Trustworthy and Sustainable Edge AI

A Research Agenda

Ding, Aaron Yi; Janssen, Marijn; Crowcroft, Jon

DOI
10.1109/TPSISA52974.2021.00019

Publication date
2021

Document Version
Accepted author manuscript

Published in
Proceedings - 2021 3rd IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications, TPS-ISA 2021

Trustworthy and Sustainable Edge AI: A Research Agenda

Aaron Yi Ding*
Delft University of Technology
Delft, Netherlands

Marijn Janssen
Delft University of Technology
Delft, Netherlands

Jon Crowcroft
University of Cambridge
Cambridge, United Kingdom

***Abstract*—As a fast evolving domain that merges edge computing, distributed systems, data analytics and AI/ML, commonly referred as Edge AI, the community of Edge AI is establishing and gradually finds its way to connect with mainstream research communities of distributed systems, IoT, and embedded machine learning. Meanwhile, despite of its well-claimed potential to transform cloud and IoT industry, Edge AI is still a complex subject that faces critical challenges from the trustworthy and sustainable concerns. To shed light on these pressing matters, this paper aims to develop a research agenda for trustworthy and sustainable Edge AI. We clarify the concepts, define the proper scoping and propose a research agenda for Edge AI to be trustworthy and sustainable. To illustrate the research agenda in practice, we highlight two active R&D projects: the SPATIAL project on trustworthy Edge AI and the APROPOS project on sustainable computing. The projects serve as concrete use cases to explore the agenda development. Our goal is to equip researchers, engineers, service providers, government and public sectors with a better understanding of the underlying concepts and raise awareness of emerging directions in trustworthy and sustainable Edge AI.**

***Index Terms*—Edge computing, Edge AI, IoT, trustworthiness, sustainable AI**

I. Introduction

Edge AI represents a fast growing domain that converges edge computing, distributed systems, data analytics and AI/ML. By offering AI functionality at the network edge, Edge AI builds on the advancement of edge computing and AI/ML (e.g., distributed, embedded and tiny ML). As illustrated in Figure 1, Edge AI is an emerging paradigm that augments cloud computing and the Internet of Things (IoT) by bringing storage, computing, and AI functionality close to the end devices/users where data (of large volume) is generated. The edge layer fulfills the gap between IoT/mobile and the cloud in terms of computing power and data intelligence [1]–[4].

Ever since the idea of 'edge' was introduced decades ago [5], research communities, cloud providers and mobile operators have been arguing on the values and relevance of Edge AI for general ICT service industry [6], [7]. Gartner recently predicted that by 2023, over 50% of the primary responsibility for data analytic stakeholders will go to data that is created, managed and analyzed in edge environments. The listed advantages include greater data management flexibility, speed, governance, and resilience [8]. In addition, the edge capabilities have been applied to use cases ranging from real-time event analytics to autonomous driving services [9]–[11].

* Corresponding author: Aaron Ding (aaron.ding@tudelft.nl)

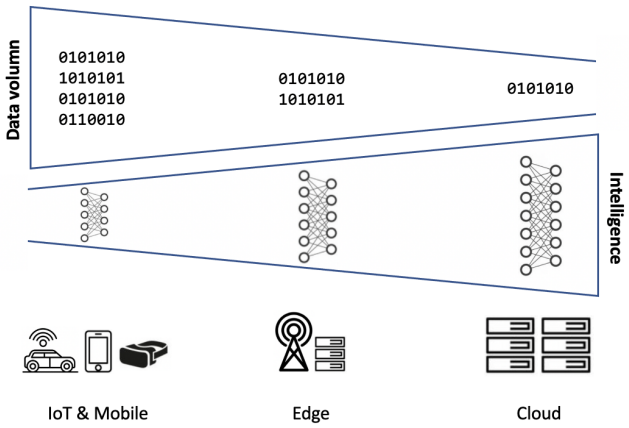

Fig. 1. Edge AI to converge IoT and Cloud.

Despite of its potential, Edge AI is facing two major challenges in recent years to scale up its deployment, coming from trustworthiness and sustainability. These challenges are becoming visible when AI functionality is migrating from an elastic provisioning of centralized platforms (the cloud) to a dynamic and decentralized hosting environment (the edge).

To shed light on the development of Edge AI on the pressing matters of trustworthiness and sustainability, this paper focuses on two essential elements:

- **Raise awareness on trustworthy sustainable Edge AI**: In Section 2 and 3, we clarify the key concepts and define the scoping for trustworthy and sustainable Edge AI, respectively. We reveal new challenges, opportunities and the potential impact.
- **Establish a research agenda**: Section 4 describes the research agenda by suggesting five meta questions that deserve further investigations. Section 5 highlights two active R&D projects, one for trustworthy Edge AI and the other one targeting at sustainable computing aspects, respectively. The on-going initiatives illustrate the research agenda in practice.

This paper is our endeavour to identify core concepts and develop a practical research agenda towards trustworthy and sustainable Edge AI.

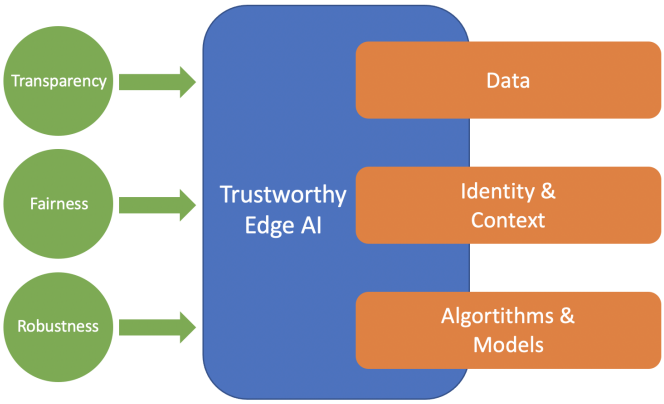

Fig. 2. Trustworthy Edge AI: key elements and scoping.

II. Towards Trustworthy Edge AI

In this section, we clarify the concepts and scoping for trustworthy Edge AI. For major actors deploying Edge AI, we further highlight open challenges and the extended impact.

A. Conceptualization of Trustworthy Edge AI

We identify three key elements contributing to Trustworthy Edge AI: **transparency**, **fairness**, and **robustness**, as highlighted in Figure 2. Given the complexity of involving diverse actors in Edge AI [4], [12], [13], it is crucial to clarify the core concepts concerning the trust related aspects of Edge AI.

First, the trust related concepts in Edge AI should be positioned and scoped. As illustrated in Table I, trust is a multi-faced concept and exists in a relationship with a trustee. For instance, if certain computing tasks are sourced to another party then this party is the trustee. A trustee is expected to make decisions in the best interest of the other parties. When there is a positive experience with the other parties, then this results in more trust, while negative experience results in the opposite. Trust generally depends on the perception of the functioning of other parties and the past experience, e.g., it influences which components/partners a service provider will select. Trust can be enhanced by providing insight into the working, conducting audits and so on. The concept of *trustworthiness* refers to the properties through which a trustee serves the interests of the trustor [14]. Carter also defined trustworthiness as the perception of conviction in the trusted entity's reliability and integrity [15]. This perception usually involves concerns related to reliability, security and privacy. Whereas trust focuses more on the organizations and person-centric systems, trustworthiness can also be about technical systems.

In the Edge AI context, trust can be perceived as the belief in a trustee developing, operating and maintaining the Edge AI based on experiences and reputation. This implies the intention to accept vulnerability [1], [16]–[19]. The trustee include both human being and cyber-physical subjects such as edge hardware and AI algorithms deployed on the edge. The trustworthiness is about the properties of the trustee like the ability, benevolence, and integrity of such trustee (e.g., to perform expected computing operations according to standards, requirements and societal values). Meanwhile, the trust propensity is the willingness to rely on others.

TABLE I
Trustworthy Edge AI Conceptualization [16], [17], [20], [21]

Concept	Definition and Notes
trust	The belief in a trustee and results based on experiences and reputation.
trustworthiness	The properties through which a trustee serves the interests of the trustor, including ability, benevolence, and integrity.
trust propensity	The willingness to rely on others.
transparency	Ability to understand what is happening in the system. Three elements for Edge AI transparency include traceability, explainability and open communication.
fairness	Impartial and just treatment for Edge AI users.
robustness	Ability of a system to keep on functioning, when changes or incidents happen.
accountability	The answerability for Edge AI systems and operators actions or inactions and to be responsible for their consequences.
traceability	The capability to keep track of the system's data, development and deployment processes, typically by means of documented recorded identification.
explainability	The ability to explain both the technical processes of the system and the related human decisions such as the application areas of the system). It entails that an explanation can be formulated.
interpretability	The extent to which a cause and effect can be observed within an Edge AI system.

Second, we need to pay attention to three key elements for trustworthy Edge AI, including **transparency**, **fairness**, and **robustness**. In general, transparency is not easy to define and numerous definitions exist [22]. The level of transparency depends on the stakeholder view and the context. Stakeholders might perceive transparency in different ways. Whereas one stakeholder might be interested in the agreements among parties, another stakeholders might be interested in the protocols used and so on. Transparency implies the ability to see what is happening in a system. Transparency-by-design is about ensuring that systems are transparency. It refers to both the design process and the outcomes of the design process for accomplishing transparency, Transparency-by-design can be defined as 'taking into account transparency in every phase of the design process" [22].

For Edge AI to be trustworthy, the fairness attribute has both a substantive and a procedural dimension [21]. The substantive dimension is on a commitment to ensure equal and just allocation of resources, and also ensure the design free from unfair bias, discrimination and stigmatisation. Fairness can be regarded as the absence of any prejudice or favoritism

toward an individual or group based on their inherent or acquired characteristics. It can refer to individual fairness, group fairness and subgroup fairness [23]. For Edge AI, the algorithmic bias is one particular concern when the bias is not present in the input data and is added purely by the algorithm.

The robustness attribute refers to a system's ability to remain functioning under disturbances. For Edge AI, this implies to cope with failures during execution and handle incorrect feedback in training process. For robustness, we must consider matching the redundancy with complexity.

In addition, for training and deployment of Edge AI in decentralized, uncontrolled environments, there are several related concepts including: accountability, traceabity, explanability, interpretability [24], and privacy in proximate communication [25]. The accountability implies a set of mechanisms, practices and attributes that sum to a governance structure which involves committing to legal and ethical obligations, policies, procedures and mechanism, explaining and demonstrating ethical implementation to internal and external stakeholders and remedying any failure to act properly [26]. As one of the key principles of General Data Protection Regulation (GDPR), an EU regulation that takes effect since 2018, 'accountability' implies that individuals have the right to explanation when using computer algorithms - especially in critical mission applications, such as autonomous driving and smart surgery, as people cannot fully trust on AI's decision without any explanations and legal liability of the actions. The term explainability in the context of AI accountability, is the level to which the internal logic of a machine learning system can be explained in human terms. Lack of explainability has also resulted in excessive time wasted in debugging work towards the cases where the AI results are incorrect, as these results are based on massive training data, which is difficult for people to check manually [27].

B. Scoping

Compared with centralized cloud AI, Edge AI benefits from its close proximity to end-devices and users where data is generated. However, due to its distributed deployment and deep penetration into personal context, the perceived trustworthiness of Edge AI services has raised concerns across numerous stakeholders including end users, developers, service providers, private and public sectors [18]. Critical technology building blocks are demanding clear scoping.

We identify three dimensions for trustworthy Edge AI, including **data**, **identity & context**, and **algorithm & model**, as highlighted in Figure 2.

- **Data**: this dimension of trustworthy Edge AI considers the entire data processing pipeline, including collecting, communicating, storing, and analysing data. Since data has become a new fuel for digital economy, protecting and retaining critical data with respect to trust, privacy and sovereignty are essential for businesses prosperity and social welfare. As highlighted in Figure 1, one visible trend is the shift of numerous connected devices from the role of data consumer towards the data producer. For example, YouTube users can contribute nearly 100 hours video contents while Instagram users posting over 2430000 photos in every single minute [28]. The growth from such 'producer' perspective is unprecedented.
- **Identity & Context**: this dimension concerns the trust across edge devices and identify of involved end-users in the given deployment context. Since identity and context relates closely to trust and data privacy concerns, especially with large-scale training over datasets that are crowdsourced from edge entities and individuals with sensitive information [29].
- **Algorithm & Model** this dimension considers the algorithmic and modeling in Edge AI to be trustworthy. For instance to achieve transparency, we can resort to three components including traceability, explainability (e.g., XAI), and open communication about the system limitation [21]. Algorithmic system transparency as for Edge AI can be global, by seeking insight into the system behaviour for any kind of input, or local, by seeking to explain a specific input-output relationship. [26].

C. Technology and Societal Impact

Trust is playing a key role for edge devices and associated clients, as they collaborate with others. i.e., trusting other edge devices, identifying suitable partners for Federated Learning in a dynamic environment. From organizational perspective, data privacy and sovereignty pave the way to create a fair, trusted environment for key stakeholders (e.g., public sectors, private companies, governments) to collaborate and respond agilely to urgent needs.

Currently we witness lots of focus on scalability and efficacy but not much on making Edge AI robust. With more logic, parameters, modules added to edge systems, failure rate is growing. Adding redundancy could be straightforward solution but the entire system become more complex and less transparent. Explainable AI (xAI) research is paving the way to address the transparency issues of wide deployment of AI intelligent systems.

To be trustworthy, Edge AI also needs to maximise the benefits of applying AI in numerous edge setting, while preventing and minimising the risks. Although there are clear benefits from trustworthy Edge AI, the societal view is far from trustworthy as to its usage in safety critical areas. There is a strong demand for tools and assessment frameworks to support both users and developers in effectively auditing the code and data of safety-critical systems and developing appropriate safeguards.

Takeaway

The trustworthiness of Edge AI is a stepping stone to establish an appropriate governance and regulatory framework, on which the promise of Edge AI can be built. Key enabling elements include transparency, fairness, and robustness.

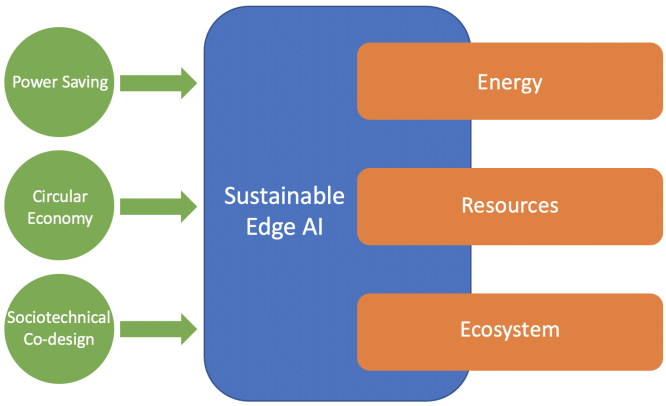

Fig. 3. Sustainable Edge AI: key elements and scoping.

III. Towards Sustainable Edge AI

This section covers the conceptualization, scoping and extended impact for sustainable Edge AI.

A. Conceptualization of Sustainable Edge AI

For Edge AI to become sustainable, we identify three enabling elements that include **power saving**, **circular economy**, and **sociotechnical co-design**, as highlighted in Figure 3.

Edge AI benefits from its latency and can enable more applications for IoT, drones, and autonomous vehicles [9]. However, the fast growth has introduced additional energy requirements at the edge of the network in terms of computing and communication energy. In addition, the devices employed in the edge ecosystem such as IoT, drones are of limited battery life. The power saving is hence a crucial enabler to achieve overall goal of sustainable Edge AI. Besides power saving techniques proposed for mobile devices [30], [31], one promising technique explored by Edge AI community is the approximate computing [32] that can trade off accuracy in computing to less power and less time consumed.

As illustrated in Table II the enabling element of circular economy is commonly referred as a regenerative system where resource and waste, emission, and energy leakage are minimised by slowing, closing, and narrowing material and energy loops. It could be achieved through long-lasting design, maintenance, repair, reuse, re-manufacturing, refurbishing, and recycling [33], [34]. As Edge AI is performed on large amount of embedded battery-powered devices, often in a uncontrolled distributed environments [35], the circular concept in Edge AI is about the minimizing the waste and promoting resource recycling and reuse throughout the edge device life cycle.

From the ecosystem perspective, we advocate the sociotechnical co-design for enabling Edge AI as the sustainability issue is further elevated by the lack of design scope, unclear sustainability requirements on embedded edge devices (e.g., energy saving, carbon emission), unclear value network, and lack of mutual understanding across ISP, cloud providers and Edge AI service providers. In addition, the Edge AI life cycle management part of sociotechnical co-design by covering governance, development, and maintenance for Edge AI resources including software, data, and hardware systems.

TABLE II
Concepts for Sustainable Edge AI [32]–[34], [36]

Concepts	Definition and Concerns
power saving	The techniques for reducing the energy consumption on edge devices.
circular economy	A regenerative system where resource and waste, emission, and energy leakage are minimised by slowing, closing, and narrowing material and energy loops
sociotechnical co-design	A proposed design approach that brings together multiple relevant actors in Edge AI to develop dedicated solutions.
value network	A dynamic network of legally independent, collaborating actors who intend to offer a specific service, and in which tangible and intangible value exchanges take place between the actors involved.
life cycle management	The governance, development, and maintenance for Edge AI resources including software, data, and hardware systems.
approximate computing	Techniques to trade off computation quality with effort expended, such as energy consumption in Edge AI.

B. Scoping

For research community, we highlight three dimensions for scoping sustainable Edge AI, including **energy**, **resources**, and **ecosystem**, as indicated in Figure 3.

- **Energy**: this dimension of sustainable Edge AI considers the optimization for energy demand and supply. At the moment, growth in energy demand due to the growth in network traffic and processing requirements is not sustainable. On the supply side, we shall explore the optimal use of renewable energy to further reduce the carbon foot print of networks and processing. Even compared with cloud computing, which has been steadily improving its energy efficiency over the past decade, the Edge AI implementation in a decentralized setting can be a hard challenge, especially it is becoming large scale, albeit its latency benefit [37].
- **Resources**: this dimension concerns the life cycle of physical and digital resources of Edge AI. Besides the energy dimension, little attention has been given to the embodied costs in the manufacture the networking and the processing equipment and the materials used in Edge AI. This sustainability dimension also considers the resource cost needed to dispose or replace the networking and processing equipment.

- **Ecosystem**: this dimension considers the long-term development of all major actors and stakeholders in the Edge AI ecosystem, including standards, business models, and governance. One factor outside the technology circle to stress is the governance where supports/restrictions from government can equally affect the development, control and maintenance of the Edge AI ecosystem.

C. Technology and Societal Impact

The power of Edge AI by using advanced neural networks implies its process is different from applying simple formulas but instead through complicated and energy & resource demanding ones. Since many existing models such as Deep Learning (DL) are often overparameterized for being more flexible, executing DL on Edge (a popular trend) could lead to more computation and hence more energy consumption to match the performance of expert models [36].

As an important goal of sustainability, the energy consumption of Edge AI needs to be optimized. The energy efficiency is crucial for Edge AI embedded infrastructures (e.g., road side units, micro base stations) to sustainably support advanced autonomous driving and Extended Reality (XR) services in the years to come.

Through the pipeline of data acquisition, transfer, computation, and storage, there exists the possibility (e.g., approximate computing) for Edge AI to trade off accuracy to less power and less time consumed. For instance, noisy inputs from numerous sensors can be selectively processed and transferred in order to save energy.

Edge AI can have a considerable societal impact by disrupting the current way of working. Yet, the societal impact is hard to determine in advance, as humans often enact technology within an institutional context. Besides potential technology disruption, the adoption of Edge AI brings policy and governance challenges, including the business model of the operators and the role of the governments, which might be different per country or region. There are a number of societal issues, including the control of the data, access to the intelligence and federated learning, security (including fraud and misuse) [38], [39]. Companies might opt to strengthen their ownership of the data, instead of a shift towards more privacy. They might include this in their terms of use for the future Edge AI services.

Control of the data is one of the main issues in Edge AI as data is collected and processed in a decentralized manner. Data privacy is nowadays regulated at the personal data levels. Using federated computing data can be processed without violating privacy. This might create a new business model that those who operate the system or steward the data might exploit. Not surprisingly, leading players often reinforce their position and use technology to strengthen their power position and revenue model and prevent new players from entering the market that might disrupt their position. The investigation of new business models and changes in the changing and dynamic landscape provides ample revenues for further research.

Another issue is the access to the infrastructure. Will the infrastructure be open for everyone? Or might the type of services be dependent on the type of hardware (high/low end) used? Inclusion might be affected, as some users might have no access, others merely access to basic facilities and the happy few have access to very advanced facilities. This digital access might influence the physical world, as, for example, advanced Edge AI might enable to drive faster and to overtake in the physical world. A societal research question is if this will be acceptable by the society.

A third issue concerns ensuring the continuous and uninterrupted operation of the Edge AI. Like any infrastructure users and companies expect that the infrastructure is secure and available. The high dependency among the components makes it crucial to have clear expectations and the use of standards. Furthermore, procedures for restarting are needed. Also in case of security breaches, swift cooperation among the operating parties might be needed. The infrastructure should ensure that hacks and breaches are detected and measures should be in place to respond and to resolve the issues.

From ecosystem angle, governments might further pose govern the development, control and maintenance of the Edge AI. Regulatory functions consist of three main groups, 1) the architecture design; 2) establishing rules governing; and 3) the interactions between users [40]. In addition, government might pose requirements on the architectural design, like the ability to update the Edge devices to ensure security or the ability to ensure that Edge devices are energy efficient to contribute to sustainability. Regulation to pose requirements on the ability to update and keep connected devices secure for a minimum period of time are already being discussed in the EU already. Also governing rules concerning what users can expect from the Edge AI providers might be established, including the way is deals with data and what is expected when a system is hacked or fraud conducted. Finally, the interactions among users might be addressed, such as how interaction passing a countries boundary is deal with or when users want to move to another Edge AI provider.

Takeaway

Sustainable Edge AI is crucial for the Edge ecosystem that spans across energy, resource and sociotechical dimensions. Key enabling elements include power saving techniques, circular economy, and sociotechnical co-design.

IV. Research Agenda

The core of the research agenda is towards multidisciplinary collaboration given the complexity resulted from the trustworthy and sustainable concerns (Section II and III). It is necessary for Edge AI engineers and researchers to team up with social science, ethics, energy and public sectors to make Edge AI both trustworthy and sustainable.

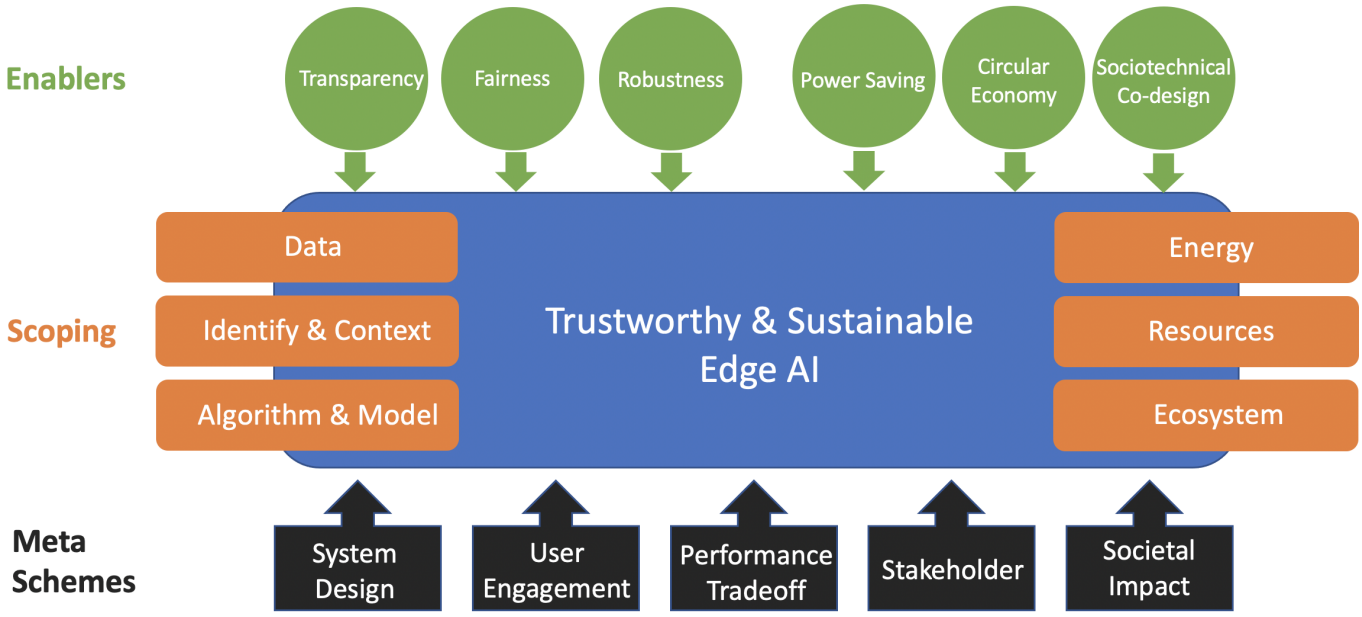

Fig. 4. Overview of Edge AI Meta Schemes.

Meta Schemes for Trustworthy & Sustainable Edge AI

We advocate 5 meta schemes for trustworthy and sustainable Edge AI, as illustrated in Figure 4, including the realm of **system design**, **user engagement**, **performance tradeoff**, **stakeholder**, and **societal impact** that deserve further investigations by the communities.

1) The system design scheme is to investigate how to embed trustworthy and sustainable concerns coherently into the software and hardware co-design for Edge AI. Given the latency benefit of Edge AI, we also need to explore the lightweight by design [41] to match the resource and battery limits on edge devices.
2) The user engagement scheme concerns how to engage and involve user-centric design given the growing complexity to deal with user data privacy, usability, user acceptance, fairness, explainability, transparency in Edge AI services. This scheme is not merely about end customers but shall embrace all relevant actors that rely on Edge AI such governing bodies, developers, data scientists [18], [23].
3) The performance trade-off scheme is centered on which performance metric to trade for sustainability and trustworthiness and what are the qualities to consider beyond functional features. For instance, the power saving target has to be balanced with privacy and security considerations as privacy and security enforcement all entail energy overhead and performance impact (e.g., delay).
4) The stakeholder scheme considers how to maintain and include stakeholders to build a healthy ecosystem for the long run. We hence need to answer who shall be involved, who has been ignored at the moment? Can a holistic infrastructure design to accommodate multi-stakeholders e.g., cloud and cellular operators, and how to promote the convergence between standardization bodies and research communities [42].
5) The societal impact scheme considers how to enable technology and organizational co-evolving within the existing legal framework and governance? How to align with key values for societal impact? How to ensure inclusion and access control of data and secure and continuous operations? What kinds of regulations and governance are needed for ensuring these values? What are the qualities to consider beyond functional features? How to educate and train future engineers and scientists? From a critical sense, the answers are not clear and might be different than anticipated. For instance, in making Edge AI trustworthy and sustainable, we still lack of widely perceived metric defined to measure the various attributes like robustness, fairness, and sustainability. The challenge remains that these terms are subjective and dependent on user/application context. The societal impact is hence hard to justify.

As one example, Fountain's technology enactments framework (TEF) can be used to the of technology by investigating the organizational structure and institutional arrangements [43]. The technology of Edge AI, as a decentralized approach, is shaped by human beings who are operating in different operational and institutional context. This can explain why different companies use the Edge AI in different ways. Furthermore technologies can co-evolve with the organizations that are using Edge AI. As each human has only part of the system in mind, unintended consequences might be generated. For example, what looks like a good idea to increase the

revenue by a company might result in the exclusion of parts of the society, and hence affects the trust and sustainability. Governments role is to warrant those societal values. Therefore the whole ecosystem made up of many heterogeneous stakeholders should be considered to understand the societal implications.

Takeaway

To shape the research agenda of trustworthy and sustainable Edge AI, we need to focus on five meta schemes including system design, user engagement, performance tradeoff, stakeholder, and societal impact.

V. New Initiatives and Outlook

We highlight two large EU projects to explore the agenda development on trustworthy and sustainable edge AI. To inspire future research, we share an outlook on the emerging directions and potential impacts.

A. Project SPATIAL for Trustworthy Edge AI

SPATIAL project (Security and Privacy Accountable Technology Innovations, Algorithms, and Machine Learning) [44] is an EU H2020 funded initiative that aims to develop resilient accountable metrics, privacy-preserving methods, verification tools and system framework for achieving trustworthy AI in security solutions. The project consortium consists of 12 partners from 8 EU member states, and led by TU Delft in the Netherlands. The three-year project of circa 5 million EUR has started in Autumn 2021 to run till 2024.

Ambition of SPATIAL project is to strengthen Europe's ambition of a human-centric approach to AI by taking a holistic approach towards technology development. As illustrated in Figure 5, the project is in line with three key pillars of trustworthy AI identified by EC: lawful, ethical, and robust [21]. Given growing complexity, it is clear that each of the three elements alone is insufficient to make AI fulfill both individual and collective trust. The challenge has motivated SPATIAL project to develop tangible measures that can connect and embed those three pillars into trustworthy AI, specifically focusing on privacy, accountability and resilience, which contribute to the **Transparency** and **Robustness** as described in Figure reffig:trust-scope.

Trustworthy Edge AI Gaps that are tackled in SPATIAL project include: 1) Data issues of AI concerning bias, privacy and data poisoning, since AI model training such as supervised machine learning typically requires large amounts of data, so the quality of data raises an important question on how data is influencing the behavior of the AI systems. 2) Opaque-box AI, where transparency (e.g., explainability, traceability) is missing from current security solutions that adopt AI. In particular, existing ML techniques are opaque for decision makers of private and public sectors in terms of understanding how the systems are making decisions.

Agenda Development: SPATIAL project intends to cover four meta themes (Section IV) including: system design, user engagement, performance trade-off, and societal impact. On system design, SPATIAL intends to develop dedicated XAI mechanisms to ensure transparency in Edge AI distributed context. On user engagement, the project aims to develop communication framework that enables accountable and transparent understanding of AI applications for users, software developers and security service providers. In addition, the project fills the gap by defining dedicated metrics that can quantify the level of explainability to the relevant users or stakeholders. For performance trade-off in Edge AI context, the metrics defined by SPATIAL project can enhance the understanding of what attackers can achieve and with what resources and capabilities. Therefore, developers and service providers will be in a better position to make optimal choices when facing trade-offs and to give relevant evidence of their systems' resilience properties to customers, standardisation bodies and regulators. A special focus of SPATIAL is on the trade-offs between training parameters of deep learning, data quality and data privacy. With such insight, we can tune the deployment for the most significant parameters that influence the overall behavior. Concerning the societal impact, SPATIAL project will develop requirements in a way as to constitute concrete tangible inputs for regulation authorities, governance, standardization and certification bodies. In addition, the project will generate education modules that can reorient AI engineering knowledge within and beyond the cybersecurity domain towards ethical practices. In particular, SPATIAL project is an endeavour in Edge AI to raise awareness that non-functional values such legal ethics values are at least as important as economic profits, as explored in software industry [45].

Fig. 5. SPATIAL Project on Trustworthy AI [44]

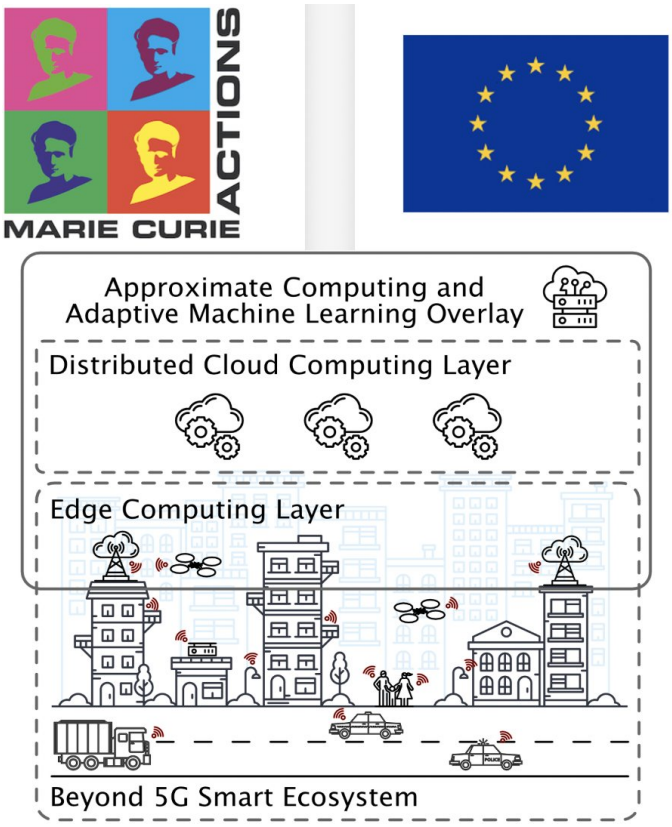

Fig. 6. APROPOS Project for Sustainable Edge AI [46]

B. Project APROPOS for Sustainable Edge AI

APROPOS project for Approximate Computing for Power and Energy Optimisation [46] is a EU funded initiative of the Marie Skłodowska-Curie Innovative Training Networks (ITN). The project aims to tackle the challenges of energy efficiency in future embedded and high-performance computing. During the 4-year project duration 2020-2024, APROPOS will train 15 PhD researchers across 9 European countries in energy-accuracy trade-offs on circuit, architecture, software and system-level solutions. The project creates a well-balanced consortium with 12 industrial companies and 12 academic institutes.

Ambition of APROPOS project is to toward a more sustainable and greener computing by decreasing energy consumption in both distributed computing and communications for cloud-based cyber-physical systems. As highlighted in Figure 6, the project is exploring the adaptive Approximate Computing to optimize energy-accuracy trade-offs in the context of 5G-beyond ecosystem supported by distributed cloud and edge AI.

Sustainable Edge AI Gaps tackled by APROPOS include: 1) Energy saving mechanisms not yet inline with the growing energy consumption by AI computing and network communications. For instance, although not yet a contributor to energy consumption in the global scale, the battery-operated IoT devices are the most dependable on low energy instrument. With powerful edge devices equipped with accelerators such as GPU being deployed, there is strong call for dedicated energy optimization for edge and IoT. 2) Lack of hardware/software co-design strategy that covers every aspect across the computing stack in Edge AI. The key is to effectively translate future applications' characteristics and user behaviors into system level parameters for both training and execution.

Agenda Development: APROPOS tackles the meta schemes of system design, performance trade-off, ecosystem, and societal impact. In specific, the project will build system software to support intelligent resource allocation for energy efficiency. On the performance dimension, APROPOS aims to design and implement hardware modules to enable accuracy-performance-energy trade-offs by embedding and formalizing algorithmic level error tolerance. From the ecosystem perspective, the project is to enhance European industry's competences in an emerging area, by cultivating collaboration between industrial and academia through this joint initiatives. On the societal impact, APROPOS will raise the awareness of energy issues (greener society values) and bring economic impact because of energy savings. Besides improving energy efficiency of cyber-physical infrastructure, it can also create ecological impact (win-win for the economy and nature) since natural resources for electricity production can be saved (in part) because of the APROPOS project.

C. Outlook and Concluding Remarks

As industry is swiftly picking up the potential business and service models of Edge AI [4], [7], we expect more novel development in terms of trustworthy and sustainable aspects that can facilitate building the ecosystem of Edge AI. Given the new commitments made by computing industry due to the financial pressures on climate risk and expectations from investors, peers, and customers, the sustainable development of Edge AI is in line with the general trend [47].

Due to a lack of metrics and evaluation frameworks that are appropriate in trustworthy and sustainable Edge AI, little is known about how performance varies across application, and which non-functional factors can influence performance variation. In this regard, evaluation toolkit for evaluating Edge AI in terms of trustworthiness and sustainability will be of high demand.

Compared with prior research on model-centric approaches for better trustworthy and energy efficiency, we advocate that a data-centric view of Edge AI and call for future studies on the overall Edge AI data pipeline (e.g., pre-processing parameters, training data offloading). Based on recent results [13], the data-centric perspective will play a more important role in balancing the various metrics to achieve trustworthy and sustainable Edge AI.

By identifying key concepts, clarifying the scope, and introducing five meta schemes, this work is our endeavour to contribute to the growing body of knowledge in Edge AI and serve as a call for more research to make Edge AI trustworthy and sustainable.

ACKNOWLEDGEMENTS

This work is supported by SPATIAL project funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101021808, and APROPOS project under the Marie Skłodowska-Curie grant agreement No. 956090.

REFERENCES

[1] T. Rausch and S. Dustdar, "Edge intelligence: The convergence of humans, things, and ai," in *2019 IEEE International Conference on Cloud Engineering (IC2E)*, 2019, pp. 86–96.

[2] S. Deng, H. Zhao, W. Fang, J. Yin, S. Dustdar, and A. Y. Zomaya, "Edge intelligence: The confluence of edge computing and artificial intelligence," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 7457–7469, 2020.

[3] E. Peltonen, M. Bennis, M. Capobianco, M. Debbah, A. Ding, F. Gil-Castiñeira, M. Jurmu, T. Karvonen, M. Kelanti, T. Kliks, Adrian Leppänen, L. Lovén, T. Mikkonen, A. Rao, S. Samarakoon, P. Seppänen, Kari Sroka, S. Tarkoma, and T. Yang, "6G White Paper on Edge Intelligence," *6G RESEARCH VISIONS, NO. 8*, 2020.

[4] D. Xu, T. Li, Y. Li, X. Su, S. Tarkoma, T. Jiang, J. Crowcroft, and P. Hui, "Edge intelligence: Empowering intelligence to the edge of network," *Proceedings of the IEEE*, vol. 109, no. 11, pp. 1778–1837, 2021.

[5] D. D. Clark and et al., "Making the world (of communications) a different place," *SIGCOMM Comput. Commun. Rev.*, vol. 35, no. 3, p. 91–96, Jul. 2005.

[6] B. Varghese, E. de Lara, A. Y. Ding, C.-H. Hong, F. Bonomi, S. Dustdar, P. Harvey, P. Hewkin, W. Shi, M. Thiele, and P. Willis, "Revisiting the arguments for edge computing research," *IEEE Internet Computing*, vol. 25, no. 5, pp. 36–42, 2021.

[7] "AWS CEO Admits Some Workloads Will 'Never Move' To Cloud," 2021. [Online]. Available: https://www.crn.com/news/data-center/aws-ceo-admits-some-workloads-will-never-move-to-cloud

[8] Gartner, "Top 10 Trends in Data and Analytics, 2021," 2021.

[9] W. Shi and S. Dustdar, "The promise of edge computing," *Computer*, vol. 49, no. 5, pp. 78–81, 2016.

[10] M. Satyanarayanan, "The emergence of edge computing," *Computer*, vol. 50, no. 1, pp. 30–39, 2017.

[11] M. Satyanarayanan and N. Davies, "Augmenting cognition through edge computing," *Computer*, vol. 52, no. 7, pp. 37–46, 2019.

[12] W. Toussaint and A. Y. Ding, "SVEva Fair: A Framework for Evaluating Fairness in Speaker Verification," *CoRR*, vol. abs/2107.12049, 2021. [Online]. Available: https://arxiv.org/abs/2107.12049

[13] W. Toussaint, A. Mathur, A. Y. Ding, and F. Kawsar, "Characterising the role of pre-processing parameters in audio-based embedded machine learning," in *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems (SenSys '21)*. ACM, 2021, p. 439–445.

[14] M. Levi and L. Stoker, "Political trust and trustworthiness," *Annual Review of Political Science*, vol. 3, no. 1, pp. 475–507, 2000.

[15] L. Carter and F. Belanger, "The utilization of e-government services: citizen trust, innovation and acceptance factors," *Information Systems Journal*, vol. 15, no. 1, pp. 5–25, 2005.

[16] J. A. Colquitt, B. A. Scott, and J. A. LePine, "Trust, trustworthiness, and trust propensity: a meta-analytic test of their unique relationships with risk taking and job performance," *The Journal of applied psychology*, vol. 92, no. 4, 2007.

[17] M. Janssen, M. Hartog, R. Matheus, A. Y. Ding, and G. Kuk, "Will algorithms blind people? the effect of explainable ai and decision-makers' experience on ai-supported decision-making in government," *Social Science Computer Review*, December 2020.

[18] W. Toussaint and A. Y. Ding, "Machine learning systems in the iot: Trustworthiness trade-offs for edge intelligence," in *2020 IEEE Second International Conference on Cognitive Machine Intelligence (CogMI)*, 2020, pp. 177–184.

[19] CPS Public Working Group, "Framework for Cyber-Physical Systems : Volume 1 , Overview," National Institute of Standards and Technology, Tech. Rep., 2017.

[20] N. C. Roberts, "Keeping public officials accountable through dialogue: Resolving the accountability paradox," *Public Administration Review*, vol. 62, no. 6, pp. 658–669, 2002.

[21] AIHLEG, "Ethics Guidelines for Trustworthy AI," 2019.

[22] M. Janssen, R. Matheus, J. Longo, and V. Weerakkody, "Transparency-by-design as a foundation for open government," *Transforming Government: People, Process and Policy*, vol. 11, no. 1, pp. 2–8, 2017.

[23] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, "A survey on bias and fairness in machine learning," *ACM Comput. Surv.*, vol. 54, no. 6, 2021. [Online]. Available: https://doi.org/10.1145/3457607

[24] L. H. Gilpin, D. Bau, B. Z. Yuan, A. Bajwa, M. Specter, and L. Kagal, "Explaining explanations: An overview of interpretability of machine learning," in *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*, 2018, pp. 80–89.

[25] M. Haus, M. Waqas, A. Y. Ding, Y. Li, S. Tarkoma, and J. Ott, "Security and privacy in device-to-device (d2d) communication: A review," *IEEE Communications Surveys Tutorials*, vol. 19, no. 2, pp. 1054–1079, 2017.

[26] EPRS, "A governance framework for algorithmic accountability and transparency," 2019.

[27] I. Stoica, D. Song, R. A. Popa, D. Patterson, M. W. Mahoney, R. Katz, A. D. Joseph, M. Jordan, J. M. Hellerstein, J. E. Gonzalez, K. Goldberg, A. Ghodsi, D. Culler, and P. Abbeel, "A Berkeley View of Systems Challenges for AI," 2017. [Online]. Available: http://arxiv.org/abs/1712.05855

[28] J. Zhang, B. Chen, Y. Zhao, X. Cheng, and F. Hu, "Data security and privacy-preserving in edge computing paradigm: Survey and open issues," *IEEE Access*, vol. 6, pp. 18 209–18 237, 2018.

[29] L. Yu, L. Liu, C. Pu, M. E. Gursoy, and S. Truex, "Differentially private model publishing for deep learning," in *2019 IEEE Symposium on Security and Privacy (SP)*, 2019, pp. 332–349.

[30] A. Y. Ding, B. Han, Y. Xiao, P. Hui, A. Srinivasan, M. Kojo, and S. Tarkoma, "Enabling energy-aware collaborative mobile data offloading for smartphones," in *IEEE International Conference on Sensing, Communications and Networking (SECON)*, 2013, pp. 487–495.

[31] S. Tarkoma, M. Siekkinen, E. Lagerspetz, and Y. Xiao, *Smartphone energy consumption: modeling and optimization*. Cambridge University Press, 2014.

[32] S. Mittal, "A survey of techniques for approximate computing," *ACM Comput. Surv.*, vol. 48, no. 4, 2016.

[33] W. R. Stahel, "The circular economy," *Nature*, vol. 531, 2016.

[34] M. Geissdoerfer, P. Savaget, N. M. Bocken, and E. J. Hultink, "The circular economy – a new sustainability paradigm?" *Journal of Cleaner Production*, vol. 143, pp. 757–768, 2017.

[35] M. Haus, A. Y. Ding, and J. Ott, "Managing iot at the edge: The case for ble beacons," in *Proceedings of 3rd Workshop on Experiences with the Design and Implementation of Smart Objects*, ser. MobiCom SMARTOBJECTS '17. New York, NY, USA: ACM, 2017, pp. 41–46. [Online]. Available: http://doi.acm.org/10.1145/3127502.3127510

[36] N. C. Thompson, K. Greenewald, K. Lee, and G. F. Manso, "Deep learning's diminishing returns: The cost of improvement is becoming unsustainable," *IEEE Spectrum*, vol. 58, no. 10, pp. 50–55, 2021.

[37] X. Chen, H. Song, J. Jiang, C. Ruan, C. Li, S. Wang, G. Zhang, R. Cheng, and H. Cui, "Achieving low tail-latency and high scalability for serializable transactions in edge computing," in *Proceedings of the Sixteenth European Conference on Computer Systems*, ser. EuroSys '21. Association for Computing Machinery, 2021, p. 210–227.

[38] I. Hafeez, M. Antikainen, A. Y. Ding, and S. Tarkoma, "Iot-keeper: Detecting malicious iot network activity using online traffic analysis at the edge," *IEEE Transactions on Network and Service Management*, vol. 17, no. 1, pp. 45–59, 2020.

[39] A. Y. Ding, "Mec and cloud security," in *Wiley 5G Ref*. John Wiley and Sons, 2020.

[40] F. Di Porto and M. Zuppetta, "Co-regulating algorithmic disclosure for digital platforms," *Policy and Society*, vol. 40, no. 2, pp. 272–293, 2020.

[41] R. Morabito, V. Cozzolino, A. Y. Ding, N. Beijar, and J. Ott, "Consolidate iot edge computing with lightweight virtualization," *IEEE Network*, vol. 32, no. 1, pp. 102–111, Jan 2018.

[42] A. Y. Ding, J. Korhonen, T. Savolainen, M. Kojo, J. Ott, S. Tarkoma, and J. Crowcroft, "Bridging the gap between internet standardization and networking research," *ACM SIGCOMM CCR*, vol. 44, no. 1, p. 56–62, 2014.

[43] J. Fountain, "Bureaucratic reform and e-government in the united states: An institutional perspective," *Routledge handbook of Internet politics*, pp. 115–129, 2008.

[44] "EU H2020 SPATIAL Project," 2021. [Online]. Available: https://cordis.europa.eu/project/id/101021808

[45] V. del Bianco, L. Lavazza, S. Morasca, and D. Taibi, "A survey on open source software trustworthiness," *IEEE Software*, vol. 28, no. 5, pp. 67–75, 2011.

[46] "EU H2020 APROPOS Project," 2021. [Online]. Available: https://cordis.europa.eu/project/id/956090

[47] A. A. Chien, "Good, better, best: How sustainable should computing be?" *Communications of the ACM*, vol. 64, no. 12, 2021.