Prominence of occurrence accorded to data science and other related terms in Philippine newspapers

Dan Anthony Dorado

2023-12-05

In the contemporary era marked by the digital revolution, data science has emerged as a pivotal discipline globally. This study examines the prominence and frequency of data science and its associated terminologies within Philippine newspapers. By analyzing a comprehensive selection of major Philippine dailies over 20 years, we assessed how often and in what context these terms appear. Our findings indicate a steadily increasing trend, reflecting the growing significance and understanding of data science in the country. Notably, the words were most frequently associated with business, technology, and academic news sections. However, there was variability in the depth and accuracy of discussions surrounding these terms. The media's portrayal, in many instances, was foundational, suggesting a burgeoning stage of familiarity with data science concepts among the general populace. This study underscores the role of print media in disseminating and shaping perceptions about emerging technological fields in the Philippines. It offers insights into potential areas for enhanced media literacy and collaboration.

Introduction

Background of the study

Global rise of data science and its influence.

Data science has become increasingly important in decision-making processes across industries. The advancements in data volume and computational capabilities have played a significant role in the growth and application of data science [1]. This growth has led to the recognition of data science as a crucial discipline in driving decision-making. Companies have realized the need to hire data scientists, and academic institutions now offer data science programs [2].

1

The application of data science techniques has also expanded to fields such as computer and network security [3]. In sustainable development, data science has become crucial in meeting human development needs while protecting the earth's life support systems [4]. By utilizing knowledge systems, scientists, technologists, policymakers, and communities at all levels can work together to find sustainable solutions. Data science has also made a significant impact on global health. Efforts by organizations like Sana (MIT) have played a vital role in promoting data science in global health [5]. The use of health information technology as a premise for data science in global health has opened up opportunities for improving healthcare outcomes worldwide. In asthma research, data science has helped understand the worldwide prevalence of asthma symptoms. Studies have shown that the global burden of asthma is continuing to rise, with increasing prevalence in Africa, Latin America, and parts of Asia [6]. However, it is essential to note that there has been a lack of new worldwide data on asthma prevalence since 2003 [7]. The global macroeconomic impacts of COVID-19 have also been analyzed using data science techniques. Different scenarios have been examined to understand their effects on macroeconomic outcomes and financial markets [8]. This analysis provides valuable insights for policymakers and economists in navigating the challenges posed by the pandemic. In glaucoma research, data science has been used to analyze the distribution pattern and subject domain knowledge of worldwide glaucoma research [9]. This scientometric review provides a comprehensive understanding of the research landscape in this area. Scholarly production analysis has also benefited from data science methodologies. Studies have compared the performance of databases such as Google Scholar, Scopus, and Web of Science regarding citation analysis and cross-disciplinary coverage [10]. These analyses contribute to understanding the impact and reach of scholarly publications worldwide.

Overall, the global rise of data science has had a transformative influence on various fields. It has enabled researchers, policymakers, and practitioners to gain valuable insights, make informed decisions, and address complex challenges in a data-driven manner. However, it is essential to note that the emphasis on data science education primarily focuses on computer science and statistics, with limited training in ethics and domain-specific knowledge [11]. This highlights the need for a more comprehensive approach to data science education, including ethics and domain-specific training. Additionally, the application of data-driven research methods has extended beyond computer science to disciplines such as social sciences and humanities [12].

Importance of media in disseminating new terminologies and concepts.

The media is crucial in disseminating new terminologies and concepts to the public. It serves as a platform for introducing and explaining unfamiliar terms and ideas, making them accessible and understandable to a broad audience. A critical aspect of the media's role in disseminating new terminologies and concepts is its ability to reach a large audience. The mass media, including mainstream newspapers, television, and online news outlets, have a broad reach and can effectively communicate new terminologies and concepts to diverse people [13]. This

widespread dissemination helps to increase awareness and understanding of these new terms and ideas. Furthermore, the media can shape public discourse and influence public opinion. Through news articles, opinion pieces, and other media content, the media can introduce and promote new terminologies and concepts, framing them in a way that resonates with the audience [14]. This framing can help to shape public perceptions and attitudes towards these new terms and ideas. In addition, the media's role in disseminating new terminologies and concepts extends to specialized fields such as healthcare and medicine. Health journalism, for example, plays a crucial role in sharing health information with the public [15]. The media can introduce new medical terminologies and concepts, helping to educate the public about important health issues and advancements in medical research.

However, it is essential to note that the media's role in disseminating new terminologies and concepts is not without challenges. One challenge is the potential for the media to misinterpret or misrepresent these new terms and ideas. The media's framing of new terminologies and concepts can sometimes be influenced by various factors, such as political or economic interests [13]. This can distort the original meaning or intent of these terms and ideas. Another challenge is the potential for the media to perpetuate stereotypes or biases through new terminologies and concepts. For example, in the context of Islam-related vocabularies, the media's use of specific terms can influence the audience's perception of Islam as a religion of terrorism [16]. This highlights the importance of responsible and accurate reporting when introducing and disseminating new terminologies and concepts.

Objective of the research

Data science and other related terms have emerged as a crucial field in the era of big data and advanced analytics. It encompasses various techniques and methodologies for extracting insights and knowledge from large datasets. The prominence of data science in the Philippines has been a topic of interest, as it plays a significant role in driving innovation, economic growth, and decision-making processes. This research objective aims to assess the frequency and prominence of data science and related terms in Philippine newspapers.

Literature Review

Data science: Definition and global trends

Data science is a field dedicated to the principled extraction of knowledge from complex data [17]. It combines principles from statistics, which focuses on learning relationships from data, and computer science, which emphasizes efficient computing algorithms [18]. Data science aims to extract generalizable knowledge from data [19]. It is a rapidly growing field becoming increasingly important in various domains, including critical care in medicine [17]. Data science is closely related to other concepts, such as big data and data-driven decision-making [2]. It

is intricately intertwined with these concepts and plays a crucial role in analyzing and making sense of large volumes of data [2]. Data science is often used to rapidly derive knowledge and discover trends and novel research topics from vast literature [20].

While data science builds on knowledge from various disciplines, such as computer science, mathematics, and statistics, it is a unique field with challenges and mysteries to unlock [21]. It is establishing itself as a profession and has the potential to make significant contributions to society [22]. Data science is also characterized by its generic and vaguely defined nature [23]. There is no universally accepted definition of data science, and different perspectives exist. Some define data science as the study of the generalizable extraction of knowledge from data [19]. In contrast, others emphasize its interdisciplinary nature and intersection with statistics, computer science, and substantive fields [24].

Media's role in popularizing technical jargon and scientific terms

The role of media in introducing and normalizing new terminologies and concepts to the public is pivotal in today's rapidly evolving world. As new scientific and technological terms such as 'data science', 'big data', and 'artificial intelligence' emerge, media outlets serve as the primary source for the public's initial exposure to these concepts [25]. Moreover, the media is crucial in providing context and explanation for these complex terms, making them accessible and understandable to a non-specialist audience. This involves decoding technical jargon and presenting information using metaphors, analogies, and simplifications to convey complex ideas [26].

The influence of media in shaping public understanding extends beyond the introduction of new concepts. It also encompasses disseminating knowledge through social media platforms, rapidly gaining popularity worldwide. The speed of uptake of social media applications highlights their significance as tools for knowledge sharing, further emphasizing the central role of media in disseminating information [27]. Additionally, the Internet and social media development has expanded the speed and scope of information dissemination. However, not all widely published information is accurate, emphasizing the need to critically evaluate information sources [28].

Furthermore, the media's influence on public perception is evident in the context of scientific and technological advancements. The press operates at the interface between science and society, reporting on scientific advances and technological developments in ways that shape public perceptions of new technologies and their applications. This underscores the media's role in framing and agenda setting, particularly about novel technologies, influencing public views and perceptions [26].

Theoretical Framework

Agenda-setting theory is a well-established and influential theoretical framework in communication research. It focuses on the role of the media in shaping public opinion and influencing

the salience of societal issues. According to the theory, the media can determine which issues are important and worthy of public attention, thus setting the agenda for public discourse [29].

A critical aspect of agenda-setting theory is the transfer of salience, which refers to the media's ability to determine which issues are important and shape how people think about those issues. This salience transfer occurs through the media's selection and emphasis on specific topics, influencing the public's perception and understanding of those issues [29]. Furthermore, the media's agenda-setting power extends not only to the salience of the problems but also to the salience of their attributes [30]. In other words, the media can shape how people think about the characteristics and qualities of specific issues.

Agenda-setting research has evolved over the years, with scholars refining and expanding the scope of the theory. For example, researchers have explored intermedia agenda-setting effects, which examine how different media outlets influence each other in shaping the public agenda. This expansion has led to a more sophisticated understanding of agenda-setting effects and their impact on public opinion [31].

Moreover, agenda-setting theory has applications in various fields, including politics, journalism, and public relations. It has been used to study the influence of the media on political campaigns, the framing of news stories, and the communication of consent [32–34]. The theory has also been examined in online news media, where researchers have analyzed agenda-setting effects among different platforms [31].

In recent years, there has been a growing interest in meta-analyses of agenda-setting research, which provide a comprehensive overview of the existing empirical literature. These meta-analyses analyze the evolution of agenda-setting theory in terms of research focus, subject areas, and research methods. By synthesizing and analyzing statistical indicators from many studies, these meta-analyses contribute to a deeper understanding of the current state of agenda-setting research [35].

Methodology

Selection of newspapers

The selection of newspapers for a study focused on data science-related terms is crucial for comprehensive coverage and depth of analysis. This research has chosen the Philippine Daily Inquirer website for several reasons. Firstly, the Philippine Daily Inquirer is one of the most widely read newspapers in the Philippines, ensuring the analysis is relevant to a broad online audience. Additionally, the extensive archives of the newspaper's online version are essential for historical data analysis, allowing for an exploration of the evolution of data science terminology over time. The diverse array of articles published on the website, including news, opinion pieces, features, and editorials, is vital for understanding the different contexts in

which data science-related terms are used. As a continually updated digital news source, the Philippine Daily Inquirer website offers the most current perspective on using and understanding data science terms. Moreover, the newspaper's national scope reflects national interests and discourse, providing a comprehensive overview of how data science is perceived and discussed nationwide. The editorial stance and policies of the newspaper provide insight into how data science topics are framed and prioritized, which is an essential aspect of media analysis.

Time frame of the study

The time frame for the study, spanning from 1993 to 2023, is crucial for capturing the evolution of data science in the Philippine media landscape. This extended period allows for identifying long-term trends and patterns in media coverage related to data science [36]. Additionally, the chosen period should align with significant events or developments in the field of data science, such as the introduction of data science courses, government initiatives, or the establishment of data science firms in the Philippines.

The start and end dates of the data collection period should be justified based on the emergence of critical events or milestones in the Philippine data science landscape [37]. Furthermore, dividing the time frame into segments, such as monthly or quarterly intervals, allows for a detailed temporal analysis, enabling the examination of changes in reporting frequency or contextual shifts around specific events.

The selected time frame may influence the results and the interpretation of trends in the data, as it allows for the assessment of any immediate effects of significant global or local events on the presence and representation of data science in the press. Considerations for seasonal variations or particular circumstances, such as elections or natural disasters, should also be factored in, as they might affect media coverage and public interest in data science.

If prior data is available, a comparative analysis with previous periods can provide valuable insights into the growth, decline, or cyclical patterns in the coverage of data science-related terms. Anticipating the impact of the time frame on the results is essential, considering how seasonal variations or specific events, such as elections or natural disasters, might influence media coverage and public interest in data science.

Data Collection Process

This process begins with the careful selection of primary and secondary keywords, such as "**Data Science**", "**Big Data**", "**Analytics**", "**Artificial Intelligence**", along with related terms like "**Robotics**". The search for these keywords is facilitated by using the search textbook of https://www.inquirer.net/. A manual review process is implemented to validate the accuracy and relevance of the automated search results. This involves random checks to ensure articles are contextually related to data science and its associated fields. Moreover, the

entire process is conducted with strict adherence to ethical and legal considerations, ensuring compliance with copyright laws and respect for privacy and ethical guidelines, especially when handling articles that include interviews or personal data. Through these comprehensive keyword search techniques, the study aims to gather a wide range of data from the Philippine Daily Inquirer, offering detailed insights into the media's portrayal and emphasis on data science and related terms.

Findings

Computational Environment and Software

In this study, we employed a robust computational environment to ensure the reliability and reproducibility of our data processing and analysis. Our analysis pipeline was constructed with a focus on using widely-recognized, open-source software to enhance the accessibility and verifiability of our methodologies.

The primary statistical analyses and data processing tasks were conducted using `R` [38], a language and environment specifically designed for statistical computing and graphics. The R environment is renowned for its extensive package ecosystem and ability to handle various data types and analytical procedures. For this study, we used `R` version 4.3.0, ensuring that all analyses benefit from the latest improvements and bug fixes.

Complementing `R`, we utilized `RStudio` [39], an integrated development environment (IDE) that provides a user-friendly interface for `R`. `RStudio` version 2023.06.1+524 was employed, which offers enhanced coding tools, debugging capabilities, and data visualization features. The combination of `R` and `RStudio` provided a versatile and efficient platform for coding, data manipulation, and analysis.

The `rvest` package [40] in `R` was utilized for web scraping and HTML data extraction. This package simplifies the process of reading and manipulating HTML content, making it invaluable for gathering data from web sources. Version 1.0.3 of rvest was explicitly used.

Data manipulation and transformation tasks were streamlined using the `tidyverse` collection of `R` packages [41]. This suite of packages, including `dplyr`, `tidyr`, and `ggplot2`, facilitates efficient data processing and visualization. We specifically employed `tidyverse` version 2.0.0.

The primary operating system used for all data processing tasks was `Windows 11 x64 (build 22621)`. This version of Windows offers enhanced security features, improved performance, and a user-friendly interface, which significantly contributed to the efficiency of our data handling processes. The 64-bit architecture was particularly beneficial for handling large datasets, ensuring memory-intensive operations were executed smoothly.

All scripts and processed data files have been made publicly available for transparency and to facilitate the reproducibility of our findings. They are hosted on GitHub, a widely recognized

version control and code-sharing platform. The repository can be accessed at https://github.com/dddorado/ds_term_prominence. This repository includes all the scripts used in the data processing and analysis phases, along with detailed documentation on the usage and functionality of these scripts [42].

Data Scrapping Process

In this study, we aimed to systematically gather data from https://www.inquirer.net/, focusing on identifying and downloading web pages that contain specific keywords of interest. This task was necessary due to the nature of the website's content presentation. Like many contemporary news and information websites, Inquirer.net leverages modern web technologies to enhance user experience and content delivery. A vital aspect of these technologies is the use of `JavaScript` to load and display content dynamically. This approach, commonly employed in web applications developed with frameworks like `Angular`, `React`, or `Vue`, poses a specific challenge for web scraping tools like `rvest`. Once the pages are downloaded, we can proceed to harvest its content.

Running the provided code in R will load the rvest and tidyverse packages into your R session.

```
library(rvest)
library(tidyverse)
library(lubridate)
library(tm)
library(tidytext)
```

The code starts by creating a vector `robotics_url` containing a URL list. Each URL points to an HTML page that presumably includes robotics-related articles.

```
robotics_url <- c('robotics_articles/robotics01.html',
                  'robotics_articles/robotics02.html',
                  'robotics_articles/robotics03.html',
                  'robotics_articles/robotics04.html',
                  'robotics_articles/robotics05.html',
                  'robotics_articles/robotics06.html',
                  'robotics_articles/robotics07.html',
                  'robotics_articles/robotics08.html',
                  'robotics_articles/robotics09.html',
                  'robotics_articles/robotics10.html')
```

An `extractor` function is defined to process each URL. This function uses `read_html` from the `rvest` package to parse the URL's HTML content. It then uses a series of `rvest` functions

to extract specific data:

- `html_element('div.gsc-expansionArea')`: Selects a division in the HTML document with the class `gsc-expansionArea`.
- `html_nodes('a.gs-title')`: From this division, it finds all hyperlinks (`<a>` tags) with the class `gs-title`, extracting the text from these nodes to get the titles of the articles.
- `html_nodes('div.gs-bidi-start-align.gs-snippet')`: Similarly, it extracts text snippets from divisions with the classes `gs-bidi-start-align` and `gs-snippet`, which likely contain article summaries or content.
- The `unique()` function is applied to the titles to ensure no duplicates.
- The function returns a `tibble` (a type of data frame in R) with two columns: `title` and `data`, containing the extracted titles and text snippets, respectively.

```r
extractor <- function(url){
    page <- read_html(url)
    title <- page %>% html_element('div.gsc-expansionArea') %>%
      html_nodes('a.gs-title') %>%
      html_text() %>%
      unique()
    data <- page %>%
      html_element('div.gsc-expansionArea') %>%
      html_nodes('div.gs-bidi-start-align.gs-snippet') %>%
      html_text()
    tibble(title = title, data = data)
}
```

The `lapply` function applies the `extractor` function to each URL in the `robotics_url` vector. This results in a list of tibbles, where each tibble contains data scraped from one URL.

```r
robotics_scraped <- lapply(robotics_url, extractor)
```

The scraped data from all URLs are combined into a single tibble using `do.call` with `rbind`, which row-binds the individual tibbles. This creates a consolidated dataset.

```r
robotics_combined_data <- do.call(rbind, robotics_scraped)
```

Finally, the code adds a new column named `category` to the combined tibble, assigning the value 'robotics' to each row. This categorization step indicates that all scraped data in this dataset are related to robotics.

```r
robotics_combined_data <- robotics_combined_data %>%
    mutate(category = 'robotics')
```

Repeat the process for other downloaded websites. Combine the different datasets (`robotics_combined_data`, `ai_combined_data`, `analytics_combined_data`, `big_data_combined_data`, `data_science_combined_data`) into a single dataset using the `rbind` function. This creates a comprehensive dataset (`compiled_dataset`) containing data from all these categories.

```
compiled_dataset <- rbind(robotics_combined_data,
                          ai_combined_data,
                          analytics_combined_data,
                          big_data_combined_data,
                          data_science_combined_data)
```

It then processes the `data` column to extract dates using `str_extract` with a regex pattern that matches date formats (e.g., "Jan 1, 2020"). These extracted dates are then parsed into a standardized format using `parse_date_time`.

```
compiled_dataset <- compiled_dataset %>%
    mutate(date = str_extract(data, "\\b[A-Z][a-z]{2} \\d{1,2}, \\d{4}\\b"))
compiled_dataset$date <- parse_date_time(compiled_dataset$date, orders = "mdY")
```

Frequency and Temporal Analysis

```
additional_stopwords <- c("ai",
                          "inquirer",
                          "artificial",
                          "intelligence",
                          "analytics",
                          "ph",
                          "data",
                          "robotics",
                          "science",
                          "inquirernet")
all_stopwords <- c(stopwords("en"), additional_stopwords)
```

We started with a script that sets up a list of additional stopwords specific to the context (like "ai", "inquirer", etc.) and combines them with the default English stopwords provided by the `tm` package.

A `clean_text` function is then defined to preprocess the text. This function converts text to lowercase, removes punctuation and stopwords, and strips extra whitespace. This function is applied to the `title` column of the dataset.

```r
clean_text <- function(text) {
    text <- tolower(text) # Convert to lowercase
    text <- removePunctuation(text) # Remove punctuation
    text <- removeWords(text, all_stopwords) # Remove stopwords
    text <- stripWhitespace(text) # Remove extra white spaces
    return(text)
}

compiled_dataset$clean_title <- sapply(compiled_dataset$title, clean_text)
```

We then tokenizes the cleaned titles and calculate term frequencies. It uses `unnest_tokens` to break down the titles into words and then counts the frequency of each word within each category.

```r
compiled_dataset_clean <- compiled_dataset %>%
    unnest_tokens(word, clean_title) %>%
    count(category, word, sort = TRUE)
```

```
Warning: Outer names are only allowed for unnamed scalar atomic inputs
```

We select the top 10 terms for each category based on their frequency using `top_n`.

```r
top_terms <- compiled_dataset_clean %>%
    group_by(category) %>%
    top_n(10, n) %>%
    ungroup()
```

Finally, we use `ggplot2` for data visualization. It creates a bar plot showing the top 10 terms for each category, with words on the x-axis, frequency on the y-axis, and different colors for each type. The plot uses `facet_wrap` to create separate plots for each category and `coord_flip` for better readability of the terms.

```r
ggplot(top_terms, aes(x = reorder(word, n), y = n, fill = category)) +
    geom_bar(stat = "identity") +
    facet_wrap(~ category, scales = "free") +
    coord_flip() +
    labs(title = "Top 10 Terms in Titles per Category", x = "Terms", y = "Frequency") +
    theme_minimal()
```

Top 10 Terms in Titles per Category
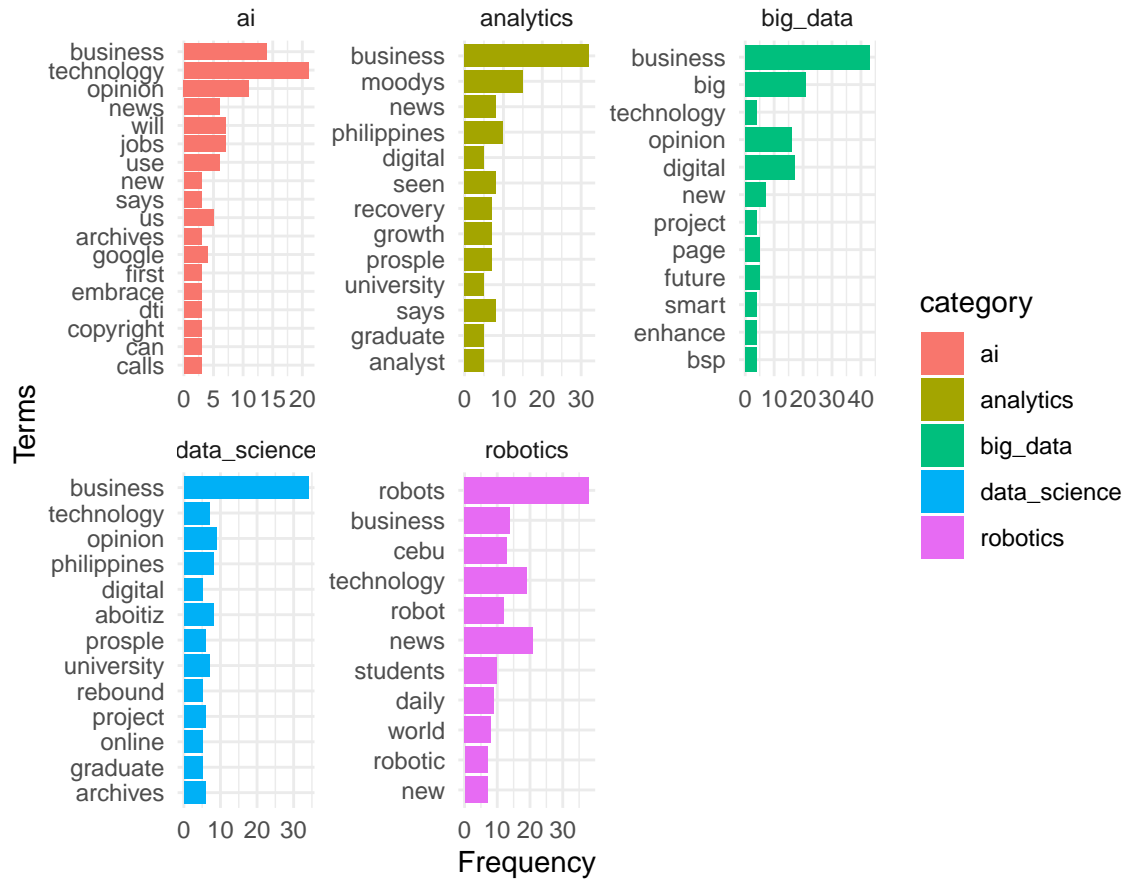
We then perform a temporal analysis of the occurrence of news articles related to the keywords across the years.

```
compiled_dataset$year <- year(compiled_dataset$date)
```

The code starts by manipulating the `compiled_dataset`, specifically by adding a new column named `year`. This column is created by extracting the year from the `date` column in the dataset. The `year` function from the `lubridate` package, is used for this extraction. This step is crucial for temporal analysis, as it allows the dataset to be analyzed on a yearly basis.

```
yearly_data <- compiled_dataset %>%
    group_by(year, category) %>%
    summarise(count = n())
```

```
`summarise()` has grouped output by 'year'. You can override using the
`.groups` argument.
```

The first part of the code uses `dplyr` for data processing. The `group_by(year, category)` function groups the `compiled_dataset` by two columns: `year` and `category`. This means the dataset is partitioned into subsets based on the combination of year and category. The `summarise(count = n())` function then calculates the count (`n()`) of rows (articles) in each group, effectively giving the number of articles per category for each year. The result of this operation is stored in `yearly_data`.

```
ggplot(yearly_data, aes(x = year, y = count, group = category, color = category)) +
    geom_line() +
    labs(title = "Yearly Temporal Analysis by Category", x = "Year", y = "Count") +
    theme_minimal()
```

```
Warning: Removed 5 rows containing missing values (`geom_line()`).
```
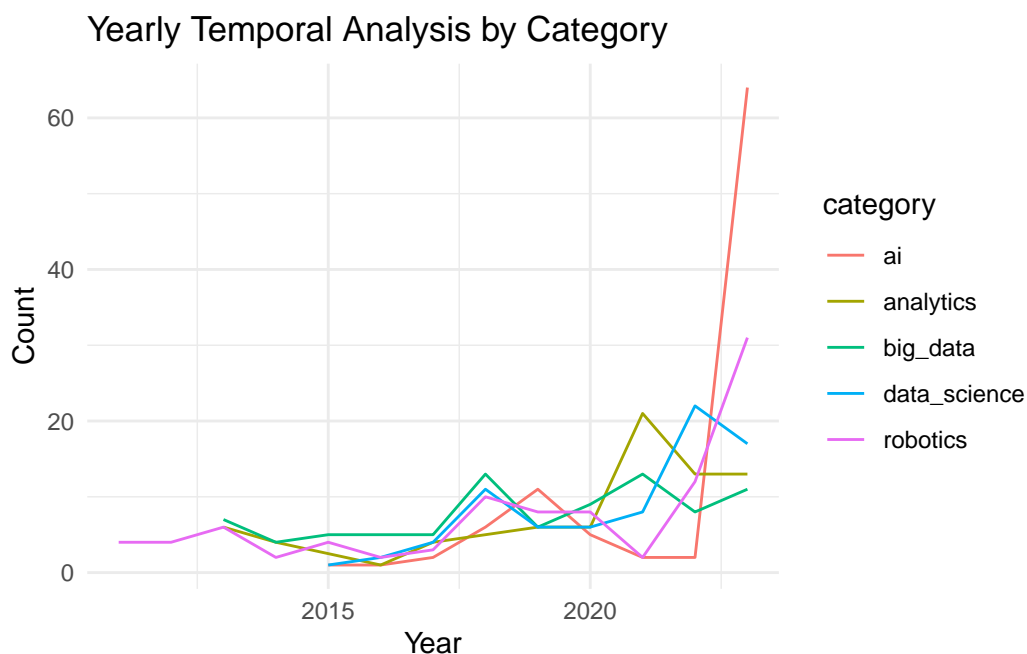

The second part of the code creates a bar chart using `ggplot2`. The `aes(x = year, y = count, fill = category)` sets up the aesthetics for the plot, specifying that the x-axis will represent years, the y-axis will represent counts, and different colors (fill) will represent different categories. The `geom_bar(stat = "identity", position = "dodge")` adds bar geometries

to the plot, with bars positioned side-by-side (dodged) for each year-category combination rather than stacked. This makes it easier to compare the count of articles across categories for each year. `theme_minimal()` is used to apply a minimalistic theme to the plot, a stylistic choice for a clean and uncluttered presentation. The `labs()` function adds labels to the plot, setting the title of the plot as "Yearly Article Count by Category" and naming the x and y-axes as "Year" and "Count," respectively.

The prominent occurrence of terms related to data science and similar fields in Philippine newspapers suggests several vital implications. Firstly, it indicates a growing interest and focus on technology and data science within the country, reflecting the global trend towards data-driven decision-making and artificial intelligence. This prominence could indicate heightened public awareness and an educational shift towards these disciplines. Additionally, it might signal the alignment of the Philippines' economic and industrial trends with global technological advancements.

Furthermore, the frequent coverage of these topics could reflect governmental focus and investment in the tech industry, hinting at potential public policy shifts and infrastructure development to support technological growth. The media's role in shaping public perception is also crucial, as this increased focus could influence more individuals to pursue careers in data science and related fields. This trend might also suggest a vibrant business and start-up environment in the tech sector of the Philippines, highlighting growth areas for investment and job creation.

Moreover, the frequent discussions around data science and associated fields might point to societal challenges and opportunities, such as data privacy, ethical AI use, and the need for skilled professionals. Finally, this trend could signify the Philippines' increasing integration into the global tech landscape, indicating its responsiveness to international technological advancements. While the term frequency analysis underscores the prevalence of these topics, it is essential to conduct further qualitative research to understand the context and sentiment behind this media coverage fully.

Recommendations

In light of the findings, several recommendations can be proposed. Firstly, educational institutions and policymakers are advised to emphasize emerging technologies like data science, AI, and robotics. This could entail updating educational curricula to include advanced technology subjects and promoting STEM education more vigorously. Additionally, there is an opportunity for media outlets and educational bodies to collaborate on public awareness initiatives to enhance understanding and appreciation of these fields among the general public.

Investments in local research and development within these technological domains should be a priority for both the government and the private sector, aiming to keep pace with global trends and establish competitive advantages in specific areas of these technologies. Furthermore, the

nurturing and support of technology startups, through funding and mentorship, can stimulate innovation and economic growth in these sectors.

An equally important recommendation is the need to focus on the ethical implications of data science and AI advancements. This could involve forming regulatory bodies to ensure these technologies' responsible development and implementation. Strengthening industry-academia collaboration would also be beneficial, as it would lead to more practical research outcomes, internship opportunities, and enhanced job prospects for students in these fields.

Finally, the Philippines could greatly benefit from international collaborations and partnerships in research and policy-making in data science and technology, bringing new perspectives, technologies, and investments to the country. These recommendations are geared towards leveraging the growing media interest in data science and related technologies, to foster technological advancement, education, and ethical practices.

Reference

1.   Hey, T. The Fourth Paradigm – Data-Intensive Scientific Discovery. In *Communications in computer and information science*; Springer Berlin Heidelberg, 2012; pp. 1–1.

2.   Provost, F.; Fawcett, T. Data Science and Its Relationship to Big Data and Data-Driven Decision Making. *Big Data* **2013**, doi:10.1089/big.2013.1508.

3.   Tewari, S.H. Necessity of Data Science for Enhanced Cybersecurity. *International Journal of Data Science and Big Data Analytics* **2021**, doi:10.51483/ijdsbda.1.1.2021.63-79.

4.   Cash, D.W.; Clark, W.C.; Alcock, F.; Dickson, N.M.; Eckley, N.; Guston, D.H.; Jäger, J.; Mitchell, R.B. Knowledge Systems for Sustainable Development. *Proceedings of the National Academy of Sciences* **2003**, doi:10.1073/pnas.1231332100.

5.   Agha-Mir-Salim, L.; Sarmiento, R.F. *Health Information Technology as Premise for Data Science in Global Health: A Discussion of Opportunities and Challenges*; 2020;

6.   Pearce, N.; Aït-Khaled, N.; Beasley, R.; Mallol, J.; Keil, U.; Mitchell, E.A.; Robertson, C.F. Worldwide Trends in the Prevalence of Asthma Symptoms: Phase III of the International Study of Asthma and Allergies in Childhood (ISAAC). *Thorax* 2007.

7.   Asher, M.I.; García-Marcos, L.; Pearce, N.; Strachan, D.P. Trends in Worldwide Asthma Prevalence. *European Respiratory Journal* **2020**, doi:10.1183/13993003.02094-2020.

8.   McKibbin, W.J.; Fernando, R. The Global Macroeconomic Impacts of COVID-19: Seven Scenarios. *SSRN Electronic Journal* **2020**, doi:10.2139/ssrn.3547729.

9.   Ichhpujani, P.; Kalra, G.; Kaur, R.; Bhartiya, S. Evolution of Glaucoma Research: A Scientometric Review. *Journal of Current Glaucoma Practice* **2020**, doi:10.5005/jp-journals-10078-1286.

10. Harzing, A.; Alakangas, S. Google Scholar, Scopus and the Web of Science: A Longitudinal and Cross-Disciplinary Comparison. *Scientometrics* **2015**, doi:10.1007/s11192-015-1798-9.

11. Oliver, J.C.; McNeil, T. Undergraduate Data Science Degrees Emphasize Computer Science and Statistics but Fall Short in Ethics Training and Domain-Specific Context. *Peerj Computer Science* **2021**, doi:10.7717/peerj-cs.441.

12. Weichselbraun, A.; Kuntschik, P.; Francolino, V.; Saner, M.; Dahinden, U.; Wyss, V. Adapting Data-Driven Research to the Fields of Social Sciences and the Humanities. *Future Internet* **2021**, doi:10.3390/fi13030059.

13. Herman, E.S.; Chomsky, N. Manufacturing Consent: The Political Economy of the Mass Media. *Contemporary Sociology a Journal of Reviews* **1989**, doi:10.2307/2074220.

14. Shanahan, D. Amusing Ourselves to Death: Public Discourse in the Age of Show Business. *The Social Science Journal* **1988**, doi:10.1016/0362-3319(88)90048-1.

15. Keshvari, M.; Yamani, N.; Adibi, P.; Shahnazi, H. Health Journalism: Health Reporting Status and Challenges. *Iranian Journal of Nursing and Midwifery Research* **2018**, doi:10.4103/ijnmr.ijnmr_158_16.

16. Younes, Z.B.; Hassan, I.; Azmi, M.N.L. A Pragmatic Analysis of Islam-Related Terminologies in Selected Eastern and Western Mass Media. *Arab World English Journal* **2020**, doi:10.24093/awej/vol11no2.6.

17. Sanchez-Pinto, L.N.; Luo, Y.; Churpek, M.M. Big Data and Data Science in Critical Care. *Chest Journal* **2018**, doi:10.1016/j.chest.2018.04.037.

18. Deo, R.C. Machine Learning in Medicine. *Circulation* **2015**, doi:10.1161/circulationaha.115.001593.

19. Song, I.-Y.; Zhu, Y. Big Data and Data Science: What Should We Teach? *Expert Systems* **2015**, *33*, 364–373, doi:10.1111/exsy.12130.

20. Shen, S.; Cheng, C.; Yang, J.; Yang, S. Visualized Analysis of Developing Trends and Hot Topics in Natural Disaster Research. *Plos One* **2018**, doi:10.1371/journal.pone.0191250.

21. Wing, J.M. Ten Research Challenge Areas in Data Science. **2020**, doi:10.1162/99608f92.c6577b1f.

22. Walker, M. The Professionalisation of Data Science. *International Journal of Data Science* **2015**, doi:10.1504/ijds.2015.069048.

23. Peng, R.D.; Parker, H.S. Perspective on Data Science 2021.

24. Zhang, Z.; Zhang, D. What Is Data Science? An Operational Definition Based on Text Mining of Data Science Curricula. *Journal of Behavioral Data Science* **2021**, *1*, 1–16, doi:10.35566/jbds/v1n1/p1.

25. Krosnick, J.A.; Holbrook, A.L.; Lowe, L.; Visser, P.S. The Origins and Consequences of Democratic Citizens' Policy Agendas: A Study of Popular Concern about Global Warming. *Climatic Change* **2006**, *77*, 7–43, doi:10.1007/s10584-006-9068-8.

26. Fitzgerald, A.; Halliday, J.; Heath, D. Environmental DNA as Novel Technology: Lessons in Agenda Setting and Framing in News Media. *Animals* **2021**, *11*, 2874, doi:10.3390/ani11102874.

27. Naeem, M. Uncovering the Role of Social Media and Cross-Platform Applications as Tools for Knowledge Sharing. *VINE Journal of Information and Knowledge Management Systems* **2019**, *49*, 257–276, doi:10.1108/vjikms-01-2019-0001.

28. Wang, X.; Li, Y.; Li, J.; Liu, Y.; Qiu, C. A Rumor Reversal Model of Online Health Information During the Covid-19 Epidemic. *Information Processing & Management* **2021**, *58*, 102731, doi:https://doi.org/10.1016/j.ipm.2021.102731.

29. McCombs, M.E.; Shaw, D.L. The Evolution of Agenda-Setting Research: Twenty-Five Years in the Marketplace of Ideas. *Journal of Communication* **2006**, *43*, 58–67, doi:10.1111/j.1460-2466.1993.tb01262.x.

30. Vu, H.T.; Guo, L.; McCombs, M.E. Exploring "the World Outside and the Pictures in Our Heads": A Network Agenda-Setting Study. *Journalism & Mass Communication Quarterly* **2014**, *91*, 669–686, doi:10.1177/1077699014550090.

31. Lim, J. A Cross-Lagged Analysis of Agenda Setting Among Online News Media. *Journalism & Mass Communication Quarterly* **2006**, *83*, 298–312, doi:10.1177/107769900608300205.

32. Caldwell, M. Public Opinion and the Communication of Consent: Theodore l. Glasser and Charles t. Salmon, Eds. New York: Guilford, 475 Pp., 49.95*hardcover*,24.95 Softcover. *Public Relations Review* **1996**, *22*, 85–86, doi:https://doi.org/10.1016/S0363-8111(96)90076-4.

33. Meraz, S. Using Time Series Analysis to Measure Intermedia Agenda-Setting Influence in Traditional Media and Political Blog Networks. *Journalism & Mass Communication Quarterly* **2011**, *88*, 176–194, doi:10.1177/107769901108800110.

34. McCombs, M.; Llamas, J.P.; Lopez-Escobar, E.; Rey, F. Candidate Images in Spanish Elections: Second-Level Agenda-Setting Effects. *Journalism & Mass Communication Quarterly* **1997**, *74*, 703–717, doi:10.1177/107769909707400404.

35. Li, Y.; Jiang, Y. Research Transmutation of Agenda Setting Theory (2010-2022)—a Meta-Analysis Based on Journalism Studies. *Open Journal of Social Sciences* **2022**, *10*, 148–155, doi:10.4236/jss.2022.1011011.

36. Ho, R.T.H.; Ng, S.M.; Ho, D.Y.F. The Sage Handbook of Qualitative Research. *Asian Journal of Social Psychology* **2007**, *10*, 277–279, doi:10.1111/j.1467-839x.2007.00237.x.

37. Suppli, C.H.; Hansen, N.D.; Rasmussen, M.; Valentiner-Branth, P.; Krause, T.G.; Mølbak, K. Decline in HPV-Vaccination Uptake in Denmark – the Association Between HPV-Related Media Coverage and HPV-Vaccination. *BMC Public Health* **2018**, *18*, doi:10.1186/s12889-018-6268-x.

38. R Core Team *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2023;

39. RStudio Team *RStudio: Integrated Development Environment for r*; RStudio, PBC: Boston, MA, 2023;

40. Wickham, H. *Rvest: Easily Harvest (Scrape) Web Pages*; 2022;

41. Wickham, H.; Averick, M.; Bryan, J.; Chang, W.; McGowan, L.D.; François, R.; Grolemund, G.; Hayes, A.; Henry, L.; Hester, J.; et al. Welcome to the tidyverse. *Journal of Open Source Software* **2019**, *4*, 1686, doi:10.21105/joss.01686.

42. Dorado, D.A. Prominence of occurrence accorded to data science and other related terms in Philippine newspapers 2023.