Trustworthy Distributed Intelligence for Smart Cities

Xiaoli Liu*, Satu Tamminen†, Sasu Tarkoma*, Xiang Su‡†
* University of Helsinki, Helsinki, Finland
† University of Oulu, Oulu, Finland
‡ Norwegian University of Science and Technology, Gjøvik, Norway

***Abstract*—The future of smart cities has been significantly impacted by Internet of Things (IoT) and distributed intelligence, where a large scale of data are collected from massive amounts of heterogeneous devices and distributed intelligence brings storage, computing, and Artificial Intelligence (AI) functionality close to the end devices where data are generated for providing novel services and applications. However, AI empowered systems face many challenges due to the inscrutability of complex AI models which weakens the trust of users. This paper provides a general understanding of the underlying concepts and challenges in trustworthy distributed intelligence. A use case of district heating network is illustrated to explore the proposed concepts, technologies, and challenges for enabling trustworthy distributed intelligence for smart cities.**

***Index Terms*—Distributed Intelligence, Trustworthy, Smart Cities**

I. Introduction

Smart cities are pioneers in developing innovative, trustworthy, sustainable, and integrated solutions to become greener, more efficient, and better places. Novel services and applications in smart cities heavily rely on large scale data collected from massive amounts of heterogeneous devices to make informed decisions based on Artificial Intelligence (AI) technologies. We observe the opportunities offered by AI technologies to significantly improve smart cities by unitizing big data collected from sensing devices, such as in smart city applications of smart education [1], smart traffic lights [2], and smart grid [3]. Furthermore, many projects related to smart city and big data analytic have been carried out to improve the living standards of the citizens and maintain the society's sustainability. Both opportunities and challenges have been introduced, ranging from managing city infrastructure, effectively and efficiently analyzing sensing data, developing novel services for citizens and stakeholders, to data privacy and trust.

Although AI empowered systems enable interpreting and understanding of data, the inscrutability of complex AI models weakens the trust of users, especially in contexts where the consequences are significant [4]. Trust is an emerging major concern for the development of AI models in smart cities. Trust is usually associated with a trustee, which is responsible for making the best choice on behalf of other parties. Trust between the trustee and other parties could be affected by the past experience. To involve human being into the interaction loops and increase the trustworthiness, smart city systems require deep understanding of the complex data. Understandable AI can be provided via user-friendly cognitive interface that enables the automated interpretation of the AI models, which helps human being to understand the functionality of Machine Learning (ML) model, prediction and control feedback to increase the system trustworthiness. Moreover, the sensitive data collected by the AI systems raises huge concerns on safety and privacy, such as user identity leakage. Data anonymization and synthesis provide potential solutions to ensure that no individual can be recognized through data analysis, which faces the risk of re-identification [5]. Furthermore, anonymization techniques can impact data distribution and variable relationships which further could result in misinterpretation of the causality [6].

In this paper, we investigate trustworthy distributed intelligence for smart cities, which combines the innovations from 1) trustworthy AI; 2) distributed intelligence leveraging federated learning (FL); and 3) data synthesis and privacy preservation techniques. We analyse the key aspects for trustworthy distributed intelligence including explainability, transparency, fairness, robustness, and privacy preservation. Explainable AI provides the methods to help human understand and trust the results of AI systems. Distributed intelligence with FL plays a significant role in privacy preservation. Leveraging a reliable distributed AI system based on edge-cloud city-scale Internet of Things (IoT) architecture, the synthesized sensor data can be collected across the entire system. FL enables model training locally, where training dataset for user profiling stays in local servers and only a collection of model parameters updates to a centralized server. We optimize the balance between user privacy preservation and synthesized data analysis with usefulness and correctness measurement, realize distributed AI based on edge-cloud architectures, guarantee data security utilizing FL, and improve users' trust in AI systems via user-friendly cognitive interface.

This paper provides a general understanding of the underlying concepts and raises awareness of emerging directions in trustworthy distributed intelligence, which benefit researchers, engineers, service providers, government, and public sectors. The paper is organized as follows. We introduce the key aspects and supporting technologies for trustworthy distributed intelligence in Section II and challenges for enabling trustworthy privacy-preserving distributed intelligence in Section III. A motivation example is presented in Section IV. Finally, the

paper is summarized with a conclusion in Section V.

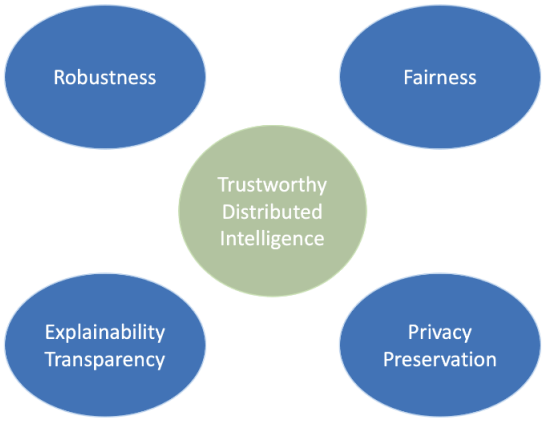

Fig. 1. Key aspects for trustworthy distributed intelligence.

II. Towards trustworthy distributed intelligence for smart cities

In this section, we introduce the key aspects for trustworthy distributed intelligence. As shown in Figure 1, these aspects include explainability, transparency, fairness, robustness, and privacy preservation. For addressing challenges of trustworthy distributed intelligence, we propose key supporting technologies and further highlight the extended impact.

A. Trustworthy distributed intelligence

Existing distributed AI systems are found vulnerable to imperceptible attacks, biased against underrepresented groups, lacking in user privacy protection, which not only degrades user experience but also erodes the society's trust in all distributed AI systems. Therefore, various aspects of distributed AI systems should be carefully considered to improve their trustworthiness.

1) Explainability and Transparency: In general, the information opacity of AI systems inevitably harms its trustworthiness. Explainability and transparency of distributed AI systems address the concerns of AI opaqueness and are recognized to help build the trust of the society on AI technology [7]. Explainability aims to verifying and understanding the output decision made by AI agents and algorithms, which serves as a fundamental factor that determines the trust in AI technology. The motivation of AI explainability pertains to various aspects. From the perspective of scientific research, it is meaningful to understand all bits of the data, parameters, procedures, and outcomes in AI systems, which provides insight into the nature of both machine intelligence and human intelligence. From the perspective of building AI products, explainability helps understand the AI system in outlining its defects and also reduce the by-product trade-offs. From the ethical point of view, explainability also helps alleviate the concern about responsibility distinction of the black box model [8].

Transparency requires the information disclosure of a system and is a recognized requirement in the area of software engineering. In the AI industry, this requirement of transparency naturally covers the lifecycle of an AI system and helps stakeholders confirm that appropriate design principles are reflected.

2) Fairness and Robustness: AI system can exhibit bias towards certain factors and thus, needs to be evaluated for fairness. Fairness here is tested by verifying if the bias is valid as per pre-established ethical principles. Distributed AI systems amplify this challenge. Data bias significantly affects the data quality and could lead to discrimination in decision making. The data bias is hard to avoid and can exists in the whole process of data analytics, from data collection and data processing to modelling. Therefore, it is crucial to design the fairness-awareness ML algorithms to achieve the fairness of decisions against illegal discrimination.

Robustness measures the ability of a system to continue to function when changes or incidents happen. For example, the system could give accurate prediction in the normal situation, while if happens sudden accident in network, whether the system could give the accurate prediction. In context of distributed intelligence, designing the robustness algorithms to deal with unexpected/anomaly situations is crucial. The algorithmic robustness might exist in different situations and the definition of algorithmic robustness could have minor difference. For example, algorithm robustness could be defined that the algorithm has quite similar performance on the testing data and training data that are close [9], which means if a model is trained on dataset A and if a small perturbation is added on dataset A, it will not greatly affect the model's performance on the perturbed dataset. However, this is not always the truth. Many works have shown that ML models, such as deep neural network models are vulnerable to adversarial examples [10], where an attacker intentionally designed to cause the model to make mistakes. Goodfellow et al. [11] has shown that it is easily for the model to classify the panda as gibbon by adding an unnoticeable perturbation. The algorithm robustness should also take into the consideration of heterogeneity, such as statistical heterogeneity and systems heterogeneity, to make the model more robustness.

3) Privacy preservation: Another emerging concern for smart systems is privacy. Citizens begin to pay attention to the privacy caused by the collected data. Recognizing privacy as a key obstacle to full promise of digital services and many aspects of user privacy need to be well addressed. For example, rules related to data collection, data processing, and data aggregation should be clearly defined and access right to data should be controlled. Many countries have issued regulation to protecting users privacy. EU commission published General Data Protection Regulation (GDPR), which allows users to control their own data, including data their devices generate. The GDPR states the need for trust to be built into personal data services through a combination of transparency, interchangeability, public governance, respectable companies, public awareness, and secure technology. Following GDPR, IoT systems require local transform and store personal data in a way that user concerned sensitive information is protected

without losing the cloud services' utility.

B. Key supporting technologies for trustworthy distributed intelligence

To enable above-mentioned key aspects for trustworthy distributed intelligence, we summarize key supporting technologies for deployment of trustworthy distributed intelligence in smart cities, including edge-cloud computing, data preparation, and dynamic prediction models.

1) FL systems for smart city: Emerging edge and fog computing take a first step with moving the computation from the central high-powered cloud or server machine to the edge of the networks in smart cities. The development of an edge-cloud IoT system for smart city involves several aspects. Edge facilities reside closer to data sources and end users to provide faster data processing locally. Cloud clusters can be managed by the system provision company in the city to execute global system monitoring and control. Generally, the system deploys numerous sensors and IoT systems connecting sensors with wired/wireless networks. The edge-cloud system collects data from the IoT in real-time and enables comprehensive data analytics to improve the system performance and user experience. The system aims at development of features, including 1) server deployment efficiency, i.e., optimizing the edge server deployment by maximizing the utilization efficiency; 2) communication efficiency, i.e., optimize the system data transport by minimizing the transmission latency; and 3) data utilization efficiency, i.e., optimize the data collection and aggregation by selecting and transporting only the most important data.

To fully excavate the value of the data, the edge-cloud system can be empowered with 1) modelling and analyzing collected IoT data with ML algorithms; 2) supporting personalized services while guaranteeing privacy preservation; and 3) improving local and global system performance and user experience. To realize above-mentioned functions, FL can be deployed in the edge-cloud system. FL is a new ML paradigm, which enables training and inference in a distributed manner. Instead of aggregating the data to train a centralized model, the central cloud distributes a global model to the edge servers on which the model is trained and updated. The edge servers send the updated parameters back to the central server, which then aggregates the local updates into the global model. In practice, FL is a form of distributed optimization, and the procedure is iterated for as many rounds of communication as needed.

2) Data preparation for machine learning and trustworthy AI: Smart evidence based decision support is based on statistical prediction models. When data is collected from various heterogeneous sources, the quality of data is playing an essential role. The prior knowledge of the physical systems can be utilized to improve the reliability of the data. Dynamic models rely on high quality data. To support distributed intelligence in IoT systems, the first task is to estimate sensor measurement reliability and to ensure that adaptation is based on high quality data. In addition, context information should be used to detect reasons for abrupt change in sensor value distribution. The causality of the stability-plasticity dilemma and context change can be solved by building separate models for each context, for example.

The ability to increase the understandability of ML models requires reference information that enables to find out what has changed or how specific user groups differ from each other. For example, consumer profiling and similarity measures for different groups may serve this purpose. The replacement of highly sensitive consumer data with synthesized data protects the privacy of the customers. In addition, the presence of sensitive data in IoT systems should be considered. The safety of the data and privacy of people can be compromised when data need to be transferred to a cloud for modelling, for example. Data processing should be performed at the edge when computation resources allow. Data synthesis enables the use of artificial data with similar distributional properties for modelling purposes. Then, the sensitive user data need not to be transferred from the collection location.

3) Dynamic prediction models for smart city applications: In order to produce models with a long life cycle for smart city applications, the models should continually learn in dynamic environments. Lifelong learning paradigm enables learning continuously, accumulating the knowledge learned in the past, and adapting it to help future learning and problem solving. Depending on the speed mechanism of the change in data, the models should adapt slowly or quickly to the change. This challenge is called the stability-plasticity dilemma. One of the main challenges is the concept drift, which is caused by changes in the data distribution. Another source of concept drift, when dealing with indirect measurements, is falsely labelled streaming data that is used to update a prediction model. Concept drift will be handled by updating models only using high quality data, and by studying how much dynamic learning is allowed to change the initial model. A concept drift detection identifies different contexts where the model can be updated safely. This way, data coming from the wrong context would not confuse the whole recognition process. Human AI collaboration can also handle concept drift when falsely predicted labels will be replaced with user defined, correct labels, but the human interventions should be minimized. This requires studying how to define prediction as unreliable, for instance this can be done based on posterior values or by training a separate model to predict reliability of the prediction. Moreover, it is highly important to understand the studied phenomenon, as it helps to understand when human intelligence is needed.

III. Challenges

We discuss some prominent challenges for enabling trustworthy and privacy-preserving distributed intelligence for smart cities, including system optimization, privacy-preserving techniques, and trustworthy AI. It is import to address those challenges in order to seek novel solutions.

A. Privacy-preserving techniques

Privacy-preserving techniques need to be designed and implemented not only to preserve sensitive raw data and data

streams of users, but also developed AI models in the edge and cloud. FL is regarded as a privacy-preserving solution in edge-cloud system with distributing ML on edge machines without sending the data to the Could. The edge-cloud architecture needs to leverage the FL in edge architectures for achieving the best possible performance. Many factors, such as data heterogeneity, system heterogeneity, communication cost, and model performance, need to be considered to build reliable and efficient edge-cloud FL.

FL methods develop a global model for all organization/users that is trained with sufficient data and distributed to all organization/users. Such a process lacks personalization if we consider the health scenarios. The process does not fully address the situation of imbalanced or sparse dataset in different domains. Techniques, such as transfer learning and model-agnostic meta-learning (MAML), should be combined together with FL to realize accurate user profiling while guaranteeing preservation of user privacy. Transfer learning addresses this problem and makes it possible to use commonalities between the different domains in order to efficiently train more accurate models, with smaller amounts of data needed per domain. MAML use a different approach but with similar idea to have the initial model for new organization/users and then adapt the initial model for personalization using its local data with one or few step of gradient descent. Meanwhile, enhanced privacy should be guaranteed by combing FL with other methods, such as Multi-party computation, Homomorphic encryption, and Differential Privacy.

B. Trustworthy AI

Interpretation of AI models helps the users to understand functionalities of the ML models and the reasoning behind the predictions. However, it should be noted that the explanations are based only on the data in hand. It is important to consider how to weigh the model and data distribution to the local explanation and how to have clear explanations for causality between variables. Regarding to the fairness of AI system, different fairness-awareness ML methods could be applied in different process to achieve free discrimination. Techniques, such as massaging, reweighing and sampling, have been used in preprocessing for discrimination-free classification. However, achieving fairness in ML methods is not simple due to different definition of fairness. It is important to understand what kinds of fairness that the system wants to achieve beforehand to design the corresponding fairness algorithms to improve the system trustworthy. It is challenging to design the explainable, fairness, and robust algorithms under the FL setting.

C. Network optimization in FL

Many FL systems in smart cities require minimization of data processing latency, conservation of the bandwidth consumption, collecting and securing data across a wide geographic area, and addressing security, privacy, and system reliability concerns. Design and implement an efficient FL system based on real-world infrastructures is important for the realization of the smart cities. The system needs to be optimized by taking many factors into consideration, such as edge server deployment, task offloading between edge and cloud, and data processing efficiency. For example, for the task-offloading, load balancing mechanism should be deployed to decide the amount of tasks need to be offloaded to the cloud server based on the quality of service (QoS). There are different criteria to define the QoS, for example optimisation of resource usage, energy efficiency, and processing delay. It is important that the system could take the specific requirements of the use case into consideration while can also be easily adapted to other scenarios. Furthermore, data on the edge are heterogeneous and some data suffer from low quality, for example, the data can be repetitive or even are not relevant to the learning tasks. Data importance mechanisms should be designed to optimize the whole network.

IV. A Motivating Example

Trustworthy distributed intelligence has a great potential to be implemented to various scenarios of smart cities. In this paper, we present the use case of district heating network, as shown in Figure 2. The goal of this scenario is to predict the energy consumption of customers in district heating system of smart city and the serving capability of the network in different situations, such as different weather conditions, building type, size and age of the building, and time and calendar in the year. The data was collected in Oulu area from Oulun Energia district heating network during 2015-2021 and it contains hourly measurements of energy consumption, supply and return temperature, and the flow of the water from over 10 000 customers in different locations in Oulu area. In addition, information about the buildings (location and zip code, volume and area, year of construction, number of apartments, and purpose of use) and outside temperature was collected. In total, there are over 140 Million measurements in the data set. Main stakeholders are the customers, the energy provider, and city of Oulu. Customers can adjust their consumption and save energy with tools that advice for greener habits, energy provider get valuable information about network's operational status, and city of Oulu can use the information in urban planning, for example, optimal locations for energy partners producing waste heat for circulation in district heating system.

In the distributed cognitive AI empowered decision support system for heating network, the edge layer is mainly responsible for data synthesis, data pre-processing, and local model training, and the cloud is for global model, data storage, and energy management applications. In real-world data, the ability to clean data before actual use determines the success of developed AI tools. Especially, if models learn on the edge. FL enables to build models that are based on individual energy consumption and weather and calendar data for customers without compromising privacy. Personalized energy saving advising systems can be developed with these models.

With understandable ML and cognitive AI, it is possible to provide useful energy consumption information for the users and advise them to choose more environmentally healthy

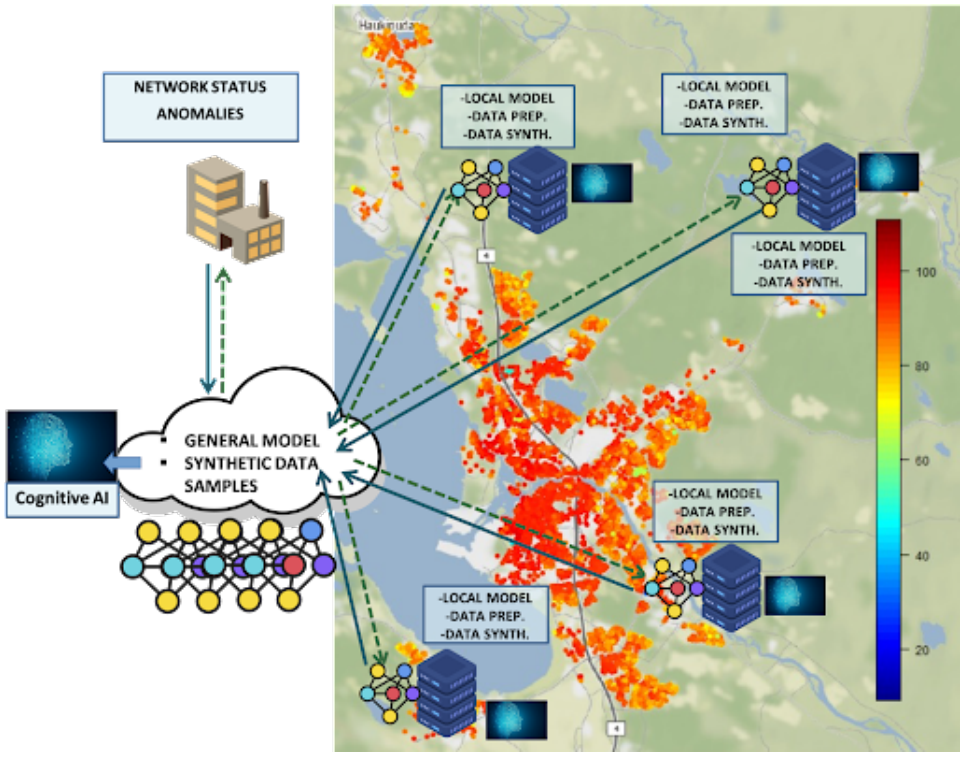

Fig. 2. The Cognitive AI empowered decision support system for distributed heating network use case.

habits. Meanwhile, the environment changes which requires the model need to be learned continuously to maintain the robustness of the model. The network's serving capability information and anomaly information is valuable for the energy company. This information can be used to assist in network optimization and also in urban planning with collaborative companies or the city. The district heating network is an energy platform that enables the building an ecosystem of different stakeholders around it.

In order to utilize the ML methods and data mining in their full potential, the user should have advanced skills in data analytics. Cognitive AI will present from all produced information the contextually relevant incidents that require preventive actions or that are useful to the end user. In order to increase the trustworthiness, there is a need for an user friendly UI to help customers understand how the system functions with understandable AI. For decision support tools, it is crucial to provide information about the reasoning behind the predictions. Transparent and understandable ML is enabled with visualization techniques that can be utilized in order to derive information that help end users to understand how the predictions have been made. The role of the end users determine what kind of information is needed. Utilization of complex ML models may not require advanced understanding of statistics or ML techniques. Instead, cognitive AI can perform tasks that require model interpretation or refining the information and linking it to related phenomena which is usable knowledge to the user.

AI attracts a wide attention from stakeholders of smart city. Although stakeholders and users look forward to leverage AI techniques, the trustworthiness and explainability of AI remain questionable and thus delay its deployment. Therefore, it is a responsibility of AI developers not only develop complex ML techniques to provide comprehensive data analysis ability and improve the performance of our use case, but also to guarantee trustworthiness with solid provement.

V. Conclusion

Trustworthy is recognized as a major concern for the development of smart cities, because AI system can only be successful if it is trusted by humans. This involves several aspects, such as explainability, transparency, fairness, robustness, and privacy preservation. For example, distributed AI system needs to be able to explain its actions and why it ends up in its current state. Secondly, the intelligent system should act lawfully respecting all applicable laws and regulations, ethically respecting the right principles and values, and technically robust and fair while considering its social environment. Thirdly, the systems should involve humans when they are needed.

We propose key supporting technologies and challenges for trustworthy AI-aided decision making and present a real-world use case about distributed heating network. Smart city IoT networks produce vast amounts of data and the utilization of it is impossible for humans. For example, AI can recognize abnormal situations or changes for preventive actions. However, people do not trust AI if they do not know the grounds behind the decisions. There are some trust-related issues that slow

down the digital transformation in smart cities. Complex ML models should be made more trustful by using techniques that increase the interpretability and understandability. There has also been some critical concerns towards IoT networks as well because of sensitive data. The data privacy can be protected by using synthetic data in functionalities that locate in the cloud. The lifecycle of models is important and continually learn models can adapt to changes to lengthen their lifecycle.

REFERENCES

[1] Zhu, Z. T., Yu, M. H., Riezebos, P. (2016). A research framework of smart education. Smart learning environments, 3(1), 1-17.

[2] Kanungo, A., Sharma, A., Singla, C. (2014). Smart traffic lights switching and traffic density calculation using video processing. In 2014 recent advances in Engineering and computational sciences (RAECS), 1-6, IEEE.

[3] Fang, X., Misra, S., Xue, G., & Yang, D. (2011). Smart grid—The new and improved power grid: A survey. IEEE communications surveys & tutorials, 14(4), 944-980.

[4] Rai, A. (2020). Explainable AI: from black box to glass box. Journal of the Academy of Marketing Science, 48, 137–141. https://doi.org/10.1007/s11747-019-00710-5

[5] Rocher, L., Hendrickx, J., Montjoye, Y-A (2019). Estimating the success of re-identifications in incomplete datasets using generative models. Nature Communications 10, 1-9.

[6] Badr, W. (2019). Why feature correlation matters... A lot!. Towards Data Science, 2019.

[7] Lipton, Z. C. (2018). The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery. Queue, 16(3), 31-57.

[8] Li B., Qi P., Liu B., Di S., Liu J., Pei J., Yi J., Zhou B. (2021). Trustworthy AI: From Principles to Practices. arXiv preprint arXiv:2110.01167

[9] Xu, H. and Mannor, S. (2012). Robustness and generalization. Machine learning, 86(3), 391-423.

[10] Xu, H., Ma, Y., Liu, H.C., Deb, D., Liu, H., Tang, J.L. and Jain, A.K. (2020). Adversarial attacks and defenses in images, graphs and text: A review. International Journal of Automation and Computing, 17(2), 151-178.

[11] Goodfellow, I.J., Shlens, J. and Szegedy, C. (2014). Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572.

[12] Wang, P., Yang, L. T., & Li, J. (2018). An edge cloud-assisted CPSS framework for smart city. IEEE Cloud Computing, 5(5), 37-46.

[13] Hossain, S. A., Rahman, M. A., & Hossain, M. A. (2018). Edge computing framework for enabling situation awareness in IoT based smart city. Journal of Parallel and Distributed Computing, 122, 226-237.

[14] Digital transformation (2020). An introduction to the Digital Transformation of Industries initiative. http://reports.weforum.org/digital-transformation/an-introduction-to-the-digital-transformation-initiative/

[15] Nahavandi, S. (2019). Industry 5.0 – A human-centric solution. Sustainability 11, 4371: 1-13.

[16] Paschek, D., Mocan, A., Draghici (2019). Industry 5.0 – The expected impact of next industrial revolution. Management Knowledge Learning International conference. 125-132.

[17] Ruvolo, P., Eaton, E. (2013). Active task selection for lifelong machine learning. In Proceedings of the 27th AAAI Conference on Artificial Intelligence.

[18] Silver, D. L. (2011). Machine lifelong learning: challenges and benefits for artificial general intelligence. In International Conference on Artificial General Intelligence, Springer Berlin Heidelberg. 370-375.

[19] Siirtola, P., Röning, J. (2019). Incremental learning to personalize human activity recognition models: the importance of human AI collaboration. Sensors 9(23), 5151.

[20] Losing, V., Hammer, B., Wersing, H. (2018). Incremental on-line learning: A review and comparison of state of the art algorithms. Neurocomputing, 275, 1261-1274.

[21] Bifet, A., Gavalda, R. (2007). Learning from time-changing data with adaptive windowing. In Proceedings of the 2007 SIAM international conference on data mining. 443-448.

[22] Du, M., Liu, N., Hu, X. (2019). Techniques for interpretable machine learning. Communications of the ACM 63(1), 68-77.

[23] Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J., Müller, K. R. (2020). Toward interpretable machine learning: transparent deep neural networks and beyond. arXiv preprint arXiv:2003.07631.