Issues in Information Systems
Volume 25, Issue 2, pp. 379-396, 2024

DOI: https://doi.org/10.48009/2_iis_2024_130

Big data usage, natural language processing, and ethics in developing countries: instrument development using systematic literature review

Assion Lawson-Body, *University of North Dakota, assion.lawsonbody@und.edu*
Jason Jensen, *University of North Dakota, jjensen@business.und.edu*
Laurence Lawson-Body, *University of North Dakota, laurence.lawsonbody@und.edu*

Abstract

Artificial intelligence (AI) and big data analytics technology are widely used globally. However, the challenges both present to organizations and individuals differ in developing and developed countries. This study aims to develop measurement instruments using a systematic literature review to assess the usage of big data analytics in developing countries, considering the ethical and AI components of natural language processing (NLP) challenges. The PRISMA method, a systematic review protocol, was considered for developing measurement instruments using structural (that relates the latent constructs) and measurement (that relates the latent constructs to their indicators) frameworks. About 136 articles were extracted from the publication databases. After exclusion and inclusion criteria were used, 99 papers were extensively processed and analyzed. The findings revealed 29 items measuring big data analytics usage in developing countries when NLP is moderating. Specifically, we found five items measuring the latent construct of smartphone record data, six items measuring the latent construct of big data location analytics, five items measuring the latent construct of big data analytics family ties, seven items measuring the latent construct of big data analytics ethics, and six items measuring the latent construct NLP in developing countries. Theoretical and practical contributions were offered .

Keywords: big data analytics, developing countries, ethics, NLP, measurement development, systematic review

Introduction

Artificial intelligence (AI) and big data analytics technology are not just widely applied in business, healthcare, engineering, etc., but they are a global phenomenon (Zhou, 2024). Big data analytics, the process of examining large and varied data sets to uncover hidden patterns, unknown correlations, market trends, customer preferences, and other helpful information, spans the entire globe and has become a technology accessible in both developed and developing countries (Lawson-Body et al., 2024). AI represents a new digital frontier that fundamentally reshapes our lives and work (Guliyev et al., 2023). AI has become critical to processing big data because not all traditional tools can do so due to increased data volume, diversification of data types and sources, and the data change rate in society (Li et al., 2022). In this sense, big data analytics use data extraction, transformation, and loading techniques to extract AI from big data collected from social media, the Internet of Things, sensor technologies, radio frequency identification (RFID) technologies, near-field communication, etc. (Li et al., 2022). Individuals in developing countries have enormous opportunities to improve their lives using big data and AI's natural

language processing (NLP). In recent efforts, they have used mobile phones, social media platforms, geospatial technologies, and crowdsourcing tools to redefine how they share, collect, and analyze intelligence (Dubey et al., 2022).

However, the challenges big data brings to organizations and individuals are reflected in its analytics usage. Those challenges are not the same in both developing and developed countries. For instance, in developing countries, the lack of infrastructure and resources can hinder the effective use of big data analytics (Lawson-Body et al., 2024). Additionally, economic growth is more prevalent in developed than developing countries. Such discrepancies occur due to poor socioeconomic conditions, low education, culture, and social trends in developing countries (Iqbal et al., 2022). Due to the lower cost of big data analytics, poverty is no longer a massive barrier for individuals to adopt big data analytics in developing countries. Individuals must overcome challenges to effectively derive value from big data analytics in developing countries, such as aligning AI's NLP and big data strategies and factoring in ethical and cultural issues (Dwivedi et al., 2023).

In developed countries, debates have centered on protecting individual rights in a big data environment, which often involves machine learning and AI (Mullins et al., 2021). These debates are nonexistent in developing countries. The United States (US) government has proposed a regulatory framework to control the widespread use of AI (Guidici et al., 2024). An example is the American AI risk management framework. Another example is in Europe, where the General Data Protection Regulation (GDPR) framework exists to protect the access and processing of personal data (Lepri et al., 2021). None of these frameworks exist in developing countries to protect big data analytics user rights. Most people in developing countries do not write or read. Therefore, NLP allows them to speak and talk to big data applications, making it more convenient to use big data in those countries. However, this convenience comes with ethical implications. For instance, using big data in developing countries could lead to privacy breaches, discrimination, or misuse of personal information.

Therefore, researchers must answer two questions: Is big data usage in developing countries ethical? Does the advent of NLP exacerbate these ethical issues of big data usage? What are the new instruments for measuring big data analytics dimensions, including big data analytics ethics and NLP, and how do they relate to developing countries' big data analytics usage?

This study aims to develop measurement instruments using a systematic literature review to assess the usage of big data analytics in developing countries, considering the ethical and NLP challenges. More specifically, this study examines the impact of big data dimensions such as big data location analytics, big data analytics family ties, smartphone record data, and big data analytics ethics on big data usage in developing countries. It also aims to investigate the moderating role of the AI technique of NLP on the relations between these big data dimensions and big data usage in the same countries.

As a result, this study contributes to the literature by first building a bridge between big data analytics and AI. Second, it is among the first to analyze big data analytics ethics in developing countries. Third, this research sheds light on NLP's moderating role in using big data dimensions and analytics in developing countries. Fourth, this research adds a new measurement instrument to the list of existing measurement tools of big data analytics and AI. Fifth, this study used a systematic literature review based on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) model to propose instruments that organizations and governments can use as a first step or beginning point toward the development of regulations and ethical guides for big data analytics and AI in developing countries.

This paper is organized as follows. Section 2 outlines related work. Section 3 proposes the structural framework. Section 4 presents the research methodology. Section 5 formulates the data analysis and results including the measurement framework. Section 6 details the discussion. Finally, section 7 provides the conclusion.

Related Work

In terms of relations between AI and big data analytics, the critical role of AI is to generate value by providing insights from large data sets and capturing structured interpretations of large unstructured data sets (Dubey et al., 2022). This is because training AI models requires immense volumes of data (Li et al., 2023). In addition, AI is trained iteratively with more accurate data (Li et al., 2023). As a result, big data analytics usage enables the discovery of new functions and relationships, which determine the performance of AI. In this sense, big data will have no usage without AI applications (Dubey et al., 2022); simultaneously, AI heavily depends on big data. However, the economic impact of AI and big data analytics varies across different regions, influenced by each region's specific economic activities and technology readiness (Guliyev et al., 2023). In developing countries, individuals use big data to help create value by tackling complex issues and problems in less time at a relatively low cost (Dubey et al., 2022).

There have been a few AI applications and big data analytics in developing countries compared to developed countries. For instance, Zhou (2024) found that big data analytics and AI can help improve the quality and dissemination efficiency of news content in developing countries. This author added that AI has high learning and reasoning abilities, while big data technology has real-time and value-density characteristics (Zhou, 2024). On their side, Guliyev et al. (2023) found that implementing big data and AI technologies in economic processes can effectively reduce employment rates and boost wages by creating job opportunities. AI technologies used historical data related to environmental pollution to build predictive models (Fu et al., 2023). In this sense, AI technologies gather all kinds of ecological environment monitoring data. AI technologies enable comparing and communicating big data (Fu et al., 2023).

The AI technique of NLP is a key global driver of business opportunities for individuals in developing countries. Extracting valuable data from NLP using unsupervised and supervised machine learning techniques is a fundamental task (Li et al., 2020). The big data behind NLP comes in many different forms, such as keyword extraction, key-phrase extraction, and named-entity recognition (Li et al., 2020). NLP has been facilitating the development of Chatbots, which are changing and redefining the global landscape. In particular, chatbots are conversational or interactive agents that respond instantly to users (Okonkwo and Ade-Ibijola, 2021). Chatbots and big data are connected because Chatbots support unlimited and high-volume conversations coming from different sources and various formats. The introduction of AI in developing countries allows the integration of Chatbot systems into multiple aspects of life and businesses. For instance, deep learning and machine learning techniques improved healthcare in sub-Saharan Africa (Mbunge and Batani, 2023). These AI models analyzed and derived inferences from massive health data for early detection and diagnosis of diseases (Mbunge and Batani, 2023).

Big Data Analytics Dimensions

Tsumura et al. (2022) and Onyeji-Nwogu et al. (2020) found that smartphone record data in developing countries is one factor in the big data analytics dimensions. Mir et al. (2022), Goel et al. (2021), and Otterbach et al. (2021) found that big data location analytics in developing countries is another factor in big data analytics usage. Stahl et al. (2023) found that big data ethics are precursors of culture in developing countries. In turn, these factors are adopted in this study to measure big data analytics dimensions. Tsumura et al. (2022) found that the potential of call detail records can be used to improve transport planning in

developing countries. Based on the findings from Tsumura et al. (2022), smartphone record data has become a dimension of big data analytics measurement in developing countries. Family ties have been increasingly discussed in the literature (Varvarigos, 2021) because big data applications facilitate exchanges between immigrant families around the world.

Due to the specific IT context in developing countries, such as limited Internet access to population groups, insufficient addressing systems, smartphone record data, big data analytics family ties, and big data location analytics are necessary to solve most of these context-specific problems. The big data analytics dimensions in developing countries depend on big data location analytics, big data analytics ethics, smartphone record data, and big data analytics family ties. NLP is the bridge between big data analytics and AI in this research. Also, it moderates the relationship between big data analytics dimensions and big data analytics usage. The significant contribution of this study is to use a systematic literature review to identify the measurement instrument of the impact of big data analytics dimensions on its usage. Each component of the big data analytics dimensions and NPL are discussed below.

Big Data Analytics Ethics

Ethics is a philosophic discipline that explores what is good from what is wrong and on what grounds such judgments are made (Stahl et al., 2023). Technological developments have led theorists to apply ethics to the information systems discipline. Therefore, ethics of technology, computer ethics, information ethics, and big data analytics ethics have emerged. Concerns have emerged in developing countries about Governments, private organizations, non-profit entities, and individuals ignoring ethics in favor of political and commercial priorities (Rodgers et al., 2023; Akter et al., 2023). Meanwhile, big data analytics ethics are vulnerable to weak data formats, inconsistent data sources, and cultural beliefs.

In developing countries, the proliferation of islands of data or data silos contributes to false analytics, which is a precursor to ethical issues. In those countries, individuals believe in local divinities and that human talent is predestinated. In this sense, all wrongdoing related to big data analytics comes from local divinities. The difficulty associated with big data analytics ethics is cultural considerations, which are the belief that individuals are attached to local divinities. Additionally, one can imagine self-driving cars in developing countries. Those countries believe in divinities and that natural forces drive the vehicles. Also, AI is not a technology; it is manipulated by invisible forces, which can raise ethical issues. Laws could not solve this issue because it is not illegal.

In developing countries, individuals and organizations have big data tools to share data, information, and intelligence. However, using those tools could be hampered by ethical issues, mainly related to misinterpreting results and insight from big data analytics. Big data analytics ethics in developed countries are related to privacy and data misuse. In developing countries, big data analytics ethical violations are beyond the characteristics of data governance. Another ethical issue is the data on criminal activities developed in those countries. Those hackers mainly targeted the assets of people in developed countries. This is because rich people in developing countries do not use big data analytics to manage their assets like in developed countries. Technological development in developed countries is another example of the exploitation of criminal or unethical behavior by developing countries. This is why the target is developed countries.

Moreover, ethical issues are big data analytics tools used for political contestations in developing countries. Governments usually shut down the Internet connection to prevent the uprising organizers from communicating with the participants. Sometimes, smartphone companies must filter, diminish the

bandwidth, or disturb individuals' transmissions based on the injunctions they receive from Governments in developing countries.

Furthermore, the lack of continuous training to understand the complex interfaces of big data visualization tools promotes unethical behavior in developing countries. Hence, a lack of appropriate data protection and unauthorized access considerations can expose sensitive information. In this sense, Governments can spy on individuals easily. Individuals can eavesdrop on their relatives, siblings, friends, etc. Companies use big data analytics to control the behavior of their employees. Another example is money transfer fraud using big data analytics, which is vital in developing countries. Simultaneously, these activities generate the most unethical behavior in developing countries because of insufficient data cleaning and poor data governance processes, which generally lead to faulty.

Finally, efforts have been made in developed countries to combat unethical behavior derived from big data analytics usage. Ethical values guide acceptable and desirable behavior beyond mere compliance with regulations and laws (Albahri et al., 2023). Growing public concerns over AI's potentially harmful societal effects have prompted Adams et al. (2023) to identify four new ethical principles unique to K-12 education: pedagogical appropriateness, children's rights, AI literacy, and teacher well-being. They found that policy needs to address the long-term pedagogical and developmental implications for children's cognitive, physical, socio-emotional, political, and cultural adaptations. While there is a growing body of theorists working on the ethics of AI, the current ethics principles of AI are limited (Stahl et al., 2023). The ethical issues raised by AI touch policy areas such as research and innovation, data protection, taxation, and intellectual property. Policy and legislative proposals have been analyzed in parliaments and governments that may influence how ethical issues of AI can be accomplished (Stahl et al., 2023).

However, existing ethical and regulatory processes do not match the demands of AI developers and users or guarantee public confidence in AI and big data analytics usage (Albahri et al., 2023). These efforts are not available in developing countries to protect individuals' rights, data, privacy, and even security created by big data analytics ethical violations.

Big Data Analytics Family Ties

Family ties are an essential dimension in African sub-Saharan countries because many people in these countries have immigrated to Western countries like the USA, Canada, France, the UK, Germany, and others. Those immigrants have left their families at home and need to communicate with them regularly. Many refugees in Western countries come from these countries and keep family ties with their siblings back home. Using the theory of economic growth that incorporates intergenerational income transfers among family members, Varvarigos (2021) built a framework where a culture of family ties characterizes income transfers from adult children to their old parents. They found that economic development is negatively related to the strength of family ties.

Other studies contradicted the findings of Varvarigos (2021) and found that financial support and help adults provide for their old parents and family members are positive indicators of economic development. Mare et al. (2020) considered the importance of religion, trust, and political participation in family ties. They found that in countries where family ties are more robust, the level of trust is lower, and the underground economy is higher. Stronger family ties are associated with higher financing constraints for family members in developing countries (Mertzanis, 2019). Afulani et al. (2016) examined the association between cross-border ties and cross-border separation of sub-Saharan African migrant adults in France. They found that having a child abroad is associated with poor health among women but not men.

The ubiquitous dissemination of affordable big data analytics applications on mobile devices also contributes to potential economic growth and family ties in developing countries. For instance, family members in developing countries have strong connections with family abroad using WhatsApp (Tsumura et al., 2022). Big data analytics connected to video conference applications like Zoom and Microsoft Teams facilitate these family ties between developing and developed countries.

Hanson and Block (2020) examined Hostede's national culture model and found that family involvement improves performance outcomes in countries with more collectivistic cultures. Individuals from developing countries living abroad use big data analytics to take advantage of the collectivistic concept of Hofstede cultural dimensions when they operate a business with their family members in developed countries. Big data analytics provide transparency in business management and support better financial performance in developing and developed country family ties relationships.

Moreover, the diffusion of big data analytics technology in the most remote and isolated areas of developing nations provides a facility to bring rural producers into contact with consumers in the city and links with family members that can help knowledge transfer and facilitate investments at the international level (Goel et al., 2021). Furthermore, the diaspora becomes a safety net for their family back home. The diaspora uses big data analytics to send billions of dollars back home to their family to pay for food, funerals, schooling, health care, houses, etc. Sometimes, the diaspora opens businesses and makes investments in their home countries. Others build their houses and send cars, technological devices like smartphones, laptops, and gifts to their home countries. Finally, big data analytics allow immigrant families in Western countries to manage their businesses in their home countries in sub-Saharan Africa. In this sense, big data analytics creates more investment opportunities for family members desiring to help their relatives in developing countries.

Big Data Location Analytics

Big data location analytics uses data from social media, travel data, and geographical positioning information to detect and track the accurate position of objects and persons in the atmosphere. Additionally, big data location analytics use sensor technologies to identify objects in space. It is based on topological and coordinated mechanisms for mapping addresses to physical locations (Mir et al., 2022). With the advent of big data, big data location analytics uses data to determine the geographical location of mobile phone users (Tsumura et al., 2022). An example of big data analytics usage includes big data location analytics techniques for predicting and monitoring COVID-19 cases in developing countries (Mir et al., 2022). Additionally, big data location analytics has been used to identify hotspot areas and slow disease transmission (Mir et al., 2022). Moreover, Goel et al. (2021) found that big data location analytics is utilized for intelligent agriculture in developing countries. Furthermore, Otterbach et al. (2021) used Google data to identify restaurant and food store locations in South Africa. Terrestrial and aerial sensors, satellites, and surveillance equipment collect, analyze, and utilize location data from diversifying sources for intelligent agriculture in developing countries (Goel et al., 2021).

Smart Phone Record Data

Many authors used different expressions or names to identify smartphone record data. Some authors called it phone record data, mobile phone data, call detail records (Tsumura et al., 2022), and call record data (Onyeji-Nwogu et al., 2020). Tsumura et al. (2022) used mobile phone call records to understand people's daily travel patterns in developing countries. Smartphones record data containing the cell tower location for each activity, such as phone calls or text messages (Tsumura et al., 2022). Onyeji-Nwogu et al. (2020) mentioned that call record data is another source of big data relevant to managing the electricity

infrastructure in developing countries. The call record dataset included the total call duration, number of calls, total number of text messages, etc. (Onyeji-Nwogu et al., 2020).

In developing countries, individual transport needs become more diverse (Tsumura et al., 2022). Consequently, the traditional approach to transportation needs is obsolete and unadaptable. For instance, individuals used to leave their homes and stand in front of roads, raise their hands, or make a sign to the cab or taxi driver to stop, excluding all essential safety and security considerations. After the cab driver stops, the customer must indicate their destination and bargain on the price. If there is an agreement, the customer must enter the taxi car and be satisfied. This old approach has become outdated with the advent of big data analytics tools. With the new approach, individuals download an app on their smartphone, save it on their phone record, and use it to contact the taxi drivers' smartphones. Sometimes, individuals must launch the app from their homes to search for available taxis. Once the smartphone record detects a cab, the taxi driver can use the big location data to find the customer's address and offer him the appropriate service. This new approach solves the baggage handling and management-specific problems for the customers because they do not need to lift their luggage on the road before finding a taxi driver. The cost of transportation is no longer a major issue.

Natural Language Processing

NLP is an AI component that allows computers to converse with humans (Lawson-Body et al., 2022). In this sense, the machine imitates human voices and exchanges writing with humans (Lawson-Body et al., 2022). After receiving appropriate training, the machine can interpret, manipulate, and understand human language and thoughts. The applications of NLP in developed countries are viral. However, its application in developing countries needs to be demonstrated. This is because most NLP applications are tailored based on the needs and requirements of developed countries' cultures. Because of their increasingly growing population, the future prosperity of NLP remains in developing countries. Organizations increasingly employ NLP to deliver more tailored services to individuals and businesses (Berryhill et al., 2019) in these countries. An application of NLP is the intelligent search engine services that allow text standardization using various methods, including stemming, normalization, and tokenization (Hajipour et al., 2023). Intelligent question-answering is another application of NLP, including sentiment analysis, word weighing, spell checking, keyword extraction, and word lemmatization (Hajipour et al., 2023).

The Chatbot and Audio Bot services are two primary components of intelligent question-answering (Hajipour et al., 2023). The Chatbot and Audio Bot service allows individuals to send questions to governments or organizations via voice message and receive suitable answers (Hajipour et al., 2023). Keiper et al. (2023) found that ChatGPT produces grammatically quality, complete, topically accurate outputs for short answer questions. ChatGPT is an open source chatbot introduced by OpenAI in 2022 (Keiper et al., 2023). Chatbot allows humans to interact verbally or in writing with computers as if communicating with real people (Keiper et al., 2023). However, NLP must deal with the challenge of a lack of training algorithms in languages other than English (Berryhill et al., 2019). Therefore, governments in developing countries could not make predictions considering individuals who speak languages other than English.

Li (2022) introduced NLP applications as AI recognition tools and mentioned that their roots were in voice control technology, which used speech recognition to control the communication of intelligent devices. NLP recognition tools allow us to compare human speech with smartphone command sets (Li, 2022). NLP recognition technologies can be classified into inanimate detection and animate detection (Li, 2022). There are three types of inanimate detection AI technologies: barcode identification, smart card, and radio frequency identification. There are also three animate detection AI technologies: voice recognition, face

recognition, and fingerprint recognition. In most developing countries, individuals use Apple Siri and Amazon Alexa as virtual assistants. NLP elicited the relationship between online discourse and political violence in Kenya (Berryhill et al., 2019). The NLP model can help predict the increase and decrease in causality in advance (Berryhill et al., 2019). However, the language barrier or obstacle does not allow most individuals to take advantage of Siri and Alexa because they do not speak English.

Big Data Analytics Dimensions Structural Framework

The structural framework shows the relationships between the study's latent variables or constructs. In this study, the framework has three parts. The first part deals with big data analytics dimensions such as big data location analytics, big data analytics family ties, big data analytics ethics, and smartphone record data. The second part bears on AI's component of NLP, which systematizes the bridge between big data analytics and AI in this research. In other words, NLP moderates big data analytics dimensions and its usage in developing countries. The third part is the usage of big data analytics in developing countries. This framework is a starting point for generating measurement items using a systematic literature review. Figure 1 shows the structural framework.

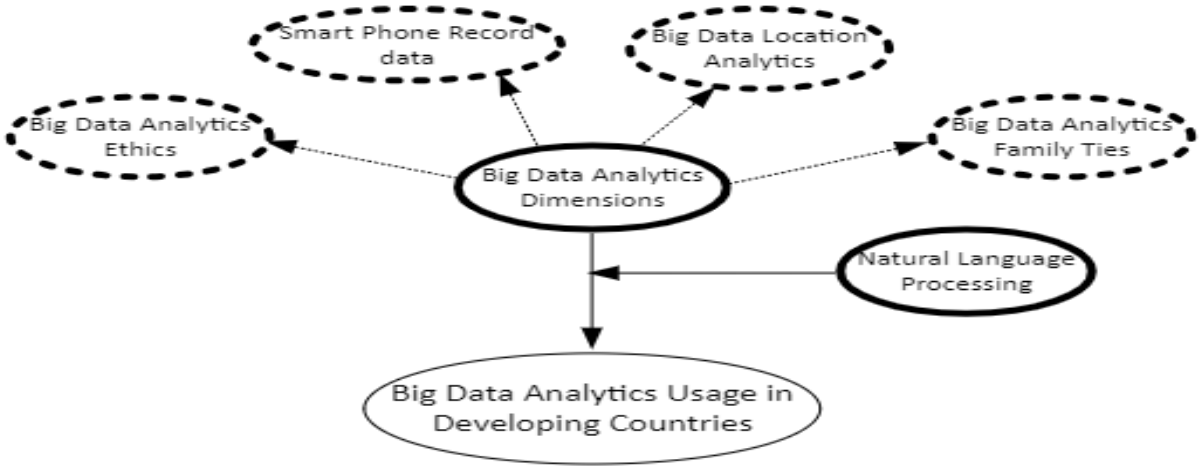

Figure 1: Structural Framework

Research Methodology

The qualitative method PRISMA was considered for developing measurement instruments (as shown in Figure 2) and was adopted herein (Mbunge and Batani, 2023). This method provides an entirely systematic literature review approach and was adopted because it better fits the objective of this research than other methodologies.

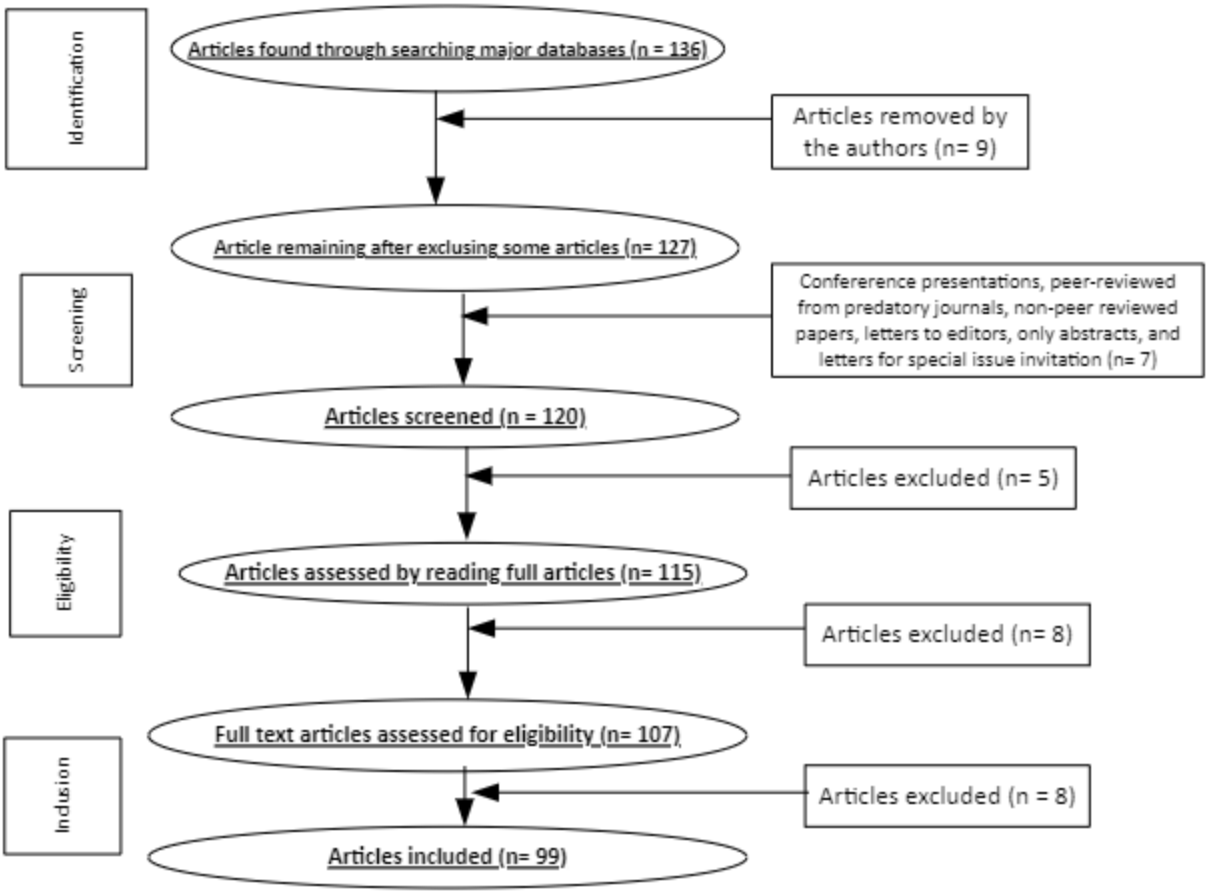

Figure 2: PRISMA Model (Adapted from Mbunge and Batani, 2023)

Systematic Literature Review

Systematic literature reviews are essential activities in scientific research, which have the goal of delivering the synthesis of existing knowledge and supporting the research agenda (Hendricks and Mwapwele, 2024; Mbaidin et al., 2023). The research methodology herein, crucially based on the interpretive and qualitative approaches, aligns with the perspectives of Hendricks and Mwapwele, (2024) and Mbaidin et al (2023). These esteemed authors have underscored that the systematic literature review, a pivotal tool to synthesize studies, finds its foundation in the interpretivism theory. Moreover, the systematic literature review evaluates the existing body of knowledge in a field, propelling research forward (Hendricks and Mwapwele, 2024). Based on the systematic literature review, the authors in this study analyzed and critically assessed the related literature and proposed papers related to big data analytics, NLP, big data location analytics, smartphone record data, big data analytics family ties, and big data analytics ethics. Consistent with the structural framework, the systematic review was used to identify items to establish measurement instruments for big data analytics dimensions and NLP in developing countries.

Search strategy

The online digital libraries were considered during this initial search process, consistent with Chonkker et al. (2017). The thematic analysis covered peer-reviewed, SCOPUS-indexed, or ABDC (Australian Business

Dean Council) journals. More importantly, the thematic keyword search was done in social media using ScienceDirect, Google Scholar, Google Search, SCOPUS, and SCImago. Additionally, a keyword-based search was used, consistent with the study of Samad et al. (2022). The search was guided by the following keywords: big geospatial data, big data location analytics, smartphone record data, ethics in data analytics, family ties, big data in developing countries, AI in developing countries, AI, and big data in Africa. Papers published between 2016 and 2024 were selected. The 2016 and 2024 selections were used because of the advent of big data and the newness of AI.

Study Selection

We searched the databases for articles that mentioned terms and concepts related to big data usage in developing countries, AI techniques of NLP, big data ethics, and big data dimensions. The following exclusion criteria were used: 1) conference presentations, 2) peer-reviewed from predatory journals, 3) non-peer review papers, 4) letters to the editors, 5) only abstracts, and 6) letters for special issue invitations. The following inclusion criteria were retained: peer-reviewed publications and conference peer-reviewed proceedings publications. The selected articles were screened based on titles, abstracts, introductions, conclusions, and full text consistent with the study of Samad et al. (2022).

Data Extraction

After the literature search, 136 articles were extracted. The authors screened the articles and papers collected using titles and abstracts. Nine articles were removed from the pool because of duplication and redundancy. The introductions and conclusions of the rest of the articles were screened. After this step, seven articles were excluded. The methodology rigor of the remaining pool of articles was examined and assessed to discard 21 more papers. Finally, 99 papers were considered in this study.

Analysis and Results

Figure 3 shows the results of the systematic literature review below. The measurement framework presents the relationships between the latent variables and their respective indicators or items. All items were graded on a scale from 1 (“strongly disagree”) to 7 (“strongly agree”).

We found:

- five items measuring the latent variable smartphone record data.
- Six items measure the latent variable of big data location analytics.
- Five items measuring the latent variable big data analytics family ties.
- Seven items measure the latent variable of big data analytics ethics.
- Six items measure the latent variable NLP.

Tables 1 through 5 also explain how each factor or latent variable is connected to its measurement items.

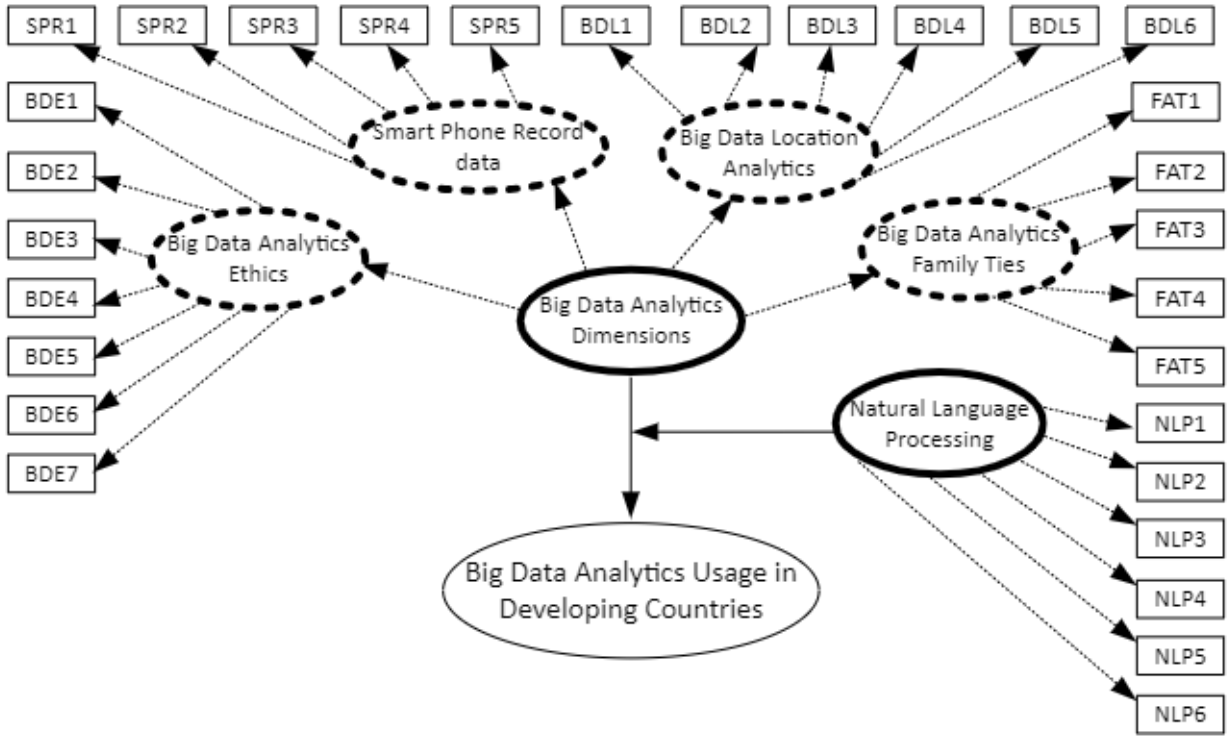

Figure 3: Measurement Frameworks with generated Items from the systematic review

Big Data Analytics Ethics Measurement Instrument

Big data analytics ethics is the extent to which individuals in developing countries 1) use big data applications to do as much as it has been programmed to perform, 2) use big data applications that do not have the property of freedom of will, 3) recognize AI and big data applications as moral judgment, 4) use AI and big data applications that can be morally responsible for good actions, 5) use AI and big data applications that can be morally responsible for wrong actions, 6) use AI and big data applications that could respect human values better than human beings, and 7) use AI and big data actions that could not be moral or immoral because they are machinery actions with any moral quality (Das et al., 2023). Table 1 shows the measurement items for big data analytics family ties.

Table 1: Big Data Analytics Ethics Measurement Items

Coded Items	Items and Variable
BDE	Big Data Analytics Ethics
BDE1	AI applications can do as much as they have been programmed to perform.
BDE2	AI applications do not have the property of freedom of will.
BDE3	People recognize AI as moral judgment.
BDE4	AI applications can be morally responsible for good actions.
BDE5	AI applications can be morally responsible for wrong actions.
BDE6	AI could respect human values better than human beings.
BDE7	AI actions could not be moral or immoral because they are machinery actions with any moral quality.

Big Data Analytics Family Ties Measurement Instrument

Big data analytics family ties are the extent to which individuals in developing countries use big data and AI applications 1) to receive transferred materials and money from their family or friends overseas, 2) like WhatsApp, FaceTime, Zoom, Chatbots, etc. to facilitate online family reunion, 3) to participate in religious, entertainment, and social gathering to facilitate family ties, 4) to create crowdsourcing to collect funds for funerals and social issues for their community, and 5) to manage businesses my relatives undertake in his home countries to create more investment opportunities. Table 2 shows the measurement items for big data analytics family ties.

Table 2: Big Data Analytics Family Ties Measurement Items

Coded Items	Items and Variable
FAT	Big Data Analytics Family Ties
FAT1	I use big data apps to receive transferred materials and money from my family or friends overseas.
FAT2	I use big data apps like WhatsApp, FaceTime, Zoom, etc., to facilitate online family reunions.
FAT3	I use big data and AI applications to participate in religious, entertainment, and social gatherings to facilitate family ties.
FAT4	I use big data and AI applications to create crowdsourcing campaigns to collect funds for funerals and social issues in their community.
FAT5	I use big data and AI applications to manage businesses my relatives undertake in their home countries to create more investment opportunities.

Smartphone Record Data Measurement Instrument

Smartphone record data is the extent to which individuals in developing countries 1) use call records to understand patterns, 2) use call records to manage the total call duration and the number of calls made and received from people in their community, 3) review phone records to build and enhance relations with people in the community, 4) remove irrelevant data from big data, and 5) transfer photos, pictures, images, and videos from big data storage into iCloud. Table 3 shows the measurement items for smartphone record data.

Table :. Smartphone Record Data Measurement Items

Coded Items	Items and Variable
SPR	Smartphone Record Data
SPR1	I use call records to understand patterns.
SPR2	I use call records to manage the total duration of calls and the number of calls I make and receive from the community.
SPR3	I review phone records to build and enhance relations in the community.
SPR4	I remove irrelevant data from big data.
SPR5	I transfer photos, pictures, images, and videos from big data storage into iCloud.

Big Data Location Analytics Measurement Instrument

Big data location analytics is the extent to which individuals in developing countries use analytics to 1) determine the location of people related to them, 2) monitor the progression of diseases in their community, 3) access the emergency information from their government officials, 4) identify the location of service

providers in their cities, 5) predict the weather conditions in their cities, and 6) manage their travel and transportation needs effectively. Table 4 shows the measurement items for big data location analytics.

Table 4: Big Data Location Analytics Measurement Items

Coded Items	Items and Variable
BDL	Big Data Location Analytics
BDL1	I use big data analytics to determine the location of people related to me.
BDL2	I use big data analytics to monitor the progression of diseases in my community.
BDL3	I use big data analytics to access emergency information from our government officials.
BDL4	I use big data analytics to identify the location of service providers in my city.
BDL5	I use big data analytics to predict the weather conditions in my city.
BDL6	I use big data analytics to manage my travel and transportation needs effectively.

Natural Language Processing (NLP) in Developing Countries Measurement Instrument

NLP is the extent to which individuals in developing countries use AI systems 1) to communicate with a computer in their native language, 2) to translate speech or text from one language to another, 3) to reduce multiple words to their base or root (stemming), 4) to categorize a block of text in a sentence (Tokenizing), 5) to distinguish between words that have more than one meaning, and 6) analyze the unstructured content of Web pages. Table 5 shows the measurement items for NLP in developing countries.

Table 5. Natural Language Processing in Developing Countries Measurement Items

Coded Items	Items and Variables
NLP	Natural Language Processing in Developing Countries
NLP1	I use AI systems to communicate with a computer in their native language.
NLP2	I use AI systems to translate speech or text from one language to another.
NLP3	I use AI systems to reduce multiple words to their base or root (stemming).
NLP4	I use AI systems to categorize a text block in a sentence (Tokenizing).
NLP5	I use AI systems to distinguish between words with multiple meanings.
NLP6	I use AI systems to analyze the unstructured content of Web pages.

Discussion

This paper investigated the relationships between these big data analytics dimensions and big data analytics usage. Moreover, this study examines the moderating role of NLP in developing countries on big data analytics. Most importantly, the research performed a systematic literature review and found 29 items measuring big data analytics usage in developing countries when NLP is moderating.

The results indicated that big data analytics family ties are appropriate in developing countries when consistent measurement instrument is used to assess them. Many big data applications and technologies facilitate lower-cost transfers of materials and money to developing countries, which justifies considering family ties as a big data analytics dimension in these countries.

This study's findings indicated that big data analytics ethics have a measurement instrument in developing countries. The application of big data analytics in developing countries raises ethical questions about data protection, decision-making, and substituting human work with autonomous machines, which could affect public sector activities (Souza et al., 2019). Most AI applications generating big data are not yet available in developing countries. For instance, self-driving cars, which use sensory technologies to produce tremendous amounts of big data, are unavailable in developing countries.

This research provides an effective measurement tool to assess the relationship between big data location analytics and usage. Location analytics is one of the most prominent big data analytics uses in developing countries. This dimension of big data analytics has successfully monitored COVID-19 cases and intelligent agriculture, tracked people, and found stores, restaurants, and hospitals.

The results of this study present evidence that smartphone record data should be measured with efficient tools, allowing individuals in developing countries to manage voluminous data and exploit them to create opportunities and identify patterns.

Similarly, NLP presents many opportunities in developing countries. Its measurement became vital because it plays a moderating role in big data analytics. For instance, people speak many languages in sub-Saharan Africa. The current NLP applications understand only English and some Western countries' languages. However, using interesting measurement instruments can allow people to use NLP to translate more languages in order to do business.

Conclusion

This research uses a systematic literature review to develop measurement instruments for big data analytics family ties, ethics, big data location analytics, smartphone record data, and NLP in developing countries. Therefore, two big data analytics frameworks were developed: structural and measurement. The structural frameworks allowed us to identify the study's boundaries, which permitted a systematic literature review using the PRISMA model. The measurement framework showed the systematic review results, indicating the measurement items and their relationships with their latent variables or constructs. After selecting and extracting data from the publication databases above, the next phase involves screening the selected papers for inclusion and exclusion. Paper titles and abstracts were used to discard and exclude unfit papers from the systematic review during the selection and review. Paper titles that did not match the search keywords were removed from the studies.

The authors could not remove some papers based on unmatched titles and abstract criteria. In that scenario, the authors pursued their investigation by analyzing those papers' introduction and conclusion sections before making a final exclusion decision. In some rare cases, authors reviewed the subtitles and headings within the papers and made final recommendations to exclude those from the systematic literature review. Additional exclusion criteria include papers that were not published in English, non-peer-reviewed papers, special issue announcements, only the paper abstract is available, inaccessible papers, and duplicated research. The findings revealed five items measuring the latent variable of smartphone record data, six items measuring the latent variable of big data location analytics, five items measuring the latent variable of big data analytics family ties, seven items measuring the latent variable of big data analytics ethics, and six items measuring the latent variable NLP in developing countries.

Theoretical contributions

There are four notable theoretical contributions to the literature herein. First, this study demonstrates a theoretical relationship between big data and AI. It provides an opportunity to elicit big data support for the NLP AI technique. Second, it applies ethics to big data analytics and NLP in developing countries. Third, this research is one of the first to use NLP as a moderating variable in the relationship between big data dimensions and analytics in developing countries. Fourth, it added a new measurement instrument to the existing list of AI and big data measures in the literature. Fifth, most existing measurement instruments for big data and AI are anecdotal. This study used qualitative empirical methods to create an instrument to measure big data dimensions and analytics. Sixth, this study used an innovative method based on a systematic literature review to propose a measurement instrument. Seventh, this research sheds light on the PRISMA procedure for developing measures.

Practical Contributions

This study proposes that a newly developed instrument for big data analytics should include ethics and NLP in developing countries. Most of those countries speak a myriad of languages. The results of this study would allow them to use AI to translate languages to languages easily. People in developing countries use Western languages, specifically English, as an official language. Most AI apps based on NLP are dedicated to English; therefore, big data analytics has become convenient for those populations. Governments worldwide need to develop regulations to manage AI and big data analytics. This study's results provide a huge opportunity for developing country governments to race with developed countries to propose laws to govern AI and big data analytics.

Future Research and Limitations

This research used a systematic literature review to propose a measurement for big data analytics and AI. Future research should expand the research model using variables like regulation, privacy, trust, and security. Quantitative survey data can be collected to conduct exploratory and confirmatory analyses. Those data can help perform validity and reliability analyses, including convergent and discriminant validity. A limitation of this study is that the measurement instruments were developed using qualitative analysis based on a systematic literature review. It would be fruitful to complete the survey with a quantitative analysis.

References

Adams, C., Pente, P., Lemermeyer, G. & Rockwell, G. (2023). Ethical principles for artificial intelligence in K-12 education, *Computers and Education: Artificial Intelligence*, 4, Article e100131. https://doi.org/10.1016/j.caeai.2023.100131

Afulani, P., Torres, J., Sudhinaraset, M. & Asunka, J. (2016). Transnational ties and sub-Saharan African migrants: the moderating role of gender and family separation, *Social Science and Medicine*, 168, 63-71. https://doi.org/10.1016/j.socscimed.2016.09.009

Akter, S., Sultana, S., Mariani, M., Wamba, S. Spanaki, K. & Dwevedi. Y. (2023). Advanced algorithm bias management capabilities in AI-driven marketing analytics research, *Industrial Marketing Management*, 114, pp. 243-261. https://doi.org/10.1016/j.indmarman.2023.08.013

Albhari, A., Duhaim, A., Fadhel, M., Alnoor, A., Baqer, N., Alzubaidi, L. & Albhari, O. (2023). A systematic review of trustworthy explainable artificial intelligence in healthcare Assessment of

quality, bias risk, and data fusion, *Information Fusion*, 96, 156-191. https://doi.org/10.1016/j.inffus.2023.03.008

Berryhill, J., Heang, K. K., Clogher, & McBride, K. R. (2019). Hello World: Artificial Intelligence and its Use in Public Sector, *OECD Working papers on public governance*, No 36. https://dx.doi.org/10.1787/726fd39d-en

Burton, J. (2023). Algorithm Extremism? The securitization of artificial intelligence (AI) and its impact on radicalism, polarization, and political violence, *Technology in Society*, 75, Article e102262. https://doi.org/10.1016/j.techsoc.2023.102262

Chonkker, M. S., Verma, D., & Kar, A. K. (2017). Review of technology adoption framework in mobile commerce. Information Technology and Quantitative Management. *Procedia Computer Science*. 122. 888-895. https://doi.org/10.1016/j.procs.2017.11.451

Das, K., Pattanaik M., Basantia, S., Mishra, R., Das, D., Sahoo, K. & Paital, B. (2023). Informatics in social view and need for ethical interventions for wellbeing via interference of artificial intelligence, *Telematic and Informatics Reports*, 11, Article e100065. https://doi.org/10.1016/j.teler.2023.100065

DeSouza, W., Demelo, E., Bermejo, P., Farias, R., & Gomes, A. How and where is artificial intelligence in the public sector going? A literature review and Research Agenda. Government Information Quarterly. 36. 2019. https://doi.org/10.1016/j.giq.2019.07.004

Dubey R, Bryde DJ, Dwivedi YK, Graham G, Foropon C. (2022). Impact of Artificial Intelligence-Driven Big Data Analytics Culture on Agility and Resilience in Humanitarian Supply Chain: A Practice-Based View. *International Journal of Production Economics*. https:// doi.org/10.1016/j.ijpe.2022.108618.

Dwivedi, R., Nerur, S., & Balijepally, V. (2023). Exploring artificial intelligence and big data scholarship in information systems: A citation, bibliographic coupling, and co-word analysis, *International Journal of Information Management Data Insights*, 3, Article e100185. https://doi.org/10.1016/j.jjimei.2023.100185

Fu, L., Li, J., & Chen, Y. (2023). An innovative decision-making method for air quality monitoring based on big data-assisted artificial intelligence technique, *Journal of Innovation and Knowledge*, 8, Article e100294. https://doi.org/10.1016/j.jik.2022.100294

Garg, S., & Krishnamurphy, R. (2023). A CNN Encoder decoder LSTM model for sustainable wind power predictive analytics. *Sustainable Computing: Informatics and Systems*, 38. https://doi.org/10.1016/j.suscom.2023.100869

Giudici, P., Centurelli, M., & Turchetta, S. (2024). Artificial Intelligence Risk Measurement, *Expert Systems with Applications*, vol. 235, Jan. 2024. https://doi.org/10.1016/j.eswa.2023.121220

Goel RK, Yadav CS, Vishnoi S, Rastogi R. (2021). Smart Agriculture - Urgent need of the Day in Developing Countries. Sustainable Computing: Informatics and Systems, 30, Article e100512. https://doi.org/10.1016/j.suscom.2021.100512

Guliyev, H., Huseynov, N., & Nuriyev, N. (2023). The relationship between artificial intelligence, big data, and unemployment in G7 countries: New insights from dynamic panel data model. *World Development Sustainability,* 3, Article e100107. https://doi.org/10.1016/j.wds.2023.100107

Hajipour, V., Hekmat, S. & Amini, S. A (2023). value-oriented artificial intelligence-as-a-service business plan using integrated tools and services, *Decision Analytics Journal*. 8. https://doi.org/10.1016/j.dajour.2023.100302

Hendricks, S. & Mwapwele, S. D. (2023). A systematic Literature Review on the factors influencing e-commerce adoption in developing countries. *Data and Information Management*. https://doi.org/10.1016/j.dim.2023.100045

Iqbal, A., Hassan, S., Mahmood, H. & Tanveer, M. (2022). Gender equality, education, economic growth, religious tensions nexus in developing countries: A spatial analysis approach, *Heliyon*, 8, Article e11394. https://doi.org/10.1016/j.heliyon.2022.e11394

Keiper, M., Fried, G., Lupinek, J. & Nordstrom, H. (2023). Artificial intelligence in sport management education: playing the AI game with ChatGPT. *Journal of Hospitality, Leisure, Sport, and Tourism Education*, 33, https://doi.org/10.1016/j.jhlste.2023.100456

Lawson-Body, A., Illia, A., Lawson-Body, L., Rouibah, K., Akalin, G. & Tamandja, E. (2024). Big data analytics and culture: Newly validated measurement instruments for developing countries' value proposition, *Journal of Organizational and End User Computing*, 36(1), https://doi.org/10.4018/JOEUC.344453

Lawson-Body, A., Lawson-Body, L., Illia, A. & Willoughby, L. (2022). Impact of natural language processing on CRM and Trust: An integrated framework, *Issues in Information Systems*, 23(1), 306-315.

Li, J., Herdem, M., Nathwani, J., Wen, J. (2023). Methods and applications for artificial intelligence, Internet of things, blockchain in smart energy management, *Energy and AI*, 11, Article e100208. https://doi.org/10.1016/j.egyai.2022.100208

Li, L., Geissinger, J., Ingram, W., & Fox, E. (2020). Teaching natural language processing through big data text summarization with problem-based learning, *Data and Information Management*, 4(1), 18-43. https://doi.org/10.2478/dim-2020-0003

Li, Y. (2022). Research on neural network algorithm in artificial intelligence recognition, *Sustainable Energy Technologies and Assessments*, 53, Article e102691.

Lepri, B., Olivier, N., & Pentland, A. (2021). Ethical Machines: The human-centric use of Artificial Intelligence, *iScience Review*, vol. 24, March 19, 2021, https://doi.org/10.1016/j.isci. 2021.102249.

Mbaidin, H., Alsmairat, M. & Al-Adaileh, R. (2023). Blockchain adoption for sustainable development in developing countries: Challenges and opportunities in the banking sector, *International Journal of Information Management Data Insights*, 3, Article e100199. https://doi.org/10.1016/j.jjimei.2023.100199

Mbunge, E. & Batani, J. (2023). Application of deep learning and machine learning models to improve healthcare in sub-Saharan Africa: Emerging opportunities trends, and implications, *Telematic and Informatics Reports*, 11, Article e100097. https://doi.org/10.1016/j.teler.2023.100097

Mertzanis, C., (2019). Family ties, institutions and financing constraints in developing countries, *Journal of Banking and Finance*, 108, Article e105650. https://doi.org/10.1016/j.jbankfin.2019.105650

Mir SA, Bhat MS, Rather GM, Mattoo D. (2022). Role of Big Geospatial Data COVID-19 Crisis. *Data Science for COVID-19*, 2:589-609. https://doi.org/10.1016/B978-0-323-90769-9.00031-1

Mullins, M., Holland, C., & Cunneen, M. (2021). Creating ethics guidelines for artificial intelligence and big data analytics customers: The case of consumer European insurance market, *Patterns*, 2, Article e100362. https://doi.org/10.1016/j.patter.2021.100362

Okonkwo, C. & Ade-Ibijola, A. (2021). Chatbot applications in education: A systematic review, *Computers and Education: Artificial Intelligence*, 2, Article e100033. https://doi.org/10.1016/j.caeai.2021.100033

Onyeji-Nwogu I, Brazilian M, Moss T. (2020). Big Data and Electricity Sector in African Countries. *World Development Perspectives*, 17: Article e100170. doi:https://doi.org/10.1016/j.wdp.2020.100170

Otterbach S, Oskorouchi HR, Rogan M, Qaim M. (2021). Using Google Data to Measure the Role Big Food and Fast Food in South Africa's Obesity Epidemic. *World Development*, 140: Article e105368. https://doi.org/10.1016/j.worlddev.2020.105368

Samad, S., Ahmed, F., Naher, S., Kabir, M., Das, A., Amin, S., & Islam, S. (2022). Smartphone apps for tracking food consumption and recommendation: evaluating artificial intelligence-based functionalities, features and quality of current apps, *Intelligence Systems with Applications*, 15, Article e200103. https://doi.org/10.1016/j.iswa.2022.200103

Stahl, B., Brooks, L., Hatzakis, T., Santiago, N. & Wright, D. (2023). Exploring ethics and human rights in artificial intelligence – A Delphi study, *Technology Forecasting & Social Change*, 191, Article e122502. https://doi.org/10.1016/j.techfore.2023.122502

Rodgers, W., Murray, J., Stefanidis, A., Degbey, W. & Tarba, S. (2023). An artificial intelligence algorithm approach to ethical decision-making in human resource management processes. *Human Resource Management Review*, 33. Article e100925. https://doi.org/10.1016/j.hrmr.2022.100925

Tsumura Y, Asada Y, Kanasugi H, Arai A, Shibasaki R, Kawaguchi H, Yamada K. (2022). Examining Potentials and Practical Constraints of Mobile Phone Data for Improving Transport Planning in Developing Countries, 8: Article e100043. https://doi.org/10.1016/j.eastsj.2021.100043

Varvarigos, D. (2023). Cultural persistence in corruption, economic growth, and the environment, *Journal of Economic Dynamic and Control*, 147, Article e104590. https://doi.org/10.1016/j.jedc.2022.104590

Wang, C. & Lin, J. (2023). Utilizing artificial intelligence to support analyzing self-regulated learning: a preliminary mixed-methods evaluating from a human-centered perspective, *Computer in Human Behavior*, 144. https://doi.org/10.1016/j.chb.2023.107721

Zhou, W. (2024). Dilemma and coping strategies of news communication based on artificial intelligence and big data, *Heliyon*, 10, Article e25398. https://doi.org/10.1016/j.heliyon.2024.e25398