents, Mail, and the Web." Microsoft Support. Accessed 13 August 2026. https://support.microsoft.com/en-us/word/microsoft-editor-checks-grammar-and-more-in-documents-mail-and-the-web.

OpenAI. 2020. "Language Models are Few-Shot Learners." 28 May 2020. https://openai.com/index/language-models-are-few-shot-learners/.

Pollitt, Christopher, and Geert Bouckaert. 2017. *Public Management Reform: A Comparative Analysis—Into the Age of Austerity*. 4th ed. Oxford University Press.

Regan, Priscilla M., and Karl Maschino. 2021. "Ethics and AI: A Practical Guide for Public Administration." *Ethics in Public Administration* 12 (2): 300–315.

Shark, Alan R., and Anders Shropshire. 2019. "Teaching AI in Public Policy and Administration." *Journal of Public Affairs Education* 25 (2): 175–90.

Shrum, Karen, and Gordon, Lisa. 2022. "The Future of Work: AI's Impact on Public Administration." *Public Administration Review* 82 (1): 45–58.

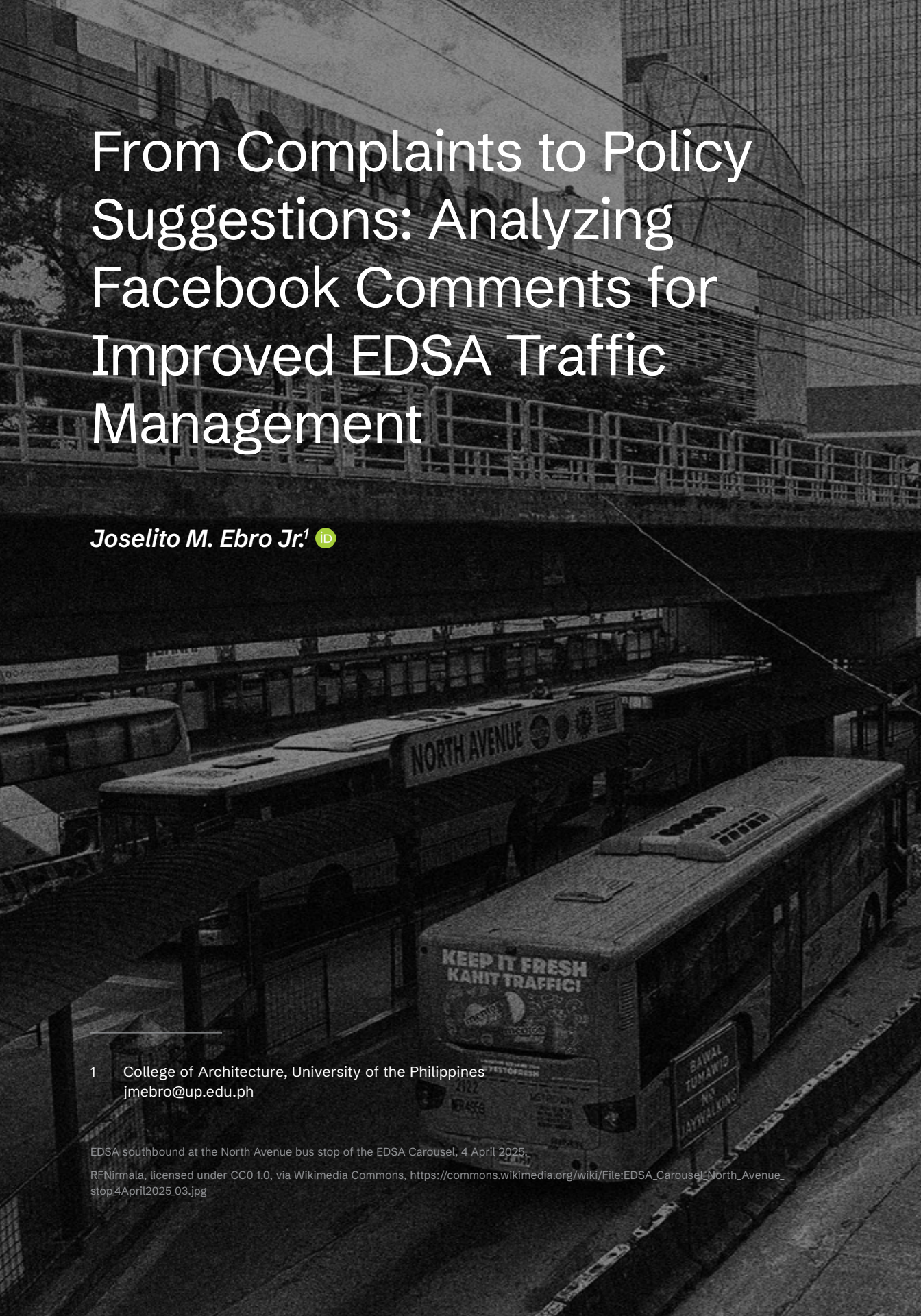
Stryker, Cole. n.d. "What Are Large Language Models (LLMs)?" IBM. Accessed 13 August 2026. https://www.ibm.com/think/topics/large-language-models.

West, Darrell M. 2018. *The Future of Work: Robots, AI, and Automation*. Brookings Institution Press.

From Complaints to Policy Suggestions: Analyzing Facebook Comments for Improved EDSA Traffic Management

Joselito M. Ebro Jr.[1]

1 College of Architecture, University of the Philippines
jmebro@up.edu.ph



EDSA southbound at the North Avenue bus stop of the EDSA Carousel, 4 April 2025.

RFNirmala, licensed under CC0 1.0, via Wikimedia Commons, https://commons.wikimedia.org/wiki/File:EDSA_Carousel_North_Avenue_stop_4April2025_03.jpg



Abstract

This study investigates the potential of social media discourse analysis to inform and enhance traffic management on Epifanio de los Santos Avenue (EDSA), a primary thoroughfare of Metro Manila. Despite numerous interventions, EDSA continues to face significant congestion challenges, necessitating innovative approaches to policy development. Using a multi-method qualitative approach combining a semi-structured interview with a Metropolitan Manila Development Authority (MMDA) representative and social media discourse analysis of 2,878 Facebook comments from the official MMDA account, the study examines public discourse surrounding the postponed ₱8.7 billion rehabilitation program to answer the question: What forms of policy-relevant discourse emerge in Facebook comments on the MMDA's EDSA rehabilitation announcements, and how might the MMDA integrate this discourse into traffic management policymaking?By analyzing public reactions to the rehabilitation program's announcements, proposed coding schemes, and eventual suspension, the findings reveal the untapped potential of social media comments. This "enhanced public discourse" encompasses three forms of engagement: technical critique of the policy, linking lived experiences to systemic problems, and the proposal of actionable interventions. These findings contribute to data-informed governance literature by proposing an integrative framework that leverages digital public discourse to develop more responsive and effective urban traffic management policies.

Keywords: Social media discourse, Transportation policy, Digital governance, MMDA, Citizen engagement

Introduction

The sharing of public sentiments towards transportation issues has shifted into the digital world. In recent years, social media platforms such as Facebook, X (formerly Twitter), and YouTube have played a significant role in hosting public grievances relating to traffic congestion. In the Philippines, social media pages such as Gadget Addict, popular for sharing traffic management updates along EDSA, for instance, have generated a total of 1.9 million followers on Facebook and 429,000 subscribers on YouTube. Similarly, VISOR, a Facebook page that addresses motorist-related concerns, received over 1.2 million followers. In these digital spaces, participants update traffic conditions and share grievances on the broader transportation issues that have hounded the Philippines in recent decades. With pages like these continuing to increase, exploring the impact of digital platforms on enhancing traffic management remains timely and relevant, leading to the question of what implications and uses government offices draw from these spaces.

The Metropolitan Manila Development Authority (MMDA) has been a pioneer among Philippine government agencies in embracing digital platforms, maintaining an active presence across multiple channels, and establishing specialized units such as the Public Concerns and Response Management Unit (PCRMU) for systematic public engagement. This forward-thinking approach has already yielded concrete results, such as the implementation of MMDA Regulation 23-002, increasing the penalties for Bus Way or EDSA Carousel violators. The decision was made after the agency noticed an increase in busway violations, along with the prevailing social media comments suggesting that violators were willing to pay the cheap penalty. Such information prompted the agency to raise the penalty for the first offense from ₱1,000 to ₱5,000. Subsequent offenses would merit harsher penalties, including a ₱30,000 penalty and revocation of the driver's license for those who committed the violation four times (MMDA representative 2024). While this example demonstrated MMDA's capacity to leverage social media insights, it also suggests significant untapped potential in the sophisticated policy discourse occurring on these platforms. In addition, it raises important questions about how systematic analysis of such digital conversations could further enhance traffic management policies.

This study explores this untapped potential by examining the case of the postponed rehabilitation plan for EDSA. Given EDSA's historical and commercial significance, it has long battled severe congestion challenges, earning the moniker "Carmageddon" for its extreme traffic conditions that affect hundreds of thousands of daily commuters (Boquet 2013). With the MMDA managing the daily flow of 464,000 vehicles (Dizon 2024) on this vital corridor, the agency's capacity to leverage digital insights becomes particularly critical for effective traffic management. In addition, the postponement of its ₱8.7 billion rehabilitation program and the subsequent

public discourse surrounding proposed traffic mitigation measures provide a well-bounded case to capture and analyze public discourse, thereby improving traffic management policies.

This research aims to bridge the gap between MMDA's current social media capabilities and the rich policy intelligence embedded in public discourse by answering the question: What forms of policy-relevant discourse emerge in Facebook comments on the MMDA's EDSA rehabilitation announcements, and how might the MMDA integrate this discourse into traffic management policymaking? Using a multi-method qualitative approach, which combines an institutional interview with systematic thematic analysis of 2,878 Facebook comments from MMDA's official page, this study demonstrates how government agencies can transform social media platforms from reactive complaint channels into proactive policy development tools. Moreover, this research provides a replicable framework for extracting actionable policy insights from digital public discourse, contributing to data-informed governance, offering MMDA and similar agencies a practical pathway toward more responsive, citizen-centric traffic management policies.

The EDSA Rehabilitation Program Crisis

The Marcos Jr. administration has identified the comprehensive rehabilitation of EDSA as a priority infrastructure project under the "Build Better More" program, illustrating the administration's commitment to improving traffic management in Metro Manila. With a substantial budget of 8.7 billion pesos, the initiative calls for the complete reconstruction of the 23.8-kilometer thoroughfare stretching from Caloocan City to Pasay City (Cabuenas 2025). These improvements encompass the replacement of concrete pavement to enhance traffic flow, including the modernization of drainage systems to mitigate flooding and the development of pedestrian infrastructure, transforming EDSA into a more walkable and pedestrian-friendly space.

For some, the rehabilitation was too late, given that since 2014, the Japan International Cooperation Agency (JICA) has mentioned that Metro Manila loses ₱2.4 billion a day and could increase to ₱6 billion by 2030 if no action is taken to address traffic congestion (Torres 2014). President Marcos Jr., however, emphasized that it is needed and requires a cohesive solution to a lasting problem, as implementing such a massive project would deal with numerous challenges. Recognizing the potential for widespread disruption, the administration formulated a multipronged strategy to manage the project's impact. Key mitigation measures include implementing a new traffic coding scheme and providing free access to the Skyway to divert vehicular volume and ease congestion. To support public transportation capacity, the plan involves deploying 100 additional buses on alternative routes and supplementing the Metro Rail Transit Line 3 (MRT-3) with additional train sets.

On the part of the citizens, the people, along with other advocacy groups, voiced frustrations on social media regarding the proposed new traffic coding scheme, with many motorists finding it confusing and poorly explained. This confusion generated online criticism, as social media comments express fears that the rehabilitation could inadvertently worsen congestion. As the policy went viral, intensified pressure on the Marcos administration for greater clarity, more robust public consultation, and a well-defined plan before proceeding with the monumental undertaking became more prominent.

Despite its high-priority status, the rehabilitation was officially postponed on 1 June 2025, pending a solid rerouting plan (PCO 2025).[2] The announcement of the delay, coupled with the proposed mitigation plans, quickly sparked considerable public discourse, particularly across social media platforms.

From Digital Complaints to Policy Intelligence

As the goal of this study is to explore and suggest practical approaches for utilizing social media discourse to enhance traffic management on EDSA, this research is anchored in the growing body of literature that recognizes social media as a credible and potent tool for transport planners, operators, and policymakers. Contrary to the prevailing impression that they are inundated with traffic complaints from their users, this section aims to highlight not only the process of generating traffic policy suggestions from social media but also to argue for the needed paradigm shift in viewing social media data as sources of "enhanced public discourse" framed to enhance traffic policies.

Technology has not only changed the way we live, but also the way we interact with one another and with institutions. In the field of digital governance, information and communication technologies (ICT) have enhanced citizen engagement in public policymaking by fostering more effective, responsive, and inclusive models. Driven by the high reach and rapid escalation capabilities of these platforms, digital platforms such as X (formerly Twitter), Facebook, and Instagram have become the go-to platforms for people to express their opinions, making them significant platforms for people to voice their complaints about products and services.

2 Rehabilitation eventually began on 24 December 2025 under a revised plan. Roadworks on the first phase, from Roxas Boulevard to Orense, were substantially complete by April 2026, about two months ahead of schedule, and work on the second phase was under way by October 2026 (Romero-Alcazar 2026; Serquiña 2026). The discourse analyzed in this study concerns the original program, from its announcement in January 2025 to its suspension in June 2025.

On the institutional side, however, managing and addressing the flow of these opinions requires a more efficient approach. The field of communication research adopts two basic positions: passive monitoring strategy and active dialog strategy (Stauss and Seidel 2019). While active dialog strategy provides a more engaged way of addressing grievances via prompt and detailed responses to social media complaints, Golmohammadi et al. (2021) find that potential exposure of complaints via "complaint publicization" results in a negative impact not only on people's perceived value and quality but also increases future complaints. Applying these scenarios to the field of transportation studies raises similar concerns, given the widespread use of social media in managing traffic worldwide.

The convergence of social media and transportation research has evolved through distinct phases. As noted by Zayet et al. (2021), research activity in transportation-related research based on social media analysis (TRR-SMA) has grown exponentially, with studies focusing on real-time traffic monitoring, event detection, and commuter feedback analysis. Transportation agencies are increasingly leveraging social media analytics to enhance their operations, communication, and decision-making processes. These key applications appear in real-time information sharing (Alam and Sadri 2024; Cottrill et al. 2017), public engagement and feedback (Giancristofaro and Panangadan 2016; Manetti et al. 2017), and transportation planning and decision-making (Sarram and Ivey 2022).

However, most studies have concentrated on operational applications rather than exploring the potential for policy development. This gap represents an opportunity in modern governance, where government agencies, like the MMDA, can leverage proactive policy intelligence embedded in citizen discourse. This disconnect not only underutilizes valuable public intelligence but also perpetuates top-down policy development approaches that may miss ground-level innovations and community-generated solutions. By bridging this gap, transportation agencies can transform social media platforms from passive complaint repositories into active policy co-creation spaces, thereby advancing citizen-centric governance models that harness collective intelligence for more informed and responsive urban traffic management.

As social media empowers public engagement and communication through user interactions, direct communication between traffic management authorities and the public enables the dissemination of actionable information and the collection of user feedback (Wang et al. 2017; Wojtowicz and Wallace 2016). In China, local governments have implemented various models that promote e-participatory policymaking through increased transparency, soliciting citizens' input, and facilitating substantive discussions (Xin and Huang 2025).

These developments are grounded in evolving digital governance theories that conceptualize technology as transforming the relationship between citizens and government institutions. Bekkers et al. (2013) identified social media monitoring as a form of "responsive governance," where digital platforms enable governments to move beyond traditional reactive approaches toward proactive engagement with citizen concerns. This framework positions social media not merely as communication tools but also as institutional mechanisms for continuous democratic dialogue that can inform policy development processes. The participatory governance literature supports the concept of Nam (2012), known as "citizen-sourcing," where government agencies systematically capture and integrate people's insights, bringing citizen co-production to the forefront of policy development in the digital age.

These approaches demonstrate that digital platforms have a positive impact on citizen engagement by providing opportunities for individuals to contribute to the development of public policy. In the context of this study, the feature of collecting and analyzing social media insights through public sentiment and discourse offers untapped potential for informing the design and implementation of traffic management policies. This endeavor, however, is not without challenges.

Building on Dong and Lian's (2021) Social Media–Based Public Opinion (SMPO), this study addresses four critical challenges they identified: data collection complexity, managing unstructured data, methodological rigor, and ethical considerations. To address these concerns, the authors promoted leveraging multiple data sources, optimizing keyword selection for data collection, using digital tools to accelerate data mining, and paying more attention to the uses of social media in various societal applications. These recommendations serve as anchors for this study, which aims to transform raw and complaint-perceived public discourse into actionable policy insights that MMDA and other government agencies can leverage to enhance traffic management policies by leveraging the untapped potential of social media data.

By employing a systematic methodological approach, such as Braun and Clarke's (2006) thematic analysis, this study advances beyond TRR-SMA applications by conceptualizing social media platforms as sites of "enhanced public discourse" encompassing characteristics such as sophisticated policy analysis that goes beyond complaints to technical critiques, the linking of lived experiences to systemic problems, and the proposal of actionable interventions. This framework positions social media as a dynamic space for policy co-creation, rather than a passive feedback mechanism, thereby bridging the identified gap to extend the digital capabilities of government agencies. The following section outlines the methodology applied in the study, aiming to guide in generating policy-oriented discourse from social media.

Methodology

The systematic analysis of public discourse surrounding the EDSA rehabilitation program employed a multi-method qualitative approach combining an institutional interview with thematic analysis of social media comments to examine the gap between current practice and potential policy applications. The research methodology began with a qualitative phase to establish an institutional baseline. An interview was conducted with a representative from the MMDA's Public Concerns and Response Management Unit (PCRMU) to understand how the agency currently leverages social media data in its daily operations. This initial step provided essential context on the MMDA's existing public engagement and data management practices, setting the stage for the subsequent collection of public discourse.

Following the interview, the data collection phase focused on capturing public discourse from MMDA's official Facebook page during the project's key period, from its announcement on 8 January 2025 to its suspension on 1 June 2025. Using the keywords "edsa" and "edsa rehab," the top 10 posts with the highest number of top-level comments were selected, yielding a dataset of 2,878 comments. The dataset consists of top-level comments only; replies were excluded. The comments were retrieved manually from May to June 2025 using screenshots of comments. All were publicly posted on the agency's official page in response to its announcements and were collected without interaction with commenters. Usernames and other identifying details are omitted from this study, and comments written in Filipino or *Taglish* are translated into English by the author. To reduce traceability, three English-language comments in Table 1 are presented as close paraphrases. The data shows that public engagement peaked around critical events, particularly on 26 May 2025, when the new coding scheme was announced (1,182 comments), and on 1 June 2025, following the news of the project's suspension (612 comments). This targeted approach ensured the collected data was rich with reactions to specific policy decisions.

The 2,878 Facebook comments were analyzed using MAXQDA, informed by Braun and Clarke's (2006) thematic analysis. The analysis process involved four key stages. First, a word cloud was generated to identify the most frequently occurring terms, revealing key concepts such as "EDSA" (571 mentions), "coding" (206 mentions), and "MMDA" (269 mentions). Next, using MAXQDA's auto-coding function, 67 distinct codes were generated based on identified keywords, with a minimum threshold of 25 mentions to ensure analytical rigor. Code-occurrence and relationship networks were visualized to identify interconnected discourse patterns and thematic clusters. Lastly, the data were systematically grouped and synthesized into nine overarching themes, revealing three dimensions of public engagement: technical critique of the policy, linking lived experiences to systemic problems, and the proposal of actionable interventions.

This methodological approach successfully captured both the institutional context of MMDA's current social media practices and the discourses occurring on their platforms during critical policy moments. The systematic progression from keyword identification to thematic synthesis revealed interconnected patterns of citizen engagement, demonstrating notable untapped potential for policy development. These analytical findings, presented in the following results section, illustrate how citizens, through Facebook comments, engage in three distinct dimensions of policy discourse that could significantly enhance traffic management strategies.

Results: Institutional Capacity and Citizen Policy Intelligence

The findings are presented in two interconnected parts: first, an institutional analysis of MMDA's current social media practices, establishing the baseline for digital engagement capabilities; and second, a comprehensive thematic analysis of public discourse surrounding the EDSA Rehabilitation Program, revealing the untapped potential for enhanced public discourse instrumental for improving policy intelligence.

MMDA's Social Media and Traffic Management Approach

The MMDA, as the primary agency mandated with traffic management in the National Capital Region, has established itself as a digital governance pioneer among Philippine government institutions. The agency maintains an active and strategic presence across multiple social media platforms, including Facebook, X (formerly Twitter), TikTok, Instagram, and Viber, demonstrating institutional commitment to digital citizen engagement. This multi-platform approach reflects an understanding that effective communication requires meeting citizens where they are most active online (MMDA representative 2024).

Central to this digital strategy is the Public Concerns and Response Management Unit (PCRMU), a specialized institutional innovation that serves as the systematic bridge between citizen feedback and agency operations. The PCRMU represents more than a customer service function; it embodies a strategic institutional design for continuous public engagement. With a mandate that encompasses collating complaints, concerns, and suggestions from all agency communication channels, the PCRMU functions as a critical component in the agency's digital operations.

The operational architecture of the PCRMU reveals sophisticated institutional capacity for digital engagement. The unit operates through a team of twenty personnel working in three rotating shifts, ensuring 24/7 monitoring and response capabilities (MMDA representative 2024). This continuous operation demonstrates

institutional recognition that public discourse and traffic management concerns operate beyond traditional business hours. However, there is more to enhance.

The agency's current digital practices are characterized by three primary functions: information dissemination, real-time traffic updates, and basic feedback collection (MMDA 2023). While functionally effective for immediate incident response, this approach requires a more sophisticated, data-driven policy generation. The agency primarily responds to specific citizen reports, such as misconduct by traffic enforcers or immediate traffic incidents, creating one-way communication that, while valuable, underutilizes the rich policy discourse occurring on these platforms.

Looking toward institutional evolution, the MMDA leadership has expressed clear intentions to enhance its methodological approach through systematic analysis, particularly in data mining from social media (MMDA representative 2024). This forward-thinking perspective aims to transform the agency's digital platforms from basic complaint mechanisms into sophisticated tools for policy development and evaluation. The existing PCRMU infrastructure provides a strong foundation for this enhancement, suggesting that the institutional capacity for advanced social media analytics already exists within the agency's operational framework.

Enhanced Public Discourse and its Three Dimensions

The systematic thematic analysis of 2,878 Facebook comments fundamentally challenges conventional assumptions about social media discourse. Rather than serving as a simple repository of complaints and grievances, the platform hosted sophisticated policy discussions that demonstrated commenters' capacity for nuanced analysis, strategic thinking, and constructive engagement with complex traffic management challenges.

The analytical process, using MAXQDA, informed by Braun and Clarke's (2006) thematic analysis, revealed remarkable depth in public discourse. Initial word frequency analysis revealed key engagement patterns, with "EDSA" appearing 571 times, "coding" 206 times, and "MMDA" 269 times, indicating sustained and focused discussion rather than scattered complaints. The systematic progression from 67 distinct codes to nine overarching themes generated 3,446 coded segments, with many comments spanning multiple thematic categories, demonstrating the interconnected nature of citizen thinking about traffic policy. Table 1 outlines the descriptions, representative quotes, and the specific number of coded segments for each identified theme.

Table 1. Thematic Analysis Results

Theme	Description	Representative Quote	Coded segments
EDSA	Comments focused on EDSA as infrastructure, crossing permissions, and the geographic scope of policies.	"Is the odd-even scheme on EDSA? What about other roads? Is the number coding scheme retained there?" (trans.)	529
MMDA	Direct critiques and questions about MMDA's authority, policy clarity, and coordination with other agencies.	"Coding never fixed traffic; it just pushed drivers to buy another car. And now the MMDA believes more coding will help? Brilliant." (par.)	333
Coding	Technical analysis of coding schemes, window hours, and policy implementation confusion.	"What's happening with the existing number coding scheme, MMDA? Will it still be in effect when the odd-even scheme is implemented on June 16?" (trans.)	273
Infrastructure	Alternative routes, road capacity, and construction impact on daily operations.	"Is this the only solution you've found? Have you provided alternative routes for motorists?" (trans.)	98
Social Dynamics and Enforcement	Motorist-government relations, enforcement concerns, and consultation processes.	"So, that means motorists are the sacrifice for your experiment? You should also make any violations that occur during that time a dry run." (trans.)	199
Bike/Motor	Two-wheeled vehicle policies, lane sharing, and motorcycle-specific concerns.	"Put a motorcycle lane on EDSA so that motorcycles won't pass through the bike lane... supposedly." (trans.)	86
Commuter Experience	Window hours, fairness concerns, and daily commuting challenges	"You're not the only ones who work, MMDA! Three times a week and 24 hours! That is not fair!" (trans.)	143
Traffic Management	Comprehensive policy questions, exemptions, and systematic traffic solutions	"If you really want to ease congestion on the main roads, start by improving public transport." (par.)	667

Theme	Description	Representative Quote	Coded segments
Vehicle Technology	Mixed theme covering various vehicle types, delivery services, and transport technology	"The scheme was dropped from Pasig's roads for a reason: it didn't work." (par.)	1,118

Analysis of 2,878 Facebook comments from MMDA's official page (January–June 2025) using MAXQDA software and Braun and Clarke's (2006) thematic analysis. Coded segments represent individual comment portions that contained thematic content, with many comments spanning multiple themes.

The analysis revealed that citizen discourse demonstrated systematic policy thinking through interconnected thematic relationships. As mentioned previously, words like "EDSA," "coding," "MMDA," and "traffic" appeared together consistently, while "bike/motor" connected with "coding" and "infrastructure," indicating citizen understanding of policy interdependencies that traditional mechanisms might address only in fragments. Rather than thinking about transportation issues in isolation, commenters demonstrated the capacity for systems thinking that connects infrastructure, regulation, enforcement, and behavioral change into a comprehensive policy framework.

The findings complicate assumptions about social media as a space for uninformed complaint. Instead, they suggest that online spaces can serve as a repository of sophisticated policy intelligence that can inform evidence-based decision-making processes. These analytical findings demonstrate how citizens engage in three distinct dimensions of policy discourse that could significantly enhance MMDA's traffic management strategies, as explored in the following discussion section.

The Three Dimensions of Enhanced Public Discourse

The analysis of 2,878 Facebook comments surrounding the EDSA rehabilitation program suggests social media platforms serve as repositories of "enhanced public discourse" that extends far beyond the conventional depiction of being a space for digital complaints. By moving beyond the operational focus that exists in current transportation social media research (Zayet et al. 2021), citizens demonstrate the ability to elicit discourses that move beyond simple complaints to more constructive criticism, founded on commenters' ability to provide technical critiques, link personal experiences, and create actionable interventions. These findings advance beyond current literature limitations by providing empirical evidence that social media platforms contain rich policy intelligence that can inform traffic management decision-making. The following analysis examines each dimension in detail, demonstrating how commenters engage with policy complexities in ways that can significantly enhance MMDA's traffic management strategies.

Offering Technical Critique of Policy

Commenters demonstrated sophisticated engagement with technical policy details, moving far beyond simple complaints to offer nuanced critiques that reveal a deep understanding of policy mechanics and systemic implications. Drawing on Bekkers et al. (2013, 336), this study suggests that social media discourse can serve as a "communicative approach" with the ultimate goal of coproducing policies on a shared understanding of problems. The comments in this study went beyond simple feedback, focusing on technical policy details, infrastructure solutions, technological innovations, and operational complexities.

For example, one commenter offered a nuanced critique of the electric vehicle exemption in the coding scheme, arguing that its purpose should be volume reduction, not just pollution: "MMDA, I hope that even full-electric and hybrid vehicles will be included in these same coding schemes. I believe the point behind this is to control the volume of cars, not to lessen pollution. . . ." This comment demonstrates a sophisticated understanding of policy objectives, distinguishing between environmental goals and traffic management goals.

Another user contextualized EDSA's traffic within broader national issues of urban planning and rural development, suggesting that decongestion requires addressing the root causes, such as migration patterns. Additionally, commenters engaged with the technical complexities of policy implementation, as seen in comments questioning the coordination between different coding schemes: "What's happening with the existing number coding scheme, MMDA? Will it still be in effect when the odd-even scheme is implemented on June 16?" These examples reveal commenters who understand policy interdependencies and implementation challenges rather than simply expressing frustration with traffic conditions.

This sophisticated engagement aligns with Chung and Zeng's (2016) framework of "social media-based public policy informatics." Their approach establishes social media as a legitimate source for policy analysis, providing systematic methods for extracting insights from complex online discourse. Like the electric vehicle exemption critique comment, citizens can demonstrate technical understanding of policy objectives while analyzing relationships between environmental goals and traffic management. These findings reveal significant opportunities for enhancing MMDA's strategy. The agency's PCRMU already operates 24/7 monitoring with twenty personnel, demonstrating a strong institutional foundation. While current practice effectively handles information dissemination and feedback collection, the sophisticated discourse in 2,878 comments represents valuable untapped potential. By establishing a dedicated analytical unit within PCRMU, MMDA can expand social media capabilities from personal communication into comprehensive policy development tools. Beyond technical policy analysis, commenters demonstrated a

higher level of engagement by systematically connecting their lived experiences to broader policy implications and infrastructure solutions.

Linking Lived Experiences to Systemic Problems

The discourse revealed commenters' capacity to transform personal commuting experiences into comprehensive policy recommendations that address systemic infrastructure challenges across multiple domains. This engagement pattern illustrates the policy-relevant potential that Dong and Lian (2021) identified in social media discourse, where systematic analysis can yield actionable insights for policy adjustments and informed decision-making.

Commenters demonstrated interconnected thinking where personal experiences were directly linked to infrastructure needs and policy suggestions. One commenter proposed systematic work schedule changes: "We should change the work hours in Manila. If we adjust the start and end times for work, vehicles on EDSA . . . won't all converge at the same time... Additionally, we should implement designated nighttime hours for truck deliveries." Another user connected the proliferation of private vehicles directly to the inadequacies of public transportation, arguing that policy should address the root cause rather than penalizing the effect: ". . . the increase in the number of private vehicles . . . isn't because people want to buy cars . . . It's because public transportation in the Philippines . . . is unreliable, inefficient, and broken."

These comments are consistent with Wang et al.'s (2017) finding that social media enables direct communication between traffic management authorities and the public for actionable information collection, transforming individual commuting frustrations into systemic policy recommendations. This networking effect aligns with Lv et al.'s analysis of how social media creates collective intelligence for transportation research. Their work describes social media as a real-time source that treats "every user as a social sensor to detect the real world at any time" (2017, 19), effectively aggregating individual experiences into patterns that reveal systemic transportation issues. The EDSA comments exemplify this collective intelligence, where personal transportation experiences aggregate into comprehensive policy recommendations.

These proactive engagements present valuable opportunities for MMDA's policy development process. Commenters' ability to connect personal experiences with systemic solutions provides the agency with ground-level intelligence that traditional consultation methods often miss. Additionally, the agency's 24/7 monitoring capacity can be enhanced to identify and categorize experience-based policy recommendations systematically. Rather than treating such comments as individual complaints, MMDA can aggregate similar suggestions to identify recurring themes

and viable policy alternatives. This approach transforms citizen experience into a continuous policy feedback loop, enabling the agency to develop more responsive and contextually grounded traffic management strategies.

The most striking dimension of citizen engagement emerged when discourse moved beyond analysis and proactive suggestions to concrete, actionable solutions for complex infrastructure challenges.

Proposing Actionable Interventions

Commenters demonstrated a remarkable capacity to propose specific, actionable interventions rather than merely identifying problems. This solution-oriented approach represents the practical culmination of SMPO framework implementation, where systematic analysis reveals commenters offering concrete solutions and detailed suggestions that can directly inform policy action.

One commenter proposed a specific construction methodology: "MMDA should do it by section and one lane at a time, a 500-meter stretch, with concrete pouring done at night . . . Nothing is complicated here if only MMDA would think outside the box and without corruption being the hindrance." Another identified the poor quality of road maintenance as the primary cause of congestion, questioning the logic of proposed detour routes. This discourse pattern addresses the gap identified by Wojtowicz and Wallace (2016), who found that transportation agencies struggle to utilize incoming social media messages for actionable solutions. The EDSA comments demonstrate exactly the solution-oriented engagement that agencies need systematic frameworks to capture and analyze.

This solution-oriented discourse represented the highest-value intelligence for MMDA's strategic planning processes. Commenters' concrete recommendations can provide the agency with experience-based insights that complement technical assessments. These solutions emerge from lived experience with traffic patterns, infrastructure failures, and operational inefficiencies, offering MMDA practical alternatives that traditional planning processes might overlook.

MMDA's institutional framework is well-positioned to capture and evaluate these solutions systematically. The PCRMU's existing analytical capabilities can be enhanced to create a solution repository, categorizing citizen recommendations by feasibility, cost, and alignment with agency priorities. This approach transforms social media from a feedback channel into a continuous innovation pipeline. By establishing protocols to evaluate and pilot citizen-proposed solutions, MMDA can leverage collective public intelligence to accelerate problem-solving while building stronger community engagement in traffic management improvements.

Toward Data-Informed Governance

This study answers the question: What forms of policy-relevant discourse emerge in Facebook comments on the MMDA's EDSA rehabilitation announcements, and how might the MMDA integrate this discourse into traffic management policymaking? The analysis of 2,878 Facebook comments reveals that commenters engage in sophisticated, solution-oriented discussions that extend far beyond simple complaints, encompassing technical policy critiques, infrastructure recommendations, and operational innovations.

The three dimensions of enhanced public discourse, namely technical critique of the policy, linking lived experiences to systemic problems, and the proposal of actionable interventions, demonstrate the untapped potential of systematic social media analysis for policy development. As this framework promotes proactive analysis of social media data, government agencies can adopt protocols that transform basic communication channels into a dynamic policy intelligence system. For instance, MMDA's existing institutional capacity, particularly through the PCRMU's 24/7 monitoring infrastructure, provides a strong foundation for implementing systematic discourse analysis. By establishing dedicated analytical protocols to capture and evaluate citizen-generated policy insights, MMDA can leverage collective public intelligence to enhance the decision-making process while building stronger community engagement in traffic management.

These findings should be read within the study's limits. The dataset comprises top-level comments on ten posts from a single agency page during one policy episode, and its commenters are self-selected; because much of the discussion concerned a coding scheme for private vehicles, motorists' concerns are likely overrepresented relative to those of public transport users. The institutional account rests on the one interview with a PCRMU representative, conducted in May 2024, before the period analyzed. The keyword-based coding identifies the forms that policy-relevant discourse takes but not how common they are, and the quotations discussed above were chosen to illustrate these forms rather than to represent the typical comment. The study also does not examine whether any comment influenced the MMDA's subsequent decisions. Within these limits, the study shows that comments on an agency's page can carry policy-relevant reasoning worth collecting systematically. Future research could estimate how prevalent such reasoning is across more posts and platforms, and trace whether and how it enters agency decisions.

References

Alam, Md. Rakibul, and Arif Mohaimin Sadri. 2024. "Examining the Communication Pattern of Transportation and Transit Agencies on Twitter: A Longitudinal Study in the Emergence of COVID-19 on Twitter." *Transportation Research Record* 2678 (12): 13–29. https://doi.org/10.1177/03611981221082564.

Bekkers, Victor, Arthur Edwards, and Dennis de Kool. 2013. "Social Media Monitoring: Responsive Governance in the Shadow of Surveillance?" *Government Information Quarterly* 30 (4): 335–42. https://doi.org/10.1016/j.giq.2013.05.024.

Boquet, Yves. 2013. "Battling Congestion in Manila: The EDSA Problem." *Transport and Communications Bulletin for Asia and the Pacific*, no. 82: 45–59. https://www.unescap.org/sites/default/files/bulletin82_Article-4.pdf.

Braun, Virginia, and Victoria Clarke. 2006. "Using Thematic Analysis in Psychology." *Qualitative Research in Psychology* 3 (2): 77–101. https://doi.org/10.1191/1478088706qp063oa.

Cabuenas, Jon Viktor D. 2025. "Marcos: No EDSA Rehab Until Rerouting Plans 'Solid,' LGUs Ready." *GMA News Online*, 8 June 2025. https://www.gmanetwork.com/news/topstories/nation/948726/marcos-edsa-rehab-metro-manila/story/.

Chung, Wingyan, and Daniel Zeng. 2016. "Social-Media-Based Public Policy Informatics: Sentiment and Network Analyses of U.S. Immigration and Border Security." *Journal of the Association for Information Science and Technology* 67 (7): 1588–606. https://doi.org/10.1002/asi.23449.

Cottrill, Caitlin, Paul Gault, Godwin Yeboah, John D. Nelson, Jillian Anable, and Thomas Budd. 2017. "Tweeting Transit: An Examination of Social Media Strategies for Transport Information Management During a Large Event." *Transportation Research Part C: Emerging Technologies* 77: 421–32. https://doi.org/10.1016/j.trc.2017.02.008.

Dizon, David. 2024. "EDSA Overload: 464,000 Vehicles Plying EDSA Daily." *ABS-CBN News*, 9 December 2024. https://www.abs-cbn.com/news/nation/2024/12/9/edsa-overload-464-000-vehicles-plying-edsa-daily-0919.

Dong, Xuefan, and Ying Lian. 2021. "A Review of Social Media-Based Public Opinion Analyses: Challenges and Recommendations." *Technology in Society* 67:101724. https://doi.org/10.1016/j.techsoc.2021.101724.

Giancristofaro, G. T., and A. Panangadan. 2016. In "Predicting Sentiment Toward Transportation in Social Media Using Visual and Textual Features." *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, 2113–18. IEEE. https://doi.org/10.1109/ITSC.2016.7795898.

Golmohammadi, Alireza, Taha Havakhor, Dinesh K. Gauri, and Joseph Comprix. 2021. "Complaint Publicization in Social Media." *Journal of Marketing* 85 (6): 1–23. https://doi.org/10.1177/00222429211002183.

Lv, Yisheng, Yuanyuan Chen, Xiqiao Zhang, Yanjie Duan, and Naiqiang Li. 2017. "Social Media Based Transportation Research: The State of the Work and the Networking." *IEEE/CAA Journal of Automatica Sinica* 4 (1): 19–26. https://www.ieee-jas.com/en/article/id/52b08389-b6dd-4cf6-a9de-115a729ddc95.

Manetti, Giacomo, Marco Bellucci, and Luca Bagnoli. 2017. "Stakeholder Engagement and Public Information Through Social Media: A Study of Canadian and American Public Transportation Agencies." *American Review of Public Administration* 47 (8): 991–1009. https://doi.org/10.1177/0275074016649260.

MMDA (Metropolitan Manila Development Authority). 2023. *PAS-PIO Operations Manual*. Unpublished internal document.

MMDA representative. 2024. Personal interview, conducted online, 31 May 2024.

Nam, Taewoo. 2012. "Suggesting Frameworks of Citizen-Sourcing via Government 2.0." *Government Information Quarterly*, 29 (1): 12–20. https://doi.org/10.1016/j.giq.2011.07.005.

Presidential Communications Office. 2025. "PCO (Presidential Communications Office). 2025." Press release, 8 June 2025. https://pco.gov.ph/news_releases/pbbm-edsa-rehab-remains-suspended-without-solid-rerouting-plans/.

Romero-Alcazar, Maria Bernadette. 2026. "EDSA Rehab Enters 2nd Phase, Motorists Face Lane Closures." *Daily Tribune*, 17 October 2026. https://tribune.net.ph/2026/10/17/edsa-rehab-enters-2nd-phase-motorists-face-lane-closures.

Sarram, Golnaz, and Stephanie S. Ivey. 2022. "Evaluating the Potential of Online Review Data for Augmenting Traditional Transportation Planning Performance Management." *Journal of Urban Management* 11 (1): 123–36. https://doi.org/10.1016/j.jum.2022.01.001.

Serquiña, Mariel Celine. 2026. "Phase 1 of EDSA Rehab Nears Completion; DPWH Starts Bidding for Phase 2." *GMA News Online*, 6 April 2026. https://www.gmanetwork.com/news/topstories/metro/982777/phase-1-of-edsa-rehab-nears-completion-dpwh-starts-bidding-for-phase-2/story/.

Stauss, Bernd, and Wolfgang Seidel. 2019. *Effective Complaint Management: The Business Case for Customer Satisfaction*. 2nd ed. Springer International Publishing. https://doi.org/10.1007/978-3-319-98705-7.

Torres, Ted P. 2014. "NEDA Approves JICA Transport Infra Plan." *Philippine Star*, 3 September 2014. https://philstar.com/metro/2014/09/03/1364753/neda-approves-jica-transport-infra-plan.

Wang, Di, Ahmad Al-Rubaie, Sandra Stinčić Clarke, and John Davies. 2017. "Real-Time Traffic Event Detection from Social Media." *ACM Transactions on Internet Technology* 18 (1): 9:1–23. https://doi.org/10.1145/3122982.

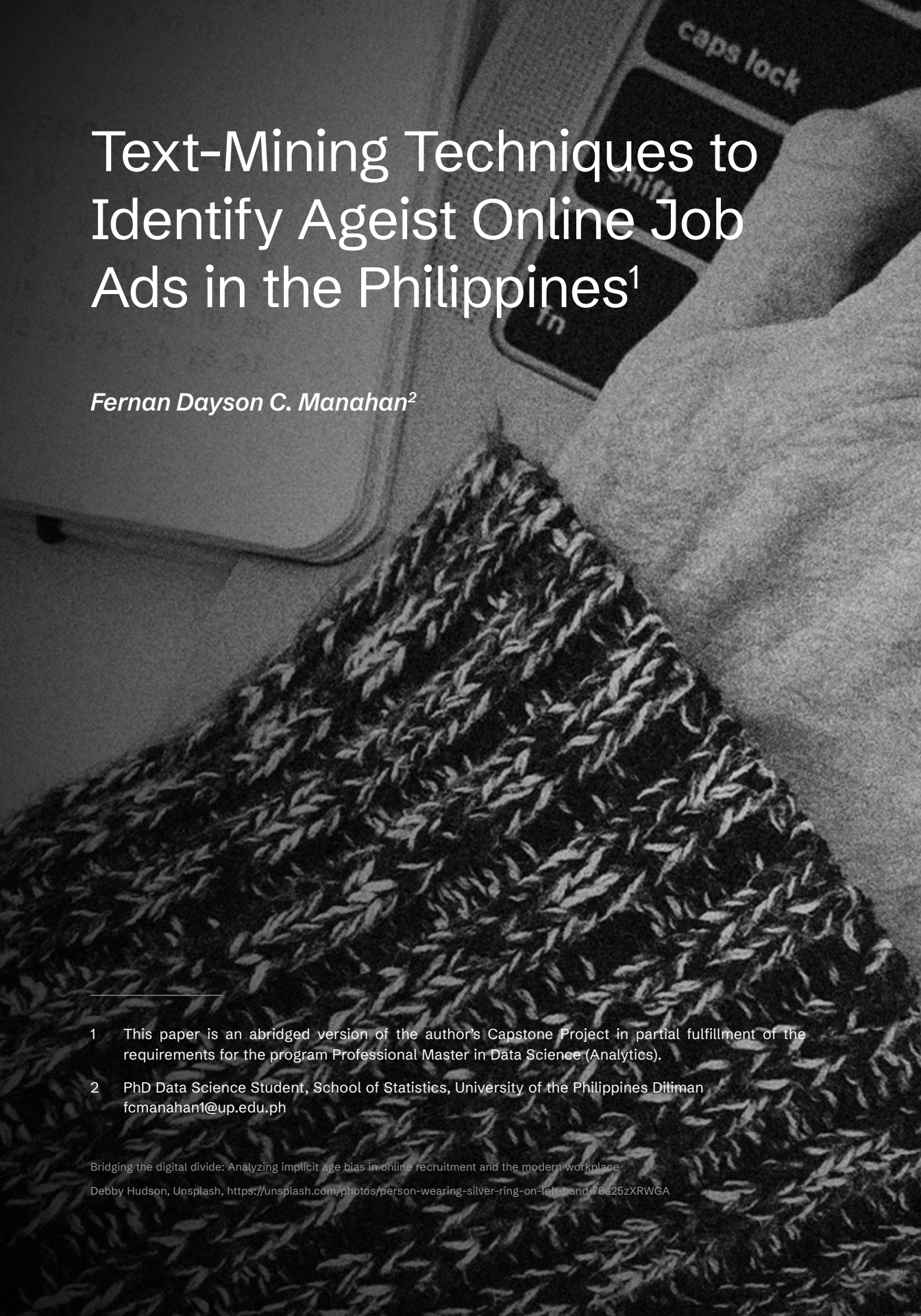
Wojtowicz, Jeffrey, and William A. Wallace. 2016. "Use of Social Media by Transportation Agencies for Traffic Management." *Transportation Research Record* 2551 (1): 82–89. https://doi.org/10.3141/2551-10.

Xin, Ge, and Jie Huang. 2025. "Making the People's Voice Heard: Pathways of E-participative Governance in China." *Journal of Chinese Governance* 10 (1): 33–56. https://doi.org/10.1080/23812346.2024.2400634.

Zayet, Tasnim M. A., Maizatul Akmar Ismail, Kasturi Dewi Varathan, et al. 2021. "Investigating Transportation Research Based on Social Media Analysis: A Systematic Mapping Review." *Scientometrics* 126 (8): 6383–421. https://doi.org/10.1007/s11192-021-04046-2.

Text-Mining Techniques to Identify Ageist Online Job Ads in the Philippines[1]

Fernan Dayson C. Manahan[2]

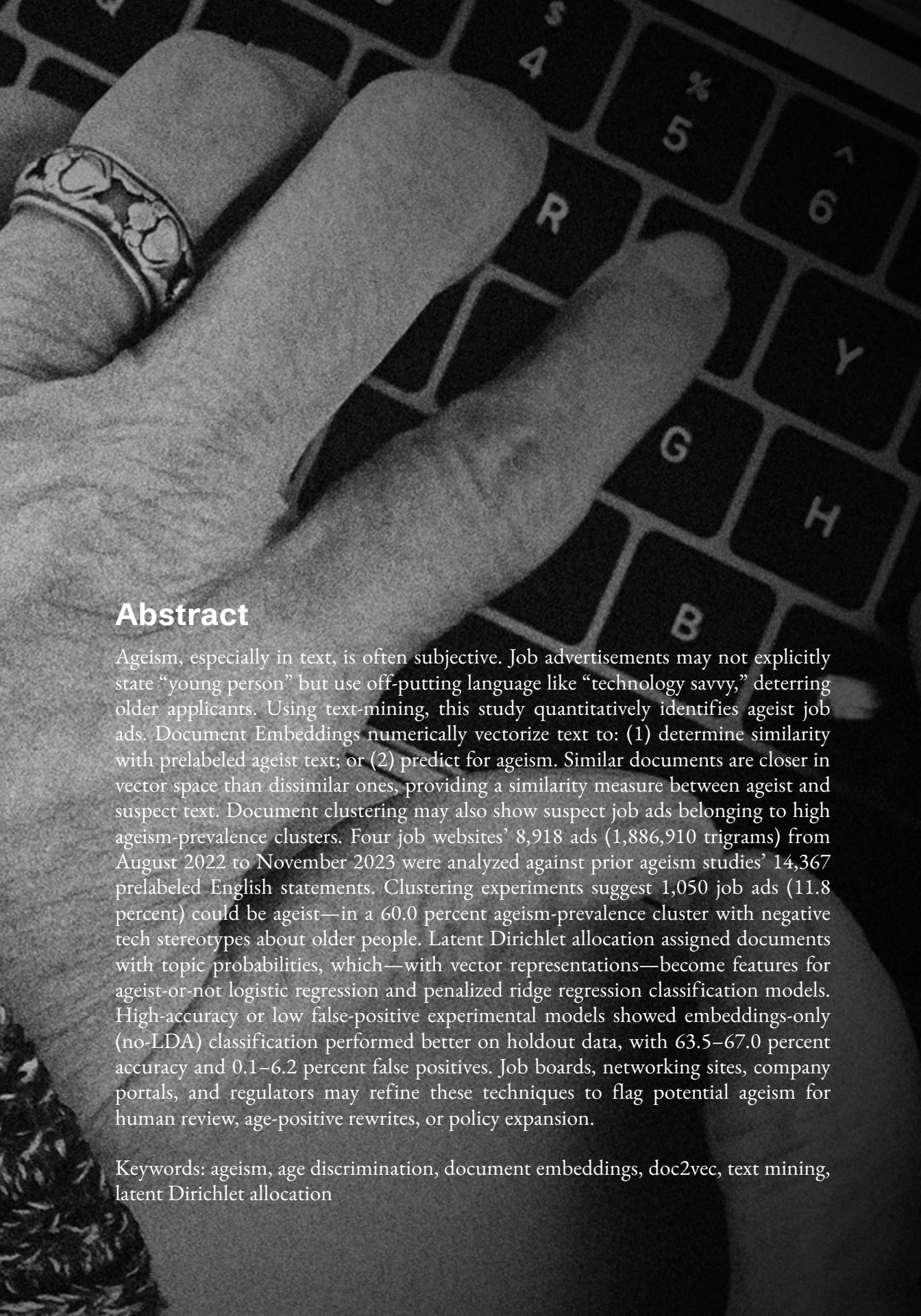


1 This paper is an abridged version of the author's Capstone Project in partial fulfillment of the requirements for the program Professional Master in Data Science (Analytics).

2 PhD Data Science Student, School of Statistics, University of the Philippines Diliman
fcmanahan1@up.edu.ph

Bridging the digital divide: Analyzing implicit age bias in online recruitment and the modern workplace

Debby Hudson, Unsplash, https://unsplash.com/photos/person-wearing-silver-ring-on-left-hand-7Ca25zXRWGA



Abstract

Ageism, especially in text, is often subjective. Job advertisements may not explicitly state "young person" but use off-putting language like "technology savvy," deterring older applicants. Using text-mining, this study quantitatively identifies ageist job ads. Document Embeddings numerically vectorize text to: (1) determine similarity with prelabeled ageist text; or (2) predict for ageism. Similar documents are closer in vector space than dissimilar ones, providing a similarity measure between ageist and suspect text. Document clustering may also show suspect job ads belonging to high ageism-prevalence clusters. Four job websites' 8,918 ads (1,886,910 trigrams) from August 2022 to November 2023 were analyzed against prior ageism studies' 14,367 prelabeled English statements. Clustering experiments suggest 1,050 job ads (11.8 percent) could be ageist—in a 60.0 percent ageism-prevalence cluster with negative tech stereotypes about older people. Latent Dirichlet allocation assigned documents with topic probabilities, which—with vector representations—become features for ageist-or-not logistic regression and penalized ridge regression classification models. High-accuracy or low false-positive experimental models showed embeddings-only (no-LDA) classification performed better on holdout data, with 63.5–67.0 percent accuracy and 0.1–6.2 percent false positives. Job boards, networking sites, company portals, and regulators may refine these techniques to flag potential ageism for human review, age-positive rewrites, or policy expansion.

Keywords: ageism, age discrimination, document embeddings, doc2vec, text mining, latent Dirichlet allocation

Introduction

The World Health Organization (WHO) (2021) defines ageism as age-based thoughts, feelings, and actions towards others or oneself. Ageism can be positive or negative, with stereotypes about young people being energetic but naive, and old people being wise but slow. However, ageism is generally considered negative and often directed towards older people (Butler 1969; Ayalon and Tesch-Römer 2018).

The Commission on Human Rights of the Philippines (CHRP 2019, 2021) states that denying older persons access to employment deprives them of confidence and autonomy. Eliminating ageism, especially in employment, is also crucial to the United Nations' Sustainable Development Goals of ending poverty, promoting inclusive employment, and reducing inequalities (United Nations 2023a, 2023b, 2023c).

This study focuses on ageism in Philippine Recruitment advertising, specifically in online job ads. It shows that statistical techniques can quantitatively identify ageist language in job ads; uses previous studies' techniques, such as word embeddings and cosine similarity (Burn et al. 2020; Centre for Ageing Better 2021a, 2021b); and explores alternatives like document clustering and using latent Dirichlet allocation topics as ageism-prediction features.

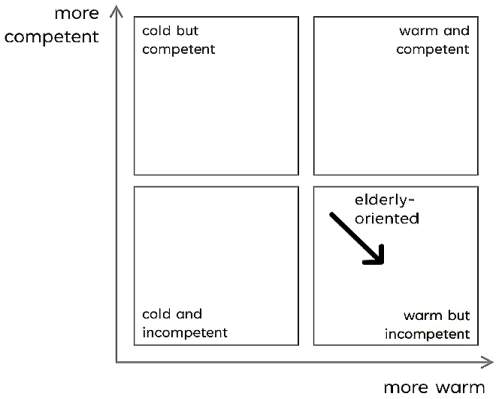
The study reports potential prevalence of ageist language in particular industries, job types, or job ad sections. It also provides ageism-detection alternatives to embeddings-based techniques and applies them to the Philippine setting.

Specific to the Philippine setting and to English job ads from 2022 to 2023, this study does not cover other forms of discrimination or stages of Recruitment beyond job ads. Job portals, hiring companies, and regulators can use this study's techniques to detect ageism in job ads and expand age-discrimination policies to include implicit ageist language.

Age Construct and Determining Ageism

Ayalon and Tesch-Römer (2018) provide theories explaining the origins of ageism, suggesting that certain words or phrases are youth- or elderly-oriented. The Stereotype Content Model classifies social groups based on warmth and competence, with the elderly typically seen as warm but incompetent. Text describing age-based social groups can be classified accordingly. For example, text near the word "old" in a vector space aligns with the "warm but incompetent" stereotype (Ayalon and Tesch-Römer 2018).

Figure 1. An illustration of the Stereotype Content Model of Ayalon and Tesch-Römer (2018) as a construct for elderly-oriented or youth-oriented text



Burn et al. (2020) analyzed job ads for ageism, using a word2vec model to determine cosine similarity between known ageist phrases and job-ad trigrams. They identified ageist trigrams and used them to predict callbacks for job applications. The Centre for Ageing Better (2021a, 2021b) found that younger-oriented words like "innovative" are more common in job ads than older-oriented words like "knowledgeable." They studied the impact of age-stereotypical language on job applicants' perceptions.

Díaz et al. (2018) found that sentiment analysis tools rated documents describing older people less positively. They suggested that bias could be reduced by resampling data from larger datasets. Their data is available for further research (Díaz 2020).

Combining these datasets with job-related ageist phrases from Philippine-based research (Lustañas 2019; DOLE 2017; Jobstreet Philippines Content Team 2022a, 2022b; CHRP 2019; Ramos 2022; Valdez et al. 2013) can create a large, job-related, and Philippine-specific textual dataset.

Ageism in the Philippines

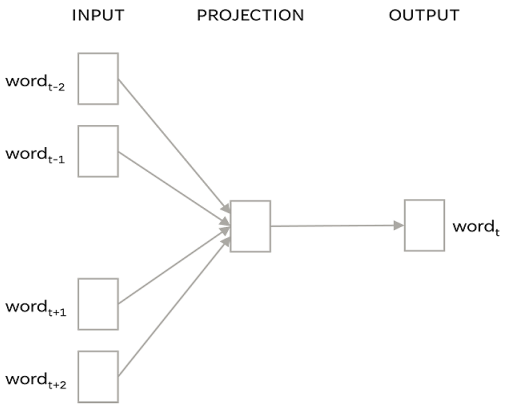
Lustañas (2019) used a critical discourse analysis (CDA) linguistic framework to study 30 online news articles and identified attributes associated with elderly Filipinos, such as weakness, susceptibility to crimes and accidents, incompetence, dependence on others, and poverty. The Philippine Department of Labor and Employment (DOLE 2017) issued rules under Republic Act (RA) No. 10911, prohibiting age-based preferences, limitations, or specifications in employment advertisements. Jobstreet Philippines Content Team (2022a, 2022b) reinforced these rules, advising employers to hire based on abilities, knowledge, skills, and qualifications, and job applicants to avoid ageist questions. Headhunter Manila Recruitment supported RA No. 10911 with tips for avoiding ageist language and promoting diversity and inclusivity (Ramos 2022).

The Commission on Human Rights of the Philippines (2019) stated that characterizing older Filipinos as weak, sickly, or frail constrains their access to jobs. Valdez et al. (2013) interviewed 12 Filipinos aged 40 to 59 and found mostly positive but some negative and neutral perceptions of aging, including responsibility, maturity, promise, autonomy, productivity, positivity, physiological and physical decline, and apprehension.

Word Representations in Vector Space (word2vec)

Mikolov et al. (2013) proposed two neural network model architectures to compute continuous vector representations of words, where the relative positions and operations on these word vectors have semantic significance, such as the analogy *king* − *man* + *woman* = *queen*. The Continuous Bag of Words (CBOW, illustrated in figure 2) model predicts the current word based on its context (the n words before and after it) using one-hot encoded vectors of each word in the context.

Figure 2. Continuous Bag of Words (CBOW) model architecture of Mikolov et al. (2013)



The model observes every word in the corpus and whether every other word is in its context window (see matrix A in figure 3).

Figure 3. Matrix A: Observations of words and whether other words are in its context window, demonstration by ritvikmath. (2021) of Mikolov et al. (2013)

Main	Context	Label
word$_t$	word$_{t\pm1}$	1
word$_t$	...	1
word$_t$	word$_{t\pm n}$	1
word$_t$	word$_{t\pm(n+1)}$	0
...	...	...

Two vector spaces are randomly initialized: one for each word as a main word (Matrix M in Figure 4) and another for each word as a context word (see matrix C in figure 4).

Figure 4. Illustration (using hypothetical values) of Matrices M and C: Vector spaces of each word as a main word and as a context word, respectively, demonstration by ritvikmath. (2021) of Mikolov et al. (2013).

Main

Word	V1	V2	...	Vk
$word_1$	0.06862	1.083	...	0.98084
$word_2$	0.66341	1.064	...	2.44544
...			...	
$word_t$	0.25467	2.2	...	1.03863
...	...	...	...	...

Context

Word	V1	V2	...	Vk
$word_1$	0.46883	1.94360	...	2.21079
$word_2$	1.87593	2.37459	...	1.77115
...			...	
$word_t$	0.21451	1.20045	...	0.28110
...	...	...	...	...

For each observation (Main-Context word pair) in matrix A, a score is computed as the sigmoid of the dot product of the vector of the main word and the vector of the context word:

$$\text{Score} = \sigma(\boldsymbol{V}_{\text{main}} \cdot \boldsymbol{V}_{\text{context}})$$

where:

- $\boldsymbol{V}_{\text{main}}$ is the vector embedding of a word as a main word
- $\boldsymbol{V}_{\text{context}}$ is the vector embedding of a word as a context word
- $\boldsymbol{V}_{\text{main}} \cdot \boldsymbol{V}_{\text{context}} = \sum_{i=1}^{k} \boldsymbol{V}_{\text{main}_i} \boldsymbol{V}_{\text{context}_i}$

An error is computed as the difference between the Label (1 or 0) and the computed score (as in figure 5).

Figure 5. Illustration (using hypothetical values) of Score and Error for each Main-Context word pair, demonstration by ritvikmath. (2021) of Mikolov et al. (2013)

Main	Context	Label	Score	Error
$word_1$	$word_2$	1	xxxx	xxxx
$word_1$	...	1	...	...
$word_1$	$word_n$	1		
$word_1$	$word_{n+1}$	0		
...	...	...		
$word_t$	$word_{t\pm1}$	1		
$word_t$	...	1		
$word_t$	$word_{t\pm n}$	1		
$word_t$	$word_{t\pm(n+1)}$	0		
...	...	...		

Matrix M: Main

Word	V1	V2	...	Vk
$word_1$	0.06862	1.083	...	0.98084
$word_2$	0.66341	1.064	...	2.44544
...			...	
$word_t$	0.25467	2.2	...	1.03863
...	...	...	...	...

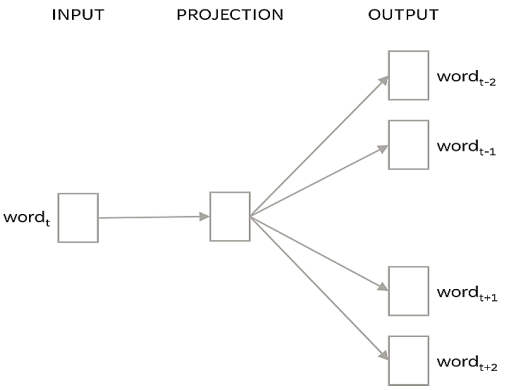
Matrix C: Context

Word	V1	V2	...	Vk
$word_1$	0.46883	1.94360	...	2.21079
$word_2$	1.87593	2.37459	...	1.77115
...			...	
$word_t$	0.21451	1.20045	...	0.28110
...	...	...	...	...

The vector spaces are iteratively adjusted during training to minimize the error. The resulting vector space of each word as a main word (matrix M) will be the final vectors for each word in the corpus.

The Skip-Gram model architecture (illustrated in figure 6) works similarly, except that context words are predicted given the main word. CBOW runs faster, while Skip-Gram performs better on rare words.

Figure 6. Skip-Gram model architecture of Mikolov et al. (2013)



Word2vec vector coordinates do not have an inherent per-observation meaning on their own but have meaning relative to the vector coordinates of other words in the corpus. They position words among other words in the corpus, placing similar words nearer and dissimilar words farther apart. Mikolov et al. (2013) noted the need to increase the word2vec training data and its vector dimensionality to maximize prediction accuracy. For smaller datasets, the greatest increase in accuracy occurs when raising vector dimensionality to 100, with smaller improvements beyond 300. Word2vec and doc2vec R package author Wijffels (n.d.) concurs that for smaller datasets, higher vector dimensionality does not always mean better.

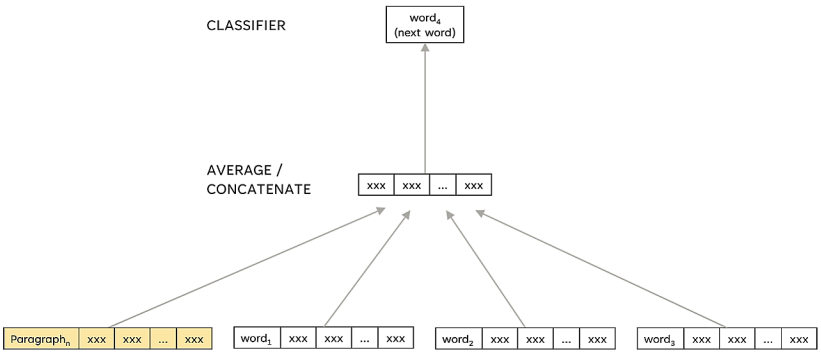
Paragraph Representations in Vector Space (doc2vec)

Le and Mikolov (2014) introduced the Paragraph Vector (PV) to generalize word embeddings to entire document embeddings. Unlike bag-of-words techniques, which lose semantic context, PV represents variable-length documents as fixed-length vectors. Paragraph vectors serve as "memory" of how words are used in various contexts, capturing differences in word usage across paragraphs.

Like word2vec, doc2vec has two model frameworks. The Paragraph Vector–Distributed Memory (PV-DM, illustrated in figure 7) framework uses word vectors and the randomly initialized paragraph vector to predict the next word in the sentence.

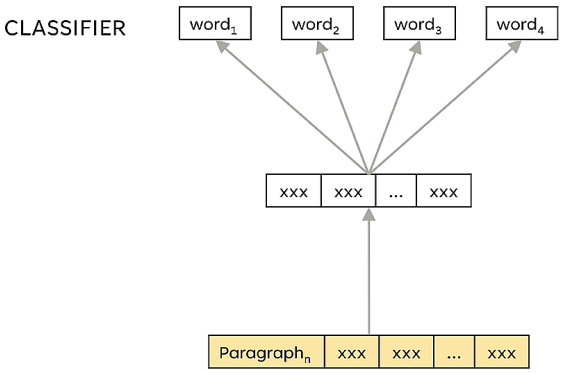
(As in word2vec, the process is iterated to minimize the error, and the resulting paragraph vector will be the final one.)

Figure 7. Paragraph Vector - Distributed Memory (PV-DM) model framework of Le and Mikolov (2014)



The Paragraph Vector–Distributed Bag of Words (PV-DBOW, illustrated in figure 8) architecture trains the paragraph vector to predict the words inside a context window.

Figure 8. Paragraph Vector - Distributed Bag of Words (PV-DBOW) model framework of Le and Mikolov (2014)



Wijffels (2021) authored an R implementation of Le and Mikolov's (2014) paragraph/document-to-vector algorithm in the form of the doc2vec R package. This project utilized this R implementation.

Latent Dirichlet Allocation

While extremely popular and effective, the doc2vec technique of Le and Mikolov (2014) is not the only method to represent text as numerical vectors. One of the stalwarts of text-mining techniques is latent Dirichlet allocation (LDA), described by Blei et al. (2003). LDA determines the topics of documents based on the words in

these documents. Documents can be a mixture of more than one topic, and topics are assumed to be a mixture of words. LDA is a generative model that attempts to find out how documents were generated. The best LDA model for a corpus is the one whose settings give the highest likelihood of generating documents in the corpus. The probability of generating a particular document is shown in figure 9.

Figure 9. Probability of generating a document, according to LDA (Blei et al., 2003)

$$P(\underline{w}, \underline{z}, \underline{\theta}, \underline{\psi} | \underline{\alpha}, \underline{\beta}) = P(\underline{\theta}, \underline{\alpha}) P(\underline{\psi}; \underline{\beta}) \prod_{t=1}^{N} P(z_t | \theta) P(w_t | \underline{\psi}_{z_t}, z_t)$$

Prior distributions. Generates probabilities of topics in a doc and words in a topic.

Prob. of selecting a topic

Prob. of selecting a word given selected topic.

where:

- w is the probability of selecting a word
- z is the probability of selecting a topic
- θ is the probability of a topic
- ψ is the probability of a word
- α is the Dirichlet parameter characterizing how single-topic or multi-topic the documents in the corpus are
- β is the Dirichlet parameter characterizing how single-word or multi-word the topics are
- N is the number of words in a document

This expression also reflects the process of how a document is generated. Given how single-topic or multi-topic the documents in the corpus are (i.e., given that the probability θ of a topic follows a Dirichlet distribution: $\theta \sim Dir(\alpha)$) and with the probability z_n of selecting any particular topic ($z_n \sim Mult(\theta)$), randomly select a topic.

Then, given how the topics comprise certain words (i.e., given that the probability Ψ of a word also follows a Dirichlet distribution: $\Psi \sim Dir(\beta)$) and with the probability of selecting any particular word in the previously selected topic ($\Psi z_{n,n} = p(w_n | z_n)$), randomly select a word from the selected topic.

Repeat this process until one has N words ($N \sim Poi(\xi)$) to build a document.

The output of LDA is a document-by-topic matrix containing the proportion of each document that is made up of words from each topic, as illustrated in table 1. These proportions are called topic gammas.

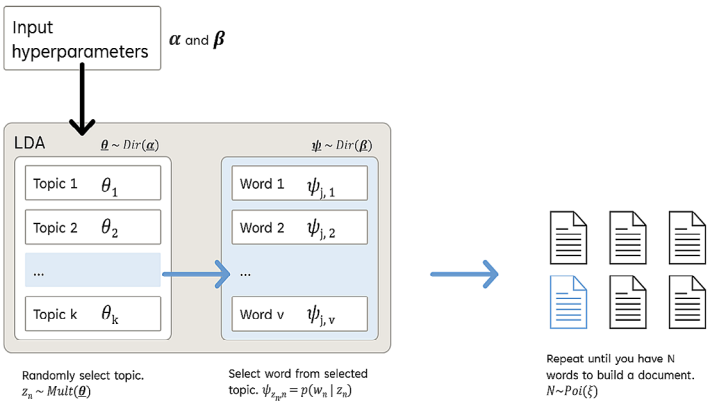
Table 1. Example 10 documents' topic assignments (based on their gammas) in a 25-topic LDA model on prelabeled ageism phrases.

Document ID	lda25_topic1	lda25_topic_k	lda25_topic25
jobstreet_2	0.04012185	...	0.039953
bbcn_11	0.04014714	...	0.04036888
centre_2021c_litrev_8	0.0399992	...	0.04009208

LDA is commonly used to interpret and determine the topics of documents. Typically, one looks at the words that a topic comprises and gives that topic a human-understandable label. However, rather than interpreting and determining documents' topics, Jones (2019) demonstrated that one may be more interested in LDA's output (i.e., topic gammas) as features for a downstream analytical task, such as classification. This project used LDA's output as input to an ageist-or-not classification task, rather than assigning human-understandable labels to the LDA-determined topics.

Vayansky and Kumar (2020) demonstrated LDA model performance peaking at 30 topics, prompting this study's LDA experiments to stay in the vicinity of 30 topics. This study will also emulate the LDA hyperparameter values of Griffiths and Steyvers (2004), which have been found to work well in several scenarios: topic-word distribution parameter value $\beta = 0.1$ and document-topic distribution parameter value $\alpha = 50/T$ where T is the number of topics.

Figure 10. LDA generative process: how a document is created, according to LDA (Blei et al., 2003)



Classification Techniques

This project uses two classical techniques to build ageist-or-not classification models. Logistic regression (Cox 1958) predicts the probability of a binary target (thus, is appropriate for determining ageist-or-not job ads) based on one or more predictors.

Penalized ridge regression (Hoerl and Kennard 1970), on the other hand, resolves multicollinearity and overfitting through a penalty term added to the Ordinary Least Squares (OLS) objective function.

This project's use of penalized ridge regression in a downstream classification task using doc2vec vectors as predictors emulates an illustrated example by Wijffels (n.d.).

Clustering Documents Through Top2Vec

The doc2vec R package by Wijffels (2021) includes an implementation of Top2Vec (Angelov 2020), which clusters documents by first reducing their dimensionality using Uniform Manifold Approximation and Projection (UMAP) (McInnes et al. 2020) and then applying Density-Based Spatial Clustering for Applications with Noise (DBSCAN) (Ester et al. 1996). Each cluster, except the noise cluster, represents a topic.

Angelov (2020) recommends the following UMAP hyperparameters:

- Number of nearest neighbors: Balances the preservation of global vs. local structure in the reduced-dimension embedding. The recommended value is 15.
- Distance metric: Measures distance between points in high-dimensional space. Cosine similarity is commonly used for document vectors.
- Embedding dimension: Refers to the number of dimensions in the low-dimensional embedding. The recommended value is 5, though further tuning may be needed.

For DBSCAN clustering, Schubert et al. (2017) highlight the ease of tuning the minPts hyperparameter, which is the number of neighbors within an Ɛ distance for a point to be considered a core point. They reiterate Sander's (1998) heuristic of twice the dimensionality, suggesting higher values for large and noisy datasets (e.g., minPts = 200 for a 100-vector doc2vec dataset). Indicators of poor DBSCAN parametrization include the largest cluster containing more than half the observations and more than 30 percent of observations in the noise cluster (Schubert et al. 2017).

This project considers all the above suggestions for Top2Vec, UMAP, and DBSCAN parametrization.

Comparison with Other Ageism Research

Table 2 compares the central topic, method of ageism determination, and modeling data of this study with those of Díaz et al. (2018), Díaz (2020), Burn et al. (2020), and Centre for Ageing Better (2021a, 2021b).

Table 2. Comparison of this study with other ageism research

	Díaz et al. (2018); Díaz (2020)	**Burn et al. (2020)**	**Centre for Ageing Better (2021a and 2021b)**	**This Project**
Topic	age bias in sentiment analysis tools	age bias in recruitment, predicted by ageist job ads	age bias in UK job ads	age bias in Philippine job ads
Ageism determination method	human (older adults') annotation re: ■ age-related (Y/N) ■ sentiment (-level scale from Very Negative to Very Positive)	cosine similarity to prelabeled ageist text in a word2vec vector space	human annotation (both younger and older adults') re: ■ effects on likelihood of applying, perception of getting interviewed, perception of getting hired, perception of fit (six-level scales) ■ appealing or off-putting	■ classification model using doc2vec and LDA as predictors ■ clustering on document vectors
Human annotation used?	Yes	No	Yes	No

	Díaz et al. (2018); Díaz (2020)	Burn et al. (2020)	Centre for Ageing Better (2021a and 2021b)	This Project
Embeddings and cosine similarity used?	No	Yes	Yes	Yes
LDA topics used?	No	No	No	Yes
Clustering used?	No	No	No	Yes
Classification used?	No	Yes	No	Yes
Modeling/ training data	human annotated tweets and blog posts containing "old" or "young"	English Wikipedia articles	■ human-annotated words, phrases, and job ads ■ UK job ads	■ Díaz (2020) ■ Burn et al. (2020) ■ Centre for Ageing Better (2021a and 2021b) ■ Philippine sources

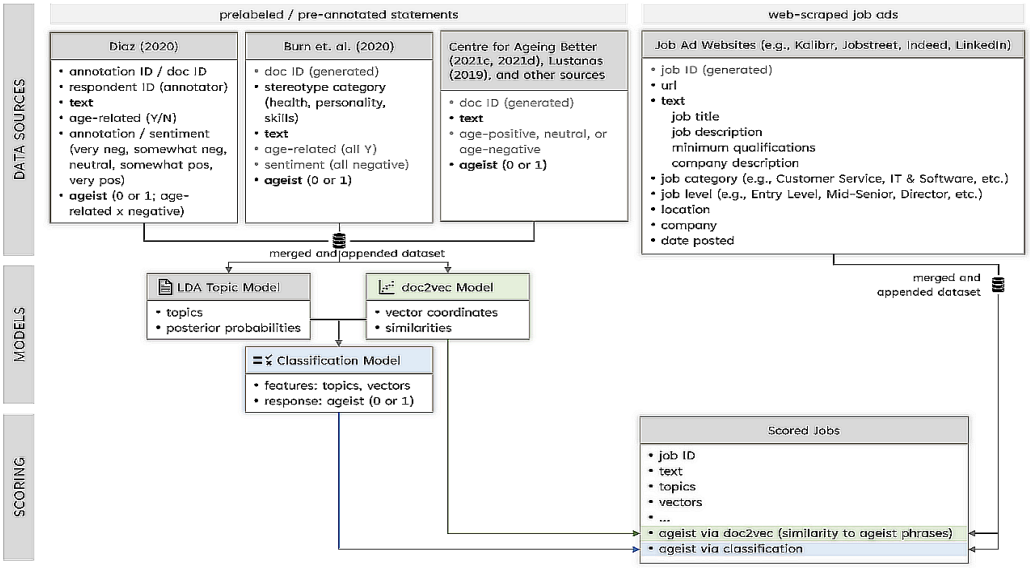
This study's contribution and distinctiveness are apparent in its: (1) focus on the Philippines; (2) alternative techniques different from—but built on—the other ageism research; and (3) leveraging of data and results from the other studies.

Methodology

Figure 11 illustrates the project's framework, comprising three parts: data sourcing, modeling, and scoring. In data sourcing, prelabeled text (documents)—that would be used to train and build an LDA topic model, a doc2vec (paragraph vector) model, and a classification model—were gathered. Additionally, text from online job advertisements between August 2022 and November 2023 was scraped from Kalibrr, Jobstreet, Indeed, and LinkedIn—four of the most popular job ad websites in the Philippines (Moneymax Editorial Team 2023; Similarweb 2023).

In the modeling phase, LDA topic models and doc2vec models were built. Their outputs (topic assignments or probabilities, and document vectors) were used as features in ageist-or-not classification models. These models were built, experimented on, and rated on accuracy and false-positive rate.

Figure 11. Overall framework for this study, comprising Data Sourcing, Modeling, and Scoring



In the scoring phase, job ads were scored on the classification model, LDA topic model, and doc2vec model to determine if they were ageist. Another doc2vec model was created on both prelabeled documents and job ads, clustering the entire set of documents. Job ads in clusters with a predominance of prelabeled ageist documents were shortlisted as ageist. Clustering models were evaluated based on DBSCAN parametrization guidelines of Schubert et al. (2017).

Detailed Procedure

A dataset of variable-length documents prelabeled with age-relatedness and sentiment was gathered. Age-related documents with negative sentiment were labeled "ageist" (Díaz 2020; Burn et al. 2020; Centre for Ageing Better 2021a, 2021b; CHRP 2019; Lustañas 2019; Valdez et al. 2013). Table 3 shows examples of pre-annotated texts.

Table 3. Examples of text prelabeled whether ageist or not

Source	Phrase	Ageist
Centre for Ageing Better (2021b)	knowledgeable	0
DOLE (2017)	age-based early retirement	1
Ramos (2022)	lower energy	1
Centre for Ageing Better (2021b)	ambitious	0
Centre for Ageing Better (2021b)	welcome applications from everyone	0

Topic models (e.g., LDA) were generated on these prelabeled documents, assigning topic probabilities and at least one topic. Doc2vec models created vector representations of each document. Table 4 shows sample documents and their first five vector coordinates in a 100-vector PV-DBOW doc2vec model.

Table 4. An example of documents and their first five vectors in a 100-vector PV-DBOW doc2vec model

Document ID	V_1	V_2	V_3	V_4	V_5
bbcn_28	0.01471602	0.12163511	-0.017993355	-0.14885184	0.021494599
chrp_2	0.02534879	-0.13588424	0.001587876	-0.15118743	-0.014123488
dole_7	0.04622494	-0.17081921	0.001124491	-0.16459785	-0.029896287
centre_2021c_wordprev_47	-0.02379648	-0.09205699	0.081412666	-0.17564282	0.049024981

These models determined which documents were most similar to specified words (word2doc) or phrases (sent2doc). Table 5 shows documents most similar to the word "young." Table 6 shows documents most like the phrase "fast paced environment."

Table 5. Prelabeled documents in the 100-dimension PV-DBOW doc2vec model that are most similar to the word "young", associating it with "handsome," "child," "hope," "cute," "bachelor"

Term	Prelabeled Document	Similarity	Rank
Young	hey youre handsome young man refresh see put positive way	0.5561293	1
Young	thank yes much count really fabo help boost quite young filmmakersso great alround	0.5461619	2
Young	child young know slightly much interest youre welcome	0.5397393	3
Young	surround significant people life especial family	0.5338619	4
Young	hope youngins enjoy wisdom	0.5291179	5
Young	think maybe young understand although kid understand good age	0.5291013	6
Young	okay hangover frackin awesome large life definitely younglings though	0.5256644	7
Young	aww cute youre cranky youre right though wouldnt kill 1in5 young people use product	0.5246596	8
Young	just see people hot bachelor 2009 jealous young single hollywood girl shoot em	0.5236568	9
Young	exact youre look	0.5213542	10

Table 6. The 10 documents used in the 100-dimension PV-DBOW doc2vec model that are most similar to the phrase "fast paced environment," associating it with "grow," "prejudice," and "enthusiastic"

Term	Prelabeled Document	Similarity	Ageist
fast paced environment	patience	0.8669859	0
fast paced environment	go old picsi grow fast	0.8622276	1
fast paced environment	prejudice	0.8536614	1
fast paced environment	mentor	0.852326	0
fast paced environment	none desire criteria	0.8518094	1
fast paced environment	reliable	0.8502892	0
fast paced environment	loyal	0.8492736	0
fast paced environment	careful	0.8483287	0
fast paced environment	enthusiastic	0.8458295	1
fast paced environment	dedicate	0.8441755	1

Ageist-or-not classification models were created using LDA topic gammas and vector representations as features. The models were trained and tested on a data frame similar to table 7.

Table 7. Sample dataframe for classification model with LDA topic gammas and document vectors as predictors

Document ID	Ageist	lda25_topic1	...	lda25_topic25	V1	...	V100
bbcn_116	1	0.04013757	...	0.03983601	0.002329581	...	-0.152559742
dole_5	1	0.04014391	...	0.04011021	0.071558222	...	-0.20970349
bbcn_40	0	0.03996642	...	0.03996259	-0.168001562	...	0.098400161
centre_2021c_wordprev_86	1	0.0400874	...	0.03984739	0.028569793	...	-0.164757639
valdez_45	1	0.03991346	...	0.0400346	-0.063052066	...	-0.009805218

Job ads were scraped from online job boards and professional networking sites, generating trigrams from various sections. Table 8 shows trigrams extracted from the Minimum Qualifications section of one job ad. These trigrams were scored with the LDA topic models, doc2vec models, and ageist-or-not classification models. Alternatively, another doc2vec model was created from both prelabeled documents and job ads, clustering them based on vector representations. Job ads in clusters with a high concentration of ageist documents were shortlisted as ageist.

Table 8. Trigrams extracted from the Minimum Qualifications section of one pre-scraped job ad

[1] "very good attention"	"good attention detail"
[3] "attention detail good"	"detail good project"
[5] "good project management"	"project management skills"
[7] "management skills self"	"skills self motivated"
[9] "self motivated team"	"motivated team player"
[11] "team player ability"	"player ability work"
[13] "ability work fast"	"work fast paced"
[15] "fast paced environment"	"paced environment good"
[17] "environment good organizational"	"good organizational time"
[19] "organizational time management"	"time management abilities"
[21] "management abilities third"	"abilities third fourth"
[23] "third fourth year"	"fourth year business"
[25] "year business management"	"business management majors"
[27] "management majors equivalent"	"majors equivalent available"
[29] "equivalent available commit"	"available commit least"
[31] "commit least 2"	"least 2 months"
[33] "2 months full"	"months full time"
[35] "full time internship"	"time internship us"
[37] "internship us fresh"	"us fresh graduates"

Prelabeled Ageism-Related Phrases

This study addresses the lack of a large, job-related, and Philippine-inclusive corpus of documents prelabeled as ageist by gathering 14,367 unique words, phrases, or sentences from Díaz (2020), Burn et al. (2020), and Centre for Ageing Better (2021c, 2021d), supplemented with Philippine sources. The dataset includes both ageist and non-ageist documents, with 36.6 percent labeled as ageist. Table 9 shows the number of documents contributed by each source and the proportion of ageist-labeled documents.

The mean word count of the prelabeled phrases varies, with recruitment-related or Philippine-specific datasets averaging between two and three words.

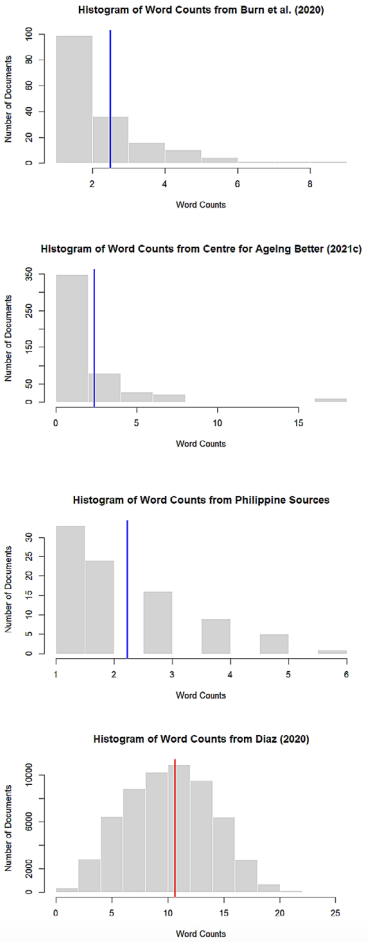
Table 9. No. of prelabeled ageism-related phrases gathered per source

Source	Documents	% Ageist
Diaz (2020)	14,002	35.8%
Burn et. al. (2020)	141	78.7%
Centre for Ageing Better (2021c)	128	64.8%
Valdez et. al. (2013)	50	36.0%
Ramos (2022)	12	50.0%
Burn et. al. (2020)...Centre for Ageing Better (2021c)	9	11.1%
DOLE (2017)	6	83.3%
CHRP (2019)	5	100.0%
Jobstreet Philippines Content Team (2022a and 2022b)	5	20.0%
Lustanas (2019)	5	100.0%
Centre for Ageing Better (2021c)...Valdez et. al. (2013)	2	0.0%
Centre for Ageing Better (2021c)...Ramos (2022)	1	100.0%
Diaz (2020)...Jobstreet Philippines Content Team (2022a and 2022b)	1	100.0%

- Burn et al. (2020): 2.5 words
- Centre for Ageing Better (2021b): 2.4 words
- various Philippine sources: 2.2 words
- Díaz (2020): 10.6 words

Figure 12 presents word counts of prelabeled documents from various sources.

Figure 12. Histograms of word counts of prelabeled documents from various sources



This study defines ageist documents as those labeled age-related or age-stereotypical and negative by older human annotators. Figure 13 displays the number of documents labeled per annotator cohort and sub-cohort.

Figure 13. Counts of documents labeled per cohort and subcohort

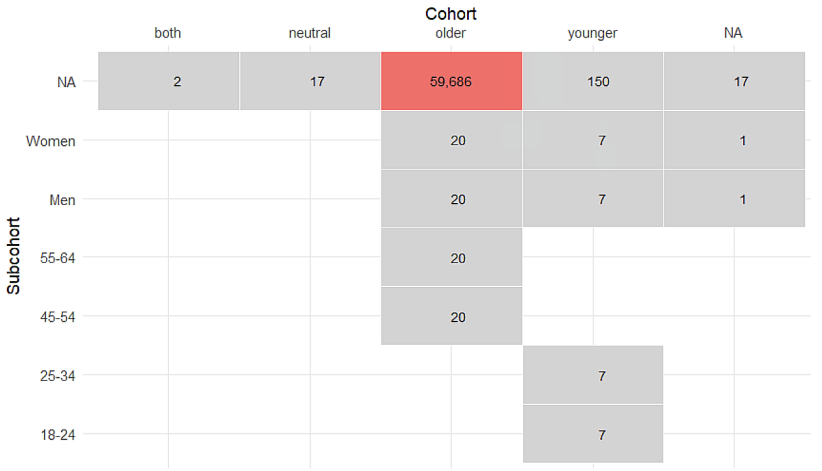


Table 10 indicates the counts and ageist proportions of documents labeled per category and subcategory.

Table 10. Counts and ageist proportions of documents labeled per category and subcategory

Category	Subcategory	Documents	% Ageist
Personality	Negative Personality	23	100.0%
Health	Less Physically Able	20	100.0%
Skills	Worse with Technology	11	100.0%
Health	Hard of Hearing	7	100.0%
Skills	Less Productive	6	100.0%
Health	Less Attractive	5	100.0%
Health	Worse Memory	4	100.0%
Personality	Less Creative	2	100.0%
Skills	Worse Communication Skills	17	94.1%
Personality	Less Adaptable	17	88.2%
Skills	Lower Ability to Learn	13	84.6%
employer perception of fit	NA	27	48.1%
getting interview	NA	27	48.1%
getting offer	NA	27	37.0%
applicant perception of fit	NA	27	29.6%
applying	NA	135	22.2%
NA	NA	59,571	13.4%
Personality	Dependable	17	0.0%
Personality	Careful	7	0.0%
Personality	Warm Personality	7	0.0%
Skills	Better Communication Skills	5	0.0%
Skills	More Experienced	5	0.0%
Skills	More Productive	2	0.0%

Table 11 shows the most frequently mentioned prelabeled phrases and the proportion of mentions where they were annotated as ageist.

Table 11. Most frequently mentioned/labeled phrases from sources

Processed Phrase	Frequency Mentioned/Labeled	% Labeled Ageist	Sources
old	35	11.4%	Diaz (2020)...Jobstreet Philippines Content Team (2022a and 2022b)
dynamic	26	23.1%	Centre for Ageing Better (2021c)
rain pour old man snore	25	4.0%	Diaz (2020)
ambitious	24	33.3%	Centre for Ageing Better (2021c)
dependable	24	8.3%	Centre for Ageing Better (2021c)
generous pension contribute	22	9.1%	Centre for Ageing Better (2021c)
innovate	22	27.3%	Centre for Ageing Better (2021c)
technological savvy	22	40.9%	Centre for Ageing Better (2021c)
active build diverse team welcome apply everyone	18	0.0%	Centre for Ageing Better (2021c)
energetic	17	58.8%	Centre for Ageing Better (2021c)
mcfly lose im fan now like music miss old music	17	17.6%	Diaz (2020)
enthusiastic	15	46.7%	Centre for Ageing Better (2021c)
experience	15	26.7%	Burn et. al. (2020)...Centre for Ageing Better (2021c)
dedicate	13	7.7%	Centre for Ageing Better (2021c)
fun	13	23.1%	Centre for Ageing Better (2021c)
get old	13	15.4%	Diaz (2020)
knowledgeable	13	7.7%	Centre for Ageing Better (2021c)
little worry one old close friend go hospital open heart surgery afternoon x	13	46.2%	Diaz (2020)
miss old time	13	38.5%	Diaz (2020)
patient	13	53.8%	Centre for Ageing Better (2021c)
adapt	12	58.3%	Centre for Ageing Better (2021c)
back use plain old digsby alpha flatlined	12	8.3%	Diaz (2020)
name	12	0.0%	Diaz (2020)
recent graduate	12	100.0%	Centre for Ageing Better (2021c)
thank bday wish feel love work morning feel old	12	41.7%	Diaz (2020)
active build diverse team welcome apply people background men women people age sexual orientations nationality religion beliefs	11	18.2%	Centre for Ageing Better (2021c)
flexible work opportunity	11	9.1%	Centre for Ageing Better (2021c)
twist sister wanna rock okay admit im date n old still wanna rock though	11	18.2%	Diaz (2020)

Table 12 zooms in on the most ageist job-related prelabeled phrases in the dataset.

Table 12. Mentions and mean ageist classifications of 'most ageist' job-related prelabeled phrases

Processed Phrase	Frequency Mentioned/Labeled	% Labeled Ageist	Sources
recent graduate	12	100.0%	Centre for Ageing Better (2021c)
worse memory	4	100.0%	Burn et. al. (2020)
techsavvy	3	100.0%	Centre for Ageing Better (2021c)...Ramos (2022)
3 5 year work experience	9	77.8%	Centre for Ageing Better (2021c)
energetic	17	58.8%	Centre for Ageing Better (2021c)
adapt	12	58.3%	Centre for Ageing Better (2021c)
patient	13	53.8%	Centre for Ageing Better (2021c)

The study highlights the need for an open-source dataset to advance age-discrimination research and provides a bootstrap method for building such a dataset.

Philippine Online Job Ads

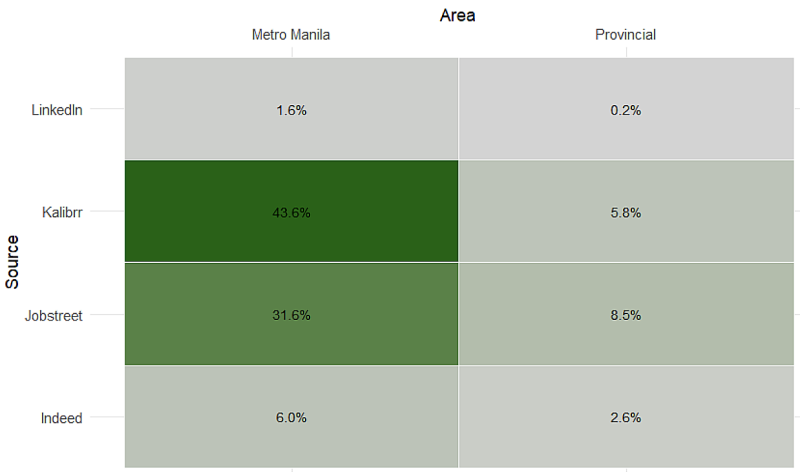
This project offers a dataset of 8,918 online job advertisements from four popular job boards in the Philippines, scraped using the rvest R package (Wickham 2022) and Octoparse between August 2022 and November 2023. The count of job ads, estimated dates of scraping, and scraping tool for each source job board are summarized in Table 13.

Table 13. Summary of job ad scraping activity per source job board

Job Ad Website	Date Scraped	No. of Job Ads Scraped	Scraping Tool Used
Jobstreet Philippines	2023-Nov-05	3,576	rvest
Kalibrr Philippines	2022-Aug-01	2,625	rvest
Kalibrr Philippines	2023-Sep-03	1,779	rvest
Indeed Philippines	2023-Nov-10	771	Octoparse
LinkedIn Philippines	2023-Nov-10	167	Octoparse

Despite nonstandard categorical descriptors, certain insights could be gleaned. For instance, three-quarters of job ads scraped were Metro Manila jobs advertised in either Jobstreet or Kalibrr. Figure 14 shows the proportion of job ads scraped per area and source job board.

Figure 14. Proportion of job ads scraped per area and source job board



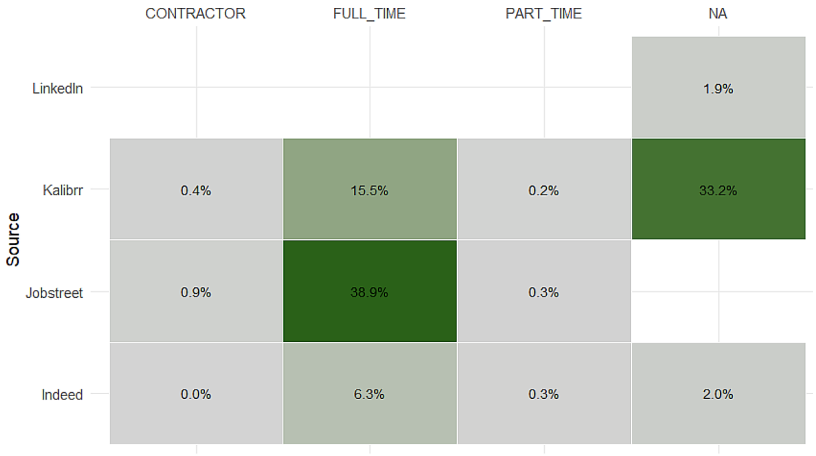
Job ads for Metro Manila and unspecified provinces (blank) make up 89.3 percent of our scraped job ads. Beyond these top two, the other top provinces of the scraped job ads are in table 14.

Table 14. No. of job ads from most-scraped provinces (sans Metro Manila and blank)

Province	Ads
Cebu	201
Davao	181
Batangas	112
Pampanga	68
Cavite	61
Bulacan	50
Laguna	48
Rizal	42

More than half of the job ads scraped are labeled full-time roles, while more than one-third are unlabeled in terms of employment type. Figure 15 shows the proportion of job ads scraped per employment type and source job board.

Figure 15. Proportion of job ads scraped per employment type and source job board



Where the educational requirement for a job could be extracted, this study found that a good portion of the scraped job ads required a "bachelor's degree graduate." Table 15 details the number of job ads scraped per educational requirement and source job board.

Table 15. No. of job ads scraped per educational requirement and source job board

Source	Educational Requirement	Ads
Kalibrr	Bachelor's degree graduate	4,167
Jobstreet	NA	3,576
Indeed	NA	771
LinkedIn	NA	167
Kalibrr	Graduated from high school	128
Kalibrr	Completed associate's degree	47
Kalibrr	Completed vocational course	39
Kalibrr	Master's degree graduate	18
Kalibrr	Doctoral degree graduate	5

After considerable cleansing, consolidation, and standardization of job categories, this project finds the top job categories of the scraped job ads to be: (1) IT, Software & Communication Technology; (2) Sales, Marketing & Communications; and (3) Call Centre & Customer Service. Table 16 shows the number and proportion of job ads scraped per job category.

Table 16. No. and proportion of job ads scraped per job category

Job Category	Ads	% of Dataset
IT, Software & Communication Technology	1,481	16.6%
Sales, Marketing & Communications	1,350	15.1%
Call Centre & Customer Service	1,318	14.8%
Accounting & Finance	787	8.8%
Administration, Coordination & Office Support	603	6.8%
Human Resources & Recruitment	472	5.3%
Design, Architecture & Engineering	458	5.1%
NA	353	4.0%
Manufacturing, Production, Transport & Logistics	315	3.5%
Healthcare & Medical	296	3.3%
Advertising, Arts, Media & Creatives	268	3.0%
Education & Training	188	2.1%
Retail & Consumer Products	151	1.7%
Hospitality & Tourism	100	1.1%
Supply Chain	99	1.1%
Skilled Trades & Services	89	1.0%
Banking & Financial Services	88	1.0%
Management, Consulting & Strategy	82	0.9%
Legal	78	0.9%
Writing & Content	77	0.9%
General Services	63	0.7%
Science & Technology	53	0.6%
Construction	36	0.4%
Farming, Animals & Conservation	28	0.3%
Safety & Security	25	0.3%
Real Estate & Property	24	0.3%
Public Service & NGOs	18	0.2%
CEO & General Management	11	0.1%
Arts & Sports	3	0.0%
Insurance & Superannuation	3	0.0%
Mining, Resources & Energy	1	0.0%

The mean word count of job ads' narrative fields is 307 words (215 words post-processing), further described for each source website in Table 17.

Table 17. Narrative free-text fields' mean word counts per job ad per source website

Source	Mean Word Count per Job Ad After Processing
Kalibrr	275
Indeed	205
Jobstreet	151
LinkedIn	57

Trigrams (three-word phrases) were extracted from the full narrative text of job ads, resulting in a total of 1,886,910 trigrams. Table 18 shows the total trigrams extracted per source website.

Table 18. Total trigrams extracted per source website

Source	Trigrams
Kalibrr	1,187,916
Jobstreet	533,350
Indeed	156,574
LinkedIn	9,070

Table 19. Most frequent trigrams from scraped job ads

Trigram	Frequency
bachelor degree graduate	4,214
bachelor s degree	1,285
hunter s hub	813
relevant work experience	741
experience require role	728
require role work	715
year work experience	706
right require role	698
role work right	698
work right require	698
year relevant work	682
summary role requirement	668
high school graduate	666
communication presentation skill	661

Doc2Vec Models

This project built six doc2vec models of varying dimensionality (100, 200, and 300 dimensions) and architectures (Distributed Bag of Words and Distributed Memory), in keeping with the findings of Le and Mikolov (2014) and Mikolov et al. (2013). To illustrate these doc2vec models, one can take some frequent job-ad trigrams and determine which ageist-prelabeled documents the trigrams are most similar to.

Notable results (in table 20) come out when one extracts the prelabeled ageist documents (labeled ageist more than 50 percent of the time in our dataset) that are most similar (i.e., the top 100 most similar prelabeled documents via doc2vec) to the frequent job-ad trigram "bachelor degree graduate."

Table 20. Prelabeled ageist documents most similar to "bachelor degree graduate" using the 100-dimension PV-DBOW model

Trigram	Similarity	Prelabeled	% Labeled Ageist
bachelor degree graduate	0.8880582	recent graduate	100.0%
	0.8353815	role suit recent graduate	100.0%
	0.7814786	18 25	100.0%
	0.7796524	must relevant post graduate qualify	100.0%
	0.7682048	low energy	100.0%
	0.7675754	low creativity	100.0%
	0.7624123	poor	100.0%
	0.7588589	patient	53.8%
	0.7587355	hungry	100.0%
	0.7586968	middle son 18th birthday today feel old	75.0%
	0.7544356	birthday today feel old	60.0%
	0.7536912	anxiety depression	100.0%
	0.7523112	high performance rate positive relate you	100.0%

It is not surprising to find that prelabeled phrases containing "graduate" are rated similarly to "bachelor degree graduate." Interestingly, though, prelabeled age-related phrases such as "18 25" and "18th birthday" are also rated similarly to "bachelor degree graduate," suggesting that the doc2vec model has rather learned the semantic relationship between that age range and the concept of graduation. On the flip side, the model also rated semantically unrelated phrases such as "patient" and "anxiety depression" as similar to "bachelor degree graduate."

Another interesting set of results ensues when one looks at the prelabeled ageist documents most similar to "relevant work experience," as illustrated in table 21 for a 100-dimension PV-DBOW model.

Table 21. Prelabeled ageist documents most similar to "relevant work experience" using a 100-dimension PV-DBOW model

Trigram	Similarity	Prelabeled	% Labeled Ageist
relevant work experience	0.7897334	3 5 year work experience	77.8%
	0.7747199	talk slow	100.0%
	0.7523395	talk past	100.0%
	0.7429368	fulfilment service	100.0%

Trigram	Similarity	Prelabeled	% Labeled Ageist
	0.7329472	dependent	100.0%
	0.7326579	prejudice	100.0%
	0.7297582	ideal first job	100.0%
	0.7245371	focus away future toward past	100.0%
	0.7244952	passionate	100.0%
	0.7239527	low creativity	100.0%
	0.7228774	volunteer good turn volunteer work everyone want young lady work find annoy	75.0%
	0.7224135	positivity	100.0%
	0.7198892	lose track talk	100.0%
	0.719634	ambition	100.0%
	0.7175448	demand	100.0%
	0.7158317	less energy	100.0%
	0.7156081	vulnerable sickness	100.0%
	0.7143649	low activity	100.0%
	0.7127285	bubbly	100.0%
	0.7127139	shaky hand	100.0%
	0.7105685	friendly	100.0%
	0.7096761	lone	100.0%
	0.7079928	flexible advantage	100.0%
	0.7055119	fragile	100.0%
	0.7055091	weaken cognitive function	100.0%
	0.7045839	forward think	100.0%
	0.7041513	creative	100.0%
	0.7040421	annoy	100.0%
	0.7030308	tech whizz	100.0%

Apart from showing that the doc2vec model is able to detect the obvious (e.g., a document containing "work experience" too is rated to "relevant work experience"), these results also provide guidance about phrases that should be avoided: known older stereotypes ("talk slow," "talk past," "dependent," "low creativity," "fragile") and younger stereotypes that may be off-putting to older applicants ("passionate," "volunteer," "positivity," "ambition," "bubbly," "flexible advantage," "forward think,"

"creative," and "tech whizz"). In short, "relevant work experience" could be just as bad as using these negative older stereotypes or off-putting younger stereotypes.

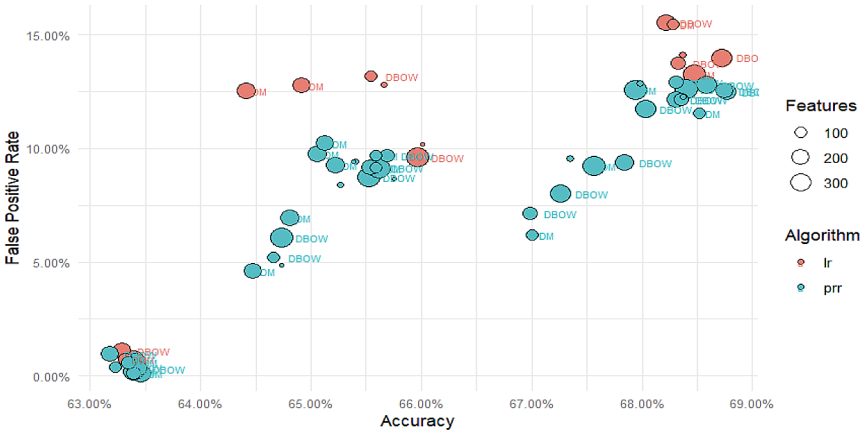
In summary, the doc2vec model experiments yielded the following findings:

- The lowest dimension (100) PV-DBOW model has associated age-indicative phrases with "bachelor degree graduate." Meanwhile, it takes more dimensions (300) for the PV-DM model to start doing that. PV-DBOW could possibly be a more ageism-sensitive model architecture than PV-DM (using our training data).
- An increase in dimensions in the PV-DBOW model from 100 to 200 has no noticeable semantic improvements. An increase from 100 to 300 dimensions for PV-DBOW, however, shows the start of learning conceptual relationships (e.g., "graduate"—"graduation").
- PV-DBOW similarity results for the frequent job-ad trigram "relevant work experience" could be viewed as guidelines for negative older stereotypes (e.g., "talk slow," "without ambition," "less energy") or off-putting younger stereotypes (e.g., "ideal first job," "passionate," "bubbly," "flexible," "forward think," "creative," "tech whizz") that should be avoided.
- PV-DM similarity results are much fewer than PV-DBOW results, again suggesting that PV-DM is more ageism-selective, while PV-DBOW is more ageism-sensitive.
- Two of the most frequent job-ad trigrams—"bachelor degree graduate" and "relevant work experience"—are reasonable job requirements, yet they seem to be viewed as ageist by sensitive PV-DBOW doc2vec models. This casts doubt on the viability of using embeddings' cosine similarity as a practical means in determining ageist job-ad phrases.

Determining Ageist Job Ads Through Classification Models

A good ageism classifier should accurately identify ageist phrases or documents and not be overly sensitive to non-ageist phrases. This project trained 60 classification models on 10,054 ageism-prelabeled documents and tested them on a holdout dataset of 4,310 prelabeled documents—a 70/30 train/test split. These models varied by classification algorithm (logistic regression or penalized ridge regression), LDA topic model predictors (25, 30, or 35 topics), and doc2vec dimensions (100, 200, or 300) and architecture (PV-DBOW or PV-DM). Their performance is shown in figure 16.

Figure 16. Accuracy and false-positive rate of classification models developed



The figure suggests that more accurate models tend to have more false positives. Four types of classifiers were created: preferable models (more accurate, not too falsely suspicious), collateral damage models (more accurate, more falsely suspicious), snowflake models (less accurate, more falsely suspicious), and careful models (less accurate, barely falsely suspicious). The champion models are the most careful preferable model (penalized ridge regression with 100-dimension PV-DM doc2vec, 67.0 percent accuracy, 6.2 percent false-positive rate) and the most accurate careful model (penalized ridge regression with 300-dimension PV-DM doc2vec, 63.5 percent accuracy, 0.1 percent false-positive rate), as shown in figures 17, 18, and 19.

Figure 17. Accuracy and false-positive rate of classification models developed, with champion models identified

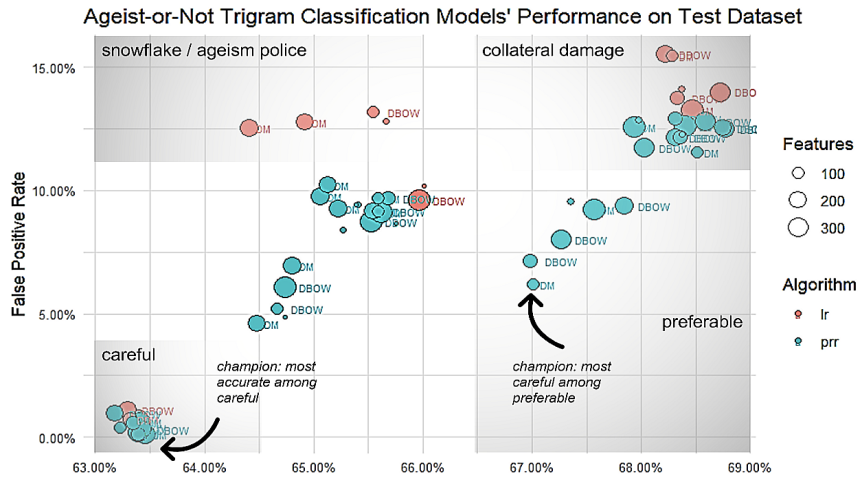


Figure 18. Zoom-in on the "preferable" classification models (higher accuracy but not too high false-positive rate)

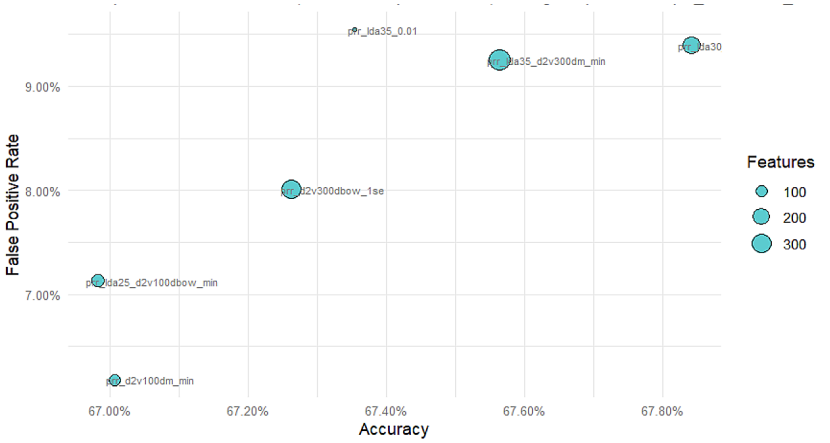
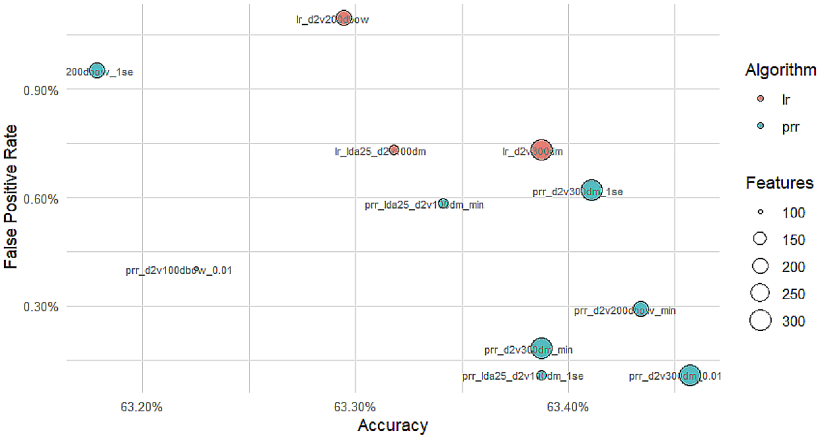


Figure 19. Zoom-in on the "careful" classification models (lower accuracy but also lower false-positive rate)



Both champion models are penalized ridge regression models that do not use LDA topic gammas as features and use only PV-DM doc2vec dimensions as predictors.

Most Careful Model Among the Preferable

Using this "most careful among preferable" model, 4,565 job ads scraped, or 51.2 percent of our 8,918 job ads, contained at least one ageist trigram. On average, 0.8% of the trigrams in a job ad are ageist.

Two measures of a job ad's ageism could be used. The first is binary: if a job ad contains just one ageist trigram, then the whole job ad could be considered ageist. The second is continuous: the number of ageist trigrams in a job ad as a percentage of the total number of trigrams in the job ad.

Using the "most careful among preferable" model, Kalibrr and Indeed contained the highest rate of ageist job ads and trigrams among the scraped samples, as shown in table 22.

Table 22. Summary of ageism predictions per job ad website, using "most careful among preferable" model

Source	Job Ads Scraped	Ads with Ageist Trigram	% Ads with Ageist Trigram	Trigrams Extracted	Ageist Trigrams	% Ageist Trigrams
Kalibrr	4,404	3,104	70.5%	1,187,916	10,085	0.8%
Indeed	771	515	66.8%	156,574	2,325	1.5%
Jobstreet	3,576	931	26.0%	533,350	2,911	0.5%
LinkedIn	167	15	9.0%	9,070	38	0.4%

Table 23 breaks these results down per job category, with LinkedIn Healthcare & Medical job ads; Indeed Sales, Marketing & Communications ads; and Kalibrr Hospitality & Tourism ads being most ageist. Meanwhile, there are no traces of ageism in Mining, Resources & Energy job ads from Jobstreet; as well as LinkedIn job ads for Accounting & Finance; Administration, Coordination & Office Support; Call Centre & Customer Service; IT, Software & Communication Technology; and Sales, Marketing & Communications.

Table 23. Percentage of ageist job ads per job category and source website

Source	Job Category	% Ads with Ageist Trigram	% Ageist Trigrams
LinkedIn	Healthcare & Medical	100.0%	2.4%
Indeed	Sales, Marketing & Communications	86.9%	2.3%
Kalibrr	Hospitality & Tourism	91.7%	2.2%
Indeed	Healthcare & Medical	97.4%	1.4%
Indeed	Call Centre & Customer Service	63.6%	1.2%
Kalibrr	General Services	65.1%	1.1%
Kalibrr	Science & Technology	80.0%	1.1%
Kalibrr	Call Centre & Customer Service	87.1%	1.1%
Kalibrr	Management, Consulting & Strategy	77.5%	1.1%
Kalibrr	Writing & Content	56.3%	1.0%
Kalibrr	Sales, Marketing & Communications	68.7%	1.0%
Kalibrr	Advertising, Arts, Media & Creatives	65.8%	1.0%
Jobstreet	Healthcare & Medical	22.7%	0.9%
Jobstreet	CEO & General Management	27.3%	0.9%
Indeed	IT, Software & Communication Technology	100.0%	0.9%
Kalibrr	Education & Training	85.5%	0.9%
Kalibrr	Human Resources & Recruitment	67.6%	0.8%
Kalibrr	IT, Software & Communication Technology	68.3%	0.8%
Indeed	Human Resources & Recruitment	50.0%	0.8%
Kalibrr	Accounting & Finance	70.1%	0.8%
Kalibrr	Healthcare & Medical	72.3%	0.8%
LinkedIn	Human Resources & Recruitment	27.3%	0.8%
Indeed	Accounting & Finance	50.0%	0.8%
Jobstreet	Management, Consulting & Strategy	54.5%	0.7%
Kalibrr	Administration, Coordination & Office Support	67.0%	0.7%
Kalibrr	Legal	84.8%	0.7%
Jobstreet	Sales, Marketing & Communications	28.1%	0.7%
Kalibrr	Manufacturing, Production, Transport & Logistics	60.2%	0.7%
Jobstreet	Accounting & Finance	27.2%	0.7%
Jobstreet	Advertising, Arts, Media & Creatives	32.3%	0.7%
Jobstreet	IT, Software & Communication Technology	42.9%	0.6%
Kalibrr	Skilled Trades & Services	75.0%	0.6%
Jobstreet	Human Resources & Recruitment	37.0%	0.6%
Jobstreet	Retail & Consumer Products	27.8%	0.6%
Jobstreet	Insurance & Superannuation	66.7%	0.6%
Indeed	NA	45.8%	0.6%

Jobstreet	Banking & Financial Services	42.0%	0.5%
Kalibrr	Public Service & NGOs	61.1%	0.5%
Jobstreet	Manufacturing, Production, Transport & Logistics	27.1%	0.5%
Kalibrr	Supply Chain	61.6%	0.5%
Kalibrr	Design, Architecture & Engineering	80.3%	0.5%
Jobstreet	Farming, Animals & Conservation	14.3%	0.5%
Jobstreet	Legal	18.8%	0.5%
Jobstreet	Design, Architecture & Engineering	18.4%	0.5%
Jobstreet	Call Centre & Customer Service	23.5%	0.5%
LinkedIn	NA	8.7%	0.4%
Indeed	Administration, Coordination & Office Support	40.6%	0.4%
Jobstreet	Skilled Trades & Services	13.0%	0.4%
Jobstreet	Science & Technology	32.6%	0.4%
Kalibrr	Safety & Security	80.0%	0.4%
Jobstreet	Hospitality & Tourism	22.7%	0.4%
Indeed	Writing & Content	33.3%	0.4%
Jobstreet	Administration, Coordination & Office Support	17.4%	0.3%
Jobstreet	Real Estate & Property	12.5%	0.3%
Kalibrr	Arts & Sports	66.7%	0.3%
Jobstreet	Education & Training	4.8%	0.2%
Jobstreet	Construction	8.3%	0.1%
Jobstreet	Mining, Resources & Energy	0.0%	0.0%
LinkedIn	Accounting & Finance	0.0%	0.0%
LinkedIn	Administration, Coordination & Office Support	0.0%	0.0%
LinkedIn	Call Centre & Customer Service	0.0%	0.0%
LinkedIn	IT, Software & Communication Technology	0.0%	0.0%
LinkedIn	Sales, Marketing & Communications	0.0%	0.0%

On average, a job ad has 212 trigrams. One may say that the most ageist job ads are those with the highest percentage of ageist trigrams, provided that they have at least 212 trigrams. (This prevents very short job ads from being the most ageist ad simply by virtue of, for instance, a single ageist trigram.) Tables 24 through 27 show the most ageist job ads (truncated for brevity) from each source website, using the "most careful among preferable" model. These tables contain the proportion of ageist trigrams in each ad, a truncated version of the ads themselves, and the three highest ageist-scored trigrams in each ad.

Table 24. Most ageist above average-length job ad from Kalibrr

Source	Ageist Trigrams	Total Trigrams	% Ageist Trigrams
Kalibrr	18	271	6.6%

Sales, Marketing & Communications
Sales Admin Associate | Delivery and Collection

Join Our Talent Community
There's a whole world of opportunities in front of you.

Our Purpose

We are Ardent World Inc., an FMCG, in the business of personal care and retail vanity products. AWI is the product of a passionate desire to bring quality, affordable skincare solutions to everyone. We combine passion with science and in-depth analytics to bring products to market that people can feel and see the difference.

A World of Opportunities

What's in it for you?

Play a key role in increasing income and revenue by managing and negotiating with clients, generating leads, qualifying prospects and managing product sales;

Champion customer service by building relationships in your assigned account;

Fill in the gap between the product and the market through insightful market and competitive research;

Develop and utilize your expertise as we gear towards digital transformation;

Maximize career growth opportunities within the organization.

AWI Professionals in our talent community ...

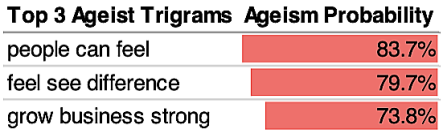


Table 25. Most ageist above average-length job ad from Jobstreet

Source	Ageist Trigrams	Total Trigrams	% Ageist Trigrams
Jobstreet	15	235	6.4%

Retail & Consumer Products
Key Accounts Manager - Modern Trade Accounts

Join Our Talent Community
There's a whole world of opportunities in front of you.

Our Purpose

We are Ardent World Inc an FMCG in the business of personal care and retail vanity products AWI is the product of a passionate desire to bring quality affordable skincare solutions to everyone We combine passion with science and in depth analytics to bring products to market so that people can feel and see the difference.

A World of Opportunities

AWI Professionals in our talent community share the same passion for excellence and thrive well with these qualities;

Goal driven mindset;

Firm understanding of business and economics to develop strategies and grow the business Strong creative and problem solving skills;

Strong management and leadership skills able to get the best of people and inspire them to follow;

Excellent communication skills can interact with different organizational levels Learning attitude to keep up with the digital transformation of the organization

What's in it for you

Play a key ...

Top 3 Ageist Trigrams	Ageism Probability
people can feel	83.7%
feel see difference	79.7%
grow business strong	73.8%

Table 26. Most ageist above average-length job ad from Indeed

Source	Ageist Trigrams	Total Trigrams	% Ageist Trigrams
Indeed	7	228	3.1%

Sales, Marketing & Communications
BPI AIA Sales Executive - Kamuning, Timog Avenue, Quezon Avenue, Roosevelt

BPI AIA Life Assurance Corporation (BPI AIA) is a strategic alliance between two leading financial companies in the Philippines: AIA Philippines Life and General Insurance Company, Inc. (AIA Philippines) and Bank of the Philippine Islands (BPI). As a combined brand, Filipinos can trust BPI AIA to help achieve their dreams through solutions that are accessible, affordable, and personalized.

We at BPI AIA understand that the Filipinos needs continue to evolve as they go through the different life stages and they want to be able to easily access solutions that help them live healthier, longer and better lives.

To help every Filipino family achieve financial security as we provide solutions to help customers live healthier, longer, and better lives.

Duties and Responsibilities:

Solicit new business, cross-sell and up-sell identified BPI AIA products to the bank s customer base

Establish, maintain, and grow relationships with the bank officers and staff in the branch/es they are assigned ...

Top 3 Ageist Trigrams	Ageism Probability
grow relationship bank	68.5%
grow relationship prospective	62.1%
maintain grow relationship	53.6%

Table 27. Most ageist above average-length job ad from LinkedIn

Source	Ageist Trigrams	Total Trigrams	% Ageist Trigrams
LinkedIn	5	356	1.4%

NA
Cabin Crew Records & Admin

Ensure adherence to quality, safety and service standards in accordance with the Quality and Safety Management System of Cebu Pacific and with the applicable regulatory, statutory and Company policies and procedures.. Participate actively in CEB SMS by ensuring that highest reasonable level of safety performance is achieved by complying with all applicable safety and security laws, regulations, and standards in cluding CEB policies and procedures and by reporting all accidents, incidents, and/or hazards.. Update APE result (Fit to Fly Certificate) in AIMS. Update and follow up Medical Certificate of Cabin Crew at the medical clinic.Ensure Medical Certificates are scanned and filed. Coordinate with HRD thru email regarding changes in attendance report on maternity and extended sick leave. Ensure Cabin Crew on Maternity Leave reports on scheduled date of check. Prepare, update, and encode the following: APE request letter to Health Service Provider and advise Cabin Crew with pending APE ...

Top 3 Ageist Trigrams	Ageism Probability
sick leave ensure	62.9%
body handle distribution	54.3%
regulatory body handle	54.3%

Most Accurate Model Among the Careful

Alternatively, using the "most accurate among careful" model, 4,949 job ads scraped, or 55.5 percent of our 8,918 job ads, contained at least one ageist trigram. On average, 1.1 percent of the trigrams in a job ad are ageist.

Once again, Kalibrr and Indeed contained the highest rate of ageist job ads and trigrams among the scraped samples, as shown in table 28.

Table 28. Summary of ageism predictions per job ad website, using "most accurate among careful" model

Source	Job Ads Scraped	Ads with Ageist Trigram	% Ads with Ageist Trigram	Trigrams Extracted	Ageist Trigrams	% Ageist Trigrams
Kalibrr	4,404	3,339	75.8%	1,187,916	13,845	1.2%
Indeed	771	531	68.9%	156,574	3,448	2.2%
Jobstreet	3,576	1059	29.6%	533,350	3,958	0.7%
LinkedIn	167	20	12.0%	9,070	52	0.6%

Indeed Sales, Marketing & Communications job ads are the most ageist in this "most accurate among careful" model, as shown in Table 29. Once again, there are no traces of ageism in Mining, Resources & Energy job ads from Jobstreet; as well as LinkedIn job ads for Accounting & Finance; Administration, Coordination &

Office Support; IT, Software & Communication Technology; and Sales, Marketing & Communications.

Table 29. Percentage of ageist job ads per job category and source website

Source	Job Category	% Ads with Ageist Trigram	% Ageist Trigrams
Indeed	Sales, Marketing & Communications	94.6%	3.5%
Indeed	IT, Software & Communication Technology	100.0%	2.2%
Kalibrr	Hospitality & Tourism	75.0%	2.2%
Indeed	Healthcare & Medical	97.4%	1.9%
Kalibrr	Call Centre & Customer Service	89.7%	1.8%
Indeed	Call Centre & Customer Service	72.7%	1.7%
Kalibrr	General Services	85.7%	1.6%
Kalibrr	Science & Technology	80.0%	1.5%
Kalibrr	Management, Consulting & Strategy	81.7%	1.5%
Kalibrr	Sales, Marketing & Communications	76.0%	1.5%
Kalibrr	Arts & Sports	100.0%	1.4%
Kalibrr	Healthcare & Medical	84.2%	1.3%
LinkedIn	Healthcare & Medical	33.3%	1.2%
Kalibrr	Writing & Content	69.0%	1.2%
Kalibrr	Education & Training	83.1%	1.2%
Jobstreet	CEO & General Management	36.4%	1.2%
Kalibrr	Human Resources & Recruitment	73.4%	1.1%
LinkedIn	Human Resources & Recruitment	36.4%	1.1%
Jobstreet	Management, Consulting & Strategy	72.7%	1.1%
Kalibrr	Accounting & Finance	72.6%	1.1%
Kalibrr	Advertising, Arts, Media & Creatives	69.6%	1.0%
Indeed	Accounting & Finance	46.2%	1.0%
Kalibrr	Manufacturing, Production, Transport & Logistics	75.9%	1.0%
Jobstreet	Sales, Marketing & Communications	37.0%	1.0%
Kalibrr	Administration, Coordination & Office Support	76.1%	1.0%
Kalibrr	IT, Software & Communication Technology	69.9%	1.0%
Jobstreet	Banking & Financial Services	44.3%	1.0%
Jobstreet	Retail & Consumer Products	39.7%	0.9%
Indeed	Administration, Coordination & Office Support	43.8%	0.9%
Jobstreet	Accounting & Finance	29.9%	0.9%
Jobstreet	Human Resources & Recruitment	38.5%	0.9%
Kalibrr	Legal	87.0%	0.9%
Kalibrr	Skilled Trades & Services	85.0%	0.8%
Jobstreet	Advertising, Arts, Media & Creatives	25.8%	0.8%
Indeed	NA	39.8%	0.8%
Indeed	Human Resources & Recruitment	40.0%	0.8%
Kalibrr	Design, Architecture & Engineering	86.5%	0.8%
Jobstreet	Skilled Trades & Services	14.5%	0.8%
Kalibrr	Public Service & NGOs	66.7%	0.7%
Jobstreet	Call Centre & Customer Service	32.6%	0.7%
Jobstreet	Legal	21.9%	0.7%
LinkedIn	NA	12.5%	0.7%
Jobstreet	Real Estate & Property	16.7%	0.6%
Jobstreet	Design, Architecture & Engineering	21.2%	0.6%
Jobstreet	Construction	27.8%	0.6%
Jobstreet	Manufacturing, Production, Transport & Logistics	24.6%	0.6%
Jobstreet	Healthcare & Medical	14.3%	0.6%
Jobstreet	Insurance & Superannuation	66.7%	0.6%
Kalibrr	Safety & Security	84.0%	0.5%
Jobstreet	Farming, Animals & Conservation	10.7%	0.4%
Jobstreet	Science & Technology	37.2%	0.4%
Jobstreet	Hospitality & Tourism	25.0%	0.4%
Jobstreet	Administration, Coordination & Office Support	17.4%	0.4%
LinkedIn	Call Centre & Customer Service	11.8%	0.3%
Indeed	Writing & Content	16.7%	0.3%
Jobstreet	Education & Training	6.7%	0.3%
Jobstreet	Mining, Resources & Energy	0.0%	0.0%
LinkedIn	Accounting & Finance	0.0%	0.0%
LinkedIn	Administration, Coordination & Office Support	0.0%	0.0%
LinkedIn	IT, Software & Communication Technology	0.0%	0.0%
LinkedIn	Sales, Marketing & Communications	0.0%	0.0%

Tables 30 through 33 show the most ageist, above average-length job ads from each source website, using the "most accurate among careful" model. Except for the most ageist job ad from LinkedIn and higher ageism scores for trigrams, these results are the same as the "most careful among preferable" model. Once again, these tables contain the proportion of ageist trigrams in each ad, a truncated version of the ads themselves, and the three highest ageist-scored trigrams in each ad.

Table 30. Most ageist above average-length job ad from Kalibrr

Source	Ageist Trigrams	Total Trigrams	% Ageist Trigrams
Kalibrr	18	271	6.6%

Sales, Marketing & Communications
Sales Admin Associate | Delivery and Collection

Join Our Talent Community
There's a whole world of opportunities in front of you.

Our Purpose

We are Ardent World Inc., an FMCG, in the business of personal care and retail vanity products. AWI is the product of a passionate desire to bring quality, affordable skincare solutions to everyone. We combine passion with science and in-depth analytics to bring products to market that people can feel and see the difference.

A World of Opportunities

What's in it for you?

Play a key role in increasing income and revenue by managing and negotiating with clients, generating leads, qualifying prospects and managing product sales;

Champion customer service by building relationships in your assigned account;

Fill in the gap between the product and the market through insightful market and competitive research;

Develop and utilize your expertise as we gear towards digital transformation;

Maximize career growth opportunities within the organization.

AWI Professionals in our talent community ...

Top 3 Ageist Trigrams	Ageism Probability
people can feel	91.9%
feel see difference	86.2%
grow business strong	78.8%

Table 31. Most ageist above average-length job ad from Jobstreet

Source	Ageist Trigrams	Total Trigrams	% Ageist Trigrams
Jobstreet	14	235	6.0%

Retail & Consumer Products
Key Accounts Manager - Modern Trade Accounts

Join Our Talent Community
There's a whole world of opportunities in front of you.

Our Purpose

We are Ardent World Inc an FMCG in the business of personal care and retail vanity products AWI is the product of a passionate desire to bring quality affordable skincare solutions to everyone We combine passion with science and in depth analytics to bring products to market so that people can feel and see the difference.

A World of Opportunities

AWI Professionals in our talent community share the same passion for excellence and thrive well with these qualities;

Goal driven mindset;

Firm understanding of business and economics to develop strategies and grow the business Strong creative and problem solving skills;

Strong management and leadership skills able to get the best of people and inspire them to follow;

Excellent communication skills can interact with different organizational levels Learning attitude to keep up with the digital transformation of the organization

What's in it for you
Play a key ...

Top 3 Ageist Trigrams	Ageism Probability
people can feel	91.9%
feel see difference	86.2%
grow business strong	78.8%

Table 32. Most ageist above average-length job ad from Indeed

Source	Ageist Trigrams	Total Trigrams	% Ageist Trigrams
Indeed	11	243	4.5%

Sales, Marketing & Communications
BPI AIA Sales Executive - Kamuning, Timog Avenue, Quezon Avenue, Roosevelt

BPI AIA Life Assurance Corporation (BPI AIA) is a strategic alliance between two leading financial companies in the Philippines: AIA Philippines Life and General Insurance Company, Inc. (AIA Philippines) and Bank of the Philippine Islands (BPI). As a combined brand, Filipinos can trust BPI AIA to help achieve their dreams through solutions that are accessible, affordable, and personalized.

We at BPI AIA understand that the Filipinos needs continue to evolve as they go through the different life stages and they want to be able to easily access solutions that help them live healthier, longer and better lives.

To help every Filipino family achieve financial security as we provide solutions to help customers live healthier, longer, and better lives.

Duties and Responsibilities:

Solicit new business, cross-sell and up-sell identified BPI AIA products to the bank s customer base

Establish, maintain, and grow relationships with the bank officers and staff in the branch/es they are assigned ...

Top 3 Ageist Trigrams	Ageism Probability
grow relationship bank	75.9%
grow relationship prospective	65.2%
life stage want	61.8%

Table 33. Most ageist above average-length job ad from LinkedIn

Source	Ageist Trigrams	Total Trigrams	% Ageist Trigrams
LinkedIn	4	237	1.7%

NA
Operations Director

Lives by and promotes the ClearSource Core Values (Customer First, Personal Accountability, Humble Courage, Hungry, Happy and Healthy). Attend weekly one on one with their leader and implement decisions and execute action items that result. Hold weekly one-on-one meetings with direct report leaders to:. Build relationships of mutual trust. Create accountability around their dreams and career goals. Teach and assess their implementation of the ClearSource leadership training tools (the Role Definition, One-on-One, Coaching Process, People Analyzer, and Performance Self-Assessment). Confirm attainment of the KPIs outlined in their respective Role Definitions. Follow up on commitments made. Update your People Analyzer and, monthly, administer the Performance Self-Assessment to direct reports. Daily, review performance metrics by program and leader in the following areas:. Client satisfaction (contractual obligations and quality of service). Financial health (revenue and profitability). Em ...

Top 3 Ageist Trigrams	Ageism Probability
make update people	64.3%
coach process people	61.7%
people analyzer performance	58.9%

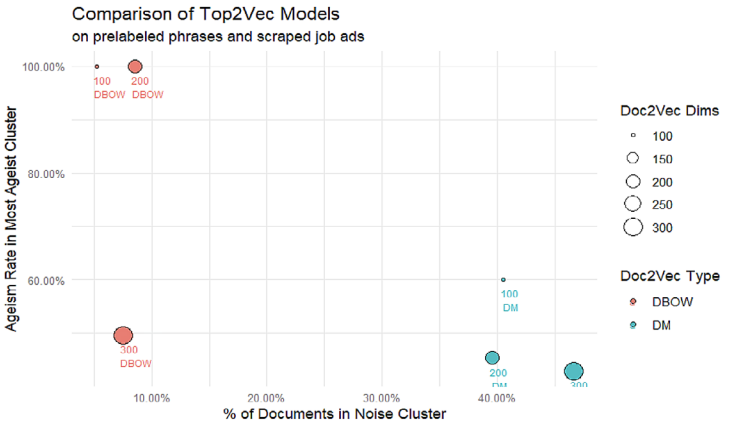
Remarkably, the themes around "people," "growth," "feeling," and "relationships" in these ageism-classification results coincide with the "warm but incompetent" age construct (Ayalon and Tesch-Römer, 2018). Meanwhile, the ageist classification of the job ad containing "sick leave" and "body" coincides with stereotype of the old being sickly. All these results illustrate that doc2vec models of various parameters could be used as predictors for explainable ageism-detection (or ageism-shortlisting) models with a certain degree of success.

Clustering on Document Vectors (Embeddings)

Using Top2Vec (Angelov 2020) and DBSCAN (Ester et al. 1996), this project created six clustering models on 23,277 documents, combining 14,367 ageism-prelabeled documents and 8,910 job ads. The models varied by doc2vec dimensions (100, 200, or 300) and architecture (PV-DBOW or PV-DM), following Angelov (2020) and Sander (1998) recommendations. The document clusters are described in the Appendix.

Applying criteria for a good clustering model (minimal noise, maximal ageist documents in any cluster), the most appropriate model is shown in figure 20.

Figure 20. Comparison of Top2Vec document clustering models



Two PV-DBOW models met the criteria, but their 100 percent ageist clusters contained only one prelabeled ageist document. In light of this, the next-best model, the 100-dimension PV-DM model, was chosen as the champion, despite having more documents in the noise cluster (40.5 percent) and a 60.0 percent ageism rate in the most ageist cluster, as seen in figure 21 and table 34.

Figure 21. Comparison of Top2Vec document clustering models with champion model identified

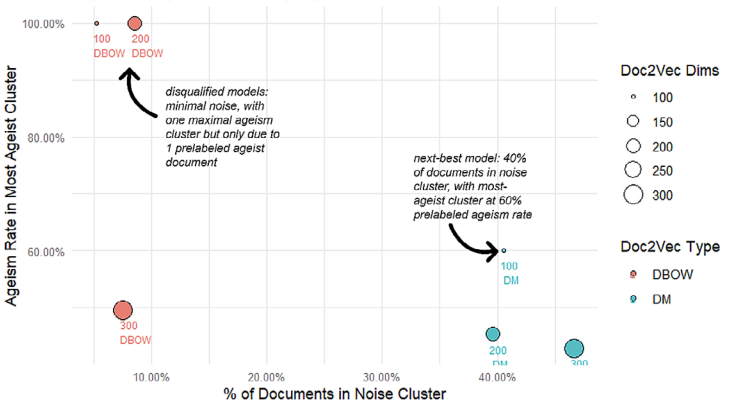


Table 34. Clusters of selected champion model

Cluster	Docs in Cluster	% of Docs	Job Ads	Prelabeled Docs	Ageist Docs	% Ageist
0	9,436	40.5%	3,927	5,509	1,943	35.3%
1	356	1.5%	20	336	142	42.3%
2	1,060	4.6%	1,050	10	6	60.0%
3	938	4.0%	916	22	9	40.9%
4	471	2.0%	465	6	2	33.3%
5	268	1.2%	266	2	1	50.0%
6	515	2.2%	503	12	6	50.0%
7	2,002	8.6%	1,481	521	217	41.7%
8	8,231	35.4%	285	7,946	2,926	36.8%

The most-ageist cluster had 6 ageist documents out of 10, illustrating the stereotype of older people and technology, as seen in table 35.

Table 35. Prelabeled ageist phrases in most-ageist cluster of champion model show ageist stereotype of older people struggling with technology

Prelabeled Ageist Phrase
lack capacity deal new technology
please note apply can consider cv support state submitted
problem technology
start repel technology mean youre get old
wireless 3g blackberry lounge just coffeebean n good old gsm web20 shoestring wwwpingfm
give back cant around old one yet hasnt feline leukemia test

There are 1,050 job ads in this cluster, suggesting a 60.0 percent chance of being ageist. Kalibrr had the highest potential ageism rate among the sources, as seen in table 36.

Table 36. Potential ageism rate per job-ad source, based on champion Top2Vec model

Source	Job Ads Scraped	Potentially Ageist Ads	Potential Ageism Rate
Kalibrr	4,396	785	17.9%
Jobstreet	3,576	238	6.7%
LinkedIn	167	10	6.0%
Indeed	771	17	2.2%

Table 37 shows that IT, Software & Communication Technology job ads from Kalibrr, LinkedIn, and Jobstreet had the highest potential ageism rate, consistent with the technology-themed ageist phrases in their cluster.

Table 37. Potential ageism rate per job-ad source and job category, based on champion Top2Vec model

Source	Job_Category	Job Ads Scraped	Potentially Ageist Ads	Potential Ageism Rate
Kalibrr	IT, Software & Communication Technology	1,118	741	66.3%
LinkedIn	IT, Software & Communication Technology	2	1	50.0%
Jobstreet	IT, Software & Communication Technology	359	179	49.9%
Jobstreet	Management, Consulting & Strategy	11	3	27.3%
Jobstreet	Science & Technology	43	11	25.6%
LinkedIn	Call Centre & Customer Service	17	2	11.8%
Kalibrr	Skilled Trades & Services	20	2	10.0%
Kalibrr	Management, Consulting & Strategy	71	6	8.5%
Kalibrr	Safety & Security	25	2	8.0%
LinkedIn	NA	104	7	6.7%
Indeed	NA	249	16	6.4%
Jobstreet	Administration, Coordination & Office Support	253	13	5.1%
Kalibrr	Design, Architecture & Engineering	207	9	4.3%
Kalibrr	Call Centre & Customer Service	307	10	3.3%
Jobstreet	Education & Training	105	3	2.9%
Jobstreet	Design, Architecture & Engineering	250	7	2.8%
Kalibrr	Legal	46	1	2.2%
Jobstreet	Call Centre & Customer Service	881	13	1.5%
Jobstreet	Manufacturing, Production, Transport & Logistics	207	3	1.4%
Jobstreet	Healthcare & Medical	154	2	1.3%
Jobstreet	Human Resources & Recruitment	192	2	1.0%
Kalibrr	Healthcare & Medical	101	1	1.0%
Kalibrr	Administration, Coordination & Office Support	306	3	1.0%
Kalibrr	Manufacturing, Production, Transport & Logistics	108	1	0.9%
Indeed	Call Centre & Customer Service	110	1	0.9%
Kalibrr	Sales, Marketing & Communications	805	7	0.9%
Jobstreet	Sales, Marketing & Communications	235	2	0.9%
Kalibrr	Accounting & Finance	430	2	0.5%

Table 38 contains sample tech-themed job ads from Kalibrr, LinkedIn, and Jobstreet that are potentially ageist for being clustered with known ageist phrases.

Table 38. Sample tech-themed ads clustered alongside known ageist phrases

Source	Full Job Text
Jobstreet	Engineer Technical Writer Job Overview Create and sustain service documentation for the company s products based on information supplied by the Product Specialist Create manuals drawings technical documents service software and other service files and publish them in the internal knowledge database Requirements 1 3 years of experience Knowledgeable in Uninterruptible Power Supply and Power Products Precision Cooling Systems preferred Knowledge of Data Centers preferred Working knowledge of web technologies and documentation technologies preferred Working knowledge of version control technologies platforms plus Experienced assembling source content into structured documentation plus Familiarity with Structured Query Language SQL plus Willing to work on site in Mandaluyong We request your consent to collect process and share your personal data with our hiring clients for application purposes in accordance with the Data Privacy Act of 2012 of the Philippines Your information will be handled confidentially and used solely for the recruitment process By proceeding with this application you acknowledge and agree to our data privacy policy Summary of role requirements Looking for candidates available to work Monday Morning Tuesday Morning Wednesday Morning Thursday Morning Friday Morning2 3 years of relevant work experience required for this role Working rights required for this role Expected start date for role 15 November 2023...

Source	Full Job Text
Kalibrr	Solution Development and Delivery - Write well designed, testable, efficient code by using best software development practices - Create website layout/user interface by using standard HTML/CSS practices - Integrate data from various back-end services and databases - Gather and refine specifications and requirements based on technical needs - Create and maintain software documentation - Be responsible for maintaining, expanding, and scaling our site - Stay plugged into emerging technologies/industry trends and apply them into operations and activities - Cooperate with web designers to match visual design intent - Understands standard systems development lifecycle processes and applies our methodology effectively on client engagements. Project Execution - Understands project and development plans and is able to clearly articulate roles, project goals, and timelines - Adheres to coding standards defined by technical management. - Accurately employs our development tools. - Establishes responsible deadlines and personal work plans. *We are currently on a temporary work-from-home setup and will start a hybrid setup. The Position We are looking for a full-time Software Engineer who looks forward to taking on challenging tasks over repetitive ones; and who can participate in developing our services and products. The chosen candidate will be responsible for defining, implementing, testing, troubleshooting new features and functionality, and fixing any issues. Requirements At least two years of work experience in web development At least one year of experience working with any of the following major MVC frameworks: - Ruby on Rails (preferred) - Laravel PHP - Django Preferably with at least one year of experience working with any of the following major frontend frameworks: - React (preferred) - Vue - Angular Knowledgeable in responsive web design and version control (e.g., Git) Knowledgeable in testing and DevOps A solid understanding of how web applications work, including security, session management, and best development practices Adequate knowledge of relational database systems, Object-Oriented Programming, and web application development Aggressive problem diagnosis and creative problem-solving skills Hands-on experience with network diagnostics and network analytics tools is an advantage. CREATIVEHOPE PH is a subsidiary company of a Technology and Web Consultancy Company in Japan called CREATIVEHOPE JP. [www.creativehope.co.jp] We offer services in Fintech, Datahub, and Marketing & Sales among others. CREATIVEHOPE PH is a growing Technology Consultancy company that aims to provide the best products and services for its clients using the latest technologies. Bachelor's degree graduate...

Interestingly, all these samples contain one or more of the following characteristics: (1) technical jargon expected of a technology job; (2) reference to a certain number of years of experience (usually a small number); and (3) mention of "graduate" (which is in doc2vec proximity to "recent graduate" and which—from Centre for Ageing Better (2021b)—is off-putting to older applicants).

Similar to the ageism-classification models, these Top2Vec results illustrate that the clustering method could also be a viable and explainable technique to shortlist potentially ageist job ads.

Tables 39 to 48 show the five most ageist and five least ageist trigrams in the five most frequent job categories, according to the two champion classification models. The most ageist trigrams all contain the word "feel," consistent with the warm-but-incompetent older-age construct. On the other hand, the least ageist trigrams all mention positive words such as "love," "good", or "great," suggesting that positive phrasing is one way to avoid implied ageism. Once again, some of the ageist-scored job-ad trigrams could come from the hiring company description more than the job description.

Table 39. Five most ageist and five least ageist trigrams in IT, Software & Communication Technology, using the "most careful among the preferable" model

Trigram	Ageism Probability	Job titles using this trigram	Hiring companies using this trigram
finger handle feel	0.842395804	1	1
feel reach hand	0.838780116	1	1
people can feel	0.837158133	1	1
people think feel	0.837158133	1	1
feel appreciate value	0.836222561	1	1
love build cool	0.04782069	1	1
good talent can	0.046287173	18	2
business owner love	0.044951843	1	1
love job opportunity	0.037379384	1	1
love technology benefit	0.034567091	1	1

Table 40. Five most ageist and five least ageist trigrams in Sales, Marketing & Communications, using the "most careful among the preferable" model

Trigram	Ageism Probability	Job titles using this trigram	Hiring companies using this trigram
people can feel	0.837158133	1	1
feel pain sale	0.835303822	3	1
believe take feel	0.834828642	4	1
help people feel	0.830459321	2	1
feel home may	0.824564781	1	1
love business opportunity	0.037513925	1	1
can create great	0.034524189	1	1
love company will	0.034505298	1	1
love learn new	0.03445106	1	1
love great copy	0.033512752	1	1

Table 41. Five most ageist and five least ageist trigrams in Call Centre & Customer Service, using the "most careful among the preferable" model

Trigram	Ageism Probability	Job titles using this trigram	Hiring companies using this trigram
feel customer care	0.878686396	1	1
feel value care	0.876930073	1	1
help client feel	0.844520428	1	1
finger handle feel	0.842395804	1	1
culture everyone feel	0.839356267	15	1
love ability tap	0.045912715	1	1
program expect love	0.044961934	1	1
love join team	0.03687548	1	1
love learn new	0.03445106	1	1
love enjoy flexible	0.026876289	1	1

Table 42. Five most ageist and five least ageist trigrams in Accounting & Finance, using the "most careful among the preferable" model

Trigram	Ageism Probability	Job titles using this trigram	Hiring companies using this trigram
believe take feel	0.834828642	4	1
feel value matter	0.834293658	3	1
feel welcome value	0.828837558	2	2
feel include value	0.826903475	3	1
feel home may	0.824564781	2	1
create great financial	0.053899485	1	1
decision wed love	0.053791644	1	1
good talent can	0.046287173	4	1
financial service love	0.042348217	1	1
love meet skill	0.028575774	1	1

Table 43. Five most ageist and five least ageist trigrams in Administration, Coordination & Office Support, using the "most careful among the preferable" model

Trigram	Ageism Probability	Job titles using this trigram	Hiring companies using this trigram
believe take feel	0.834828642	1	1
feel home may	0.824564781	5	1
goal everyday feel	0.822374341	1	1
feel opportunity present	0.821794896	1	1
feel free browse	0.799650907	1	1
create great financial	0.053899485	1	1
decision wed love	0.053791644	1	1
good talent can	0.046287173	4	1
financial service love	0.042348217	1	1
love meet skill	0.028575774	1	1

Table 44. Five most ageist and five least ageist trigrams in IT, Software & Communication Technology, using the "most accurate among the careful" model

Trigram	Ageism Probability	Job titles using this trigram	Hiring companies using this trigram
people think feel	0.919016808	1	1
people can feel	0.919016808	1	1
feel reach hand	0.90619803	1	1
finger handle feel	0.902534178	1	1
feel appreciate value	0.899095672	1	1
create great positive	0.035339899	1	1
business owner love	0.034793187	1	1
role create good	0.033655742	1	1
love job opportunity	0.026044786	1	1
love technology benefit	0.025889905	1	1

Table 45. Five most ageist and five least ageist trigrams in Sales, Marketing & Communications, using the "most accurate among the careful" model

Trigram	Ageism Probability	Job titles using this trigram	Hiring companies using this trigram
people can feel	0.919016808	1	1
feel pain sale	0.896356442	3	1
goal everyday feel	0.892764166	7	1
share value feel	0.89020273	1	1
help people feel	0.889639137	2	1
love business opportunity	0.027541937	1	1
love company will	0.027398801	1	1
love lord god	0.02623294	1	1
love great copy	0.025743557	1	1
show ability good	0.025665512	1	1

Table 46. Five most ageist and five least ageist trigrams in Call Centre & Customer Service, using the "most accurate among the careful" model

Trigram	Ageism Probability	Job titles using this trigram	Hiring companies using this trigram
feel customer care	0.917149406	1	1
feel value care	0.916392281	1	1
improve fast feel	0.909914696	2	1
help client feel	0.903155589	1	1
finger handle feel	0.902534178	1	1
bring follow good	0.034505043	1	1
order form good	0.033977829	1	1
love learn new	0.02877662	1	1
love join team	0.026100746	1	1
love enjoy flexible	0.018937496	1	1

Table 47. Five most ageist and five least ageist trigrams in Accounting & Finance, using the "most accurate among the careful" model

Trigram	Ageism Probability	Job titles using this trigram	Hiring companies using this trigram
feel value matter	0.894759894	3	1
goal everyday feel	0.892764166	5	1
feel include value	0.89238937	3	1
feel welcome value	0.890863669	2	2
share value feel	0.89020273	1	1
search great talent	0.036936882	1	1
record youre good	0.036785739	1	1
financial service love	0.035474576	1	1
create great financial	0.03538812	1	1
love meet skill	0.021380894	1	1

Table 48. Five most ageist and five least ageist trigrams in Administration, Coordination & Office Support, using the "most accurate among the careful" model

Trigram	Ageism Probability	Job titles using this trigram	Hiring companies using this trigram
goal everyday feel	0.892764166	1	1
feel opportunity present	0.886866846	1	1
feel free browse	0.884055551	1	1
believe take feel	0.871613729	1	1
point list feel	0.864895501	1	1
good talent can	0.038835543	1	1
bring forward idea	0.037710466	1	1
company wed love	0.028603866	1	1
love appreciate work	0.028525294	1	1
love gaming can	0.025707413	1	1

Summary of Results of Ageism-Identification Techniques

Table 49 compares results of this study's techniques, showing classification models being viable trigram-level ageism-identification techniques with only 0.8 to 1.1 percent of all trigrams flagged as ageist. At the job-ad level, however, where the presence of only one ageist trigram would make an entire job ad ageist, these classification models cease to be viable as they unrealistically flag ageist more than half of the job ads scraped.

Table 49. Summary of results of this study's various ageism-identification techniques

Technique Used	Ageist Job Ads	Total Job Ads	% Ageist Job Ads	Ageist Trigrams	Total Trigrams	% Ageist Trigrams	Discussion
Embeddings-Based: "bachelor degree graduate"	4,214	8,918	47.3%	n/a	1,886,910	n/a	close to half of job ads being ageist is unrealistic; "bachelor degree graduate" is a reasonable job requirement
Embeddings-Based: "relevant work experience"	741		8.3%	n/a		n/a	not viable, as "relevant work experience" is a reasonable job requirement
Classification: most careful among preferable	4,565		51.2%	15,359		0.8%	viable at trigram-level but not at job ad-level, where more than half of job ads are flagged ageist
Classification: most accurate among careful	4,949		55.5%	21,303		1.1%	viable at trigram-level but not at job ad-level, where more than half of job ads are flagged ageist
Top2Vec Clustering	1,050		11.8%	n/a		n/a	viable in surfacing job ads that may need further review for ageist language

Top2Vec clustering, on the other hand, appears to be a viable technique at the job ad level, flagging a modest 11.8 percent of the job ads scraped for potential ageism review.

All other techniques, due to either their flagging of reasonable job requirements or their sweeping ageism sensitivity (e.g., half of job ads scraped flagged ageist), appear to be unviable.

Conclusion and Policy Implications

This project combines traditional, alternative, and recent text-mining techniques to quantitatively determine implicit ageism in Philippine online job advertisements. Techniques include gathering training data, preprocessing text, and multiple layers of model-building and scoring (LDA, doc2vec, logistic regression, penalized ridge regression, and Top2Vec). Practical alternatives for selecting a champion model when an ideal ageism-classification model is not apparent include the "most careful among the preferable" and the "most accurate among the careful." The best-performing classification models were penalized ridge regression models with doc2vec vectors from Paragraph Vector–Distributed Memory architectures as features. These models identified ageist job ads consistent with warm-but-incompetent stereotypes about older persons (Ayalon and Tesch-Römer 2018).

Future work should experiment with other classification algorithms and further tune hyperparameters. Researchers should also consider the overall thematic undertones of job ads rather than focusing solely on ageist-scored trigrams. The study also shortlisted potentially ageist job ads through document embeddings and clustering, identifying job ads grouped into predominantly ageist clusters.

Future research should separate the Job Description section from the Company Description section of job ads to allow more targeted rewrites. Additional considerations include more domain-specific, Philippine-based, and job-related prelabeled text. Combining predictions from all techniques may also be explored.

This project contributes uniquely to ageism research by using various machine learning techniques beyond the word2vec or doc2vec embeddings of previous studies and by focusing on the Philippine setting. It found that trigram-level classification models and job ad-level Top2Vec clustering models are more viable for ageism detection than are cosine similarity scores of job ads to known ageist text. Practically, the study suggests avoiding unnecessary technical jargon, focusing on desired outcomes, carefully phrasing physical conditions, leveraging positive language, and using verbiage that appeals to both older and younger applicants.

Implications to policy include recognizing implicit—not just explicit—forms of ageism and requiring age-inclusive language in employment advertising.

Appendix

Link to Data, Code, and doc2vec Models Developed

R data, code, and binary files of doc2vec models used in this study are available at https://tinyurl.com/ageism-capstone.

Acknowledgment

The author extends gratitude to Associate Professor Francisco N. de los Reyes, UP Diliman School of Statistics, for his invaluable contribution as adviser to this project.

References

Angelov, Dimo. 2020. "Top2Vec: Distributed Representations of Topics." *arXiv.org*, 19 August 2020. https://arxiv.org/abs/2008.09470.

Ayalon, Liat, and Clemens Tesch-Römer, eds. 2018. *Contemporary Perspectives on Ageism*. SpringerOpen. https://doi.org/10.1007/978-3-319-73820-8.

Badana, Adrian N. S., and Ross Andel. 2018. "Aging in the Philippines." *Gerontologist* 58 (2): 212–18. https://doi.org/10.1093/geront/gnx203.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3: 993–1022. https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf.

Burn, Ian, Patrick Button, Luis Munguia Corella, and David Neumark. 2020. "Older Workers Need Not Apply? Ageist Language in Job Ads and Age Discrimination in Hiring." *IZA Institute of Labor Economics Discussion Paper Series*, no. 13506, July 2020. Social Science Research Network. https://doi.org/10.2139/ssrn.3660252.

Butler, Robert N. 1969. "Age-Ism: Another Form of Bigotry." *Gerontologist* 9 (4 Part 1): 243–46. https://doi.org/10.1093/geront/9.4_part_1.243.

Centre for Ageing Better. 2018. *Becoming an Age-Friendly Employer*. 10 September 2018. https://ageing-better.org.uk/resources/becoming-age-friendly-employer.

Centre for Ageing Better. 2021a. *Ads for All Ages: How Age-Biased Are Job Adverts in the UK, and What Can We Do About It?* 22 April 2021. https://ageing-better.org.uk/resources/ads-for-all-ages.

Centre for Ageing Better. 2021b. *Good Recruitment for Older Workers: Understanding and Improving Recruitment Language, Imagery and Messaging*. April 2021. https://ageing-better.org.uk/sites/default/files/2021-04/understanding-recruitment-language-GROW.pdf.

Commission on Human Rights of the Philippines. 2021. "Inputs on Ageism and Age-Discrimination." 30 April 2021. https://www.ohchr.org/Documents/Issues/OlderPersons/AgeismAgeDiscrimination/Submissions/NHRIs/Human-Righs-Commission-Philippines.docx.

Cox, D. R. 1958. "The Regression Analysis of Binary Sequences." *Journal of the Royal Statistical Society*. Series B 20 (2): 215–42. https://www.jstor.org/stable/2983890.

Data Mining, Text Mining, Künstliche Intelligenz. 2022. "CLUD5: DBSCAN Parameterization." YouTube. 24 May 2022. https://www.youtube.com/watch?v=PoJmJ8wu6YI.

de Guzman, Allan B., Erika Largo, Lillianne Mandap, and Virnalyn Mae Muñoz. 2014. "The Mediating Effect of Happiness on the Job Satisfaction of Aging Filipino Workers: A Structural Equation Model (SEM)." *Educational Gerontology* 40 (10): 767–82. https://doi.org/10.1080/03601277.2014.882108.

Díaz, Mark. 2020. "Age Bias Training and Testing Data." Data set, version 1. Harvard Dataverse. https://doi.org/10.7910/DVN/F6EMTS.

Díaz, Mark, Isaac Johnson, Amanda Lazar, Anne Marie Piper, and Darren Gergle. 2018. "Addressing Age-Related Bias in Sentiment Analysis." *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, paper no. 412. Association for Computing Machinery. https://doi.org/10.1145/3173574.3173986.

Department of Labor and Employment. 2017. "Age Does Not Matter in Employment: DOLE Issues Rules on Anti-Age Discrimination in Employment." 11 February 2017. https://www.dole.gov.ph/news/age-does-not-matter-in-employment-dole-issues-rules-on-anti-age-discrimination-in-employment/.

Durrheim, Kevin, Maria Schuld, Martin Mafunda, and Sindisiwe Mazibuko. 2022. "Using Word Embeddings to Investigate Cultural Biases." *British Journal of Social Psychology* 62 (1): 617–29. https://doi.org/10.1111/bjso.12560.

Ester, Martin, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. 1996. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, 226–31. https://doi.org/10.5555/3001460.3001507.

Eustaquio, John D. 2021a. "Building, Checking, and Applying Logistic Regression Models." PowerPoint slides, School of Statistics, University of the Philippines, 11 March 2021.

Eustaquio, John D. 2021b. "Methods for Classification." PowerPoint slides, School of Statistics, University of the Philippines, 25 March 2021.

Griffiths, Thomas L., and Mark Steyvers. 2004. "Finding Scientific Topics." *Proceedings of the National Academy of Sciences of the United States of America* 101(suppl_1): 5228–35. https://doi.org/10.1073/pnas.0307752101.

Hoerl, Arthur E., and Robert W. Kennard. 1970. "Ridge Regression: Biased Estimation for Nonorthogonal Problems." *Technometrics* 12 (1): 55–67. https://doi.org/10.2307/1267351.

Jobstreet Philippines Content Team. 2022a. "Here Are 5 Ways to Deal With Age Discrimination During the Job Hunt." JobStreet, 13 April 2022. https://www.jobstreet.com.ph/career-advice/article/here-are-5-ways-to-deal-with-age-discrimination-during-the-job-hunt.

Jobstreet Philippines Content Team. 2022b. "Jobseekers, Be Informed: Anti-Age Discriminatory Laws Are Now in Effect!" Jobstreet, 13 April 2022. https://www.jobstreet.com.ph/career-advice/article/jobseekers-be-informed-anti-age-discriminatory-laws-are-now-in-effect.

Jones, Tommy. 2019. Optimizing Topic Models for Classification Tasks." Lander Analytics. 3 December 2019. Video of lecture, 20:08. https://www.youtube.com/watch?v=k6kCSzWigiI.

Le, Quoc V., and Tomas Mikolov, T. 2014. "Distributed Representations of Sentences and Documents." *arXiv.org*, version 2, 22 May 2014. https://doi.org/10.48550/arXiv.1405.4053.

Lustañas, Bridgette M. 2019. "Against the Old?: A Critical Discourse Analysis of Philippine Online News Articles' Ageist Ideologies." *Asian Journal of English Language Studies* 7: 84–110. https://doi.org/10.59960/7.a5.

McInnes, Leland, John Healy, and James Melville. 2020. "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction." *arXiv.org*, version 3, 18 September 2020. https://doi.org/10.48550/arXiv.1802.03426.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. "Efficient Estimation of Word Representations in Vector Space." *arXiv*, version 3, 7 September 2013. https://doi.org/10.48550/arXiv.1301.3781.

Moneymax Editorial Team. 2023. "Looking for Work? Check Out These Legit Online Job Sites." *Moneymax*, 28 July 2023. https://www.moneymax.ph/lifestyle/articles/online-job-sites.

Octoparse. 2024. "Scrape Jobs from Linkedin." 20 December 2024. https://helpcenter.octoparse.com/en/articles/6471036-scrape-jobs-from-linkedin.

Octoparse. 2026. "Octoparse Pricing Plans." Accessed 29 July 2026. https://www.octoparse.com/pricing.

Ramos, Arvin. 2022. "How HR Managers can Tackle Age Discrimination in the Workplace." *Manila Recruitment*, 10 February 2022. https://manilarecruitment.com/manila-recruitment-articles-advice/managers-tackle-age-discrimination/.

ritvikmath. 2021. "Word2VEC : Natural Language Processing." 23 June 2021. Educational video, 13:16. https://www.youtube.com/watch?v=f7o8aDNxf7k.

Robles, Desiree R. 2023. "Web Scraping of Commodities for Consumer Price Index in the National Capital Region, Philippines." Paper presented at the 2nd Philippine Data Festival, Manila, Philippines, 4 October 2023. https://psa.gov.ph/sites/default/files/ss_ssd/web-scraping-of-commodities-for-consumer-price-index-in-the-national-capital-region-philippines.pdf.

Schubert, Erich, Jörg Sander, Martin Ester, Hans Peter Kriegel, and Xiaowei Xu. 2017. "DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN." *ACM Transactions on Database Systems* 42 (3): 1–21. https://doi.org/10.1145/3068335.

Serrano, Luis. (2020, March 19). "Latent Dirichlet Allocation (Part 1 of 2)." Luis Serrano Academy. Educational video, 26:56. https://www.youtube.com/watch?v=T05t-SqKArY.

Similarweb. n.d. "Jobs and Career Category Leaders Sep 2023 - Nov 2023 (3 Months)." Accessed 26 June 2026. https://pro.similarweb.com/#/digitalsuite/markets/webmarketanalysis/mapping/Jobs_and_Career/608/3m?webSource=Total.

Supranes, Michael Van. 2021. "Introduction to Topic Modeling." Lecture presented at the School of Statistics, University of the Philippines, 18 May 2021.

Supranes, Michael Van, and Lansangan, Joseph Ryan. n.d. "Regularization." Lecture presented at the School of Statistics, University of the Philippines.

Tilburg Science Hub. 2024. "Text Pre-processing in R - Tilburg Science Hub." Tilburg University. https://tilburgsciencehub.com/topics/manage-manipulate/manipulate-clean/textual/text-preprocessing/.

Treadway, Andrew. 2019. "BeautifulSoup vs. Rvest." *R-bloggers*, 22 July 2019. https://www.r-bloggers.com/2019/07/beautifulsoup-vs-rvest/.

United Nations. 2023a. "Goal 1: End Poverty in All Its Forms Everywhere." United Nations Sustainable Development, 7 September 2023. https://www.un.org/sustainabledevelopment/poverty/.

United Nations. 2023b. "Goal 10: Reduce Inequality Within and Among Countries." United Nations Sustainable Development, 11 September 2023. https://www.un.org/sustainabledevelopment/inequality/.

United Nations. 2023c. "Promote Inclusive and Sustainable Economic Growth, Employment and Decent Work for All." United Nations Sustainable Development, 11 September 2023. https://www.un.org/sustainabledevelopment/economic-growth/.

Valdez, Jana Patricia M., Julio L. Angeles, Ernesita Pareja-Corpuz, and Christopher P. Hernandez. 2013. "Perceptions Towards Aging Among Selected Filipino Adults." *International Journal of Research Studies in Psychology* 2 (3): 89–96. https://doi.org/10.5861/ijrsp.2013.458.

Vayansky, Ike, and Sathish A. P. Kumar. 2020. "A Review of Topic Modeling Methods." *Information Systems* 94 (101582). https://doi.org/10.1016/j.is.2020.101582.

Wickham, Hadley. 2022. *Rvest: Easily Harvest (Scrape) Web Pages*. R package version 1.0.3, 20 August 2022. https://doi.org/10.32614/CRAN.package.rvest.

Wijffels, Jan. 2021. *Doc2vec: Distributed Representations of Sentences, Documents and Topics*. R package version 0.2.0, 28 March 2021. https://CRAN.R-project.org/package=doc2vec.

Wijffels, Jan. n.d. *Bnosac:: Open Analytical Helpers - Doc2vec in R*. https://www.bnosac.be/index.php/blog/103-doc2vec-in-r.

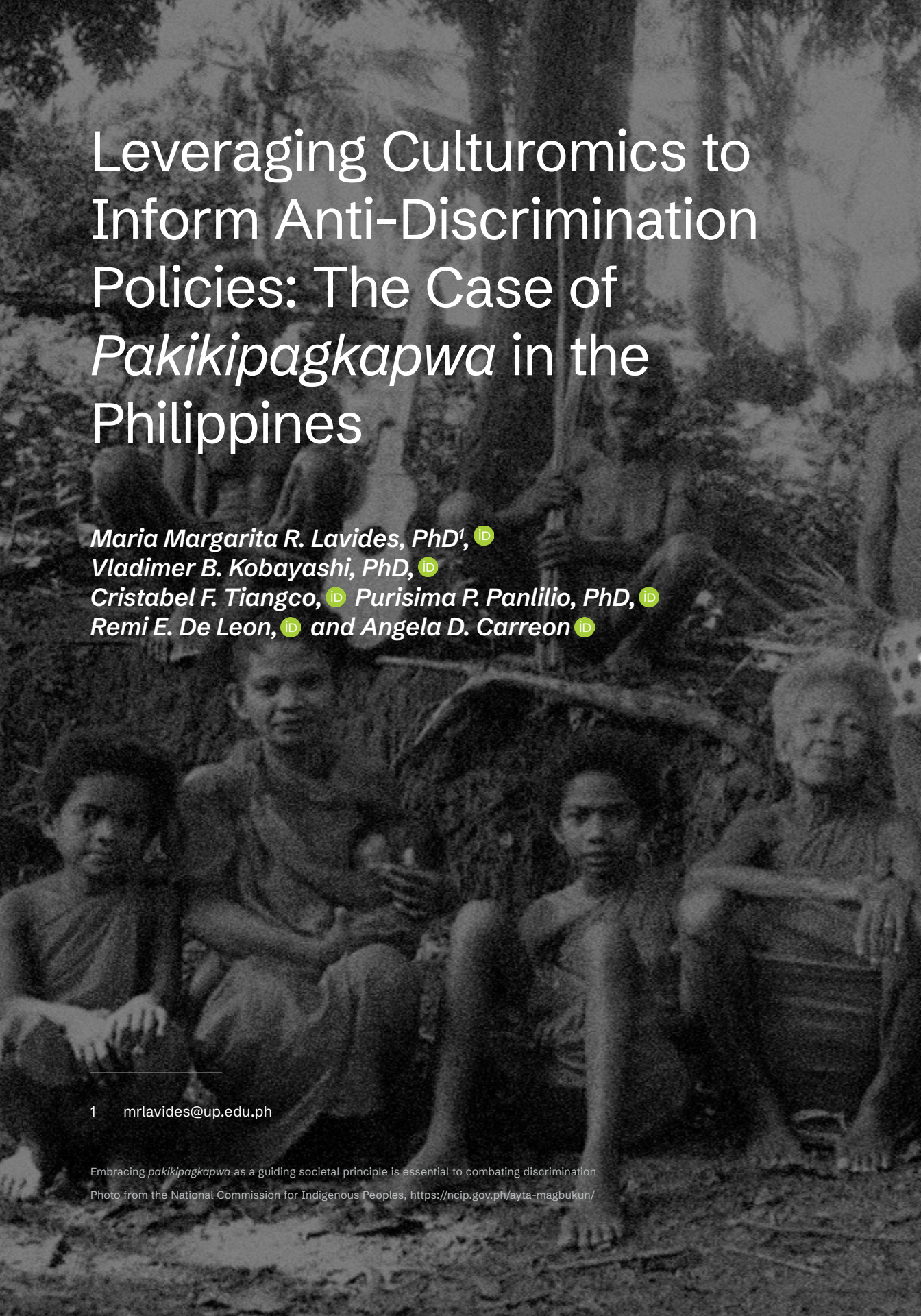
Wijffels, Jan. n.d. *Bnosac:: Open Analytical Helpers - Word2vec in R*. http://www.bnosac.be/index.php/blog/100-word2vec-in-r.

World Health Organization. 2021. "Ageing: Ageism." 18 March 2021. https://www.who.int/news-room/questions-and-answers/item/ageing-ageism.

Leveraging Culturomics to Inform Anti-Discrimination Policies: The Case of *Pakikipagkapwa* in the Philippines

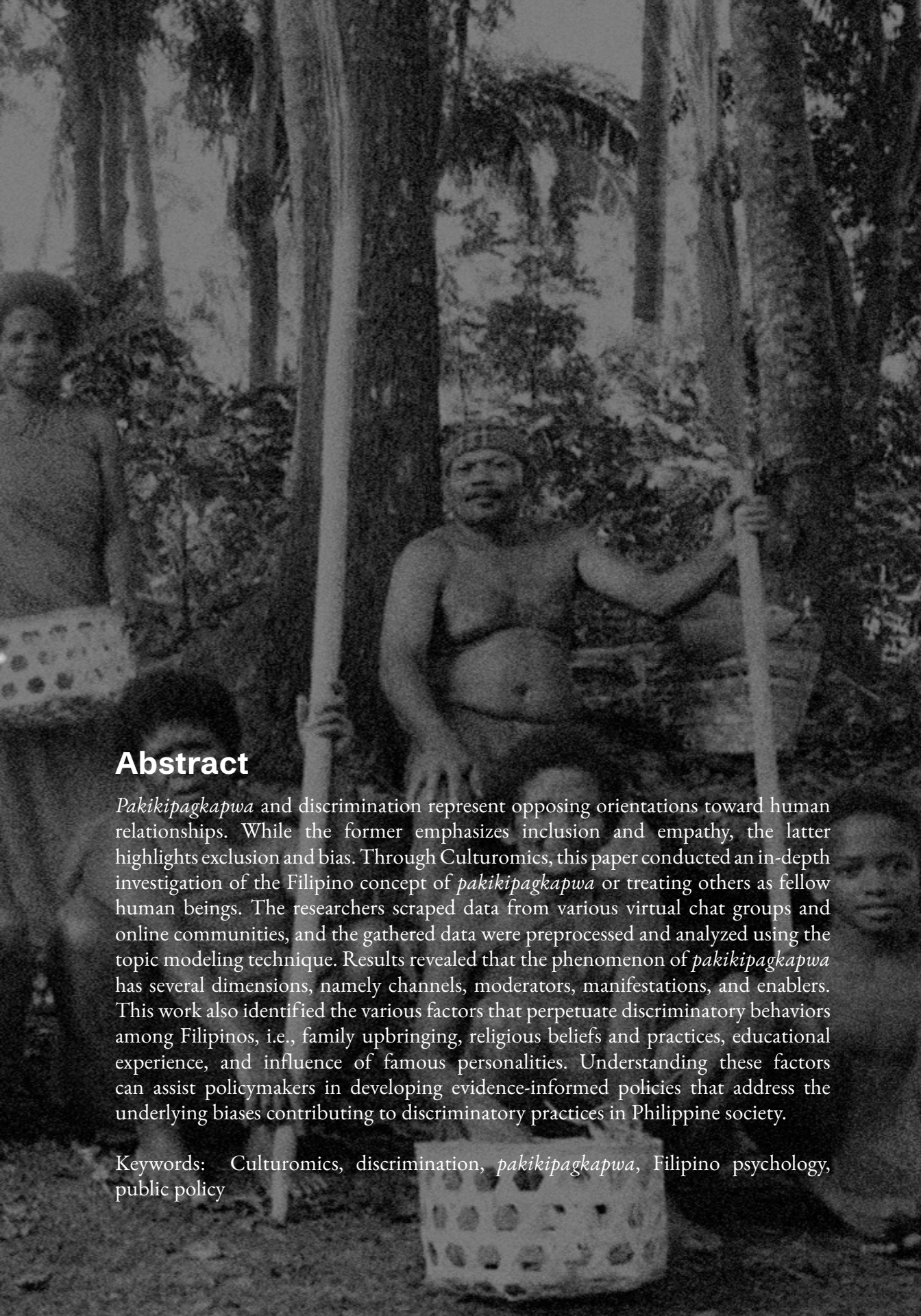
Maria Margarita R. Lavides, PhD[1],
Vladimer B. Kobayashi, PhD,
Cristabel F. Tiangco, Purisima P. Panlilio, PhD,
Remi E. De Leon, and Angela D. Carreon

1 mrlavides@up.edu.ph



Embracing *pakikipagkapwa* as a guiding societal principle is essential to combating discrimination

Photo from the National Commission for Indigenous Peoples, https://ncip.gov.ph/ayta-magbukun/



Abstract

Pakikipagkapwa and discrimination represent opposing orientations toward human relationships. While the former emphasizes inclusion and empathy, the latter highlights exclusion and bias. Through Culturomics, this paper conducted an in-depth investigation of the Filipino concept of *pakikipagkapwa* or treating others as fellow human beings. The researchers scraped data from various virtual chat groups and online communities, and the gathered data were preprocessed and analyzed using the topic modeling technique. Results revealed that the phenomenon of *pakikipagkapwa* has several dimensions, namely channels, moderators, manifestations, and enablers. This work also identified the various factors that perpetuate discriminatory behaviors among Filipinos, i.e., family upbringing, religious beliefs and practices, educational experience, and influence of famous personalities. Understanding these factors can assist policymakers in developing evidence-informed policies that address the underlying biases contributing to discriminatory practices in Philippine society.

Keywords: Culturomics, discrimination, *pakikipagkapwa*, Filipino psychology, public policy

Introduction

Discrimination is pervasive in the Philippines, with reported cases based on gender (Human Rights Watch 2017), age (Handog 2021), physical appearance (Dayrit and Alibudbud 2022; Lavides 2024, 123–24), and health conditions (Mrakovcic 2024, 62), among others. It not only causes social divisiveness (Khan, Combaz, and McAslan 2015, 3) but also anxiety and trauma to those subjected to it (Williams et al. 2019, 1383). Hence, the government must mitigate factors that pave the way for such prejudiced practices.

The Filipino concept of *pakikipagkapwa*, understood as recognizing others as fellow human beings, naturally stands in opposition to discriminatory practices. *Pakikipagkapwa* is a component of Dr. Virgilio Enriquez's *kapwa* model value structure (1992, 75), as well as Clemente et al.'s reformulated value system of the Filipino adolescent (2008, 25). The limited number of empirical studies on *pakikipagkapwa* is perhaps due to the difficult and laborious process of collecting data that permits thorough investigations. Typically, information is collected from volunteer respondents who complete a Likert-type questionnaire or through interviews. Various challenges also reduce the generalizability and validity of studies about *pakikipagkapwa*, including difficulty in recruiting respondents; ensuring an adequate sample size and the resulting representativeness; questionnaire reliability and validity; obtaining honest responses (people may say one thing but do the opposite thing); and consistent participation of respondents (who may later opt out of the study).

Presently, researchers are turning to online data to study people's values, behaviors, and preferences. The exponential growth of internet-generated data enabled the emergence of a new field called Culturomics, or the use of big data analytics to mine digital traces of human activity or interaction that reflect people's thoughts and behaviors (Michel et al. 2011; Luo et al. 2024). Filipinos are among the heaviest online users and consumers. They use the digital space to interact, voice opinions, and support various advocacies. Due to the prevalent use of digital technologies among Filipinos, there is now an abundance of data. These data further explicate the concept of *pakikipagkapwa*, following the approach of culturomics, which studies human language, cultures, and behavior within a given set of digital texts to deduce emerging cultural phenomena from large bodies of online data. This also opens the possibility of applying data science methods to study Filipino culture, values, and traditions.

This paper makes an in-depth investigation of the Filipino concept of *pakikipagkapwa* using data science methods. The researchers scraped data from various virtual chat groups and online communities, and the gathered data were

preprocessed and analyzed using the topic modeling technique. Results revealed that the phenomenon of *pakikipagkapwa* has several dimensions, namely channels, moderators, manifestations, and enablers. Channels refer to modes through which *pakikipagkapwa* is expressed. Moderators are factors that affect the extent to which *pakikipagkapwa* is exercised. Manifestations are the tangible representations of *pakikipagkapwa*. Enablers are the catalysts of *pakikipagkapwa*. Aside from presenting an emergent model of *pakikipagkapwa*, this work also unearthed the various factors relevant to understanding the prevalence of discriminatory behaviors in the Philippines—i.e., family upbringing, religious beliefs and practices, educational experience, and the influence of famous personalities. Awareness of such factors can guide government decision-makers in developing policies that can address the proliferation of biases that promote discriminatory practices in the country.

Literature Review

Kapwa, or shared identity, is at the heart of *Sikolohiyang Pilipino* (Filipino psychology). It is regarded as a stable construct integral to social interactions among Filipinos. Often, it is perceived as a value. Values are internalized cognitive structures from which people draw for decision-making. These reflect one's basic principles of morality (i.e., right versus wrong) and sense of priorities (e.g., whether to forgive a transgressor or not) (Oyserman 2015, 36). Additionally, values can be studied either on the individual level or on the group/societal level through norms, priorities, and guidelines.

In the Philippines, discussions of Filipino values initially revolved around *utang na loob* (gratitude), *hiya* (shame), and *pakikisama* (yielding to the leader or majority). These discussions were based on studies in the 1960s by Racelis and Lynch, among others. However, Virgilio Enriquez, known as the father of *Sikolohiyang Pilipino*, contends that *utang na loob*, *hiya*, and *pakikisama* are merely colonial/accommodative surface values, with *kapwa* being the core value.

Enriquez partly objects to the use of an outsider's point of view (i.e., research by non-Filipino, usually Western, academics) in analyzing and identifying Filipino values without contextual information that an insider may have. Meanwhile, Mendoza and Perkinson describe *kapwa* as:

> A clear instance of a core value of a non-Western people locked inside its own culture in titanic struggle, with an invasive counter-value that does not submit readily to its own ethic of reciprocity . . . The surface perception of Filipino culture as smooth-seeking and serenity-entreating, in this particular case, is not merely a Western misreading, but also a protective Filipino misleading, safeguarding a secret strength that offers its code only to a particular mode of participation (2003, 291)

The proposed value structure can be seen in figure 1, which is classified into four major categories: (a) surface (colonial/accommodative and confrontative) values; (b) pivotal interpersonal value; (c) core value; and (d) foundation of human values (linking socio-personal and associated societal values). *Pakiramdam* (capacity to feel) functions as the processor from which the core value of *kapwa* spins off toward the surface values; it is the pivotal value of shared inner perception (Enriquez 1992, 75–76). Should *pakiramdam* be absent, "*kapwa*, in all its manifestations, becomes inoperative since it ceases to have a means of expression or communication" (Mendoza and Perkinson 2003, 289).

Figure 1. Behavior, patterns, and value structure: surface, core, and societal. Reproduced from Enriquez (1992, 93).

Colonial/Accommodative Surface Value	*hiya* (propriety/ dignity)	*utang na loob* (gratitude/ solidarity)	*pakikisama* (companion-ship/esteem)
Confrontative Surface Value	*bahala na* (determination)	*sama/lakas ng loob*(resent-	*pakikibaka* (resistance) ment/guts)
Pivotal Interpersonal Value	*Pakiramdam (pakikipagkapwa-tao)* (shared inner perception)		
CORE VALUE	***KAPWA (Pagkatao)*** (shared identity)		
Linking Socio-Personal Values	*Kagandahang-loob (Pagkamakatao)* (shared humanity)		
Associated Societal values	*karangalan* (dignity)	*katarungan* (justice)	*kalayaan* (freedom)

Analyses of social interaction using the Filipino language yielded eight levels and modes of behavior. These are the following: *pakikitungo* (level of amenities/ civility); *pakikisalamuha* (level of "mixing"); *pakikilahok* (level of joining/ participating); *pakikibagay* (level of conforming); *pakikisama* (level of adjusting); *pakikipagpalagayang-loob* (level of mutual trust/rapport); *pakikisangkot* (level of getting involved); and *pakikiisa* (level of fusion, oneness, and full trust) (Santiago and Enriquez 1976, as mentioned in Enriquez 1992, 49–51).

The different levels of interpersonal relations have different behavioral manifestations. For example, *pakikitungo*, *pakikisalamuha*, *pakikilahok*, *pakikibagay*, and *pakikisama* typically describe how one relates to others as an outsider, or as *ibang tao* (literally, another person). Meanwhile, *pakikipagpalagayang-loob*, *pakikisangkot*, and *pakikiisa* are typical of interactions with those in the one-of-us category, or *hindi ibang tao* (not another person). *Kapwa* is a superordinate concept that embraces both categories of outsider and one-of-us. Meanwhile, *pakikipagkapwa* embraces all levels of both categories (Enriquez 2018, 289).

The levels imply that *pakikipagkapwa* is not simply about engaging in social interaction. Before being treated as an equal, or *hindi ibang tao*, one needs to interact first as *ibang tao*; the process goes from *pakikitungo* to *pakikisalamuha*, then *pakikilahok* to *pakikibagay*, and lastly to *pakikisama*, which is the final level of social interaction before one gets considered in the *hindi-ibang-tao* category. Thus, *pakikipagkapwa* requires interpersonal skills in dealing with others, whether *ibang tao* or *hindi ibang tao*.

The spirit of *alay-kapwa* (*alay* means offer or offering) was concretized by the emergence of community pantries during the COVID-19 pandemic. Despite several restrictions, people contributed food to communities, which supplemented the diminishing supply of government relief. Community pantries were driven by a spirit of fraternal charity and a sense of social responsibility for the suffering of others. Organizers and contributors recognized that people in need are fellow humans, whose "existence is closely [tied] to their own existence, and the humanity of the needy is the same humanity of those who are giving" (Canete, Rocha, and Dolosa 2022, e296).

Recent empirical investigations on *kapwa* attempted to address the conceptual and operational ambiguity of the term and the lack of evidence of its being a core value. Labor and Gastardo-Conaco (2021, 16) used a bottom-up approach through language and argued for a more nuanced definition of *kapwa*: a "view of a highly positive social other to which one has strong connections and interactions." This was seen as empirical support for the notion of *kapwa* as a social relational construct rather than as a value. A separate social category for *di-kapwa* was also found, in contrast to Enriquez's model, where everyone is considered as *kapwa*, even outsiders. People in the *di-kapwa* category are described as having negative characteristics and engaging in undesirable behaviors. Since they are outside the *kapwa* boundaries, the acceptable modes of social interaction may differ, in which an individual may display unpleasant behavior toward a *di-kapwa*.

In another study, Clemente et al. (2008, 20) asked 47 students enrolled in a course, Introductory Psychology, to create a diagram/map showing links among the thirteen values identified in their updated *kapwa* value list. Included here are:

- *maka-Diyos* (Godly/religious);
- *pakikibaka* (struggling for a cause);
- *lakas ng loob* (inner strength, bravery, willfulness);
- *katarungan* (justice);
- *kalayaan* (freedom);

- *karangalan* (honor);
- *pakikipagkapwa* (a sense of shared identity);
- *pakikiramdam* (the ability to sense or feel what others are experiencing);
- *pakikisama* (Getting along with others);
- *utang na loob* (A deep sense of gratitude);
- *paggalang* (respect);
- *kagandahang-loob* (virtue); and
- *hiya* (shame)

From the original values proposed by Enriquez, *bahala na* was endorsed the least and was replaced with "*maka-Diyos*" and "*paggalang*." The data from the multidimensional scaling (MDS) analysis did not conform to the a priori categories by Enriquez. No core value emerged. The MDS also derived distinct dimensions (instead of categories), including *hindi ibang tao* to *ibang tao*, and *sarili* (self) to *grupo* (group) to *lipunan* (society) (see figure 2).

Figure 2. Reformulated Value System of the Filipino Adolescent (reproduced from Clemente et al. 2008, 25)



Like Labor and Gastardo-Conaco (2021) and Clemente et al. (2008), Bernardo, Clemente, and Liem (2014) did not find *kapwa* as a core value. In a study with an etic approach, 752 students from two secondary schools and one university in Metro Manila completed the Portrait Values Questionnaire (PVQ). The results showed that "benevolence" received a high ranking (Bernardo, Clemente, and Liem 2014, 6), which is consistent with the *kapwa* model. Additionally, priority was given to "universalism," which is also consistent with the shared identity model. Moreover, there were other values that were reported by the Filipino adolescent participants to be more salient, including "self-direction," which ranked the highest.

Meanwhile, the findings of Yacat (2017, 18) imply that the hierarchy of values proposed by Enriquez applies only to the *ibang-tao* category and not to the *hindi-ibang-tao* category. In his study, actions that go against *pakikipagkapwa-tao* were considered more serious. They could terminate a relationship if the transgressor is considered as *ibang tao*, and not *hindi ibang tao*.

Methodology

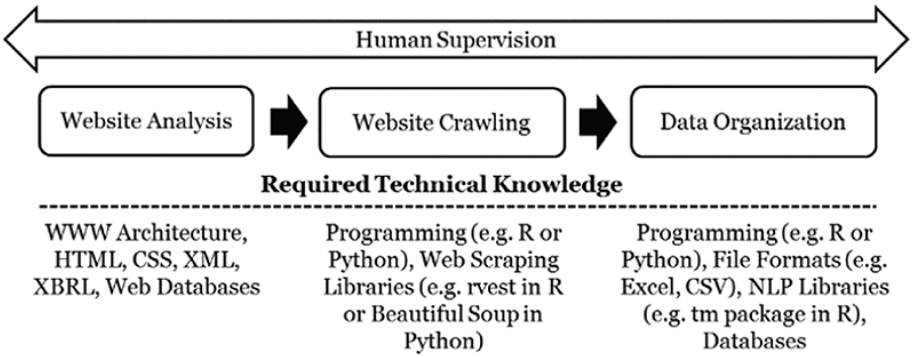
To address the limitations of empirical research on Enriquez's *kapwa* model, and to examine the initial assertions of Clemente et al., this study applied data analytics techniques to answer its research questions. These techniques include web scraping, text mining processes (e.g., topic modeling), and developing semantic networks.

Web Scraping

The data are from popular social media platforms, namely YouTube, X (formerly Twitter), and various Philippine forums (see column labels of table 2). They consist of social media posts, comments, and tweets. Data were collected through web scraping. Web scraping, also known as screen scraping, web data extraction, or web harvesting, automatically extracts or retrieves data from the World Wide Web (WWW), as opposed to manually copying and saving them to a file or database (Singrodia et al. 2019; Zhao 2017). The method uses various technologies and tools to automate collection and organization, such as programming languages (e.g., R or Python) and software libraries (e.g., Selenium, Beautifulsoup, Snscrape, etc.). This involves three phases: website analysis, website crawling, and data organization, as shown in Figure 3 (Krotov and Silva 2018).

Website analysis involves a basic understanding of HTTP (Hypertext Transfer Protocol) and HTML (Hypertext Markup Language). The structure of a website must be examined first to understand how the data are embedded and how to scrape them. Meanwhile, website crawling involves developing and running a script using programming languages (e.g., R and Python) (Krotov and Silva 2018; Krotov and Tennyson 2018). The present study particularly used the Python programming language and Python packages. Python packages include Beautiful Soup for parsing HTML and XML documents, Selenium for browser automation, and the snscrape Python library for scraping from social networking sites such as X (formerly Twitter).

Figure 3. Website scraping phases, as lifted from Krotov and Tennyson (2018)



In the case of some websites or applications, there is no need to use web scraping because they have application programming interface (API) services. The API is an interface that allows two software components to communicate with each other, following a set of terms and conditions.16 Developers of Facebook, YouTube, and X provide APIs that would allow users to access data. To use APIs, we must create an account, go to the developer's website to register an application, and then obtain an application ID (AppID), application secret (AppSecret), or access tokens.

Comments, posts, and replies were gathered using web scraping and APIs. The objective is to examine whether the *pakikipagkapwa* value is manifested or mentioned in these discussions. Sources for this study include the following: two social media platforms (YouTube and X) and five online forums (the r/Philippines subreddit, PinoyExchange (PEx), Philippines Expats Forum, Expat.com, and Skyscraper City). YouTube comments were extracted from Fernando Poe Jr.'s (FPJ's) *Ang Probinsyano* videos, which were posted on the official ABS-CBN YouTube channel. The authors specifically chose *Ang Probinsyano* after hypothesizing that the popular television series exhibits the *pakikipagkapwa* value (and other related values) in many of its episodes. For the other platforms, the authors used keywords related to *pakikipagkapwa* to gather posts or comments that are potentially relevant. The use of keywords helps filter out irrelevant discussions. The keywords are listed in the first column of table 1.

Table 1. List of the Total Number of Scraped Comments from Different Sources Related to *Pakikipagkapwa*

Keywords	Data Sources				
	Twitter	Philippine Expats Forum	Skyscraper City	Pinoy Exchange	Expat.com
Pakikipagkapwa (Regarding others as fellow human beings)	3,040	295	5,732	3,766	73

Keywords	Data Sources				
	Twitter	Philippine Expats Forum	Skyscraper City	Pinoy Exchange	Expat.com
Kapwa (Fellow human being)	60,000	0	0	0	0
Pakikitungo (Amenities or civility)	15,000	0	69,437	0	0
Pakikisalamuha (Literally "mixing" or engaging)	0	0	13,382	0	0
Pakikilahok (Joining or participating)	15,000	0	16,453	0	0
Pakikibagay (Conforming)	499	0	0	0	0
Pakikisama (Comradeship)	10,000	0	15,279	0	0
Pakikipagpalagayang-loob (Establishing mutual trust/ rapport)	32	0	0	0	0
Pakikisangkot (Getting involved)	245	0	5,078	0	0
Pakikiisa (Solidarity)	12,206	0	0	0	0
Pakikibaka (Joining in the struggle)	5,000	0	12,236	0	0
Paggalang (Act of respect)	20,000	0	0	0	0
Pagbibigay-galang (Showing respect)	436	0	0	0	0
Pagrespeto (Respect)	8,311	0	9,382	0	0
Pagtulong (Helping)	6,809	0	4,679	0	0
Pagbibigay-tulong (Offering assistance		0		0	0
Pag-aalaga (Caring or nurturing)	5,196	0	11,016	0	0
Pag-aaruga (Nurturing)		0		0	0
Pagdamay (Commiseration)	14,149	0	0	0	0
Pakikiramay (Showing compassion)	16,129	0	0	0	0
Sumaklolo (Providing aid or support)	3,538	0	0	0	0
Saklolohan (Providing aid or support)		0	3,854	0	0

Keywords	Data Sources				
	Twitter	Philippine Expats Forum	Skyscraper City	Pinoy Exchange	Expat.com
Tulong-tulong (Collective assistance)	15,000	0	0	0	0
Paglilingkod (Act of service)	44,778	0	0	0	0
Pag-aabuloy (Giving donation)	923	0	0	0	0
Pag-alalay (Providing support)	2,950	0	0	0	0
Magtustos (To provide)	227	0	0	0	0
Tustusan (To provide)		0	0	0	0
Magsustento (To provide financial support)	6,659	0	0	0	0
Sustentuhan (To provide financial support)		0	0	0	0
Pagtataguyod (to provide support)	4,333	0	0	0	0
Nagtaguyod (Provided support)		0	4,918	0	0
Pagsasakripisyo (Sacrificing)	7,309	0	0	0	0
Pagbibigay (Giving)	15,000	0	0	0	0
Pagseserbisyo (Giving service)	2,819	0	0	0	0
Pagsisilbi (Serving)	4,310	0	0	0	0
Bayanihan (Cooperative effort)	60,000	0	0	0	0
Total	359,898	295	171,446	3,766	73

The study used the YouTube API to extract comments from YouTube videos. The main challenge is that comments can only be extracted one video at a time due to the need for each video's unique identification (ID) when running the Python code. To streamline the process, we listed all links to obtain their video IDs. In this way, the video ID can be copied from the list and pasted into the script before running the code. The running time took a few seconds to five minutes, depending on the number of comments. Extracted data include usernames, content/comments, the timestamp of comments, and the number of likes and replies.

For extracting tweets, the study did not use the Twitter API since it cannot scrape tweets posted more than three weeks prior (Hernandez-Suarez et al. 2018). To avoid

such a restriction, we used the snscrape Python library, which allowed the retrieval of tweets from 2006. This scraping tool is for scraping information, such as user profiles, hashtags, searches, threads, and posts for social networking services (Desai 2021). We imported the snscrape library and created a variable for the keyword/s (see table 2), including the time of the tweets to be scraped. We also set a limit for the number of tweets, made a for loop22 to retrieve the tweets based on the keywords, and listed the data (e.g., username, tweet content, date and time), and created a data frame to be saved as a CSV (comma-separated value) file.

Table 2. Total Number of Scraped Comments from YouTube Videos of FPJ's *Ang Probinsyano* and Subreddit Philippines Online Forums

Data Source	Number of Scraped Comments
YouTube	54,749
r/Philippines subreddit	51,500

We also used two methods to scrape online fora. These included: an API to scrape the subreddit r/Philippines; and Selenium for the other online fora (i.e., Pinoy Exchange, Philippines Expat Forums, Expat.com, and Skyscraper City). Scraping comments from the subreddit is the same as the process for YouTube. One must create an account and create a Reddit application to obtain an API key (or the Reddit ID) and a secret key. We set a limit of 60,000 posts and replies due to our uncertainty about the exact number of posts and replies in the r/Philippines subreddit community.

Moreover, the Selenium Python library was used to scrape the four other fora: Pinoy Exchange, Philippines Expat Forums, Expat.com, and Skyscraper City. Before performing the three phases of web scraping (see figure 3), we first created an account for each forum. Then, we used the keywords or search words (as shown in table 1) to narrow down the online discussions related to *pakikipagkapwa*. After identifying the threads of these discussions, we performed web analysis to understand the structure of each website. Selenium made the process convenient, given the dynamic nature of online fora. Then, we used the following libraries: WebDriver from Selenium for browser automation, CSV for saving data into a file, and the time library for setting up a waiting period while the browser was loading. After importing the libraries, we opened the Chrome browser and located all corresponding elements in the HTML code of the website using locators (i.e., XPath), such as class name, id, tag name, link text, or CSS selector. We then extracted the necessary data, including the username, contents, replies to posts, and date posted. After that, we saved them as a CSV file. For websites with multiple pages, we created a for loop and set the range, depending on the total number of pages of a website.

Text Analysis

The collected data were subjected to text analysis. Each comment, post, or reply was treated as one document (the unit of analysis). We decided to conduct a separate text analysis for each data source. This approach determined which website most frequently discusses or manifests *pakikipagkapwa*, and further explored the concept across different platforms. For the analytical technique, we applied a topic modeling technique called the latent Dirichlet allocation model (LDA), which automatically extracts topics/themes from a corpus of texts. The LDA assumes that each document is composed of several topics, in which each one is a distribution of words. The number of topics, denoted as k, is set in advance but can also be determined empirically. In this study, we set the number to 100 for every run of LDA. Gibbs sampling was also used to estimate the topic proportion per document and the word distribution per topic.

Results and Discussion

After applying the procedures for web scraping and using the API, we managed to scrape a total of 54,749 comments from FPJ's *Ang Probinsyano* videos on YouTube, 51,500 posts and comments from the r/Philippines subreddit, and 359,898 tweets from X. From the other forums, we extracted a total of 295 comments from Philippine Expats Forum, 171,446 comments from Skyscraper City, 3,766 comments from Pinoy Exchange, and 73 comments from Expat.com. In total, 641,727 texts were collected from seven sources.

Before applying text analytics techniques, we first preprocessed the text. Preprocessing typically involves text cleaning, through which it is transformed into a form that is suitable for the application of analytical techniques. During text cleaning, we converted all characters to lowercase. Next, we removed English and Tagalog stop words since some texts were in English and others were in Tagalog. We used precompiled lists of stop words for both languages. We also removed punctuation marks, emojis, and numbers. Then, we made a new set of stop words specific to this context. The processed corpus was represented as a document-by-term matrix (DTM), in which the rows are the documents, and the columns are the terms. The entries of the DTM are the raw frequencies of words found in each document. To reduce the size of the dictionary and consequently the dimensionality of the DTM, we deleted words (or terms) with a frequency less than three. No further weighting of the terms was applied; we used raw frequencies as input. Although we removed some words, those that were semantically relevant were retained, so there was no major alteration of meaning. The same steps have also been applied in other studies (Kobayashi et al. 2018).

We ran separate topic models for each source. Each one was labeled as a "topic" and was categorized as to whether it pertained to *pakikipagkapwa* or not. In this study, six experts independently performed labeling. Afterwards, we computed the reliability of their labeling using Fleiss' kappa, which measures the reliability of agreement between a fixed number of raters. The possible values of this measure range between 0 and 1 (inclusive), where values close to zero indicate poor agreement and values close to 1 signify almost perfect agreement. Table 3 presents the computed Fleiss' kappa for each source. There is a slight to fair agreement among the six experts.

Table 3. Computed Fleiss' Kappas for the Labeling of Topics for Each Source

Source	Fleiss Kappa
PinoyExchange	0.15
Philex	0.30
X (formerly Twitter)	0.20
Skyscraper	0.09
Reddit	0.19

The final labels were determined by obtaining the majority label of each topic. In case of a tie, the experts discussed among themselves to come up with a different label. Topics that have no relevance to *pakikipagkapwa* were deleted. In summary, there are 20 topics from PinoyExchange that are relevant to *pakikipagkapwa*. From other sources, there are 6 topics from Philex, 73 topics from X, 3 topics from Skyscraper, and 3 topics from Reddit. In total, there are 105 topics labeled positively. Each of them is further categorized according to what aspect or dimension of *pakikipagkapwa* is expressed. Table 4 shows some topics related to *pakikipagkapwa*.

Table 4. Example of Topics and Labels from Selected Online Sources

Source	Topic (With Top 15 Terms)	Label
PinoyExchange	Pinoy, Philippines, respect, worked, Pinas, you're, referred, friends, *pulubi* (beggar), choice, manager, find, subordinates, house, govt (government)	Social Interest and Concern for Community
PinoyExchange	*alam* (known), *kilala* (known), *kità* (visible), personal, *talaga* (really), *ano* (what), *totoo* (true), *diyan* (there), waste, *pati* (also), *mahal* (precious/beloved), *pagmamahal* (love), *raw* (she/he/they said), *madalas* (often)	Better Social Connections
PinoyExchange	God, love, things, life, kind, pray, hope, you're, good, religious, power, mass, long, posted, don't	Religious Priming

Source	Topic (With Top 15 Terms)	Label
PinoyExchange	parents, family, money, don't feel, you're, mom, job, support, college, make, relatives, pay, child, lot	Family-Centeredness
X (formerly Twitter)	*pagbibigay* (the act of giving), mayor, Sotto (Pasig City Mayor Victor Ma. Regis "Vico" Nubla Sotto), *patuloy* (continuing/continuance), Vico (Mayor Sotto), Maria, inang (mother), *taga-* (comes from), Isko (Manila Mayor Francisco "Isko" Moreno Domagoso), Tito (Senator Vicente "Tito" Castelo Sotto III), *Maynila* (Manila), congratulations, *mabuhay* ("live," a cheer analogous to Spanish *viva*), Manila, *punô* (full)	Governance
X (formerly Twitter)	*patuloy* (continuing/continuance), *pandemya* (pandemic), *gitna* (middle), *kabila* ("other side"), *iba* (different/other), *pagsunod* (the act of following), *ngunit* (however), *bansa* (nation), *krisis* (crisis), *balita* (news) *kabayanan* (township), DZRH Nationwide (AM radio station DZRH, 666 kHz), *aral* (lesson), *dulot* (to cause), *karunungan* (knowledge)	Crisis and Technology

We did not extract any meaningful topics related to *pakikipagkapwa* from the television series *Ang Probinsyano*. Most comments were either about the actors or anticipations for the next episode. Hence, the results excluded texts from the show.

An Emergent Model of Pakikipagkapwa

Our results revealed that the various topics generated through data mining can be grouped into different clusters, depending on their relationship with *pakikipagkapwa*. While the model builds on the "classic" Enriquez structure, it does not limit itself to values associated with *kapwa*. Instead, this framework also includes behavioral manifestations, mediums, enablers, and moderators of *pakikipagkapwa*.

First, the "Mediums" cluster consists of variables that serve as potential channels for *pakikipagkapwa*. These variables include technology, social media, mass media, nongovernmental organizations (NGOs), and government agencies. Second, the "Manifestations" cluster includes topics that depict how *pakikipagkapwa* is demonstrated. Third, the "Enablers" cluster consists of events that initiate the process of *pakikipagkapwa*, or actions that serve as catalysts for its occurrence. Enablers include the presence of suffering, crisis, and natural disasters. Lastly, the "Moderators" cluster comprises variables that affect the extent of *pakikipagkapwa* and/or influence its nature. These variables include families, religious organizations, educational institutions, and celebrities or famous personalities in the Philippines.

Figure 4. The emergent model of *pakikipagkapwa* from social media data



Validating the Emergent Model of Pakikipagkapwa

To validate the findings of this study, the authors conducted a survey (administered both online and offline). The questions were focused on affirming whether the variables mentioned in the model were indeed relevant to explain *pakikipagkapwa*. The survey questions were divided into four sections, corresponding to the four dimensions of the emergent model of *pakikipagkapwa* such as channels, moderators, manifestations, and enablers. An initial set of survey questions were formulated by the authors and pilot tested on several individuals. The results and feedback provided by the people from the pilot test were then examined and led to further reformulation of the questions. Hence, the questions underwent several rounds of revisions before being finalized.

A total of 329 people responded to the survey and were predominantly female, at 61.3 percent, while male respondents were pegged at 34.8 percent. The remaining portion came from non-binary genders and those who did not prefer to reveal their genders. Students emerged as the largest group of respondents, accounting for 56.1 percent of the total, followed by private sector employees at 18.2 percent and government employees at 12.2 percent. The rest of the respondents were teachers, self-employed individuals, and the unemployed. The dominance of students in the survey was also reflected in the age distribution, with the 18-30 age group being the most prevalent, followed by the 41-50 and 31-40 age groups at 14.8 percent and 13.1 percent respectively.

Facebook emerged as the most popular social media platform among all age groups, with a 98.5 percent usage rate. Other significant platforms included Instagram (75.9

percent), YouTube (69.8 percent), TikTok (61 percent), X (50.3 percent), and Reddit (25 percent).

Based on the emergent model, "Government," "Mass Media Companies with Humanitarian Units," and "Technology" are channels for *pakikipagkapwa*. These were all confirmed by the survey results. 256 (78.05 percent) of the respondents asserted that their interactions with government programs that help those affected by disasters were a way of showing concern for others, with 129 agreeing and very slightly fewer, 127, strongly agreeing (six respondents did not answer the question).

Meanwhile, 282 (85.96 percent) indicated that Mass Media Organizations with linked foundations, such as GMA 7 and ABS-CBN, were excellent channels for *pakikipagkapwa* in the form of providing assistance to victims of disasters, calamities, or crises. One hundred thirty-three respondents agreed with this sentiment, while 149 strongly agreed (eight respondents did not answer the question).

Respondents confirmed that they used technology in the form of online banking apps to help others and social media as a channel to connect with relatives and friends. 244 (74.39 percent) reported that they use GCash, PayMaya or other similar online banking apps to help others, with 118 strongly agreeing, and 126 agreeing (six respondents did not answer the question).

Moreover, a remarkable 91.77 percent of the respondents (301 of 328) said that they use Facebook and other social media to connect with relatives and friends, especially those who are far away— with a remarkable majority of 228 "strongly agreeing," and 73 indicating that they "agree" (seven respondents did not answer the question).

Asked if the way their parents or guardians raised them has influenced how they interact with others, 57.3 percent of the respondents strongly agreed and 33.3 percent agreed. As to whether the teachings of their religion affected the way they treat others, 38.3 percent agreed while 31.2 percent strongly agreed. Asked if what they learned and experienced in school during their youth has influenced how they interacted with others today, most of the respondents (48.9 percent) strongly agreed followed by those who agreed (40.2 percent).

Regarding the question to the respondents if what they see in famous personalities has influenced how they interact with others: 7.6 percent strongly agreed, 28.7 percent agreed, 28 percent neither agreed or disagreed, 22.6 percent disagreed, and 13.1 percent strongly disagreed. This outcome still shows that more than one-third of those polled have been influenced by said personalities in interacting socially with others. Through their social media platforms, notable personalities and celebrities communicate and promote their social advocacies which can positively affect the attitudes and behaviors of the public toward pro-social endeavors.

Despite the varied forms by which *pakikipagkapwa* is expressed, the core intent driving these manifestations is that of conveying care and concern for others. Statements pertaining to behaviors in specific contexts were given to participants to rate in terms of their agreement. Among these, it was the statement related to behaviors online during disaster situations ("Actions like tweeting or sharing posts about disasters and encouraging others to pray and donate relief goods demonstrate solidarity") that most participants (81 percent) agreed with. Only 6.5 percent disagreed with the statement while the rest neither agreed nor disagreed.

Survey results confirmed that "witnessing the suffering" of others enables the practice of *pakikipagkapwa* whether or not the person concerned is a relative although they are more likely to practice *pakikipagkapwa* if those people who are suffering are their family members and friends. Asked if they would assist any family member who needs help, a large majority or 91 percent of the respondents agreed, with 56 percent strongly agreeing. Moreover, asked if they would provide assistance to their relatives or friends who would experience hardship due to a disaster, calamity, or crisis, an overwhelming 92 percent agreed, with 54 percent strongly agreeing. On the other hand, when asked if they would assist someone who is not a relative or acquaintance, 76 percent agreed, while only 26 percent strongly agreed. Also, asked if they would provide assistance to victims of disaster, calamity, or crisis who are suffering but are not their relatives or acquaintances, 86 percent agreed, with only 34 percent strongly agreeing.

The survey provided face validity that supports the analytics-based model of *pakikipagkapwa* that emerged from this study. The results established the four dimensions included in the *pakikipagkapwa* model i.e., channels, moderators, manifestations, and enablers. It showed that modern-day technology paved the way for the practice of *pakikipagkapwa*. The findings affirmed the influence of family, religion, education, and famous personalities in determining the level of *pakikipagkapwa*. The authors were able to prove that *pakikipagkapwa* has various manifestations as well as enablers which are all embedded in the Filipino culture. The survey confirms that data science methods can be a valuable tool for understanding intricate cultural phenomena.

Implications for Policy

By showing the various factors that moderate *pakikipagkapwa*, this study implies areas of action where policymakers and stakeholders can intervene to address discrimination. Although the study involved data science methods and focused on online interactions, these instances are nevertheless as "real" as "real-life" incidents of discrimination. As such, they can also be treated under, or subject to, general policymaking. Recognizing the need for intervention, this paper offers practical

recommendations to address discrimination while also examining existing legislations known for antidiscrimination provisions for any gaps they may have. The authors argue that antidiscrimination practices must be incorporated in the following areas: families, religion, schools, and online spaces.

Family Upbringing and Discrimination

While discrimination often arises from the unfavorable attitudes and actions of a group to which the target does not belong, it can also exist in primary groups, such as the family. Ideally, a family functions as a safe base for supporting development; however, older members may transmit prejudicial attitudes and model discriminatory behaviors to younger members. A systematic review of studies on adolescents found a significant family transmission process of ethnic prejudice (Zagrean et al. 2022, 12). Children as young as three years old acquire prejudicial attitudes from their parents, which may later manifest as discriminatory behaviors (Pirchio et al. 2018, 6).

One source of discrimination within Filipino families is the undervaluation of female members, particularly due to the greater weight given to decisions made by male members over those made by female members in certain matters. This can be seen in the Family Code of the Philippines (1987), particularly in its sections on the exercise of parental consent to marry (art. 14), on the administration and enjoyment of community property or conjugal partnership (arts. 96, 124), and on the exercise of parental authority over the person of common children and legal guardianship over their properties (arts. 211, 225).

Some family practices that also discriminate against women include: (1) the prevalence of girls aged 15 to 19 who are married, divorced, or widowed in an informal union; and (2) the negative attitude toward working mothers (Organisation for Economic Co-operation and Development, n.d.). In particular, the former reflects the notion that children are the property of their parents, especially in the case of child brides. Although Republic Act No. 11596 prohibits and criminalizes marriage when at least one of the partners is below 18 years of age, reports of early and forced marriages among children, particularly in Mindanao, still persist (United Nations Population Fund 2024).

Another type of discrimination is the nonacceptance and, at times, abuse experienced by LGBTQ (lesbian, gay, bisexual, transgender, and queer) family members. In a study funded by the United Nations Development Programme, it was found that at least one out of ten Filipino LGBT respondents had been a victim of violence and abuse, mostly committed at home by their parents (Agence France-Presse 2014).

Religious Affiliation and Discrimination

The majority of Filipinos identify themselves as Catholics. According to the Philippine Statistics Authority (PSA 2023), out of 108,667,043 people across the country in 2020, 78.8 percent are Roman Catholics; 6.4 percent are Muslims; 2.6 percent are Iglesia ni Cristo; 0.8 percent each are Seventh-day Adventists and Aglipay; 0.6 percent are Iglesia Filipina Independiente; 0.5 percent are members of Bible Baptist Church; 0.4 percent each are members of United Church of Christ in the Philippines, Jehovah's Witness, and Church of Christ; and other religious affiliations comprise the remaining percentage.

Colonization by the male-dominated Spanish Catholic Church led to the formation of biases that promoted discrimination against women and children. By using the Catholic image of the Virgin Mary and the Holy Family, women were made to believe that engaging in sexual activity prior to marriage considered them "impure." Meanwhile, children born out of wedlock were deemed outcasts (Lavides 2024, 124–26).

On the contrary, being part of a religious minority can also lead to discrimination. In a 2020 survey by International Alert, the majority of young Filipino Muslims living in Taguig and Baguio reported being discriminated against because of their religion (Saludes 2020). Some individuals found it difficult to get employed due to having Arab-sounding names; or even when hired, they were not allowed to wear hijab (International Alert 2018; Takumi 2017). This discrimination stems from the perception that Muslims are violent and untrustworthy, leading them to be labeled as terrorists (International Alert 2018; Saludes 2020).

Educational Training/Environment and Discrimination

Philippine schools nurture children to build a just and humane society. But, over time, these institutions have become loci of discriminatory practices. In his study on Filipino high school students, Datu (2018) reveals that everyday discrimination is connected to negative emotions and academic achievement. He found that experiencing everyday discrimination is related to increased negative emotions, which, in turn, negatively affect one's perceived academic achievement.

As previously noted, the LGBTQ community is vulnerable to discriminatory behaviors, particularly among students involving teachers and parents. Abulencia et al. (2023) explore the experiences and practices of different stakeholders in several Marikina schools. The results show that discriminatory behaviors included teachers inquiring about students' sexual orientation, inappropriate touching by volunteer staff, and a lack of parental support for their children's sexuality, all of which led to

emotional distress. The study concludes that gender issues, involving bullying and microaggressions, extend beyond the male-female binaries and stereotypes.

Meanwhile, Almeda and Ballesteros (2021) highlight other forms of discrimination in schools. Indigenous people face overtly discriminatory policies, including the use of English in schools, despite fluency varying between communities. Another issue is smart shaming, or anti-intellectualism, which leads to the invalidation of unconventional contributions and ideas. Similarly, students with mental health issues face discrimination due to the stigma surrounding mental health. They become reluctant to seek help, feel isolated, and face limited growth opportunities.

Influence of Famous Personalities and Discrimination

The parasocial interactions Filipinos form with famous personalities play a significant role in combating discrimination, especially in online contexts. Kobayashi et. al. (2024) note that online audiences develop parasocial relationships with figures from show business ("showbiz"), politics, and news documentaries. This observation echoes Centeno's (2016, 445) assertion that the parasocial relationship Filipinos develop with media personalities extends into areas like politics. Moreover, Filipinos who participate in celebrity events are not just passive spectators. Their interactions with celebrities influence how these events are perceived (Pertierra 2020). Kobayashi et al. (2024) also observe the difference between political figures and showbiz personalities. Political figures are often associated with specific services and programs, while showbiz personalities are linked to broader terms of *pakikipagkapwa*, such as *pakikisama* and *pakikiisa*. Nevertheless, both types of public figures are expected to use their platform to promote just and fair treatment, thereby countering discriminatory behaviors.

This confirmation of the influence of parasocial interactions affirms that development sector advocates—including communication units of the government, line agencies, and cabinet agencies involved in development communication—have good reasons to include celebrities in advocacy campaigns against discriminatory behaviors and policies. It also suggests that more research should be done regarding the various roles celebrities might play in facilitating participatory approaches to development communication, which should be nondiscriminatory by their nature. Furthermore, it suggests that various measurable aspects of *pakikipagkapwa* would be useful in formulating the terms of reference in contracts that celebrities sign when recruited for government advocacy campaigns, ensuring that their statements and behaviors do not promote discrimination.

Conclusion

The authors of this paper argue that pakikipagkapwa must be the core and anchor of all Philippine government policies that seek to address discriminatory practices in the country. Through the emergent model of *pakikipagkapwa* that resulted from this study, the government leaders and policymakers are presented with an intricate blueprint that facilitates an in-depth understanding of such cultural phenomena. The analytics-based framework of *pakikipagkapwa* also provides a roadmap that could guide the development of programs that are not only more inclusive and humane but also more reflective of the Filipino worldview.

References

Abulencia, Arthur S., et al. 2023. "A Qualitative Cross-Sectional Study on Gender Issues in the Schools Division of Marikina: Basis for GAD Policy Recommendations." *The Normal Lights*, 17 (1), 31–54. https://po.pnuresearchportal.org/ejournal/index.php/normallights/article/download/1890/558.

Agence France-Presse. 2014. " Philippines Tolerates Gays but Abuses Continue, Says UN-Backed Study." *Inquirer.net*, 12 May 2014. https://newsinfo.inquirer.net/601549/ph-tolerates-gays-but-abuses-continue.

Almeda, Angel, and K Ballesteros. 2021. "Safety First: Discrimination at Philippine Schools and Work Places." *#MentalHealthPH*, 20 November 2021. https://mentalhealthph.org/11-20/.

Amnesty International Philippines. 2021. "Pass the Comprehensive Anti-Discrimination Bill." 24 September 2021. https://www.amnesty.org.ph/2020/06/pass-the-comprehensive-anti-discrimination-bill/.

Bernardo, Allan B. I., Jose Antonio R. Clemente, and Gregory Arief D. Liem. 2014. "Describing the Values of Filipino Adolescents: A Comparison with Pan-cultural Norms." *Journal of Tropical Psychology* 4 (e2). https://doi.org/10.1017/jtp.2014.2.

Canete, Jonathan James O., Ian Christopher Rocha, and Jian Dennis P. Dolosa. 2022. "The Filipino Community Pantries: A Manifestation of the Spirituality of 'Alay Kapwa' in the Time of the Pandemic." *Journal of Public Health* 44 (2): e295–96. https://doi.org/10.1093/pubmed/fdab210.

Centeno, Dave De Guzman. 2016. "Parasociality and Habitus in Celebrity Consumption and Political Culture: A Philippine Case Study on JSTOR." *Asian Journal of Social Science* 44 (4–5): 441–84. https://www.jstor.org/stable/43954178.

Clemente, Jose Antonio, Delia Belleza, Angela Yu, Effie Vinia Diane Catibog, Goyena Solis, and Jason Laguerta. 2008. "Revisiting the Kapwa Theory: Applying Alternative Methodologies and Gaining New Insights." *Philippine Journal of Psychology* 41 (2): 1–32. https://www.researchgate.net/publication/326589382_Revisiting_the_Kapwa_Theory_Applying_Alternative_Methodologies_and_Gaining_New_Insights.

Datu, Jesus Alfonso D. 2018. "Everyday Discrimination, Negative Emotions, and Academic Achievement in Filipino Secondary School Students: Cross-sectional and Cross-lagged Panel Investigations." *Journal of School Psychology* 68: 195–205. https://doi.org/10.1016/j.jsp.2018.04.001.

Dayrit, Julienne Celina, and Rowalt Alibudbud. 2022. "The Unbearable Struggle for Beauty: Physical Appearance Perfectionism, Mental Health, and Discrimination Among Heterosexual Cisgender and Sexually Diverse Youth in the Philippines." *Journal of LGBT Youth* 20 (1): 179–97. https://doi.org/10.1080/19361653.2022.2145404.

Desai, Rashi. 2021. "How to Scrape Millions of Tweets using snscrape [WWW Document]." *DataSeries*, 21 June 2021. https://medium.com/dataseries/how-to-scrape-millions-of-tweets-using-snscrape-195ee3594721.

Enriquez, Virgilio G. 1992. *From Colonial to Liberation Psychology: The Philippine Experience*. University of the Philippines Press.

Enriquez, Virgilio G. 2018. "Kapwa: A Core Concept in Filipino Social Psychology." In *Handbook of Filipino Psychology* (Vol. 1: Perspectives and Methodology), edited by Rogelia Pe-Pua, 278–92. University of the Philippines Press.

Handog, Marj Casal. 2021. "How Bad Is Ageism in PH? Survey Says 1 in 2 Women Refuse to Say Age." *Rappler*, 13 July 2021. https://www.rappler.com/brandrap/health-beauty-and-wellness/reasons-women-do-not-like-talking-about-age/.

Hernandez-Suarez, A., G. Sanchez-Perez, K. Toscano-Medina, V. Martinez-Hernandez, V. Sanchez, and H. Perez-Meana. 2018. "A Web Scraping Methodology for Bypassing Twitter API Restrictions." *arXiv* 1803.09875. https://doi.org/10.48550/arXiv.1803.09875.

Human Rights Watch. 2017. *Just Let Us Be: Discrimination Against LGBT Students in the Philippines*. Human Rights Watch. https://www.hrw.org/sites/default/files/report_pdf/philippineslgbt0617_web.pdf.

Indigenous Peoples' Rights Act of 1997. RA No. 8371 (29 October 1997). https://www.officialgazette.gov.ph/1997/10/29/republic-act-no-8371/.

International Alert. 2018. "Muslim Youth and Women Break Fast Together to Fight Discrimination." June 2018. https://www.international-alert.org/stories/muslim-youth-and-women-break-fast-together-fight-discrimination-philippines/.

Khan, Seema, Emilie Combaz, and Erica Fraser McAslan 2015. *Social Exclusion: Topic Guide*. Rev. ed. Governance and Social Development Resource Centre (GSDRC), University of Birmingham. https://gsdrc.org/wp-content/uploads/2015/08/SocialExclusion.pdf.

Kobayashi, Vladimer, Maria Margarita Lavides, Cristabel Tiangco, Purisima Panlilio, Remi de Loeon, and Angela Carreon. 2024. *Understanding Pakikipagkapwa Through Analytics: Exploring Filipino Relational Communication in Online Spaces*. University of the Philippines Center for Integrative and Development Studies.

Kobayashi, Vladimer B., Stefan T. Mol, Hannah A. Berkers, Gábor Kismihók, and Deanne N. Den Hartog. 2018. "Text Mining in Organizational Research." *Organizational Research Methods* 21 (3): 733–65. https://doi.org/10.1177/1094428117722619.

Krotov, Vlad, and Leiser Silva. 2018. "Legality and Ethics of Web Scraping Abstract." *Communications of the Association for Information Systems* 47: 555–79. https://doi.org/10.17705/1CAIS.04724.

Krotov, Vlad, and Matthew F. Tennyson. 2018. "Scraping Financial Data from the Web Using the R Language." *Journal of Emerging Technologies in Accounting* 15 (1): 169–81. https://doi.org/10.2308/jeta-52063.

Labor, Paul Danielle P., and Maria Cecilia C. Gastardo-Conaco. 2021. "Viewing Your Kapwa : Elaboration of a Social-Relational Construct through Language." *Philippine Social Science Journal* 4 (4): 10–19. https://doi.org/10.52006/main.v4i4.418.

Lavides, Maria Margarita R. 2024. "Discrimination Experienced by Biracial Children in Angeles City, Philippines." *Khazanah Sosial* 6 (1): 120–31. https://doi.org/10.15575/ks.v6i1.33776.

Luo, Siyang, Hang Yuan, Wang Yin, and Michael Harris Bond. 2024. "Culturomics: Taking the Cross-Scale, Interdisciplinary Science of Culture Into the Next Decade." *Neuroscience & Biobehavioral Reviews* 167 (December): 105942. https://doi.org/10.1016/j.neubiorev.2024.105942.

Mendoza, S. Lily, and Jim Perkinson. 2003. "Filipino 'Kapwa' in Global Dialogue: A Different Politics of Being-With the 'Other.'" *Intercultural Communication Studies*

12 (4): 277–93. https://www.researchgate.net/publication/242488077_Filipino_Kapwa_in_Global_Dialogue_A_Different_Politics_of_Being-With_the_Other.

Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, Joseph P. Pickett, et al. 2011. "Quantitative Analysis of Culture Using Millions of Digitized Books." Science 331 (6014): 176–82. https://doi.org/10.1126/science.1199644.

Mrakovcic, Emily. 2024. "HIV in the Philippines: A Persisting Public Health Crisis Closely Tied to Social Stigma." *Brown Undergraduate Journal of Public Health* 3: 61–69. https://bpb-us-w2.wpmucdn.com/sites.brown.edu/dist/e/218/files/2024/04/BUJPH-2024-bda8dd60a1079393.pdf.

Oyserman, Daphna. 2015. "Values, Psychology of." In *International Encyclopedia of the Social & Behavioral Sciences*, 2nd ed., vol. 25, edited by James D. Wright, 36–40. Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.24030-0.

Pertierra, Anna Christina. 2020. "Philippines Beyond Clichés Series 1 #1: Celebrity Culture." *New Mandala*, 26 March 2020. https://www.newmandala.org/philippines-beyond-cliches-series-2-1-celebrity-culture.

Pirchio, Sabine, Ylenia Passiatore, Angelo Panno, Fridanna Maricchiolo, and Giuseppe Carrus. 2018. "A Chip Off the Old Block: Parents' Subtle Ethnic Prejudice Predicts Children's Implicit Prejudice." *Frontiers in Psychology* 9 (110). https://doi.org/10.3389/fpsyg.2018.00110.

PSA [Philippine Statistics Authority]. 2023. "Religious Affiliation in the Philippines (2020 Census of Population and Housing)." 22 February 2023. https://psa.gov.ph/content/religious-affiliation-philippines-2020-census-population-and-housing.

Saludes, Mark. 2020. "Young Filipino Muslims Feel Discriminated Against because of Religion." *Licas News*, 26 June 2020. https://philippines.licas.news/2020/06/26/survey-young-filipino-muslims-feel-discriminated-against-because-of-religion/.

Singrodia, Vidhi, Anirban Mitra, and Subrata Paul. 2019. "A Review on Web Scrapping and its Applications." *2019 International Conference on Computer Communication and Informatics (ICCCI)*, Coimbatore, India, 23–25 January 2019. https://www.researchgate.net/publication/335577015_A_Review_on_Web_Scrapping_and_its_Applications.

Takumi, Rie. 2017. "Muslims Experience Workplace, Hiring Discrimination." *GMA News Online*, 2 August 2017. https://www.gmanetwork.com/news/topstories/nation/620441/muslims-experience-workplace-hiring-discrimination/story/.

Williams, David R., et al. 2019. "Understanding How Discrimination Can Affect Health." *Health Services Research* 54 (S2): 1374–88. https://doi.org/10.1111/1475-6773.13222.

Yacat, Jay A. 2017. "Walang Pakisama o Walang Kapwa-Tao?: Isang Sikolohikal na Pagsusuri sa Tindi ng Paglabag at Ugnayan sa Relasyong Panlipunan." *DIWA E-Journal* 5: 1–22. https://www.pssp.org.ph/diwa/wp-content/uploads/2017/11/1-Artikulo-Yacat.pdf.

Zagrean, Ioana, Daniela Barni, Claudia Russo, and Francesca Danioni. 2022. "The Family Transmission of Ethnic Prejudice: A Systematic Review of Research Articles with Adolescents." *Social Sciences* 11 (6): art. 236. https://doi.org/10.3390/socsci11060236.

Zhao, Bo. 2017. "Web Scraping." In *Encyclopedia of Big Data*, edited by Laurie A. Schintler and Connie L. McNeely, 1–3. Springer.

Python Codes

Pistocop. 2022 "subreddit-comments-dl." GitHub, last updated 16 February 2022. https://github.com/pistocop/subreddit-comments-dl.

Raka-Raka. 2021. "download-youtube-comments-with-python." GitHub, 13 February 2021. https://github.com/Raka-Raka/download-youtube-comments-with-python.

Does the ASEAN Smart Cities Network Promote Climate Resilience?

Louie Ayade[1], Charlyn Capulong[2], Martha Joy Abing[3], Maria Rizalia Teves[4], and Gerrich Aldin Babanto[5]

1 Department of Economics, Mindanao State University-Iligan Institute of Technology
louie.ayade@g.msuiit.edu.ph

2 Department of Economics, Mindanao State University-Iligan Institute of Technology
charlyn.maybituin@g.msuiit.edu.ph

3 Department of Economics, Mindanao State University-Iligan Institute of Technology
College of Economics, Business, and Accountancy (CEBA), MSU-IIT
marthajoy.abing@g.msuiit.edu.ph

4 Department of Economics, Mindanao State University-Iligan Institute of Technology
marizalia.teves@g.msuiit.edu.ph

5 Department of Economics, Mindanao State University-Iligan Institute of Technology
gerrich.babanto@g.msuiit.edu.ph



Manila Skyline, Philippines

fazon1, iStock, https://www.istockphoto.com/photo/manila-skyline-philippines-gm649114716-118527161?phrase=manila%20skyline&searchscope=image%2Cfilm

Abstract

As Southeast Asia becomes increasingly vulnerable to climate change, smart city initiatives like the ASEAN Smart Cities Network (ASCN) have emerged as strategic efforts to integrate climate mitigation and adaptation into urban development. However, the actual impact of these initiatives on climate resilience remains underexplored, particularly in the Philippine context. This study investigates the impact of ASCN on climate resilience by employing a Difference-in-Differences approach, using panel data from 2014 to 2023 across 30 Highly Urbanized Cities in the Philippines. The results reveal the following: (1) ASCN significantly reduces climate resilience, suggesting insufficient integration of smart city strategies into climate resilience planning, (2) mechanism analysis further shows that ASCN participation significantly diminishes information technology (IT) capacity, although IT capacity positively influences climate resilience, it fails to mediate the effect, and (3) heterogeneity analysis indicates that cities with stronger capacity of school services experience less adverse impacts, implying that educated population mitigates the negative effects of ASCN. Overall, the findings suggest that while ASCN promotes technological innovation, its current implementation may fall short in enhancing local climate resilience. Improved coordination between national and local governments is necessary to ensure that smart city initiatives effectively support climate resilience and sustainable urban development.

Keywords: smart city, ASEAN Smart Cities Network (ASCN), climate resilience, urban planning, policy effect

Introduction

Southeast Asia is one of the most diverse regions in the world, with variations in culture, politics, economics, and social views shaping its complexity. However, evidence suggests that the region may share more common traits in the future. Kong and Woods (2021) identified two "megatrends" that could serve as points of alignment: urbanization and technological advancement. These trends are interconnected, with urbanization presenting development challenges and technology offering potential solutions. In response, the Association of Southeast Asian Nations (ASEAN) launched the ASEAN Smart Cities Network (ASCN) during the 32nd ASEAN Summit in Singapore on 28 April 2018 (ASEAN 2026). Philippine cities such as Manila, Cebu, and Davao were among the first-batch members, all experiencing rapid urbanization and population growth.

While rapid urbanization has supported economic development, it has also produced a range of urban problems, including excessive carbon emissions, population growth, traffic congestion, and environmental pollution (Dong et al. 2019). One of the most pressing challenges is climate change, which further constrains economic development (Gera 2018), as rapidly urbanizing areas face significant difficulties in terms of sensitivity and capacity to manage its impacts (Garschagen and Romero-Lankao 2015). The smart city model offers a promising approach to addressing these challenges, providing a new pathway for urban development that supports climate resilience, provided fiscal resources are effectively managed (Jiang et al. 2021). Previous studies have examined the effects of smart cities on innovation, employment, economic growth, and air quality (Lv and Gao 2023; Guo and Zhong 2022; Dharmarathne et al. 2024; Liu and Harrison 2024), yet limited research has investigated the integration of environmental protection with economic growth, with most existing literature focusing on the Global North and China (Jiang et al. 2021). Despite positive findings in these contexts, Jamilurahman et al. (2024) highlighted that implementing climate change adaptation through smart city initiatives in developing countries faces significant governance, policy, and technological barriers, which resonates with the Philippine context where stronger integration of environmental and economic strategies is needed (Alvarez et al. 2020; Katigbak and Credo 2022). Moreover, the Asian Development Bank (2009) identified Southeast Asia as one of the "world's most vulnerable" regions to climate change. Therefore, as urban areas in developing nations become vulnerable to the impacts of climate change, it is essential to investigate the role of smart cities in enhancing climate resilience in these regions.

Treating ASCN as a quasi-natural experiment, this study uses the two-way fixed effect difference-in-differences (TWFE-DID) approach to measure the impact of smart cities on climate resilience. The first batch of ASCN (smart cities) in the Philippines

is selected as the treatment group for analysis. To understand the mechanism, this study examines how information technology mediates the impact of smart cities on climate resilience through a mediation-effect model. Finally, the study explores the heterogeneous effects of smart cities on climate resilience, focusing on the educated population.

It is important to note that smart cities pursue various objectives, but the ASCN primarily emphasizes improvement of development plans to pursue climate resilience and to address regional climate vulnerabilities. This paper contributes to the literature by examining the impact of smart cities on climate resilience, providing lessons particularly in Southeast Asia and developing countries in Asia, an area previously underexplored. The paper is structured as follows. Section 2 presents the policy background and research hypotheses. Section 3 details the methodology, variables, and data. Section 4 discusses empirical results, robustness checks, mechanism analysis, and further covers heterogeneity analysis. Section 5 concludes with policy implications and recommendations.

Background and Research Hypothesis

This section first presents the background of the ASEAN Smart Cities Network (ASCN). It then discusses the impact of smart cities on climate resilience planning and proposes Hypothesis 1. Finally, it analyzes the mechanism through which ASCN influences urban planning for climate change and presents Hypothesis 2.

Policy Background of ASCN

The ASEAN Smart Cities Network (ASCN) was launched in April 2018 as part of ASEAN's broader integration agenda. It was initiated to address the challenges of rapid urbanization in Southeast Asia by fostering collaboration among cities to develop sustainable and smart urban environments. ASCN includes Davao, Cebu, and Manila. Historically, the region has faced significant environmental challenges, including climate change impacts such as rising sea levels and extreme weather events. ASCN aims to enhance climate resilience planning by promoting sustainable infrastructure and innovative urban solutions that help cities adapt to these challenges.

ASCN aims to strengthen climate resilience by promoting sustainable urban development, fostering economic growth, and improving living standards for its residents. The network encourages the adoption of smart technologies to enhance energy efficiency, transportation systems, and public services, all of which contribute to making cities more resilient to climate change. ASCN also emphasizes city diplomacy, where local governments play a crucial role in regional integration efforts.

By integrating these aspects, ASCN supports ASEAN member states in achieving sustainable development while addressing environmental challenges. It started with 26 pilot cities, expanded to 31 in 2024, and 34 in 2025, with member cities sharing knowledge and best practices for climate resilience.

In the case of the Philippines, the ASEAN Smart Cities Network (ASCN) primarily supports improvements in urban security, transportation systems, smart health and education, and the modernization of command centers for real-time monitoring using smart technologies. These focus areas are intended to help cities adapt to and reduce the effects of climate change. Overall, the ASCN aims to support existing national development plans or assist in creating new ones that promote climate-resilient urban planning, while upholding human rights and fundamental freedoms as outlined in the ASEAN Charter (ASEAN 2018).

Effect of Smart Cities on Climate Resilience

Smart cities have become a global trend, evolving from improving infrastructure and service delivery through sensors and technology to using data analytics, machine learning, and artificial intelligence for city-wide decision-making (Yoon et al. 2020). These advancements help address climate change issues and enhance predictive models for better solutions (Neo et al. 2023), especially in Asia, where many large cities face unavoidable impacts of climate change on urban systems (Tyler and Moench 2012). By analyzing large datasets, municipalities can design urban spaces that reduce carbon footprints, optimize energy use, and improve resilience to climate impacts (Bokolo 2023). In general, smart cities address climate change through mitigation (reducing emissions) and adaptation (preparing for unavoidable consequences) (Rolnick et al. 2022). In mitigation, smart cities focus on reducing greenhouse gas emissions and transitioning to more sustainable systems to lower environmental impact while improving quality of life (Tyagi et al. 2024). They use data-driven mobility solutions, such as real-time traffic monitoring, optimized routes, and smart traffic management systems integrated with public transit, to reduce vehicle emissions by easing congestion and cutting idle time (Abdelati 2024), which aligns with the smart mobility goals of Davao, Cebu, and Manila under ASCN. Smart cities also reduce carbon footprints by optimizing energy supply chains and minimizing fossil fuel dependence (Zadeh and Garay-Rondero 2023). In addition, they leverage vertical green structures as a mitigation strategy, promoting urban sustainability through technological integration and community engagement (Ogut et al. 2024). For adaptation, smart cities equip urban areas to cope with the unavoidable consequences of climate change by leveraging innovative technologies and advanced urban planning strategies. They combine smart technologies with robust planning and community engagement to address geographic and socioeconomic challenges (Ramirez Lopez and Grijalba Castro 2020), as seen in cities like Sydney, Melbourne, and Brisbane

(Liu and Harrison 2024). Smart mobility solutions also play a role by improving air quality and strengthening cities' ability to adapt, supporting efficient transportation networks and enabling better responses to challenges like flooding and heatwaves through real-time information and adaptive traffic management (Dharmarathne et al. 2024). There is, however, a contrast in developing countries because even though these initiatives use data-driven strategies to adapt to climate change, challenges still remain due to weak governance, limited technology, and the need for better coordination between national and local governance levels to implement effective strategies that maximize the potential of smart cities (Jamilurahman et al. 2024; Hurlimann et al. 2021).

The impact of smart cities on climate resilience should be reflected in their comprehensive development plans, as it is supposed to improve a locality's mitigation, adaptation, and recovery efforts. This suggests that smart cities are expected to be more climate resilient than other highly urbanized cities. Based on this, the following hypothesis is proposed:

> Hypothesis 1: The ASEAN Smart Cities Network promotes climate resilience in the Philippines.

Information Technology as Mechanism

Information technology (IT) plays a crucial role in enhancing the impact of smart cities on climate resilience by using information and communication technology (ICT) tools to improve service efficiency and residents' quality of life (Argyroudis et al. 2022; Davila Delgado et al. 2017). To ensure accurate, digital, and flexible management, smart cities must gather and analyze data from various aspects (Guo et al. 2022). Smart cities can leverage open data services and citizen engagement to achieve climate resilience, with IT serving as a crucial enabler for these processes (Ahdekivi et al. 2023). For instance, digital twin technology, 3D city modeling, and early warning systems are being explored for their potential in managing climate-related risks in urban areas, particularly in coastal regions (Riaz et al. 2023), which is parallel to the command center upgrade of ASCN in the Philippines, as they are coastal cities. These technologies support the development of early warning systems that provide stakeholders with accurate and up-to-date information, helping them respond effectively to extreme climate events, while also using data analytics and sensors to monitor and manage urban infrastructure for better prediction and response to climate-related challenges (Riaz et al. 2023; Li 2024). Innovation was found to be a mechanism that influences environmental integration in economic development (Jiang et al. 2021). IT is also used as a mechanism for other smart city outcomes such as employment, which shows that it has positive impacts on better achieving the goal of smart cities (Lv and Gao 2023).

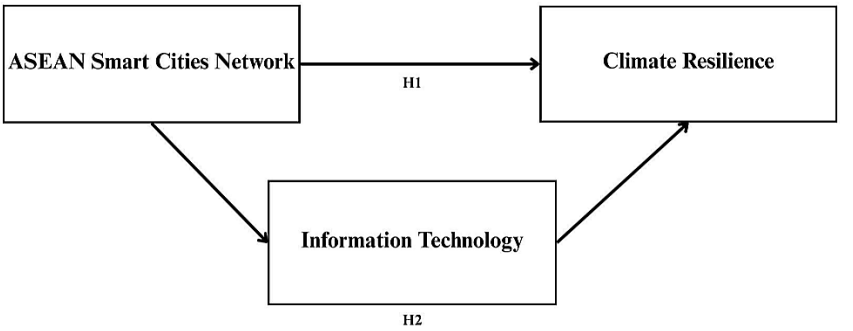
In summary, it drives innovation in information technology, further enhancing climate resilience. Therefore, the second hypothesis is suggested:

> Hypothesis 2: The ASEAN Smart Cities Network promotes climate resilience in the Philippines by facilitating information technology.

Conceptual Framework

Figure 1 presents the conceptual framework of this paper.

Figure 1. Conceptual Framework



Methodology, Variables, and Data

Data Collection and Sample Criteria

The ASEAN Smart Cities Network (ASCN) was introduced in 2018, with some Philippine highly urbanized cities (HUCs) included. The treatment group consists of HUCs in the ASCN, while the control group includes those HUCs not part of the network. This setup allows a comparison of climate resilience outcomes before and after ASCN participation. HUCs with missing data were removed, and the data were sourced from the Cities and Municipalities Competitiveness Index (CMCI) (Department of Trade and Industry [DTI] 2024). Therefore, we obtained panel data for 30 HUCs covering the years 2014 to 2023. Figure 2 shows the spatial distribution, while Table 1 shows the numerical distribution of treated and control cities.

Figure 2. Spatial Distribution of Sample HUCs

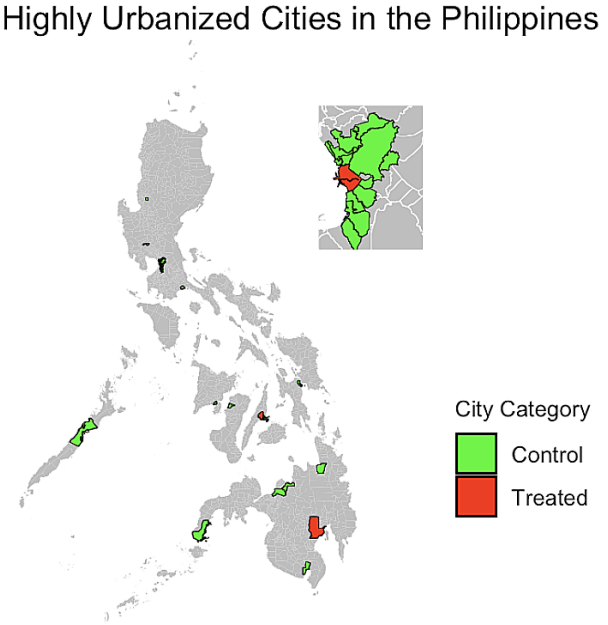


Table 1. Numerical Distribution of Sample Data

	Treated	Control	Total
Luzon	1	18	19
Visayas	1	5	6
Mindanao	1	4	5
Total	3	27	30

Model Setting

1. DID Model

In order to determine if the ASCN (smart city) has a causal effect on climate resilience, the two-way fixed effect (TWFE) difference-in-differences (DID) model will be implemented. According to Roth et al. (2023), if the panel is balanced, then the TWFE specifications yield easily interpretable results.

Equation 1

$$CR_{it} = \alpha_i + \phi_t + D_{it}\beta_1 + \beta_2 X_{it} + \varepsilon_{it}$$

In Equation 1, i represents the city, t represents the year, and CR represents climate resilience. D_{it} is the interaction term between the city and time dummy

variables; it is set to 1 if a city is part of ASCN in 2018 or later, and 0 otherwise. β_1 measures the impact of smart city on climate resilience. X_{it} represents the covariates in this study, such as local economic growth, productivity, education infrastructure, capacity of school services, and presence of an investment promotion unit. α_i denotes city fixed effects, ϕ_t denotes time fixed effects, and ε_{it} is the random disturbance term. If $\beta_1 > 0$, it indicates that smart city promotes climate resilience; if $\beta_1 < 0$, it suggests the opposite.

2. Mediating Effect Model

As discussed, this paper examines whether information technology is a mediating variable between smart city and climate resilience. Following Guo et al. (2022) and Lv and Gao (2023), the mediating effect models will be:

Equation 2

$$IT_{it} = \gamma_1 D_{it} + \gamma_2 X_{it} + \alpha_i + \phi_t + \varepsilon_{it}$$

Equation 3

$$CR_{it} = \varphi_1 D_{it} + \varphi_2 IT_{it} + \varphi_3 X_{it} + \alpha_i + \phi_t + \varepsilon_{it}$$

Where *IT* represents information technology. Equation 2 is initially employed to measure the impact of smart cities on information technology. Following that, equation 3 examines the combined effect of smart cities and IT, the mediating variable, on climate resilience. Then, the impact of smart cities is β_1 in equation 1, and the direct effect on IT is γ_1 in equation 2. If φ_2 is significant in equation 3, then there is a mediating effect.

Variables

The data of this study were sourced from Cities and Municipalities Competitiveness Index (CMCI), an annual ranking survey conducted to assess the competitiveness of cities and municipalities across the Philippines, from 2014 to 2023.

1. Independent Variable

The interaction term Post × Treat serves as the key independent variable in this study. Here, Treat equals 1 for HUCs that are part of the ASCN, and 0 otherwise. Post equals 1 for years in which the HUC is part of the ASCN treatment (2018 and above), and 0 for the pre-treatment period. The treatment group comprises HUCs that are part of the ASCN, while the control group consists of HUCs that are not part of the network.

2. Dependent Variable

Climate resilience (CR) is the dependent variable of this study that is from the Compliance to National Directive index. According to the CMCI, this reflects the prioritization of revising comprehensive land use plans by local chief executives to ensure urban planning that enhances climate resilience in response to the increasing impacts of climate change.

3. Covariates

Local economy growth, productivity, education infrastructure, capacity of school services, and presence of investment promotion unit are the covariates for this study.

4. Mediating Variable

Information Technology (IT) capacity reflects a locality's ICT readiness.

5. Indicators

The variables in this study are methodologically weighted by CMCI from the following indicators:

Table 2. Variable Indicators

Variable	Indicators
Climate Resilience	Presence and regular updating of the Comprehensive Development Plan, a key tool for urban planning. This is crucial in addressing the adverse impacts of climate change
Local Economy Growth	Gross sales of registered firms, Total capitalization of new businesses
Productivity	Gross sales of registered firms, Number of declared employees for business renewals
Education Infrastructure	Number of public and private secondary schools and classrooms, Number of public and private tertiary schools, Number of public and private technical vocational education and training
Capacity of School Services	Number of teachers, students, and graduates in public and private secondary and tertiary education, and technical vocational education and training
Presence of Investment Promotion Unit	Presence of the Local Investment Incentives Code, Presence of the equivalent of an Investment Promotions Unit (Physical Office), Presence of Staff manning the IPU, Presence of local executive order or ordinance that mandates the implementation of the LIIC or the setting up of an IPU
Information Technology Capacity	Number of cell sites, cable, internet, telephone companies, mobile service providers in the locality

Descriptive Statistics

The descriptive statistics of the variables are presented in table 3. For the period 2014–2023, the treatment and control groups exhibit nearly identical mean values for climate resilience, at 2.34 and 2.30, respectively. However, before the policy implementation, the treatment group recorded a higher mean of climate resilience (2.99) compared to 2.83 for the control group. After the policy implementation, the treatment group showed a slightly lower mean climate resilience of 1.91, compared to 1.95 for the control group. This preliminary observation suggests that the ASEAN Smart Cities Network may have failed to promote climate resilience.

Table 3. Descriptive Statistics

Variables	Year	Full Sample			Treated Cities			Control Cities		
		Obs	Mean	Std. dev	Obs	Mean	Std. dev	Obs	Mean	Std. dev
CR	2014–2023	300	2.31	0.79	30	2.34	0.74	270	2.30	0.80
	2014–2017	120	2.84	0.74	12	2.99	0.46	108	2.83	0.76
	2018–2023	180	1.95	0.60	18	1.91	0.16	162	1.95	0.61
eco	2014–2023	300	0.76	0.48	30	0.68	0.37	270	0.77	0.49
prod	2014–2023	300	0.41	0.64	30	0.63	0.75	270	0.38	0.63
inf	2014–2023	300	0.55	0.48	30	1.25	0.46	270	0.48	0.42
educ	2014–2023	300	0.72	0.46	30	0.92	0.53	270	0.70	0.45
pui	2014–2023	300	2.39	0.76	30	2.40	0.80	270	2.39	0.76

Empirical Analysis Results

Benchmark Regression Using DID

Table 4 presents the results of the baseline regression examining the impact of ASCN on climate resilience in Philippine highly urbanized cities. In both specifications, the coefficient for Dit, which captures the effect of being part of the ASEAN Smart Cities Network, was negative and statistically significant. Specifically, column 1 reported an estimated coefficient of −0.204 (significant at the 5 percent level), while column 2, which included covariates, showed a slightly larger negative effect of −0.285, significant at the 1 percent level. These findings are consistent with Alvarez et al. (2020) and Ferraris et al. (2020), who argued that smart city initiatives in developing countries are often less effective and require further improvement to achieve their intended goals. Similarly, Gera (2018) and Jones (2007) explained that the Philippine case has been constrained by persistent political dysfunction, limiting

cities' capacity to address pressing development challenges effectively. Therefore, the evidence does not support Hypothesis 1.

The interpretation of covariates in column 2 shows that the capacity of school services (*educ*) significantly promoted climate resilience, with an estimated coefficient of 0.162 at the 1 percent level. This suggests that higher student enrollment enhances resilience by equipping communities with knowledge on disaster risks and water management (Anderson 2012) and by fostering mitigation through sustainable behaviors and youth-led innovations (Borde et al. 2022). In contrast, education infrastructure (*inf*) did not significantly promote climate resilience, indicating a possible misallocation of funds. This finding is consistent with evidence of a dysfunctional and inefficient public procurement system in the Philippines (Jones 2007), which often prioritizes short-term political gains over sustainable development (Trinidad 2023), as well as with the country's weak education infrastructure and performance, underscoring the need for integration of education infrastructure into climate resilience strategies (Haw et al. 2021). Finally, economic growth (*eco*), productivity (*prod*), and the presence of an investment unit (*pui*) also did not significantly promote climate resilience, as economic priorities frequently outweigh the long-term requirements of adaptation and mitigation (Suprayitno et al. 2024).

Table 4. Results of Baseline Regression

	(1) Climate Resilience	(2) Climate Resilience
D_it	−0.204** (0.099)	−0.285*** (0.081)
eco		0.038 (0.066)
prod		0.032 (0.104)
inf		−0.279 (0.207)
educ		0.162*** (0.052)
pui		0.106 (0.087)
FE: City	YES	YES
FE: Year	YES	YES
Observations	300	300
Adj. R^2	0.691	0.704
Robust standard errors in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$		

PSM-DID Estimation Result

To address potential endogeneity arising from sample self-selection and to control for systematic differences between the treatment and control groups, this study applies the propensity score matching–difference-in-differences (PSM-DID) approach as a robustness check to equation 1. Propensity scores are estimated through a logit regression using the same covariates from equation 1: *econ*, *prod*, *inf*, *educ*, and *pui*. Following Stuart and Rubin (2008), the caliper is set at one-quarter of the standard deviation of the estimated logit scores. Nearest neighbor matching with two neighbors ($M = 2$) is then implemented, and the results of the balance test are reported in Table 5.

After matching, the standardized mean differences (SMD) of all covariates fall within the ±10 percent threshold, except for *educ*. However, the post-matching *p*-value for *educ* is not statistically significant, indicating that the covariate is adequately balanced. Thus, the balance test requirement is satisfied, confirming that PSM effectively reduces systematic disparities between treatment and control units.

Equation 1 is subsequently reestimated using the matched sample, and the results of the PSM-DID estimation are presented in table 6. The estimated coefficients remain statistically significant, suggesting that participation in the ASCN reduces climate resilience. These findings reinforce the baseline regression results and indicate that the main conclusion is robust even after accounting for potential endogeneity through the PSM-DID framework.

Table 5. Balancing test of PSM

Variables	Unmatched			Std. Mean Diff.	*T*-test *p*-value
	Matched	Treated	Control		
eco	U	0.678	0.767	−23.914	0.238
	M	0.733	0.711	5.917	0.869
prod	U	0.629	0.382	32.969	0.091
	M	0.646	0.602	5.228	0.840
inf	U	1.250	0.475	169.65	0.000
	M	1.182	1.160	4.876	0.720
educ	U	0.920	0.702	41.787	0.035
	M	0.884	1.044	−32.547	0.298
pui	U	2.397	2.388	1.191	0.951
	M	2.395	2.463	−7.692	0.779

Table 6. PSM-DID Estimation

	(1) Climate Resilience	(2) Climate Resilience
D_{it}	− 0.356** (0.160)	− 0.522*** (0.157)
eco		0.045 (0.121)
prod		0.014 (0.053)
inf		0.344** (0.149)
educ		0.480** (0.204)
pui		0.002 (0.036)
FE: City	YES	YES
FE: Year	YES	YES
Observations	53	53
Adj. R^2	0.722	0.775
Robust standard errors in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$		

Robustness Test

Event Study

It's important to verify that the treatment and control groups followed the same trend before the policy was introduced, before applying the DID model. If this assumption is not satisfied, the estimated policy effect may be biased. To ensure the model's reliability, this study modifies the independent variable in equation 1 and constructs the following equation:

Equation 4

$$CR_{it} = \sum_{s=-4, s\neq -1}^{s=5} \beta_s treat_i \times year_s + \mu_1 X_{it} + \alpha_i + \phi_t + \varepsilon_{it}$$

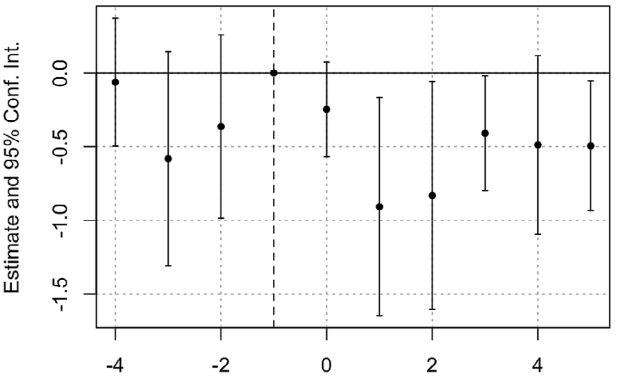
Where $treat_i$ shows whether city i is a pilot city, and $year_s$ for $s \in (-4, -2) \cup (0, 5)$ stands for dummy variables for the three years before 2018 and the five years after 2018, excluding $s = -1$. The coefficients $\beta_{-4}, \beta_{-3}, \ldots, \beta_{-5}$ (except β_{-1}) show the policy effect for the years 2014, 2015, . . . , 2023, except for the year 2017.

Table 7 and figure 3 are the results of the event study, with the latter plotting the result in 95 percent confidence interval. The estimations show that the interaction terms before 2018 are not significant, which supports the parallel-trend assumption. In the five post periods after treatment, the coefficient on climate resilience is negative throughout and statistically significant, except in *post_4*. The drop is largest immediately after ASCN implementation and then gradually shrinks. Therefore, there's a sharp, significant decline right after ASCN. This also signifies that the long-term effect of smart city might be more beneficial, even though the result can be short-term if there is enough city efficiency. This is evident in the historical shortcomings of urban development planning in the Philippines, where misalignment between outgoing and incoming city administrations, along with weak integration among local, regional, and national councils, often disrupts the continuity of previously signed policies (Gera 2018).

Table 7. Event Study

Climate Resilience			
pre_4	– 0.062 (0.212)	*post_1*	−0.907** (0.362)
pre_3		*post_2*	−0.831** (0.377)
pre_2		*post_3*	−0.410** (0.190)
2018		*post_4*	−0.488 (0.296)
		post_5	−0.494** (0.215)
FE: City	YES		
FE: Year	YES		
Covariates	YES		
Observations	300		
Adj. R^2	0.703		
Robust standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$			

Figure 3. Event Study



Placebo Test Through Fake Time

Following the method of Jiang et al. (2021), this study assigns artificial start dates for the ASCN and adjusts the panel years to reestimate Equation 1. If the coefficient of $D_{i\mathrm{t}}$ is not statistically significant, it would suggest that the negative impact observed in table 4 can be attributed to the ASCN; otherwise, the result would not be robust. Specifically, the years 2016 and 2017 are designated as placebo implementation years, with 2014–2023 and 2015–2022 serving as the corresponding time windows.

The placebo test results, reported in table 8, show that when the implementation year and policy duration are altered, the coefficient of $D_{i\mathrm{t}}$ becomes statistically insignificant. This finding suggests that the negative impact identified in table 4 is indeed attributable to ASCN and that the baseline results are robust.

Table 8. Placebo Test Fake Time

Climate Resilience								
Fake Year: Panel Year:	**2017 (2014-2023)**	**2017 (2014-2023)**	**2016 (2014-2023)**	**2016 (2014-2023)**	**2017 (2015-2022)**	**2017 (2015-2022)**	**2016 (2015-2022)**	**2016 (2015-2022)**
D_{it}	−0.097 (0.099)	−0.128 (0.093)	−0.164 (0.152)	−0.150 (0.128)	−0.014 (0.091)	0.030 (0.115)	−0.053 (0.163)	−0.122 (0.249)
eco		0.038 (0.067)		0.040 (0.068)		0.010 (0.055)		0.011 (0.056)
prod		0.031 (0.104)		0.028 (0.105)		−0.044 (0.098)		−0.051 (0.101)
inf		−0.272 (0.208)		−0.262 (0.211)		−0.018 (0.105)		−0.001 (0.116)
educ		0.141** (0.056)		0.140** (0.056)		0.065 (0.048)		0.068 (0.049)
pui		0.108 (0.088)		0.108 (0.088)		0.083 (0.087)		0.082 (0.087)

Climate Resilience								
Obs	300	300	300	300	240	240	240	240
Adj. R^2	0.689	0.702	0.690	0.702	0.706	0.705	0.706	0.705
Robust standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$								

Elimination of Confounding Cities

In 2016, the Department of Information and Communications Technology launched the Next Wave Cities program. Because the panel covers the period 2014–2023, cities included in this program may influence the analysis, as they generally exhibit lower vulnerability to natural disturbances. In addition, the special income class city may introduce bias due to its unique fiscal status. To eliminate potential interference from these cases, Equation 1 is reestimated after excluding highly urbanized control cities belonging to the Next Wave Cities program as well as the special income city. Column (1) of table 9 shows that, even after these exclusions, the coefficient of interest remains negative and statistically significant, confirming the robustness of the baseline results.

Spillover by Land Border

To address potential spillover effects, we conduct a robustness check by excluding control cities that share a land border with treated cities. Since the dependent variable is the level of climate resilience, it is reasonable to expect that nearby cities may be influenced by the policy initiatives, planning frameworks, or institutional reforms adopted in treated cities. Such spillovers could lead neighboring control cities to improve their planning capacity in parallel, thereby biasing the estimated treatment effect. By restricting the sample to nonadjacent control cities, we reduce this risk and ensure a cleaner counterfactual. If the results remain robust, this strengthens the validity of our findings. Column (2) of table 9 reports that after removing them, the result remains robust.

Spillover by Region

To further ensure the validity of the results, we eliminate control cities that belong to the same region as the treated cities. The rationale is that cities within the same region often share policy structures, planning institutions, and development programs. Consequently, the urban planning capacity of control cities in the same region may improve indirectly through regional spillovers, shared policy frameworks, or coordinated infrastructure investments triggered by the smart city initiative. Such region-level influences could blur the distinction between treated and control groups, leading to biased estimates of the treatment effect. By removing these regionally adjacent controls, we obtain a stricter counterfactual and test whether our results are

robust to potential intra-regional spillovers. Column (3) of table 9 reports that after removing them, the result remains robust.

Phase I Lag Control Variables

In the benchmark model, there is a possibility of reverse causality between the dependent variable and independent variables, as the level of climate resilience may itself shape local economic, social, or institutional conditions. To address this potential endogeneity and avoid simultaneity bias, all covariates are lagged by one period. This ensures that only predetermined values of the covariates are included, mitigating reverse causality concerns. The DID model is then reestimated with lagged covariates as a robustness check to verify the stability of the results. Column (4) of table 9 reports that after removing them, the result remains robust.

Table 9. Report of Additional Robustness

	(1) Confounding	(2) Land Border	(3) Region	(4) Lag Covariates
D_{it}	−0.215** (0.104)	−0.208** (0.089)	−0.286** (0.122)	−0.168** (0.067)
eco	−0.001 (0.070)	0.035 (0.094)	0.034 (0.120)	−0.088 (0.066)
prod	−0.040 (0.083)	0.054 (0.042)	0.019 (0.046)	0.044 (0.058)
inf	−0.021 (0.117)	−0.109 (0.134)	−0.126 (0.188)	−0.014 (0.134)
educ	0.106** (0.051)	0.116* (0.060)	0.107 (0.086)	0.031 (0.051)
pui	0.032 (0.067)	0.079 (0.069)	0.153 (0.134)	0.008 (0.046)
Obs	270	230	170	270
Adj. R^2	0.731	0.754	0.722	0.736
Robust standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$				

Mechanism Test

Table 10 presents the results of Equations 2 and 3. First, ASCN participation is found to significantly reduce IT capacity in the Philippines. Second, while D_{it} continues to have a significant negative effect on climate resilience, IT capacity itself does not exert a significant positive impact. These findings indicate that Hypothesis 2 is not supported, suggesting that ASCN's current implementation may divert

resources away from IT infrastructure. As a result, IT capacity does not function as a mediating channel, even though it is critical for strengthening command centers within the ASCN framework.

Prior studies support this interpretation. Ferraris et al. (2020) argued that limited human resources can undermine the effectiveness of smart city initiatives when residents lack the capacity to adopt technologies aimed at climate mitigation and adaptation. Similar challenges are evident in Valencia, where IT interventions failed to enhance early warning systems and thus did not improve climate resilience (Kurbalija 2025), and in Cauayan, where technology integration in water management remains insufficient (Alvarez et al. 2020). Meanwhile, the variable *educ* consistently shows a significant positive effect on both IT capacity and climate resilience, showing the role of educated population in enhancing adaptive capacity.

Table 10. Mechanism Analysis

	Information Technology	**Climate Resilience**
D_{it}	−0.417** (0.200)	− 0.522*** (0.157)
IT		
eco		0.045 (0.121)
prod		0.014 (0.053)
inf		0.344** (0.149)
educ		0.480** (0.204)
pui		0.002 (0.036)
FE: City	YES	YES
FE: Year	YES	YES
Observations	53	53
Adj. R^2	0.722	0.775
Robust standard errors in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$		

Heterogeneity Analysis

Since the capacity of school services (educ) has consistently shown a positive role in enhancing climate resilience, it is important to test for heterogeneity. Prior research has emphasized that smart cities require an educated population to maximize their effectiveness (Jiang et al. 2021; Lv and Gao 2023). To account for this, equation 1 was reestimated by including the interaction term D_{it} × educ. The results, reported in Table 8, indicate that smart cities with higher educ levels exhibit a stronger positive impact on climate resilience compared to those with lower educ, significant at the 10 percent level. This finding highlights the critical role of an educated population in raising awareness of climate risks. Schools provide essential knowledge and skills that help communities prepare for and respond to climate challenges (Borde et al. 2022). By integrating climate change education into the Philippine curriculum, schools can empower individuals to adopt sustainable practices, mitigate risks, and adapt to environmental disruptions. This aligns with broader smart city resilience goals and suggests that the government can strengthen both ASCN implementation and climate resilience by investing in school service capacity.

Table 8. Heterogeneity Result

	Climate Resilience	
D_{it}	−0.740**	(0.273)
D_{it} × *educ*	0.548*	(0.272)
Controls	YES	
Adj. R^2	0.705	
Robust standard errors in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$		

Conclusion and Recommendations

The idea of smart cities has been around for many years, and their development has mostly shown benefits for economic and social growth, but these benefits have been concentrated in the Global North and China. It is still unclear whether smart cities can help improve climate resilience planning in the Philippines and other developing countries in Asia. Important questions remain unanswered: Can the ASEAN Smart Cities Network help? How strong is the effect? Through what mechanism does it work? Does the impact differ across cities? These questions have not received enough attention from researchers, who continue to focus on the Global North and China. The ASEAN Smart Cities Network (ASCN) only began in 2018, much later than in more developed countries. At the same time, climate issues have become more serious in the Global South, especially in Southeast Asia, where urban areas are growing quickly. This makes it important to study whether smart city projects are working

for climate mitigation and adaptation, and how they can be improved. Doing so can help ASCN improve its plans by further managing targeted goals and provide lessons to other developing countries in Asia. Therefore, this study empirically measured the effect of the ASEAN Smart Cities Network (smart city) on climate resilience using the TWFE-DID model and mediating effect model, focusing on transmission pathway and urban heterogeneity, with 30 highly urbanized cities in the Philippines as the research sample.

The analysis yields the following main results:

1. ASCN (smart city) significantly reduces climate resilience in the Philippines: compared with non-ASCN, their level of climate resilience planning decreased by 0.204 (without covariates) and 0.285 (with covariates). The conclusion is still robust in PSM-DID, parallel trend analysis, placebo test through fake time and adjusted time period, removal of interfering cities, spillover through land border and shared region, and lag of covariates. The results also consistently further revealed that the capacity of school services improves climate resilience.

2. The mechanism test reveals that information technology (IT) capacity does not significantly act as a mediating variable in climate resilience. As much as unwelcoming, empirical evidence in this study shows that ASCN significantly reduces IT capacity of smart cities by 0.417. When IT is added, it promotes climate resilience by 0.039, but was not significant. It further shows that when IT is integrated, ASCN significantly reduces climate resilience planning by 0.269.

3. There was significant heterogeneity in urban characteristics among Philippine highly urbanized cities, such that those with higher capacity of school services are better off. This signifies that an educated population could potentially help climate resilience planning.

The findings of this study enrich the academic community's understanding of smart cities, especially that there has been a huge literature gap between developed and non-developed countries. Climate change poses a continuing risk, and although literature mainly focuses their outcome variable on carbon emission, climate resilience planning has never been explored ideally. This study empirically explores the impact of smart cities on climate resilience and offers insights and recommendations for the development of smart cities. Based on the conclusions, here are the targeted recommendations.

Recommendations

First, to address the counterproductive impact of ASCN on climate resilience planning in the Philippines, the government should recalibrate ASCN's infrastructure investments to prioritize climate mitigation and adaptation. Also, the government should strengthen its efficiency, as Gera (2018) mentioned that its political dysfunction hinders development in the Philippines, and that the prioritization operations of the next administration will not continue that of the replaced one. The study also reveals that current infrastructure projects do not significantly promote climate action goals, exacerbating resilience gaps and misalignment of government spending. Aligning ASCN's framework with localized climate risk assessments, such as those emphasized in ASEAN's urban resilience initiatives, could ensure infrastructure upgrades that directly enhance adaptive capacity. Strengthening enforcement mechanisms through capacity-building for local government units (LGUs), as suggested by Katigbak and Credo (2022), would further bridge the misalignment between ASCN's technical goals and climate resilience outcomes. Investment in the capacity of school services also becomes necessary as it continues to improve climate resilience.

Secondly, the negative effect of ASCN on IT capacity and its non-significant mediating role in climate resilience planning necessitates a shift toward integrating climate-smart IT solutions. Targeted investments in IT systems for disaster (early warning, which is part of command center upgrades of ASCN), data-sharing platforms, and digital public services could reverse this trend. Leveraging ASEAN's focus on rapid urbanization, the Philippines should collaborate with ASCN partners to deploy scalable IT tools that empower LGUs to monitor climate risks and coordinate responses (Crumpton et al. 2021). This aligns with broader ASEAN efforts to address urbanization challenges through smart city frameworks while countering the observed IT capacity decline in the Philippines ASCN cities.

Thirdly, strengthening the smart people pillar through the expansion and enhancement of educational services is essential for climate mitigation and adaptation. This offers a pathway to bolster climate resilience in Philippine cities. The heterogeneity analysis highlights that cities with higher school service capacity perform better in resilience planning. Therefore, enhancing educational programs on climate literacy, technical skills for green jobs, and participatory governance can cultivate a population equipped to support ASCN's climate goals.

Finally, this study provides important lessons for developing countries in Asia that aim to align smart city development with climate action goals. It shows that without clear climate resilience standards, urban technology projects may unintentionally weaken the ability of cities to mitigate and adapt to climate risks. The study also highlights the need to strengthen education by offering educational programs that combine

climate risk management with information technology skills. Additionally, it points out the importance of having separate funding and governance systems for digital infrastructure to maintain and improve IT systems that are crucial for managing disaster risks and transportation system enhancement. Therefore, policymakers in Asia should use performance-based investment measures for all smart city projects to ensure they support climate resilience. They should also include climate risk management education from primary to higher education levels to build a skilled workforce. Finally, they need to implement financing tools for resilience technologies along with strong oversight to protect essential ICT resources. Together, these steps can create a practical model for sustainable urban development.

References

Abdelati, Mohamed H. 2024. "Smart Traffic Management for Sustainable Development in Cities: Enhancing Safety and Efficiency." *International Journal of Advanced Engineering and Business Sciences* 5 (1): 49–62. https://doi.org/10.21608/ijaebs.2024.242361.1088.

Ahdekivi, Viljam, Hadi Ghanbari, and Matti Rossi. 2023. "Building Climate Resilience in Smart Cities Using Open Data Services." *Hawaii Internationla Conference on System Sciences*. University of Manoa. https://scholarspace.manoa.hawaii.edu/bitstreams/16f2c8d2-9e0a-498d-a687-8ae70477d26c/download.

Anderson, Allison. 2012. "Climate Change Education for Mitigation and Adaptation." *Journal of Education for Sustainable Development* 6 (2). https://doi.org/10.1177/0973408212475199.

Argyroudis, Sotirios A., Stergios Aristoteles Mitoulis, Eleni Chatzi, et al. 2022. "Digital Technologies Can Enhance Climate Resilience of Critical Infrastructure." *Climate Risk Management* 35 (100387). https://doi.org/10.1016/j.crm.2021.100387.

Asian Development Bank. 2009. *The Economics of Climate Change in Southeast Asia: A Regional Review*. Asian Development Bank. https://www.adb.org/sites/default/files/publication/29657/economics-climate-change-se-asia.pdf.

Association of Southeast Asian Nations. 2026. "ASEAN Smart Cities Network." ASEAN Secretariat. https://asean.org/our-communities/asean-smart-cities-network/.

Bokolo, Anthony Jnr. 2023. "Data Driven Approaches for Smart City Planning and Design: A Case Scenario on Urban Data Management." *Digital Policy, Regulation and Governance* 25 (4): 351–367. https://doi.org/10.1108/DPRG-03-2022-0023.

Borde, Badin, Pierre Léna, and Lydie Lescarmonti er. 2022. "Education as a Strategy for Climate Change Mitigation and Adaptation." In *Handbook of Climate Mitigation and Adaptation*, edited by Maximilian Lackner, Baharak Sajjadi, and Wei-Yin Chen, 3089–113. Springer. https://doi.org/10.1007/978-3-030-72579-2_149.

Ceccato, Vania, Marco Antonio Jorge, and Gabriel Gliori. 2024. "Assessing the Impact of Safety Interventions on Open Drug Markets in a Small Swedish Municipality Using Synthetic Control Methods." *Journal of Experimental Criminology* 22: 275–308. https://doi.org/10.1007/s11292-024-09652-7.

Chen, Shuxing, Xu Zhang, Wei Wang, Yunhao Liang, Wei He, and Zhixiong Tan. 2024. "The Effect of Smart City Policies on City Innovation—A Quasi-Natural Experiment from the Smart City Pilot Cities in China." *Sustainability* 16 (18): 8007. https://doi.org/10.3390/su16188007.

Crumpton, Charles David, Supawatanakorn Wongthanavasu, Peerasit Kamnuansilpa, John Draper, and Eva Bialobrzeski. 2021. "Assessing the ASEAN Smart Cities Network (ASCN) via the Quintuple Helix Innovation Framework, with Special Regard to Smart City Discourse, Civil Participation, and Environmental Performance." *International Journal of Urban Sustainable Development* 13 (1): 97–116. https://doi.org/10.1080/19463138.2020.1827411.

Davila Delgado, Juan M., Liam J. Butler, Niamh Gibbons, Ioannis Brilakis, Mohammed Z. E. B. Elshafie, and Campbell Middleton. 2017. "Management of Structural Monitoring Data of Bridges Using BIM." *Proceedings of the Institution of Civil Engineers—Bridge Engineering* 170 (3): 204–18. https://doi.org/10.1680/jbren.16.00013.

Department of Trade and Industry. 2024. *Cities and Municipalities Competitiveness Index 2023 Rankings Data Portal*. https://cmci.dti.gov.ph/data-portal.php.

Dharmarathne, Gangani, A. O. Waduge, Madhusha Bogahawaththa, Upaka Rathnayake, and D. P. P. Meddage. 2024. "Adapting Cities to the Surge: A Comprehensive Review of Climate-Induced Urban Flooding." *Results in Engineering* 22 (102123). https://doi.org/10.1016/j.rineng.2024.102123.

Dong, Feng, Jingyun Li, Yue Wang, Xiaoyun Zhang, Shengnan Zhang, and Shuaiqing Zhang. 2019. "Drivers of the Decoupling Indicator Between the Economic Growth and Energy-Related CO2 in China: A Revisit from the Perspectives of Decomposition and Spatiotemporal Heterogeneity." *Science of the Total Environment* 685: 631–58. https://doi.org/10.1016/j.scitotenv.2019.05.269.

Ferraris, Alberto, Gabriele Santoro, and Anna Claudia Pellicelli. 2020. "'Openness' of Public Governments in Smart Cities: Removing the Barriers for Innovation and Entrepreneurship." *International Entrepreneurship and Management Journal* 16 (4): 1259–80. https://doi.org/10.1007/s11365-020-00651-4.

Garschagen, Matthias, and Patricia Romero-Lankao. 2015. "Exploring the Relationships Between Urbanization Trends and Climate Change Vulnerability." *Climatic Change* 133: 37–52. https://doi.org/10.1007/s10584-013-0812-6.

Gera, Weena. 2018. "Scalar Politics in Philippine Urban Disaster Management: Reframing Metropolitan Governance for Local Resilience and Sustainability." *Erdkunde* 72 (4): 287–312. https://www.jstor.org/stable/26554735.

Guo, Qingbin, Yong Wang, and Xiaobin Dong. 2022. "Effects of Smart City Construction on Energy Saving and CO2 Emission Reduction: Evidence from China." *Applied Energy* 313 (118879). https://doi.org/10.1016/j.apenergy.2022.118879.

Guo, Qingbin, and Jinrong Zhong. 2022. "The Effect of Urban Innovation Performance of Smart City Construction Policies: Evaluate by Using a Multiple Period Difference-in-Differences Model." *Technological Forecasting and Social Change* 184 (122003). https://doi.org/10.1016/j.techfore.2022.122003.

Gupta, Namita. 2023. "Smart Cities Challenges, Technologies and Trends." *Frontiers in Big Data* 6 (1258051). https://doi.org/10.3389/fdata.2023.1258051.

Haw, Joseph Y., Ronnel B. King, and Jose Eos R. Trinidad. 2021. "Need Supportive Teaching Is Associated with Greater Reading Achievement: What the Philippines Can Learn from PISA 2018." *International Journal of Educational Research* 110 (101864). https://doi.org/10.1016/j.ijer.2021.101864.

Hurlimann, Anna, Sareh Moosavi, and Geoffrey R. Browne. 2021. "Urban Planning Policy Must Do More to Integrate Climate Change Adaptation and Mitigation Actions." *Land Use Policy* 101 (105188). https://doi.org/10.1016/j.landusepol.2020.105188.

Issa Zadeh, Seyed Behbood, and Claudia Lizette Garay-Rondero. 2023. "Enhancing Urban Sustainability: Unravelling Carbon Footprint Reduction in Smart Cities through Modern Supply-Chain Measures." *Smart Cities* 6 (6): 3225–50. https://doi.org/10.3390/smartcities6060143.

Jamilurahman, Faizi, Sharifi Ahmad Zia, Momand M. Yusuf, and Habibi Zargay. 2024. "Climate Change Adaptation through Smart Cities in Developing Countries: A Review." *Nangarhar University International Journal of Biosciences*, special issue 2024: 327–30. https://www.researchgate.net/publication/382528476_Climate_

Change_Adaptation_through_Smart_Cities_in_Developing_Countries_A_Review.

Jiang, Hongli, Pengcheng Jiang, Dong Wang, and Jiahui Wu. 2021. "Can Smart City Construction Facilitate Green Total Factor Productivity? A Quasi-Natural Experiment Based on China's Pilot Smart City." *Sustainable Cities and Society* 69 (102809). https://doi.org/10.1016/J.SCS.2021.102809.

Jones, David S. 2007. "Public Procurement in Southeast Asia: Challenge and Reform." *Journal of Public Procurement* 7 (1): 3–33. https://doi.org/10.1108/JOPP-07-01-2007-B001.

Katigbak, Jovito Punzalan, and Jeremie P. Credo. 2022. "City Diplomacy Amid Regional Integration: A Philippine Perspective on the ASEAN Smart Cities Network." *International Journal of Research and Innovation in Social Science* 6 (7): 210–15. https://doi.org/10.47772/IJRISS.2022.6715.

Kong, Lily, and Orlando Woods. 2021. "Scaling Smartness, (De)provincialising the City? The ASEAN Smart Cities Network and the Translational Politics of Technocratic Regionalism." *Cities* 117 (103326). https://doi.org/10.1016/j.cities.2021.103326.

Kurbalija, Jovan. 2025. "Valencia Flooding: Why Did One of the Leading Smart Cities Fail on Basic Safety?" *Diplo*, updated 22 October 2025. https://www.diplomacy.edu/blog/valencia-flooding-smart-city-safety-failure/.

Li, Wei. 2024. "Research on the Construction of Smart, Safe and Resilient Cities." *Journal of Civil Engineering and Urban Planning* 6 (1): 92–97. https://doi.org/10.23977/jceup.2024.060113.

Liu, Simin, and Michael James Harrison. 2024. "Climate Change Adaptation Strategies in Coastal Cities of Australia: A Comparative Study of Sydney, Melbourne, and Brisbane." *Studies in Social Science & Humanities* 3 (5): 64–71. https://doi.org/10.56397/SSSH.2024.05.06.

Lv, Rongjie, and Hao Gao. 2023. "Effects of Smart City Construction on Employment: Mechanism and Evidence from China." *Empirical Economics* 65 (5): 2393–425. https://doi.org/10.1007/s00181-023-02429-3.

Neo, En Xin, Khairunnisa Hasikin, Khin Wee Lai, et al. 2023. "Artificial Intelligence-Assisted Air Quality Monitoring for Smart City Management." *PeerJ Computer Science* 9 (e1306). https://doi.org/10.7717/peerj-cs.1306.

Ogut, Ozge, Julia Nerantzia Tzortzi, and Chiara Bertolin. 2024. "Creating a Roadmap to Forecast Future Directions in Vertical Green Structures as a Climate Change Mitigation Strategy: A Critical Review of Technology-Driven Applications." *Sustainability* 16 (11): 4543. https://doi.org/10.3390/su16114543.

Ramirez Lopez, Leonardo Juan, and Angela Ivette Grijalba Castro. 2020. "Sustainability and Resilience in Smart City Planning: A Review." *Sustainability* 13 (1): 181. https://doi.org/10.3390/su13010181.

Riaz, Khurram, Marion McAfee, and Salem S. Gharbia. 2023. "Management of Climate Resilience: Exploring the Potential of Digital Twin Technology, 3D City Modelling, and Early Warning Systems." *Sensors* 23 (5): 2659. https://doi.org/10.3390/s23052659.

Rolnick, David, Priya L. Donti, Lynn H. Kaack, et al. 2022. "Tackling Climate Change with Machine Learning." *ACM Computing Surveys (CSUR)* 55 (2): 42. https://doi.org/10.1145/3485128.

Roth, Jonathan, Pedro H. C. Sant'Anna, Alyssa Bilinski, and John Poe. 2023. "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature." *Journal of Econometrics* 235 (2): 2218–44. https://doi.org/10.1016/j.jeconom.2023.03.008.

Stuart, Elizabeth A., and Donald B. Rubin. 2008. "Best Practices in Quasi-experimental Designs." In *Best Practices in Quantitative Methods*, edited by Jason Osborne, 155–76. SAGE Publications. https://doi.org/10.4135/9781412995627.d14.

Suprayitno, Degdo, Sofyan Iskandar, Keristian Dahurandi, Totok Hendarto, and Freddy Johanis Rumambi. 2024. "Public Policy in the Era of Climate Change: Adapting Strategies for Sustainable Futures." *Migration Letters* 21 (S6), 945–58. https://migrationletters.com/index.php/ml/article/view/8068.

Teresa, Ma. Teresa S., Agnes M. Ramos, Eva U. Cammayo, and Jeoffrey Lloyd R. Bareng. 2020. "Benchmarking of the Water Supply and Wastewater Management System of a Smart City: A Case Study of Cauayan City, Philippines." *TEST: Engineering and Management* 83: 4768–75. http://testmagzine.biz/index.php/testmagzine/article/view/4394.

Trinidad, Mariel. 2023. "Imperialism and the Political Economy of Climate Adaptation in the Philippines." Undergraduate research thesis, Ohio State University. https://kb.osu.edu/server/api/core/bitstreams/7449b4d7-02b2-4d75-b31d-93aa45491b49/content.

Tyagi, Ruchi, Amrita Chaiurasia, and Suresh Vishwakarma. 2024. "Climate Change Adaptation and Mitigation Strategies at Future Smart Cities." *IET Conference Proceedings* 2024 (37): 478–85. https://doi.org/10.1049/icp.2025.0959.

Tyler, Stephen, and Marcus Moench. 2012. "A Framework for Urban Climate Resilience." *Climate and Development* 4 (4): 311–26. https://doi.org/10.1080/17565529.2012.745389.

Wang, Li, Qian Xie, Fei Xue, and Zongxin Li. 2022. "Does Smart City Construction Reduce Haze Pollution?" *International Journal of Environmental Research and Public Health* 19 (24): 16421. https://doi.org/10.3390/ijerph192416421.

Yoon, Seok Yong, Hong Soo Lee, Thilo Zelt, Ulf Narloch, and Elliot Aguirre. 2020. "Smart City Pathways for Developing Asia: An Analytical Framework and Guidance." *ADB Sustainable Development Working Paper Series*, no. 71. https://www.adb.org/sites/default/files/publication/673441/sdwp-071-smart-city-pathways-developing-asia.pdf.

Service to the People? Mapping and Measuring the University of the Philippines' Impact on the Philippine Policy Process

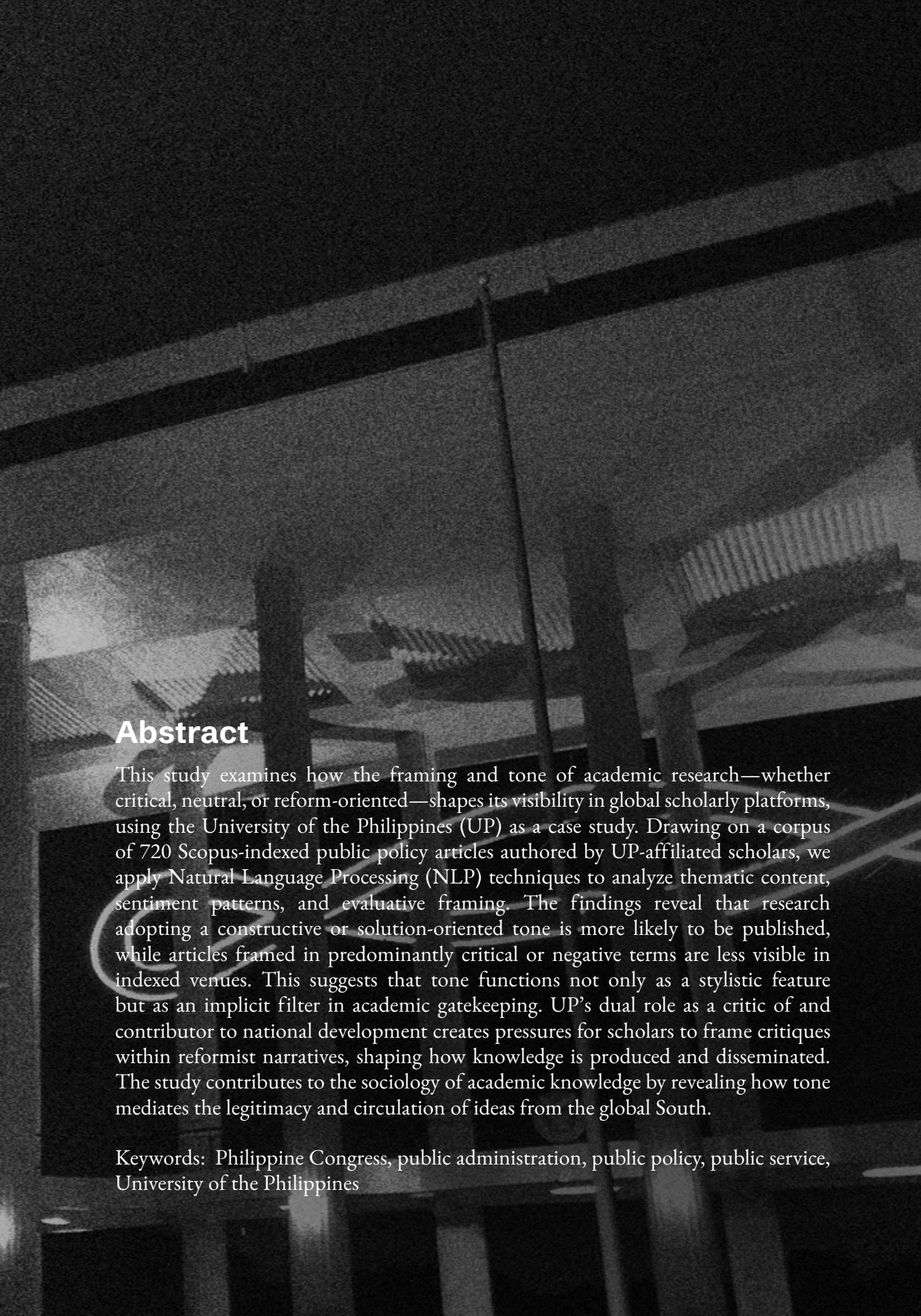
Ronald A. Pernia, PhD[1] *and Rogelio Alicor L. Panao, PhD*[2]

1 rpernia21@gmail.com

2 rlpanao@up.edu.ph

The Oblation and Quezon Hall at night, University of the Philippines Diliman. Completed in 1950 and designed by National Artist Juan Nakpil, Quezon Hall houses the administration of the UP System. As the country's leading producer of public policy research, UP embodies the tension between critique and service that this study examines.

Patrick Roque (2017), via Wikimedia Commons, licensed under CC BY-SA 4.0, https://en.wikipedia.org/wiki/File:University_of_the_Philippines_-_Quezon_Hall_and_oblation_%28UP_Diliman,_Quezon_City%29%282017-07-14%29_%281%29.jpg



Abstract

This study examines how the framing and tone of academic research—whether critical, neutral, or reform-oriented—shapes its visibility in global scholarly platforms, using the University of the Philippines (UP) as a case study. Drawing on a corpus of 720 Scopus-indexed public policy articles authored by UP-affiliated scholars, we apply Natural Language Processing (NLP) techniques to analyze thematic content, sentiment patterns, and evaluative framing. The findings reveal that research adopting a constructive or solution-oriented tone is more likely to be published, while articles framed in predominantly critical or negative terms are less visible in indexed venues. This suggests that tone functions not only as a stylistic feature but as an implicit filter in academic gatekeeping. UP's dual role as a critic of and contributor to national development creates pressures for scholars to frame critiques within reformist narratives, shaping how knowledge is produced and disseminated. The study contributes to the sociology of academic knowledge by revealing how tone mediates the legitimacy and circulation of ideas from the global South.

Keywords: Philippine Congress, public administration, public policy, public service, University of the Philippines

Introduction

What determines whether academic research gets published, cited, or taken seriously by policymakers?

Beyond methodological rigor or theoretical sophistication, emerging scholarship suggests that the dissemination of scientific findings can be influenced by various factors beyond the content itself. A scientist's facial appearance, for example, can influence public perception, with perceived competence and morality boosting interest and quality, while attractiveness increases interest but lowers perceived quality (Gheorghiu et al. 2017). The tone and framing of scientific discoveries, on the other hand, can play a role in their sharing among non-scientists (Canestaro et al. 2017; Dickersin, Min, and Meinert 1992; Milkman and Berger 2014). However, in political and policy contexts, especially within asymmetrical power environments, can research afford to be agnostic about its positioning without affecting its reach in both scholarly and policy circles?

This paper explores the underexamined relationship between evaluative framing and the politics of academic publication. Specifically, we argue that how research is positioned—whether as a critique of existing systems, a neutral analysis of conditions, or a reform-oriented proposal—can condition its likelihood of being accepted for publication in peer-reviewed outlets. The tone of scholarship, often seen as a stylistic or rhetorical choice, in fact reflects deeper strategies of intellectual positioning that shape access to academic platforms. In contexts where academic institutions are expected to contribute to national development goals or respond to policy imperatives, this relationship between tone and publication becomes even more pronounced.

This study investigates the dynamics of scholarly framing within the context of the University of the Philippines (UP), the nation's premier academic institution and a historically significant hub for public scholarship. Given UP's enduring commitment to critical engagement, nation-building, and policy-oriented research, it serves as an exemplary case for examining how academics navigate the tension between intellectual independence, institutional imperatives, and publication incentives in the Philippine setting. How do UP scholars frame their work in ways that facilitate publication? Does a reformist or constructive tone increase the odds of dissemination in peer-reviewed venues? Is overt critique penalized in favor of more policy-aligned or optimistic framings?

To answer these questions, we use natural language processing (NLP) techniques to analyze a corpus of 720 peer-reviewed journal articles authored by UP-affiliated scholars and indexed in Scopus. We focus specifically on studies related to public

policy, governance, and public administration. Our analytical approach includes sentiment analysis to determine the evaluative tone of each article, frequency and keyword analysis to identify thematic patterns, and topic modeling to uncover recurring clusters of ideas. The sentiment dimension allows us to assess how often UP research expresses positive, neutral, or negative stances in relation to the state, institutions, or public challenges—and whether certain tones correlate with more frequent appearance in the literature.

Our findings show that research combining critical analysis with a constructive tone—acknowledging problems while also presenting possible solutions—is more likely to be published in indexed journals. Articles that discuss issues such as governance failures, inequality, or institutional weakness often include positively framed terms like "sustainable," "support," "reform," and "development." In contrast, studies that rely heavily on negative language—focusing solely on "corruption," "crisis," or "failure" without suggesting reforms—appear less frequently. This suggests that a balanced or forward-looking tone may act as a tacit publication strategy, improving a manuscript's chances of passing editorial and peer review, particularly in policy-oriented and development-related disciplines.

This does not necessarily imply editorial bias. Instead, they likely reflect the broader operational context for universities such as UP. These institutions are not solely hubs for critical thought; they are also deeply integrated into national development agendas, influenced by funding priorities, and assessed by publication metrics that might favor specific research framings. In this environment, scholars may strategically frame their work to enhance its institutional relevance and publication potential, effectively balancing critique and collaboration.

This paper contributes to the growing literature on the sociology of knowledge and the political economy of academic publishing. It challenges the assumption that peer-reviewed publication is a purely meritocratic process, highlighting instead how tone functions as a form of gatekeeping, influencing not just what knowledge is produced, but what knowledge circulates. By emphasizing the role of evaluative framing, we aim to provoke deeper reflection on how scholars in the global South, in particular, navigate the subtle pressures of academic legitimacy, policy relevance, and critical engagement. Foregrounding tone as an analytical category, this study offers a new way of understanding the politics of academic visibility—and the quiet compromises that often underwrite it.

The paper begins with a review of relevant literature on how universities influence and are influenced by public policy contexts, and how academic impact is increasingly shaped by communication strategies. It then presents the research design, which involves NLP techniques applied to a corpus of 720 peer-reviewed journal articles

authored by UP scholars, all focused on public policy. Using frequency analysis, sentiment analysis, and topic modeling (through latent Dirichlet allocation or LDA), the study systematically maps the themes and tone of UP's research output. The findings demonstrate that UP scholars frequently address governance, equity, and sustainability, often framing their work in ways that blend critique with constructive policy engagement. Compared to other universities, UP research tends to emphasize social justice and democratic accountability. The paper concludes by arguing that universities are most influential when their research is interdisciplinary, publicly engaged, and responsive to real-world governance challenges, calling for more support for collaborative scholarship and deliberate alignment between research and policy priorities.

Universities as Pillars of Public Policy and Development

The role of universities in shaping public policy and governance has been widely acknowledged and extensively studied. Far from being passive producers of academic knowledge, universities serve as engaged institutions that collaborate with policymakers and influence the design of governance systems and public decision-making. Burstein (1998), for example, notes that although academic research and public opinion both significantly affect policy processes, their interaction is often insufficiently explored. They emphasize that while public sentiment has a notable impact on policy direction, it is academic research that offers the empirical foundation necessary for informed decision-making.

The relationship between academia and industry further highlights the importance of universities in the policy domain. Kochenkova, Grimaldi, and Munari (2016) point out that public policy plays a vital role in enabling effective knowledge transfer between these sectors. Their study calls attention to the need for systematic assessment of how research outputs, particularly in innovation and technology, translate into meaningful policy interventions. This underscores universities' position as intermediaries connecting theoretical insights with practical societal solutions.

In addition to shaping policy directly, universities significantly influence regional development trajectories. Trippl, Sinozic, and Lawton Smith (2015) provide a detailed account of how universities contribute to regional growth, presenting a framework that illustrates how different types of university activities can be harnessed to support development goals. They review models such as that of the entrepreneurial university and the regional innovation system (RIS), each highlighting distinct roles and outputs. Their findings indicate that the support of public policy institutions is critical in driving these roles, with national differences evident in cases such as Austria, Sweden, and the United Kingdom. In a similar vein, Harrison and Turok (2017) argue that regional development research must place universities at the center

of analysis, applying the same critical scrutiny to higher education institutions that is typically directed at other social and economic actors. However, Uyarra (2010) offers a more skeptical perspective by outlining five models that reflect the varied and sometimes conflicting ways universities engage with regional innovation systems. Their critique draws attention to the lack of coherence in innovation policy frameworks and the challenges that arise when trying to align them with universities' diverse and evolving roles.

Universities in developing countries also navigate complex institutional environments shaped by global dynamics. Saad and Zawdie (2011), in their exploration of the triple helix framework, examine how universities, industries, and governments collaborate in emerging economies. Their analysis highlights the necessity of policy mechanisms that foster synergy among these sectors to promote national and regional innovation. In parallel, globalization has intensified the application of market-oriented principles in higher education governance, with the introduction of market-based and quasi-market policy instruments intended to enhance economic performance and expand access to education (Dill 1997, 2014; Dill and Soo 2005).

Taken together, these studies reaffirm the indispensable role of universities in the public policy sphere. Their functions are not limited to knowledge generation but extend to shaping discourse, formulating policies, and influencing social and economic outcomes. As St. John, Daun-Barnett, and Moronski-Chapman (2018) contend, higher education institutions have consistently been central to policy reform efforts, contributing to broader discussions on governance, development, and equity. Within this context, universities like UP exemplify how academic institutions can play a critical role in addressing national challenges—not only by informing education policy but also by contributing to broader governance debates and public sector decision-making.

Data and Method

This study applies natural language processing (NLP) to a corpus of 720 peer-reviewed journal articles authored by scholars affiliated with UP, with the aim of analyzing how language use shapes scholarly visibility. Specifically, we investigate whether certain framings—particularly those that balance critique with constructive or reform-oriented narratives—are more likely to appear in indexed academic platforms. By examining evaluative tone, thematic focus, and recurring language patterns, the study contributes methodologically to the use of computational tools in policy research and addresses a broader theoretical concern: how discourse framing mediates academic visibility and the politics of scholarly dissemination in the global South.

The corpus was constructed by retrieving research articles from the Scopus database using the keywords "Philippine*" AND "public policy" OR "governance" OR "public administration." The wildcard (*) was used to capture word stems and variants, while the search was filtered to exclude conference papers, book chapters, and review articles in line with prior methodological recommendations (Pernia 2024). This yielded 720 articles that reflect the breadth of scholarly engagement with policy issues in the Philippines.

NLP has become a foundational tool in the analysis of large-scale text corpora, allowing for the systematic extraction of patterns, themes, and sentiment from unstructured data (Grimmer, Roberts, and Stewart 2022). Consistent with prior work by Roberts et al. (2013) and Grimmer and Stewart (2013), we employ techniques such as topic modeling and sentiment analysis to examine the underlying structure of academic discourse. These methods are particularly suited for analyzing how research institutions engage with issues of governance, reform, and policy relevance.

All analyses were performed in R, a statistical programming environment well-suited for text mining and data visualization. Three main NLP techniques were employed:

1. Term Frequency-Inverse Document Frequency (TF-IDF)

   TF-IDF analysis was used to assess the prominence of terms across articles by calculating the relative importance of a word in a single document compared to its occurrence in the entire corpus. This allows us to identify keywords and concepts that are central to institutional research outputs from UP, Ateneo de Manila University (hereinafter Ateneo), and De La Salle University (DLSU). As demonstrated by Manning (2008), TF-IDF is a powerful method for highlighting distinctive vocabulary and topical focus in textual data.

2. Latent Dirichlet Allocation (LDA) Topic Modeling

   To uncover latent themes in the corpus, we employed LDA topic modeling following Roberts et al. (2013). This unsupervised learning technique identifies clusters of co-occurring terms that represent recurring policy themes. Topics such as disaster governance, sustainable development, and institutional reform emerged as dominant themes across the dataset. The application of LDA aligns with the broader use of unsupervised models to capture both surface and latent content within academic texts (Grimmer and Stewart 2013), helping to illuminate how institutions like UP articulate their contributions to governance discourse.

3. Sentiment Analysis

   Sentiment analysis was conducted to evaluate the overall tone of the articles. By using established sentiment lexicons, we assigned sentiment scores to individual words and aggregated them at the document level. Terms with positive

connotations (e.g., "reform," "sustainable") were contrasted with negative terms (e.g., "poverty," "disaster") to examine how the tone of research aligns with or critiques prevailing policy conditions. This enables an interpretive layer in the analysis of how academic discourse reflects or reframes public challenges.

All preprocessing and analytical tasks were conducted using R packages such as tm (Feinerer 2013; Feinerer, Hornik, and Feinerer 2015), quanteda (Benoit et al. 2018), and topicmodels (Grün and Hornik 2011). These tools allowed for the efficient handling, analysis, and visualization of large text datasets, in line with best practices in computational text analysis (Welbers, Van Atteveldt, and Benoit 2017).

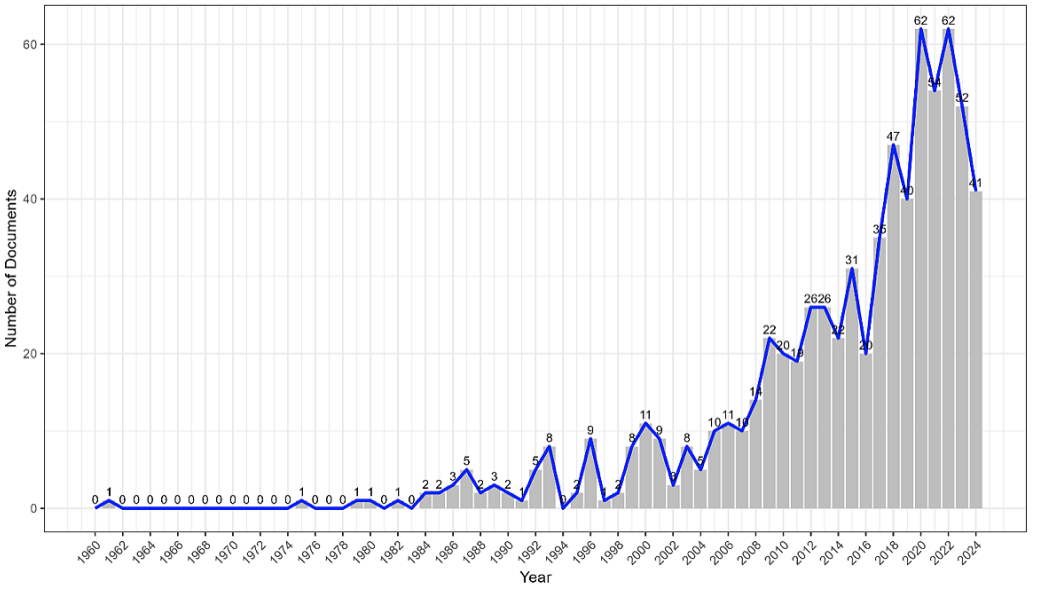
Taken together, the study provides an empirical basis for understanding how evaluative language, topic choice, and institutional framing influence the dissemination of policy knowledge in academic platforms. It underscores the importance of textual framing in shaping both scholarly impact and the politics of knowledge production in the Global South.

Results

This section presents the empirical results of the study, with a particular focus on how evaluative tone—whether positive, negative, or neutral—functions as a framing device that may influence the visibility and publication of research in academic platforms. The analysis begins with a descriptive overview of the dataset, followed by results from NLP methods, including frequency analysis, TF-IDF, sentiment analysis, and latent Dirichlet allocation (LDA) topic modeling.

Figure 1 illustrates a pronounced upward trend in document production within this field, beginning in the mid-2000s. Prior to this, from 1961 to the early 2000s, scholarly output remained relatively modest, with fewer than 10 documents published annually. This low volume is likely attributable to limited access to publication platforms and a slower pace of academic dissemination (Odlyzko 2002; Solomon, Laakso, and Björk 2013). However, starting around 2010, there was a notable increase in document production, with annual publications rapidly peaking at approximately 60 documents per year by the late 2010s and early 2020s (Odlyzko 2002; Solomon, Laakso, and Björk 2013). This significant surge reflects a broader, global trend of accelerating academic productivity, driven by advancements in digital research tools, the rise of open-access publishing, and improved global academic connectivity (Ding et al. 2010; Lee and Bozeman 2005).

Figure 1. Number of documents per year of the Public Policy Corpus in the Philippines. Source: Scopus Database.



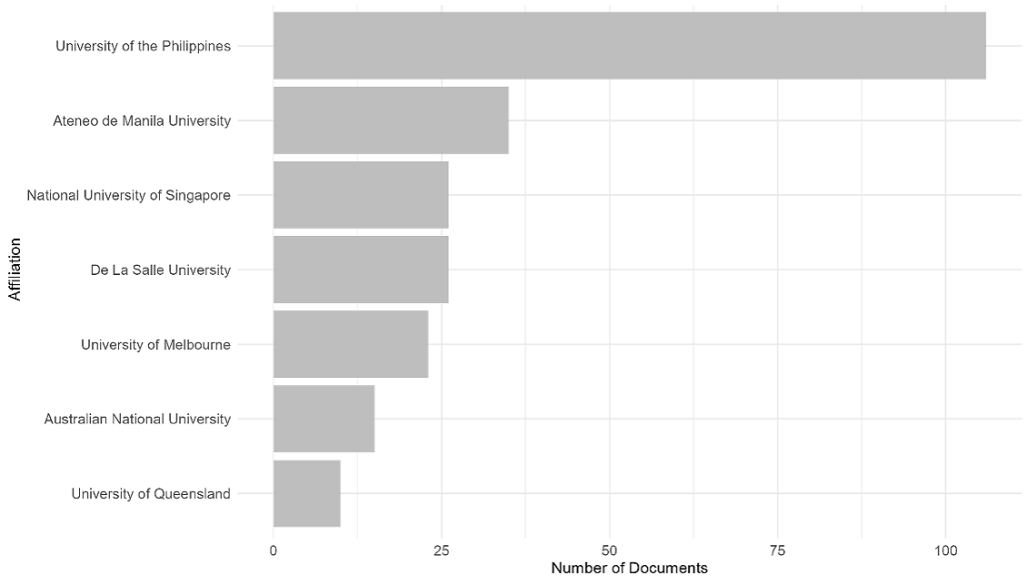
Note: The figures were generated in *R*.

The substantial increase in scholarly output can also be linked to the expansion of higher education, increased research funding, and the growing importance of interdisciplinary studies, which have opened new publication avenues. This evolving academic landscape likely encourages a wider array of research framings, ranging from positive assessments of policy success to critical analyses of failures, all of which can potentially find suitable publication outlets. The slight decline observed in recent years (Figure 1) may suggest the influence of temporary external factors, such as global economic slowdowns, shifting research priorities, or the disruptive impact of the COVID-19 pandemic, which led to a reallocation of research efforts (Krukowski, Jagsi, and Cardel 2021; Staniscuaski et al. 2021; Arora et al. 2021; Ellinas et al. 2022). Despite these short-term fluctuations, the overall trend reflects an era of expanding academic scholarship that shows no signs of long-term decline, continuing to accommodate diverse research narratives.

Figure 2 presents document production by institutional affiliation, highlighting contributions from various entities engaged in public policy research. UP clearly stands as the leading contributor, producing the largest volume of documents and reinforcing its prominent role in shaping public policy discourse. UP's dominance in scholarly output is notable, reflecting its sustained commitment to research and policy development within the Philippines. This prominence may also indicate an institutional tendency to frame research in ways that align with prevailing policy interests or national development goals, potentially enhancing its likelihood of being

published. Following UP, other leading Philippine institutions, including Ateneo and DLSU, also contribute significantly to public policy research, albeit with lower document counts. These institutions, while important contributors, collectively illustrate a national hierarchy in research output within the Philippines' public policy landscape.

Figure 2. Number of documents by affiliation of the Public Policy Corpus in the Philippines. Source: Scopus Database.



Note: The figures were generated in *R*.

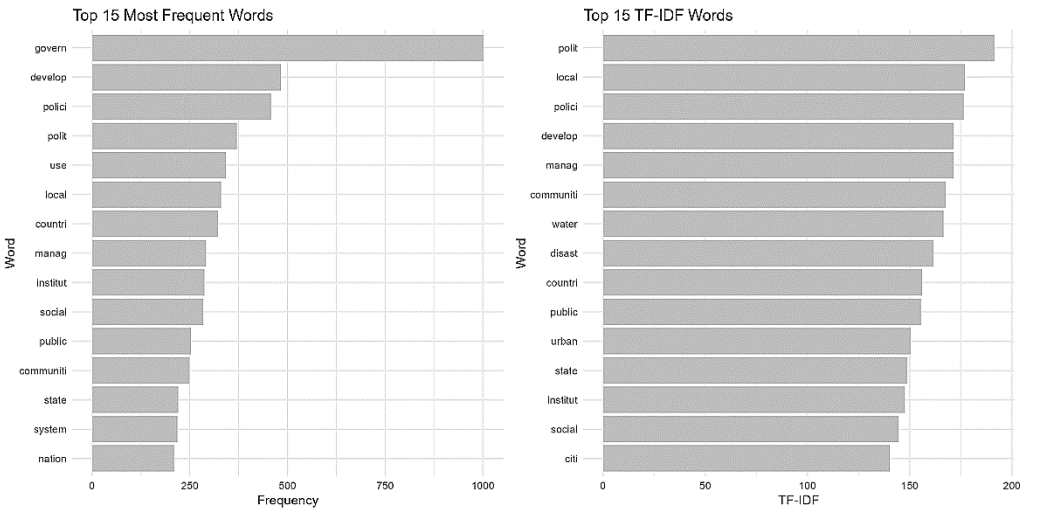
Furthermore, the presence of international universities such as the National University of Singapore, the University of Melbourne, and the Australian National University among the top affiliations points to robust international collaborations and the broader influence of global academic networks in Philippine public policy scholarship. This international dimension highlights the Philippines' role as a key player in Southeast Asian public policy research and suggests increasing interconnectedness among academic institutions globally. Such collaborations can also enrich local public policy discourses by integrating diverse perspectives and global best practices, potentially leading to research framed with a global-comparative or best-practice orientation, which is often highly appealing to international journals.

These insights offer a broader understanding of the factors driving both local and international research outputs, underscoring the importance of collaboration and the evolving role of digital tools in shaping modern academic scholarship. The adoption of specific research tones and framings—whether problem-focused, solution-oriented, or critically analytical—can significantly influence a paper's appeal

to various academic outlets. In the next section, we will present general statistical insights derived from the Scopus database using NLP approaches, building on this preliminary overview.

Figure 3 presents an analysis of the most frequent words and TF-IDF terms from the public policy text corpus. The high-frequency word chart reveals that terms such as "govern," "develop," "policy," "polit," and "use" are dominant, reflecting a strong focus on governance, development, and policy-related themes. The prominence of words like "local" and "community" further indicates significant interest in decentralized governance and the management of community-level issues, consistent with literature highlighting the importance of local governance in policy effectiveness (Wright et al. 2016; Cheema and Rondinelli 2007; Bardach and Patashnik 2023). These high-frequency terms suggest that the corpus is heavily concentrated on addressing governance challenges and policy formulation at both local and national levels (Newig and Koontz 2014; Krahmann 2003; Thompson 2008; Bauer, Feichtinger, and Steurer 2012; Gupta 2007). This consistent framing around core governance and development themes likely ensures a steady stream of publishable work within these well-established domains.

Figure 3. High-frequency words and TF-IDF terms in the Public Policy Corpus in the Philippines. Source: Scopus Database.



Note: The figures were generated in *R*.

Additionally, words like "manage" and "institution" highlight the importance of institutional capacity and social system management in the analyzed texts, echoing findings from North (1990) and DiMaggio and Powell (1983) regarding institutions' critical role in governance and policy implementation. Research framed around institutional analysis, whether it highlights positive outcomes (e.g., successful

reforms) or identifies negative ones (e.g., institutional barriers), often finds a receptive audience.

The TF-IDF analysis, conversely, reveals slightly different insights. While "policy," "develop," and "local" still rank high, terms such as "disaster" and "water" emerge, indicating that certain documents emphasize specific topics like disaster management and water resource governance. This aligns with recent studies that underscore the need for tailored policies in disaster risk reduction and water resource management (Knutsson and Ostwald 2023; Whelchel et al. 2018; Ishiwatari 2022). These terms, though less frequent across the entire dataset, are pivotal in particular contexts, emphasizing the need for more focused policies in areas such as environmental resilience and crisis management. Framing research around urgent, specific issues like "disaster" or "water" can significantly enhance its policy relevance and, consequently, its likelihood of publication, especially in outlets seeking timely interventions.

This distinction between frequent and TF-IDF terms underscores TF-IDF's value in uncovering specialized themes within the corpus that are not necessarily frequent but are highly relevant in particular discussions (Silge 2017). The combined analysis suggests a corpus broadly concerned with governance and development, while also addressing specialized topics like disaster management and environmental policy in specific contexts, each potentially warranting distinct rhetorical strategies for publication to maximize their scholarly impact.

Framing UP's Academic Contributions

After extracting the TF-IDF, we proceeded with topic modeling using LDA, visualizing the top terms across 10 topics as shown in Figure 4, focusing exclusively on the works of UP System scholars. The topic modeling analysis of UP reveals a broad range of key themes reflecting the institution's diverse academic and research focus areas. The way these topics are framed—whether emphasizing successes, identifying challenges, or providing neutral analyses—contributes to their distinct appeal and subsequent publication pathways.

Topic 1 centers on governance issues related to sustainable urban development, with key terms including "govern" (β = 0.12), "develop" (β = 0.10), "policy" (β = 0.09), "city" (β = 0.08), and "sustain" (β = 0.07). This highlights UP's focus on shaping national policy directions. The significant presence of "govern" and "develop" suggests that research in this area critically evaluates the effectiveness of governance mechanisms in fostering sustainable urban growth. Existing studies emphasize the importance of collaborative governance approaches in addressing urban challenges, indicating that effective governance models can enhance sustainability outcomes in urban settings (Huntjens 2021; Allam et al. 2022; Vanli 2023). Research framed with

a positive or neutral outlook on "sustainability" and "development" is generally well-received in academic publications.

Topic 2 focuses on water resource management and local governance, represented by terms such as "water" (β = 0.15), "govern" (β = 0.11), "urban" (β = 0.09), and "local" (β = 0.08). The high beta coefficient for "water" signifies its critical role in local policy discussions, emphasizing the urgency of sustainable water governance practices. Scholarly work underscores the need for integrated water resource management (IWRM) frameworks that consider local contexts, suggesting that local governments must engage communities in decision-making processes to effectively manage water resources (Giordano and Shah 2017; Al Radif 1999; Nagata et al. 2022). Studies that frame water issues as urgent problems often see quicker publication due to their direct policy relevance and societal impact.

Topic 3 addresses policies related to Indigenous rights and social equity, with prominent terms including "indigenous" (β = 0.13), "people" (β = 0.12), "community" (β = 0.10), and "social" (β = 0.09). This reflects UP's commitment to social inclusion and Indigenous rights. The substantial presence of "indigenous" and "community" indicates robust engagement with social justice issues and the recognition of Indigenous peoples in policy frameworks. Smith (2021) argues that effective indigenous policy requires participatory approaches that respect the rights and voices of Indigenous communities, highlighting the need for collaboration between policymakers and Indigenous populations to develop equitable solutions. Framing research as advocating for marginalized groups or exposing systemic inequities can lead to impactful, though sometimes challenging, publications in social science journals.

Topic 4 highlights political dynamics in coastal management and environmental justice, with key terms such as "coastal" (β = 0.14), "politics" (β = 0.11), and "justice" (β = 0.09). The emphasis on "coastal" points to the significance of marine environments in policy discussions, while "justice" reflects a growing awareness of equitable access to coastal resources. This aligns with prevailing works emphasizing that environmental justice frameworks must address disparities faced by vulnerable coastal communities, advocating for policies that prioritize their rights and needs in coastal governance (Engen et al. 2021; Lau, Gurney, and Cinner 2021; Jurjonas and Seekamp 2020). A critical framing of "justice" issues within environmental policy often finds a receptive audience in specialized journals focusing on social and environmental equity.

Topic 5 centers on policies for managing aquatic ecosystems and fishery resources, indicated by terms like "lake" (β = 0.10), "fisheries" (β = 0.12), "govern" (β = 0.09), and "security" (β = 0.08). The presence of "fisheries" and "security" underscores the importance of sustainable management practices to ensure food security and

ecological health. Green et al. (2014) highlight the critical role of well-designed marine reserve networks in enhancing fisheries management and biodiversity conservation while addressing the impacts of climate change and habitat destruction. Similarly, Naiman and Latterell (2005) propose ecological and social principles for defining fish habitat, emphasizing that effective management strategies must adapt to the dynamic nature of ecosystems and the interconnections between aquatic and terrestrial environments. Research framed around "security" and "sustainable management" has broad appeal across policy-oriented and environmental science publications.

Topic 6 reflects research on community-level security policies and responses to local conflicts, characterized by terms such as "security" ($\beta = 0.13$), "threat" ($\beta = 0.10$), and "community" ($\beta = 0.09$). The prominence of "security" suggests a focus on ensuring the safety and well-being of local populations in the face of various threats, further emphasizing the university's focus on peace and community well-being. Papers that adopt a "problem-solution" framing for community security issues are often highly publishable due to their direct applicability and potential for policy impact.

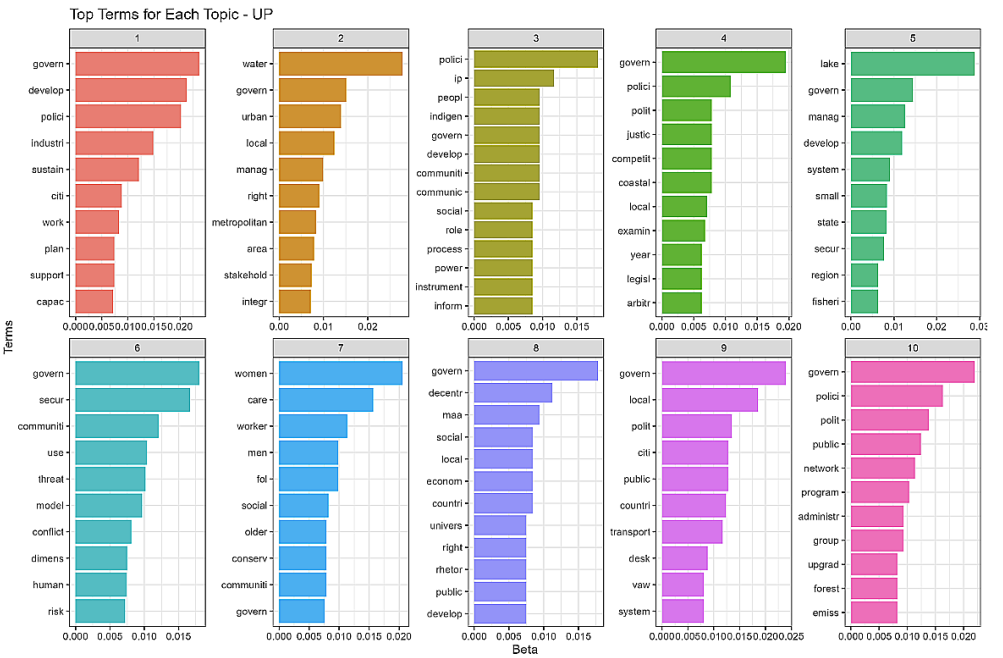
Topic 7 emphasizes gender-sensitive policy and social care issues, particularly affecting women and caregivers, with key terms including "women" ($\beta = 0.12$), "care" ($\beta = 0.11$), and "worker" ($\beta = 0.09$). The focus on "women" and "care" underscores the need for policies that address gender disparities and support caregiving roles, showcasing UP's engagement with social justice and gender equality (Ocampo and Binondo 2022). Research that frames issues around "gender equity" or highlights social disparities is generally well-regarded in social science and policy journals that prioritize inclusive approaches.

Topic 8 explores how decentralization shapes social and economic policies, with prominent terms such as "decentralization" ($\beta = 0.12$), "social" ($\beta = 0.10$), "country" ($\beta = 0.09$), and "economy" ($\beta = 0.08$). The significant presence of "decentralization" indicates a growing interest in how power dynamics affect governance outcomes, reflecting the institution's focus on empowering local governance structures and fostering economic growth at the grassroots level. This topic aligns with the broader literature on fiscal federalism, suggesting that decentralization can enhance efficiency and responsiveness in governance by allowing local governments to tailor policies to community needs (Oates 2004, 2008). In the Philippine context, however, it is more nuanced: fiscal decentralization, reliant on unconditional central transfers, often fosters fiscal dependency among local governments, undermining their autonomy and hindering effective governance (Panao 2021). Here, a nuanced framing that highlights both the theoretical benefits and empirical challenges of decentralization offers a compelling narrative for publication due to its contextual depth and critical perspective.

Topic 9 covers research on urban transit planning and infrastructure, indicated by terms like "transport" ($\beta = 0.11$), "city" ($\beta = 0.10$), "local" ($\beta = 0.09$), "vaw" (violence against women) ($\beta = 0.08$), and "system" ($\beta = 0.08$), highlighting UP's critical engagement with both infrastructure development and social protection measures. Integrating "social protection" with "infrastructure development" can offer a unique and highly publishable interdisciplinary framing.

Finally, Topic 10 examines the structure and efficacy of public administration frameworks, characterized by terms like "program" ($\beta = 0.12$), "network" ($\beta = 0.10$), and "policy" ($\beta = 0.09$), reflecting the university's involvement in environmental governance and sustainable development. Studies offering neutral or analytical framings of administrative efficacy are foundational to public policy scholarship and consistently find publication venues.

Figure 4. Top terms from LDA topic modeling from the Public Policy Corpus for University of the Philippines. Source: Scopus Database.



Note: The figures were generated in *R*.

The diversity of these topics illustrates UP's interdisciplinary approach to addressing a wide array of national and local challenges. Governance, sustainability, and social justice emerge as recurring themes, with scholars focusing on issues ranging from Indigenous rights and urban development to environmental conservation and gender equity. This broad engagement highlights UP's central role in contributing to policy discussions and fostering societal change across multiple sectors, with each topic

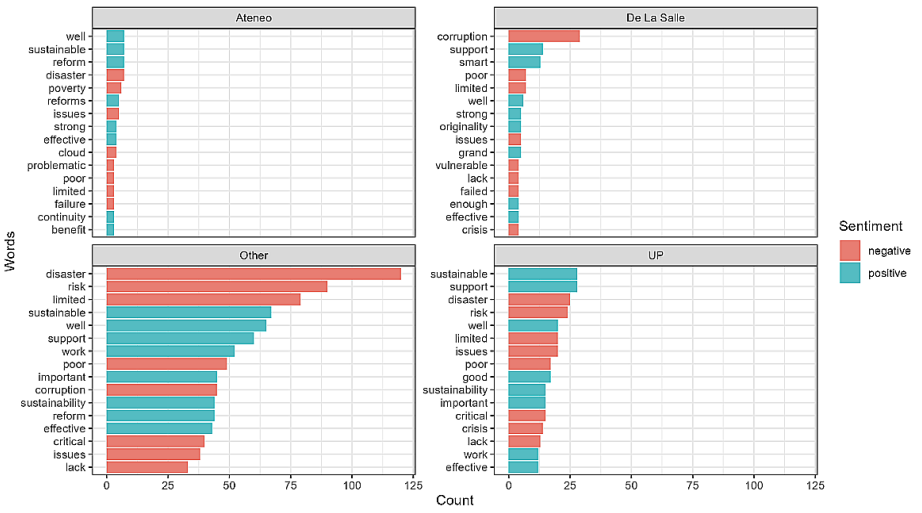
benefiting from strategic framing—emphasizing successes, identified problems, or balanced analyses—to maximize publication prospects.

The Link Between Research Tone and Visibility

This section analyzes the tone of published articles within the public policy corpus using sentiment analysis. To compare the University of the Philippines with other universities, we conducted separate sentiment analyses for each institutional affiliation. In this approach, all universities, excluding the top three in the Philippines, were aggregated into an "Other" category for comparison. The observed sentiment in published work often reflects a deliberate framing strategy, directly influencing its acceptance by different journals and its appeal to various audiences.

Figure 5 highlights the most frequent positive and negative words associated with public policy or governance discussions, illustrating institutional approaches to these topics. UP's discourse exhibits a balanced sentiment. Prominent positive words such as "sustainable," "support," "well," and "good" suggest that UP's discussions emphasize sustainability, support for reforms, and optimism regarding effective governance. Their repeated appearance indicates UP actively promotes long-term development goals and societal reforms. This predominantly positive or solution-oriented framing likely enhances the publishability of UP's research, particularly in outlets seeking constructive policy contributions. Negative words like "disaster," "risk," "issues," and "crisis" appear less frequently, reflecting critical evaluations and an acknowledgment of both successes and shortcomings in governance. Overall, UP's sentiment is slightly more positive, demonstrating a balanced yet progressive stance on public policy.

Figure 5. Comparative sentiment analysis of top universities in public policy discourse in the Philippines. Source: Scopus Database.



Note: The figures were generated in *R*.

In contrast, Ateneo presents a discourse with mixed positive and negative sentiments. "Well," "sustainable," and "effective" dominate the positive side, indicating a focus on effective governance and sustainability. However, negative terms such as "problematic," "poor," and "failure" suggest Ateneo also critiques policy inefficiencies and failures. This balanced approach, presenting both achievements and critical assessments, allows Ateneo to target a broader range of journals, from policy-oriented to more critical academic ones.

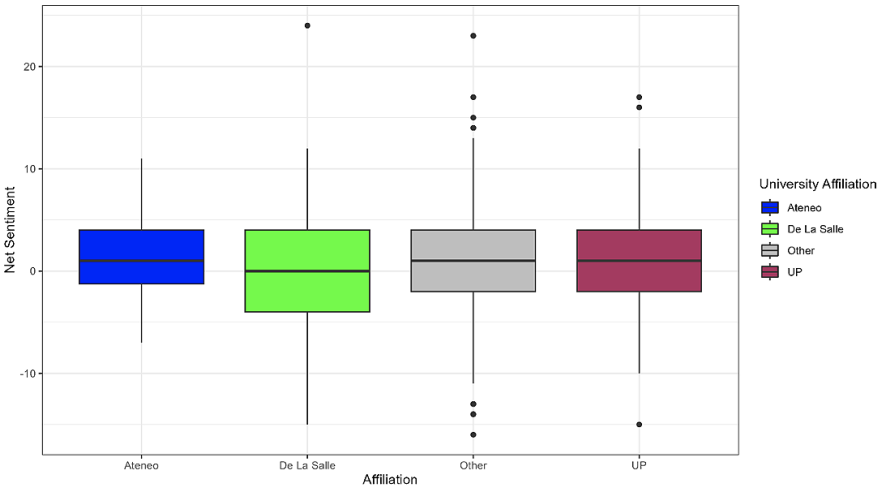
De La Salle's discourse, conversely, leans more critically, with negative terms such as "corruption," "poor," and "limited" standing out. This suggests their discussions heavily focus on governance failures, particularly issues like corruption and ineffective policies. While positive words like "support" and "effective" appear, they are less frequent, indicating a more critical tone than other universities. Research with a predominantly critical or "problem-focused" framing, like DLSU's, often finds a niche in journals emphasizing accountability or systemic issues, though it might face more scrutiny from outlets preferring policy solutions.

The "Other" institutions generally display a more negative sentiment, with "disaster," "risk," and "poor" being the most frequent terms. This suggests a significant focus on risks, policy failures, and governance issues. However, positive terms like "sustainable" and "well" indicate they also contribute to discussions on public policy improvement, albeit from a more critical standpoint. While a negative framing can highlight pressing issues, it might also influence target publication venues, favoring those prioritizing problem identification over proposed solutions.

Sentiment analysis provides key insights into how different universities approach public policy discussions. UP stands out for its balanced engagement, focusing on reforms and sustainability while critically addressing challenges. This suggests UP's leadership in public policy discourse, emphasizing long-term solutions while monitoring governance risks. Ateneo offers a balanced discourse with a slight emphasis on policy effectiveness, while De La Salle is heavily critical, focusing on corruption. The "Other" category shows a predominantly negative sentiment, highlighting risks and failures.

Interestingly, comparing net sentiment scores (positive minus negative) across affiliations (see figure 6) reveals minimal variation in median scores. DLSU and the "Other" category show slightly higher median net sentiments compared to Ateneo and UP. However, the "Other" category exhibits a wider score distribution, reflected by a larger interquartile range and notable outliers.

Figure 6. Comparative net sentiment of the public policy discourse in the Philippines by affiliation. Source: Scopus Database.



Note: The figures were generated in *R*.

This broader distribution likely stems from the diverse affiliations within this group, encompassing various disciplines and global issues. This suggests that diverse framings—from highly positive to highly negative—can all find publication, provided they are well-argued and empirically grounded.

Including institutions with diverse research agendas like engineering and social sciences introduces more sentiment variation. Research with practical or urgent policy implications (e.g., public health, environmental studies) may evoke stronger positive or negative reactions based on the issues addressed. This variability not only reflects the multidisciplinary nature of the "Other" category but also suggests different fields can trigger distinct emotional responses, influencing the overall sentiment landscape and, by extension, the strategic framing authors employ to maximize their work's appeal to specific journals and audiences.

Insights from sentiment analysis highlight universities' critical role in shaping public discourse, driving policy innovation, and addressing complex societal challenges. The variation (or lack thereof in net scores) suggests diverse university approaches focusing on governance, social equity, and environmental sustainability, underscoring the need for a multifaceted strategy for progress. Ultimately, the choice of research tone and framing—whether positive (solution-oriented), negative (problem-identifying), or neutral (analytical)—is a strategic decision that can profoundly impact a study's likelihood of publication and its resonance within academic and policy communities.

Conclusion

This study advances a critical reflection on the politics of academic visibility by demonstrating that research tone—far from being a mere stylistic choice—constitutes a strategic axis around which publication likelihood is conditioned. Drawing on a corpus of 720 peer-reviewed public policy articles authored by scholars from the University of the Philippines, our analysis reveals that works adopting a constructive or reform-oriented tone are more likely to be published, while those framed in predominantly critical or negative terms appear less frequently. This pattern holds even for studies addressing structurally important but politically sensitive topics such as corruption, inequality, or state capacity, suggesting that evaluative tone acts as a tacit filter in the dissemination of policy-relevant knowledge.

Importantly, this tendency does not necessarily reflect explicit editorial bias but rather mirrors the broader institutional and political context in which UP operates—a context in which universities are expected to support national development goals while maintaining a critical public voice. The findings thus challenge the assumption of content-neutral publication processes and underscore the subtle forms of gatekeeping that govern academic inclusion. UP's dual commitment to public service and scholarly excellence positions its researchers at the intersection of critique and collaboration, where tone becomes a navigational tool for balancing intellectual independence with institutional legitimacy.

By foregrounding tone as an analytic category, this paper contributes to the sociology of knowledge and the study of academic labor in the global South. It offers an empirical basis for understanding how scholars respond to the imperatives of publishability without surrendering critical engagement. Future research should extend this inquiry through inferential methods and comparative frameworks to determine whether these dynamics hold across other institutions and policy domains. Ultimately, making visible the invisible politics of tone invites renewed attention to how ideas circulate—and which ones are allowed to matter—in the evolving architectures of academic and policy discourse.

References

Arora, Vineet M., Charlie M. Wray, Avital Y. O'Glasser, Mark Shapiro, and Shikha Jain. 2021. "Leveling the Playing Field: Accounting for Academic Productivity during the COVID-19 Pandemic." *Journal of Hospital Medicine* 16 (2): 120–23. https://doi.org/10.12788/jhm.3558.

Burstein, Paul. 1998. "Bringing the Public Back In: Should Sociologists Consider the Impact of Public Opinion on Public Policy?" *Social Forces* 77 (1): 27–62. https://doi.org/10.1093/sf/77.1.27.

Canestaro, William J., Nathaniel Hendrix, Aasthaa Bansal, Sean D. Sullivan, Emily B. Devine, and Josh J. Carlson. 2017. "Favorable and Publicly Funded Studies Are More Likely to be Published: A Systematic Review and Meta-analysis." *Journal of Clinical Epidemiology* 92: 58–68. https://doi.org/10.1016/j.jclinepi.2017.08.004.

Dickersin, Kay, Yuan-I Min, and Curtis L. Meinert. 1992. "Factors Influencing Publication of Research Results: Follow-up of Applications Submitted to Two Institutional Review Boards." *JAMA* 267 (3): 374–78. https://doi.org/10.1001/jama.1992.03480030052036.

Dill, David. D. 1997. "Higher Education Markets and Public Policy." *Higher Education Policy* 10 (3): 167–85. https://doi.org/10.1016/S0952-8733(97)00011-1.

Dill, David D. 2014. "Public Policy Design and University Reform: Insights into Academic Change." In *Reforming Higher Education: Public Policy Design and Implementation*, edited by Christin Musselin and Pedro N. Teixeira, 21–37. Springer Netherlands. https://doi.org/10.1007/978-94-007-7028-7_2.

Dill, David D., and Maarja Soo. 2005. Academic Quality, League Tables, and Public Policy: A Cross-National Analysis of University Ranking Systems. *Higher Education*, 49(4), 495–533. https://doi.org/10.1007/s10734-004-1746-8.

Ellinas, Elizabeth H., Tavinder K. Ark, Kristina Kaljo, Katherine G. Quinn, Cassandre R. Krier, and Amy H. Farkas. "Winners and losers in academic productivity during the COVID-19 pandemic: is the gender gap widening for faculty?." *Journal of Women's Health* 31, no. 4 (2022): 487-494. https://doi.org/10.1089/jwh.2021.0321.

Gheorghiu, Ana I., Mitchell J. Callan, and William J. Skylark. 2017. "Facial Appearance Affects Science Communication." *Proceedings of the National Academy of Sciences* 114 (23): 5970–75. https://doi.org/10.1073/pnas.1620542114.

Harrison, John, and Ivan Turok. 2017. "Universities, Knowledge and Regional Development." *Regional Studies* 51 (7): 977–81. https://doi.org/10.1080/00343404.2017.1328189.

Kochenkova, Anna, Rosa Grimaldi, and Federico Munari. 2016. "Public Policy Measures in Support of Knowledge Transfer Activities: A Review of Academic Literature." *Journal of Technology Transfer* 41 (3): 407–29. https://doi.org/10.1007/s10961-015-9416-9.

Krukowski, Rebecca A., Reshma Jagsi, and Michelle I. Cardel. "Academic productivity differences by gender and child age in science, technology, engineering, mathematics, and medicine faculty during the COVID-19 pandemic." *Journal of Women's Health* 30, no. 3 (2021): 341-347. https://doi.org/10.1089/jwh.2020.8710.

Milkman, Katherine L., and Jonah Berger. 2014. "The Science of Sharing and the Sharing of Science." *Proceedings of the National Academy of Sciences* 111 (supplement_4): 13642–49. https://doi.org/10.1073/pnas.1317511111.

Saad, Mohammed, and Gima Zawdie. 2011. "Introduction to Special Issue: The Emerging Role of Universities in Socio-economic Development through Knowledge Networking." *Science and Public Policy* 38 (1): 3–6. https://doi.org/10.3152/030234211X12960315267453.

Staniscuaski, Fernanda, Livia Kmetzsch, Rossana C. Soletti, Fernanda Reichert, Eugenia Zandonà, Zelia MC Ludwig, Eliade F. Lima et al. "Gender, race and parenthood impact academic productivity during the COVID-19 pandemic: from survey to action." *Frontiers in psychology* 12 (2021): 663252. 10.3389/fpsyg.2021.663252.

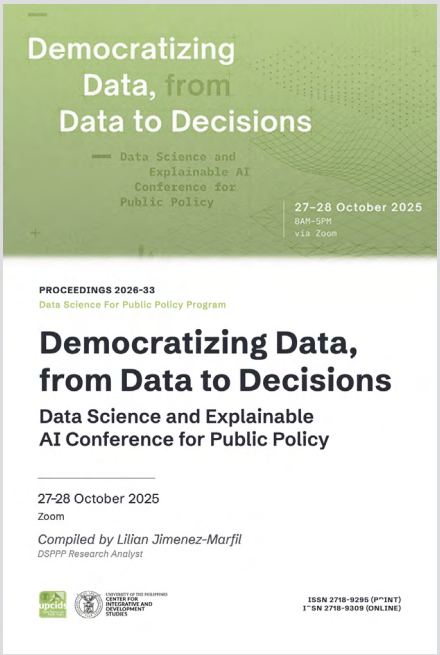
St. John, Edward P., Nathan J. Daun-Barnett, and Karen M. Moronski-Chapman. 2018. *Public Policy and Higher Education: Reframing Strategies for Preparation, Access, and College Success*. Routledge. https://doi.org/10.4324/9781315622453.

Trippl, Michaela, Tanja Sinozic, and Helen Lawton Smith. 2015. "The Role of Universities in Regional Development: Conceptual Models and Policy Institutions in the UK, Sweden and Austria." *European Planning Studies* 23 (9): 1722–40. https://doi.org/10.1080/09654313.2015.1052782.

Uyarra, Elvira. 2010. "Conceptualizing the Regional Roles of Universities, Implications and Contradictions." *European Planning Studies* 18 (8): 1227–46. https://doi.org/10.1080/09654311003791275.

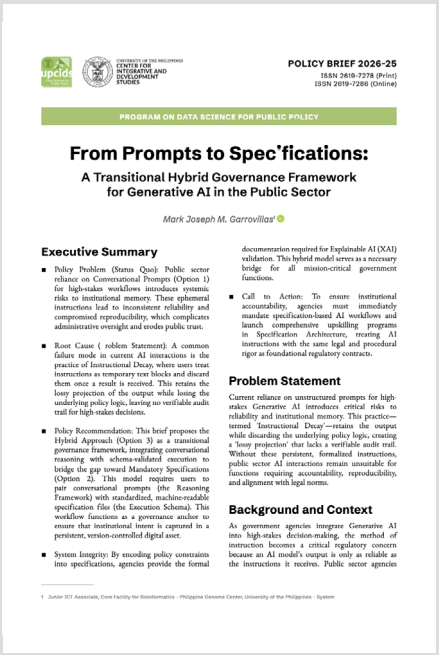
Download related policy papers for FREE

cids.up.edu.ph/collection-database



PROCEEDINGS

Democratizing Data, From Data to Decisions: Data Science and Explainable AI for Public Policy



POLICY BRIEF

From Prompts to Specifications: A Transitional Hybrid Governance Framework for Generative AI in the Public Sector

THE UP CIDS MONOGRAPH SERIES

Monographs contain findings on issues that are aligned with the core agenda of the research programs under the University of the Philippines Center for Integrative and Development Studies (UP CIDS).

Center For Integrative and Development Studies

Established in 1985 by University of the Philippines (UP) President Edgardo J. Angara, the UP Center for Integrative and Development Studies (UP CIDS) is the policy research unit of the University that connects disciplines and scholars across the several units of the UP System. It is mandated to encourage collaborative and rigorous research addressing issues of national significance by supporting scholars and securing funding, enabling them to produce outputs and recommendations for public policy.

The UP CIDS currently has twelve research programs that are clustered under the areas of education and capacity building, development, and social, political, and cultural studies. It publishes policy briefs, monographs, webinar/conference/forum proceedings, and the Philippine Journal for Public Policy, all of which can be downloaded free from the UP CIDS website.

The Program

The Program on Data Science for Public Policy (DSPPP) aims to build the data science knowledge and capacities of academics, researchers, and policymakers, as well as, decision-makers from various sectoral stakeholders, and apply this learned skill to public policy and governance. DSPPP strives to engage a community of researchers within the university and encourage the pursuit of interdisciplinary problem-oriented research using high-level quantitative analysis. It seeks to convene multidisciplinary teams of social scientists, humanists, and scientists to research issues in the public sector.

Editorial Board

Rosalie Arcala Hall
Editor-in-Chief

Honeylet L. Alerta
Deputy Editor-in-Chief

Program Editors

Education and Capacity Building Cluster

Dina S. Ocampo
Lorina Y. Calingasan
Education Research Program

Rosalie Arcala Hall
Program on Higher Education Research and Policy Reform

Romylyn Metila
Marlene Ferido
Assessment, Curriculum, and Technology Research Program

Ebinezer R. Florano
Program on Data Science for Public Policy

Social, Political, and Cultural Studies Cluster

Rogelio Alicor L. Panao
Program on Social and Political Change

Darwin J. Absari
Islamic Studies Program

Rosalie Arcala Hall
Strategic Studies Program

Development Cluster

Annette O. Balaoing-Pelkmans
Program on Escaping the Middle-Income Trap: Chains for Change

Antoinette R. Raquiza
Julius Lustro
Political Economy Program

Eduardo C. Tadem
Maria Dulce Natividad
Program on Alternative Development

Iris Thiele Isip-Tan
Program on Health Systems Development

New Programs

Maria Angeles O. Catelo
Food Security Program

Weena S. Gera
Urban Studies Program

Benjamin M. Vallejo, Jr.
Conservation and Biodiversity

Rosalie Arcala Hall
Local and Regional Studies Network

Editorial Staff

Jheimeel P. Valencia
Bryan Patrick Garcia
Copyeditor

Alexa Samantha R. Hernandez
Editorial Assistant

Mikaela Anna Cheska D. Orlino
Layout Artist

Get your policy papers published. Download open-access articles.

The Philippine Journal of Public Policy: Interdisciplinary Development Perspectives (PJPP), the annual peer-reviewed journal of the UP Center for Integrative and Development Studies (UP CIDS), welcomes submissions in the form of full-length policy-oriented manuscripts, book reviews, essays, and commentaries. The PJPP provides a multidisciplinary forum for examining contemporary social, cultural, economic, and political issues in the Philippines and elsewhere. Submissions are welcome year-round.

For more information, visit cids.up.edu.ph.
All issues/articles of the PJPP can be downloaded for free.

Get news and the latest publications.

Join our mailing list to get our publications delivered straight to your inbox! Also, you'll receive news of upcoming webinars and other updates.

bit.ly/signup_cids

We need your feedback.

Have our publications been useful? Tell us what you think.

bit.ly/dearcids

ISSN 2719-0722

UNIVERSITY OF THE PHILIPPINES
CENTER FOR INTEGRATIVE AND DEVELOPMENT STUDIES
Lower Ground Floor, Ang Bahay ng Alumni, Magsaysay Avenue
University of the Philippines Diliman, Quezon City 1101

Telephone (02) 8981-8500 loc. 4266 to 4268 (02) 8426-0955
Email cids@up.edu.ph cidspublications@up.edu.ph
Website cids.up.edu.ph