PAPA for IoT - Role of Data Accuracy in IoT Deployment and Data-centric Decision-Making

Mikko Vermanen*, Juhani Naskali*, Ville Harkke†, Jani Koskinen*
* University of Turku/Information Systems Science, Turku, Finland
† Åbo Akademi/Information Systems, Turku, Finlad

***Abstract*—Data has become a central asset supporting decision-making in modern business environments. Internet of Things (IoT) solutions in particular are designed to gather information from a large variety of monitorable entities, providing companies with seemingly unlimited abilities to measure and analyze factors such as performance and productivity.**

Considering the reliability and trustworthiness in data-driven decision-making, the accuracy of the leveraged information is critical. While the development of IoT solutions makes them continuously more accurate, caution must be applied not only in terms of potential inaccuracy caused by technical failures but also by human error. Thus, some of the key questions to consider when investigating the accuracy and trustworthiness of data are: is the gathered information itself correct, is the leveraged data contextually correct, and what are the potential issues caused by human error.

From an ethical perspective, the justifiability of leveraging potentially incorrect data needs to be addressed. When decisions related to individual employees are being made based on data, it is necessary to not only consider the possible inaccuracy of the data but also to prevent manipulating or distorting the results by applying too limited or falsely selected data, creating a biased basis for analysis and decisions.

***Keywords*—Internet of Things, PAPA-model, Accuracy, Ethics**

I. INTRODUCTION

The data economy, Digital society, Smart cities etc. are terms that are used to describe the change that society is experiencing by collecting and using information in large scale. Businesses and the whole society are more and more dependent on data and this phenomenon shows no sign of slowing down in the future. IoT has provided ways to collect information from different sources, offering possibilities to connect almost everything together, and thus turning things to data sources. This dependence of data, however, sets demands for the data and especially for its quality. As data is a central asset nowadays, the demand for data accuracy is crucial, if expectations for its use are to be met. Without data accuracy, we lose the fundamental benefit that technology offers us—automation of tasks—and are forced back to manual checks and routine data work conducted by humans; the very thing we were aiming to surpass with digital technologies.

While the correctness of the actual data is a central factor determining its quality [1], the impact of human factors appears to be a significant but easily overlooked issue in organisations [2]. Even if the concrete data gathered from monitorable objects are correct, the responsiblity for avoiding risks related to selecting the right type of data, conducting correct analysis, making justified conclusions etc. belongs to the human decision-makers and should be systematically governed [3]. This calls for not only understanding on how to correctly apply data in decision-making, but also how to retain ethicality in the related practices in business [4].

In addition to potentially causing harm to the company itself by drawing false conclusions based on inaccurate data, the individuals from whom the data is collected are at risk, and there is a demand for protecting users by ethical design that aims to empower users who are interacting with IoT [5]. For example, when IoT solutions are utilised for performance monitoring purposes and decisions are being made based on the gathered data, the accuracy of the information is one—but not the only—of the key factors to be taken into account. For this purpose, the PAPA (Privacy, Accuracy, Property, Accessibility) framework, developed for addressing the potential issues of information age [6], is leveraged. While the principle idea of PAPA framework still applies in modern technical environments, modifications and additions are required to thoroughly address all aspects of IoT data. This paper focuses on the effects of IoT on data accuracy. A more detailed examination of the other PAPA categorizations (Privacy, Property and Accessability) is outside the scope of this paper, and are examined in depth in other publications [7], [8].

In chapter II, the concept of IoT and its characteristics as a tool for producing data are explained. In chapter III, we introduce the ethical PAPA framework and its notions regarding data accuracy. In chapter IV we address four central factors defining the reliability of the conclusions driven by data. Firstly, the correctness of data and the factors defining it are introduced. Secondly, the role of selecting the right data is explained. Thirdly, we share our insight on the role of data analysis. Finally, the accuracy of governance is addressed. The chapter V provides views on how the research on IoT data accuracy should be widened to provide more thorough understanding on not only fundamental level, but also in form of concrete requirements needed to ensure accuracy. In chapter VI, we introduce our conclusions.

II. INTERNET OF THINGS

The Internet of Things (IoT) is a constantly evolving technology enabling its users to collect data from a large variety of monitorable objects. [9] The fundamental idea of IoT is built around three involved actors/aspects: the monitored entity as a data source, internet as a channel for transferring the data, and a human actor to whom the collected information is delivered in a readable form [10]. IoT solutions are widely used and found to provide concrete benefits in different professional environments, such as manufacturing, medical care, agriculture, retail, logistics, and transportation. [11], [12]

One of the beneficial measurement factors in business environments is performance, which can be monitored based on a variety of data types collected with IoT devices in differnt contexts — allthought adopting the possibilities is still challenging [13], [14]. The collected information is often connected to a certain individual operating or interacting with the object of data collection, such as machinery or vehicles, and thus, it is simultaneously possible to investigate personal productivity. Personal monitoring, including the collection of data and its use as a tool supporting decision-making is a central topic in this paper. Thus, we aim to address the potential risk factors related to this phenomenon and how companies can avoid negative outcomes.

The relatively low prices of IoT solutions compared to the amount and quality of data they are capable of producing has a large role when investigating the reasons behind its expansion. Almost unlimited amounts of information can be gathered from different data sources, enabling companies to observe and monitor the points of interest on a very detailed level. However, while the capabilities are vast and the technology itself is under constant development, accuracy of the gathered information cannot be taken for granted. Two main causes for potential issues are technical shortcomings, and perhaps even more importantly, human factors, which are described in more detail in chapter IV.

III. PAPA MODEL

The basis for investigating the role and impacts of data accuracy in decision-making, especially from the ethical standpoint, is built around the PAPA framework introduced by Richard O. Mason [6]. The original purpose of PAPA is to describe the ethical issues of the information age—it spans privacy, accuracy, property, and accessibility. In this chapter, each issue category and its relation to accuracy are described.

A. Privacy

Privacy is described as a right to seclude and protect information connected to an individual. It is suggested that individuals should be able to decide if, when, how, and to whom information about themselves can be revealed [6], [15]. Furthermore, central questions related to privacy are: what information should be revealed, under what conditions and with what safeguards? As the value of data grows alongside the developing potential for leveraging it for productivity and/or performance purposes, it is crucial that the companies aim to retain the privacy of employees when considering what information gets collected. Related to data accuracy and the possible lack thereof, the importance of limiting the visibility of individual data is emphasized further to avoid both privacy infractions and the distribution of false information.

Regulatory and legal measures have been taken to protect individual privacy, with the General Data Protection Regulation Act (GDPR) as a prominent example [16]. While regulations manage to provide some protection for individuals, a grey area does exist, where questionable approaches towards privacy can be found ethically questionable, yet legally allowed. [4] The connection between privacy and accuracy includes topics such as the justifiability of collecting and using potentially inaccurate and individually identifiable performance data, possibly leading to false conclusions and negative impacts towards the involved individual.

B. Accuracy

Mason associated the concept of accuracy with the responsibility of authenticity, fidelity, and accountability of information, as well as the accountability for errors and harmful events. Here, information accuracy is bound together with system accuracy, emphasizing that developers have a significant role in ensuring avoidance of errors. While this statement can be still be applied in modern business environments, it must be taken into account that the IoT ecosystems can be highly complex, which makes the distribution of responsibility more difficult, as we need to consider both the technological standpoint and the entire sosio-technical system. In this paper, we focus largely on the management's responsibility, but simultaneously underline that the whole network of actors must be taken into account.

Furthermore, a central topic of this paper is related to the potential harm related to the position of an employee caused by data inaccuracy. In this context, Mason used the terms 'misuse' and 'misinterpretation', which are bound to each of the accuracy-affecting areas introduced in the following chapter: accuracy of data, selection of data, accuracy of analysis, and the accuract of governance. Additionally, it covers situations where data is used falsely either intentionally of unintentionally, resulting in personal discrimination. According to Mason, employees should be aware and able to control data collection and distribution and to be protected from discrimination resulting from leveraging erroneous or unethically collected data—emphasizing that discrimination can appear in several different forms. Again, while regulations provide employees with a certain degree of protection, we claim that creating ethical guidelines is a necessary addition to thoroughly protect the individuals. The issue of combining data that is professionally irrelevant also needs to be brought up, including situations where the border between professional and personal is being crossed [17], as this enables companies to conduct ethically questionable profiling [18].

In addition to traditional data accuracy, some thought needs to be given to system accuracy as well. Besides the data collection itself, organisational structures and information processes affect data and its accuracy. The surrounding business ecosystem, deployment planning and implementation, parameter selection, rule definitions and other implicit decisions made during the whole life-cycle of data collection and analysis all have major implications on what results data-driven decision-making will have.

C. Property

The concept of property involves matters related to the ownership of the collected data. The investigation of data property revolves around questions such as who owns the information, what are the just and fair prices for is exchange, who owns the channels through which the information is transmitted and how the access should be allocated.

It is under debate whether the rights to and ownership of the collected information attached to a specific individual should remain fully or partially as the observed individual's property. Again, if inaccurate data is collected from an employee and owned by the company - or an external party - issues affecting an individual's personal position may emerge. This binds together multiple issue categories introduced in PAPA: if the data is owned by the company (see property), is the data owner allowed to reveal (see privacy and accessibility) it without permission?

D. Accessiblity

Accessibility is a category partly intertwined with privacy, including further considerations related to who can access the data. Briefly explained, privacy focuses on data collection, while accessibility concentrates on data distribution.

Obvious issues can be found in situations where inaccurate information related to an individual is collected and distributed among any group of individuals. Without careful consideration regarding with whom the data should be shared with, the potential of negative information about the individual being leveraged to weaken their position raises. As the IoT ecosystems often involve not only the company's internal personnel, but also external actors such as service providers, partners and/or customers, it is apparent that significant issues can arise if corrupted information gets shared.

IV. FACTORS DEFINING IOT DATA ACCURACY

The original PAPA model has been further developed, as need arises, to envelop new technologies and environments of the digital age. For example, topics such as behavioral surveillance, interpretation, and governance have been identified as new ethical issues specifically surrounding big data [19]. Other emergent technologies, such as AI, also require their own additions [20]. In this chapter we examine the different dimensions of IoT data as it relates to data accuracy.

A. Accuracy of Data Collection

The accuracy of the actual data is a crucial factor, considering the reliability and justification of data-driven decision-making. For example, in cases where employees' performance is being measured and actions taken in terms of career development, recruitment or layoffs, leveraging inaccurate data will have a negative impact on the position of both the employees and the company.

From the company's perspective, inaccurate data can be used both intentionally and unintentionally. Intentional use of inaccurate data results in ethical issues, which can do significant harm towards the company's image and internal relations. On the other hand, unintentional use of inaccurate data will likely reduce or remove the aimed benefits of data-driven decision-making.

Considering the factors affecting the accuracy of the collected data, the quality of the utilised IoT solutions is important. The pricing of the solutions varies, and so does their quality. While multipurpose IoT devices do offer companies a cost-efficient opportunity to leverage a variety of beneficial features, compromises in quality tend to be inevitable. [21] In addition to a need for cost-efficiency, different types of optimization requirements, such as a need for smaller devices or low energy consumption may negatively affect the quality of the data. Additionally, reusing the same components in multiple different models of sensors will complicate cross-validating results, creating the possibility of systematic errors even when validating/calibrating different models of sensors against each other. The data quality issues can manifest themselves as problems in actual measurement accuracy, timeliness, completeness, relevance (utility), volume or concordance between data sources. [22]

B. Selection of Data

When selecting the type, source and collection time of data, companies have a responsibility to follow ethically just practices. Correct data collection decisions simultaneously enhance the potential to gain useful and reliable information about the performance of employees and reduce the possibility of harming individuals' through an incorrect collection of information and/or distortions to analysis.

Some of the key questions to consider in this context are: are the variables selected in a manner conducive to reliable and rational analysis? Do the variables actually measure the real-world phenomena tracked? Is the granularity of measurements correct? Can we rely on the accuracy of each selected and measured attribute? If measurements are utilized in tandem, are they comparable and compatible? The importance of accuracy is emphasized when multiple types of data are utilized in tandem, as one distorted attribute can eliminate the reliability of the entire analysis.

Wrong data selection practices may cause distortions in later analysis, even if the analysis itself is conducted perfectly. Simple mistakes, such as tracking daily averages instead of granular per-second values, may cause a complete loss of meaningful information, as the important bits are lost to miscalibration. Wrongfully selecting which attributes to track may create a false sense of accuracy, as everything from data gathering to data management and analysis is functional, but the measurements taken do not accurately represent the facts decision-makers think they represent.

C. *Accuracy of Analysis*

Even when the data has been gathered and selected based on correct practices, falsely conducted analysis may invalidate the results. Again, the issues can be a result of either intentional or unintentional practices. While good practices for data analysis have been covered in literature, it cannot be taken for granted that companies have the knowledge and expertise to conduct a practical and ethically correct analysis.

A thoroughly considered selection of data can reduce the potential for false analysis results caused by complexity, however, it may also dismiss relevant parameters. For instance, investigating solely the usage level of industrial machinery in a specific timeframe to observe individual performance levels while disregarding the lack of material available for processing is likely to result in false conclusions regarding the employee's performance. Furthermore, it is possible that maximizing machinery usage time is not the right goal to begin with, and deciding to optimize other metrics, such as the amount of produced end-product, product quality or amount of coordination between different employees, would result in bigger profits. This emphasizes that conducting reliable analysis often requires not only IoT data but also supportive information and effective control processes. It is also worth mentioning that leveraging artificial intelligence (AI) for data analysis is a growingly common approach. While this may reduce the number of potential issues caused by human error, AI cannot be trusted blindly, and thus cannot be considered a tool for eliminating issues caused by human behaviour.

D. *Accuracy of Governance*

As noted before, data does not reside in a vacuum, and many human factors affect the accuracy of data-driven decision-making. From lower-level data governance to higher-level organisational (or inter-organisational), governance processes can affect data accuracy in a multitude of ways [23], [24].

Data governance includes many information processes that ensure that data is managed in a way that ensures its high quality. Whether loss of data due to physical corruption or communication disconnections, alteration of data due to accidental edits or migration mishaps, or intentional deletion or manipulation by bad actors, bad data governance enables multiple avenues for data inaccuracies. The existing data quality management frameworks such as Wangs Total Data Quality Management [25] are suitable for governing IOT data but specific IOT data quality management frameworks are still in their infancy. [26]

Furthermore, organisation-level governance highly affects the outputs of data-centric decision-making [27], [28]. Whether the underlying data is an accurate representation of reality or not, the decision-making process can lead to unreliable or undesirable outcomes, if there are problems in the governance process. Therefore, governance should note the ethical issues of data use as data use may be affecting the individuals from where data is collected. Accuracy at all levels is an issue that should be governed to achieve the possibility to make the right decisions based on the right data instead of a false one.

V. DISCUSSION

This paper attempts to provide the reader with a fundamental understanding of the potential pitfalls related to data accuracy in IoT use, in terms of correctness and ethicality. To thoroughly understand the mechanics behind these issues and to find concrete actions needed to ensure accuracy, technical-level research is needed. That said, our principal intention is to investigate the risk factors involving human actors and to address the issues on which they have an impact on, namely decision-making. As the sensor data itself is produced by technical devices without the capability of logical reasoning, the accountability of information cannot be fully guaranteed by the companies implementing these solutions. The importance of this matter is especially high as the IoT is a rapidly developing and growing area of technology [4]. As the trustworthiness of the gathered data increases thanks to the development of the devices and solutions, we get closer to a situation where the mechanical pitfalls diminish, yet the impact of human factors likely retains its relevancy.

Another topic for further inspection is how to implement the suggested approaches in the companies applying IoT solutions. Especially considering the smaller companies, technical expertise cannot be taken for granted, and neither their ability to put the suggested approaches into practice. This leads to the question of how the responsibility of retaining accuracy and ethicality should be realistically distributed. While the IoT solution providers can direct this development to some degree, the human actor will always have a critical impact and the ability to harm individuals. That said, we see that providing all-encompassing guidelines is not currently a realistic goal, and thus we should aim to gradually decrease the realization of the introduced risks.

VI. CONCLUSIONS

While IoT solutions provide companies with significant benefits due to their seemingly unlimited data collection capacity with often moderable expenses, the reliability of the data cannot be taken for granted, nor the drawn conclusions. The lifecycle of information from unprocessed sensor data to a trustworthy tool supporting decision-making involves a large variety of pitfalls. This paper contributes to the research gap by addressing the issues relevant specifically to IoT environments. A holistic overview of potential pitfalls regarding IoT data accuracy is introduced, as well as ideas on how to mitigate these issues.

While the accuracy of the actual data is the most obvious point of interest when considering the reliability of the gathered information, human factors have a large role as well. The two key areas in which human error can potentially damage the desired outcomes are the selection of data and the analysis phase. Firstly, the types of gathered data must be carefully selected to produce reliable conclusions, as either ignoring relevant parameters or including irrelevant ones can distort the results of the analysis phase. Thus, careful planning is critical before the solution implementation: what are the attributes necessary to cover to draw trustworthy conclusions? Can the accuracy of each selected data type be relied on considering the capability of the used solution?

It turns out that regulatory protection alone cannot safeguard employees to a sufficient degree, as opportunities for intentional and unintentional misuse are hidden throughout the decision-making lifecycle. This supports the idea that an ethical framework addressing the issues related to IoT use is needed, offering a fundamental understanding of the requirements of ethically sustainable practices covering the phases of design, implementation and actual use. This paper addressed the topic of accuracy, being a part of a larger study in which the applicability of the original PAPA categories is investigated in the IoT context. As a result, a separate publication will be released, in which the modernized and IoT-specific revision of the framework is introduced, including a modified representation of the four PAPA categories, accompanied by complementary issue categories enabling full coverage of the IoT ecosystem.

REFERENCES

[1] N. Côrte-Real, P. Ruivo, and T. Oliveira, "Leveraging internet of things and big data analytics initiatives in european and american firms: Is data quality a way to extract business value?" *Information & Management*, vol. 57, no. 1, p. 103141, 2020.

[2] P. Brous, M. Janssen, and P. Herder, "The dual effects of the internet of things (iot): A systematic review of the benefits and risks of iot adoption by organizations," *International Journal of Information Management*, vol. 51, p. 101952, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0268401218309022

[3] S.-I. Chang, L.-M. Chang, and J.-C. Liao, "Risk factors of enterprise internal control under the internet of things governance: A qualitative research approach," *Information & Management*, vol. 57, no. 6, p. 103335, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S037872062030272X

[4] M. Vermanen, M. M. Rantanen, and V. Harkke, "Ethical framework for iot deployment in smes: individual perspective," *Internet Research*, vol. 32, no. 7, pp. 185–201, 2022.

[5] G. Baldini, M. Botterman, R. Neisse, and M. Tallacchini, "Ethical design in the internet of things," *Science and engineering ethics*, vol. 24, pp. 905–925, 2018.

[6] R. O. Mason, "Four ethical issues of the information age," *MIS quarterly*, pp. 5–12, 1986.

[7] M. Vermanen, M. M. Rantanen, and J. Koskinen, "Privacy in internet of things ecosystems–prerequisite for the ethical data collection and use by companies," *Human Choice and Digital by Default: Autonomy vs Digital Determination: 15th IFIP International Conference on Human Choice and Computers, HCC 2022, Tokyo, Japan, September 8–9, 2022, Proceedings*, pp. 18–26, 2022.

[8] M. Vermanen, J. Koskinen, and V. Harkke, "Internet of things (iot) data accessibility: ethical considerations," *Well-Being in the Information Society. Fruits of Respect: 8th International Conference, WIS 2020, Turku, Finland, August 26–27, 2020, Proceedings 8*, pp. 197–208, 2020.

[9] D. Miorandi, S. Sicari, F. De Pellegrini, and I. Chlamtac, "Internet of things: Vision, applications and research challenges," *Ad hoc networks*, vol. 10, no. 7, pp. 1497–1516, 2012.

[10] J. Jin, J. Gubbi, S. Marusic, and M. Palaniswami, "An information framework for creating a smart city through internet of things. ieee internet things j. 1 (2), 112–121 (2014)," *IEEE Internet of Things Journal April 2014*, 2013.

[11] J. Kaur and K. Kaur, "Internet of things: A review on technologies, architecture, challenges, applications, future trends." *International Journal of Computer Network & Information Security*, vol. 9, no. 4, 2017.

[12] P. Tadejko, "Application of internet of things in logistics–current challenges," *Ekonomia i Zarządzanie*, vol. 7, no. 4, pp. 54–64, 2015.

[13] I. Nappi and G. de Campos Ribeiro, "Internet of things technology applications in the workplace environment: A critical review," *Journal of Corporate Real Estate*, vol. 22, no. 1, pp. 71–90, 2020.

[14] H. Tran-Dang, N. Krommenacker, P. Charpentier, and D.-S. Kim, "The internet of things for logistics: Perspectives, application review, and challenges," *IETE Technical Review*, vol. 39, no. 1, pp. 93–121, 2022.

[15] H. J. Smith, T. Dinev, and H. Xu, "Information privacy research: an interdisciplinary review," *MIS quarterly*, pp. 989–1015, 2011.

[16] European Parliament and Council of the European Union, "Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation)," *Official Journal of the European Union*, 2016.

[17] E. Oriwoh, D. Jazani, G. Epiphaniou, and P. Sant, "Internet of things forensics: Challenges and approaches," in *9th IEEE International Conference on Collaborative computing: networking, Applications and Worksharing*. IEEE, 2013, pp. 608–615.

[18] S. Wachter, "Normative challenges of identification in the internet of things: Privacy, profiling, discrimination, and the gdpr," *Computer law & security review*, vol. 34, no. 3, pp. 436–449, 2018.

[19] J. Young, T. J. Smith, and S. H. Zheng, "Call me big papa: An extension of mason's information ethics framework to big data," *Journal of the Midwest Association for Information Systems (JMWAIS)*, vol. 2020, no. 2, p. 3, 2020.

[20] B. C. Stahl, "From papa to papas and beyond: Dealing with ethics in big data, ai and other emerging technologies," *Communications of the Association for Information Systems*, vol. 49, no. 1, p. 20, 2021.

[21] M. Vermanen and V. Harkke, "Findings from multipurpose iot solution experimentations in finnish smes: common expectations and challenges," *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 2019.

[22] C. Liu, P. Nitschke, S. P. Williams, and D. Zowghi, "Data quality and the internet of things," *Computing*, vol. 102, no. 2, pp. 573–599, 2020.

[23] R. Abraham, J. Schneider, and J. vom Brocke, "Data governance: A conceptual framework, structured review, and research agenda," *International Journal of Information Management*, vol. 49, pp. 424–438, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0268401219300787

[24] M. Janssen, P. Brous, E. Estevez, L. S. Barbosa, and T. Janowski, "Data governance: Organizing data for trustworthy artificial intelligence," *Government Information Quarterly*, vol. 37, no. 3, p. 101493, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0740624X20302719

[25] R. Y. Wang and D. M. Strong, "Beyond accuracy: What data quality means to data consumers," *Journal of management information systems*, vol. 12, no. 4, pp. 5–33, 1996.

[26] L. Zhang, D. Jeong, and S. Lee, "Data quality management in the internet of things," *Sensors*, vol. 21, no. 17, p. 5834, 2021.

[27] P. Brous, M. Janssen, and R. Vilminko-Heikkinen, "Coordinating decision-making in data management activities: a systematic review of data governance principles," in *Electronic Government: 15th IFIP WG 8.5 International Conference, EGOV 2016, Guimarães,*

Portugal, September 5-8, 2016, Proceedings 15. Springer, 2016, pp. 115–125.
[28] D. G. Broo and J. Schooling, "Towards data-centric decision making for smart infrastructure: Data and its challenges," *IFAC-PapersOnLine*, vol. 53, no. 3, pp. 90–94, 2020.