VISUAL REVIEW | Vol. 16, No. 5, 2024 | ISSN 2695-9631
International Visual Culture Review / Revista Internacional de Cultura Visual
https://doi.org/10.62161/revvisual.v17.5288

INTEGRATION OF TEXT MINING AND COMPLEX SOCIAL NETWORK ANALYSIS For Quantitative Characterization of Discursive Evolution in Artificial Intelligence Through Streaming Platforms

FRED TORRES-CRUZ [1] , YUDI JANEH YUCRA-MAMANI [1] , VLADIMIRO IBANEZ-QUISPE [1] ,
MARCO FIDEL MAYTA-QUISPE [1]
[1] Universidad Nacional del Altiplano Puno, P.O. Box, 291, Puno-Perú

KEYWORDS

Artificial Intelligence
Discursive Evolution
Text Mining
Social Network Analysis
Streaming Platforms

ABSTRACT

This study provides a quantitative characterisation of the discursive evolution of Artificial Intelligence (AI) on streaming platforms (YouTube), integrating text mining and complex social network analysis. A comprehensive corpus was obtained from leading platforms, and natural language processing algorithms were implemented to analyse the textual content and identify patterns, emerging themes, and changes in AI discourses. Concurrently, a social network analysis was conducted to examine the interaction structures and the influence of different actors in the dissemination of information. The findings reveal significant trends in the representation and perception of AI, indicating the evolution of specific topics, differences in perception between different groups, and the influence of factors such as technological advances and global events. This analysis offers a more profound comprehension of the communication and perception of AI in the digital domain, providing valuable insights for academics, communicators and policymakers engaged in the field of AI.

Received: 20 / 03 / 2024
Accepted: 18 / 07 / 2024

1. Introduction

The advent of Artificial Intelligence (AI) has been identified as an agent of profound change in the contemporary era. The impact of AI on various domains, including industry, process and the way in which technology is used by humans, has been considerable. As the application of AI expands, so too does the public and academic discourse surrounding it, evolving with each technological advance and its integration into everyday life. Nevertheless, characterising this discursive evolution, particularly on pervasive digital platforms such as streaming, has proven to be a significant challenge due to the scale and dynamism of the data generated (Morchid et al., 2013). This study addresses these challenges by providing a quantitative and analytical view of how AI discourse has evolved on these platforms.

The advent of streaming platforms such as YouTube, Twitch and others has led to a notable increase in the quantity and variety of AI-related content, as well as in the ways in which it is presented and perceived. These platforms serve not only as a medium for the dissemination of information, but also as a space for public debate, community building, and opinion formation (Campis, 2023). A comprehensive analysis of the discursive evolution of these platforms can therefore provide valuable insights into public perception and communication around AI.

In order to achieve this goal, the present study integrates text mining and complex social network analysis methods, utilising computational power and advances in natural language processing (NLP) to analyse large textual datasets. (Li et al., 2024). The application of text mining enables the identification of patterns, emerging themes and changes in discourse. In contrast, complex social network analysis provides insight into the interaction structures and influence of different actors in the dissemination of AI information.

This multidisciplinary methodological approach enables an in-depth quantitative characterisation of the discursive development of AI on streaming platforms (Bernstein et al., 2019). By collecting and analysing a comprehensive corpus of textual content, the study identifies key trends in the representation and perception of AI, the evolution of specific themes, and how different groups perceive these changes (Gonzales et al., 2020). This analysis not only elucidates the dynamics of AI discourse, but also demonstrates how external factors, such as technological advances and global events, can influence this discursive evolution. By examining the intersection of technology, communication and society, this work not only enriches the academic discourse on AI, but also provides critical tools for navigating its future development and integration into the social fabric.

2. Methodology

The methodology employed in this descriptive study to analyse the discursive evolution of AI on streaming platforms, specifically YouTube, combines advanced text mining techniques and complex social network analysis (Lee, 2023). This interdisciplinary methodological approach allows for a detailed and quantitative exploration of the discursive dynamics and influence of different actors in the field of AI (Lau et al., 2014).

2.1. Participants

In the data collection phase, 5301 videos published on YouTube were analysed. A total of 16 YouTube channels were selected as being relevant to the field of AI (Table 1). A comprehensive corpus of videos and their descriptions, comments, and related metadata, such as publication date and number of views, was extracted. The selection was made with the aim of including a wide range of perspectives within the AI community, including academics, enthusiasts and industry professionals.

Table 1. Description of study participants

N	YouTube Channel	Number of Subscribers	Number of Videos
1	IBM Technology	685000	852
2	Lex Fridman	3770000	783
3	Ringa Tech	259000	48
4	Data School	236000	140

N	YouTube Channel	Number of Subscribers	Number of Videos
5	Arxiv Insights	91800	13
6	Tech Dummies Narendra L	155000	60
7	Artificial Intelligence	77700	110
8	Two Minute Papers	1530000	868
9	Machine Learning 101	2540	4
10	StatQuest with Josh Starmer	1120000	268
11	Tech With Tim	1460000	974
12	Code Bullet	3110000	59
13	Siraj Raval	765000	466
14	3Blue1Brown	5970000	172
15	Dot CSV	843000	208
16	LearnIA with Ligdi Gonzalez	65800	276
Total		20140840	5301

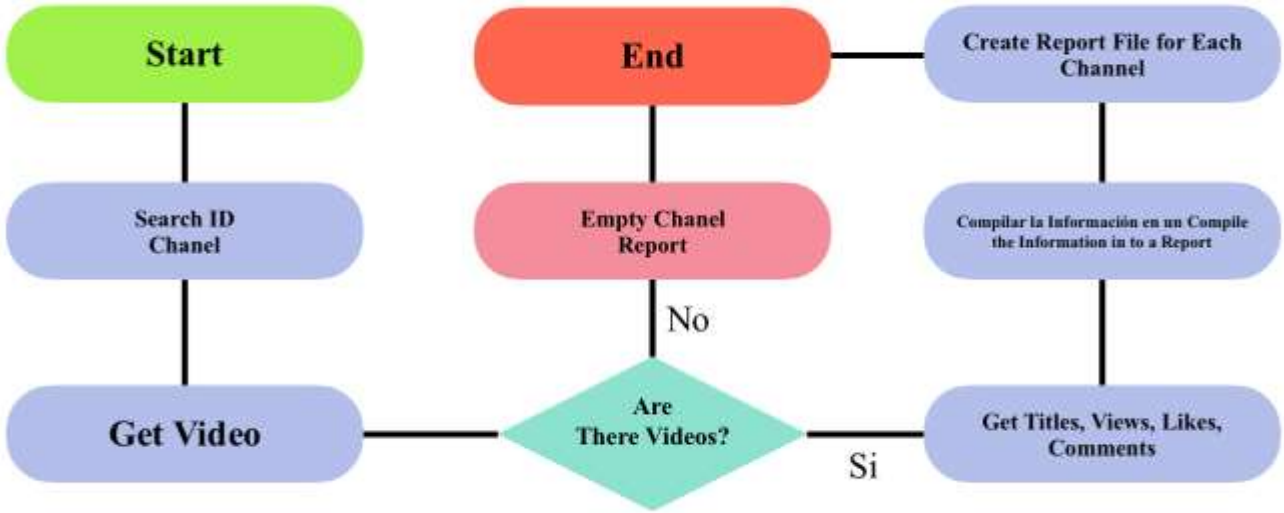
Source: Own elaboration, 2024.

2.2. Techniques and Instruments

Information Gathering

A search and selection process were conducted to identify YouTube channels that were specifically related to the field of AI. Subsequently, the channel identifiers were obtained, followed by the extraction of videos from each channel. (D. Li et al., 2024) The decision as to whether content is available on the channel leads to two possible routes: the collection of video identifiers, titles, views, likes and comments, or the creation of an empty report in the event that no videos are available. (Srikumar & Srikumar, 2021) The data thus collected is then collated into a report, which is subsequently used to create a file for each channel. This completes the data collection phase of the study, as illustrated in Figure 1.

Figure 1. Information Gathering Process



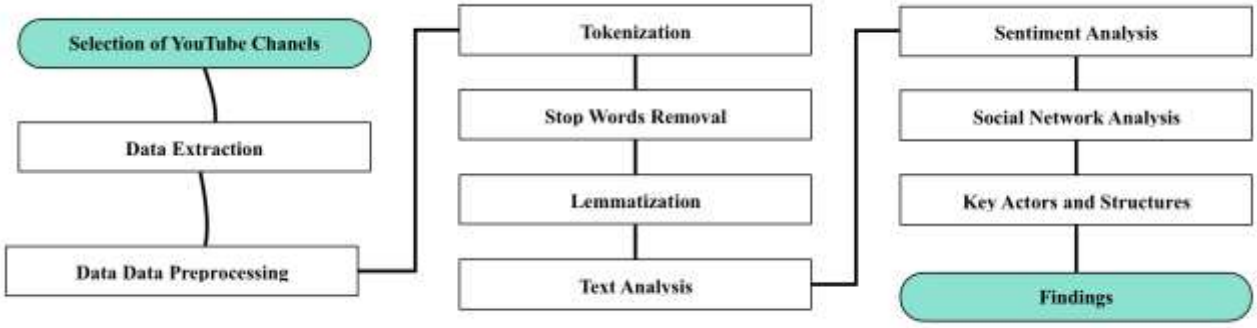
Source: Own Elaboration

Text Processing and Analysis

The Python programming language was employed for the processing and analysis of the extracted text, leveraging its robust ecosystem of libraries for data analysis and natural language processing (NLP). (Torres-Cruz & Mamani, 2022) The following libraries were of critical importance in the course of our analysis:

- **Pandas:** Used for the manipulation and analysis of data. This process facilitated the organisation of the text corpus and associated metadata into efficient data structures for further analysis.
- **Gensim:** Particularly its corpora module and LdaModel, permitted the construction of latent Dirichlet distribution (LDA) models to identify dominant themes and their evolution over time within the corpus.
- **NLTK (Natural Language Toolkit):** The library was employed for text pre-processing operations, including tokenisation, stop word removal, and lemmatisation. These operations were conducted using the WordNetLemmatizer. This was done in order to prepare the text for more effective analysis.
- **Matplotlib:** The software facilitated the visualisation of the results of the analysis, allowing the generation of graphs and diagrams for a more effective interpretation of the data.

Figure 2. Information Process Flow chart



Source: Own Elaboration

Social Media Analysis

In conjunction with the text analysis, an analysis of the interaction structures and influence of different actors within the AI community on YouTube was carried out (Campis, 2023). This analysis entailed the identification of networks of co-occurrence of topics and the influence of specific channels on information dissemination and discourse formation.

Presentation of Results

The Python reportlab library was employed to generate reports in PDF format (Gonzales et al., 2023). This process involved the creation of stylised tables and graphs, as well as the writing of interpretative paragraphs of the results, using SimpleDocTemplate, Table, TableStyle, and customised Paragraph styles to structure the final document in a professional manner.

2.3. Study Context

The context of this study is situated in a historical moment in which AI has not only made significant technological advances but has also deeply permeated public and academic discourse. As AI becomes increasingly integrated into various aspects of everyday and professional life, streaming platforms are emerging as crucial spaces for the dissemination and discussion of knowledge, opinions and predictions related to this technology (Campbell et al., 2000). These platforms, most notably YouTube, not only act as communication channels between experts and enthusiasts, but also play a fundamental role in shaping public perceptions of AI. This phenomenon has generated a rich amalgam of content that

reflects, informs and in turn shapes the discursive development of AI. In this context, the study aims to quantify and analyse this discursive evolution using an interdisciplinary approach that integrates text mining and complex social network analysis. The objective is to provide a broad and detailed perspective on how discussions about AI have changed over time and how these changes have impacted public perception (Torres & Moreta, 2023).

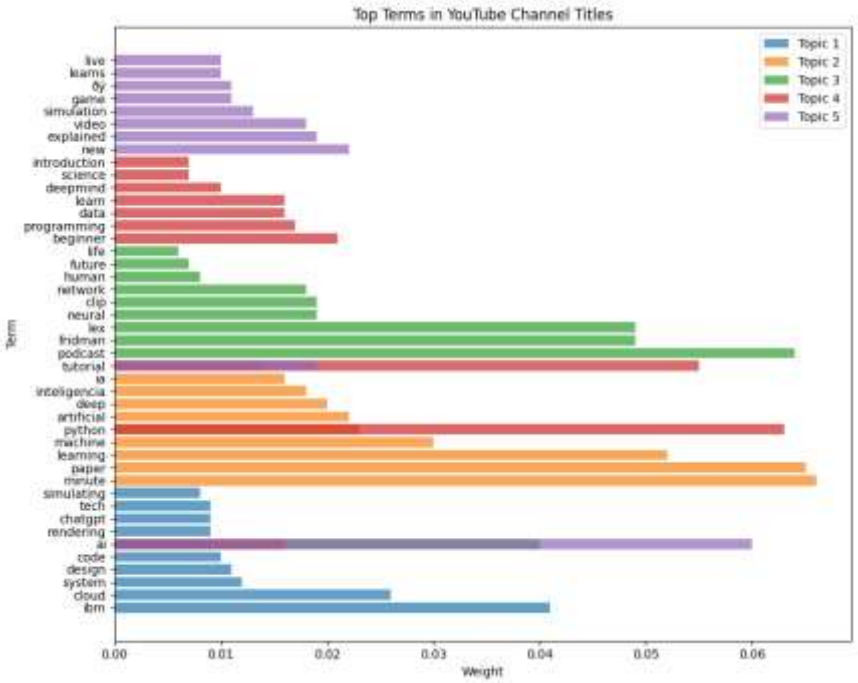
3. Objectives

The overarching objective of this study is to provide a quantitative characterisation of discursive developments in the field of AI across streaming platforms, with a particular focus on YouTube. This entails the integrated application of text mining methodologies and complex social network analysis to examine the manner in which AI-related topics are presented and perceived in digital discourse. By analysing a comprehensive corpus of content generated on these platforms, the study seeks to identify patterns, emerging themes and shifts in AI discourse. Furthermore, the study aims to explore the interaction structures and influence of different actors in the dissemination of information about AI. This approach will provide valuable insights into the evolution of communication and public perception of AI, offering an empirical basis for future research, communication strategies, and policy in the field of AI.

4. Results and Discussion

Figure 2 presents an analysis of key terms within YouTube channel titles related to the field of AI, distributed across five different topics, identified as Topic 1 to Topic 5. The terms are presented in descending order of importance within each topic. The results were obtained through text analysis and topic modelling.

Figure 2. Discursive Analysis of YouTube Channels



Source: Own elaboration, 2024.

Figure 2 presents an analysis of key terms within YouTube channel titles related to AI. Topic modelling allows us to distinguish the most addressed topics in AI videos on the streaming platform:

Theme 1 (blue): AI technology and platforms

The most relevant terms are 'tech', 'chatgpt', 'rendering', 'ai', 'code', 'design', 'system', 'cloud' and 'ibm', suggesting that this theme is related to the technical aspects of AI, including cloud computing platforms, programming and companies such as IBM.

Theme 2 (Orange): Machine learning education and tutorials.

The strongest terms include 'simulate', 'tech', 'minute', 'paper', 'machine learning', 'python', 'ia' and 'deep intelligence', which could indicate a focus on tutorials, short presentations and academic publications, possibly related to deep learning and Python programming.

Topic 3 (Green): Neural programming and its social implications

Here we highlight 'programming', 'beginner', 'life', 'future', 'human network', 'neural', 'clip' and 'max', terms that may signal a discussion of neural programming and its future implications for human life and society, with a focus on content accessible to beginners.

Topic 4 (Red): Educational Foundations in Data Science and AI.

The words 'learn', 'data', 'programming', 'beginner', 'science', 'deepmind', 'introduction' and 'new' suggest an emphasis on educational resources for learning about data science and AI, with a mention of DeepMind, which is known for its advances in AI.

Theme 5 (purple): Interactivity and playful learning in AI

With terms such as 'live', 'learning', 'game', 'simulation', 'video' and 'explained', this theme could refer to interactive and educational content that explores AI concepts through simulations and games, possibly aimed at live or broadcast learning.

5. Conclusions

The study findings provide a comprehensive view of the presence and evolution of AI on streaming platforms:

Evolution of AI Discussion: There has been a significant evolution in the topics discussed on platforms such as YouTube. AI has moved from a specialised topic to one of general interest, broadening its scope and the diversity of its audience. Patterns of discussion reflect an increased awareness of the ethical, privacy and socio-economic implications of AI.

Public perception and communication of AI: The findings suggest that the way in which YouTube channels present AI influences public perception. Content creators who emphasise practical applications of AI and its benefits tend to attract a wider audience and foster positive perceptions of the technology.

Influence of actors in AI dissemination: Social media analysis has shown that certain YouTube channels act as key influencers in the AI community, with the ability to significantly shape the discourse. Interaction between small and large channels can create a rich and diverse content ecosystem.

Impact of events and technological breakthroughs: Global events and technological breakthroughs have been observed to act as catalysts for change in the AI discourse. These events not only generate spikes in discursive activity, but also introduce new topics and perspectives into the conversation.

Innovative methodology for discourse analysis: The successful integration of text mining and complex social network analysis has proven to be a robust methodology for studying large datasets from streaming platforms, providing detailed insights that would be difficult to obtain using traditional methods.

Directions for future research: This study opens up several avenues for future research, including longitudinal analysis of the impact of AI on different industries, the influence of AI on political decision-making, and strategic communication management on streaming platforms.

This study has provided quantitative and qualitative insights into the discursive development of AI on streaming platforms, highlighting the importance of effective communication and understanding the dynamics of influence within the online community. The findings highlight the need for careful

consideration of how AI is presented to the public, and the role that content creators play in shaping public understanding and attitudes towards this transformative technology.

References

Bernstein, R., Osadchy, M., Keren, D., & Schuster, A. (2019). LDA classifier monitoring in distributed streaming systems. *Journal of Parallel and Distributed Computing*, *123*, 156-167. https://doi.org/10.1016/j.jpdc.2018.09.017

Campbell, M., Cosgrove, J. A., Greated, C. A., Jack, S., & Rockliff, D. (2000). Review of LDA and PIV applied to the measurement of sound and acoustic streaming. *Optics & Laser Technology*, *32*(7), 629-639. https://doi.org/10.1016/S0030-3992(00)00091-8

Campis, E. S. (2023). Aplicación del Análisis de Redes Sociales para el estudio de las redes de comunicación en línea: Evidencia empírica de Twitter. *EMPIRIA: Revista de Metodología de Ciencias Sociales*. https://doi.org/10.5944/empiria.57.2023.36434

Gonzales, E. P. Z., Gallegos, W. L. A., Gallegos, W. L. A., Toia, A. M. del C., Rivera, R., & Condori, M. L. (2020). Depresión postparto en mujeres víctimas de violencia doméstica en la ciudad de Arequipa. *null*. https://doi.org/null

Gonzales, J. E., Ortega, Á. C., Blanco, R. R., & Diez, E. B. (2023). ANÁLISIS DE SENTIMIENTOS Y EMOCIONES SOBRE GASTRONOMÍA PERUANA USANDO MINERÍA DE TEXTO CON PYTHON. *Hatun Yachay Wasi*. https://doi.org/10.57107/hyw.v3i1.63

Lau, J. H., Lau, J. H., Newman, D. J., Newman, D., Baldwin, T., & Baldwin, T. (2014). Machine Reading Tea Leaves: Automatically Evaluating Topic Coherence and Topic Model Quality. *null*. https://doi.org/10.3115/v1/e14-1056

Lee, S. (2023). An Analysis on Major Issues of Medical Information on YouTube Platform Using LDA Topic Modeling. *The Journal of the Korea Contents Association*. https://doi.org/10.5392/jkca.2023.23.08.198

Li, D., Bi, Z., Han, X., Qu, Z., & Lyu, C. (2024). Extracting key topics from massive COVID-19 information on social networks: An integrated deep learning and LDA framework. *High-Confidence Computing*, 100213. https://doi.org/10.1016/j.hcc.2024.100213

Li, M., Lu, M., Akram, U., & Cheng, S. (2024). Understanding how customer social capital accumulation in brand communities: A gamification affordance perspective. *Journal of Retailing and Consumer Services*, *78*, 103761. https://doi.org/10.1016/j.jretconser.2024.103761

Morchid, M., Morchid, M., Linarès, G., & Linarès, G. (2013). A LDA-based method for automatic tagging of Youtube videos. *Workshop on Image Analysis for Multimedia Interactive Services*. https://doi.org/10.1109/wiamis.2013.6616126

Srikumar, N., & Srikumar, N. (2021). Feature Augmentation for Improved Topic Modeling of Youtube Lecture Videos using Latent Dirichlet Allocation. *null*. https://doi.org/10.21427/4ey6-qg08

Torres, V., & Moreta, O. E. (2023). Detección de ansiedad mediante minería de texto en la era de redes sociales: Revisión bibliográfica. *Ciencia Huasteca Boletín Científico de la Escuela Superior de Huejutla*. https://doi.org/10.29057/esh.v11i22.10879

Torres-Cruz, F., & Mamani, Y. J. Y. (2022). Técnicas de inteligencia artificial en la valoración de la enseñanza virtual por estudiantes de nivel universitario. *HUMAN Review International Humanities Review / Revista Internacional de Humanidades*. https://doi.org/10.37467/revhuman.v11.3853