In: 2023 IEEE 98th Vehicular Technology Conference (VTC2023-Fall), 10-13 October 2023, Hong Kong
DOI: 10.1109/VTC2023-Fall60731.2023.10333364

Quality-of-Trust in 6G: Combining Emotional and Physical Trust through Explainable AI

Chen Li, Weijie Qi, Bailu Jin, Panagiotis Demestichas, Kostas Tsagkaris, Yiouli Kritikou, Weisi Guo

***Abstract*—Wireless networks like many multi-user services have to balance limited resources in real-time. In 6G, increased network automation makes consumer trust crucial. Trust is reflect in both a personal emotional sentiment as well as a physical understanding of the transparency of AI decision making. Whilst there has been isolated studies of consumer sentiment to wireless services, this is not well linked to the decision making engineering. Likewise, limited recent research in explainable AI (XAI) has not established a link to consumer perception.**

Here, we develop a Quality-of-Trust (QoT) KPI that balances personal perception with the quality of decision explanation. That is to say, the QoT varies with both the time-varying sentiment of the consumer as well as the accuracy of XAI outcomes. We demonstrate this idea with an example in Neural Water-Filling (N-WF) power allocation, where the channel capacity is perceived by artificial consumers that communicate through Large Language Model (LLM) generated text feedback. Natural Language Processing (NLP) analysis of emotional feedback is combined with a physical understanding of N-WF decisions via meta-symbolic XAI. Combined they form the basis for QoT. Our results show that whilst the XAI interface can explain up to 98.9% of the neural network decisions, a small proportion of explanations can have large errors causing drops in QoT. These drops have immediate transient effects in the physical mistrust, but emotional perception of consumers are more persistent. As such, QoT tends to combine both instant physical mistrust and long-term emotional trends.

***Index Terms*—machine learning; deep learning; XAI; wireless; trust; sentiment; NLP; LLM;**

I. Introduction

Trust in any critical service is difficult to quantify but yet important to understand. It is the basis of relationships between operators and consumers. Wireless networks have moved from Quality-of-Service (QoS) indicators (e.g., outage probability, throughput) to Quality-of-Experience (QoE) indicators (e.g., enjoyment, video clarity, gaming smoothness) [1], [2]. As we move towards the increased use of Artificial Intelligence (AI) in network automation, ranging from automated network integration to base station deployment to context profiling. As automation by AI increases in 6G, we may see both increased trust requirements for safety-critical services such as remote surgery and vehicle control [3], as well as increased risks of mistrust growing in consumers as they compete for vital radio resources. Indeed, EU and national laws mandate the AI decision making is transparent and explainable (EU GDPR Recital 71, US Equal Credit Opportunity Act Reg B, art 1002.9) to improve trust for diverse stakeholders. These laws apply to all AI that affect human lives, especially to national critical services such as wireless networking [4]. The method of explaining AI decisions and building consumer trust is not standardized yet (although efforts are underway - see IEEE P2976, ISO/IEC CD TS 6254). Trust in AI-enabled 6G must be reflected in future 6G network design, and Quality-of-Trust (QoT) remains not widely defined and tested.

C. Li (student), B. Jin (student), W. Guo are with Cranfield University, UK. W. Qi is with Ranplan Wireless, UK. P. Demestichas, K. Tsagkaris and Y. Kritikou are with WINGS-ICT, Greece.
We acknowledge funding from EC H2020 Grant 778305 - DAWN4IoE: Data Aware Wireless Network for Internet-of-Everything, and EPSRC Grant EP/X040518/1 - CHEDDAR: Communications Hub For Empowering Distributed ClouD Computing Applications And Research.

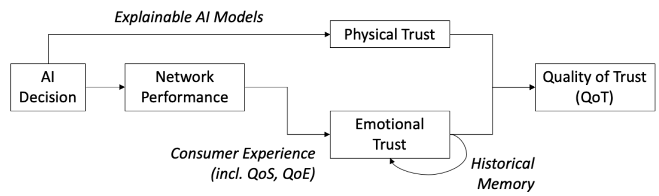

Fig. 1. Quality of Trust (QoL) derived from Emotional Trust and Physical Trust of AI decisions.

Here, trust is in how AI manages the network, e.g., asking why I am given this service level. This is different to trust mechanisms in blockchains for data provenance, or trust in cybersecurity, or trust in entity identity [5]. More recent work to use Generative Adversarial Networks (GANs) to improve trust in AI is a data-driven way to improve robustness [6], but does not create transparency on why decisions are made. Our proposal for QoT is very different to QoS and QoE, or GAN approaches [6], in that it demands more understanding of "why" rather than "what". QoT can work with traditional QoS/QoE to pro-actively optimise network performance [7] and improve customer retention [8].

A. Review of Current Consumer Trust Mechanisms

Our prior work proposed that the QoT must reflect both physical trust and emotional trust [3] (see Fig.1). We recognise that this will vary with context and ethnographically, but nonetheless we discussed some priority 6G applications. We identified that time-sensitive and life-critical applications such as remote surgery and autonomous driving might demand both high QoT and high QoS and QoE. On the other hand, smart meters and industrial robots might require reasonably high QoT, but do not necessarily demand high QoS. Many applications such as autonomous driving or piloting will combine time-sensitive actions with high levels of emotional trust [9].

In our most basic idea [3], we seek to balance physical P and emotional E trust via [8]: $\text{QoT}(m) = \alpha P(m) + (1-\alpha)E(m)$, for a particular AI model m. Here, α is a personalised parameter and we now review how physical and emotional trust functions manifest in the context of AI-enabled wireless networking.

1) Physical Trust: Physical trust is often reflected in how much we can quantify the accuracy of both the AI model m and the way we can explain its decisions. Regardless of the type of model and explainable method, we can generally say physical trust is [3]:

$$P(m) = R(m)A(m)\frac{\tau(m)}{c(m)}, \tag{1}$$

where we are interested in the robustness R and accuracy A of the model, and the transparency τ of explanations and normalised by the complexity of the XAI model c.

2) Emotional Trust: The emotional trust parameter cannot directly sensed and analyzed from the physical structure of model, but can be sensed from emotional changes collected through wearable devices and semantic feedback. Emotional trust includes memories of past via complex cognitive models and includes QoS metrics in intangible ways. As such, direct mapping of QoS metrics (e.g., throughput, outage) to QoE or Trust models assumes universal behavioural response [10], which does not reflect the diversity in different demographic of consumers and time-varying cognitive states [11]. As such, we avoid direct mapping of QoS to trust, which would otherwise involve complex multi-modal human monitoring. Instead we seek to map QoS to AI-generated human behaviour and emotional feedback first, and then analyse this emotional feedback to develop trust metrics.

To quantify emotional trust, the testing institution needs to organize a test group with participants and a daily trust baseline indicates the willingness to trust for each individual and could affect their choices in the emotional trust test. In reality, continuous testing of the individual baseline is necessary so that those with unstable moods do not participate in the emotional trust test. As shown below, accepted testing sentiment data $\gamma(t)$ will be weighted by l based on the willingness gap between the daily baseline and long-term baseline of individual participants:

$$E(m) = \sum_{t}^{T} l(t)\gamma(t)/T. \tag{2}$$

In recent years, large-scale social media analytics has enabled researchers to gather statistically significant data on consumer emotions. For example, general perceptions of 5G and correlations between social media and wireless traffic has been analysed [12]. Our own work in 2017-19 examined how we can mine emotional sentiment towards telecom subtopics (e.g., signal reception, data rate) from Twitter [13]. We expanded this work to give geographic and event contexts (e.g., O2 telecom blackout [14]). Nonetheless, there are major research gaps in mapping emotional sentiment to specific decision processes in AI-enabled wireless networks.

B. Application Area: Neural Water Filling (N-WF) Allocation

In this paper, we select one of many potential application areas for AI in wireless networking. Deep learning in AI is capable of achieving effective high-dimensional non-linear mapping that can improve the efficiency of traditional low-dimensional linear or non-linear iterative algorithms. The classic MIMO Water-filling (WF) power allocation (IEEE 802.xx series, OFDM systems) is selected as a universally understood classic example to demonstrate explainability. It's a well established mechanism of Lagrangian optimisation, which produces a WF threshold that has to be iteratively searched to achieve optimal power allocation under a fixed power budget. The essential functional role of that optimisation plays can be represented as $\boldsymbol{y} = f(\boldsymbol{x}; \lambda)$, where inputs $\boldsymbol{x} \in \mathbb{R}^{n\times 1}$ (e.g. channel gains) map to an output $\boldsymbol{y} \in \mathbb{R}^{n\times 1}$ (e.g. power allocation) via a model $f(\cdot)$.

- In classic WF, the Lagrangian optimisation produces an iterative solution to search for λ. The issue is that this is a slow process relatively to some of the high assignment demands in massive connectivity and innovations have been made to accelerate the search algorithm.
- In Neural-WF (N-WF) [15], [16], an approximate non-linear mapping of $\hat{\boldsymbol{y}} = f^*(\boldsymbol{x})$ can achieve effective power allocation $\hat{y}$ without iterative search for λ the WF level. This is useful for real-time power allocation [17] over a large number of parallel channels.

The challenge is that NNs lack the performance guarantee and explainable power of the classic WF solution. Black-box NNs cannot explain the essential features that influence actions (from PHY layer signal detection to MAC layer reinforcement learning optimisation [4]).

C. Novel Contributions & Organisation

Whilst there has been isolated studies of consumer sentiment to wireless services, this is not well linked to the decision making engineering. Likewise, limited recent research in XAI has not established a link to consumer perception. The novelty here are:

- develop and test a novel Quality-of-Trust (QoT) KPI that balances personal perception (emotional) with the quality of XAI outputs (physical).
- demonstrate this idea with an example in Neural Water-Filling (N-WF) power allocation, where the channel capacity is perceived by consumers that communicate through text feedback.
- implement explainable AI (XAI) using meta-symbolic methods that balances both native AI accuracy and robustness measures with the XAI transparency and complexity measures to create *physical trust KPI*.
- implement artificial consumer feedback via a Large Language Model (LLM) that modulates the *emotional trust KPI* based on throughput and previous experiences.
- combined spectrogram analysis of different users' Quality-of-Trust and its relation to WF performance to relate trust dynamics to wireless network performance.

We hypothesize that whilst the XAI interface can explain up to 98.9% of the neural network decisions based on previous work

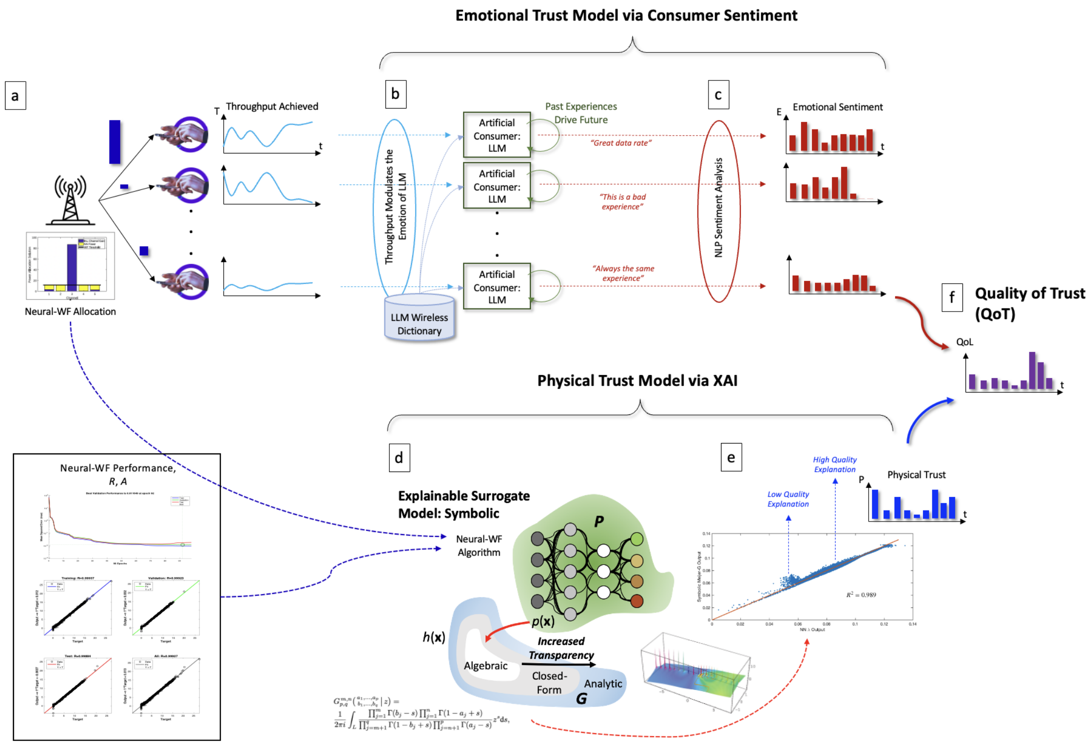

Fig. 2. System Model. (a) Wireless network with Neural-WF allocation of resources causing time-varying throughput amongst competing consumers. Emotional Trust: (b) Throughput modulates a LLM to mimic consumer behaviour, (c) sentiment analysis of LLM consumer feedback to create emotional trust. Physical Trust: (d) Neural-WF is projected to an XAI meta-symbolic function space via Meijer-G mapping, (e) which yields varying levels of physical trust based on the transparency of the explanations. (f) Combining emotional and physical trust yields the QoT.

[18], a small proportion of explanations can cause long-term drops in emotional trust.

The paper is organised as follows. In Section II, we define the wireless system and the Neural accelerated WF allocation algorithm. In Section III, we use explainable AI to define the physical trust KPI. In Section IV, we use LLMs to generate artificial consumer emotional trust KPIs. In Section V, we show combined results and analysis.

II. Wireless System and N-WF Setup

We consider a classical wireless parallel channel power allocation problem comprised of N channels with independent fading characteristics. The details of the channel model is standard 3GPP and given in Table I.

WF power allocation is well-established and we will not detail it here. Suffice to say, under the Shannon capacity assumption and a total power budget, the solution form for power in channel $n \in N$ is of (see Figure2a):

$$\rho_n = f[h_n; \lambda] = \left(\frac{1}{\lambda} - \frac{N_0}{|h_n|^2} \right)^+, \tag{3}$$

where N_0 is the AWGN power, h_n is the fading gain, and the parameter λ is the Lagrangian multiplier and a total power budget is applied to constrain the problem. The iterative search for the WF threshold $\frac{1}{\lambda}$ enables each channel to be allocated the optimal level of power. Channels with a gain that is too low will be allocated no power (i.e. $(\cdot)^+$ operator). We use the classic WF algorithm to train a 2 hidden-layer NN with 10 sigmoid neurons per hidden layer, using random data division for validation and testing [18] (see Fig.3). The input are the channel states and the output is the Lagrangian multiplier λ. The Lagrangian parameter in turn directly gives the power allocation output ($\lambda \rightarrow \rho_n$).

The implementation parameters are given in Table I. The resulting performance of the NN is as follows: training R^2 = 0.996, and out-of-sample test R^2 = 0.993, achieved convergence at 90-200 out of 1000 epochs.

TABLE I
System Setup for Wireless Network and Neural-WF Algorithm

Parameter	Value
Propagation Channel	2.4GHz 3GPP Micro
Wireless Channel	Downlink, 20MHz
Mobility Model	Gravity Random Walk [19]
Fading Channel	Rayleigh Fading
Wireless System	OFDMA, N =5, $\rho_{\max}$ =40W
Iterative WF Monte-Carlo Loops	5e4
Hidden Layers	2 with 10 Sigmoid neurons/layer
Training	Random division, $\leq$1000 epochs
Explainability	Meijer G-function
In-Sample Accuracy, A	99.6%
Out-of-Sample Robustness, R	99.3%

III. Physical Trust via Symbolic Explainability

A. Review of Explainable AI (XAI) Methods

XAI solutions that attempt to de-mystify NN mappings, building trust between machine intelligence and human expertise [4]. These XAI techniques range from visualising key hidden layer features to localised linear models (LIME) [20]. Visualising feature propagation can reveal useful component level insights of sensitivity to perturbations, but cannot construct a causal symbolic mapping of $h \rightarrow p$. LIME models offer local linear or simple symbolic models, but lack the general end-to-end insight. Work in [21] have shown that Meijer G-functions are general input output symbolic meta-representations and have the potential to explain neural networks. Our prior followup work from [21] extended the meta-symbolic framework to explain Neural-WF in wireless systems [18], where we showed the XAI method can explain up to 98.9% of the neural network decisions. We will leverage on this work to build a physical trust model, and then integrate with the emotional trust model.

B. Meijer-G Function Mapping

Meijer G-function is a general function intended to include most known special functions. A general definition is given by a line integral in the complex plane:

$$G_{p,q}^{m,n}\left(\begin{smallmatrix} a_1,\ldots,a_p \\ b_1,\ldots,b_q \end{smallmatrix} \middle| z\right) = \frac{1}{2\pi i}\int_L \frac{\prod_{j=1}^{m}\Gamma(b_j - s)\prod_{j=1}^{n}\Gamma(1-a_j+s)}{\prod_{j=m+1}^{q}\Gamma(1-b_j+s)\prod_{j=n+1}^{p}\Gamma(a_j-s)} z^s \mathrm{d}s, \tag{4}$$

where $\Gamma(\cdot)$ denotes the gamma function. As such, the Meijer G-function provides a flexible framework to discover the mapping between input variables (e.g. channel gain z) and output solution (e.g. power allocation ρ) through tuning integral residue by zeros and poles. Indeed, this was the motivation behind recent work in the biomedical domain [21].

Fitting the Meijer G-function to any set of data is almost impossible due to the large search space of the discrete hyper-parameters m, n, p, q. We have fitted the observed NN output of the Lagrangian multiplier λ (WF level) to an inverse tangent function (see Figure2d) [18]:

$$\begin{aligned} \lambda &= a \times \sum_{n=1}^{5} \tan^{-1}((|h_n|^2 + b_n)^{-2} + c_n) + d \\ &= a \times \sum_{n=1}^{5} \frac{1}{2|h_n|^2} G_{2,2}^{1,2}\left(\begin{smallmatrix} 1,\frac{3}{2} \\ 1,\frac{1}{2} \end{smallmatrix} \middle| z_1 + c_n\right) + d \end{aligned} \tag{5}$$

where

$$z_1 = G_{0,1}^{1,0}\left(\begin{smallmatrix} - \\ 2 \end{smallmatrix} \middle| |h_n|^2 + b_n\right) + G_{1,2}^{1,1}\left(\begin{smallmatrix} -1 \\ -1,-2 \end{smallmatrix} \middle| |h_n|^2 + b_n\right) \tag{6}$$

Given λ, we can find the corresponding power allocation values ρ_n^*, where the * indicates the explained power allocation as opposed to the actual. This is useful because the explained ρ_n^* can show why from the black-box neural network.

Note, whilst this explainable function was fitted with excellent agreement with coefficient of determination $R^2 = 0.989$ [18]. However, the XAI result show one of many plausible solutions and whilst the fitting is strong, if one had no intuition on the target function (WF is a well studied area), the search space can be too large.

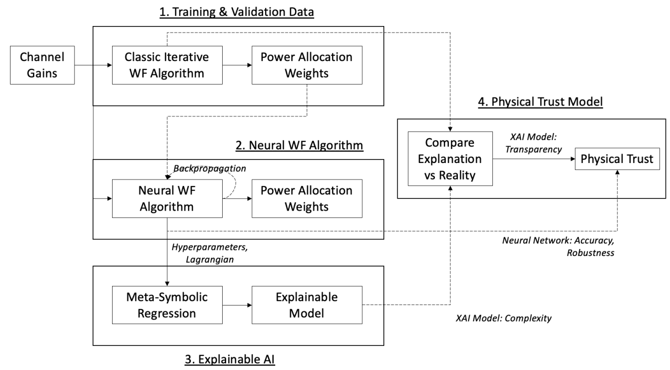

Fig. 3. Neural WF and Physical Trust Modeling Pipeline.

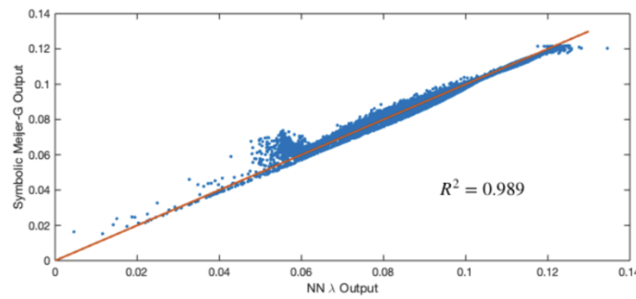

Fig. 4. Transparency τ reflected via the comparison between XAI Meijer-G output vs. Neural-WF output. Whilst overall transparency is 98.9%, there are instances where XAI error can be up to 35%.

C. Physical Trust at Individual Level

As discussed in the literature review, physical trust is reflected in how much we can quantify the accuracy of both the N-WF model m and the way we can explain its decisions [3] (see Fig.3):

$$P(m) = R(m)A(m)\frac{\tau(m)}{c(m)}, \tag{7}$$

where the instantaneous accuracy and robustness can be compared between the N-WF and WF solutions. So whilst the average in and out of sample performance is over 99%, specific instances for individual users of N-WF can have up to 10% error. The transparency of the XAI is given by the instant difference between the explainable model in Eq.(5) and the real WF model in Eq.(3): $\tau = \rho - \rho^*$ (note: we can either compare WF threshold λ or power ρ as it is a direct map).

So from Figure4 we can see that whilst the average accuracy of $\tau(m)$ is 98.9%, specific instances for individual users of XAI explanations can have up to 35% error (see Figure2e on impact). The model complexity is given by Eq.(5) and Eq.(6), where we can see that it scales linearly with the number of channels N and the number of parameters in z which is 8, and as such the general complexity scaling law of this particular XAI model for large channel numbers is $c = 8N$. Other parameters are given in Table II

TABLE II
SYSTEM SETUP FOR ARTIFICIAL CONSUMER SENTIMENT AND XAI INTERFACE

Parameter	Value
Artificial Consumer Approach	Social Simulacra [22]
Large Language Model (LLM)	ChatGPT 3.5-Turbo
Randomness Scheme	Softmax Temperature 1.0-1.4
Word Sampling Scheme	Top-p Sampling 0.5-0.95
Maximum Character Length	280
Wireless QoE Corpus	Twitter curated [13]
Grammar Model	3- or 4-gram
NLP Analysis	LSTM [13]
Explainable AI Model, m	Meijer-G [18], [21]
Transparency Accuracy, τ	up to 98.9%
Complexity, c	$8 \times N + 2$

IV. EMOTIONAL TRUST VIA ARTIFICIAL CONSUMER SENTIMENT GENERATION

A. Text Feedback Generation using LLMs

We are motivated by recent work in using LLMs to create a *Social Simulacra*, whereby artificial human personas mimic human feedback and dialogue [22]. It is important to create realistic consumer feedback, as this captures intangible emotional state dynamics in the language, as opposed to presuming that the raw QoS can be functionally mapped to a universal emotional response.

LLMs are highly advanced AI text completion algorithms designed to understand and generate human-like text based on the text input it's given. Here, we use ChatGPT v3.5T, which uses patterns in the text data it's been trained on to provide detailed, accurate, and contextually appropriate responses. It learns from a diverse range of Internet text, and this includes open access academic papers, forums, and social media. In this way, we believe at a basic level, it combines meta personalities [22]. Refinement of ChatGPT v3.5 is done in following stages: (1) a basic pre-training phase where demonstration data is used to supervise the training of a policy, (2) a fine-tuning phase based on first training a reward model and then the policy is optimised using Proximal Policy Optimization (PPO) reinforcement learning.

Here, we use a pre-existing refined ChatGPT v3.5T to generate a text feedback based two parts (see Fig.5):

1) a numerical sentiment parameter that directly scales with a time-average of recent throughput received (time-averaged approach allows the user to respond in a smooth way): $s = \overline{T}(t)$ (see Fig. 3a).
2) a topic configuration parameter informed by previous Twitter corpus [13].

The randomness of the different LLM agents giving feedback is controlled by the temperature of the softmax function and the way next word completion are sampled is given by the Top-p parameter; both of which are given in Table II. There are no frequency and presence penalties. Past answers are used to drive a prompt asking how the LLM agent feels about the current wireless service quality then produces the output for sentiment analysis below.

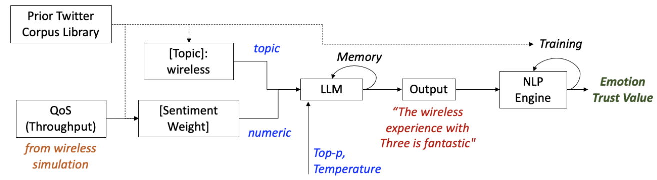

Fig. 5. LLM Feedback and NLP Pipeline: Topic and QoS drives the LLM to create artificial semantic feedback.

B. Sentiment Analysis: Corpus and Emotional Trust

In the sentiment analysis, we make topic-specific sentiment analysis using self-labeled language annotations to create a corpus based on how English-speaking users discuss wireless services in the Greater London area around 2011-16. In our previous work, we used 2 strategies to obtain the language corpus of users' qualitative experience of mobile networks:

- general large-scale data mining of QoE reporting on Twitter (validated through case studies using secondary cell signal data) [13], and
- encourage users to report QoE through dedicated social influence campaigns by @NoServiceHere Twitter account so that we can identify specific circumstances and context more accurately [14].

These studies were a joint university and industry (Ranplan UK) initiative. We were able to gather social media data from both a purpose built website account and the Twitter API, use R to process the text data to clean out spelling mistakes and typos, and then use Python code to perform the machine learning pipeline discussed below. This however was conducted over long-time scales and did not match to specific network radio engineering decisions. That is why we will use this corpus to drive our LLM and sentiment analysis, but not use it directly as data.

From previous experience (see [12], [13]) we found that 3-gram grammar models produced good accuracy. We used our previously created Corpus and Sentiment Analysis System [12], [13] and apply the following steps: (i) 3-gram TDSA model combined with corpus filtering, (ii) ML classifier with 3 choices (Naive Bayes, Support Vector Machine, and a Recurrent Neural Network). We used a relatively simple individual user's sentiment scale of positive (+1) and negative (-1) to define emotional response, and used a filter for smoothing over time to avoid sharp changes that may affect analysis. Here, as we deal with artificial consumers in the form of LLMs, we do not need to establish an emotional baseline and identify performance anomalies.

V. RESULTS

A. Time-Domain Trust Profiles

In the results, we first show a panel plot of example outputs at each stage of the pipeline. In Fig. 6a, we show an example user's downlink channel gain over 200 time samples. The corresponding N-WF performance and solutions are compared against the optimal classic WF solution in subplots b-c.

We then derive the normalised physical and emotion trust scores for 2 example users in subplots d-e, with QoT shown in

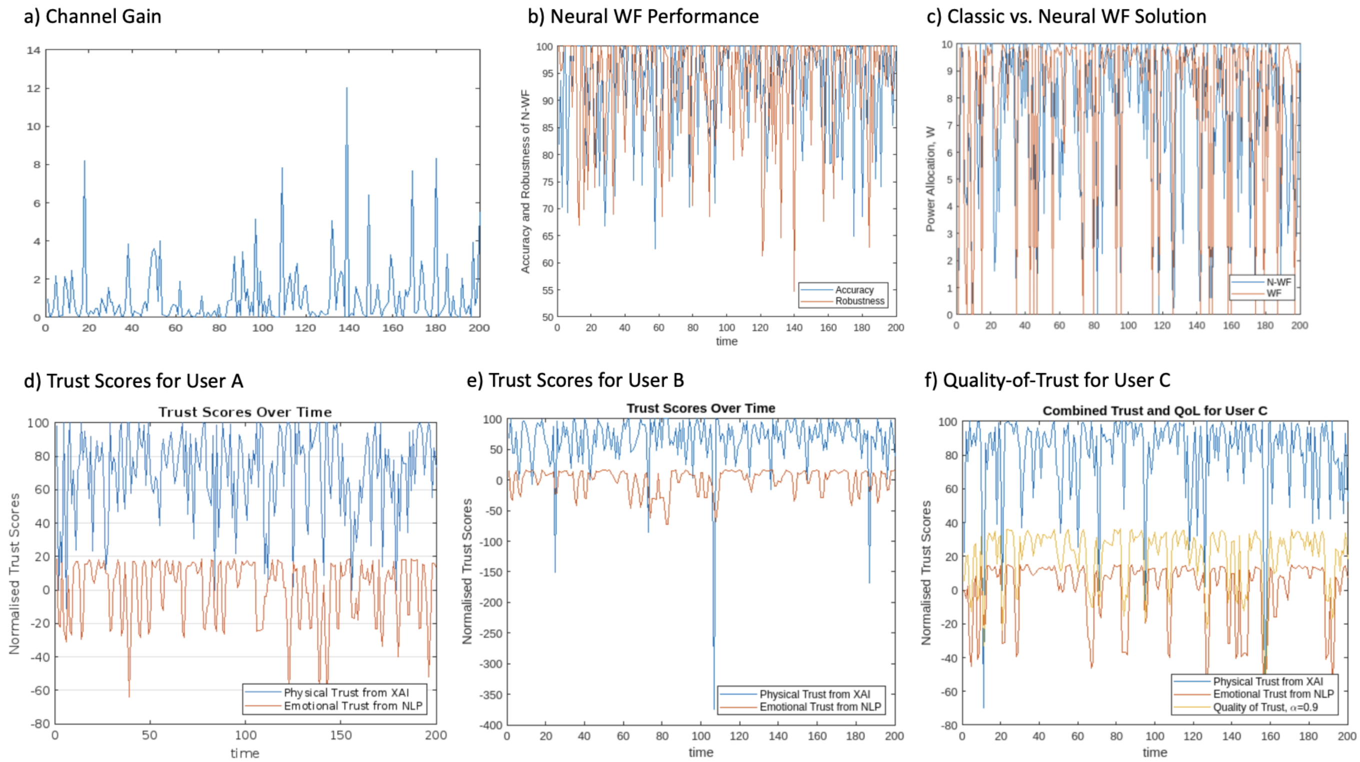

Fig. 6. Demonstrative results for different users: (a) example channel gain for a user, (b) Neural-WF Accuracy and Robustness performance, (c) comparison between optimal and neural accelerated power allocation solutions, (d-f) example trust dynamics.

subplot f. Here, our results show that whilst the XAI interface can explain up to 98.9% of the neural network decisions, a small proportion of explanations can have 35% error causing spikes in mistrust. These drops have transient effects in the emotional perception of consumers, eroding long-term QoT but with fewer spikes. Another observation we see both from past work analysing Twitter [13] and from here is that user feedback and trust dynamics tends to be negative, e.g., improved services does not cause overwhelmingly positive trends, whereas a drop sees a negative trend.

We demonstrate another example of the physical and emotional trust outputs in Fig.8, where we show how we can increase the LLM's dialogue memory to "smoothen" out the high frequency emotional reaction and create smoother trust dynamics.

B. Spectrogram Trust Profiles

We show this in Short Time Fourier Transform (STFT) spectrogram plots Fig. 7, where we can see that physical trust exhibits polarized dynamics between steady performance and spikes. This is because a poor explanation immediately leads to a spike in physical mistrust (low robustness R and/or transparency τ), but most of the performance is steady as the N-WF average performance is above 99% and the XAI average performance is above 98%.

As for emotional trust, we see it exhibits smoother dynamics with no sharp trust spikes and a relatively low power in spike events as emotions are more persistent, which is reflected in our LLM.

VI. Conclusions & Future Work

Wireless networks like many multi-user services have to balance limited resources in real-time. Consumer perception of service quality is reflect in both a personal emotional sentiment as well as a physical understanding of the decision making. Whilst there has been isolated studies of consumer sentiment to wireless services, this is not well linked to the decision making engineering. Likewise, limited recent research in explainable AI (XAI) has not established a link to consumer perception.

Here, we develop a Quality-of-Trust (QoT) KPI that balances personal perception with the quality of decision explanation. That is to say, the QoT varies with both the time-varying sentiment of the consumer as well as the accuracy of XAI outcomes. We demonstrate this idea with an example in Neural Water-Filling (N-WF) power allocation, where the channel capacity is perceived by artificial consumers that communicate through Large Language Model (LLM) generated text feedback. Natural Language Processing (NLP) analysis of emotional feedback is combined with a physical understanding of N-WF decisions via meta-symbolic XAI. Combined they form the basis for QoT. Our results show that whilst the XAI interface can explain up to 98.9% of the neural network decisions, a small proportion of explanations can have large errors causing drops in QoT. These drops have immediate transient effects in the physical mistrust, but emotional perception of consumers are more persistent. As such, QoT tends to combine both instant physical mistrust and long-term emotional trends.

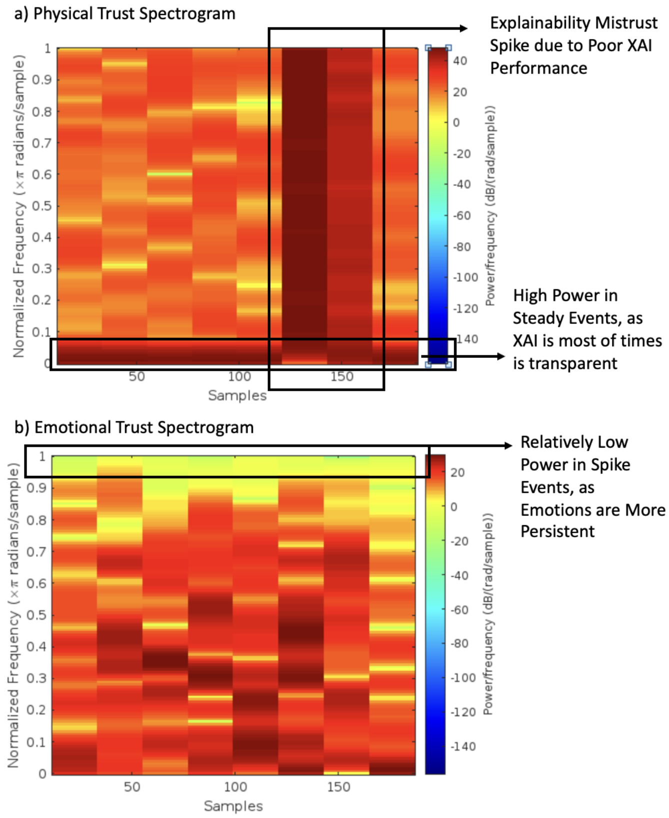

Fig. 7. Demonstrative spectrogram results: (a) physical trust exhibits polarized dynamics between steady performance and spikes, and (b) emotional trust exhibits smoother dynamics.

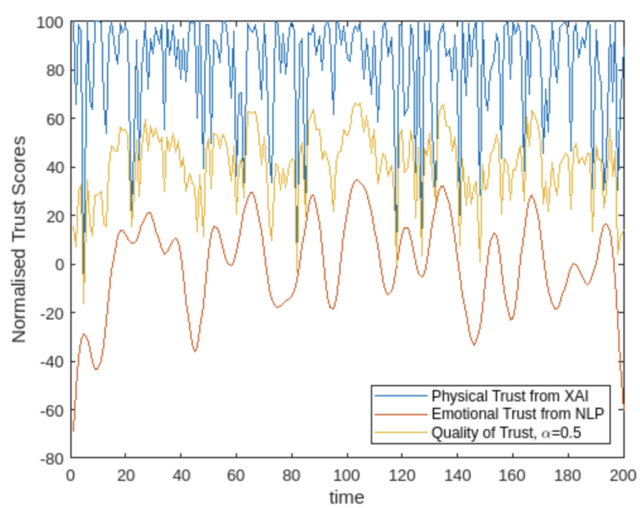

Fig. 8. Impact of increasing LLM dialogue memory in LLM.

Our future work will seek to: (1) create a feedback loop that optimises the QoT parameter α such that we can fine-tune to each individual, (2) use real-user feedback for emotional trust, and (3) perform pro-active network optimisation.

References

[1] B. Agarwal, M. A. Togou, M. Ruffini, and G.-M. Muntean, "QoE-Driven Optimization in 5G O-RAN-Enabled HetNets for Enhanced Video Service Quality," *IEEE Communications Magazine*, vol. 61, 2023.

[2] U. S. Hashmi, A. Rudrapatna, Z. Zhao, M. Rozwadowski, J. Kang, R. Wuppalapati, and A. Imran, "Towards real-time user qoe assessment via machine learning on lte network data," in *2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*, 2019, pp. 1–7.

[3] C. Li, W. Guo, S. C. Sun, S. Al-Rubaye, and A. Tsourdos, "Trustworthy deep learning in 6g-enabled mass autonomy: From concept to quality-of-trust key performance indicators," *IEEE Vehicular Technology Magazine*, vol. 15, no. 4, pp. 112–121, 2020.

[4] W. Guo, "Explainable Artificial Intelligence for 6G: Improving Trust between Human and Machine," *IEEE Communications Magazine*, 2020.

[5] W. She, Q. Liu, Z. Tian, J.-S. Chen, B. Wang, and W. Liu, "Blockchain trust model for malicious node detection in wireless sensor networks," *IEEE Access*, vol. 7, pp. 38 947–38 956, 2019.

[6] L. Yang, Y. Li, S. X. Yang, Y. Lu, T. Guo, and K. Yu, "Generative adversarial learning for intelligent trust management in 6g wireless networks," *IEEE Network*, vol. 36, no. 4, pp. 134–140, 2022.

[7] B. Ma, W. Guo, and J. Zhang, "A survey of online data-driven proactive 5g network optimisation using machine learning," *IEEE Access*, vol. 8, pp. 35 606–35 637, 2020.

[8] 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*, ser. IUI '08. New York, NY, USA: Association for Computing Machinery, 2008. [Online]. Available: https://doi.org/10.1145/1378773.1378804

[9] H. Kim, J. Ben-Othman, L. Mokdad, J. Son, and C. Li, "Research challenges and security threats to ai-driven 5g virtual emotion applications using autonomous vehicles, drones, and smart devices," *IEEE Network*, vol. 34, no. 6, pp. 288–294, 2020.

[10] N. Banović-Ćurguz and D. Ilišević, "Mapping of qos/qoe in 5g networks," in *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics*, 2019.

[11] T. Hoßfeld, P. E. Heegaard, L. Skorin-Kapov, and M. Varela, "No silver bullet: Qoe metrics, qoe fairness, and user diversity in the context of qoe management," in *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, 2017, pp. 1–6.

[12] B. Yang, W. Guo, B. Chen, G. Yang, and J. Zhang, "Estimating mobile traffic demand using twitter," *IEEE Wireless Communications Letters*, vol. 5, no. 4, pp. 380–383, 2016.

[13] W. Qi, R. Procter, J. Zhang, and W. Guo, "Mapping consumer sentiment toward wireless services using geospatial twitter data," *IEEE Access*, vol. 7, pp. 113 726–113 739, 2019.

[14] W. Qi, W. Guo, R. Procter, and J. Zhang, "Geo-tagging quality-of-experience self-reporting on twitter to mobile network outage events," in *2019 IEEE International Smart Cities Conference (ISC2)*, 2019.

[15] F. Liang, C. Shen, W. Yu, and F. Wu, "Towards optimal power control via ensembling deep neural networks," *IEEE Transactions on Communications*, pp. 1–1, 2019.

[16] L. Zhao, Y. Wang, and P. Chargé, "Efficient iterative water-filling power allocation method in MU-MIMO broadcast channels," in *Military Communications and Information Systems Conference*, Oct 2013.

[17] H. Sun, X. Chen, Q. Shi, M. Hong, X. Fu, and N. D. Sidiropoulos, "Learning to optimize: Training deep neural networks for wireless resource management," in *IEEE International Workshop on Signal Processing Advances in Wireless Communications*, July 2017.

[18] S. C. Sun and W. Guo, "Approximate symbolic explanation for neural network enabled water-filling power allocation," in *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, 2020, pp. 1–4.

[19] A. Orsino, W. Guo, and G. Araniti, "5g multiscale mobility : A look at current and upcoming models in the next technology era," *IEEE Vehicular Technology Magazine*, vol. 13, no. 1, pp. 120–129, 2018.

[20] M. T. Ribeiro, S. Singh, and C. Guestrin, ""why should i trust you?": Explaining the predictions of any classifier," in *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: ACM, 2016, pp. 1135–1144.

[21] A. Alaa and M. Schaar, "Demystifying black-box models with symbolic metamodels," in *NIPS*, 2019.

[22] J. S. Park, L. Popowski, C. Cai, M. R. Morris, P. Liang, and M. S. Bernstein, "Social simulacra: Creating populated prototypes for social computing systems," in *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, ser. UIST '22. New York, NY, USA: Association for Computing Machinery, 2022.

2023-12-11

Quality-of-Trust in 6G: combining emotional and physical trust through explainable AI

Li, Chen

IEEE

Li C, Qi W, Jin B, et al., (2023) Quality-of-Trust in 6G: combining emotional and physical trust through explainable AI. In 2023 IEEE 98th Vehicular Technology Conference (VTC2023-Fall), 10-13 October 2023, Hong Kong

https://doi.org/10.1109/VTC2023-Fall60731.2023.10333364