IEEE Vehicular Technology Magazine, Volume 15, Issue 4, December 2020, pp. 112-121
DOI:10.1109/MVT.2020.3017181

Trustworthy Deep Learning in 6G Enabled Mass Autonomy: from Concept to Quality-of-Trust KPIs

Chen Li, Weisi Guo, Schyler C. Sun, Saba Al-Rubaye, Antonios Tsourdous

***Abstract*—Mass autonomy promises to revolutionise a wide range of engineering, service, and mobility industries. Coordinating complex communication between hyper-dense autonomous agents requires new artificial intelligence (AI) enabled orchestration of wireless communication services in beyond fifth generation (5G) and sixth generation (6G) mobile networks. In particular, safety and mission critical tasks will legally require both transparent AI decision processes, and quantifiable Quality-of-Trust (QoT) metrics for a range of human end-users (consumer, engineer, legal). We outline the concept of trustworthy autonomy for 6G, including the essential elements such as how Explainable AI (XAI) can generate the qualitative and quantitative modalities of trust. We also provide XAI test protocols for integration with radio resource management and associated key performance indicators (KPIs) for trust. The research directions proposed will enable researchers to start testing existing AI optimisation algorithms and develop new ones with the view that trust and transparency should be built in from the design through to the testing phase.**

***Index Terms*—Machine Learning; Deep Learning; Trust; XAI; 6G; Mass Autonomy**

I. Introduction

As 5G networks roll out across the world, researchers are sewing the seeds for the ideas and technologies that will shape future 6G mobile networks. 6G networks are likely to be increasingly integrated with hyper-dense mass autonomy, where intelligent agents (from mechanical robots to data analytic engines) are complementing and supplementing human labour across a diverse range of local industrial, commercial, agricultural, and mobility services. Internet-of-Autonomous-Things (IoAT) will require highly tactile and robust wireless communication channels. This will increase the network complexity in safety critical operations, and therefore Quality-of-Trust (QoT) requirements are necessary from legal, safety, and ethical reasons. Although AI is anticipated to be an enabler - to optimize high dimensional network resource management from the bottom to top layer [1], many deep learning algorithms' low transparency in processing logic leads to difficulties in exploring model transparency and uncertainty. In turn, this risks undermining the human trust in AI-empowered services, and slow down their ubiquitous adoption in real systems. Here, we attempt to describe both a technology agnostic approach for 6G - adding a trust brokerage, but also give wireless specific examples to aid understanding.

The legal requirements for an all data-driven autonomy requires decisions to be explainable to human beings to enable transparency and pave the way for legal responsibility. After all, communication channels are increasingly responsible for *safety-critical* tasks such as autonomous driving, remote surgery, and manufacturing. Legally, the General Data Protection Regulation (GDPR) in EU propose "right to explanation" that request machine learning models provide reasoning through dyadic statements. As such, AI orchestration of resources (communication, computing, storage) in 5G beyond and 6G will need to offer QoT in additional to the current Quality-of-Service (QoS) and Quality-of-Experience (QoE) targets. As we expand our use of autonomous systems, trust and the associated KPIs to measure it will become increasingly important.

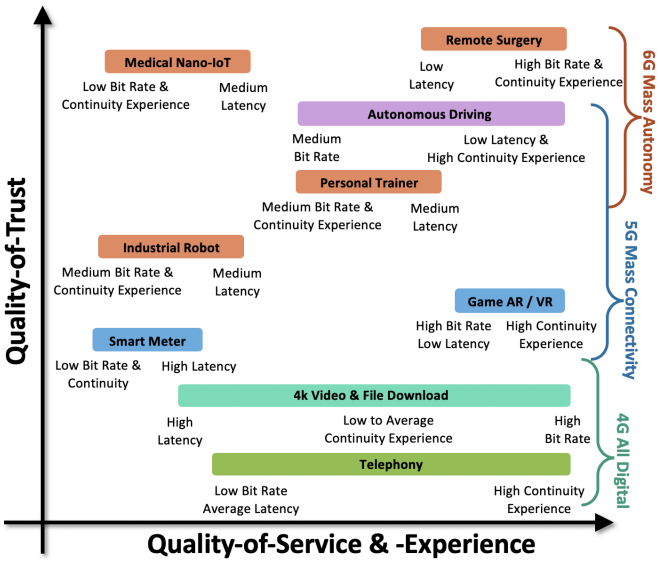

Fig. 1. Relationship Between the Proposed Quality-of-Trust (QoT) with Traditional Motions of QoS and QoE for Diverse Wireless Services.

A. Trust of XAI in 6G Autonomy

Orchestration of diverse service requirements in 5G and beyond has led to the proposed adaptation of deep learning optimisation approaches to overcome growing complexity. For example, in the PHY layer, deep learning's high-dimensional ability to achieve effective non-linear channel equalisation can enable new levels of QoS in highly complex scatter rich channels without channel state information (CSI) [2]. In the MAC layer, deep Q-network (DQN) is used to optimise a variety of high-dimensional dynamic RRM challenges including unmanned aerial vehicle (UAV) relay joint navigation and

Chen Li, Weisi Guo, Schyler C. Sun, Saba Al-Rubaye, and Antonios Tsourdous are with Cranfield University, Bedford, United Kingdom. Weisi Guo is also with the Alan Turing Institute, London, United Kingdom, and *Corresponding Author: wguo@turing.ac.uk.

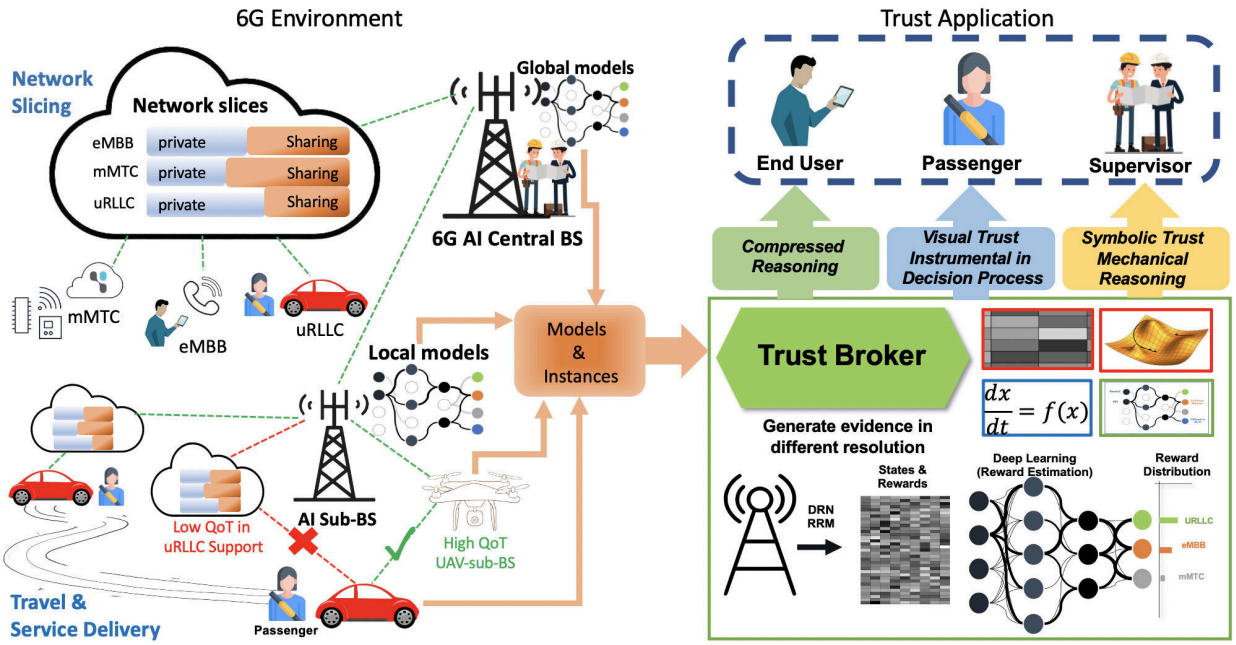

Fig. 2. 6G Network Slicing and Trust Broker for Different Applications and End User Stakeholders. The Trust Broker Translates AI Algorithms into Explainable Outputs.

communication [3]. Deep Learning (DL) can in many complex cases improve performance compared to classic approaches (DSP, SVM, Bayesian inference), especially in the absence of explicitly accurate models. However, the lack of transparency in its reasoning yields a lack of human trust. As such, whilst the design logic of DL and Deep Reinforcement Learning (DRL) is clear, the data features' propagation and the logical reasoning processes are not.

Our increased demand for mass autonomous prompt the requirement of trust metrics, such as our proposed Quality of Trust (QoT). As shown in Fig. 1, there is increased emphasis from new services on high trust - ranging from remote surgery (high trust, high QoS & QoE) to industrial robotics (medium trust, but low QoS demand). These will sit alongside current telephony and multimedia services that require very little trust, but a large variation in QoS/QoE. To achieve trust of AI wireless resource orchestration, we will propose the need for a trust broker entity in future wireless networks - see Fig. 2. This entity can produce a variety of visual, textual, symbolic explainable outputs, offering reasoning to deep learning actions embedded in the base stations. The reasoning outputs speak to human stakeholders in a variety of applications, ranging from engineering experts to end-users. As such, a range of KPIs and test scenarios should be developed. Considerations should include human psychology and philosophy aspects and a high-quality XAI model should have the ability to clarify itself in human-understandable ways (different modes) based on their purpose.

B. XAI and QoT in Future 6G Network Slicing

Mass autonomy in 6G will demand localized sub-network slicing for diverse and dynamic service demands (incl. trust in safety critical multi-modal actions). Therefore, current Software-Defined Networks (SDN) in 5G will need to adapt its network slicing (NS) to meet rapid multi-modal service requirement transitions in hyper-dense autonomous system environments [4]. AI-empowered 6G is envisaged to grant BSs with *edge intelligence* by embedding high speed, precise and robust AI algorithms to ensure safety critical multi-modal mass autonomy in localized sub-network settings, as shown in Fig. 2. Current 5G network slicing has virtually split the network into different independent slices according to service types (e.g. eMBB). Future AI-empowered slicing in 6G will be more fine-grained at the sub-network level and allocate by different new human-centric requirements (e.g. QoT for safety, ethics). Simultaneously, XAI can explain behaviors of mass control systems in both individual instances and overall policy, combined with system performance as important evidence in continuous trust supervision of 6G services.

C. Current Work, Novelty, & Organisation

In current research, uncertainty propagation in neural networks with different structure is analysed in [5]. Trust in 6G physical security is analysed and defined in [4], but lack analysis of mobile resource management trust. In the work of [6], initial and continuous trust in AI are defined but lack test protocol studies, and a recent EU 'EASA Road Map' defines trustworthiness and risk profiles for aerial autonomous systems, but lack consideration of 6G and trust quantification.

In this article, we focus on reviewing the relevant concepts of trust for 6G RRM automation. We focus on mapping the technical aspects of XAI to the psychological aspects of trust in the context of wireless networks. We develop potential trust test protocols and key performance indicators (KPIs) that map AI architecture, performance, to trustworthiness. Firstly, we introduce deep learning model explainability, an

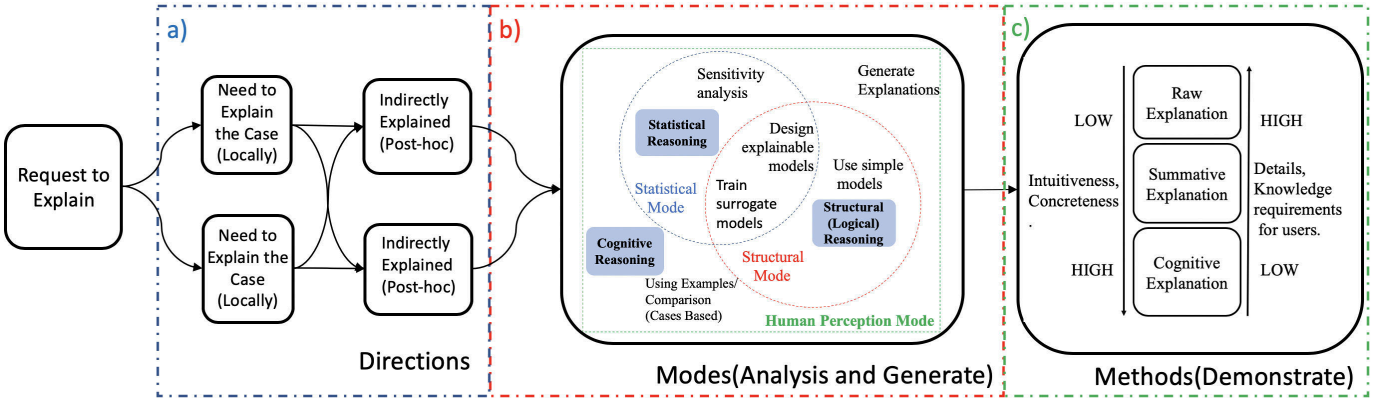

Fig. 3. Mapping between Explainable AI to Trust: Demonstration of a) Directions, b) Modes, c) Methods and Their Relationships

important concept in the trust of AI. Secondly, we articulate the methodological approaches in explainability, spanning different modalities and depth. Thirdly, we design trustworthy KPIs and the test protocols for trust in 6G enabled autonomy, factoring in both quantitative physical trust and qualitative emotional trust.

II. Explainability of AI

Globally implementation of AI in a variety of industries has raised the attention of legal issues regarding both its reliability (e.g., variability and robustness of performance to diverse circumstances) and the provenance of reasoning (e.g., which data features caused which decisions). We define the concepts of research direction, mode of analysis, methods, and KPIs for both explainability and trust in Fig. 3. Explainability is extracted from different characteristics of ML models with multi-level expression of reasoning in statistics, semantics, mathematics, and vision.

A. XAI Research Directions: Integrated and Post-hoc for Local and Global Interpretability

We envisage that a request for explanation maybe demanded either post-event or continuously during reasoning (operations) - see Fig. 3-a). In either case, the request should specify the direction or granularity in which the reasoning needs to be made (e.g., at the local or global reasoning level, and at the post-hoc or integrated level.)

Integrated interpretability directly extracts explanations from the ML model structures and processing logic [7]. The structural complexity of ML models limit current integrated interpretability to only models with low complexity. Some low-dimensional classifiers (e.g. decision-tree or Naïve Bayes based channel state prediction models) could directly be explained by its processing logic. But models with complex structures, like deep neural network (DNN) or support vector machine (SVM) based RRM optimizing models, are hard to be explained due to complex multi-layer connections or hyperplanes.

Post-hoc interpretability twin-system approach proposes to explain the AI model by using high transparency (white-box) systems (like decision trees, linear models) to mimic/mirror the AI model after training. The twin-system can roughly explain black-box models (e.g. DNN, DQN) [7] at the risk of poor approximation or over-fitting to a particular training set.

Local and global explanation indicates the granularity scope. Local explanation examines an individual prediction while global explanation explains the operational logic of the entire ML model behaviour [8]. As in Fig. 2, complex service slice handover decisions between BSs can be explained by local explanations (e.g., the mobility and data demand of one service), but can also be explained by the overall handover policy that governs the process.

B. Different Modes of Generating Explainability

The modes indicate methods to extract and generate the compositions of explainability from learning models, guides the turning of learning model into explainable learning model (Integrated), or the analysed explanation based on the output of models (post-doc) as shown in Fig. 3-b).

1) Structural Mode: **Use simple models:** low complexity in structure can generate logical explanations at the perceptual level to be accepted by users, the backtracking of the decision-making process from decision tree can directly generate explanation. Symbolic classification can also relate to physical laws or well-established optimisation results (e.g. water-filling or channel inversion).

Design explainable models: the machine learning models are reconstructed by interactive sections that process sub-tasks respectively from the overall prediction task, and each section and interaction could be inherent architecturally explainable. Sub-modular Pick Local Interpretable Model-Agnostic Explanations (SP-LIME) [9] generates reliable non-redundant explanations globally by using a set of representative-enough instances from LIME (a surrogate model introduced later) see - Algorithm 2 in Fig. 4; it iterates to greedily find the explainable set by the softmax of instance coverage then let users choose interested cases themselves and track the raw data.

2) Statistical Mode: **Sensitivity analysis:** the gradient of input features respect to a label can give out which part contains the most influential information while doing decision. For instance, Layer-Wise Relevance Propagation (LRP) [10]

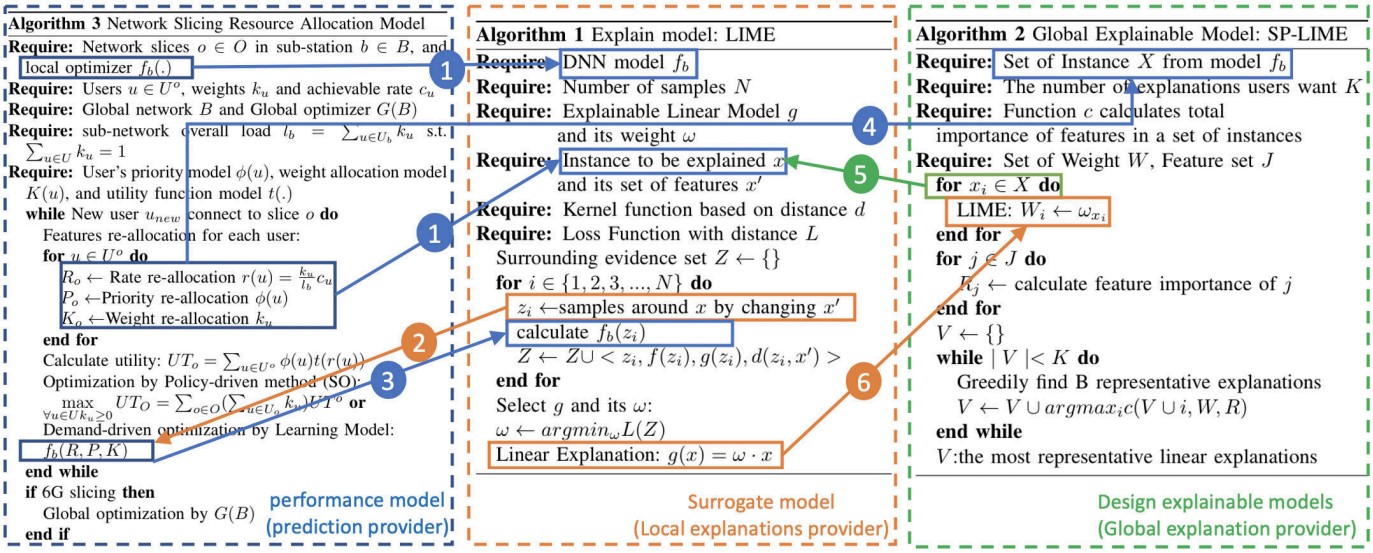

Fig. 4. Demonstration of Surrogate XAI Model: Mapping 6G Optimisation (3) to XAI Local (1) and Global Algorithms (2).

allocates each input with a relevance score to intuitively analyse the contribution of each layer in DNN by the proportion for each neuron output between each layer pair.

Train surrogate models: similar to twin-system, surrogate models [11] ideally use a simple-structured model to fit the output of a complex model. LIME [9] in Fig. 4-Algorithm 1, is a model-agnostic algorithm, explains the complex prediction by approximation of the local case via an interpretable model (e.g. linear regression). Authors in [12] proposed a method to build a soft decision tree created by DNN, to makes hierarchical decisions with the ability to provided better generalizations and robustness to unlabelled data.

3) Human Perception Mode: **Using Examples/ Comparison (Cases Based):** Similar/opposite cases can guarantee users with confidence in the reason why AI models handle the recent problem in a specific way or not. Case-Based Reasoning (CBR) could select relevant similar cases from database by selecting items with minimum distance to give out the reasons for model predictions.

Generate Explanations: Linguistic explanations could be one of the most powerful and intuitive explanations which should extract the interaction and logic from features, generate semantic sentences and cooperate with visualisation methods to demonstrate intuitive explanations. Authors in [13] proposed a method to generate linguistic explanation by using the coupling of visual recognition and text definitions, which generate an understandable high-leveled explanation of the prediction.

C. Methods of Explainability in 6G Mass Autonomy

Previously, modes defined where and how to generate explanations from ML models. Methods guide the way to demonstrate the explanation to human users, based on their individual demands and different knowledge backgrounds. We define 1) experts: sufficient background knowledge, designer for learning logic; 2) trained-users: sufficient background knowledge in specific area, designer of applications; 3) end-users: lack specific background knowledge, user of designed applications. XAI resolutions are divided into 3 methods based on different dimensions of their demonstration: raw explanation for experts, summative explanation for trained-users and cognitive explanation for end-users, which explain the learning model from abstract to concrete as shown in c), Fig. 3, details are listed as follows.

1) Raw Explanation: Data is the most direct, meaningful and detailed explanation in ML models, the raw explanation highlights the features with high contribution to the decision-making, which contains rich unbiased and unmodified raw information (e.g. Trust Broker in Fig. 2 provides activation map, gray-scales and derivatives to supervisors), but lacks expressions in logic (why and how extracted features cooperated). For example, In 6G RRM, underlying data response in the MAC layer extracted and used by ML models could help experts understand the output resource allocation, and dig out problematic channels refers to the features.

2) Summative Explanation: Summative explanation using design explainable models and surrogate models (high transparency 'white box') mentioned above, generates smoothing and fitted explanations based on the processing of statistics from low transparency 'black box'. As Auto-drive tasks in Fig. 2, the decision flow of on-board autonomous model is not visible from Raw Explanation explanations, especially for CNN encoded high dimensional features, while summative explanation could generate a fitted symbolic expression (e.g. Meijer G function) of the DL model to clarify the relationships among inputs in formula expressions. But during the extraction process, meaningful sharp data (such as outliers) could be ignored that add difficulty in data trace to experts.

3) Cognitive Explanation: For end-users who without the ability and interest to comprehend summative explanations, a clearly semantic or visualised explanation will be needed. Simple models (e.g. decision tree), high transparency surrogate models could propose the simple and basic information

needed. Cognitive explanation will conclude the evidence to clarify the prediction in human-understandable linguistic and visualized explanations, but highly concluded explanations ignore some value details and break the information integrity that important for experts and trained-users.

Explainability introduced above could increase the transparency of autonomy and help the development of trust supervision in 6G environment. Take service delivery as an example in Algorithm 3 in Fig. 4. Here, we demonstrate mobility and resource allocation, and how XAI can be used to understand AI reasoning:

- We start with user u under high mobility transfer into another sub-network (Fig. 2). As the mobility request is received, the local sub-BS b will allocate the user to uRLLC network slice. It will then analyze its priority $\phi(u)$ and achievable rate $c(u)$, and allocate the user with individual weight $k(u)$ and re-optimize the resource allocation.

- During the optimization, 5G RRM uses policy-driven optimization that calculates the utility of each slice and find solutions based on different baseline methods like Socially Optimal (SO) and Static Slicing (SS) [14]. This basically meet the requirement in different targets (e.g. SO: maximize overall utility; SS: unilaterally optimize node), but lack the balance of these factors due to modeling clashes. As such, deep learning (f_b) can overcome the clash by finding new high dimensional nonlinear models to replace these explicit policy-driven optimization. As network slicing greatly increased network efficiency, we demonstrate how XAI-modes introduced explain RRM in network slicing in Fig. 4 with the numbered flow steps:

1) To achieve XAI: the sub-BS provide LIME (surrogate model in statistical mode) for f_b and optimisation instance x;
2) LIME then generate i surrounding perturbation instances z_i, and use the model f_b to make predictions based on the set of z_i;
3) f_b send back the predictions and LIME use the predictions to find local approximate linear explanation;
4) To globally explain f_b, SP-LIME request a set of instances X used in performance model f_b;
5) SP-LIME send each instance x into LIME for local explanations;
6) LIME select local explanations to SP-LIME, and SP-LIME greedily find K most representative explanation sets V to globally explain the model f_b.

As such, we have demonstrated both local and global explanations, both of which are important for trust, e.g. local is for understanding specific feature importance in decisions, whereas global is for the overall optimisation balancing between competing demands. Now that we have algorithmic understanding of DL reasoning, we must develop trust metrics to translate between explainability and human perception of trust. As different services have different QoT requirements, trust analytic and supervision detailed in the next chapter can better guide decisions on whether trust requirements are met in a continuous manner.

III. Trust in 6G Enabled Mass Autonomy

"Trust" is a highly abstract conception, indicating the reliability of technology and willingness of users to trust the performance. In order to quantify the trust in 6G, we define *Trust*: $T(m)$ for an explainable DL model m, which is composed by a set of sub-models $M = \{m'_1, m'_2 ... m'_n\}$, calculated by a linear combination of *physical trust* $P(m)$ and *emotional trust* $E(m)$, adjusted by a coefficient α as shown in equation 1. According to the phenomenon proposed in [15] that users may not need explanations from systems with extremely high accuracy, or systems they can not participate in, Trust of highly autonomic models should depend more on physical trust while multi-model-human-interaction models should be allocated more weight on emotion trust than that of physical trust.

$$T(m) = \alpha P(m) + (1-\alpha)E(m) \tag{1}$$

A. Physical Trust of AI Model

$P(m)$ is quantified by a product of model's robustness $R(m)$, accuracy $A(m)$ and explainability, where explainability calculated by division of transparency $\tau(m)$ and complexity $C(m)$ as:

$$\begin{aligned} P(m) &= R(m)\tau(m)\frac{A(m)}{C(m)} \\ &= R(m)\tau(m)\frac{a_m^d(\prod_{m'_n \in M} a_{m'_n})}{\sum_{m'_n \in M} \Omega_{g_n}(\omega(m'_n))/n}, \end{aligned} \tag{2}$$

The parameter $A(m)$ is a combined-accuracy-indicator in $(0, 1)$, which equals to the product of accuracy for each sub-model $a_{m'_n}$ with different functionalities (inner accuracy for fitting and explaining), and overall prediction accuracy a_m (prediction accuracy) powered by an *importance-adjust factor,* d to adjust the importance of prediction accuracy in overall system performance. For example, accuracy of system in Fig. 4 is calculated by both the accuracy of NS resource allocation model and explain model LIME/SP-LIME.

Legal commercial DL model should not use confidential personal data without permitted by users. Transparency $\tau(m)$ is a rate of visible features to all input features and that of sensitive information encryption. The data used in model training may contain personal or confidential information, that affects data privacy in 6G communication, but encryption algorithm (e.g. Hash) could be imported to convert the original data into training set without losing information and will be studied in future research. In [15], Glass et al indicate that the participation of human users can also be seen as part of system transparency, which will be quantified as a part of emotional trust introduced later.

Complexity $C(m)$ is highly dependent on the inner algorithms of models with different structures and processing logic. In formula 2, G is a set of all learning models (DT, NN, DNN, DRL, etc.), g_n is the model type of m'_n, $g \in G$; function ω quantifies the structural complexity for sub-models (for DT, the depth; for DNN, the number of hidden layers); function Ω calculates the complexity indicator for sub-model m', based on its model type g_n; a balance of complexity should

be considered in Ω when multi-models cooperating to make explanations and decisions, we take the average complexity of sub-models in this article.

Assume the complexity $C(m)$ is stable, models with low accuracy in each sub-model will not have the ability to generate high accuracy in overall prediction. Models with high inner accuracy, but low in overall prediction will have lower physical trust. Whilst, if the complexity of the model could be reduced with the same performance, the physical trust will rise to indicate the advancement. Issuers should make improvements based on their prototype project by $argmax(P(m))$ and using the physical trust as an evolutionary indicator.

B. Emotional Trust from Human Experts and End Users

The emotional trust parameter $E(m)$ can not directly be sensed and analysed from the physical structure of model m, but can be sensed from emotion changes collected by brain-machine interactions in future 6G environment [8], or a questionnaire on user experiences. In order to quantify emotional trust, the testing institution needs to organise a test group with q participants $\{t_1, t_2..., t_q\} \in T$. Daily trust baseline indicates the willingness of trust for each individual, and could affect their choice in emotional trust test (emotional changes will make people highly or lowly willing to trust). Continuous testing of the individual baseline is necessary that those with unstable moods should not participate in the emotional trust test [8]. As shown in equation 3, accepted testing data $\gamma(t)$ will be fine-tuned by a factor $l(t)$ based on the willingness gap between the daily baseline and long-term baseline of individual participants.

$$E(m) = \frac{1}{q} \sum_{t \in T} l(t)\gamma(t) \tag{3}$$

The models with soft decision tree are important for visual recognition, which is a critical element in real world autonomous system safety and trust. As such, we envisage that visual data is important in 6G mass autonomy support. We analyze the physical trust of models from [12] (DNN, DT and soft-DT) to demonstrate our framework, with assumptions that the robustness and transparency of these models are the same in Table I. We take function Ω for DT as a linear function $\Omega_{DT}(x) = 1/4x$ that the explainability of DT is linearly influenced by it depth and a exponential function $\Omega_{NN}(x) = 2^x$ for DNN according to the difficulty to open network structures; and *importance-adjust factor* $d = 2$ as demonstration. Please note that these are intended only as proof of principle, the main contribution is the framework itself rather than any specific algorithm or parameter settings.

TABLE I
PERFORMANCE TABLE FOR MODELS IN [12]

Models	Accuracy	DNN	DT	Physical Trust
DT	94.45%	None	Depth:8	0.22302
DNN	96.86%	Hidden Layers:3	None	0.11727
DNN-sDT	99.22%	Hidden Layers:3	Depth:4	0.19689

According to physical trust above, roughly, we consider the pure decision tree as the best model that DNN is too difficult to be explained, although the using of DNN-sDT significantly prompt the accuracy of overall model, its physical trust result influenced by the complexity indicator with the intervention of DNN, but considering robustness, transparency and emotional trust of models in real-use, the conclusion could be different for specific tasks. As for precision machining, high accuracy is the most significant; for large-scale systems like heavy industry, equipment, labor and material dispatch, the overall explainability is important that human supervisor could fine-grained monitor the overall processes and states. Chemical plant and vehicle transport systems, both explainability and precision are needed, and scenarios in this area highly depend on physical trust.

IV. Testing Protocol for 6G Mass Autonomy

With the explainability of learning models guarantee transparency, KPI quantifies the trust of learning models, we proposed a trust testing protocol in Fig. 5 for smart products, both initial trust test (before it launched in real use) and continuous trust test (after implement in real-world environment) should be completed to guarantee security and legality. The rating of learning models should be completed by qualified third-party institutions using a uniformed criterion rather than the product issuer.

Trust Band in dashed box of Fig. 5 is designed to justify which level of trust does the model achieve, layered from high to low, as 'totally trusted'; 'totally trusted with risk'; 'highly trusted'; 'partly trusted'; 'low quality' and 'fail'. Totally trusted level contains models that could directly affect human safety (like auto-break in 6G autonomous vehicles), should achieve high accuracy and none failures while running; in continuous trust testing of totally trusted models, once failure observed by AI supervisor, the model will be layer-downed into 'totally trusted with risk', and the failure case will be reported to human supervisors to decide whether the product needs rebuilding. But in some processing industries, high accuracy is more necessary then trust band (e.g. precision machining), for whom, the trust band could be 'highly trust', with high accuracy but allow low probability of failures.

The newly released product should be analysed by the issuer, and upload the testing request to the third-party institution, with a packet contains: the product, its demo/running data, expected method of explainability and trust band. A group of experts will be organised to define the participation of different layer users in emotional test, based on the explainability methods introduced above, for example, AI-based transport control will need raw explanations in 70% cases, and 30% summative explanations, the test group should be allocated 70% developers and experts, and 30% trained users. By analysing the monitored data from the 6G test environment, whether the smart product is accepted or not will depend on the trust report generated from the physical test result and the emotional trust test result. Once accepted, the model will be implemented into real-world environment and be handed over to continuous trust supervision mentioned earlier; if not

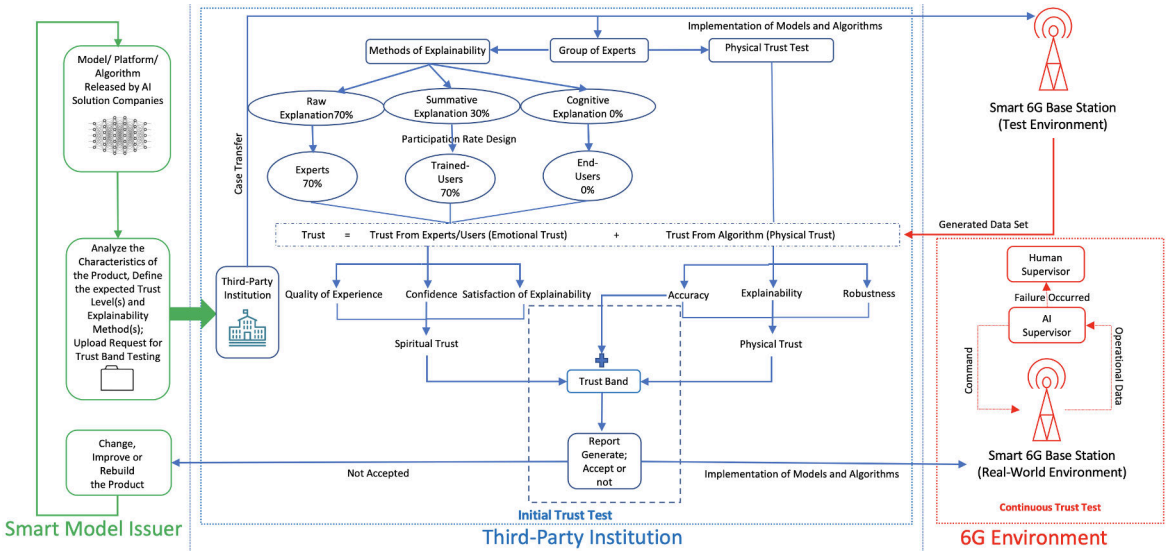

Fig. 5. Trust Testing Flow Chart for New Released AI Models and Algorithms.

accepted, or human supervisors define the model is risked after 'totally trusted' model labeled as 'totally trusted with risk', the model issuer should recall the product, modified it, and re-request for trust testing.

V. Conclusion and Further Research

In this article, we are the first to attempt to quantify trust of AI in a future wireless communication and 6G context, and outline the KPIs and testing protocols to guide its development to work alongside legal frameworks and standards. Here, we do assume a technology agnostic approach for 6G, adding a trust brokerage alongside current and new wireless technologies. The KPI and test protocol guarantee universality that the KPI and testing protocol could be used in all learning models and scenarios. We outline a number of promising local and global XAI methods, ranging from post-hoc explainability to integrated design. Our proposed KPIs factor in both AI model accuracy and complexity, as well as their explainability and human emotional trust.

For future research, the measurement of model complexity ω and Ω in function 2 based on different algorithm structures should be defined at finer scales and these functions need to catch up with the rapid development of AI. The importance of applying brain-machine interactions in emotional trust is significant, and the influence factor $l(t)$ in function 3 needs to be clearly defined based on the long-term emotional trust gap.

References

[1] M. Elsayed and M. Erol-Kantarci, "Ai-enabled future wireless networks: Challenges, opportunities, and open issues," *IEEE Vehicular Technology Magazine*, vol. 14, no. 3, pp. 70–77, Sep. 2019.

[2] K. Burse, R. N. Yadav, and S. Shrivastava, "Channel equalization using neural networks: A review," *IEEE transactions on systems, man, and cybernetics, Part C (Applications and Reviews)*, vol. 40, no. 3, pp. 352–357, 2010.

[3] Z. Du, Y. Deng, W. Guo, A. Nallanathan, and Q. Wu, "Green deep reinforcement learning for radio resource management: Architecture, algorithm compression and challenge," *arXiv preprint arXiv:1910.05054*, 2019.

[4] M. Ylianttila, R. Kantola, A. Gurtov, L. Mucchi, I. Oppermann, Z. Yan, T. H. Nguyen, F. Liu, T. Hewa, M. Liyanage *et al.*, "6g white paper: Research challenges for trust, security and privacy," *arXiv preprint arXiv:2004.11665*, 2020.

[5] C. Li, C. Sun, S. Al-Rubaye, A. Tsourdous, and W. Guo, "Uncertainty propagation in neural network enabled multi-channel optimisation," in *IEEE Vehicular Technology Conference 2020-Spring Recent Results and Workshops*, 2020.

[6] K. Siau and W. Wang, "Building trust in artificial intelligence, machine learning, and robotics," *Cutter Business Technology Journal*, vol. 31, no. 2, pp. 47–53, 2018.

[7] F. K. Došilović, M. Brčić, and N. Hlupić, "Explainable artificial intelligence: A survey," in *2018 41st International convention on information and communication technology, electronics and microelectronics (MIPRO)*. IEEE, 2018, pp. 0210–0215.

[8] W. Guo, "Explainable artificial intelligence (xai) for 6g: Improving trust between human and machine," *arXiv preprint arXiv:1911.04542*, 2019.

[9] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should i trust you?: Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. ACM, 2016, pp. 1135–1144.

[10] A. Shrikumar, P. Greenside, and A. Kundaje, "Learning important features through propagating activation differences," in *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org, 2017, pp. 3145–3153.

[11] E. M. Kenny and M. T. Keane, "Twin-systems to explain artificial neural networks using case-based reasoning: comparative tests of feature-weighting methods in ann-cbr twins for xai," in *Twenty-Eighth International Joint Conferences on Artifical Intelligence (IJCAI), Macao, 10-16 August 2019*, 2019, pp. 2708–2715.

[12] N. Frosst and G. Hinton, "Distilling a neural network into a soft decision tree," *arXiv preprint arXiv:1711.09784*, 2017.

[13] L. A. Hendricks, Z. Akata, M. Rohrbach, J. Donahue, B. Schiele, and T. Darrell, "Generating visual explanations," in *European Conference on Computer Vision*. Springer, 2016, pp. 3–19.

[14] P. Caballero, A. Banchs, G. De Veciana, and X. Costa-Pérez, "Network slicing games: Enabling customization in multi-tenant mobile networks," *IEEE/ACM Transactions on Networking*, vol. 27, no. 2, pp. 662–675, 2019.

[15] A. Glass, D. L. McGuinness, and M. Wolverton, "Toward establishing trust in adaptive agents," in *Proceedings of the 13th international conference on Intelligent user interfaces*. ACM, 2008, pp. 227–236.

2020-09-30

Trustworthy deep learning in 6G-enabled mass autonomy: from concept to quality-of-trust key performance indicators

Li, Chen

IEEE

Chen L, Guo W, Sun SC, et al., (2020) Trustworthy deep learning in 6G-enabled mass autonomy: from concept to quality-of-trust key performance indicators. IEEE Vehicular Technology Magazine, Volume 15, Issue 4, December 2020, pp. 112-121

https://doi.org/10.1109/MVT.2020.3017181