WIRELESS WORLD
RESEARCH FORUM®

OUTLOOK

Visions and research directions for the Wireless World

Artificial Intelligence in Future Wireless Communication Systems

May 2019, No 19

WWRF WG-WAI: Artificial Intelligence for Wireless Communications

White Paper
Artificial Intelligence in Future Wireless Communication Systems

Editors:
Fisseha Mekuria, Council for Scientific & Industrial Research
Jean-Paul Pallois, EU Product Management, Huawei France
Miguel Ángel Vázquez, Centre Tecnològic de Telecomunicacions de Catalunya
Sudhir Dixit, Basic Internet Foundation, Norway and University of Oulu, Finland

Version 1.0

Project website address: www.wwrf.ch

Contributors

Andreas Georgakopoulos	WINGS ICT Solutions
Christos Masouros	University College London
Fisseha Mekuria	Council for Scientific & Industrial Research
Javan Erfanian	Bell Canada
Jean Paul Pallois	Huawei
Kostas Tsagkaris	Incelligent
Merouane Debbah	Huawei
Miguel Ángel Vázquez	Centre Tecnològic de Telecomunicacions de Catalunya
Sudhir Dixit	Basic Internet Foundation, Norway and University of Oulu, Finland
Panagiotis Demestichas	Incelligent
Tony Kenyon	University College London
Vino Vinodrai	Vinodrai & Associates Inc
Vinod Kumar	WWRF
Yansha Deng	King's College London

Executive Summary

Artificial Intelligence (AI), Machine Learning (ML) are hot topics nowadays and have achieved huge breakthroughs in the recent years and months. AI is transformative and will be embedded in each link of the overall chain of the communications system. We envision to see AI almost everywhere: in connected devices, machines, objects, automobiles and many connected things but also in all elements making up the network, such as base stations, controllers, core network equipment, operations and management system. However, we need to be aware that this AI proliferation and pervasiveness will generate potential entanglement which will need to be solved. Nevertheless, we can expect AI in each device to process data close to the source to complement the cloud, especially for latency-sensitive and mission-critical applications. This intelligence at the micro level will need to be abstracted to the higher levels for macro control and holistic view of the network. AI in the device will also be useful when the device is disconnected from the network and the cloud.

However, AI is not seen as the holy grail of technology by everyone. Some major issues may slow down the adoption of AI by the industry. The first reason may be the difficulty to fully understand the real benefits that AI can bring in daily work despite the current market trend. Another reason is also human, emotional reason: people in the industry may perceive AI as a potential threat to their jobs. Eventually, the AI proponents will have to deal with the above challenges, but even if the AI advocates can overcome these barriers, the need to adapt existing operations processes will be a heavy, cumbersome and tedious task and will need support from the top management to adopt AI. The last obstacle will be the expected transparency of AI. Indeed, delivering AI as a black box may be enough in some cases for basic actions (e.g. like how to optimize parameter settings during a roll-out phase) but would not be acceptable during the optimization or healing phases. Operators who could not be in the position to explain what happened and why it happened could be caught up in explaining to the public, regulators and may suffer legal consequences. Therefore, it would be necessary to be able to explain AI to all stakeholders. In short, for AI in network to be successful, all the above key issues would need to be addressed.

Table of Contents

1. Introduction

The success of artificial intelligence (AI) in tackling long term problems of image processing started a new era where AI is being studied and used in many fields from natural language processing to military defence systems. In this context, the wireless communications stakeholders have shown a strong interest in including AI in their daily life engineering problems. Apart from being a simple yet powerful tool for trying to solve current and future problems that appear in the wireless communication field, AI introduces a crucial change in the ecosystem. Indeed, as the techniques need to rely on data, the owners of the data become a key player in the use of AI in the wireless infrastructure.

Bearing in mind this paradigm change, it is crucial to understand the potentials of AI in supporting different areas of the wireless communications systems in order to encourage the data owners to facilitate its different usages. This is done in Section 2 of this white paper where we introduce AI approaches to deal with different wireless communications problems. As it is described, the introduced AI techniques shall be fuelled with data coming from different parts of the wireless infrastructure.

Once the AI potential is identified, a careful assessment of its deployment shall be addressed. This is presented in Section 3, where the tentative shortcomings AI of implementation in wireless systems are identified. Those barriers include the ethics which might become a key element of data owners when rolling out AI tools. It should be noted that the scope of this white paper is on the use of AI in the access part of the network and not on the core Internet as it is dealt with already by the core network equipment manufacturers and fixed network providers.

2. AI for the Wireless World

2.1 Intelligent Spectrum management

Standards organizations, telecom operators, research institutes and service providers agree that a new spectrum management paradigm is required to replace the static spectrum allocation which awards spectrum chunks to specific users in specific areas for a specified period of time. Intelligent spectrum management (ISpM) techniques are therefore needed to support the complex future wireless network resource and service requirements including the high bandwidth demand use cases in emerging heterogeneous network technologies evident in 5th generation and beyond.

It is expected that 5G mobile networks will be the platforms to test and deploy such ISpM techniques ubiquitously. Existing spectrum management techniques require evolutionary transformation towards ISpM techniques. Such techniques can utilize recent developments in network technology standards, smart spectrum sharing techniques, machine learning based interference mitigation and AI inspired network management systems. Specifically, ISpM will cater to the dynamic and heterogeneous nature of future wireless network technologies and service requirements. Such ISpM techniques will contribute to research and innovations in smart spectrum sharing (S^3) and produce results with profound impact in: (1) Efficient utilization of available spectrum resources and providing sufficient spectrum to the 5G networks for the enhanced mobile broadband (eMBB) and other 5G use cases (such as affordable broadband, IoT & rural long range 5G); (2) Reduced interference probability and provision of quality-of-service (QoS) and/or quality-of-experience (QoE) guarantees; (3) Affordable 5G broadband deployment based on geo-location and unlicensed spectrum. Technological progress, standards and new spectrum technical regulatory rules, enhanced interference mitigation are also being proposed to enable sustainable deployment of networks and improved operational aspects following the rules defined for ISpM techniques.

Existing spectrum allocation and management process is performed manually in a static manner whereby spectrum divided and allocated to service providers in a rigid and exclusively licensed bands per national or regional boundaries. This human-driven process is not adaptive to the dynamics of supply and demand in future wireless networks, and thus cannot exploit the full potential capacity of the spectrum. Future developments of intelligent spectrum management systems will need the necessary technical co-existence framework which will utilize AI inspired interference mitigation and regulatory mechanisms to make smart spectrum sharing work in all bands of interest.

The development of technologies for software defined radios and spectrum sensing are also the other drivers in the quest for dynamic access to spectrum and cognitive radio based networks. The solution for this is to implement the ISpM system which will realize the 3 aspects described in section 2.1. Intelligent spectrum management will use a unified intelligent geo-location spectrum database [1], augmented by spectrum sensing, spectrum management technologies using AI & machine learning ML tools. Integrating adaptive spectrum databases with inputs from AI & ML tools [2]-[3] for spectrum sensing and band specific standards/regulatory rules is the way forward. Such techniques will help us realize intelligent spectrum management and smart spectrum sharing, and support the realization of advanced use cases such as the 5G eMBB and provision of affordable broadband to the digitally excluded. The next step in intelligent spectrum management is the enablement of reconfigurable radio networks based on software defined radios, thereby allowing heterogeneous networks to dynamically and collaboratively share a given spectrum band, including machine learning based real-time RF interference mitigation.

2.2 Radio Air Interface

Massive MIMO is going to be a key technology in new networks to deliver higher capacity. However, there are technical road blocks to efficiently tune massive MIMO. One of the challenges comes from a fast enough selection and optimization of the appropriate beam pattern. KPIs such as the antenna tilt, the interference and noise (INR), the Received Signal Strength (RSS) and the reference signal received power (RSRP) can be monitored and analyzed. To select the best beam pattern, AI enables massive MIMO to be effective. Indeed, AI allows a learning and iterative approach, to select the best beam pattern. This is a topic on which the Huawei Paris Research Center works as well by adopting a reinforcement learning (RL) algorithm. Indeed, RL enables, during the exploration phase – exploration here means “let’s discover my codebook, test the best response and select the best pattern”, a very efficient discovery method. RL can be seen as a kind of unsupervised learning algorithm as the learning phase is done “in situ”. RL includes two distinct phases: 1) Exploration and when convergence is reached and 2) Exploitation. The algorithm uses a smart approach, based on a technique called “Random Forest” to find out, as fast as possible, the best beam pattern.

The scaling-up of the networks in future generations will necessitate low-cost IoT devices that will involve low-specification hardware, such as analog-to-digital converters, power amplifiers to name a few. These are prone to hardware failures such as malfunctions in antennas or RF chains during the network operation, which may jeopardize the reliability of the MTC links. Hardware impairments and losses in analogue components are often mathematically intractable and commonly approximated with additive noise models. As a result, existing analytically-driven approaches to hardware impairments, are based on oversimplified models and are suboptimal in practice. This model mismatch is largely overlooked and can only be revealed through the application of *data-driven* approaches to the design of communications systems, as opposed to traditional *system-model driven* designs, and through experimental validation.

In this scope, learning-based approaches appear to be prominent candidates for addressing such mathematically intractable paradigms, owing to a number of advantages [4], [5]: Firstly, a learning-based communication link allows an optimal data-driven architecture to be designed without the need for the system model to be analytically tractable, making it essential for the emerging communication scenarios above. Secondly, it can be optimized for the end-to-end performance, rather than independent block-by-block design that may be sub-optimal from a holistic point of view. Finally, through layered training It is able to adapt to unpredictable changes, e.g. due to channel/hardware impairments/failures, to provide optimal performance without the need to re-design any system models. Accordingly, the application of learning-based approaches to dealing with hardware imperfections and failures, promises an exciting and fruitful research direction in the AI context.

2.3 Radio Access

With the rapid proliferation of innovative applications in the paradigm of massive Internet of Things (mIoT), such as smart city, smart home, smart industrial, and vehicular communication, the demand of data traffic in wireless networks has grown explosively. In view of this, providing reliable wireless access for the mIoT network becomes challenging due to its nature of massive IoT devices and diversification of data traffic [7]. IoT devices perform Random Access (RA) procedure to request channel resources for uplink transmission in the cellular-based mIoT network, where the massive mIoT access will impose enormous collisions during the RA. It has already been shown in [8][9] that the conventional RA schemes, such as Access Class Barring (ACB), back-off, and power ramping schemes, become inefficient in terms of RA success probability as the number of IoT devices increases.

To effectively support the emerging massive Internet of Things (mIoT) ecosystem, the 3rd Generation Partnership Project (3GPP) partners have standardized a new radio access technology, namely NarrowBand-IoT (NB-IoT) [10]. To support various traffic with different coverage requirements, NB-IoT supports up to three Coverage Enhancement (CE) groups of IoT devices sharing the uplink resource in the same band [11]. At the beginning of each uplink Transmission Time Interval (TTI), eNB selects a system configuration that specifies the radio resource allocated to each group in order to accommodate the RACH procedure along with the remaining resource for data transmission. The key challenge is to optimally balance the allocation of channel resources between the RACH procedure and data transmission so as to provide maximum success accesses and transmissions in massive IoT networks. Unfortunately, dynamic RACH and data transmission resource configuration optimization is an unsolved problem in cellular networks.

Generally speaking, the eNB observes the transmission receptions of both RACH (e.g., number of successfully received preambles and collisions) and data transmission (e.g., number of successful scheduling and unscheduling) for all groups at the end of each TTI. This historical information can be potentially used to predict traffic from all groups and to facilitate the optimization of future TTIs' configurations. Even if one knew all the relevant statistics, tackling this problem in an exact manner would result in a Partially Observable Markov Decision Process (POMDP) with large state and action spaces, which would be generally intractable. The complexity of the problem is compounded by the lack of a prior knowledge at the eNB regarding the stochastic traffic and unobservable channel statistics (i.e., random collision, and effects of physical radio including path-loss and fading).

In order to consider more complex and practical formulations of radio access techniques, Reinforcement Learning (RL) [12] emerges as a natural solution given the availability of feedback in the form of number of successful and unsuccessful transmissions per TTI. To solve the high-dimensional configurations problem in the multi-parameter multi-group scenario as defined in NB-IoT standard, Deep RL [13] can be the convergent capability of Q-learning by sacrificing the accuracy in resource configuration [14]. In addition, 5G radio access will involve connecting simultaneously with several radio access technologies (from micro-cellular to macro-cellular), and AI can play a major role in determining which radio access should be used at a particular time and place and how the intra- and inter-technology handovers should take place considering multiple parameters.

2.4 Radio resource management

Both academia and industry agree that early adoptions of the 5G infrastructure shall accommodate a flexible network configuration able to provide services to an extremely large variety of services and clients. This network architecture design is coined as network slicing [15]-[16] and, in contrast to current pre-5G networks, it aims to provide granular and heterogeneous services. That is, the whole network is separated into different slices each of them tackling a specific service requirement (internet of things, mobile edge computing,…), which may be customised for a particular customer or data type (video, audio, data) or industry verticals.

Efficiently guaranteeing certain service level agreements (SLAs) among the different slices is a cumbersome and computationally demanding task. First, dynamic demands impose different and heterogonous slices over time, promoting a very short time-to-react to the network orchestrator. In this context, the network slicing optimization has to rely on heuristic approaches that have a very low computational complexity yet are stable for a period of time. Second, although SLAs are currently being established with clear metrics, tenants might impose other requirements such as the quality-of-experience perceived by the end user whose modelling hinders the overall network optimization.

In light of the above discussion, network orchestrators shall consist of tools with low processing time that are able to manage the different network slices while maintaining an efficient radio resource usage.

Current network slicing optimizers' proposals (e.g. [17], [18]) consist of computationally demanding operations, leading to slow reactive network responses to slices in response to time varying user requests. In order to conceive quicker reactive network slicing controllers yet providing efficient network usages, the use of AI tools might be one of the key elements [19].

In one hand, supervised deep learning regression method can tackle the power control optimization [20] and the user scheduling process [21]. This machine learning model assumes that

the network controller is able to perfectly model the communication system and a synthetic data set could be available for designing the deep neural network. The main rationale of this approach entails that a deep neural network shall properly infer the results of other computationally demanding algorithms with a substantial decrease of the computational time.
On the other hand, when modelling the system involves challenging technical issues, the system designer can resort to model-free reinforce learning mechanisms [22],[23] that are able to provide efficient slices configurations via an iterative process. In contrast to the supervised learning methods, reinforce learning techniques only requires access to the performance results of the network given a configuration. That is, a fully mathematical model of the slices performances is not required. This characteristic makes this machine learning model adequate when dealing with difficult quality-of-experience metrics or complicated network deployment where simple metrics such as delay or rate might be influenced by different external agents.
A very precise location technique enables other interesting features like mobility prediction. To achieve this, multiple predictors being local or user-specific predictors must be combined together to make the prediction of the destination possible. The knowledge of the end-user profile, his or her habits and contextual information, like the day of the week, the time of the day and location, will be one of the key elements to make this happen. This mobility prediction is a key element and can be used as a “primitive” feature for many optimizations in wireless networks. Handover (HO), Received Signal Strength (RSS) or QoS in a wireless system for instance, can be anticipated accordingly with the estimated trajectory. Lastly, any AI tools and techniques will need to be bounded on both extremities to avoid any race conditions which may cause failures of a catastrophic type and sound alerts for human intervention.

2.5 Pre-emptive network automation

Communication Service Providers (CSP) own critical and valuable data which can be seen as the new oil for the industry. Unfortunately, and this for a long time, CSPs were unable to extract and leverage this data. Time has changed and CSPs have clearly understood the importance of these data jewels. On top of delivering value-added services, CSPs expect to differentiate themselves thanks to the QoE and QoS delivered to the subscribers. The old way of operating networks, where network administrators were applying configuration changes, were monitoring the outcome of the changes and were tuning up again the network, is still in function in operational centres. However, this optimization loop and process will not be enough to face the upcoming network revolution where video consumption will be at its highest and where very large number of objects of different nature will be connected to the network.
CSPs handle many daily customer interactions due to network quality issues. Alarm events can be notifications of network faults and provide indication of network quality degradation. Handling of alarms has great impact on both the operational costs and the quality of services offered to customers. The alarms are highly heterogeneous and this trend is steadily increasing as the network infrastructure is upgraded and expanded. Due to these factors, managing the alarms constitutes a challenging task that cannot be performed manually at an optimal manner. Therefore, through the appropriate management of such alarms it will be possible to proceed to improved customer management and assistance for optimal content delivery through preventive analytics and AI techniques. Specifically, it should be noted that the provisioning of a superior QoE to users is a very challenging task, as it depends not only on the content itself, but also on the network quality, at the specific location during the specific time. Moreover, the device status may also play a role. If users experience significant delays or other transmission issues, they will complain and may abandon the service permanently, leading to a loss of revenue for operators and content providers.
In fact, a differentiation in terms of QoE delivered by the operators to subscribers or partners will be achieved if network problems do not simply occur. The best approach is not healing or even self-healing, but predicting and thus preventing. One way to avoid network faults is to anticipate potential network degradation or outages. Predictive analysis of the performance of systems and networks, dynamic and proactive allocation of resources are critical components of next generation Network Manager Systems (NMS). The ability to collect, store, move, model data in a structured or

unstructured data storage and reliably managing data flows while monitoring, displaying and reporting, in real time, the performance of the network and its resources is really important to succeed in the 5G era. In order to achieve this goal, CSPs will have to leverage powerful reporting and/or closed-loop optimization systems maximizing the usage data collected from the network.
Some industry professionals are still skeptical about the benefits brought by AI in network operations. They all have good reasons to mention like:

- "*Why should we change our current process? If there is something wrong then we should fix it, no need to use AI!*" or
- "*What is the true value brought by AI?*" or also
- "*AI will be running in a black box, so if the network becomes faulty, it could be uncontrollable really quickly, so why should I introduce something that could damage my network with no means to re-stabilize it easily?*"

Some people think that rule-based automation systems (no learning capabilities) will be enough to deal with 5G challenges successfully. The automation of configuration, maintenance and optimization tasks work well and is a key step to reduce operational expenditures (OPEX) and cope with the increased complexity of the network. The mode of operation to automate these tasks is usually based on pre-defined, hand-crafted rulesets created during the network deployment or engineering phases. However, this mode has a major drawback: it is too static. This mode can be qualified as a sort of static intelligence or crafted knowledge while 5G requires agility, flexibility and a highly dynamic behavior.
This is where AI enters the dance: AI brings the power to transform the way networks are operated. It transforms networks that are descriptive, reactive into predictive and later to prescriptive networks thanks to efficient network analytics.
AI based location technique is exactly a use case where AI can definitely play a role and help delivering geo-fencing services using contextual insights, but also be an enabler for precise mobility predictions. Leveraging true intelligence to differentiate dynamically and accurately indoor vs. outdoor positions or predicting mobility paths would be a real plus for the industry and the market. The usage of AI helps in this case to identify connected devices outside or inside buildings with a higher accuracy.

2.6 Users Mobility Handling

The central role of AI in wireless is in how it handles users' and connected devices' mobility. This encompasses virtually all mechanisms and enablers discussed above.

Automated and autonomous use cases involve various degree of mobility, from automated factories to autonomous vehicles. This requires a level of operation, which is proactive and increasingly predictive. The use of AI allows modelling, enables recognition, provides predictive insights, and enforces action, in a highly dynamic and heterogeneous environment with potentially irregular patterns. The collection of data and use of learning methods are expected to enable these scenarios, while governed by network slicing, management and resource allocation, and integral to an architecture design and operation that meets use case requirements such as ultra-low latency and high reliability. Obvious examples include the mobility in robotics and those for autonomous vehicles.

An important consideration in the increasingly mobile and connected society is the enterprise mobility. This will demand new business and behaviour modelling, including workforce and organizational modelling that leverages AI in the context of user mobility. The research, innovations, standardization, and deployment scenarios in use of AI for wireless in general, and user mobility in particular, must also make it feasible and sustainable in large scale and in long term. This requires high performance, with high level of reliability, security and availability, at a cost and energy efficiency, appropriate for the particular service environment.

The enabling environment includes distributed intelligence such as interaction between central and localized clouds with edge clouds. This is in a potentially dense architecture with dynamic user density and irregular traffic patterns such as those in vehicular scenarios. Intelligent automated operation using AI mechanisms are required to make the future mobility possible, feasible, reliable and sustainable.

2.7 Summary

Wireless Topic	Threats	AI support
Spectrum management	✓ Guaranteeing an SLA to the spectrum incumbents in a dynamic environment.	✓ Spectrum assignment based on predictive QoS to incumbents.
Air interface	✓ Ultra-fast beam assignment in massive MIMO. ✓ Robustness to hardware impairments.	✓ Beam management via reinforcement learning. ✓ Data-based air interface design.
Radio access	✓ RACH decision and user scheduling in mIoT scenarios. Network slicing in the RAN	✓ Deep reinforcement learning based on the number of successful and unsuccessful transmissions.
Radio resource management	✓ Network slicing optimization in heterogeneous networks.	✓ Mimicking computationally demanding optimization techniques via deep learning regression.
Network management	✓ Currently unpredictable network failures and outages.	✓ Data based QoE predictions in order to pre-emptively manage the network malfunctionings.
Users mobility	✓ Communications to autonomous vehicles with heterogeneous degree of mobility.	✓ Data based predictable mobility patterns.

Table 1 Summary of potentials in deploying AI for wireless communications

3. Challenges of artificial intelligence for wireless communications

3.1 Deploying AI mechanisms

Future wireless networks will be characterized by an unprecedented level of complexity, which is beyond the capabilities of traditional design techniques in terms of both performance and complexity. A promising way to face this challenge, and thus an enabler of future wireless networks, is AI, and in particular the use of deep learning based on artificial neural networks (ANN), which has the potential to provide a novel design framework able to achieve near-optimal system performance with a complexity in line with the real-time constraints of online implementations [24]. However, this area of research and deployment is nascent and will progress rapidly in the next decade. The main barrier towards this goal is represented by the large amount of data required by ANNs approaches, which might be difficult to acquire in the wireless context, due to both privacy and cost problems. However, at the same time, the wireless application of ANNs also offers the possibility to significantly reduce the amount of measurements and live data that is needed. Indeed, as opposed to other fields of science where mathematical models are scarce and deep learning relies only on data, mathematical models for communication networks

optimization are very often available, even though they may be simplified and inaccurate. It was shown in [25],[26] that (possibly inaccurate) analytical models can be used in synergy with AI-based, data-driven methods, leading to a significant reduction of the amount of data required to carry out performing system designs, as well as striking a much better complexity-performance trade-off compared to traditional resource management techniques.
However, in order to enable these gains, an open issue is the integration of AI-based resource management in future wireless networks, which poses several challenges. A big question in this context is whether it is better to take a cloud/centralized approach, in which all data is stored in a central "artificial brain" where a large ANN carries out all computations and operates the whole network, or whether it is more convenient to take a distributed approach, spreading the intelligence across all network segments and devices. It is our opinion that, although a cloud-based AI would be conceptually simpler, its unique use is not viable for three major reasons:

- Already present networks are required to ensure strict end-to-end communication latencies, which will become stricter in the future. This makes cloud-based AI not feasible due to the additional delays to wait for cloud processing and feedback.
- Future wireless networks will have privacy and security as key requirements. This makes it problematic for end-user devices to share information/data with the cloud.
- In order to ensure ubiquitous service delivery, connectivity should be ensured even in areas and/or times in which only a poor connection to the cloud exists. Therefore, solely relying on cloud intelligence is not recommended.

For these reasons, a *mobile AI* approach is needed, where each network device will be endowed with intelligence, and will thus be able to self-organize based on local information and data. In this regard, we envision two possible approaches.

Distributed AI: Each network component (e.g. base station, access point, user equipment) is an independent decision-maker, which is equipped with its own ANN, that is trained by a local dataset that is built by local measurements. This approach does not require any interaction among the network infrastructure and edge-terminals, as far as data sharing and processing are concerned, thus significantly reducing feedback overheads. On the other hand, it makes it difficult to control the evolution of the whole wireless network and might lead to network failures. Letting each edge component operate based on local information and experience could potentially lead to inconsistencies in the behavior of the different network components, leading to malfunctioning and even network crashes. Research in this context should be aimed at developing new mechanisms to study the evolution of AI-based wireless networks, understanding whether suitable equilibria points naturally exist, or whether the system can be forced to evolve towards desirable operating points.
The figure below depicts a possible architecture of a cell wherein all devices associated with the access points are fully self-organizing through the use of their own ANN. No connection exists between the devices and the access points as far as ANN operation is concerned.

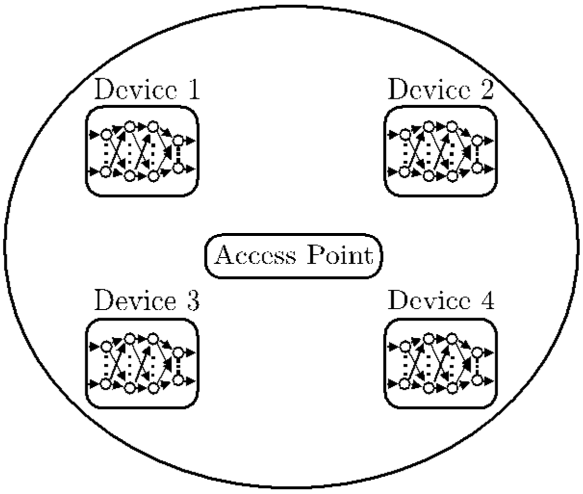

Figure 1 Distributed wireless infrastructure for AI

2) **Federated AI.** An alternative approach towards a mobile AI architecture is the framework of *federated learning* [22]. The aim is to balance between the purely distributed AI approach and the unfeasible and sole cloud-based AI approach. Federated learning distributes the data and computation tasks among a federation of local network devices that are coordinated by a central node. Like in the distributed AI approach, each member of the federation has its own ANN, which is locally trained based on local information/data. After this first phase, each member of the federation transmits to the coordinator the settings of its own ANN, which are then merged to come up with a refined model that is shared among all members of the federation. The amount of feedback between the local nodes and the coordinator should be tuned trading-off between the level of overhead that can be accepted and the performance that must be achieved.
The figure below depicts a possible architecture of a cell in a wireless network where the federated learning approach is employed. Each device of the federation, as well as the coordinator, has its own ANN. Each local device trains its own ANN based on local information, and then feeds it back to the coordinator information about the resulting ANN configuration. By leveraging these local configurations, the coordinator ANN is trained.

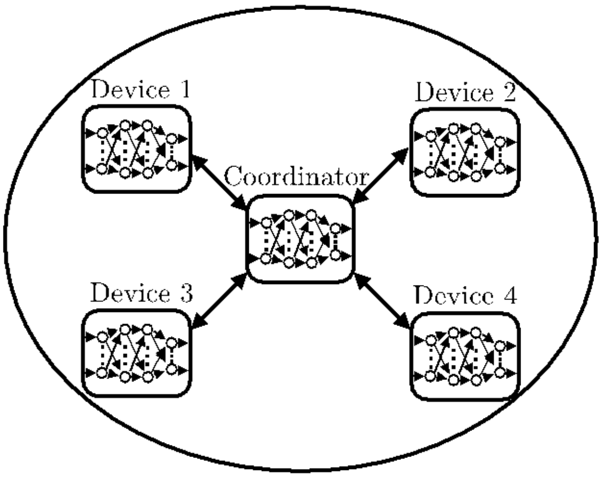

Figure 2 Federated wireless infrastructure for AI

As there is growing interest in machine learning (ML) and artificial intelligence (AI), ML relies on Big Data that is mined to gain information and knowledge. This approach is a reasonable candidate for example in detecting malicious behavior of a remote entity. There are other needs in networking that need “intelligence” such as self-configuration or managing complexity. Besides Big

Data, AI relies on the abundance of computing power (more on this in section 3.2). 6G will use the increasing computing power for coping with the higher bitrates but also for gaining added flexibility.

3.2 AI Challenges

Artificial intelligence in communication networks has a huge potential in telecommunications. Telecom operators see this technology as almost the unavoidable technology enabling to maintain or even reduce the OPEX significantly while delivering higher quality of service (QoS) to the end-users. Communication service providers will be looking at AI for various reasons. On top of enhancing their business ability, the adoption of AI will be driven by the increased complexity of the network. Indeed, with the advent of 5G, networks will be handling more spectrum, more and varied bandwidths, additional radio technologies, dealing with lower latencies enabling to reach new business territories like on-line tactile applications while benefiting from always more computing power. AI has the power to change the networks from reactive networks to predictive and finally proactive networks. All of these reasons will boost the wide spread use of AI in telecommunications.

However, AI is not seen as the holy grail by everyone. A few factors, not minor, may slow down the adoption of AI by the industry. The first reason may be the difficulty to fully assess the real benefits that AI can bring in daily work inspite of the emerging market trends to support it. Another reason is also human and emotional: people in the industry may (or may not) perceive AI as a potential threat to jobs. Rightly or wrongly, this is a reality the AI proponents will have to deal with. But even if the AI advocates overcome these barriers, the need to adapt existing operations processes will be a heavy, cumbersome and tedious task and will need support from the top management in organisations. An additional obstacle will be the expected transparency of AI. Indeed, delivering AI as a black box may be enough in some cases for basic actions (e.g. like how to optimize parameter settings during a roll-out phase) but would not be acceptable during certain (such as healing) phases. Operators who could not be in the position to explain what happened and why it happened could be liable. Herein lies the next obstacle, which is the training data set accuracy and inclusiveness, which might result in biased decision making and can be detrimental to the company and its customer base. That is why AI that can be explained in plain language would be necessary. All these factors will need, to be addressed for a successful adoption of AI by the market and the telecommunications industry.

It is important to note that in the dynamic wireless transmission scenarios of the future, the training overhead for learning based AI solutions may pose a particular challenge. As an example, typical learning tools available have been developed for applications such as computer vision, speech recognition, or natural language processing, where training time and overheads are not a key factor. In a dynamic wireless environment where the transceivers will need to adopt to changing channels, hardware responses, or event-driven dynamic link connectivity, the overheads of re-training the system are quite significant. Solutions inspired by layered training such as incremental or reinforcement learning will inevitably become key in addressing such environments.

The widespread adoption of AI poses severe challenges to the speed and power consumption of existing computing systems. A recent Nature editorial put it succinctly by asking the question “Does AI have a hardware problem?”, the author’s answer being a definitive “yes” [27]. Present day von Neumann computing architectures require constant shuffling of data between storage and CPU, providing a critical performance bottleneck. Most clock cycles are wasted in moving data rather than computing, while physical separation of memory and processing builds in latency. Machine Learning (ML) systems, which strive to mimic some functions of the brain, have a significant power cost. While the human brain expends ~20W, large ML systems can consume tens of kW. Scaling up becomes prohibitive; for example, a simulation of a neural network approaching the complexity of the human brain (10^{10} neurons and 10^{14} synapses), running on the Lawrence Livermore Sequoia supercomputer, consumed 7.9MW [28]. This six order of magnitude mismatch in power efficiency is largely because of the “von Neumann bottleneck”. Optimised GPUs and Tensor Processing

Units (TPUs) offer significant benefits, but still consume orders of magnitude more power than biological systems when performing similar tasks, and do not offer the fault and noise tolerance of such systems.

The International Technology Roadmap for Semiconductors (ITRS) and the US Department of Energy Office of Science recognise this as a key challenge, stating in 2015: "*Well-supported predictions … indicate that conventional approaches to computation will hit a wall in the next 10 years … Novel approaches & new concepts are needed…*" [29].

As we begin to deploy AI systems both at the heart of communication systems and as edge computing elements for the IoT, such power consumption issues will severely limit the possibility of integrating AI and communication unless significant effort is directed to more efficient hardware solutions. While these will undoubtedly include further optimisation of coprocessors (GPUs, TPUs), more ambitious solutions, such as non-von Neumann approaches, are required in the long term.

Challenges
✓ Need to adapt the network operations processes. ✓ AI perceived as a threat for employment in the telecom industry. ✓ AI may act as a black box possibly raising liability issues between parties if mechanisms in place cannot explain what happened and why. ✓ Power consumption of existing AI hardware is too high to support both large-scale rollout or distributed edge computing for IoT. ✓ Definition and technologies not yet matured for industry consensus. ✓ KPIs and standards not in place to drive implementation and inter-operability of AI across CSP domains and across network elements from different vendors.

Table 2 Summary of challenges in deploying AI for wireless communications

3.3 Ethics in the use of AI in the Wireless Infrastructure

Artificial Intelligence in future will be pervasive in all aspects of our lives. Each day we hear new applications for AI. There is growing use of artificial intelligence (AI) by businesses to collect data which can be analysed to analyse customer behaviour. AI can collect and store information about individuals, including their profiles, payment histories and location information, allowing companies to mine that rich data. This definitely leads to security and privacy, which must be addressed as a number one priority along with the improvements in the legal and policy framework.
In recent months and weeks, some important studies have been done on the topic of ethics. One such study was conducted by the French Commission Nationale de l'Informatique et des Libertés in 2017. Amongst the six (6) policy recommendations elaborated by the CNIL, the recommendation #2 is directly related to the usage of AI in wireless networks: "*Making algorithmic systems comprehensible by strengthening existing rights and by rethinking mediation with users*".

The latest report on ethics has been issued by the European Union in April 2019. The EU has released a report including recommendations and requirements to achieve trustworthy artificial intelligence: "*AI Ethics Guidelines for Trustworthy AI – April 2019*". In this comprehensive report, one requirement on "Privacy and Data Governance" relates to the point raised above.

A few key additional needs apply directly to wireless networks as well. To realize trustworthy AI-based wireless networks, the following requirements will have to be fulfilled: 1) Technical Robustness and Safety, 2) Privacy and Data Governance, 3) Transparency and 4) Accountability.

Both reports, the one from the French CNIL and the one from the European Union, highlight the need of explainability, transparency and accountability. These requirements are paramount with the advent of 5G. 5G will enable new businesses and address vertical ones such as the automotive, transport and logistics or health industries for instance. Accountability and liability will be of the utmost importance. In case problems arise, the 5G network provider will have to explain what happened, why it happened and when the decision was taken. As a consequence, network providers will need to have access to logs, traces and understand the type of algorithms used while having access to the details to the logics that resulted in certain actions. However, it has to be noted that not all actions empowered by AI taken by the network will be needed to be fully explainable. For example, fault monitoring and analysis enhanced by AI to reveal concealed anomalies may not need to deliver a full detailed explanation of the intelligence used. On the other hand, an algorithm guaranteeing a very low latency delivered by a 5G network instantiation (5G slice) for health care and remote surgery for instance, will need to deliver details in case any issue arises due to legal liabilities given that AI does not work in a lawless world. It is the same for connected cars or intelligence transportation. As such, legal responsibilities require that the network powered by AI must be trustworthy and explainable. In short, a complete transaction trail will need to be captured and retained.

Next generation wireless networks will continue to play a major transformative role and deliver technical and social benefits to the society. AI will undoubtedly enhance the way wireless networks behave and position them as the central nerve system of the future economy. But the success of it will depend on secured and trustworthy intelligence guaranteeing the privacy of everyone.

4. References

[1] L. Mfupe, F. Mekuria, M. Mzyece "Geo-location White Space Spectrum Database: Models and Design of South Africa's First Dynamic Spectrum Access Coexistence Manager", *KSII Transactions on Internet and Information Systems (TIIS),* Journal, Vol 8, No. 11, pp. 3810-3836, 2014.

[2] "DARPA Spectrum Collaboration Challenge SC2." www.spectrumcollaboration challenge.com.

[3] T. Irnich, J. Kronander, Y. Selen, and G. Li, "Spectrum sharing scenarios and resulting technical requirements for 5G systems," IEEE Int. Symp. Pers. Indoor Mob. Radio Commun. PIMRC, pp. 127–132, 2013.

[4] T. O'Shea and J. Hoydis, "An Introduction to Deep Learning for the Physical Layer," *IEEE TCCN*, 2017.

[5] M. Chen, et al. "Machine learning for wireless networks with AI: A tutorial on NNs.", 2017 arxiv.org/abs/1710.02913

[6] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of things (IoT): A vision, architectural elements, and future directions," Future Gener. Comp. Sy., vol. 29, no. 7, pp. 1645–1660, Sep. 2013.

[7] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of things: A survey on enabling technologies, protocols, and applications," IEEE Commun. Surveys Tuts., vol. 17, no. 4, pp. 2347–2376, Nov. 2015.

[8] N. Jiang, Y. Deng, A. Nallanathan, X. Kang, and T. Q. S Quek, ''Analyzing Random Access Collisions in Massive IoT Networks', IEEE Trans. on Wireless Communications, vol. 17, no. 10, pp. 6853-6870, Oct. 2018.

[9] N. Jiang, Y. Deng, X. Kang, A. Nallanathan, "Random Access Analysis for Massive IoT Networks under A New Spatio-Temporal Model: A Stochastic Geometry Approach", IEEE Trans. on Communications, vol. 66, no. 11, pp. 5788–5830, Nov. 2018.

[10] J. Schlienz and D. Raddino, "Narrowband internet of things whitepaper," IEEE Microw. Mag., vol. 8, no. 1, pp. 76–82, Aug. 2016.

[11] Y.-P. E. Wang, X. Lin, A. Adhikary, A. Grovlen, Y. Sui, Y. Blankenship, J. Bergman, and H. S. Razaghi, "A primer on 3GPP narrowband internet of things (NB-IoT)," IEEE Commun. Mag., vol. 55, no. 3, pp. 117–123, Mar. 2017.

[12] R. Sutton and A. Barto, "Reinforcement learning: An introduction (draft)," URl: http://www.incompleteideas.net/book/bookdraft2017nov5.pdf, 2017.

[13] V. Mnih et al., "Human-level control through deep reinforcement learning," Nature, vol. 518, no. 7540, p. 529, Feb. 2015.

[14] N. Jiang, Y. Deng, A. Nallanathan, and J. A. Chambers, ''Deep Reinforcement Learning for Real-Time Optimization in NB-IoT Networks', IEEE J. Sel. Areas Commun., 2018.

[15] X. Zhou, R. Li, T. Chen and H. Zhang, "Network slicing as a service: enabling enterprises' own software-defined cellular networks," in *IEEE Communications Magazine*, vol. 54, no. 7, pp. 146-153, July 2016.

[16] S. Vassilaras et al., "The Algorithmic Aspects of Network Slicing," in IEEE Communications Magazine, vol. 55, no. 8, pp. 112-119, Aug. 2017.

[17] B. Han, J. Lianghai and H. D. Schotten, "Slice as an Evolutionary Service: Genetic Optimization for Inter-Slice Resource Management in 5G Networks," in IEEE Access, vol. 6, pp. 33137-33147, 2018.

[18] L. Zanzi, F. Giust and V. Sciancalepore, "M2EC: A multi-tenant resource orchestration in multi-access edge computing systems," 2018 IEEE Wireless Communications and Networking Conference (WCNC), Barcelona, 2018, pp. 1-6.

[19] Q. Mao, F. Hu and Q. Hao, "Deep Learning for Intelligent Wireless Networks: A Comprehensive Survey," in IEEE Communications Surveys & Tutorials, vol. 20, no. 4, pp. 2595-2621, Fourthquarter 2018.

[20] H. Sun, X. Chen, Q. Shi, M. Hong, X. Fu and N. D. Sidiropoulos, "Learning to Optimize: Training Deep Neural Networks for Interference Management," in IEEE Transactions on Signal Processing, vol. 66, no. 20, pp. 5438-5453, 15 Oct.15, 2018.

[21] Wei Cui, Kaiming Shen, and Wei Yu, "Spatial Deep Learning for Wireless Scheduling", IEEE Global Communications Conference (Globecom), December 2018.

[22] Zhao, Z., Li, R., Sun, Q., Yang, Y., Chen, X., Zhao, M. and Zhang, H., 2018. Deep Reinforcement Learning for Network Slicing. arXiv preprint arXiv:1805.06591.

[23] Chen, X., Zhao, Z., Wu, C., Bennis, M., Liu, H., Ji, Y. and Zhang, H., 2018. Multi-Tenant Cross-Slice Resource Orchestration: A Deep Reinforcement Learning Approach. arXiv preprint arXiv:1807.09350.

[24] A. Zappone, M. Di Renzo, M. Debbah, "Wireless Networks Design in the Era of Deep Learning: Model-Based, AI-Based, or Both?", IEEE Transactions on Communications, (invited paper), 2019, [Online] http://export.arxiv.org/abs/1902.02647.

[25] A. Zappone, M. Di Renzo, M. Debbah, T. T. Lam, X. Qian, "Model-Aided Wireless Artificial Intelligence: Embedding Expert Knowledge in Deep Neural Networks Towards Wireless Systems Optimization", 2018, [Online] https://arxiv.org/abs/1808.01672.

[26] J. Konečný, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, "Federated learning: Strategies for improving communication efficiency," 2017, [Online] https://arxiv.org/pdf/1610.05492.

[27] Does AI have a hardware problem?", Nature Electronics (Editorial) vol. 1 no. 4, 205, 2018

[28] T. Wong et al, *IBM Res. Rep. RJ10502* (2012)

[29] I.K. Schuller & R. Stevens, "Neuromorphic computing: from materials to systems architectures", https://science.energy.gov/~/media/ascr/pdf/programdocuments/docs/Neuromorphic-Computing-Report_FNLBLP.pdf 2015.

[30] https://www.ryerson.ca/content/dam/pbdce/papers/Response_to_AI_'Explainability'.pdf

5. Imprint

Wireless World Research Forum
c/o Format A AG
Pfingstweidstrasse 102b
CH-8005 Zürich

Secretariat:
Vinod Kumar
Alcatel-Lucent France
Centre de Villarceaux
Route de Villejuste
91 620, NOZAY
France
e-Mail: vinod.kumar@alcatel-lucent.com
Phone : + 33 1 30 77 27 37
Fax : + 33 1 30 77 61 75

The WWRF is a non-profit organisation registered in Switzerland

Chairman of the Forum:
Dr. Nigel Jefferies

Editor-in-Chief: Mr Sudhir Dixit

The WWRF Outlook Visions and research directions for the Wireless World

ISSN 1662-615X is published non-periodically by the Wireless World Research Forum http://www.wwrf.ch

Responsibility for the contents rests with the Steering Board of the Forum.