



Big Data–Driven Public Opinion Mining: A Conceptual Framework for Evidence-Informed Policymaking

Ghasem Safaeinejad¹ , Seyed Mehdi Sharifi² 

1. Ph. D. in Media Management, Alborz Campus, University of Tehran, Tehran, Iran. Email: safaeinejad@ut.ac.ir
2. Associate Prof., Department of Media Management, College of Management, University of Tehran, Tehran, Iran. Email: Sharifee@ut.ac.ir

Article Info

Article type:

Research Article

Article history:

Received: 30 September 2025

Received in revised form: 29

December 2025

Accepted: 23 February 2026

Available online: 31 July 2026

Keywords:

Opinion Mining,

Big Data,

Sentiment Analysis,

Policymaking,

Social media.

ABSTRACT

Objective: The present study was conducted with the aim of designing a conceptual model for national opinion mining based on big data. The necessity of this research arises from the fact that traditional opinion polling methods, such as questionnaires and interviews, face limitations including high costs, time delays, and restricted geographical coverage. In contrast, big data generated from citizens' digital interactions on social media and the web provides unprecedented capacity for real-time monitoring of public opinion dynamics and supporting evidence-informed policymaking.

Methods: This research is developmental–theoretical in nature and was carried out with a qualitative–inferential approach. The data used were collected from secondary sources, including peer-reviewed scientific articles and national and international policy reports. The methodology was based on theoretical analysis and conceptual synthesis. First, a systematic literature review was conducted to identify key concepts and research gaps. Then, using conceptual inference, the essential components and relationships of the opinion mining process were extracted and organized into a five-layer model: (1) data collection, (2) cleaning and preprocessing, (3) analysis and clustering, (4) interpretation and validation, and (5) feedback and application.

Results: The findings of the study indicated that the proposed model can overcome the limitations of traditional opinion polling and enable scalable, multidimensional, and real-time monitoring of public opinion. In the analysis layer, the use of natural language processing and machine learning algorithms (sentiment analysis, topic modeling, clustering, and social network analysis) made it possible to identify emotional trends, dominant topics, and key actors. In the interpretation layer, expert feedback was obtained to enhance the validity of the model. Finally, the feedback and application layer demonstrated that the model's outputs can be delivered through interactive dashboards and policy briefs to decision-making institutions, thereby serving to complement and strengthen traditional opinion polling through data-driven analysis.

Conclusions: The conceptual model of national opinion mining based on big data provides a novel framework for a gradual transition in certain functions and for complementing traditional opinion polling through data-driven analysis. By reducing costs and time, increasing accuracy and coverage, and enabling real-time insights, this model can serve as a strategic tool for evidence-informed policymaking at the national level. Nevertheless, challenges such as the need for robust computational infrastructure, adherence to ethical and privacy principles, and optimization of natural language processing algorithms for Persian must be taken into account. Overall, the findings of this study can serve as a foundation for developing indigenous opinion mining tools and enhancing data-driven decision-making capacity in Iran.

Cite this article: Safaei Nejad, Gh. & Sharifi, S.M. (2026), Big Data–Driven Public Opinion Mining: A Conceptual Framework for Evidence-Informed Policymaking, *News Science*, 15 (2), 35-39. <http://doi.org/10.22034/lrsi.2026.550018.1443>



© The Author(s) retain the copyright.

DOI: <http://doi.org/10.22034/lrsi.2026.550018.1443>

Publisher: Press Institute for Contemporary International Journalism (PICIJ)



The Journal of News Science

Vol. 15, No. 2, Ser.58, summer 2026, P. 35- 39

Journal homepage: <https://www.mjourcom.ir/>

DOI : <http://doi.org/10.22034/lrsi.2026.550018.1443>

Open Access

ORIGINAL ARTICLE

Big Data–Driven Public Opinion Mining: A Conceptual Framework for Evidence-Informed Policymaking

Ghasem Safaeinejad¹ , Seyed Mehdi Sharifi² 

3. Ph.D. in Media Management, Alborz Campus, University of Tehran, Tehran, Iran. Email: safaeinejad@ut.ac.ir
4. Associate Prof., Department of Media Management, College of Management, University of Tehran, Tehran, Iran. Email: Sharifee@ut.ac.ir

Received: September 30, 2025

Accepted: February 23, 2026

EXTENDED ABSTRACT

Introduction:

The rapid expansion of digital ecosystems and social media platforms has fundamentally transformed the production and circulation of public information. Citizens continuously generate vast quantities of textual, behavioral, and networked data through their interactions on social media, messaging applications, news websites, and other digital environments. These data offer unprecedented opportunities for understanding public attitudes, concerns, affective states, and emerging social issues. Concurrently, conventional public opinion polling methods-which rely predominantly on questionnaires, interviews, and field-based sampling-face several inherent limitations, including relatively high operational costs, time-intensive data collection procedures, restricted geographical and temporal coverage, declining response rates, and susceptibility to various response biases.

Against this backdrop, the present study aimed to develop a conceptual framework for national opinion mining grounded in big data analytics, and to examine how data-driven approaches can complement-and, in certain functional domains, progressively substitute for-traditional opinion polling. Particular attention was devoted to the contextual requirements of applying such a framework within the Iranian setting,

including the linguistic idiosyncrasies of Persian-language digital content, constraints in accessing platform data, ethical and privacy considerations, the representativeness of social media data, and the imperative to translate computational outputs into actionable intelligence for policymakers. Accordingly, the study sought to establish an integrated framework that links data acquisition, computational analysis, contextual interpretation, validation, and policy application into a coherent operational model.

Method:

This study was developmental–theoretical in purpose and adopted a qualitative–inferential approach. A systematic qualitative meta-synthesis was employed to integrate findings from prior research and policy documents, and to derive the constituent components of the proposed conceptual framework. The study population comprised peer-reviewed scientific publications and credible policy reports addressing big data, opinion mining, public opinion analysis, social media analytics, and evidence-informed policymaking.

A systematic literature search was conducted across international databases-including Google Scholar, Web of Science, ScienceDirect, and Scopus-as well as Iranian repositories such as Noormags, Magiran, and the Comprehensive Portal of Humanities. The search covered publications from 2010 to 2025. Keywords comprised combinations of "Opinion Mining," "Sentiment Analysis," "Topic Modeling," "Big Data," "Social Media Data," "Social Network Analysis," "Evidence-Informed Policy," "Conceptual Model," and "Framework," along with their Persian equivalents.

The initial search yielded 245 documents. Following the removal of duplicates and screening of titles and abstracts, 98 documents of direct relevance to the research problem were retained. Full-text assessment was subsequently conducted based on criteria of scientific credibility, procedural orientation, and comprehensiveness, culminating in the final selection of 29 documents for in-depth analysis and coding.

Data were analyzed using qualitative content analysis with a combined deductive–inductive strategy. The analytical procedure encompassed open coding, axial coding, and selective coding. Extracted concepts were documented in an evidence matrix to ensure traceability of conceptual inferences. To enhance credibility, portions of the source material were independently coded by two researchers; discrepancies were resolved through discussion and inter-coder consensus. The resulting conceptual model was subsequently reviewed by six experts in data science and media management, and their feedback was incorporated into the final structure.

Findings:

The synthesis of the selected studies yielded a five-layer conceptual framework for big data–driven national opinion mining. The first layer, *data collection and ethical monitoring*, concerns the continuous multi-source acquisition of data from domestic and international social media platforms, messaging applications, news websites, and relevant databases. Diversity of sources is emphasized to mitigate representational bias, while privacy-by-design principles and data anonymization protocols are embedded within the data acquisition process.

The second layer, Persian-language data cleaning and preprocessing, addresses noise reduction and the linguistic complexities inherent in Persian digital content. It encompasses the removal of duplicate and bot-generated content, character normalization, stemming and lemmatization, and the development and ongoing updating of indigenous sentiment lexicons capable of recognizing colloquial expressions, idioms, irony, and culturally specific semantic nuances.

The third layer, analysis and modeling, transforms processed data into structured analytical intelligence. Core techniques include sentiment analysis, analysis of sentiment dynamics, topic modeling, clustering, social network analysis, and anomaly detection. These methods enable the identification of dominant and emerging topics, emotional trajectories, influential actors, sudden shifts in public attention, and potentially coordinated or non-organic communication patterns.

The fourth layer, contextual interpretation and conceptual validation, underscores the role of interdisciplinary expertise in interpreting computational outputs. Social, cultural, generational, and geographical contexts must be taken into account to identify possible biases and to prevent purely algorithmic interpretations of public opinion. Expert validation serves as an additional mechanism for assessing the conceptual coherence, applicability, and trustworthiness of the model.

The fifth layer, feedback and policymaking, connects analytical findings to decision-making processes. Outputs can be presented through interactive management dashboards, concise policy briefs, and data visualization tools. The framework also incorporates a feedback loop through which the consequences of policy interventions can inform subsequent data collection procedures and analytical models, thereby enabling a continuously adaptive learning system.

Conclusion:

The proposed conceptual framework demonstrates that big data-driven opinion mining can provide a scalable, multidimensional, and near-real-time complement to conventional public opinion polling. By integrating computational techniques with contextual interpretation, expert validation, ethical safeguards, and policy-oriented feedback mechanisms, the framework transcends purely technical approaches to social media analytics and establishes a more comprehensive model for data-informed public decision-making.

Nevertheless, big data-driven opinion mining should not be regarded as an unconditional replacement for conventional surveys. Digital data may underrepresent population segments with limited online participation and may contain demographic, geographical, platform-related, and behavioral biases. Therefore, particularly during early implementation phases, combining big data analytics with traditional survey results can improve validation and facilitate the estimation and correction of potential biases.

Implementation of the proposed model in Iran requires reliable data infrastructures, advanced computational resources, trained interdisciplinary teams, enhanced Persian natural language processing tools, transparent ethical and legal protocols, and institutional mechanisms for translating analytical findings into policy-relevant evidence. Subject to these prerequisites, the proposed framework can provide a foundation for developing indigenous national opinion-mining systems and strengthening Iran's capacity for evidence-informed and data-driven policymaking.

Data Availability Statement.

Data available on request from the authors.

Acknowledgements

The authors would like to thank anonymous reviewers.

Ethical considerations

Not applicable.

Funding

Not applicable.

Conflict of interest

The authors declare no conflict of interest.

References:

- Abdalla, H. B. (2022). A brief survey on big data: Technologies, terminologies and data intensive applications. *Journal of Big Data*, 9, Article 107, 1–36. <https://doi.org/10.1186/s40537-022-00659-3>
- Akbas, E. (2017). Opinion mining on non-English short text. In M. Kryszkiewicz, A. Appice, D. Ślęzak, H. Rybinski, A. Skowron, & Z. Raś (Eds.), *Foundations of Intelligent Systems: ISMIS 2017* (Lecture Notes in Computer Science, Vol. 10352, pp. 1–10). Springer, Cham. https://doi.org/10.1007/978-3-319-60438-1_41
- Aladeemy, A. A., Alzahrani, A., Algarni, M. H., Khalaff, O. I., Wong, W.-K., & Aqburi, S. (2024). Advancements and challenges in Arabic sentiment analysis: A decade of methodologies, applications, and resource development. *Heliyon*, 10(21), e39786. <https://doi.org/10.1016/j.heliyon.2024.e39786>

- Albalawi, R., Yeap, T. H., & Benyoucef, M. (2020). Using topic modeling methods for short-text data: A comparative analysis. *Frontiers in Artificial Intelligence*, 3, Article 42. <https://doi.org/10.3389/frai.2020.00042>
- Atkin, C., & Gaudino, J. (1984). The impact of polling on the mass media. *The ANNALS of the American Academy of Political and Social Science*, 472(1), 119–128. <https://doi.org/10.1177/0002716284472001011>
- Bovet, A., Morone, F., & Makse, H. (2018). Validation of Twitter opinion trends with national polling aggregates: Hillary Clinton vs Donald Trump. *Scientific Reports*, 8, Article 6789, 1–16. <https://doi.org/10.1038/s41598-018-26951-y>
- Cen, Y., & Wang, Y. (2016). Social public opinion analysis and decision making support with big data. *Data Analysis and Knowledge Discovery*, 32(7), 3–11. <https://doi.org/10.11925/infotech.1003-3513.2016.07.02>
- Cortis, K., & Davis, B. (2021). Over a decade of social opinion mining: A systematic review. *Artificial Intelligence Review*, 54(7), 4873–4965. <https://doi.org/10.1007/s10462-021-10030-2>
- Farrokhi, M. (2004). Public opinion polling in Iran. Majlis Research Center (Iranian Parliament Research Center). (In Persian)
- Farzindar, A., & Louis, A. (2016). Natural language processing for social media [Book review]. *Computational Linguistics*, 42(4), 833–836. https://doi.org/10.1162/COLI_r_00270
- Featherstone, J. D., & Barnett, G. A. (2020). Validating sentiment analysis on opinion mining using self-reported attitude scores. In 2020 Seventh International Conference on Social Networks Analysis, Management and Security (SNAMS) (pp. 1–4). Paris, France. <https://doi.org/10.1109/SNAMS52053.2020.9336540>
- Heidari, E. (2020). A study of big data in social networks. *New Research Quarterly in Humanities, New Series*(26), 109–130. (In Persian)
- Lasswell, H. D. (1948). The structure and function of communication in society. In L. Bryson (Ed.), *The communication of ideas* (pp. 37–51). Harper & Row.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage Publications.
- Miller, P. V. (2017). Is there a future for surveys? *Public Opinion Quarterly*, 81(S1), 205–212. <https://doi.org/10.1093/poq/nfx008>
- Patil, A. S. (2016). Opinion mining techniques for non-English languages: An overview. *International Journal of Computational Linguistics (IJCL)*, 7(2), 1–16.
- Rahmanian, P., Mousavi, S. M. R., & Sadraldini, M. H. (2025). Multimodal sentiment analysis in Persian using transformer-based models and sentiment recognition capsules. *Theoretical and Applied Research in Machine Intelligence*, 3(1), 51–67. <https://doi.org/10.22034/abmir.2025.23347.1137> (In Persian)
- Rajabi, Z., Velvi, M. R., & Hourali, M. (2022). A review of sentiment analysis methods in Persian texts. *Signal and Data Processing*, 10(52), 107–132. (In Persian)
- Sandelowski, M., Barroso, J., & Voils, C. I. (2007). Using qualitative metasummary to synthesize qualitative and quantitative descriptive findings. *Research in Nursing & Health*, 30(1), 99–111. <https://doi.org/10.1002/nur.20176>
- Schaefer, T., & Brooker, R. (2005). *Public opinion in the 21st century: Let the people speak?* Houghton Mifflin Harcourt Publishing Company.
- Shang, S., Shi, M., Shang, W., & Hong, Z. (2015). Research on public opinion based on big data. In 2015 IEEE/ACIS 14th International Conference on Computer and Information Science (ICIS) (pp. 559–562). <https://doi.org/10.1109/ICIS.2015.7166655>
- Shi, Z., & Agrawal, R. (2025). A comprehensive survey of contemporary Arabic sentiment analysis: Methods, challenges, and future directions. In *Findings of the Association for Computational Linguistics: NAACL 2025* (pp. 3760–3772). Albuquerque, NM: Association for Computational Linguistics.
- Stahl, B. C. (2025). The ethics of data and its governance: A discourse theoretical approach. *Information*, 16(6), 497. <https://doi.org/10.3390/info16060497>
- Temporão, M. (2019). *Measuring public opinion using big data: Applications in computational social sciences** [Doctoral dissertation, Laval University]. Canada.
- Wang, Y., Guo, J., Yuan, C., & Li, B. (2022). Sentiment analysis of Twitter data. *Applied Sciences*, 12(22), 11775. <https://doi.org/10.3390/app122211775>
- Zahed, A., & Sakhi, M. R. (2016, December 19). A review of big data techniques [Conference presentation]. First National Conference on Electrical and Computer Engineering, Distributed Systems and Smart Networks, Islamic Azad University, Kashan Branch, Iran.
- Zhao, Z., Liu, W., & Wang, K. (2023). Research on sentiment analysis method of opinion mining based on multi-model fusion transfer learning. *Journal of Big Data*, 10, Article 155. <https://doi.org/10.1186/s40537-023-00837-x>

علوم خبری، دوره ۱۵، شماره ۲ (۱۴۰۵)



افکار کاوی ملی مبتنی بر داده‌های کلان: چارچوبی مفهومی برای سیاست‌گذاری مبتنی بر شواهد

قاسم صفائی نژاد^۱ | اسید مهدی شریفی^۲

۱. دانش آموخته دکترا مدیریت رسانه، پردیس البرز، تهران، ایران. (نویسنده مسئول)، رایانامه: safaeinejad@ut.ac.ir

۲. دانشیار، گروه مدیریت رسانه و ارتباطات کسب‌وکار، دانشکده مدیریت کسب‌وکار، دانشکده‌گان مدیریت، دانشگاه تهران، تهران، ایران. رایانامه: Sharifee@ut.ac.ir

اطلاعات مقاله

چکیده

نوع مقاله: مقاله پژوهشی

تاریخ دریافت: ۱۴۰۴/۷/۸

تاریخ بازنگری: ۱۴۰۴/۱۰/۸

تاریخ پذیرش: ۱۴۰۴/۱۲/۴

تاریخ انتشار: ۱۴۰۵/۵/۹

کلیدواژه‌ها:

افکار کاوی،

داده‌های کلان،

تحلیل احساسات،

سیاست‌گذاری،

شبکه‌های اجتماعی.



هدف: پژوهش حاضر با هدف طراحی چارچوب مفهومی افکار کاوی ملی مبتنی بر داده‌های کلان انجام شده است. ضرورت این تحقیق از آنجا ناشی می‌شود که روش‌های سنتی افکارسنجی، نظیر پرسشنامه و مصاحبه، با محدودیت‌هایی همچون هزینه‌بر بودن، تأخیر زمانی، و عدم پوشش گسترده جغرافیایی مواجه‌اند. در مقابل، داده‌های کلان حاصل از تعاملات دیجیتال شهروندان در شبکه‌های اجتماعی و فضای وب، ظرفیت بی‌سابقه‌ای برای رصد لحظه‌ای تحولات افکار عمومی و پشتیبانی از سیاست‌گذاری مبتنی بر شواهد فراهم می‌آورند.

روش تحقیق: این پژوهش از نوع توسعه‌ای-نظری و با رویکرد کیفی-استنتاجی انجام شد. داده‌های مورد استفاده از منابع ثانویه شامل مقالات علمی معتبر و گزارش‌های سیاست‌گذاری ملی و بین‌المللی گردآوری شده است. روش تحقیق مبتنی بر تحلیل نظری و سنتز مفهومی بود. در گام نخست، مرور نظام‌مند ادبیات برای شناسایی مفاهیم کلیدی و شکاف‌های پژوهشی صورت گرفت. سپس با بهره‌گیری از روش استنتاج مفهومی، مولفه‌ها و روابط اساسی فرایند افکار کاوی استخراج و در قالب یک مدل پنج‌لایه شامل: (۱) گردآوری داده، (۲) پاک‌سازی و پیش‌پردازش، (۳) تحلیل و خوشه‌بندی، (۴) تفسیر و اعتبارسنجی، و (۵) بازخورد و کاربرد، سازماندهی شد.

یافته‌ها: یافته‌های پژوهش نشان داد که مدل پیشنهادی قادر است محدودیت‌های افکارسنجی سنتی را برطرف کرده و امکان رصد لحظه‌ای، مقیاس‌پذیر و چندبعدی افکار عمومی را فراهم سازد. در لایه تحلیل، استفاده از الگوریتم‌های پردازش زبان طبیعی و یادگیری ماشین (تحلیل احساسات، مدل‌سازی موضوعی، خوشه‌بندی و تحلیل شبکه اجتماعی) امکان شناسایی روندهای احساسی، موضوعات غالب و بازیگران کلیدی را فراهم خواهد کرد. در لایه تفسیر، بازخورد خبرگان اخذ شد تا اعتبار مدل افزایش یابد. در نهایت، لایه بازخورد و کاربرد نشان داد که خروجی مدل می‌تواند به‌صورت داشبوردهای تعاملی و گزارش‌های سیاستی در اختیار نهادهای تصمیم‌گیر قرار گیرد تا برای تکمیل و تقویت افکارسنجی سنتی از طریق تحلیل داده‌محور مورد استفاده قرار گیرد.

نتیجه‌گیری: مدل مفهومی افکار کاوی ملی مبتنی بر داده‌های کلان، چارچوبی نوین برای گذار تدریجی در برخی کارکردها و مکمل‌سازی افکارسنجی سنتی از طریق تحلیل داده‌محور ارائه می‌دهد. این مدل با کاهش هزینه و زمان، افزایش دقت و پوشش، و امکان تحلیل لحظه‌ای، می‌تواند به ابزاری راهبردی برای سیاست‌گذاری مبتنی بر شواهد در سطح ملی تبدیل شود. با این حال، چالش‌هایی همچون نیاز به زیرساخت محاسباتی قدرتمند، رعایت اصول اخلاقی و حریم خصوصی، و بهینه‌سازی الگوریتم‌های پردازش زبان فارسی باید مورد توجه قرار گیرد. در مجموع، نتایج این پژوهش می‌تواند مبنای توسعه ابزارهای بومی افکار کاوی و ارتقای ظرفیت تصمیم‌سازی داده‌محور در ایران باشد.

استناد: صفائی نژاد، قاسم و شریفی، اسیدمهدی (۱۴۰۵). افکار کاوی ملی مبتنی بر داده‌های کلان: چارچوبی مفهومی برای سیاست‌گذاری مبتنی بر شواهد، فصلنامه علوم

<http://doi.org/10.22034/lrsi.2026.550018.1443>

خبری، ۱۵ (۲)، ۱۶۸-۱۴۸.



© نویسندگان.

ناشر: موسسه مطبوعاتی روزنامه نگاری بین‌المللی معاصر.

مقدمه

تحولات شتابان فناورانه و گسترش زیست‌بوم‌های دیجیتال، ساختار تولید و گردش اطلاعات عمومی را دگرگون ساخته است؛ امروز حجم وسیعی از داده‌های رفتاری شهروندان به‌صورت پیوسته، زمان‌مند و پراکنده تولید می‌شود. این دگرگونی، ابزارهای متعارف سنجش افکار عمومی که بر نمونه‌گیری میدانی و پرسشنامه متکی‌اند را در مقام پاسخ‌دهی به نیازهای سیاست‌گذاری به سه محدودیت اساسی دچار کرده است: هزینه بالا و زمان طولانی، پوشش مکانی و زمانی محدود و کاهش مشارکت پاسخ‌دهندگان که تعمیم‌پذیری نتایج را تضعیف می‌کند (میلر، ۲۰۱۷). در مقابل، داده‌های تولیدشده در شبکه‌های اجتماعی و جستارهای وب فرصتی برای رصد بلادرنگ و چندبعدی نگرش‌ها، احساسات و مباحث عمومی فراهم کرده‌اند (شانگ و همکاران، ۲۰۱۵؛ عبدلا، ۲۰۲۲).

شواهد تجربی بین‌المللی نشان می‌دهد که تحلیل داده‌های لحظه‌ای شبکه‌های اجتماعی در بسیاری از موارد همگرایی معناداری با نتایج پیمایش‌های سنتی دارد و در برخی موقعیت‌ها اطلاعات زمان‌مندتر و مقرون‌به‌صرفه‌تری را به سیاست‌گذار عرضه می‌کند (بووه، مورونه و ماکسه، ۲۰۱۶). با این حال، انتقال این ظرفیت‌ها به سطح ملی و به‌ویژه به متن‌های فارسی‌زبان با مجموعه‌ای از موانع فنی، زبانی و نهادی روبه‌رو است: کمبود واژه‌نامه‌ها و بانک‌های حسی بومی، پیچیدگی پردازش زبان فارسی در بافت‌های محاوره‌ای، دشواری دسترسی پایدار به منابع داده و همچنین ملاحظات حقوقی و اخلاقی مربوط به حریم خصوصی و نمایندگی داده‌ها (زاهد و سخی، ۱۳۹۵؛ حیدری، ۱۳۹۹).

بر این مبنا ضرورت پژوهش حاضر آشکار می‌شود: سیاست‌گذاری خرد و کلان در محیط کنونی نیازمند چارچوب‌هایی است که بتوانند خروجی‌های داده‌محور را به شاخص‌های قابل‌اعتماد و تصمیم‌پذیر تبدیل کنند؛ چارچوب‌هایی که هم از منظر روش‌شناختی قابلیت بازتولید و اعتبارسنجی داشته باشند و هم از منظر زبانی و نهادی متناسب با شرایط ملی طراحی شده باشند (تمپورائو، ۲۰۱۹). پاسخ به این نیاز، فراتر از انتخاب الگوریتم‌هاست و مستلزم طراحی فرایندی سازمان‌یافته است که از گردآوری و پاک‌سازی داده تا تحلیل، تفسیر و بازخورد سیاستی را پوشش دهد.

پرسش محوری این پژوهش آن است که: «چه چارچوب مفهومی‌ای می‌تواند تحلیل کلان‌داده‌های تولیدشده در زیست‌بوم‌های دیجیتال را به‌عنوان رویکردی مکمل و در برخی حوزه‌های کارکردی، جایگزین افکارسنجی سنتی ساماندهی کند، به‌گونه‌ای که خروجی‌های حاصل از آن، با سرعت بیشتر، هزینه کمتر و دقت بالاتر، قابلیت بهره‌برداری در حوزه‌های مختلف تصمیم‌سازی عمومی و تخصصی را داشته باشند؟» این پرسش ناظر بر طراحی یک چارچوب مفهومی است که بتواند فرایند افکارکاوی را به‌صورت مقیاس‌پذیر، چندلایه، و منطبق با ویژگی‌های زبانی، فرهنگی و اخلاقی جامعه ایرانی هدایت کند.

هدف این مقاله، طراحی یک چارچوب مفهومی برای افکارکاوی مبتنی بر کلان‌داده‌های تولیدشده در زیست‌بوم‌های دیجیتال است؛ چارچوبی که بتواند فرایند تحلیل افکار عمومی را به‌صورت نظام‌مند، مقیاس‌پذیر و منطبق با ویژگی‌های زبانی و فرهنگی جامعه ایرانی هدایت کند. این چارچوب بر سه محور اصلی تمرکز دارد: نخست، توجه به چالش‌های زبانی در تحلیل داده‌های فارسی‌زبان؛ دوم، پیشنهاد سازوکارهایی برای اعتبارسنجی نتایج از طریق مقایسه با منابع رسمی و مطالعات موجود؛ و سوم، تدوین ملاحظات نهادی و اخلاقی برای بهره‌برداری مسئولانه از داده‌های شهروندی. از منظر نوآوری، این مطالعه می‌کوشد مدلی تلفیقی ارائه دهد که مباحث فنی، زبانی و نهادی را هم‌زمان در نظر گرفته و مسیر توسعه افکارکاوی داده‌محور در ایران را روشن سازد.

ساختار مقاله به‌گونه‌ای تنظیم شده است که ابتدا با مرور تحلیلی ادبیات مرتبط با افکارسنجی سنتی و ظرفیت‌های داده‌محور در تحلیل افکار عمومی آغاز می‌شود؛ سپس چارچوب مفهومی پیشنهادی معرفی و مؤلفه‌های اصلی آن تبیین می‌گردد؛ در ادامه، چالش‌های اجرایی، ملاحظات اخلاقی و سازوکارهای اعتبارسنجی مورد بحث قرار گرفته و در پایان، پیشنهادهایی برای بهره‌برداری عملی و مسیرهای پژوهشی آتی ارائه می‌شود.

پیشینه پژوهش

افکارسنجی سنتی: مفاهیم، روش‌ها و محدودیت‌ها

افکارسنجی سنتی، یکی از ابزارهای اصلی در شناخت نگرش‌ها و ترجیحات عمومی در حوزه‌های اجتماعی، فرهنگی و سیاسی محسوب می‌شود. این روش‌ها عمدتاً مبتنی بر پرسشنامه‌های حضوری، تلفنی یا اینترنتی هستند و با نمونه‌گیری تصادفی از جامعه هدف، تلاش می‌کنند تصویری از افکار عمومی ارائه دهند (فرخی، ۱۳۸۳؛ شفر و بروکر، ۲۰۰۸). با وجود مزایایی چون ساختارمند بودن، امکان کنترل متغیرها و قابلیت تعمیم نتایج، افکارسنجی سنتی با چالش‌های قابل توجهی مواجه است:

نخست، محدودیت‌های نمونه‌گیری و کاهش نرخ پاسخدهی موجب شده است که نتایج بسیاری از نظرسنجی‌ها نماینده واقعی جامعه نباشند. مطالعات اخیر نشان داده‌اند که در سال‌های گذشته، به‌ویژه در انتخابات‌های ۲۰۱۶ و ۲۰۲۰ ایالات متحده، اعتماد عمومی به نظرسنجی‌ها کاهش یافته و مؤسسات افکارسنجی به بازنگری در روش‌های خود روی آورده‌اند (مرکز پژوهشی پیو، ۲۰۲۳).

دوم، هزینه‌بر بودن و زمان‌بر بودن اجرای نظرسنجی‌های گسترده، به‌ویژه در مقیاس ملی، موجب شده است که بسیاری از نهادهای سیاست‌گذار یا رسانه‌ای از انجام نظرسنجی‌های مکرر صرف‌نظر کنند یا به داده‌های ناقص بسنده نمایند (سن و وانگ، ۲۰۱۶).

سوم، خطاهای ناشی از سوگیری پاسخ‌دهندگان و عوامل محیطی، از جمله تمایل به پاسخ‌های مطابق هنجارها و ارزش‌ها یا تأثیر نحوه پرسشگری، موجب کاهش دقت و اعتبار نتایج شده است. پژوهش ژو و همکاران (۲۰۲۱) نشان داده‌اند که در برخی انتخابات، نظرسنجی‌ها به دلیل سوگیری‌های پنهان، نتوانسته‌اند نتایج واقعی را پیش‌بینی کنند.

چهارم، محدودیت در پوشش زمانی و جغرافیایی، به‌ویژه در جوامع متنوع و چندلایه، موجب شده است که افکارسنجی سنتی نتواند تحولات سریع و لحظه‌ای افکار عمومی را به‌درستی رصد کند (مرکز پژوهشی پیو، ۲۰۲۳؛ شفر و بروکر، ۲۰۰۸). در مقابل، ظهور رسانه‌های اجتماعی و تولید پیوسته داده‌های رفتاری و متنی، ظرفیت‌هایی برای تحلیل بلادرنگ و چندبعدی فراهم کرده است (ژو و همکاران، ۲۰۲۱).

در مجموع، اگرچه افکارسنجی سنتی همچنان جایگاه خود را در مطالعات اجتماعی حفظ کرده، اما محدودیت‌های آن زمینه‌ساز جستجوی رویکردهای مکمل یا جایگزین شده است؛ رویکردهایی که بتوانند با بهره‌گیری از داده‌های کلان، تحلیل سریع‌تر، دقیق‌تر و گسترده‌تری از افکار عمومی ارائه دهد.

افکار کاوی داده‌محور – ظهور، مزایا و چالش‌ها

با گسترش زیست‌بوم‌های دیجیتال و تولید انبوه داده‌های رفتاری، متنی و شبکه‌ای توسط کاربران، افکار کاوی داده‌محور به عنوان رویکردی نوین در تحلیل نگرش‌های عمومی مطرح شده است. برخلاف افکارسنجی سنتی که مبتنی بر پرسش مستقیم از افراد است، افکار کاوی داده‌محور تلاش می‌کند با تحلیل داده‌های تولیدشده در بسترهای طبیعی تعامل، از جمله شبکه‌های اجتماعی، وبسایت‌های خبری و پلتفرم‌های گفت‌وگو محور، به شناختی غیرمداخله‌گرانه و بلادرنگ از افکار عمومی دست یابد (ژائو و همکاران، ۲۰۲۳؛ شی و آگراوال، ۲۰۲۵).

یکی از مزایای اصلی این رویکرد، سرعت و مقیاس‌پذیری تحلیل است. داده‌های کلان به پژوهشگر اجازه می‌دهند تا در بازه‌های زمانی کوتاه، تغییرات افکار عمومی را در سطح وسیع رصد کند و روندهای نوظهور را شناسایی نماید. همچنین، هزینه پایین‌تر و عدم نیاز به نمونه‌گیری سنتی موجب شده است که افکار کاوی داده‌محور برای نهادهای رسانه‌ای، فرهنگی و سیاست‌گذار جذاب‌تر باشد (ژائو و همکاران، ۲۰۲۳).

با این حال، این رویکرد با چالش‌هایی نیز مواجه است. نخست، تنوع زبانی و سبک‌های نگارشی در داده‌های فارسی‌زبان -از جمله استفاده از اصطلاحات عامیانه، ایموجی‌ها، و کدهای فرهنگی- موجب شده است که بسیاری از الگوریتم‌های پردازش زبان طبیعی در تحلیل دقیق این داده‌ها ناکارآمد باشند (فرزین‌دار و اینکپن، ۲۰۱۶؛ تیم پژوهشی دانشگاه شهید بهشتی، ۱۴۰۲). دوم، مسائل مربوط به اعتبارسنجی نتایج، به‌ویژه در نبود منابع رسمی برای مقایسه، موجب شده است که خروجی‌های افکار کاوی گاه با تردید مواجه شوند. سوم، ملاحظات اخلاقی و نهادی در بهره‌برداری از داده‌های شهروندی -از جمله حفظ حریم خصوصی، رضایت ضمنی، و شفافیت در استفاده از داده‌ها- از جمله دغدغه‌های مهم در این حوزه هستند (دورن و همکاران، ۲۰۲۴؛ پاتایل، ۲۰۲۴). در مجموع، افکار کاوی داده‌محور ظرفیت‌هایی نوین برای تحلیل افکار عمومی فراهم کرده است، اما برای بهره‌برداری مؤثر از آن، نیاز به چارچوبی منسجم، بومی شده و چندلایه وجود دارد که بتواند این مزایا و چالش‌ها را هم‌زمان مدیریت کند. همین ضرورت، زمینه‌ساز طراحی چارچوب مفهومی پیشنهادی در این مقاله شده است.

پژوهش‌های مرتبط

افکار کاوی داده‌محور و تحلیل نگرش‌های عمومی با استفاده از کلان‌داده‌ها، به عنوان یکی از مهم‌ترین شاخه‌های پژوهشی علوم داده، علوم اجتماعی، و پردازش زبان طبیعی، در ده سال اخیر موجب تحول تحلیل رسانه‌های اجتماعی، سنجش احساسات جمعی، و سیاست‌گذاری مبتنی بر داده شده است. با توجه به رشد سریع حجم داده‌های تولیدشده در فضای مجازی، استفاده از کلان‌داده‌ها برای افکار کاوی و تحلیل احساسات نسبت به موضوعات مختلف، چالش‌ها و فرصت‌های جدیدی را پیش روی پژوهشگران قرار داده است. در اینجا مهم‌ترین پژوهش‌های انجام شده در جهان و ایران فهرست شده‌اند که چکیده خلاصه آن هم ارائه شده است:

جدول ۱- پیشینه پژوهشی افکار کاوی با استفاده از کلان‌داده

عنوان مقاله	نویسنده/نویسندگان	نام مجله	سال انتشار	چکیده مقاله
تحلیل احساسات چندمدلی در افکار کاوی	ژائو، لیو و وانگ	نشریه کلان‌داده	۲۰۲۳	این مقاله روش‌های یادگیری انتقالی چندمدلی را برای تحلیل احساسات در داده‌های کلان بررسی می‌کند و نشان می‌دهد که ترکیب داده‌های متنی و تصویری می‌تواند دقت افکار کاوی را افزایش دهد.
اعتبارسنجی تحلیل احساسات با داده‌های خوداظهاری	فدرستون و بارنت	مؤسسه ژنوم نوآور	۲۰۲۱	نویسندگان با مقایسه خروجی‌های تحلیل احساسات با داده‌های خوداظهاری کاربران، به بررسی اعتبار روش‌های افکار کاوی پرداخته‌اند.
افکار کاوی در متون کوتاه غیرانگلیسی	آکباش	مجموعه مقالات سیستم‌های هوشمند	۲۰۱۷	مقاله به چالش‌های تحلیل متون کوتاه در زبان‌های غیرانگلیسی پرداخته و الگوریتم‌های تطبیقی را پیشنهاد می‌دهد.
پردازش زبان طبیعی برای رسانه‌های اجتماعی	فرزین‌دار و اینکپن	زبان‌شناسی رایانشی	۲۰۱۶	این مقاله به بررسی محدودیت‌های پردازش زبان طبیعی در تحلیل داده‌های غیررسمی و محاوره‌ای در شبکه‌های اجتماعی می‌پردازد.
تکنیک‌های افکار کاوی در زبان‌های غیرانگلیسی	پاتیل	نشریه زبان‌شناسی رایانشی	۲۰۱۶	نویسندگان با مرور روش‌های موجود، به بررسی چالش‌های زبانی و فرهنگی در افکار کاوی غیرانگلیسی پرداخته‌اند.
تحلیل نگرش عمومی با داده‌های توییتر	وانگ و یوآن	نشریه تحلیل داده‌های اجتماعی	۲۰۲۲	مقاله با استفاده از داده‌های توییتر، تغییرات نگرش عمومی نسبت به سیاست‌های اقلیمی را بررسی می‌کند.
بیش از یک دهه افکار کاوی اجتماعی: یک مرور نظام‌مند	کیت کورتیس و برایان دیویس	نشریه مرور هوش مصنوعی	۲۰۲۱	نتایج نشان می‌دهد که افکار کاوی اجتماعی به‌عنوان ابزاری کلیدی برای درک نگرش عمومی در زمینه‌های سیاسی، اقتصادی و فرهنگی رشد چشمگیری داشته است. همچنین، چالش‌هایی همچون چندزبانه بودن داده‌ها، کوتاه بودن متن‌ها، و نیاز به روش‌های یادگیری عمیق و پردازش زبان طبیعی پیشرفته مورد بحث قرار گرفته است.

عنوان مقاله	نویسنده/نویسندگان	نام مجله	سال انتشار	چکیده مقاله
استفاده از روش‌های مدل‌سازی موضوعی برای داده‌های کوتاه: یک تحلیل مقایسه‌ای	رانیه ال‌بالوی، مراد بن‌یوسف و تت هین یپ	جبهه‌های هوش مصنوعی	۲۰۲۰	نویسندگان نشان می‌دهند که الگوریتم‌های سنتی مانند ال‌دی‌ای در داده‌های کوتاه عملکرد ضعیف‌تری دارند و روش‌های جدیدتر مبتنی بر بردارهای متراکم و یادگیری عمیق می‌توانند دقت بیشتری در شناسایی موضوعات و تحلیل افکار عمومی ایجاد کنند.
پیشرفت‌ها و چالش‌ها در تحلیل احساسات زبان عربی: یک دهه روش‌شناسی، کاربردها و توسعه منابع	علاذیمی و همکاران	هلیون	۲۰۲۴	این مقاله یک مرور جامع از پژوهش‌های انجام‌شده در زمینه تحلیل احساسات زبان عربی طی ده سال گذشته ارائه می‌دهد. نویسندگان به مشکلاتی مانند کمبود منابع زبانی، پیچیدگی‌های صرفی و نحوی، و نیاز به ابزارهای یادگیری عمیق برای زبان‌های راست‌چین اشاره کرده‌اند.
یک مرور جامع بر تحلیل احساسات معاصر زبان عربی	شی و آگراوال	یافته‌های انجمن زبان‌شناسی رایانشی آمریکای شمالی	۲۰۲۵	این مقاله مروری جامع بر پژوهش‌های اخیر در تحلیل احساسات زبان عربی ارائه می‌دهد. تمرکز آن بر چالش‌های خاص زبان‌های راست‌چین مانند ابهام در نشانه‌گذاری، نبود واژه‌نامه‌های استاندارد، و مشکلات پردازش متن‌های کوتاه شبکه‌های اجتماعی است.
اخلاق داده و حکمرانی آن: رویکردی مبتنی بر نظریه گفت‌وگو	برند کارستن اشتال	نشریه اطلاعات	۲۰۲۵	نویسنده با استفاده از رویکرد نظریه گفت‌وگو، چارچوبی برای تحلیل اخلاقی حکمرانی داده ارائه می‌دهد. مقاله تأکید می‌کند که حکمرانی داده باید نه تنها به کارایی و نوآوری توجه کند، بلکه به حریم خصوصی، شفافیت، رضایت شهروندان و مسئولیت‌پذیری نهادی نیز پایبند باشد.
مروری بر روش‌های تحلیل احساس در زبان فارسی	زینب رجبی و همکاران	فصلنامه پژوهش‌های توسعه اجتماعی	۱۴۰۱	این مقاله به بررسی روش‌های مختلف تحلیل احساسات در زبان فارسی پرداخته و ضمن مرور پژوهش‌های انجام‌شده، چالش‌های فنی و زبانی این حوزه را مطرح می‌کند. نویسنده نشان می‌دهد که پیچیدگی‌های صرفی و نحوی زبان فارسی، کمبود منابع زبانی استاندارد، و نبود مجموعه‌داده‌های بزرگ و برچسب‌خورده از جمله موانع اصلی توسعه ابزارهای تحلیل احساسات در این زبان هستند.
تجزیه و تحلیل احساسات چندوجهی در زبان فارسی با استفاده از مدل‌های مبتنی بر ترانسفورمر و کپسول‌های تشخیص احساس	رحمانیان و همکاران	پژوهش‌های نظری و کاربردی هوش ماشینی	۱۴۰۴	اگرچه تحقیقات اولیه در زمینه تحلیل احساسات عمدتاً تنها بر داده‌های متنی متمرکز بودند، اما مطالعات اخیر نشان داده‌اند که سیستم‌های چندوجهی که انواع مختلفی از داده‌ها را در برمی‌گیرند، می‌توانند عملکرد بهتری داشته باشند. در این پژوهش، به تحلیل احساسات چندوجهی در زبان فارسی پرداخته شده است و برای اولین بار در این زمینه، روشی براساس مدل‌های مبتنی بر ترانسفورمر پیشنهاد شده است.

شکاف پژوهشی و ضرورت مدل مفهومی جدید

با وجود رشد چشمگیر مطالعات افکارکاوی داده‌محور در سطح جهانی، در زمینه بومی‌سازی این رویکرد برای زبان فارسی و ساختار نهادی ایران، شکاف‌هایی جدی وجود دارد. بیشتر مدل‌های موجود بر پایه زبان انگلیسی و در بسترهای فرهنگی غربی طراحی شده‌اند و در مواجهه با ویژگی‌های زبانی، فرهنگی و نهادی جوامع غیرغربی -از جمله ایران- با محدودیت‌های مفهومی و عملیاتی مواجه‌اند (پاتیل، ۲۰۱۶؛ آکباش، ۲۰۱۷).

نخست، نبود چارچوب مفهومی منسجم برای افکارکاوی فارسی‌زبان موجب شده است که تلاش‌های موجود عمدتاً پراکنده، فنی‌محور و فاقد انسجام نظری باشند. در حالی که افکارکاوی در سطح ملی نیازمند مدلی چندلایه است که بتواند داده‌های متنی را از مرحله گردآوری تا مرحله بازخورد سیاستی هدایت کند، اغلب مطالعات داخلی صرفاً به تحلیل احساسات یا دسته‌بندی موضوعی محدود شده‌اند.

دوم، فقدان سازوکارهای اعتبارسنجی بومی -از جمله مقایسه با منابع رسمی، تطبیق با داده‌های میدانی، یا طراحی شاخص‌های ترکیبی- موجب شده است که خروجی‌های افکار کاوی در ایران کمتر مورد اعتماد نهادهای سیاست‌گذار قرار گیرند. در مقابل، مطالعات بین‌المللی روش‌هایی برای اعتبارسنجی خودکار و ترکیبی پیشنهاد کرده‌اند که در فضای فارسی‌زبان هنوز بومی‌سازی نشده‌اند (فدرستون و بارنت، ۲۰۲۰؛ ژائو و همکاران، ۲۰۲۳).

سوم، ملاحظات فنی خاص ایران -از جمله فیلترینگ پلتفرم‌های جهانی مانند اینستاگرام، توئیتر و تلگرام و عدم همکاری رسمی با این پلتفرم‌ها- موجب شده است که کاربران فارسی‌زبان عمدتاً با وی‌پی‌ان وارد این فضاها شوند. این امر باعث می‌شود که داده‌های رفتاری آن‌ها از نظر موقعیت جغرافیایی، هویت فنی و الگوهای مصرف، دچار تحریف شود و تحلیل‌های منطقه‌ای یا جمعیت‌شناختی با خطا همراه باشد. هرچند زبان فارسی می‌تواند نشانه‌ای قوی از ایرانی بودن باشد، اما نبود دسترسی به آی‌پی واقعی یا اطلاعات مکانی دقیق، قابلیت تعمیم تحلیل‌ها را کاهش می‌دهد.

چهارم، پلتفرم‌های بومی مانند شاد، بله، روبیکا و ای‌تا با وجود برخورداری از حجم بالای کاربران، هنوز زیرساخت‌های لازم برای تحلیل کلان‌داده را به‌صورت عمومی و پژوهشی فراهم نکرده‌اند. فقدان رابک‌های باز^۱، شفافیت داده‌ای، و مستندسازی فنی و البته فراگیر نبودن میان همه اقشار مختلف مردم موجب شده است که این منابع بالقوه، در عمل به‌عنوان منبع افکار کاوی قابل اتکا نباشند.

در مجموع، این شکاف‌ها نشان می‌دهند که برای بهره‌برداری مؤثر از ظرفیت‌های افکار کاوی داده‌محور در ایران، نیاز به طراحی یک مدل مفهومی بومی، چندلایه و تلفیقی وجود دارد؛ مدلی که بتواند هم‌زمان به چالش‌های فنی، زبانی، نهادی و اخلاقی پاسخ دهد و مسیر عملیاتی‌سازی افکار کاوی در سطح ملی را روشن سازد.

روش‌شناسی پژوهش

رویکرد کلی و نوع پژوهش

این پژوهش از منظر هدف، یک تحقیق توسعه‌ای - نظری محسوب می‌شود و رویکرد حاکم بر آن کیفی - استنتاجی است. هدف اصلی، طراحی یک چارچوب مفهومی جامع برای فرایند افکار کاوی ملی مبتنی بر داده‌های کلان است که بر پایه سنتز نظام‌مند ادبیات موجود بنا شده باشد. ماهیت توسعه‌ای-نظری پژوهش بدین معناست که تمرکز بر مدل‌سازی و استنتاج یک ساختار جدید از دل دانش پراکنده موجود است؛ لذا این پژوهش به جای آزمون تجربی فرضیه‌ها یا جمع‌آوری داده‌های میدانی (پیمایشی)، به دنبال تبیین روابط منطقی میان مؤلفه‌های داده کاوی و سیاست‌گذاری است.

روش پژوهش: فراترکیب کیفی نظام‌مند

روش اصلی به‌کار رفته در این مطالعه، فراترکیب کیفی^۲ است. بر خلاف مرورهای توصیفی رایج، هدف فراترکیب ایجاد تفسیری یکپارچه و جدید از یافته‌های مطالعات پراکنده است. (سندلوسکی، م.، و باروسو، ج.، ۲۰۰۷) در این روش، پژوهشگر با بررسی عمیق و انتقادی متون مرتبط، عناصر سازنده آن‌ها را استخراج و در قالب یک کل منسجم بازآرایی می‌کند. فرآیند اجرای پژوهش در پنج گام متوالی و تکرارشونده سازمان‌دهی شد: (۱) تنظیم سؤال پژوهش، (۲) جستجوی نظام‌مند منابع، (۳) انتخاب و پالایش منابع، (۴) استخراج داده‌ها و کدگذاری، و (۵) تلفیق و اعتبارسنجی.

جامعه آماری، نمونه و استراتژی جستجو

جامعه آماری پژوهش شامل کلیه اسناد علمی (مقالات پژوهشی) و گزارش‌های سیاستی معتبر در حوزه‌های «کلان‌داده»، «افکارسنجی» و «سیاست‌گذاری عمومی» بود. برای دستیابی به نمونه‌ای غنی و مرتبط، جستجوی نظام‌مند در پایگاه‌های اطلاعاتی زیر انجام شد:

^۱ - Open API

^۲ - Qualitative Meta-synthesis

- پایگاه‌های بین‌المللی: گوگل اسکولار^۱، وب آو ساینس^۲، ساینس دایرکت^۳، اسکوپوس^۴.
- پایگاه‌های داخلی: پایگاه مجلات تخصصی نور (نورمگز)، بانک اطلاعات نشریات کشور (مگیران) و پرتال جامع علوم انسانی.

بازه زمانی جستجو از سال ۲۰۱۰ تا ۲۰۲۵ میلادی (۱۳۸۹ تا ۱۴۰۳ شمسی) در نظر گرفته شد تا تمرکز بر مطالعات متأخر و مرتبط با ظهور فناوری‌های کلان‌داده باشد.

عبارات کلیدی جستجو بر اساس سه مفهوم اصلی پژوهش طراحی شدند:

جدول ۲- عبارات کلیدی جستجو برای استخراج مقالات از پایگاه‌های نمایه

مفاهیم اصلی	کلیدواژه‌های فارسی	کلیدواژه‌های انگلیسی
افکار کاوی و تحلیل	افکار کاوی، سنجش افکار عمومی، تحلیل احساسات، مدل‌سازی موضوعی	Opinion Mining, Sentiment Analysis, Topic Modeling
داده‌های کلان	کلان‌داده، داده‌های اجتماعی، شبکه‌های اجتماعی	Big Data, Social Media Data, Social Network Analysis
کاربرد سیاستی	سیاست‌گذاری مبتنی بر شواهد، کاربرد سیاستی، مدل‌سازی مفهومی	Evidence-Informed Policy, Conceptual Model, Framework

معیارهای ورود برای انتخاب مقالات عبارت بودند از: مقالات علمی-پژوهشی (داخلی و خارجی) و گزارش‌های معتبر سیاست‌گذاری، متون منتشرشده در بازه زمانی تعیین‌شده و تمرکز صریح بر ارتباط میان داده‌های کلان و سنجش یا تحلیل افکار عمومی. همچنین مطالعاتی که به طراحی چارچوب‌های فرایندی، الگوریتم‌های بومی یا چالش‌های زبان فارسی اشاره کرده باشند، مورد بررسی قرار گرفتند. در مورد حذف مقالات از دایره بررسی عمیق‌تر، معیارهای زیر مورد توجه قرار گرفتند: مقالات مروری غیرنظام‌مند، سرمقاله‌ها، خلاصه مقالات کنفرانس‌ها (بدون داوری معتبر)، مطالعاتی که صرفاً بر تحلیل فنی داده بدون تمرکز بر ابعاد اجتماعی-سیاستی متمرکز بودند و البته منابع دارای تاریخ انتشار خارج از بازه زمانی موردنظر.

فرآیند غربالگری: در مرحله نخست جستجو، تعداد ۲۴۵ سند شناسایی شد. پس از حذف موارد تکراری و بررسی عنوان و چکیده، تعداد ۹۸ سند که ارتباط مستقیمی با موضوع پژوهش داشتند انتخاب شدند. در مرحله نهایی، متن کامل این اسناد بر اساس معیارهای «اعتبار علمی»، «تمرکز بر فرآیندها» و «جامعیت» مورد ارزیابی دقیق قرار گرفت و در نهایت ۲۹ سند به‌عنوان منابع اصلی برای تحلیل و کدگذاری نهایی گزینش شدند.

جدول ۳- مراحل شناسایی، پالایش و انتخاب اسناد پژوهش

مرحله غربالگری	تعداد اسناد	توضیح
شناسایی اولیه	۲۴۵	نتایج جستجو در پایگاه‌های داخلی و بین‌المللی
پس از حذف موارد تکراری و بررسی عنوان/چکیده	۹۸	اسناد مرتبط با موضوع پژوهش
پس از بررسی متن کامل	۲۹	منابع نهایی وارد تحلیل و کدگذاری

به‌عنوان نمونه، در مقاله ژائو و همکاران (۲۰۲۳) به ضرورت استفاده از منابع چندگانه داده اشاره شده است. این مفهوم در مرحله کدگذاری باز با عنوان «تنوع منابع» ثبت شد. سپس در مرحله کدگذاری محوری به مقوله «کاهش سوگیری نمایندگی»

^۱ - Google Scholar

^۲ - Web of Science (WoS)

^۳ - ScienceDirect

^۴ - Scopus

تبدیل گردید. در مرحله انتخابی، این مقوله در لایه گردآوری داده‌ها جای گرفت. شواهد مربوط به این کد در دو منبع (ژائو و همکاران، ۲۰۲۳ و شی و آگراوال ۲۰۲۵) ثبت شد.

ابزار گردآوری و روش تجزیه و تحلیل (کدگذاری)

برای تحلیل داده‌ها از تکنیک تحلیل محتوای کیفی با رویکرد قیاسی-استقرایی استفاده شد. فرآیند استخراج مفاهیم طی سه مرحله کدگذاری انجام گرفت:

۱. کدگذاری باز: ابتدا متن کامل ۲۹ منبع منتخب با دقت مطالعه و گزاره‌های کلیدی مربوط به ابعاد فنی، روش‌شناختی و اخلاقی افکار کاوی شناسایی و برچسب‌گذاری شدند.
۲. کدگذاری محوری: کدهای باز بر اساس قرابت معنایی دسته‌بندی شدند تا مقوله‌های اصلی شکل گیرند (برای مثال، کدهای مربوط به «خزشگرها» و «رابک» در مقوله «گردآوری داده» جای گرفتند).
۳. کدگذاری انتخابی (مدل‌سازی): در این مرحله، با برقراری ارتباط منطقی میان مقوله‌های اصلی، ساختار نهایی مدل شکل گرفت و ترتیب توالی لایه‌ها مشخص گردید. تمامی مراحل در قالب «ماتریس شواهد» ثبت شد تا امکان ردیابی استنتاج‌ها فراهم باشد.

روایی و پایایی

از آنجا که این پژوهش از نوع کیفی و نظری است، برای تضمین کیفیت و اعتبار یافته‌ها از معیارهای لینکلن و گوبا (۱۹۸۵) استفاده شد:

- باورپذیری^۱: برای اطمینان از صحت کدگذاری‌ها، از روش سه‌سوسازی پژوهشگر^۲ استفاده شد؛ بدین صورت که نویسندگان مقاله به صورت مستقل اقدام به کدگذاری بخشی از منابع نمودند و سپس با مقایسه نتایج و بحث روی موارد اختلاف، به توافق نهایی^۳ دست یافتند.
- تأییدپذیری^۴: تمام مراحل پژوهش، از جستجو تا استخراج کدها، در جداول و ماتریس‌های جداگانه مستندسازی شد تا فرآیند پژوهش شفاف و قابل بازتولید باشد.
- تأیید خبرگان: در نهایت، ساختار استخراج‌شده (مدل نهایی) در اختیار ۶ نفر از متخصصان حوزه علم داده و مدیریت رسانه قرار گرفت و نظرات اصلاحی آنان در تنظیم نهایی روابط میان لایه‌ها اعمال گردید.

ملاحظات اخلاقی و حقوقی

اگرچه این پژوهش مبتنی بر مطالعه منابع ثانویه است و مستقیماً با آزمودنی‌های انسانی سروکار نداشته، اما ملاحظات اخلاقی در دو سطح رعایت و در مدل نهایی لحاظ گردید:

۱. اخلاق در پژوهش: رعایت کامل حقوق مالکیت فکری مؤلفان منابع اصلی و استناددهی دقیق به تمامی گزارش‌ها و مقالات مورد استفاده.
۲. اخلاق در مدل پیشنهادی (پاسخ به چالش حریم خصوصی): با توجه به حساسیت داده‌های شبکه‌های اجتماعی، در طراحی لایه‌های «گردآوری» و «تحلیل» مدل، اصل «حریم خصوصی در طراحی»^۵ به‌عنوان یک الزام در نظر گرفته شد. بدین معنا که در چارچوب استخراج‌شده، پروتکل‌هایی برای ناشناس‌سازی داده‌ها^۶ و تحلیل داده‌ها در سطح تجمعی^۷ و نه فردی گنجانده شده است تا از هرگونه نقض حریم خصوصی کاربران یا سوءاستفاده از اطلاعات جلوگیری شود.

^۱ - Credibility

^۲ - Investigator Triangulation

^۳ - Inter-coder consensus

^۴ - Confirmability

^۵ - Privacy by Design

^۶ - Anonymization

^۷ - Aggregate level

یافته‌های پژوهش

خروجی اصلی این پژوهش، که با اتکا به سنتز استقرایی^۱ و تحلیل انتقادی ۲۹ سند منتخب حاصل شده، طراحی یک چارچوب مفهومی است. این چارچوب برای هدایت نظام‌مند فرآیند افکارکاوی در مقیاس ملی و تبدیل داده‌های کلان اجتماعی به شواهد قابل اعتماد برای سیاست‌گذاری طراحی شده و در پنج لایه متوالی و متأثر از یکدیگر سازمان‌دهی شده است (شکل شماره ۱). نوآوری محوری این مدل در انعطاف‌پذیری، توجه ویژه به شرایط خاص ملی و الزامات بومی زبان فارسی و تضمین اصول اخلاقی در تمامی مراحل است، که آن را از مدل‌های عمومی و غیربومی متمایز می‌سازد.

مراحل سنتز و استخراج لایه‌های مدل

بر اساس کدگذاری محوری، مفاهیم استخراج‌شده از ادبیات موجود حول پنج محور اصلی خوشه‌بندی شدند. این پنج محور، نمایانگر فازهای ضروری یک سیستم افکارکاوی کلان‌داده‌ای هستند:

جدول ۴- مفاهیم استخراج‌شده از داده‌های ثانویه

مقوله محوری	تمرکز اصلی استخراج‌شده از متون	اهمیت در چارچوب پیشنهادی
گردآوری داده	الزامات فنی دسترسی، حجم و سرعت داده، پروتکل‌های امنیتی.	پایه و اساس مدل و تأمین جریان داده به‌صورت پیوسته.
پاک‌سازی و پیش‌پردازش	چالش‌های فنی داده‌های نامنظم ^۲ و پیچیدگی‌های زبانی (فارسی).	تضمین کیفیت داده ورودی برای جلوگیری از خطای تحلیلی.
تحلیل و مدل‌سازی	روش‌ها و الگوریتم‌های استخراج معنا از متن، شناسایی موضوع و احساسات.	قلب مدل و تبدیل داده به اطلاعات ساختاریافته.
تفسیر و اعتبارسنجی	ضرورت دخالت عامل انسانی، رفع سوگیری، تأیید صحت خروجی‌ها.	پل ارتباطی میان نتایج فنی و معنای سیاستی، تضمین قابلیت اعتماد ^۳ .
بازخورد و سیاست‌گذاری	استفاده کاربردی از نتایج در فرآیند تصمیم‌گیری و ایجاد چرخه یادگیری.	تکمیل چرخه و اتصال خروجی تحلیل به اقدام عملی.

تشریح لایه‌های چارچوب مفهومی

– لایه اول: گردآوری داده و پایش اخلاقی

این لایه بر مبنای حجم^۴، تنوع^۵ و سرعت^۶ داده‌های کلان استوار است.

- جمع‌آوری چندمنبعی: با هدف جلوگیری از سوگیری، تأکید بر جمع‌آوری داده از ترکیب شبکه‌های اجتماعی بومی، شبکه‌های اجتماعی خارجی، نظرات کاربران پلتفرم‌های خبری و پیام‌رسان‌ها است تا تنوع طیف وسیعی از نظرات تضمین شود.

^۱ – Inductive Synthesis

^۲ – Noise

^۳ – Trustworthiness

^۴ – Volume

^۵ – Variety

^۶ – Velocity

- پروتکل‌های اخلاقی: در پاسخ به چالش‌های حریم خصوصی، این مدل جمع‌آوری داده را تحت اصل «حریم خصوصی در طراحی»^۱ ضروری می‌داند. این مهم شامل اجرای دقیق پروتکل‌های ناشناس‌سازی داده‌ها^۲ و پرهیز از جمع‌آوری شناساگرهای فردی است. این الزام اخلاقی، قبل از هرگونه تحلیل، مشروعیت مدل را تضمین می‌کند.
- لایه دوم: پاک‌سازی و پیش‌پردازش زبان فارسی
این لایه حیاتی‌ترین گام در بومی‌سازی مدل است. اگرچه مدل‌های تحلیل متن در زبان انگلیسی پیشرفته‌اند، اما داده‌های فارسی با چالش‌هایی نظیر عدم یکدستی املائی، وجود اصطلاحات عامیانه و ساختارهای نوشتاری نامنظم مواجه‌اند که دقت مدل‌های تحلیلی را کاهش می‌دهد.
- عادی‌سازی و ریشه‌یابی: این مرحله شامل عادی‌سازی^۳ برای یکسان‌سازی کاراکترها (مثلاً تبدیل «ی» عربی به «ی» فارسی) و اجرای الگوریتم‌های ریشه‌یابی سطحی^۴ و ریشه‌یابی واژگانی^۵ فارسی است تا تحلیل‌های بعدی بر مبنای ساختار صحیح کلمات انجام گیرد.
- فرهنگ لغت احساسات بومی: برای افزایش دقت در تحلیل احساسات، توسعه و به‌روزرسانی مستمر یک فرهنگ لغت احساسات^۶ اختصاصی برای بافت فرهنگی و سیاسی ایران (با در نظر گرفتن کنایه‌ها و اصطلاحات خاص) ضروری است.
- لایه سوم: تحلیل و مدل‌سازی پیشرفته
پس از تصفیه داده، پردازش‌های تحلیلی برای استخراج هوشمندانه معنا و الگوهای رفتاری عمومی آغاز می‌شود.
- تحلیل دینامیک احساسات: علاوه بر تعیین قطبیت ساده (مثبت/منفی/خنثی)، تحلیل باید بر شدت احساسات و روندهای زمانی تغییرات آن تمرکز کند^۷ تا واکنش‌ها و نقاط عطف ناگهانی در افکار عمومی شناسایی شوند.
- مدل‌سازی موضوعی^۸: استفاده از الگوریتم‌های پیشرفته مانند ال‌دی‌ای^۹ و نسخه‌های توسعه‌یافته یا جایگزین آن برای متون کوتاه برای کشف موضوعات پنهان^{۱۰} و سرفصل‌های نوظهور در گفتگوهای اجتماعی، که ممکن است در نظرسنجی‌های سنتی مورد غفلت قرار گیرند.
- تشخیص ناهنجاری: سیستم باید توانایی تشخیص خودکار الگوهای نامتعارف در داده‌ها (مانند انتشار ناگهانی و هماهنگ یک پیام) را داشته باشد تا بتواند تأثیر احتمالی حملات سایبری یا کارزار رسانه‌ای هدفمند را از افکار عمومی ارگانیک تفکیک کند.
- لایه چهارم: تفسیر زمینه‌ای و اعتبارسنجی مفهومی
این لایه وظیفه دارد داده‌های پردازش‌شده را از سطح عدد و الگوریتم به سطح معنا و خرد اجتماعی ارتقا دهد. نتایج کمی و الگوریتمی باید در پرتو دانش میان‌رشته‌ای تفسیر شوند تا از جدایی از بافت فرهنگی و اجتماعی جلوگیری شود.
- تفسیر میان‌رشته‌ای: تحلیل‌ها باید توسط گروهی از متخصصان جامعه‌شناسی، علوم ارتباطات، سیاست‌گذاری و روان‌شناسی اجتماعی بررسی شوند تا معنا و پیامدهای واقعی داده‌ها روشن شود.
- توجه به بافت اجتماعی-فرهنگی: خروجی‌ها باید با در نظر گرفتن تفاوت‌های نسلی، جغرافیایی و گروه‌های اجتماعی تفسیر شوند تا سوگیری‌های احتمالی (مانند نادیده گرفتن روستاییان یا سالمندان) کاهش یابد.

^۱ - Privacy by Design

^۲ - Data Anonymization

^۳ - Normalization

^۴ - Stemming

^۵ - Lemmatization

^۶ - Sentiment Lexicon

^۷ - Sentiment Dynamics

^۸ - Topic Modeling

^۹ - LDA: Latent Dirichlet Allocation

^{۱۰} - Latent Topics

- اعتبارسنجی مفهومی: انسجام منطقی و قابلیت پیاده‌سازی مدل باید از طریق تأیید خبرگان سنجیده شود. این اعتبارسنجی تضمین می‌کند که مدل نه‌تنها نظری، بلکه عملی و قابل استفاده در سیاست‌گذاری است.
- تبدیل داده به خرد سیاستی: این لایه داده‌های خام و تحلیل‌های الگوریتمی را به توصیه‌های قابل اجرا برای سیاست‌گذاران تبدیل می‌کند.

۴. ۲. ۵. لایه پنجم: بازخورد و سیاست‌گذاری (چرخه اقدام)

- این لایه، نقطه اوج مدل و محل اتصال تحلیل به عمل است و بر پویایی و تداوم تأکید دارد.
- تسهیل تصمیم‌گیری: نتایج باید به جای گزارش‌های حجیم، در قالب داشبوردهای مدیریتی پویا، گزارش‌های سیاستی مختصر و ابزارهای بصری ارائه شوند تا بلافاصله قابل استفاده برای تصمیم‌گیرندگان باشند.
 - حلقه بازخورد: مدل پیشنهادی یک سیستم ایستا نیست. نتایج اقدام سیاستی باید مجدداً به سیستم بازگردد و باعث اصلاح پروتکل‌های جمع‌آوری داده (لایه ۱) یا بهبود الگوریتم‌های تحلیلی (لایه ۳) شود. این چرخه مداوم تضمین می‌کند که سیستم به‌صورت مستمر یادگیرنده باقی بماند.

جدول ۵- ساختار پنج‌لایه‌ای مدل تحلیل افکار عمومی مبتنی بر کلان‌داده

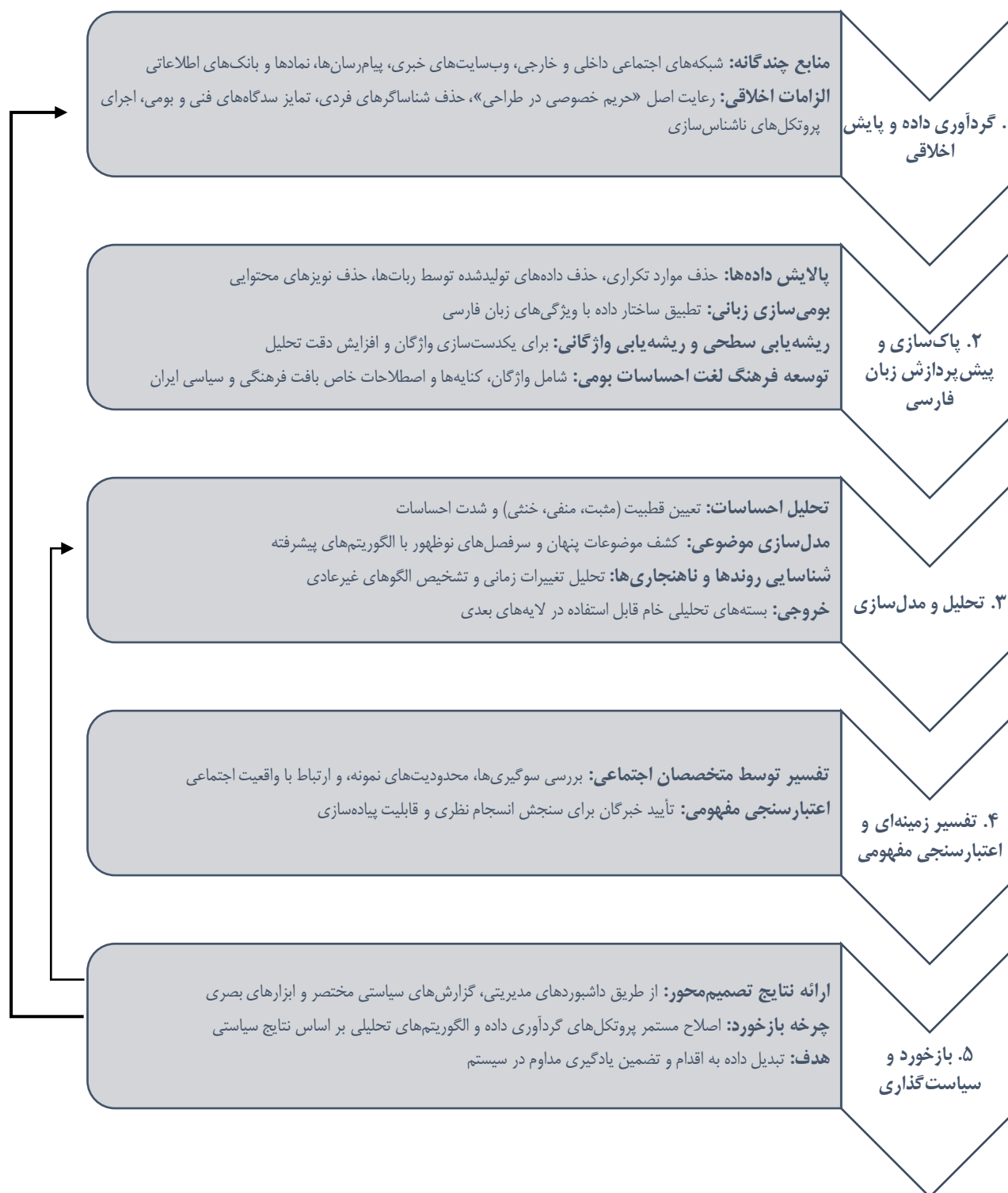
لایه مدل	هدف اصلی لایه	شاخص ارتباط با کلان‌داده (V's)	معیارهای ارزیابی عملکرد	خروجی سیاستی/عملیاتی اصلی
۱. گردآوری داده	تأمین داده با کیفیت و حجم مناسب	حجم و تنوع	پوشش منبع، نرخ موفقیت جمع‌آوری، پایداری به پروتکل اخلاقی (امتیاز رعایت اخلاق)	مجموعه داده خام
۲. پاک‌سازی و پیش‌پردازش	تضمین اعتبار و کاهش نویز داده‌های ورودی	صحت ^۱	نسبت داده قابل استفاده ^۲ ، نرخ کاهش نویز، کیفیت فرهنگ لغت بومی	مجموعه داده پاک‌سازی شده
۳. تحلیل و مدل‌سازی	تبدیل داده به اطلاعات ساختاریافته	سرعت	دقت، بازخوانی، امتیاز افیک (برای تحلیل احساسات) و شاخص‌های خوشه‌بندی (اطلاعات متقابل نرمال‌سازی شده / شاخص توافقی رند تعدیل شده) (برای مدل‌سازی موضوعی)	بینش‌های تحلیلی و مدل‌های موضوعی
۴. تفسیر و اعتبارسنجی	تضمین قابلیت اعتماد و ارتباط با واقعیت	صحت و ارزش	پایایی بین‌کدگذار (Cohen's κ)، سازگاری بین منابع، امتیاز تأیید خبرگان	شواهد تصمیم‌محور معتبر

^۱ - Veracity

^۲ - Usable Data Ratio

خروجی سیاستی/عملیاتی اصلی	معیارهای ارزیابی عملکرد	شاخص ارتباط با کلان‌داده (V's)	هدف اصلی لایه	لایه مدل
اقدام سیاستی و بهبود مستمر فرآیند	شفافیت ارائه، قابلیت تکرار نتایج، میزان زمان تصمیم‌گیری، نرخ استفاده از نتایج در سیاست‌ها	ارزش و سرعت	تبدیل شواهد به اقدام و ایجاد چرخه یادگیری	۵. بازخورد و سیاست‌گذاری

شکل ۱- چارچوب مفهومی افکار کاوی ملی مبتنی بر کلان داده



بحث

یافته‌های این پژوهش نشان می‌دهد که افکار کاوی مبتنی بر کلان داده می‌تواند افق تازه‌ای در مطالعه افکار عمومی ایران بگشاید. برخلاف رویکردهای سنتی که بر پیمایش‌های مقطعی و پرسشنامه‌های محدود متکی‌اند، این چارچوب امکان پایش مستمر، بلادرنگ و چندلایه را فراهم می‌سازد. بدین ترتیب، فاصله زمانی میان رخداد اجتماعی و تحلیل علمی کاهش یافته و ظرفیت واکنش سریع سیاست‌گذاران افزایش می‌یابد.

در ادبیات بین‌المللی، استفاده از داده‌های کلان برای تحلیل احساسات و مدل‌سازی موضوعی در شبکه‌های اجتماعی به‌عنوان ابزار پیش‌بینی روندهای انتخاباتی و اجتماعی شناخته شده است. یافته‌های این تحقیق با آن مطالعات هم‌خوانی دارد، اما تفاوت اصلی در تمرکز بر شرایط نهادی ایران و بومی‌سازی چارچوب است. چالش‌های خاص زبان فارسی (ریشه‌یابی واژگانی، فرهنگ لغت احساسات بومی، عادی‌سازی کاراکترها) و ضرورت پایش اخلاقی به‌عنوان یک لایه مستقل، این مدل را از نمونه‌های جهانی متمایز می‌سازد.

از منظر سیاست‌گذاری، نوآوری چارچوب پیشنهادی در اتصال مستقیم داده به تصمیم‌سازی است. لایه بازخورد و سیاست‌گذاری نشان می‌دهد که خروجی تحلیل صرفاً گزارش علمی نیست، بلکه به چرخه اقدام و اصلاح مستمر فرآیندها منتهی می‌شود. این ویژگی، مدل را از چارچوب‌های صرفاً فنی فراتر برده و آن را به ابزاری برای حکمرانی داده‌محور تبدیل می‌کند.

با این حال، محدودیت‌های مهمی نیز وجود دارد. داده‌های شبکه‌های اجتماعی در کلان‌شهرها به دلیل نفوذ گسترده اینترنت و دستگاه‌های هوشمند غنی‌تر و دقیق‌تر هستند، اما در مناطق کمتر برخوردار یا میان گروه‌های کم‌برخوردار، حضور محدودتر است و این امر می‌تواند به سوگیری در نتایج منجر شود. برای کاهش این انحراف، ترکیب افکار کاوی کلان داده با افکارسنجی سنتی ضروری است؛ به‌گونه‌ای که نتایج پیمایش‌ها به‌عنوان معیار اصلاحی به کار گرفته شوند و ضریب انحراف استخراج‌شده در تخمین‌های بعدی اعمال گردد.

فراتر از این، پیاده‌سازی موفق مدل نیازمند سرمایه‌گذاری در زیرساخت‌های فناوری، آموزش نیروی انسانی متخصص، تدوین چارچوب‌های اخلاقی و قانونی، و ایجاد داشبوردهای مدیریتی برای ارائه نتایج به سیاست‌گذاران است. تنها در این صورت است که افکار کاوی مبتنی بر کلان داده می‌تواند جایگزین یا مکمل پایدار افکارسنجی سنتی شود و به ابزاری مؤثر در حکمرانی ملی بدل گردد.

نتیجه‌گیری و پیشنهادها

این پژوهش با طراحی و ارائه مدل مفهومی افکار کاوی ملی مبتنی بر داده‌های کلان نشان داد که تکمیل و در برخی حوزه‌ها جایگزینی تدریجی پرسشنامه‌های سنتی با زنجیره ارزش داده‌محور نه تنها امکان رصد لحظه‌ای تحولات افکار عمومی را فراهم می‌کند، بلکه با مقیاس‌پذیری بالا می‌تواند پوشش جغرافیایی و موضوعی گسترده‌تری نسبت به روش‌های میدانی ارائه دهد. ساختار پنج‌لایه مدل شامل گردآوری، پاک‌سازی، تحلیل، تفسیر و بازخورد، پیوستگی فرایند افکار کاوی را از مرحله استخراج داده تا تصمیم‌سازی پشتیبانی می‌کند و امکان بهره‌گیری از رویکردهای احتمالاتی و فازی در توسعه‌های آتی مدل، می‌تواند اطمینان نتایج را افزایش دهد.

مدل پیشنهادی مزایای متعددی دارد: کاهش هزینه و زمان پردازش، امکان اندازه‌گیری احساسات و موضوعات واقعی کاربران، و ارائه خروجی‌های قابل فهم برای ذی‌نفعان سیاست‌گذار. در عین حال، نیاز به زیرساخت محاسباتی قدرتمند و تیم تحلیلگر متخصص، چالش حفظ حریم خصوصی و بهینه‌سازی الگوریتم‌های پردازش زبان فارسی را مطرح می‌کند. مقایسه مفهومی با افکارسنجی سنتی نشان داد که داده کاوی اجتماعی فراتر از تکمیل شکاف‌های روش‌های پرسشنامه‌ای، زمینه‌ای نوین برای تولید شاخص‌های سیاستی فراهم می‌آورد.

پیشنهادهای اجرایی

۱. توسعه و تقویت زیرساخت داده‌ای
 - سرمایه‌گذاری در مراکز داده و سرورهای قدرتمند برای پردازش حجم انبوه داده‌ها
 - به‌کارگیری پلتفرم‌های ابری با قابلیت گسترش پویا و تضمین امنیت اطلاعات
۲. ارتقای توانمندی تحلیل محتوا در زبان فارسی
 - حمایت از طرح‌های پژوهشی اختصاصی برای بومی‌سازی واژه‌نامه‌های احساس‌سنجی و موضوع‌یابی
 - برگزاری کارگاه‌های آموزشی تخصصی در حوزه NLP و تحلیل شبکه اجتماعی برای تحلیلگران
۳. طراحی چارچوب حقوقی و اخلاقی
 - تدوین دستورالعمل‌های شفاف برای دسترسی، ذخیره و کاربرد داده‌های کاربران
 - ایجاد کمیته نظارت بر اخلاق پژوهش‌های داده‌محور و رعایت حریم خصوصی
۴. گفتمان‌سازی بین نهادهای دولتی و بخش خصوصی
 - تشکیل کارگروه مشترک با نمایندگان دستگاه‌های حاکمیتی، وزارتخانه‌ها، پژوهشگاه‌ها و شرکت‌های فناوری
 - تبادل اطلاعات و داده‌های ساختاریافته برای ارتقای کیفیت مدل‌های پیش‌بینی

پیشنهادهای پژوهشی

- تعمیم مدل به حوزه‌های تخصصی: بررسی امکان تطبیق مدل در حوزه‌های سلامت، آموزش و محیط‌زیست با استخراج شاخص‌های موضوعی ویژه
- ارزیابی پویایی مدل در بحران‌ها: آزمون کاربرد مدل در مواجهه با بحران‌های ناگهانی (سیل، زلزله، اپیدمی) برای تحلیل سریع واکنش‌های عمومی
- تلفیق روش‌های داده‌محور و میدانی: طراحی مطالعات ترکیبی که نتایج مدل‌های کلان‌داده را با داده‌های میدانی کنترل‌شده مقایسه نماید
- بهبود منطق استنتاجی: بررسی کارایی شبکه‌های گراف عصبی بیزی و منطق ابری برای افزایش دقت در تحویل نتایج تحلیلی

محدودیت‌ها و چشم‌انداز آینده

اجرای کامل مدل نیازمند دسترسی پایدار به داده‌های شبکه‌های اجتماعی و منابع سیاست‌گذاری با فرمت ساختاریافته است که می‌تواند در برخی دوره‌ها با محدودیت مواجه شود. علاوه بر این، پیچیدگی الگوریتم‌های پردازش زبان فارسی ممکن است در حوزه‌های محاوره‌ای و زیربنایی با چالش دقت روبه‌رو گردد. در آینده، ادغام منابع رصد کلان‌داده با سنجش‌های میدانی و پایش رسانه‌های صوتی و تصویری می‌تواند ابعاد جدیدی از افکار عمومی را آشکار سازد و مدل را به چارچوبی فرابخشی بدل نماید. با توجه به نتایج و پیشنهادهای فوق، پیاده‌سازی مدل مفهومی افکار‌کاوی ملی مبتنی بر داده‌های کلان می‌تواند نقطه عطفی در ارتقای فرآیند سیاست‌گذاری مبتنی بر شواهد در کشور باشد و راه را برای پژوهش‌های کاربردی و توسعه ابزارهای نوآورانه در حوزه‌های مختلف هموار سازد.

تعارض منافع

بنا بر اظهار نویسندگان این مقاله تعارض منافع ندارد.

منابع

- زاهد، عطیه و سخی، محمدرضا. (۱۳۹۵). مروری بر تکنیک‌های کلان داده. اولین کنفرانس ملی برق و کامپیوتر، سیستم‌های توزیع شده و شبکه‌های هوشمند. ۲۸ آذرماه. دانشگاه آزاد واحد کاشان.
- حیدری، الهام. (۱۳۹۹). بررسی کلان داده‌ها در شبکه‌های اجتماعی. فصلنامه تحقیقات جدید در علوم انسانی. دوره جدید، شماره ۲۶، بهار، صص ۱۰۹-۱۳۰.
- رجبی، زینب؛ ولوی، محمدرضا، و حورعلی، مریم. (۱۴۰۱). مروری بر روش‌های تحلیل احساس در متون فارسی. *پژداژش علائم و داده‌ها*، ۱۰ (52)، ۱۰۷-132.
- رحمانیان، پریسا، موسوی، سید محمد رضا و صدرالدینی، محمد هادی. (۱۴۰۴). تجزیه و تحلیل احساسات چندوجهی در زبان فارسی با استفاده از مدل‌های مبتنی بر ترانسفورمر و کپسول‌های تشخیص احساس. *پژوهش‌های نظری و کاربردی هوش ماشینی*، 3(1)، 51-67. doi: 10.22034/abmir.2025.23347.1137
- فرخی، مصطفی. (۱۳۸۳). نظرسنجی در ایران، تهران: مرکز پژوهش‌های مجلس.

References:

- Abdalla, H. B. (2022). A brief survey on big data: Technologies, terminologies and data intensive applications. *Journal of Big Data*, 9, Article 107, 1–36. <https://doi.org/10.1186/s40537-022-00659-3>
- Akbas, E. (2017). Opinion mining on non-English short text. In M. Kryszkiewicz, A. Appice, D. Ślęzak, H. Rybinski, A. Skowron, & Z. Raś (Eds.), *Foundations of Intelligent Systems: ISMIS 2017 (Lecture Notes in Computer Science, Vol. 10352, pp. 1–10)*. Springer, Cham. https://doi.org/10.1007/978-3-319-60438-1_41
- Aladeemy, A. A., Alzahrani, A., Algarni, M. H., Khalaff, O. I., Wong, W.-K., & Aqbari, S. (2024). Advancements and challenges in Arabic sentiment analysis: A decade of methodologies, applications, and resource development. *Heliyon*, 10(21), e39786. <https://doi.org/10.1016/j.heliyon.2024.e39786>
- Albalawi, R., Yeap, T. H., & Benyoucef, M. (2020). Using topic modeling methods for short-text data: A comparative analysis. *Frontiers in Artificial Intelligence*, 3, Article 42. <https://doi.org/10.3389/frai.2020.00042>
- Atkin, C., & Gaudino, J. (1984). The impact of polling on the mass media. *The ANNALS of the American Academy of Political and Social Science*, 472(1), 119–128. <https://doi.org/10.1177/0002716284472001011>
- Bovet, A., Morone, F., & Makse, H. (2018). Validation of Twitter opinion trends with national polling aggregates: Hillary Clinton vs Donald Trump. *Scientific Reports*, 8, Article 6789, 1–16. <https://doi.org/10.1038/s41598-018-26951-y>
- Cen, Y., & Wang, Y. (2016). Social public opinion analysis and decision making support with big data. *Data Analysis and Knowledge Discovery*, 32(7), 3–11. <https://doi.org/10.11925/infotech.1003-3513.2016.07.02>
- Cortis, K., & Davis, B. (2021). Over a decade of social opinion mining: A systematic review. *Artificial Intelligence Review*, 54(7), 4873–4965. <https://doi.org/10.1007/s10462-021-10030-2>
- Farrokhi, M. (2004). Public opinion polling in Iran. Majlis Research Center (Iranian Parliament Research Center). (In Persian)
- Farzindar, A., & Louis, A. (2016). Natural language processing for social media [Book review]. *Computational Linguistics*, 42(4), 833–836. https://doi.org/10.1162/COLI_r_00270
- Featherstone, J. D., & Barnett, G. A. (2020). Validating sentiment analysis on opinion mining using self-reported attitude scores. In 2020 Seventh International Conference on Social Networks Analysis, Management and Security (SNAMS) (pp. 1–4). Paris, France. <https://doi.org/10.1109/SNAMS52053.2020.9336540>
- Heidari, E. (2020). A study of big data in social networks. *New Research Quarterly in Humanities, New Series*(26), 109–130. (In Persian)
- Lasswell, H. D. (1948). The structure and function of communication in society. In L. Bryson (Ed.), *The communication of ideas* (pp. 37–51). Harper & Row.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage Publications.
- Miller, P. V. (2017). Is there a future for surveys? *Public Opinion Quarterly*, 81(S1), 205–212. <https://doi.org/10.1093/poq/nfx008>
- Patil, A. S. (2016). Opinion mining techniques for non-English languages: An overview. *International Journal of Computational Linguistics (IJCL)*, 7(2), 1–16.
- Rahmanian, P., Mousavi, S. M. R., & Sadraldini, M. H. (2025). Multimodal sentiment analysis in Persian using transformer-based models and sentiment recognition capsules. *Theoretical and Applied Research in Machine Intelligence*, 3(1), 51–67. <https://doi.org/10.22034/abmir.2025.23347.1137> (In Persian)
- Rajabi, Z., Velvi, M. R., & Hourali, M. (2022). A review of sentiment analysis methods in Persian texts. *Signal and Data Processing*, 10(52), 107–132. (In Persian)
- Sandelowski, M., Barroso, J., & Voils, C. I. (2007). Using qualitative metasummary to synthesize qualitative and quantitative descriptive findings. *Research in Nursing & Health*, 30(1), 99–111. <https://doi.org/10.1002/nur.20176>
- Schaefer, T., & Brooker, R. (2005). *Public opinion in the 21st century: Let the people speak?* Houghton Mifflin Harcourt Publishing Company.
- Shang, S., Shi, M., Shang, W., & Hong, Z. (2015). Research on public opinion based on big data. In 2015 IEEE/ACIS 14th International Conference on Computer and Information Science (ICIS) (pp. 559–562). <https://doi.org/10.1109/ICIS.2015.7166655>

- Shi, Z., & Agrawal, R. (2025). A comprehensive survey of contemporary Arabic sentiment analysis: Methods, challenges, and future directions. In Findings of the Association for Computational Linguistics: NAACL 2025 (pp. 3760–3772). Albuquerque, NM: Association for Computational Linguistics.
- Stahl, B. C. (2025). The ethics of data and its governance: A discourse theoretical approach. *Information*, 16(6), 497. <https://doi.org/10.3390/info16060497>
- Temporão, M. (2019). Measuring public opinion using big data: Applications in computational social sciences* [Doctoral dissertation, Laval University]. Canada.
- Wang, Y., Guo, J., Yuan, C., & Li, B. (2022). Sentiment analysis of Twitter data. *Applied Sciences*, 12(22), 11775. <https://doi.org/10.3390/app122211775>
- Zahed, A., & Sakhi, M. R. (2016, December 19). A review of big data techniques [Conference presentation]. First National Conference on Electrical and Computer Engineering, Distributed Systems and Smart Networks, Islamic Azad University, Kashan Branch, Iran.
- Zhao, Z., Liu, W., & Wang, K. (2023). Research on sentiment analysis method of opinion mining based on multi-model fusion transfer learning. *Journal of Big Data*, 10, Article 155. <https://doi.org/10.1186/s40537-023-00837-x>