Digital Journalism and the Revitalization of Nusantara Languages: Integration of Natural Language Processing, Blockchain, and Digital Ethics for Linguistic Preservation

http://dx.doi.org/10.25008/jkiski

M. Yoserizal Saragih

Program Studi Ilmu Komunikasi, Universitas Islam Negeri Sumatera Utara
Jln. Lap. Golf No. 120, Kp. Tengah, Kec. Pancur Batu, Medan 20353 - Indonesia
Corresponding author: yoserizal@uinsu.ac.id

Abstract - In the ever-evolving digital era, regional languages in Indonesia are facing the threat of extinction due to modernization and globalization. This study aims to explore the role of digital journalism supported by artificial intelligence (AI), natural language processing (NLP), and blockchain technology in preserving Nusantara languages. The method used is descriptive analytical by reviewing literature from indexed scientific journals, UNESCO reports, and case studies related to language preservation and digital technology. The results of the study show that digital journalism plays an important role in documenting and disseminating regional languages through interactive formats such as videos, podcasts, and multimedia articles, which are more attractive to the younger generation. NLP technology helps analyze, translate, and build linguistic models of regional languages, facilitating the processing and distribution of multilingual content. In addition, the development of regional language-based chatbots opens up opportunities for fun interactive learning. On the other hand, blockchain ensures the authenticity and copyright protection of language content by permanently and encryptedly recording each documentation, and enables cultural tokenization to encourage active participation of local communities. This study confirms that the integration of digital journalism, NLP, and blockchain not only strengthens efforts to preserve regional languages, but also builds an ethical, transparent, and sustainable digital ecosystem. Collaboration between journalists, local communities, and technology developers is the key to success in ensuring that Nusantara languages remain alive and relevant in the modern era.

Keywords: Digital Journalism; Natural Language Processing; Blockchain; Language Preservation; Regional Languages

Introduction

In the rapidly developing digital era, journalism is undergoing a massive transformation, shifting from conventional formats to more dynamic and interactive digital formats. This change brings great opportunities to support the preservation of the increasingly endangered Nusantara languages. According to (UNESCO, 2022). More than 50% of the world's languages are threatened with extinction, including many regional languages in Indonesia. The presence of technologies such as Natural Language Processing (NLP), blockchain, and digital ethics principles opens up new

avenues for the revitalization and documentation of these languages through digital journalism platforms.

Language is a cultural identity that reflects the way of thinking, outlook on life, and values of a society. In Indonesia, the wealth of regional languages is very broad, with more than 700 languages living in various regions (Language Agency, 2020). However, along with the flow of globalization and the dominance of global language media, the use of regional languages has decreased drastically. The younger generation tends to be more familiar with Indonesian and foreign languages, while their mother tongue is starting to be forgotten. Digital journalism offers an effective medium to document, disseminate, and revive these languages in a more attractive and accessible format.

NLP technology is able to recognize, understand, and generate text in various languages, including less documented regional languages. Advances in NLP enable automatic text processing, translation, and mapping of more complex language structures. In the context of language preservation, NLP can help build digital corpuses of regional languages, create local language chatbots, and facilitate access to regional language-based education. Research (Zhang, 2020) shows how NLP can overcome the limitations of rare language resources with transfer learning methods and parallel data processing.

On the other hand, blockchain technology provides a solution to ensure the authenticity and durability of digitized language data. Blockchain works as a decentralized ledger that is difficult to manipulate, so that every documented language content will be recorded permanently and transparently. (Nakamoto, 2008) introduced this concept in the context of digital finance, but now blockchain is starting to be adapted to various sectors, including the preservation of cultural heritage. A study (Putra, 2022) highlights the potential of blockchain in protecting digital content from manipulation and copyright infringement, which is relevant in the context of language data preservation.

However, language preservation is not just about technology, but also ethics. Digital ethics is an important foundation in maintaining journalistic integrity, ensuring that the language preservation process is carried out with respect for local culture and the rights of native speaker communities. (Floridi, 2013) emphasized that information ethics must place respect on the owner of the information and ensure that technology does not simply exploit cultural resources without the consent of the community. Research conducted by (Aji, 2021) discusses the use of NLP in the development of digital dictionaries of regional languages, but has not integrated a journalistic approach as the main tool for dissemination. Meanwhile, research conducted by (Putra, 2022) examines the use of blockchain in protecting digital content, but has not been specifically applied to language preservation efforts. This study aims to bridge this gap by integrating the three elements of NLP, blockchain, and digital ethics in the context of digital journalism to strengthen the preservation of Nusantara languages.

Through this approach, it is hoped that regional languages will not only survive as cultural artifacts, but also live in modern communication practices. Digital journalism can be a new space for regional languages to develop, introducing the linguistic richness of the Archipelago to a global audience, while providing space for local communities to document and voice their own narratives. Integration of technology and ethics will ensure that this process is sustainable, transparent, and respects cultural rights. The role of digital journalism in supporting the preservation of regional languages is increasingly significant along with the development of the AI-based media ecosystem.

In Indonesia, more than 700 regional languages are facing the threat of extinction due to modernization and lack of documentation. In the context of digital journalism, technology-based media can play an important role in publishing, documenting, and building awareness of the preservation of regional languages. This article examines how artificial intelligence (AI), NLP, and blockchain can be utilized by journalists and academics in preserving regional languages in the digital space. The introduction title and content should be in Times Roman, font size 11. Please make sure your article meets the word count requirement of 4000 – 8000 words. Your article may be returned to you if these requirements are not met. The first paragraph in each section should be left-aligned and each following paragraph should be indented. Text should be single-spaced with no space between paragraphs.

Theoretical Framework

Digital journalism has become a new force in disseminating information globally. In the context of preserving regional languages, digital platforms have great potential to document and disseminate endangered languages (Smith, 2024). Digital media allows regional languages not only to survive but also to thrive through the presentation of more interactive and engaging content for young audiences (Rahman, 2024).

According to (UNESCO, 2023) one of the factors causing language extinction is the lack of exposure in the public space. Digital journalism, with its wide reach and rapid adaptability, opens up space for regional languages to remain relevant and heard, even amidst the dominance of global languages such as English and Mandarin.

Natural Language Processing (NLP) is a branch of artificial intelligence that aims to enable machines to understand, analyze, and produce human language effectively. In the context of language revitalization, NLP plays a crucial role in documenting, translating, and learning linguistic patterns of regional languages (Zhang & Liu, 2020). Research by Zhang, Wang, and Li (2024) highlights that the biggest challenge in applying NLP to endangered languages is the limited availability of data. However, advances in large language models are beginning to enable low-resource language processing, including regional languages in Indonesia.

Open AI (2023) also adds that the integration of NLP in digital platforms provides great opportunities in the development of interactive dictionaries, automatic translators, and audio-visual-based content to support local language learning.

Blockchain is a decentralized technology that allows for permanent and secure data storage. In the context of language preservation, blockchain can guarantee the authenticity and integrity of the collected linguistic data (Donoso & Sánchez, 2023). (Putra, 2022) emphasized that blockchain is able to protect digital content from manipulation, ensuring that language documentation remains authentic and cannot be changed by irresponsible parties. This is very important in preserving dialects and cultural narratives that are often vulnerable to modification over time. In addition, (Simamora, 2024) in his literature review explained that blockchain can be used to record the origins and cultural context of a language, so that local communities remain recognized as the rightful owners of the linguistic heritage.

In an effort to integrate technology into language preservation, digital ethics must be a primary principle. (Floridi, 2013) emphasized the importance of information ethics in ensuring responsible and fair use of technology. In this context, (Rahman, 2024) highlighted that the use of AI and blockchain must pay attention to the rights of local communities, both in recognition as cultural owners and in the use of language documentation results. This is in line with the UNESCO principle (2022) which encourages active community involvement in documenting their own languages, so that the revitalization process becomes more authentic and sustainable.

Material and Methodology

This research is descriptive analytical by reviewing the literature in depth. The focus is on identifying, comparing, and evaluating technology models that have been applied in language preservation in various parts of the world. Data The data used are secondary data in the form of scientific articles from Scopus Q1 indexed journals, official UNESCO reports, publications from language institutions, and case studies relevant to the use of NLP and blockchain technology in the context of cultural preservation (Arifin, 2020).

The researcher collected core literature discussing digital journalism, NLP, blockchain, and digital ethics. Although this research is a library study, validation is still carried out through literature triangulation (Danuri, 2019). The results of the study will be compared with current research trends and recommendations from international institutions such as UNESCO (Sundari, 2024).

Result and Discussion

Digital journalism plays an important role in introducing, documenting, and disseminating regional languages to a wider audience. Through multimedia formats such as videos, podcasts, and interactive articles, regional languages can be presented in an interesting and relevant context for the younger generation. Research by (Aji, 2021) shows that digital media can increase exposure to regional languages by up to 60% more effectively than traditional print media.

In addition, digital journalism allows local narratives to be voiced directly by native speaker communities, strengthening authenticity and public trust. Research by (Setiawan, 2023) highlights how community-based journalism platforms can revive young people's interest in local languages and cultures in various regions. With the integration of NLP and blockchain technology, this process becomes more effective and sustainable. NLP helps translate, analyze, and document regional languages automatically, while blockchain ensures the authenticity and copyright protection of the language content.

Interview with Mrs. Ratna, Old Sundanese Speaker, Cikondang Village, West Java. What is the current state of Old Sundanese usage in the community?

"When I was little, almost everyone in this village used Old Sundanese in their daily conversations, especially the older people. Now, the kids use more Indonesian, even English when playing games or watching YouTube. I'm worried that in the future only me and a few other older people will be able to speak this language."

Amidst the rapid flow of globalization, many regional languages in Indonesia are experiencing a decline in active speakers. Our investigation team visited three areas with a high risk of language extinction: West Java (Old Sundanese). News portals such as Kompas, BBC Indonesia, and Mongabay have begun actively documenting local culture, although they are still predominantly in Indonesian text and video formats. Meanwhile, innovative AI-based projects such as Google Translate and Wikitongues have begun to expand into rare languages, although they are not yet specific to the Nusantara context. Digital journalism plays an important role in introducing, documenting, and disseminating regional languages to a wider audience.

The use of AI in journalism enables automation of translation, linguistic analysis, and distribution of regional language content more widely and quickly. This technology helps translate regional languages in real-time, analyzes linguistic structures, and produces content in video, text, and audio formats that are more attractive to the younger generation. With the integration of NLP and blockchain technology, this process becomes more effective and sustainable. NLP helps translate, analyze, and document regional languages automatically, while blockchain ensures the authenticity and copyright protection of the language content.

Blockchain plays an important role in ensuring the authenticity and preservation of digitized regional languages. Research (Casado, 2018) shows that blockchain is effective in maintaining the integrity of cultural data. In this context, every documentation of regional languages, from texts, native speaker audio, to journalistic content — is recorded in an encrypted and immutable block. This technology provides copyright protection and ensures that the documentation remains authentic. Thus, blockchain becomes an ethical and technological foundation in the effort to revitalize the Nusantara language.

In the context of preserving Nusantara languages, the authenticity and credibility of language data are crucial. Blockchain, as a decentralized technology that is resistant to manipulation, offers a revolutionary solution in ensuring the authenticity of regional language documentation. Through a distributed recording system, every entry, whether in the form of text, audio, video, or NLP-based translation results, is recorded in an encrypted and immutable block without leaving a trace.

Blockchain functions as an immutable digital archive, where every documented language data will be stored permanently and publicly verified. This ensures that content cannot be manipulated or claimed by unauthorized parties. For example, interviews with native speakers in a particular regional language can be uploaded to a blockchain network, ensuring that its authenticity is recognized globally and cannot be changed for commercial purposes without the consent of the local community.

A study by (Casado, 2018) states that blockchain is able to ensure data integrity in cultural preservation projects, providing assurance that the information stored is the original and authentic version. In the context of digital journalism, blockchain functions not only as an archive keeper, but

also as a tool to protect the copyright of regional language content produced by local communities. Furthermore, blockchain can support cultural tokenization schemes, where the contribution of local communities in documenting their languages can be recognized and rewarded digitally.

This encourages the active participation of native speakers to be involved in preserving their own languages. Each contributor can receive an award in the form of a digital token or public recognition that is permanently recorded in the blockchain. Thus, the integration of blockchain in digital journalism not only strengthens the documentation of regional languages, but also provides assurance of authenticity, protection of intellectual property rights, and empowerment of the speaker community. This makes blockchain an ethical and technological foundation in efforts to revitalize the Nusantara language in the digital era.

Conclusions

This study confirms that the integration of digital technology, especially digital journalism, NLP, and blockchain, opens up great opportunities in efforts to preserve and revitalize regional languages in Indonesia. Digital journalism plays an important role in introducing and distributing regional languages to a wider audience through engaging and easily accessible multimedia content.

NLP technology strengthens this effort by enabling the automation of documentation, translation, and analysis of the linguistic structure of languages that have so far been under-documented. Meanwhile, blockchain provides a guarantee of authenticity and protection of copyright for regional language content, making it a key technology in maintaining the authenticity and integrity of linguistic data. Blockchain also enables the involvement of speaker communities through a digital incentive system that records their contributions transparently and permanently.

With an approach based on digital ethics, the active participation of local communities is very important so that the language revitalization process is not exploitative, but collaborative and sustainable. Therefore, the synergy between technology and local wisdom is the main foundation in preserving the linguistic heritage of the archipelago in this digital era.

References

Aji, A. (2021). *Digitalisasi Bahasa Daerah dalam Media: Penguatan Identitas Melalui Teknologi*. Yogyakarta: Pustaka Bahasa.

Casado, L. (2018). Blockchain technology for cultural heritage: Preserving authenticity and ensuring integrity in the digital age. *Journal of Cultural Preservation Technology*, 12(3), 45–58.

Donoso, C., & Sánchez, M. (2023). Blockchain for Linguistic Data Preservation: A New Paradigm. *Journal of Artificial Intelligence Research*, 58(4), 342–365.

Fauziah, A., Angreni, D., Pradjaningsih, A., Riski, A., & Hadi, A. F. (2020). *Support vector regression in statistical downscaling for rainfall forecasting*. IOP Conference Series: Earth and Environmental Science, 500(1), 012009. https://doi.org/10.1088/1755-1315/500/1/012009.

Floridi, L. (2013). *The Ethics of Information*. Oxford: Oxford University Press.

Izuchukwu, O. C. (2021). *Migration factors and the post-conflict reconstruction policies in Juba* [Master's thesis, Eastern Mediterranean University]. EMU Theses. http://i-rep.emu.edu.tr:8080/xmlui/handle/11129/5351.

Kemendikbudristek. (2024). *Laporan Vitalitas Bahasa Daerah di Indonesia*. Jakarta: Badan Pengembangan dan Pembinaan Bahasa.

Manurung, A. S., & Humaizi, I. Z. The Influence of Watching TikTok Content@ buiramara on Knowledge and Motivation to Complete Final Projects Among Communication Science Students at UIN Sumatera Utara.

Manurung, A. S., & Manurung, P. (2024). Tourism Communication Strategy in Developing Halal Tourism in The Lake Toba Tourism Area of North Sumatra. *Jurnal Nomosleca*, *10*(1), 128-142.

Manurung, A. S., & Pohan, S. (2024). Pengaruh Kualitas Pelayanan Publik Dan Citra Institusi Terhadap Kepuasan Mahasiswa Ilmu Komunikasi Uin Sumatera Utara Yang Berdampak Pada Word Of Mouth. *Jurnal Peurawi: Media Kajian Komunikasi Islam*, *7*(1), 129-146.

Octavianto, A. W., Priyonggo, A., & Setianto, Y. P. (2023). Framing the future: Exploring AI narratives in Indonesian online media using topic modelling. *Jurnal Komunikasi Indonesia*, 13(2), 2.

Rahman, A. (2024). Kajian Literatur: Adopsi Artificial Intelligence (AI) dalam Bidang Jurnalistik dan Dampaknya. *Journal of Digital Innovation*, 17(3), 85–102.

Setiawan, R. (2023). Revitalisasi Bahasa Daerah melalui Jurnalisme Komunitas Berbasis Digital. *Jurnal Komunikasi dan Kebudayaan Lokal*, 7(2), 102–115.

Simamora, S. (2024). Systematic Literature Review dengan Metode PRISMA: Dampak Teknologi Blockchain terhadap Periklanan Digital. *Jurnal Ilmiah M-Progress*, 14(1), 45–60.

Smith, J. (2024). The Role of AI in Digital Journalism and Language Revitalization. *International Journal of Digital Media Studies*, 52(2), 98–124.

UNESCO. (2022). *World Report on Language Endangerment*. Paris: UNESCO Publishing. Retrieved from https://unesdoc.unesco.org/

UNESCO. (2023). *Endangered Languages and Digital Preservation Strategies*. Paris: UNESCO Publishing.

Zhang, L., Wang, X., & Li, J. (2024). NLP Challenges in Low-Resource Language Processing: A Case Study in Endangered Language Documentation. *Language Resources and Evaluation Journal*, 58(1), 73–92. https://doi.org/10.1007/s10579-024-09765-8

Zhang, L., Wang, Y., & Li, H. (2024). Natural Language Processing for Endangered Languages: Challenges and Opportunities. *IEEE Transactions on Computational Linguistics*, 67(1), 112–138.

Zhang, T., & Liu, Y. (2020). Applications of Natural Language Processing in Regional Language Documentation. *Computational Linguistics Review*, 9(2), 85–98.