Mass autonomy promises to revolutionize a wide range of engineering, service, and mobility industries. Coordinating complex communication among hyperdense autonomous agents requires new artificial intelligence (AI)-enabled orchestration of wireless communication services beyond 5G and 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, and legal). We outline the concept of trustworthy autonomy for 6G, including 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 optimization algorithms and develop new ones with the view that trust
and transparency should be built in from the design through the testing phase.

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

The legal requirements for an all-data-driven autonomy oblige decisions 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 in the European Union (EU) proposes the “right to explanation,” in which machine learning models provide reasoning through dyadic statements. As such, AI orchestration of resources (communication, computing, and storage) in 5G and beyond and 6G will need to offer QoT in addition to 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.

Trust of XAI in 6G Autonomy
The orchestration of diverse service requirements in 5G and beyond has led to the proposed adaptation of DL optimization approaches to overcome growing complexity. For example, in the physical layer, DL’s high-dimensional ability to achieve effective nonlinear channel equalization can enable new levels of QoS in highly complex, scatter-rich channels without channel state information [2]. In the media access control (MAC) layer, the deep Q-network (DQN) is used to optimize a variety of high-dimensional dynamic radio resource management (RRM) challenges, including unmanned aerial vehicle (UAV) relay joint navigation and communication [3]. DL can, in many complex cases, improve performance compared to classic approaches (digital signal processing, support vector machine [SVM], and 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, while 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 autonomy prompts the requirement of trust metrics, such as our proposed QoT. As exhibited in Figure 1, there is increased emphasis from new services on high trust, ranging from remote surgery (high trust, high QoS, and high 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 in AI wireless resource orchestration, we propose the need for a trust broker entity in future wireless networks (see Figure 2). This entity can produce a variety of visual, textual, and symbolic explainable outputs, offering reasoning for DL actions embedded in the base stations (BSs). 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 purposes.

XAI and QoT in Future 6G Network Slicing
Mass autonomy in 6G will demand localized subnetwork slicing for diverse and dynamic service demands (including trust in safety-critical multimodal actions). Therefore, current software-defined networks in 5G will need to adapt their network slicing to meet rapid multimodal service requirement transitions in hyperdense autonomous system environments [4]. AI-empowered 6G is envisaged to grant BSs edge intelligence by embedding high-speed, precise, and robust AI algorithms to ensure safety-critical multimodal mass autonomy in localized subnetwork settings, as demonstrated in Figure 2.
Current 5G network slicing has virtually split the network into different independent slices according to service types (e.g., enhanced mobile broadband [eMBB]). Future AI-empowered slicing in 6G will be more fine-grained at the subnetwork level and allocate according to different new human-centric requirements (e.g., QoT for safety and ethics). To enable this, XAI can explain behaviors of mass control systems in both individual instances and overall policy, and system reasoning is important evidence in continuous trust supervision of 6G services.

**Current Work, Novelty, and Organization**

In current research, uncertainty propagation in neural networks with different structures is analyzed in [5]. Trust in 6G physical security is analyzed and defined in [4], but it lacks analysis of mobile resource management trust. In the work of [6], initial and continuous trust in AI is defined, but test protocol studies are lacking; a recent EU “EASA Road Map” defines trustworthiness and risk profiles for aerial autonomous systems but lacks 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 KPIs that map AI architecture, performance, and trustworthiness. First, we introduce DL model explainability, an important concept in the trust of AI. Second, we articulate the methodological approaches in explainability, spanning different

---

**Figure 1** The relationship of the proposed QoT with traditional motions of QoS and QoE for diverse wireless services. IoT: Internet of Things; AR: augmented reality; VR: virtual reality.
modalities and depth. Third, we design trustworthy KPIs and the test protocols for trust in 6G-enabled autonomy, factoring in both quantitative physical trust and qualitative emotional trust.

**Explainability of AI**

Globally, the implementation of AI in a variety of industries has raised legal issues regarding both its reliability (e.g., variability and robustness of performance under 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 Figure 3. Explainability is extracted from different characteristics of ML models with multilevel expression of reasoning in statistics, semantics, mathematics, and vision.

**XAI Research Directions: Integrated and Post Hoc for Local and Global Interpretability**

We envisage that a request for explanation may be demanded either postevent or continuously during reasoning (operations); see Figure 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).

*Figure 2* The 6G network slicing and trust broker for different applications and end-user stakeholders. The trust broker translates AI algorithms into explainable outputs. uRLLC: ultrareliable low-latency communication; mMTC: massive machine type communications; DRN: deep reinforcement learning.
The local and global explanation indicates the granularity scope. Local explanation examines an individual prediction, while global explanation describes the operational logic of the entire ML model behavior [8]. As in Figure 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.

Different Modes of Generating Explainability
The modes indicate methods to extract and generate the compositions of explainability from learning models and guide the turning of a learning model into an explainable learning model (integrated) or the analyzed explanation based on the output of models (post hoc), as presented in Figure 3(b).

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

Design explainable models: Machine learning models are reconstructed by interactive sections that process subtasks respectively from the overall prediction task, and each section and interaction could be inherently architecturally explainable. Submodular pick, local interpretable model-agnostic explanations (SP-LIME) [9] generates reliable nonredundant explanations globally using a set of representative-enough instances from LIME (a surrogate model introduced later); see Algorithm 2 in Figure 4, which iterates to greedily find the explainable set by the softmax of instance coverage and then lets users choose interested cases themselves and track the raw data.

Statistical Mode
Sensitivity analysis: The gradient of input features with respect to a label can explain which part contains the most influential information when making a decision. For instance, layer-wise relevance propagation [10] allocates each input with a relevance score to intuitively analyze the contribution of each layer in the DNN by the proportion for each neuron output between each layer pair.

Train surrogate models: Similar to the twin system, surrogate models [11] ideally use a simple-structured model to fit the output of a complex model. LIME [9] in Figure 4, Algorithm 1, is a model-agnostic algorithm and explains the complex prediction by approximation of the local case via an interpretable model (e.g., linear regression). The authors in [12] proposed a method to build a soft DT (sDT) created by the DNN to make hierarchical decisions, with the ability to provide better generalizations of and robustness to unlabeled data.

Human Perception Mode
Using examples/comparison (case based): Similar/opposite cases can guarantee users’ confidence in the reason that AI models handle the recent problem in a specific way. Case-based reasoning could select relevant similar cases from the database by choosing 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 and should extract the interaction and logic from features, generate semantic sentences, and cooperate with visualization methods to demonstrate intuitive explanations. The authors in [13] proposed a method to generate linguistic explanation using the coupling of visual

![Figure 3](image-url) The mapping between XAI and trust: demonstration of (a) directions, (b) modes (analysis and generation), (c) methods (demonstration), and their relationships.
recognition and text definitions, which generates an understandable, high-leveled explanation of the prediction.

**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**, who have sufficient background knowledge and are designers for learning logic; 2) **trained users**, who have sufficient background knowledge in the specific area and are designers of applications; and 3) **end users**, who lack specific background knowledge and are users of designed applications. XAI resolutions are divided into three methods based on different dimensions of their demonstration: raw explanation for experts, summative explanation for trained users, and cognitive explanation for end users, which describes the learning model from abstract to concrete, as shown in Figure 3(c). Details are listed as follows.

**Raw Explanation**

Data are 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 Figure 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 output resource allocation and dig out problematic channels referring to the features.

**Summative Explanation**

Summative explanation using design-explainable models and surrogate models (high-transparency white box), mentioned previously, generates smoothing and fitted explanations based on the processing of statistics from a low-transparency black box. As autodrive tasks in Figure 2, the decision flow of an onboard autonomous model is not visible from raw explanations, especially for convolutional neural network-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. However, during the extraction process, meaningful sharp data (such as outliers) that add difficulty in data tracing to experts could be ignored.

**Cognitive Explanation**

For end users who lack the ability and interest to comprehend summative explanations, a clearly semantic or visualized explanation will be needed. Simple models (e.g., DT) and high-transparency surrogate models could propose the simple and basic information needed. Cognitive explanations will conclude evidence to clarify the prediction in human-understandable linguistic and visualized explanations, but highly concluded explanations ignore some valuable details and break the information integrity important for experts and trained users.

---

**Figure 4** A demonstration of the surrogate XAI model: mapping 6G optimization (3) to XAI local (1) and global (2) algorithms.
**Trust is a Highly Abstract Concept, indicating the Reliability of Technology and the Willingness of Users to Trust the Performance.**

The explainability introduced previously could increase the transparency of autonomy and help the development of trust supervision in the 6G environment. Take service delivery as an example in Figure 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 subnetwork (Figure 2). As the mobility request is received, the local sub-BS $b$ will allocate the user to the ultrareliable low-latency communication (uRLLC) network slice. It will then analyze its priority $\phi(u)$ and achievable rate $c(u)$, allocate the user with individual weight $k(u)$, and reoptimize the resource allocation.

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

1) To achieve XAI, the sub-BS provides LIME (surrogate model in statistical mode) for $f_b$ and optimization instance $x$.

2) LIME then generates $i$ and surrounding perturbation instances $z_i$ and uses the model $f_b$ to make predictions based on the set of $z_i$.

3) $f_b$ sends back the predictions, and LIME uses the predictions to find a local approximate linear explanation.

4) To globally explain $f_b$, SP-LIME requests a set of instances $X$ used in performance model $f_b$.

5) SP-LIME sends each instance $x$ into LIME for local explanations.

6) LIME selects local explanations to SP-LIME, and SP-LIME greedily finds the $K$ most representative explanation sets $V$ to globally explain the model $f_b$.

As such, we have demonstrated 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 optimization balancing between competing demands. Now that we have an 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 analytics and supervision, detailed in the next section, can better guide decisions on whether trust requirements are met in a continuous manner.

**Trust in 6G-Enabled Mass Autonomy**

Trust is a highly abstract concept, indicating the reliability of technology and the willingness of users to trust the performance. To quantify the trust in 6G, we define $\text{Trust} : T(m)$ for an explainable DL model, $m$, which is composed of a set of submodels, $M = \{m_1, m_2, \ldots, m_n\}$, calculated by a linear combination of physical trust, $P(m)$, and emotional trust, $E(m)$, adjusted by a coefficient, $\alpha$:

$$T(m) = aP(m) + (1 - a)E(m). \quad (1)$$

According to the phenomenon proposed in [15], i.e., that users may not need explanations from systems with extremely high accuracy or systems they cannot participate in, trust of highly autonomic models should depend more on physical trust, while multimodel human interaction models should allocate more weight to emotional trust than physical trust.

**Physical Trust of AI Model**

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

$$P(m) = R(m) \tau(m) \frac{A(m)}{C(m)} = R(m) \tau(m) \sum_{m \in M} a_{m_i} \frac{\prod_{m \in M} a_{m_i}}{\Omega_{P,m}(\omega(m_i))/n}. \quad (2)$$

The parameter $A(m)$ is a combined accuracy indicator in $(0, 1)$, which equals the product of accuracy for each sub-model $a_{m_i}$, 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, the accuracy of the system in Figure 4 is calculated by both the accuracy of the network slicing resource allocation model and the explain model LIME/SP-LIME.

The legal commercial DL model should not use confidential personal data without permission by users. Transparency $\tau(m)$ is the rate of visible features to all input features and sensitive information encryption.
The data used in model training may contain personal or confidential information that affects data privacy in 6G communication, but an encryption algorithm (e.g., Hash) could be imported to convert the original data into a 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 (2), $G$ (superset of $g$) is a set of all learning models (DT, NN, DNN, DRL, and so on); $g_n$ is the model type of $m_n$, $g \in G$; function $\omega$ quantifies the structural complexity for submodels (for DT, the depth; for DNN, the number of hidden layers); and function $\Omega$ calculates the complexity indicator for submodel $m'$, based on its model type, $g_n$. A balance of complexity should be considered in $\Omega$ when multimodels are cooperating to make explanations and decisions, so we take the average complexity of submodels in this article.

Assuming the complexity $C(m)$ is stable, models with low accuracy in each submodel will not have the ability to generate high accuracy in overall prediction. Models with high inner accuracy but low overall prediction will have lower physical trust. 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 physical trust as an evolutionary indicator.

**Emotional Trust from Human Experts and End Users**

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

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

**The Daily Trust Baseline Indicates the Willingness to Trust for Each Individual and Could Affect Their Choices in the Emotional Trust Test.**

The models with an sDT 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 are important in 6G mass autonomy support. We analyze the physical trust of models from [12] (DNN, DT, and sDT) to demonstrate our framework, with assumptions that the robustness and transparency of these models are the same in Table 1. We take function $\Omega$ for DT as a linear function, $\Omega_{DT}(x) = 1/4x$, that the explainability of DT is linearly influenced by its depth, an exponential function, $\Omega_{NN}(x) = 2^x$, for DNN according to the difficulty of open network structures, and importance-adjusted factor, $d = 2$, as a demonstration. Please note that these are intended only as a proof of principle, and the main contribution is the framework itself rather than any specific algorithm or parameter settings.

According to physical trust described here, roughly, we consider the pure DT as the best model and that DNN is too difficult to be explained, although the use of DNN-sDT significantly improves the accuracy of the overall model. Its physical trust result is 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, in that a human supervisor could fine-grain monitor the overall processes and states. For chemical plant and vehicle transport systems, both explainability and precision are needed, and scenarios in this area highly depend on physical trust.

**Testing Protocol for 6G Mass Autonomy**

With the explainability of learning models guaranteeing transparency, KPI quantifies the trust of learning models.

**Table 1 Performance table for the 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       |
We propose a trust testing protocol in Figure 5 for smart products, and both an initial trust test (before it launched in real use) and a continuous trust test (after being implemented in the real-world environment) should be completed to guarantee security and legality. The rating of learning models should be completed by qualified third-party institutions, rather than the product issuer, using a uniform criterion.

The trust band in the dashed box of Figure 5 is designed to justify which level of trust the model achieves, layered from high to low, as “totally trusted,” “totally trusted with risk,” “highly trusted,” “partly trusted,” “low quality,” and “fail.” The totally trusted level contains models that could directly affect human safety (like autobrake in 6G autonomous vehicles) and so should achieve high accuracy and no failures while running; in continuous trust testing of totally trusted models, once failure is observed by the AI supervisor, the model will be layered-down into “totally trusted with risk,” and the failure case will be reported to human supervisors to decide whether the product needs rebuilding. However, in some processing industries, high accuracy is more necessary than trust band (e.g., precision machining), for which the trust band could be “highly trust” with high accuracy but allowing a low probability of failures.

The newly released product should be analyzed by the issuer, and the testing request should be uploaded to the third-party institution with a packet that contains the product, its demo/running data, the expected method of explainability, and trust band. A group of experts will be organized to define the participation of different layer users in emotional tests based on the explainability methods introduced previously; for example, AI-based transport control will need raw explanations in 70% of cases with 30% summative explanations, and the test group should be allocated as 70% developers and experts and 30% trained users. By analyzing the monitored data from the 6G test environment, whether or not the smart product is accepted 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 the real-world environment and handed over to continuous trust supervision mentioned earlier; if not accepted or if human supervisors redefine the model, originally defined as “totally trusted,” as “totally trusted with risk,” the model issuer should recall the product, modify it, and rerequest trust testing.

**Conclusions and Further Research**

In this article, we are the first to attempt to quantify the 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 assume a technology-agnostic approach for 6G, adding trust brokerage alongside current and new wireless technologies. The KPIs and test protocol guarantee universality, in that the KPIs 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

---

**Figure 5** The trust testing flowchart for newly released AI models and algorithms.
factor in both AI model accuracy and complexity as well as its explainability and human emotional trust.

For future research, the measurement of model complexity $\omega$ and $\Omega$ in (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 $I(t)$ in (3) needs to be clearly defined based on the long-term emotional trust gap.

**Author Information**

**Chen Li** (c.li.21@cranfield.ac.uk) is pursuing his Ph.D. in aerospace at Cranfield University, United Kingdom, based in the AI Group, Digital Aviation Research and Technology Centre, under the supervision of Prof. Weisi Guo, Prof. Antonios Tsourdos, and Dr. Saba Al-Rubaye. His research topic is swarm drone information relaying–applied AI and trust retaining, and his research interest lies in explainable artificial intelligence (AI), green AI, and AI trust.

**Weisi Guo** (weisi.guo@cranfield.ac.uk) is the chair professor of human machine intelligence at Cranfield University, United Kingdom. He has published more than 130 papers and is the principal investigator on more than $2.3$ million in research grants. His research has won him several international awards. He is a Turing Fellow at the Alan Turing Institute and fellow of the Royal Statistical Society. He is a Senior Member of IEEE. Weisi Guo is the corresponding author for this article.

**Schyler Chengyao Sun** (Schyler.Sun@cranfield.ac.uk) is currently a Ph.D. student at Cranfield University, United Kingdom. His research interests include explainable machine learning algorithms.

**Saba Al-Rubaye** (S.Alrubaye@cranfield.ac.uk) is currently a senior lecturer and leading connectivity and networking researcher in the School of Aerospace, Transport, and Manufacturing at Cranfield University, United Kingdom. Her research interests lie primarily in the area of unmanned aerial vehicle connectivity, networking, artificial intelligence, and autonomous vehicle systems. She is a Chartered Engineer and a member of the Institution of Engineering and Technology. She is a Senior Member of IEEE.

**Antonios Tsourdos** (a.tsourdos@cranfield.ac.uk) is a professor of autonomous systems and control with Cranfield University, United Kingdom. He was appointed head of the Autonomous Systems Group in 2007, head of the Centre for Autonomous and Cyber-Physical Systems in 2012, and director of Research–Aerospace, Transport, and Manufacturing in 2015. He was a member of Team Stellar, the winning team for the U.K. Ministry of Defense Grand Challenge (2008), and the IET Innovation Award (Category Team, 2009).

**References**

[1] M. Elsayed and M. Erol-Kantarci, "AI-enabled future wireless networks: Challenges, opportunities, and open issues," IEEE Veh. Technol. Mag., vol. 14, no. 3, pp. 70–77, Sept. 2019. doi: 10.1109/MVT.2019.2919236.

[2] K. Burke, R. N. Yadav, and S. Shrivastava, “Channel equalization using neural networks: A review,” IEEE Trans. Syst. Man. Cybern. C. Appl. Rev., vol. 40, no. 3, pp. 352–357, 2010. doi: 10.1109/TSMCC.2009.2038279.

[3] Z. Du, Y. Deng, W. Guo, A. Nalllanathan, and Q. Wu, “Green deep reinforcement learning for radio resource management: Architecture, algorithm compression and challenge,” IEEE Veh. Technol. Mag., to be published. doi: 10.1109/MVT.2020.3019650.

[4] M. Ylianttila et al., “6G White paper: Research challenges for Trust, Security and Privacy,” 2020, arXiv:2004.11669.

[5] C. Li, C. Sun, S. Al-Rubaye, A. Tsourdos, and W. Guo, “Uncertainty propagation in neural network enabled multi-channel optimisation,” in Proc. 91st IEEE Vehicular Technology Conf., 2020, pp. 1–4. doi: 10.1109/VTC2020-Spring48590.2020.9128702.

[6] K. Siau and W. Wang, “Building trust in artificial intelligence, machine learning, and robotics,” Cutter Bus. Technol. J., vol. 31, no. 2, pp. 47–53, 2018.

[7] F. K. Došilović, M. Brčić, and N. Hlupić, “Explainable artificial intelligence: A survey,” in Proc. 2018 41st Int. convention Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp. 0210–0215. doi: 10.23919/MIPRO.2018.8400040.

[8] W. Guo, “Explainable artificial intelligence for 6G: Improving trust between human and machine,” IEEE Commun. Mag. vol. 58, no. 6, pp. 39–45, 2020. doi: 10.1109/MCOM.001.2000050.

[9] M. T. Ribeiro, S. Singh, and C. Guestrin, “Why should I trust you?: Explaining the predictions of any classifier,” in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 1135–1144. doi:10.1145/2939672.2939778.

[10] A. Shrikumar, P. Greenside, and A. Kundaje, “Learning important features through propagating activation differences,” in Proc. 34th Int. Conf. Machine Learning, 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 Proc. 28th Int. Joint Conf. Artificial Intelligence (IJCAI), Macao, Aug.10–16, 2019, pp. 2708–2715.

[12] N. Frost and G. Hinton, “Distilling a neural network into a soft decision tree,” 2017, arXiv:1711.09784.

[13] L. A. Hendricks, Z. Akata, M. Rohrbach, J. Donahue, B. Schiele, and T. Darrell, “Generating visual explanations,” in Proc. European Conf. Computer Vision, 2016, pp. 3–19. doi:10.1007/978-3-319-46493-0_1.

[14] P. Caballero, A. Banchs, G. De Veciana, and X. Costa-Pérez, “Network slicing games: Enabling customization in multi-tenant mobile networks,” IEEE/ACM Trans. Netw., vol. 27, no. 2, pp. 662–675, 2019. doi: 10.1109/TNET.2019.2895378.

[15] A. Glass, D. L. McGuinness, and M. Wolverton, “Toward establishing trust in adaptive agents,” in Proc. 13th Int. Conf. Intelligent User Interfaces, 2008, pp. 221–236. doi: 10.1145/1578773.1578804.