



OPEN ACCESS

EDITED BY

Johann Laconte,
INRAE, France

REVIEWED BY

Nayeli Montalvo-Romero,
Misantla Higher Technological Institute
(ITSM), Mexico

*CORRESPONDENCE

Runzeng Gao
✉ 15203819753@163.com

[†]These authors have contributed equally to
this work and share first authorship

RECEIVED 04 June 2025

ACCEPTED 12 September 2025

PUBLISHED 23 September 2025

CITATION

Ren Y, Qu Y and Gao R (2025) Data quality
challenges of AIGC application in smart
agriculture.

Front. Artif. Intell. 8:1640805.
doi: 10.3389/frai.2025.1640805

COPYRIGHT

© 2025 Ren, Qu and Gao. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Data quality challenges of AIGC application in smart agriculture

Yingxue Ren^{1,2†}, Yitong Qu^{2†} and Runzeng Gao^{2*}

¹Business School, Nankai University, Tianjin, China, ²School of Economics and Management,
Tiangong University, Tianjin, China

In recent years, China's agricultural development has gradually shifted from digital agriculture to smart agriculture. At the same time, with the participation of AIGC, the decision-making system of smart agriculture is also facing numerous data challenges. In this study, we employed a comprehensive quality improvement approach to address these challenges. The methodology involves three phases: (1) Detection and removal of data noise through advanced cleaning techniques and preprocessing methods; (2) Unified data standards and formats to ensure seamless integration across diverse data sources; and (3) Strengthening agricultural infrastructure to prevent data islands and promote equitable data distribution. Our analysis reveals that data noise significantly impacts precision agriculture, leading to biased decisions and resource wastage. Data fog, resulting from heterogeneous data sources and weak inter-source correlations, complicates decision-making processes. Additionally, data islands hinder data sharing and integration, exacerbated by uneven data development across regions. Systematic implementation of standardized quality control protocols is essential for enhancing smart agricultural systems and ensuring sustainable development. This study offers a novel perspective on enhancing data quality in AIGC-driven smart agriculture by integrating the Juran quality improvement model.

KEYWORDS

smart agriculture, data quality, data noise, data fog, data islands, AIGC

1 Introduction

In October 2024, the Ministry of Agriculture and Rural Affairs issued the “National Smart Agriculture Action Plan (2024–2028)” (MARD, 2024), shifting China's agricultural development from digital agriculture to smart agriculture. By generating text, images, audio, and video content, Artificial Intelligence Generated Content (AIGC), combined with advanced technologies like natural language processing, computer vision, and machine learning (Waleed Khalid et al., 2024), can offer precision agriculture decision (Bongiovanni and Lowenberg-Deboer, 2004) and verify compliance with good agricultural practices (GAP) criteria (De Baerdemaeker, 2013), thereby enhancing agricultural production efficiency and sustainability (Ilcic et al., 2025).

However, AIGC also introduces complexities in data quality, posing challenges for smart agriculture. To start with, problems such as algorithmic bias, unstable data sources, and lack of model transparency can lead to data bias (Dehghani et al., 2024), creating data noise (Martin et al., 2024). For instance, biased training data may result in misleading predictions about crop growth or pest and disease outbreaks (Jabed and Azmi Murad, 2024). Furthermore, AIGC exacerbates data fog, as integrating and interpreting data from diverse sources and formats becomes complex. This complexity hinders agricultural producers' ability to effectively use data for decision-making (Ribeiro Junior et al., 2022). Additionally, AIGC can intensify data islands, creating barriers to data sharing and integration between systems and departments, and impeding data flow and analysis (Jakku et al., 2019). For example, a smart irrigation

system may lack access to soil moisture data from the environmental system, reducing irrigation efficiency (Morchid et al., 2024). While the detrimental effects of data noise, fog, and islands on smart agriculture are increasingly recognized, there remains a critical gap in systematically addressing these intertwined data quality challenges through a unified, process-oriented quality improvement framework within the specific context of AIGC adoption.

Addressing these challenges is crucial for AIGC application in smart agriculture. Quality loops, a conceptual model emphasizing continuous improvement from a quality perspective (Tesfay, 2021), was proposed as a lens to gain insights on data quality challenges in the AIGC application of smart agriculture (See Figure 1). This study explicitly focuses on analyzing and proposing solutions for three core data quality challenges hindering AIGC-driven smart agriculture: data noise (affecting accuracy and reliability), data fog (hindering integration and interpretation), and data islands (impeding sharing and flow). With the lens, this viewpoint analyzes the root causes of these challenges, explains the potential issues from the perspective of smart agriculture, and demonstrates their impact in the long term. By deeply exploring data quality challenges in smart agriculture, this viewpoint demonstrates typical data quality challenges in AIGC applications in the view of smart agriculture. Insights could be valuable for researchers, and practitioners, and inform future technology applications.

The remainder of this paper is organized as follows. Section 2 analyzes data noise within the quality design phase. Section 3 examines data fog in the quality control phase. Section 4 discusses data islands in the quality improvement phase. Section 5 synthesizes the findings and provides targeted suggestions for mitigating these challenges. Finally, Section 6 concludes the paper.

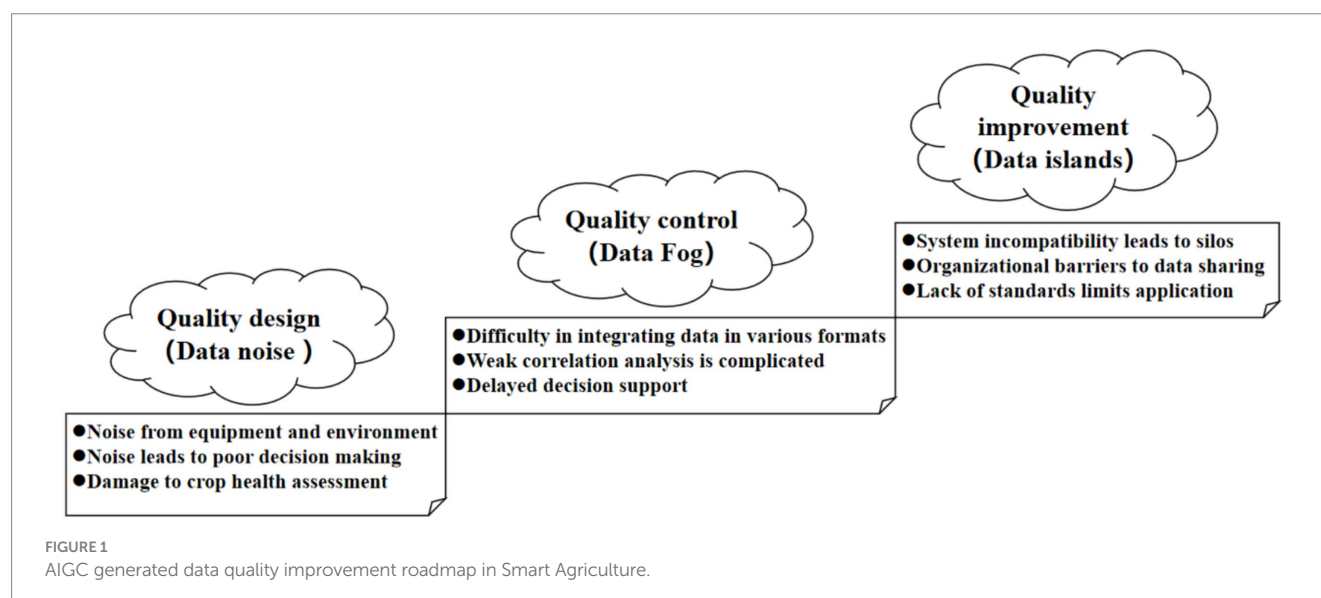
2 Data noise in the quality design phase

Noise is an unavoidable problem, which affects the data collection and data preparation processes in Data Mining applications, where

errors commonly occur (García et al., 2015). Data noise, which encompasses errors and interference within datasets, poses a significant challenge in the domain of smart agriculture, particularly with the integration of AIGC (Martín et al., 2024). Sensor faults, including those due to equipment limitations and wear from extended use, can introduce errors during data acquisition (Li et al., 2020). Additionally, environmental fluctuations such as temperature, humidity, and wind, can impact sensor readings and amplify data noise (Cai et al., 2018). Data transmission from acquisition to storage points may also be compromised by network constraints and signal degradation, leading to data corruption or loss (Brinkmann et al., 2009). Human errors during data entry and processing, especially in manual operations (Paul and Lars, 2003), are also significant sources of data noise and are inherently challenging to eliminate.

The presence of data noise is a common problem that produces several negative consequences in smart agriculture. It can result in biased agricultural decisions, particularly within precision agriculture technologies (Tey and Brindal, 2012), such as irrigation, fertilization, and pest management. Biased agricultural decisions may lead to resource wastage, including excessive water use and pesticide application, which hinder agriculture sustainability (Bongiovanni and Lowenberg-Deboer, 2004). Moreover, data noise can impair the precise assessment of compliance with GAP, which impacts the quality of agricultural production. It can also lead to increased maintenance and calibration expenses, as well as financial losses due to flawed decision-making. Consequently, data noise is a critical issue in smart agriculture, influenced by multiple factors and exerting a broad impact on agricultural operations.

A high-quality dataset is one that accurately represents real-world phenomena, is comprehensive, and is free from biases (Gong et al., 2023). In the AIGC context, addressing data noise has become an essential component of smart agriculture. The performance of smart agriculture will heavily depend on the quality of the dataset, but also on the robustness against the noise. To enhance the efficiency and effectiveness of smart agriculture, a thorough diagnostic of the sources of data noise and the implementation of cleaning methods are imperative (Xiong et al., 2006). This necessitates comprehensive



consideration of data quality control and optimization during the design, deployment, and upkeep of AIGC technology, ensuring that smart agriculture systems can be built on accurate and dependable data, thereby facilitating precision agriculture.

3 Data fog in the quality control phase

Data fog is caused by the complexity of heterogeneous data from multiple sources and the ambiguity of the relationship between data (Kumari et al., 2019). In the context of AIGC technology, this issue has become even more prominent. While AIGC technology offers volumes of data with a wide variety that can be captured, analyzed, and used for decision-making, it also adds the heterogeneity and complexity of data. In smart agriculture, data from different sources, such as sensors and robots (Wolfert et al., 2017), can become misguided in data fog if not effectively integrated, affecting the accuracy of information and the timeliness of decisions.

Data fog arises from several key issues. Data in smart agriculture comes from a variety of devices of various stakeholders. These data sets often have different formats and standards, making their integration and analysis complicated (Cheng et al., 2024). Second, the correlation between different data sources is quite weak, and the lack of standardized protocols to correlate these datasets exacerbates the difficulty of integration (Bimonte et al., 2024). Finally, existing data processing technologies may not be sufficient to handle large, multi-source, heterogeneous data, thus limiting the utilization of data (Hazra et al., 2023). The quality and usefulness of data integration depend on the existence and adoption of standards, shared formats, and mechanisms (Lapatas et al., 2015). These problems not only increase the complexity of data processing but also hinder the application of data in systematic decision-making, affecting the precision and accuracy of agricultural production.

The impact of data fog on smart farming is multifaceted. Firstly, the diverse formats and standards of agricultural data make integration and analysis difficult (Leonelli et al., 2017), preventing stakeholders from extracting valuable and intimate information. Secondly, data fog increases the difficulty of data processing, reduces efficiency, delays the time for decision-makers to obtain accurate data support, and affects the level of intelligence and precision of agricultural production. In addition, data fog can lead to unsustainable production, as growers may fail to adjust their agricultural practices based on real-time data (Wolfert et al., 2017), thus failing to achieve the goals of precision agriculture. Therefore, solving the problem of data fog and enhancing data integration and analysis capabilities are crucial to smart agriculture.

4 Data islands in the quality improvement phase

Data islands, a critical issue in smart agriculture, denote the inability to connect and share data across disparate systems or departments due to system incompatibilities, organizational barriers, and the absence of uniform standards (Philipp and David, 2020; Radauer et al., 2023). These impediments to information flow not only obstruct data integration and analysis but also precipitate decision-making errors and resource wastage. For instance, the isolation of

agricultural enterprises' sales and production data, resulting in a lack of understanding of market demand by the production department, leads to mismatched planting varieties and quantities, substantial economic loss, and even food waste.

The main cause of data islands is the uneven development of agricultural data. When data scarcity occurs, it weakens the connections and further exacerbates data fragmentation (Jones et al., 2017), ultimately leading to the formation of data islands. This disparity in data development can exacerbate economic inequalities among different regions (Wang, 2015), as farmers in less developed areas may lack access to the advanced technologies and insights available to those in more data-rich regions.

Without comprehensive data, it becomes challenging to make informed decisions about systematic problems, such as pesticide application (Pan et al., 2021). This can lead to inefficient use of resources, increased costs, and potentially lower crop yields and quality. This isolation of data island not only restricts the effectiveness of individual farming operations but also hinders the overall performance of smart agriculture on a broader scale, causing the shortage of a barrel.

5 Discussion

The analysis presented in the preceding sections underscores the profound impact of data noise, fog, and islands on the efficacy of AIGC-driven smart agriculture. These challenges are not isolated: data noise can obscure signals within individual datasets, complicating integration (fog) and rendering shared data less reliable (exacerbating island effects) (Anand et al., 2024). Data fog hinders the correlation of information necessary to overcome silos (islands). Conversely, data islands prevent access to diverse data sources needed to contextualize and clean noisy data or resolve fog ambiguities (Mishra et al., 2023). While existing research often tackles these issues individually, the quality loop perspective adopted here reveals their interconnected nature and the necessity for a holistic, phase-specific approach spanning the entire data lifecycle—from design and acquisition (noise), through integration and processing (fog), to sharing and utilization (islands). Successfully mitigating these intertwined challenges is paramount for realizing the full potential of AIGC in enabling truly precise, efficient, and sustainable smart agricultural systems (Martin et al., 2024).

6 Suggestion

The application of AIGC technology brings unprecedented changes to agricultural production, enabling more intelligent and data-driven decision-making processes. However, challenges such as data noise, data fog, and data islands have gained increasing attention from researchers, as they significantly affect the effectiveness of AIGC implementations. For instance, studies by Gupta and Gupta (2019) have highlighted the detrimental effects of data noise on prediction accuracy, while Sadri et al. (2021) discussed the complexities introduced by data fog in multi-source data environments. Additionally, Sullivan et al. (2024) emphasized the barriers posed by data islands to data sharing and collaborative agricultural management. This viewpoint offers suggestions from a unique quality

improvement perspective to analyze and mitigate these challenges, providing a structured approach to enhance data reliability and usability in smart agriculture.

To start with, in the “quality design” phase of the quality loop, it is essential to detect and remove errors and inconsistencies due to an imperfect data collection process by introducing data cleaning techniques (Xiong et al., 2006) and data preprocessing approaches (García-Gil et al., 2019). At the same time, the expert knowledge base could be combined to label and classify the data to improve the quality and availability of the data (Alonso et al., 2012). Secondly, In the “quality control” phase, unified data standards and format specifications are established to ensure that data from different sources can be effectively integrated. Standards and formats that fit various devices and could be generalized and applied are currently urgent. Finally, In the “quality improvement” phase, data islands shall be prevented by strengthening agricultural digital infrastructure in a balanced manner—such as through public-funded expansion of rural broadband and IoT networks—and by promoting even distribution of data resources via regional agricultural data platforms that integrate and openly share key information like soil moisture, weather, and market data. Meanwhile, it is crucial to strengthen data collaboration among all stakeholders, including clarifying data ownership and rights, while ensuring data security and compliance during the sharing process.

In general, through the continuous improvement of the quality loop, the challenges such as data noise, data fog, and data islands faced by the application of AIGC technology in smart agriculture shall not be ignored, the efficiency and accuracy of data processing shall be emphasized and improved.

Author contributions

YR: Data curation, Methodology, Writing – review & editing, Resources, Supervision, Writing – original draft, Funding acquisition, Conceptualization. YQ: Writing – original draft, Visualization, Investigation, Writing – review & editing, Validation, Formal analysis, Methodology, Project administration, Conceptualization, Data curation. RG: Formal analysis, Writing – original draft, Visualization, Data curation, Supervision, Validation, Writing – review & editing.

References

- Alonso, F., Martínez, L., Pérez, A., and Valente, J. P. (2012). Cooperation between expert knowledge and data mining discovered knowledge: lessons learned. *Expert Syst. Appl.* 39, 7524–7535. doi: 10.1016/j.eswa.2012.01.133
- Anand, G., Vyas, M., Yadav, R. N., and Nayak, S. K. (2024). On reducing data transmissions in fog-enabled LoRa-based smart agriculture. *IEEE Internet Things J.* 11, 8894–8905. doi: 10.1109/JIOT.2023.3321466
- Bimonte, S., Bellocchi, G., Pinet, F., Charrier, G., Sacharidis, D., Sakr, M., et al. (2024). Technological and research challenges in data engineering for sustainable agriculture. *Int. Workshop Big Data Emerg. Distribut. Environ.* 1–6. doi: 10.1145/3663741.3664786
- Bongiovanni, R., and Lowenberg-Deboer, J. (2004). Precision agriculture and sustainability. *Precis. Agric.* 5, 359–387. doi: 10.1023/B:PRAG.0000040806.39604.aa
- Brinkmann, B. H., Bower, M. R., Stengel, K. A., Worrell, G. A., and Stead, M. (2009). Large-scale electrophysiology: acquisition, compression, encryption, and storage of big data. *J. Neurosci. Methods* 180, 185–192. doi: 10.1016/j.jneumeth.2009.03.022
- Cai, Y., Zhao, Y., Ma, X., Zhou, K., and Chen, Y. (2018). Influence of environmental factors on atmospheric corrosion in dynamic environment. *Corros. Sci.* 137, 163–175. doi: 10.1016/j.corsci.2018.03.042
- Cheng, C., Messerschmidt, L., Bravo, I., Waldbauer, M., Bhavikatti, R., Schenk, C., et al. (2024). A general primer for data harmonization. *Scientific Data* 11:152. doi: 10.1038/s41597-024-02956-3
- De Baerdemaeker, J. (2013). “Precision agriculture technology and robotics for good agricultural practices” in Ifac proceedings, vol. 46, 1–4.
- Dehghani, F., Malik, N., Lin, J., Bayat, S., and Bento, M. Fairness in healthcare: assessing data Bias and algorithmic fairness. 2024 20th international symposium on medical information processing and analysis (Sipaim), (2024)
- García, S., Luengo, J., and Herrera, F. (2015). “Dealing with Noisy data” in Data preprocessing in data mining, eds. S. García, J. Luengo and F. Herrera (Cham: Springer International Publishing).
- García-Gil, D., Luengo, J., García, S., and Herrera, F. (2019). Enabling smart data: noise filtering in big data classification. *Inf. Sci.* 479, 135–152. doi: 10.1016/j.ins.2018.12.002
- Gong, Y., Liu, G., Xue, Y., Li, R., and Meng, L. (2023). A survey on dataset quality in machine learning. *Inf. Softw. Technol.* 162:107268. doi: 10.1016/j.infsof.2023.107268

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research was funded by the National Natural Science Foundation of China (NSFC), grant numbers 72261147706 and 72171166.

Acknowledgments

The authors would like to thank all respondents for their participation. We are grateful for the support from the National Natural Science Foundation of China (NSFC grant numbers 72261147706 and 72171166).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Gupta, S., and Gupta, A. (2019). Dealing with noise problem in machine learning data-sets: a systematic review. *Procedia Comput. Sci.* 161, 466–474. doi: 10.1016/j.procs.2019.11.146
- Hazra, A., Rana, P., Adhikari, M., and Amgoth, T. (2023). Fog computing for next-generation internet of things: fundamental, state-of-the-art and research challenges. *Comput Sci Rev* 48:100549. doi: 10.1016/j.cosrev.2023.100549
- Illic, A., Fuentes, M., and Lawler, D. (2025). Artificial intelligence, complexity, and systemic resilience in global governance. *Front. Artif. Intell.* 8:1562095. doi: 10.3389/frai.2025.1562095
- Jabed, M. A., and Azmi Murad, M. A. (2024). Crop yield prediction in agriculture: a comprehensive review of machine learning and deep learning approaches, with insights for future research and sustainability. *Heliyon* 10:e40836. doi: 10.1016/j.heliyon.2024.e40836
- Jakku, E., Taylor, B., Fleming, A., Mason, C., Fielke, S., Sounness, C., et al. (2019). “If they don’t tell us what they do with it, why would we trust them?” trust, transparency and benefit-sharing in smart farming. *Njas-wageningen J. Life Sci.* 90–91, 1–13. doi: 10.1016/j.njas.2018.11.002
- Jones, J. W., Antle, J. M., Basso, B., Boote, K. J., Conant, R. T., Foster, I., et al. (2017). Toward a new generation of agricultural system data, models, and knowledge products: state of agricultural systems science. *Agric. Syst.* 155, 269–288. doi: 10.1016/j.agry.2016.09.021
- Kumari, A., Tanwar, S., Tyagi, S., Kumar, N., Parizi, R. M., and Choo, K.-K. R. (2019). Fog data analytics: a taxonomy and process model. *J. Netw. Comput. Appl.* 128, 90–104. doi: 10.1016/j.jnca.2018.12.013
- Lapatas, V., Stefanidakis, M., Jimenez, R. C., Via, A., and Schneider, M. V. (2015). Data integration in biological research: an overview. *J. Biol. Res.-Thessalon.* 22:9. doi: 10.1186/s40709-015-0032-5
- Leonelli, S., Davey, R. P., Arnaud, E., Parry, G., and Bastow, R. (2017). Data management and best practice for plant science. *Nat Plants* 3:17086. doi: 10.1038/nplants.2017.86
- Li, D., Wang, Y., Wang, J., Wang, C., and Duan, Y. (2020). Recent advances in sensor fault diagnosis: a review. *Sensors Actuators A Phys.* 309:111990. doi: 10.1016/j.sna.2020.111990
- MARD. (2024). *Circular of the Ministry of Agriculture and Rural Development on the issuance of the National Action Plan on smart agriculture (2024–2028) (in Chinese)*. Available online at: https://www.gov.cn/zhengce/zhengceku/202410/content_6983057.htm (Accessed September, 2025).
- Martin, J., Sáez, J. A., and Corchado, E. (2024). Tackling the problem of noisy IoT sensor data in smart agriculture: regression noise filters for enhanced evapotranspiration prediction. *Expert Syst. Appl.* 237:121608. doi: 10.1016/j.eswa.2023.121608
- Mishra, R., Ramesh, D., Bellavista, P., and Edla, D. R. (2023). Redactable blockchain-assisted secure data aggregation scheme for fog-enabled internet-of-farming-things. *IEEE Trans. Netw. Serv. Manag.* 20, 4652–4667. doi: 10.1109/TNSM.2023.3322442
- Morchid, A., Jebabra, R., Khalid, H. M., El Alami, R., Qjidaa, H., and Ouazzani Jamil, M. (2024). IoT-based smart irrigation management system to enhance agricultural water security using embedded systems, telemetry data, and cloud computing. *Results Eng.* 23:102829. doi: 10.1016/j.rineng.2024.102829
- Pan, Y., Ren, Y., and Luning, P. A. (2021). Factors influencing Chinese farmers’ proper pesticide application in agricultural products – a review. *Food Control* 122:9. doi: 10.1016/j.foodcont.2020.107788
- Paul, B. P., and Lars, L. E. (2003). Data processing: Errors and their control appears in Chapter 7 of the book *Introduction to Survey Quality*. New Jersey, U.S.A: John Wiley & Sons Inc.
- Philipp, E. B., and David, E. (2020). Machine learning in agriculture: from silos to marketplaces. *Plant Biotechnol. J.* 19, 648–650. doi: 10.1111/pbi.13521
- Radauer, A., Searle, N., and Bader, M. A. (2023). The possibilities and limits of trade secrets to protect data shared between firms in agricultural and food sectors. *World Pat. Inf.* 73:102183. doi: 10.1016/j.wpi.2023.102183
- Ribeiro Junior, F. M., Bianchi, R. A. C., Prati, R. C., Kolehmainen, K., Soininen, J.-P., and Kamiński, C. A. (2022). Data reduction based on machine learning algorithms for fog computing in IoT smart agriculture. *Biosyst. Eng.* 223, 142–158. doi: 10.1016/j.biosystemseng.2021.12.021
- Sadri, A. A., Rahmani, A. M., Saberikamarposhti, M., and Hosseinzadeh, M. (2021). Fog data management: a vision, challenges, and future directions. *J. Netw. Comput. Appl.* 174:102882. doi: 10.1016/j.jnca.2020.102882
- Sullivan, C. S., Gemtou, M., Anastasiou, E., and Fountas, S. (2024). Building trust: a systematic review of the drivers and barriers of agricultural data sharing. *Smart Agricultural Technology* 8:100477. doi: 10.1016/j.atech.2024.100477
- Tesfay, Y. Y. (2021). “Models of continuous improvement” in *Developing structured procedural and methodological engineering designs: Applied industrial engineering tools*. ed. Y. Y. Tesfay (Cham: Springer International Publishing).
- Tey, Y. S., and Brindal, M. (2012). Factors influencing the adoption of precision agricultural technologies: a review for policy implications. *Precis. Agric.* 13, 713–730. doi: 10.1007/s11119-012-9273-6
- Waleed Khalid, A., Mohammad, O., Baydaa Sh, Z. A., and Laith, J. (2024). Smart agriculture solutions: harnessing AI and IoT for crop management. *E3S web of conferences*, 477.
- Wang, Z. (2015). “The imbalance in regional economic development in China and its reasons” in *Private sector development and urbanization in China: Strategies for widespread growth*. ed. Z. Wang (New York: Palgrave Macmillan Us).
- Wolfert, S., Ge, L., Verdouw, C., and Bogaardt, M.-J. (2017). Big data in smart farming – a review. *Agric. Syst.* 153, 69–80. doi: 10.1016/j.agry.2017.01.023
- Xiong, H., Gaurav, P., Steinbach, M., and Vipin, K. (2006). Enhancing data analysis with noise removal. *IEEE Trans. Knowl. Data Eng.* 18, 304–319. doi: 10.1109/TKDE.2006.46