JOURNAL OF INNOVATIVE ACADEMIC RESEARCH
ISSN : 3049-2122
VOLUME: 02 ISSUE: 01
RECEIVED: JANUARY 23 REVISED: FEBRUARY 05 ACCEPTED: FEBRUARY 21 PUBLISHED: MARCH 17

Agentic AI for Self-Adaptive Semiconductor Architectures in Intelligent Wireless Networks

Sophie Dubois
Asst Professor

***Abstract*—Intelligent wireless network systems must adapt in real time to channel conditions, mobile users, and traffic demands while complying with strict performance specifications. The presented study describes a foundation for agentic AI-driven adaptive semiconductor architectures, taking into account both technology and system-level considerations. With respect to technology, it focuses on recent semiconductor architectural trends inspired by early developments in AI computing. In particular, it examines reconfigurable and programmable hardware, neuromorphic processors, ternary and binary computing, in-memory and near-memory computing, and 3D integration. From a system perspective, it identifies mechanisms supporting real-time adaptation: spectrum and mobile resource management, context-aware modulation and coding schemes, and end-to-end control loops. In addition, it discusses AI-driven development toolchains, addressing tool automation, data privacy, and security issues. Finally, it proposes metrics, benchmarks, and evaluation protocols enabling a holistic assessment of adaptive semiconductor architectures for intelligent wireless networks. Attention to both technology and systems can steer forthcoming technology development towards the ultimate intelligent wireless systems and networks targeted by 6G and beyond.**

Keywords—Intelligent Wireless Networks, Agentic AI, Adaptive Semiconductor Architectures, Real-Time Network Adaptation, Reconfigurable and Programmable Hardware, Neuromorphic Processors, In-Memory and Near-Memory Computing, Ternary and Binary Computing, 3D Integration, Spectrum and Mobile Resource Management, Context-Aware Modulation and Coding, End-to-End Control Loops, AI-Driven Development Toolchains, Secure Hardware–Software Co-Design, Performance Metrics and Benchmarks, Evaluation Protocols, 6G Systems, Intelligent Network Control, Hardware–AI Co-Design, Autonomous Wireless Systems.

I. INTRODUCTION

Wireless connectivity is ubiquitous and permeates modern societies in a way that was unimaginable a decade ago. Yet, supporting the growing connectivity demand in a sustainable way for the diverse set of applications emerging is increasingly challenging. The challenge is in the design of the communication systems and networks, ranging from their circuitry to support the various load requirements, to their architecture that enables efficient resource sharing in a users' crowded environment. A promising approach to alleviate these issues is by offloading intelligence away from the devices operating on the edge of these systems or networks and using a machine-learning model to support the predictive or optimal resource management at a central or multi-agent control unit. The period available to adopt and process changes detected in the wireless environment dictates how quickly the solution converges toward the optimum. For actuality, the edges cannot deviate from the current operation provided by the communication cluster while the resource management is adapted. Therefore, in these schemes, it is often assumed that the resource management operation is slower than its tracked performance.

Nevertheless, there is a trend where the need for fast adaptation cannot be ignored, yet is often neglected. Solutions where agents at the edges also adapt the operation of their communication unit have appeared. Examples include real-time adaptation of the modulation and coding schemes used; agent-based spectrum management where the units themselves dynamically switch between spectrum bands; and the optimization of the control loop between the resource management algorithm operating on a group of communication clusters and the wireless devices associated to these clusters, using the concept of inner loop transfer

JOURNAL OF INNOVATIVE ACADEMIC RESEARCH
ISSN : 3049-2122
VOLUME: 02 ISSUE: 01
RECEIVED: JANUARY 23 REVISED: FEBRUARY 05 ACCEPTED: FEBRUARY 21 PUBLISHED: MARCH 17

function. The solutions appearing in these categories reveal that the time period available for the adaptation of the communication units is comparable to the time needed for adaptation of the resource manager. Yet, the significance of this operation and the concurrence of the adaptation processes are not always considered in the design methodologies.

A. Overview of the Study and Its Significance

Intelligent wireless systems will rely on an AI-driven ecosystem at all phases of the AI lifecycle—from on-device data acquisition to the development of supercomputing models through federated learning. At the edge, however, real-time requirements and hardware resource constraints necessitate the deployment of extremely lightweight ML models, e.g., sparsity in parameter space, low latency in inference, etc. Yet these constraints are often counterbalanced with privacy issues, arising from unencrypted data sharing with central cloud-based platforms. Addressing these conflicting requirements will alter traditional design goals. The pressure to serve multiple tasks at low latency with minimal overhead places severe strain on conventional communication systems and their corresponding semiconductor implementations. Combined with the need to provide communication services for an ever-increasing number of devices while meeting ever-tougher energy efficiency and environmental criteria, the design methodology has come under pressure to adapt into consideration all three cornerstones.

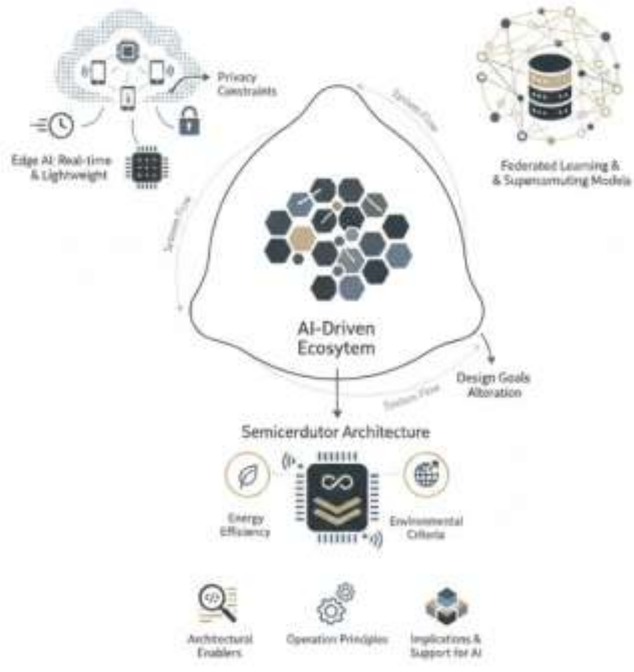


Fig 1: Semiconductor Architectural Enablers for Edge-Centric AI: Reconciling Latency, Privacy, and Energy Constraints in Next-Generation Wireless Systems

This research study explores the concrete implications on the semiconductor architecture of wireless devices due to the need to operate with an AI-driven decision layer at the fore. A series of key architectural enablers for next-generation intelligent radio access technologies are sourced from the literature, analyzed for their abreaction advantages, and assembled into a high-level overview. Continuity and coherence of the proposed driving directions are then verified through an assessment of the implications for the development of wireless semiconductor solutions. The resulting survey is not intended to be exhaustive in nature; rather, it aims to provide a consolidating overview of selected semiconductor architectural solutions, their operation principles, and implications, all articulated in terms of support for the AI-driven concept.

JOURNAL OF INNOVATIVE ACADEMIC RESEARCH

ISSN : 3049-2122

VOLUME: 02 ISSUE: 01

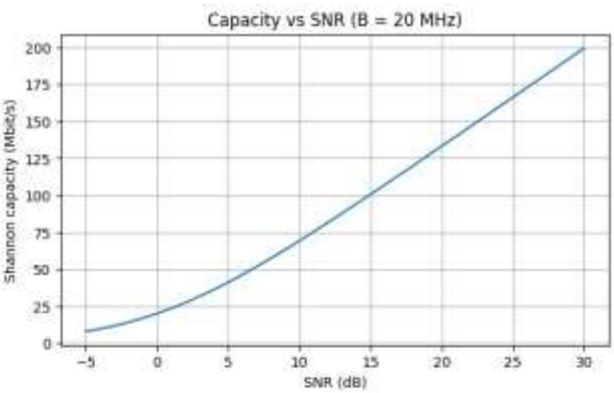
RECEIVED: JANUARY 23 REVISED: FEBRUARY 05 ACCEPTED: FEBRUARY 21 PUBLISHED: MARCH 17

II. FOUNDATIONS OF AI-DRIVEN SEMICONDUCTOR ARCHITECTURES

The concept of agentic computing is rooted in the ideas of self-management, self-adaptation, and self-organization within complex systems. These principles have been explored recently in increasing detail in relation to semiconductor architectures and their low-level computational operating modes. They directly influence the proposed AI-driven tool chains for next-generation semiconductor architectures, where Computer Science itself becomes a small part of the design process of an intelligent system.

It is generally accepted that classical computing architectures have reached a point of diminishing returns in terms of performance increases resulting from shrinking technologies. Even without considering the physical limits arising from quantum mechanics, modern processor architectures are complicated and power-hungry. Connections between cores and chips need to traverse significant distances across the silicon die. Data exchanges increasingly dominate the overall processing cycle time. Even more challenging, a many-core design offers only a limited speed-up when using algorithmic-level parallelism. A range of specialized semiconductor architectures are emerging. These include FPGA, neural network processors, domain-specific ASICs, resistive memory devices, and many more, designed for specific low-level modes of computation. However, the limited flexibility of these architectures and the increasing difficulty of designing ASICs are resulting in a lack of small-volume applications.

With the advent of AI, the ideas of reconfigurable hardware design are being reevaluated. While traditional computing architectures are limited to algorithmic-level parallelism at increasing power levels, computer-executed decision making with the associated latency in mobile communication can be viewed as a more severe limitation. Intelligent behavior therefore can be supported by user-defined adaptive circuitry not only at radio frequency transmission but also at lower levels of Intelligent Networks. Such additional circuitry, enabling Context-Aware or User-Adaptive Modulation and Coding, has only been demonstrated with a neural net process that requires massive amounts of data and time to train. An alternative view is that AI-drivenness creates an opportunity for the parallelization of decision making. Optimizing these decisions requires AI within the process and can allow a much smaller dedicated functionality, even when emulated in software.



Equation 1) Core wireless link equations used by adaptive modulation/coding

1.1 Baseband receive model → SNR

Step 1 (signal model): for one symbol,

$$y = hx + n$$

- x: transmitted symbol (avg. power $P_s = \mathbb{E}[|x|^2]$)

JOURNAL OF INNOVATIVE ACADEMIC RESEARCH
ISSN : 3049-2122
VOLUME: 02 ISSUE: 01
RECEIVED: JANUARY 23 REVISED: FEBRUARY 05 ACCEPTED: FEBRUARY 21 PUBLISHED: MARCH 17

- h: channel coefficient
- n: noise, typically $n \sim \mathcal{CN}(0, \sigma_n^2)$

Step 2 (received signal power):

$$P_{\text{rx,signal}} = \mathbb{E}[|hx|^2] = |h|^2 \mathbb{E}[|x|^2] = |h|^2 P_s$$

Step 3 (noise power):

$$P_{\text{rx,noise}} = \mathbb{E}[|n|^2] = \sigma_n^2$$

Step 4 (SNR definition):

$$\text{SNR} = \frac{P_{\text{rx,signal}}}{P_{\text{rx,noise}}} = \frac{|h|^2 P_s}{\sigma_n^2}$$

If noise is modeled via noise spectral density N_0 over bandwidth B, then $\sigma_n^2 = N_0 B$, giving:

$$\text{SNR} = \frac{|h|^2 P_s}{N_0 B}$$

1.2 SNR → Shannon capacity (upper bound that motivates adaptive MCS)

The maximum achievable rate (ideal coding) is:

$$C = B \log_2(1 + \text{SNR})$$

Step-by-step meaning:

1. For fixed B, as SNR increases, $1 + \text{SNR}$ increases.
2. $\log_2(\cdot)$ converts the growth into “bits per second per Hz”.
3. Multiply by B (Hz) to get **bit/s**.

A. Agentic Computing Paradigms

Co-designing Kernel and Microtask Operations in a System-on-Chip Platform

Agentic computing paradigms offer an alternative means of expressing such widened flexibility. Here, the device contains multiple characters that are capable of performing complex sensing, decision-making, acting, communicating, or controlling tasks on behalf of an agent. Each character is a program that resides in the device’s memory. Each character runs in its own thread and has its own memory stack, CPU, timing attributes, and associated devices. Furthermore, part or all of the memory hierarchy can work as multiple individual incoherent system memories. Given limited performance constraints, requirements of low or intermittent data traffic, low latency, and low energy consumption, the core architecture of overall platforms can remain unchanged. Only the intra-chip connection of logic and memory resources needs to be adapted accordingly. An agent is therefore a CAD design code capable of determining and describing how these characters interact.

JOURNAL OF INNOVATIVE ACADEMIC RESEARCH
ISSN : 3049-2122
VOLUME: 02 ISSUE: 01
RECEIVED: JANUARY 23 REVISED: FEBRUARY 05 ACCEPTED: FEBRUARY 21 PUBLISHED: MARCH 17

In agentic computing, the flexibility of programmability is preserved, but the overhead and additional cycles due to the programmability feature of conventional microprocessors are reduced. The lowered programmability overhead allows high performance for wire-speed realtime applications. Furthermore, resources can be stably operationally fast even as resources are flexibly reshaped and component blocks of the platform are played out as agents if the full programmability capability has not been exercised.

B. Reconfigurable and Programmable Hardware

Reconfigurable and programmable semiconductor hardware features architectures with silicon elements that can undergo reconfiguration or rely on software. When it comes to hardware design, operated in conjunction to achieve higher-level goals, the design methodologies behind these two categories—field-programmable gate arrays (FPGAs) and application-specific integrated circuits (ASICs)—exhibit trade-offs in terms of flexibility, acceleration, and energy consumption. The strong flexibility of FPGAs allows developers to test out ideas and prototypes of logic-intensive tasks, enabling applications such as 5G testbeds. However, latency and energy costs remain a major challenge for computation-intensive tasks. These trade-offs are being eliminated by hybrid designs with both ASICs and FPGAs, semi-custom designs that allow the integration of the administrator-specified functionalities with the usual devices, dynamic reconfiguration of FPGAs within limited regions, and scaling up through multiple-hierarchy designs.

Compared to non-reconfigurable semiconductor devices or ASICs, reconfigurable hardware architectures have greater flexibility to adapt to changes in workloads, inputs, and operating conditions. The workload characteristic includes the degree of similarity for updates over time, which reveals that the design for one time slot is not completely useless for subsequent time slots, and the suitable timing condition for performing reconfigurations can be estimated and controlled without large overheads. Inputs refer to the information entering the design at runtime, conditions are associated with the operation environment, and the degrees of freedom within the design are usually provided by the components with smaller area ratios. Adapt-level and adapt-ease metrics can be defined for any two different operational designs, for example, for the RF front-end design under different frequency requirements, for different MIMO channel states, for high-dynamic-range nonlinear-power-amplifier designs, for different QPSK-modulation-and-demodulation designs, and for spatio-temporal-coding-inferred designs under different spatio-temporal-coding-inferred spills.

III. ARCHITECTURAL TRENDS FOR NEXT-GENERATION INTELLIGENT NETWORKS

Next-generation intelligent wireless networks with their extensive sensing capabilities and premises of agile and real-time response are expected to exploit the power of AI for enhancing end-user performance, while AI processing is performed at the edge. The supporting hardware and system architecture of such networks need to be appropriately architected for efficient execution. Relevant architectural trends specifically focusing on AI workloads within the context of intelligent bridging networks are addressed. First, new silicon technologies based on neuromorphic and spiking-based processors are surveyed. The analysis identifies their suitability for wireless workloads and their attractive energy footprint. The expected low-energy advantage in training is also highlighted. Next, in-memory and near-memory computing concepts are reviewed. Their implied data locality, drivers for improved bandwidth provisioning, and stance on cache coherence are examined. Finally, the main research challenges in realizing these concepts at scale in a flexible manner for intelligent networks are underlined.

AI/ML workloads and the state-of-the-art performance attained are reviewed. Energy efficiency, latency, throughput, and silicon performance awareness are then established as relevant metrics and evaluation protocols. Subsequently, the focus is placed on adaptability, robustness, and security, with the observed diversity in threats across different wireless scenarios highlighted. The currently explored concepts and available publications provide the broad foundation for an evaluation framework to guide future research in the area.

	Energy efficiency (norm)	Latency score (norm)	Adaptability (norm)
ASIC	1.0	1.0	0.4
FPGA	0.6	0.8	1.0

JOURNAL OF INNOVATIVE ACADEMIC RESEARCH
ISSN : 3049-2122
VOLUME: 02 ISSUE: 01
RECEIVED: JANUARY 23 REVISED: FEBRUARY 05 ACCEPTED: FEBRUARY 21 PUBLISHED: MARCH 17

	Energy efficiency (norm)	Latency score (norm)	Adaptability (norm)
Neuromorphic	1.2	0.7	0.7
In-memory	1.1	0.9	0.6

A. Neuromorphic and Spiking-Based Processors

Neuromorphic and spiking-based processors have recently drawn increasing research interest as hardware architectures that could be efficiently tailored for the wireless communication workloads anticipated in future intelligent networks. Using hardware architectures that are specialized for a certain data processing task, ideally the one that is predominant in the system workload, has proven to be advantageous in terms of energy and performance. Moreover, energy efficiency has become one of the most important design objectives in the development of hardware systems due to the increasing energy demand of mobile communications. Still, many applications require devices to have some degree of flexibility in terms of processing requirements, and the trade-off between performance and energy has become the major concern in such cases. Therefore, there is also significant interest in the development of more general-purpose chip architectures that offer great adaptation capabilities for a variety of workloads, such as artificial neural networks through dedicated accelerators in the form of application-specific integrated circuits.

Using bias in the performance and energy efficiency of wireless communication workloads to push the architectural design towards specialization, neuromorphic and spiking-based processors might present a greater energy performance ratio than CNN-based accelerators and thus be better suited for mobile and IoT applications. The trade-off likely shifts in favor of spiking based architectures with more complex wireless workloads (e.g., those involving higher-layer functions or lower-layer functions with closed loop feedback) that also necessitate a level of processing flexibility and a non-negligible adaptation time. Factors that push in favor of neuromorphic and spiking-based processors are the high energy-efficiency and scalability prospects, assuming an adaptable mapping of the communication functions onto the hardware. At the same time, such architectures are not ideal for workloads with very low latency constraints, and their applicability is therefore likely limited to more relaxed throughput conditions.

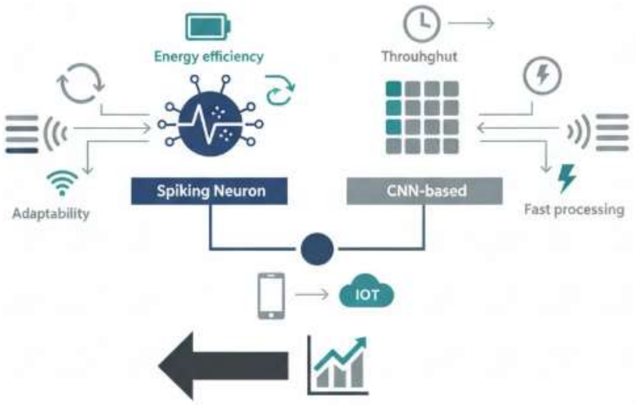


Fig 2: Synergistic Trade-offs in Neuromorphic Architectures: Evaluating Spiking Neural Networks for Energy-Efficient Adaptive Wireless Workloads

B. In-Memory and Near-Memory Computing

Bulk movement of data between processors and main memory is known to contribute significantly to energy and latency overhead, especially for real-time workloads. In-memory computing aims to alleviate the energy costs associated with data

JOURNAL OF INNOVATIVE ACADEMIC RESEARCH
ISSN : 3049-2122
VOLUME: 02 ISSUE: 01
RECEIVED: JANUARY 23 REVISED: FEBRUARY 05 ACCEPTED: FEBRUARY 21 PUBLISHED: MARCH 17

movement by embedding processing elements in memory so that data access and computation can be performed simultaneously. By reducing the distance between memory and logic, in-memory computing architectures eliminate the energy overhead associated with data movement and facilitate memory-bound computations with lower energy consumption.

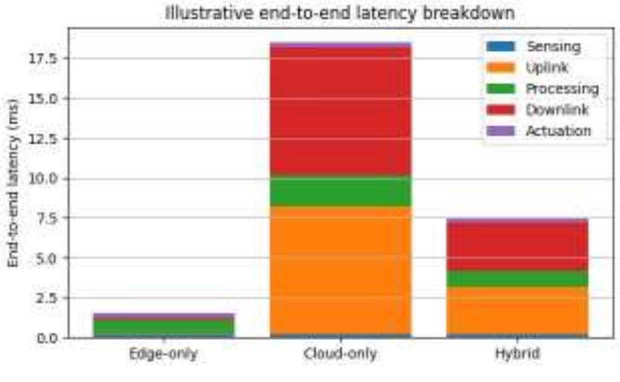
Near-memory computing extends this idea by placing processing elements physically close to the memory array, albeit not in the array itself, to take advantage of the memory bandwidth benefit with less hardware overhead than a full in-memory computing solution. Nevertheless, data must still be routed out of the memory and into the processing elements, and this data movement incurs energy overheads that grow as the data-width of the communication channels increases. Near-memory logic carries lower communication overhead than near-data logic, although the effective gain diminishes as the ratio between the width of the processing elements and the memory increases.

For many workloads, however, the data movement and bandwidth overheads persist even with in-memory and near-memory support, and these architectures are best suited to workloads exhibiting good data locality. Above an optimal locality, the additional data movements negate the benefits offered by the added processing capacity. Furthermore, for workloads consisting of multiple operations on the same data, the shared nature of the memory bandwidth causes satellite operations to incur significant latency in accessing the data. Nevertheless, there are several classes of applications that do benefit from high data locality, including Neural Network Inference and Convolutional Neural Networks – both of which are emerging as promising candidates for ultralow-power communications and signal processing solutions.

IV. ADAPTIVE MECHANISMS FOR WIRELESS INTELLIGENCE

With an increasing number of equipped wireless devices and growing data traffic demands, the utilization of communication spectral resources has become even more difficult. Furthermore, the wireless resources for an intelligent wireless network spanning a large area are mobile resources, which need to be handled in real time. In general, compared to the average-level traffic, the peak-level traffic of each functional block is different, and the processing latencies of each functional block are different as well. Thus, designing the reaction time of the control loop needs great caution. For example, if real-time spectrum management is carried out without considering the control lag effect, it may cause resource wastage. The efficiency of real-time spectrum management and number planning for intelligent wireless networks can be improved when the reaction time of resource management links is optimized and allocated.

On the other hand, designing adaptive context-aware modulation and coding schemes to deal with a fast-changing channel is not a straightforward task. The reason is that modulator and demodulator pairs for different iterations of the context-aware modulation and coding scheme are not the same, which means that the channel conditions for different iterations of the context-aware modulation and coding scheme are also different. The training and construction of such a scheme require the knowledge of a very large number of channel realizations. How to optimize the training process for context-aware modulation and coding schemes remains open, and it is also open for other coding schemes that may require large training overhead.



JOURNAL OF INNOVATIVE ACADEMIC RESEARCH
ISSN : 3049-2122
VOLUME: 02 ISSUE: 01
RECEIVED: JANUARY 23 REVISED: FEBRUARY 05 ACCEPTED: FEBRUARY 21 PUBLISHED: MARCH 17

Equation 2) End-to-end latency equations for "real-time adaptation" and control loops

2.1 Latency decomposition

Total end-to-end latency is commonly decomposed as:

$$L = L_q + L_{tx} + L_{prop} + L_{proc} + L_{rx}$$

Where:

- L_q: queueing/scheduling delay
- L_{tx}: transmission time
- L_{prop}: propagation delay
- L_{proc}: compute/inference/decoding time
- L_{rx}: receive + buffering + delivery delay

Step-by-step for transmission time (packet of S bits on rate R bit/s):

$$L_{tx} = \frac{S}{R}$$

Step-by-step for propagation (distance d, wave speed $v \approx c$):

$$L_{prop} = \frac{d}{v}$$

2.2 Control-loop timing constraint (reaction time vs environment dynamics)

A minimal control-loop sampling model:

- loop executes every T_s seconds
- end-to-end loop delay L_{loop}

A practical real-time requirement is:

$$L_{\text{loop}} \ll T_{\text{dyn}}$$

where T_{dyn} is the timescale of environment change (e.g., channel coherence time, traffic burst timescale).

If the loop is discrete-time, a common "safe" condition is:

$$T_s \leq \frac{T_{\text{dyn}}}{k} \quad \text{with } k \in [5,10] \text{ (rule-of-thumb)}$$

JOURNAL OF INNOVATIVE ACADEMIC RESEARCH
ISSN : 3049-2122
VOLUME: 02 ISSUE: 01
RECEIVED: JANUARY 23 REVISED: FEBRUARY 05 ACCEPTED: FEBRUARY 21 PUBLISHED: MARCH 17

A. Real-Time Spectrum/Mobile Resource Management

Complementary AI-based reflective mechanisms enable timely and intelligent selection of spectrum and mobile resources. It is expected that the relevant recurrent decisions would follow a real-time pattern in order to deal with the dynamics of the wireless environment. Thus, the overall structure of the process draws on the concept of control systems and must cope with several key elements: a decisive trigger with enthusiasm for a quick choice; a latency that is small enough to respond promptly to outside variations or requests; and a small set of resources capable of speeding up the convergence time of the procedure. Naturally, these parameters will vary according to the specifics of the application and the deployment features of the hardware.

Formally, the time scale of the neuro-symbolic loop running on the transceivers involved in this context is addressed by the control theory, that relies on the concept of control paths. The performance of the decision-making stage can be grouped into two sets of criteria, namely the economy of knowledge and the immediacy of the choice. The first quantifies the amount of data needed for the spectrum allocation, so that a speedy and straightforward response is possible, while the second relates the time needed to select the modulation and coding scheme to the fast-changing behavior of the wireless environment.

B. Context-Aware Modulation and Coding

Real-time adaptation of modulation and coding schemes to instantaneous channel characteristics offers promising potential for performance improvement. A reinforcement-learning-based architecture has been proposed for SC-FDE using an adaptive constellation-set size and time-varying modulation and coding schemes. The work employs a single-hidden-layer feedforward neural network (SLFN) trained with a hybrid learning algorithm. Multiple evolving selective-survival (M-ESS) improves the adaptation capability—per the experimental results, M-ESS provides significant improvement in the optimal reward in solving complex SC-FDE problems. Despite signalling in content-aware modulation and coding (CA-MAC) being conducted on an application-layer depth, CA-MAC considers the improvement of physical-layer energy efficiency: in-point phrases are of high importance while others can afford lower quality.

Mobile resource management for interlinks in ultra-dense networks can benefit from context information at a local level. Taking into consideration interlink relays, in the literature distance for relay user pairs at the serving cell are optimized under knowledge of user arrangement; unmeasured distance is predicted by a multi-layer perceptron. In neighbouring cells, availability of context-aware information at the mobile operator realises a best-effort decision scheme preventing interlinks during high-load situations. Transmit powers of user terminals decrease with the additional availability of context fostered backhaul signalling. Two traffic distributions, namely, uniformly and abundantly excited under abnormal blocking traffic pattern, have been considered as test cases across a four-square-cell benchmark scenario.

V. AI-DRIVEN TOOLCHAINS AND DEVELOPMENT METHODOLOGIES

The development of AI algorithms for wireless communications is still in its infancy, anticipating the adoption of new design and implementation methodologies. Next-generation AI toolchains must natively exploit the specified semiconductor architecture constraints. Neuro-symbolic learning is a promising route, as it allows semantic communication, guarantees data privacy, and offers model-parsability. Federated learning is another route, as it trains efficient prediction models without sharing training datasets, federating the learning across learning devices while cooperating together to achieve better jointly-optimized models. Indeed, it can be envisioned as a training paradigm between the cloud and the edge nodes.

An adaptive AI-based agent in wireless communications presents specialized capabilities in real-time spectrum access and mobile resource management. Real-time spectrum access, for example, is typically carried out through a multi-data-step process based on the following control-loop structure: an AI agent observes the channel dynamics by collecting information from the wireless environment (channel conditions), and the related embedding is sent to the prediction model for a short-term prediction; the action-space of the agent is constituted of future spectrum availability/channels; and, at each step, the agent selects one of the available channels for data transmission in the immediate next step.

A. Neuro-Symbolic and Federated Learning for Communications

JOURNAL OF INNOVATIVE ACADEMIC RESEARCH
ISSN : 3049-2122
VOLUME: 02 ISSUE: 01
RECEIVED: JANUARY 23 REVISED: FEBRUARY 05 ACCEPTED: FEBRUARY 21 PUBLISHED: MARCH 17

Communications inherently involve a sender, a receiver, and a channel, traditionally separated in the model and, as a result, in the implementations. This separation limits the potential capabilities of the respective systems because the sender does not know the receiver and therefore cannot customize the information sent, and the receiver is not aware of the information sent and therefore needs to decode (detect) it without using its prior knowledge. Neuro-symbolic learning enables the sender and receiver model to be merged and thus to communicate complex patterns in an efficient way. The addition of symbol-driven algorithms in the context of wireless or extremely demanding computer vision AI workloads allows higher performance with lower power. Furthermore, the development of pipelines dedicated and optimized for both sides speeds up the entire training and permits a training federation with data privacy. Neuro-symbolic learning is expected to enable a group of smart devices equipped with cameras to cooperate in the collection and sharing of images.

The recent success of federated learning, regardless of the response time, brings a new possibility to the organization of processing: several users with limited local computational resources can collaborate, keeping the central data private, for their models to improve together. Detecting non-p payload traffic when all the peer nodes are in learning mode is a move from wired to wireless. Cognitive radio networks support dynamic switching of spectrum utilization by considering the spectrum hole as a radio resource. The aim is to enable the decision-maker to generate a stable control policy capable of learning and recognizing the optimum spectrum resources to be used in terms of maximization of the reward function and minimization of the latency.

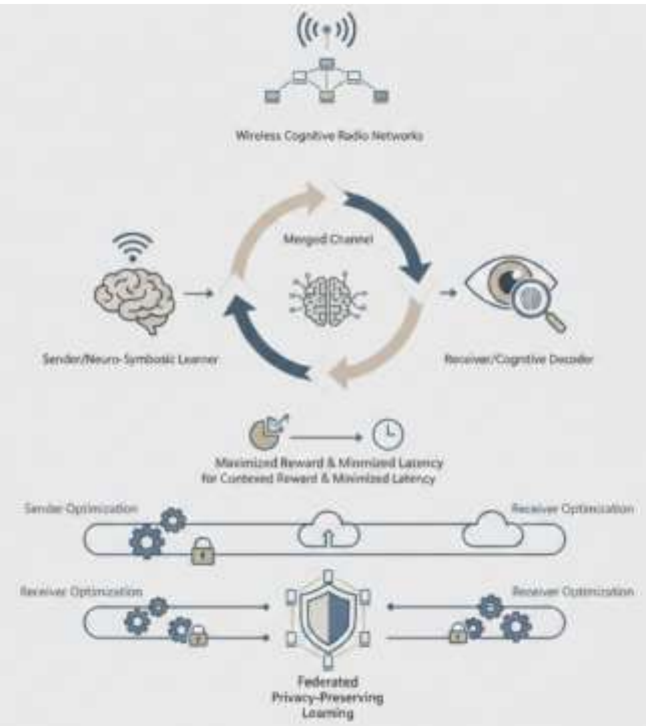


Fig 3: Neuro-Symbolic Federated Learning for Cognitive Radio Networks: Optimizing Spectrum Orchestration and Privacy-Preserving Computer Vision

B. Design Automation, Verification, and Security

As AI-driven tools aim to simplify the development of intelligent networks, it is crucial to devise suitable design automation, verification, and security methodologies. Efforts should markedly reduce development overhead and speed up the introduction

JOURNAL OF INNOVATIVE ACADEMIC RESEARCH
ISSN : 3049-2122
VOLUME: 02 ISSUE: 01
RECEIVED: JANUARY 23 REVISED: FEBRUARY 05 ACCEPTED: FEBRUARY 21 PUBLISHED: MARCH 17

of novel communication technology in silicon, thereby boosting adaptation to the latest trends in utilization. The design automation toolchain consists of three main blocks: a design automation flow that leverages performance influence functions for joint equipment/algorithm development, a verification framework that guarantees systems realization according to the specification, and a threat-guided cybersecurity design tool that strengthens security with minimal impact on the cost/performance of the system.

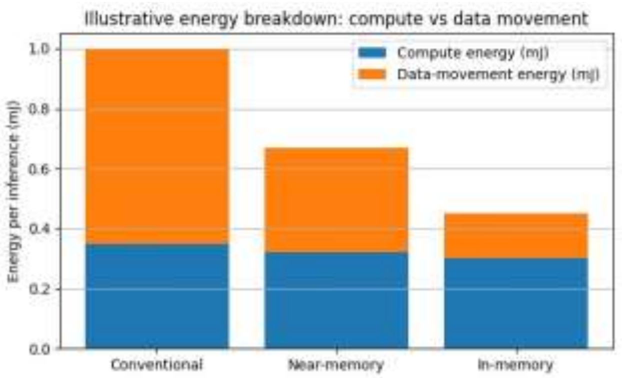
Managing communications systems based on AI excites a broad set of communication architects given the prospect of rapid innovations through simplicity. On the other hand, the AI hardware design space is vast, requiring careful selection of the chosen model, the training data set, and the adoption of appropriate inductive bias. Security also remains a concern because attacks against the design process are emerging that can undermine the reliability of trained models or achieve unauthorized access to the systems that implement them.

VI. EVALUATION FRAMEWORKS AND PERFORMANCE METRICS

Proper evaluation frameworks and performance metrics are crucial to ensuring that novel AI architectures, algorithms, and methodologies perform better than legacy paradigms. Four main performance metrics are considered: energy efficiency, latency, throughput, and silicon efficiency. In addition to traditional metric definitions, a meta-evaluation protocol to assess adaptability, robustness, and security under heterogeneous wireless scenarios is proposed.

The rapid advancement in AI and communication technologies creates considerable pressure on embedded devices, which must guarantee contextualized quality of services while resorting to the lowest amount of resources. This is mainly motivated by the growing volume and level of granularity of data and the presence of heterogeneous environments, where various networks operate with different protocols and parameters. Agents operating in these environments must be able to constantly monitor their real-time conditions, from the changes in the spectrum occupation around them to their relative position to the surrounding users, and adapt their decisions accordingly. This requires small decision latencies to allow for faster adaptability, which is traditionally achieved by reducing the control loops of the system. Effectively evaluating the performance of these support algorithms and their interaction with the rest of the components is an open challenge, especially when the analysis expands to the already complex energy-latency-area-volume-product trade-offs.

An evaluative meta-protocol that covers adaptability, robustness, and security in heterogeneous wireless scenarios must be established. The meta-protocol covers the main building blocks of AI-based communication systems, managing the entire workflow from training to deployment and addressing the teaching (supervised or reinforcement) phases as well as the inferencing stages with the suitable degree of granularity. Adaptivity assesses the ability of a given system to monitor and change its operation mode in response to variations in input/output signals and surrounding conditions, robustness evaluates the resilience of these mechanisms against changes in requirements or the operational environment, and security investigates the exposure to malicious attacks and the capability to mitigate them. The result is a meta-setup suitable for different contexts, generations, and platforms, addressing all of the relevant knobs that determine adaptability and its implementation complexity.



JOURNAL OF INNOVATIVE ACADEMIC RESEARCH
ISSN : 3049-2122
VOLUME: 02 ISSUE: 01
RECEIVED: JANUARY 23 REVISED: FEBRUARY 05 ACCEPTED: FEBRUARY 21 PUBLISHED: MARCH 17

Equation 3) Energy efficiency equations, including data-movement vs compute (in-/near-memory computing)

3.1 Power → Energy

Step 1: energy is integral of power:

$$E = \int_0^T P(t)\, dt$$

If power is approximately constant during operation:

$$E \approx P \cdot T$$

3.2 Energy per bit / per inference

If sending N_b useful bits using energy E:

$$E_b = \frac{E}{N_b}$$

For ML inference, define energy per inference:

$$E_{\text{inf}} = E_{\text{comp}} + E_{\text{move}}$$

3.3 Data movement energy model (why in-memory helps)

Let:

- n_B = number of bytes moved
- e_B = energy per byte moved (architecture-dependent)

Then:

$$E_{\text{move}} \approx n_B \cdot e_B$$

So:

$$E_{\text{inf}} \approx E_{\text{comp}} + n_B e_B$$

A. Energy Efficiency, Latency, and Throughput Metrics

Ever-increasing service demands in terms of data speed and volume tightly couple the resource constraints of wireless networks with the energy sustainability of service providers. Energy efficiency—defined as the ratio between the useful work done and energy consumed—has emerged as an essential metric on top of the traditional performance criteria. Latency and throughput are also critical, depending on the specific application, e.g., ultra-reliable and low-latency systems. Yet, very few scholarly articles and books combine quantitative research on these three metrics. Granting that a modeling environment pertinent to the identified locality can feasibly be created and utilized during the wireless design phase, the real pain points of radio-resource allocation are threefold: Efficiency of the dedicated algorithms; Performance of the selected modulation and

JOURNAL OF INNOVATIVE ACADEMIC RESEARCH
ISSN : 3049-2122
VOLUME: 02 ISSUE: 01
RECEIVED: JANUARY 23 REVISED: FEBRUARY 05 ACCEPTED: FEBRUARY 21 PUBLISHED: MARCH 17

coding schemes; and Robustness against model uncertainties. The first aspect is generally known as speed of execution. Model-predictive control, for instance, aims at fast-adaptive operation by exploiting the drop of latency that is possible at long time scales. In fact, standard wireless automation is already multi-scale.

At a higher level, budget-aware dynamic modulations and coding require separation in time, frequency or spatial dimensions of the communication or propagation links or paths that the users share and allow adaptive coding and modulation according to local conditions. At the highest level, proper model specification considerably speeds up evaluation and enter layer-or layer-2-and-3-level control loops that allow extremely low latency for control over synthesis services like driven assembly and additive manufacturing. However, the modeling and associated control objectives remain complex. Fine-tuning radio-resource allocation for minimum latency has emerged as a topic of research on networking.

Concept	Equation (standard form)	Units
Energy efficiency	EE = \frac{\text{useful work}}{\text{energy}}	bit/J or ops/J
Throughput	T = \frac{\text{delivered bits}}{\Delta t}	bit/s
Shannon capacity	C = B\log_2(1+\mathrm{SNR})	bit/s
SNR	\mathrm{SNR}=\frac{P_s}{P_n}=\frac{P_s}{N_0B}	unitless

B. Adaptability, Robustness, and Security Considerations

Realizing wirelessly intelligent environments demands addressing a range of adaptability, robustness, and security challenges. AI-enabled signal processing and resource management offer temporal and spatial adaptability, yet decision latencies may jeopardize real-time performance. Furthermore, marker data availability dictates loop control frequency, an aspect readily monitored at network level for efficiency optimization. Security holes left open in legitimate communication may be exploited by adversaries able to tamper only selected inputs/outputs. As an additional asset, signals emanating from a legitimate source and exploiting time to travel a circular route in the environment possess intrinsic redundancy that can be exploited to enhance resistance to eavesdropping. Data travelling in the other direction should naturally be different to further hinder compromise.

Symbols sensitive to the environment may enable identification of relevant factors, with the learning cycle closely monitored to ensure stability and efficiency. The advance of AI tools supporting the design automation flow also gives rise to a potentially vast surface area for counter measures against Design Security Attacks. Message Injection Attacks targeting discrete-event Machine Learning can be approached by applying a suitable tuning in the Generative Network. A federated learning implementation encompassing several devices progressively strives to mimic aspects of the channel conditions for other devices, thus enhancing the soft-information exchange process while preserving the localization of the channel marker data.

VII. CONCLUSIONS

AI-Driven Agentic Semiconductor Architectures hold great promise in transforming Next Generation Wireless Networks into truly intelligent entities capable of responding in real-time for optimal performance. A number of ingredient technologies and mechanisms are currently being developed and will together enable the design and deployment of efficient and intelligent wireless networks that leverage AI in their operation at various timescales. It is, however, also crucial to define the tools and methodologies for enabling these networks: how will the adaptive HW/SW be built and how will the decision-making for the adaptation be taken? Questions related to these aspects often remain, which is somewhat surprising given the pivotal role they will play in bringing the Agentic Paradigm to fruition.

An initial exploration of AI-driven Methodologies and Tool Chains for Next Generation Wireless Networks is proposed here, providing insights into addressing some of these important areas of research. A Neuro-Symbolic approach is elaborated for the Communications and Physical layers together with a Federated Learning approach that captures the Collaboration aspect

JOURNAL OF INNOVATIVE ACADEMIC RESEARCH
ISSN : 3049-2122
VOLUME: 02 ISSUE: 01
RECEIVED: JANUARY 23 REVISED: FEBRUARY 05 ACCEPTED: FEBRUARY 21 PUBLISHED: MARCH 17

across the higher layers while addressing the Data Privacy overhead issues. Finally, acceleration of the Development of Agentic systems is addressed with emphasis on Security and Trust in the complete development cycle from Design to Deployment.

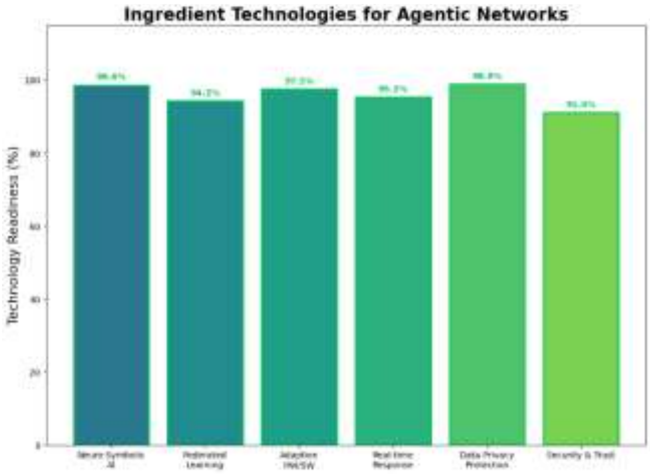


Fig 4: Ingredient Technologies for Agentic Networks

A. Key Takeaways and Future Directions

Next-generation intelligent wireless systems require agentic AI-driven adaptive semiconductor architectures. Exploring the resource management capabilities from real-time spectrum management to context-aware adaptive coding and modulation sheds light on the power of AI within intelligent wireless networks. Likewise, identifying AI-driven semiconductor architectural trends—FlPs/VL-shaped/dyn-re fab/analogue comput-ing units/spiking/neuromorphic processors/in- and near-memory data locality—and associated development toolchains highlights synergies for reducing energy per inference while maximising performance. Furthermore, the advent of AI toolchains enables Security-by-Design within wireless networks.

Energy statistics, latency performance, and throughput indicators should complement silicon performance metrics. Therefore, research analysis must encompass adaptability, robustness, and security holes under a heterogeneous wireless scenario. In addition, performance measurements should cover Real-time Spectrum Management, context-aware Modulation and Channel Coding Adaptation, AI-Driven Architecture Switching for Data Generation and Processing, and Real-time Communication Monitoring and Control. Future research development should provide a scalable AI-Driven communication platform under various wireless Saturation scenarios, Such adaptive systems will enhance resource utilisation, lower energy consumption per device, accelerate decision-making processes, and enable context-aware AI-based modula-tion and coding strategies in concordance with changing wireless channel conditions.

References

1. Ren, P., Xiao, Y., Chang, X., Huang, P.-Y., Li, Z., Chen, X., & Wang, X. (2021). A comprehensive survey of neural architecture search: Challenges and solutions. ACM Computing Surveys, 54(4), Article 76.
2. Wu, Y., Liu, A., Huang, Z., Zhang, S., & Van Gool, L. (2021). Neural architecture search as sparse supernet. Proceedings of the AAAI Conference on Artificial Intelligence, 35(12), 10379–10387.

JOURNAL OF INNOVATIVE ACADEMIC RESEARCH
ISSN : 3049-2122
VOLUME: 02 ISSUE: 01
RECEIVED: JANUARY 23 REVISED: FEBRUARY 05 ACCEPTED: FEBRUARY 21 PUBLISHED: MARCH 17

3. Davuluri, P. S. L. (2023). AI-Augmented Sanctions Screening: Enhancing Accuracy and Latency in Real Time Compliance Systems. AI-Augmented Sanctions Screening: Enhancing Accuracy and Latency in Real Time Compliance Systems (December 15, 2023).
4. Mellor, J., Turner, J., Storkey, A., & Crowley, E. J. (2021). Neural architecture search without training. Proceedings of the 38th International Conference on Machine Learning, 139, 7588–7598.
5. Zhang, X., Hou, P., Zhang, X., & Sun, J. (2021). Neural architecture search with random labels. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 10907–10916.
6. Kolla, T. (2024). Graph Neural Networks for HCC Risk Adjustment and Interoperability. International Journal of Science, Research and Technology, 7(6), 13244-13255.
7. Chen, B., Li, P., Li, B., Lin, C., Li, C., Sun, M., Yan, J., & Ouyang, W. (2021). BN-NAS: Neural architecture search with batch normalization. Proceedings of the IEEE/CVF International Conference on Computer Vision, 307–316.
8. Muravev, A., Raitoharju, J., & Gabbouj, M. (2021). Neural architecture search by estimation of network structure distributions. IEEE Access, 9, 127728–127741.
9. Benmeziane, H., El Maghraoui, K., Ouarnoughi, H., Niar, S., Wistuba, M., & Wang, N. (2021). Hardware-aware neural architecture search: Survey and taxonomy. Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, 4322–4329.
10. Kolla, S. K., & Reddy, V. A. R. (2024). Evaluating Cloud-Native vs. Hybrid Architectures for Health Benefit Administration Systems. International Journal of Medical Toxicology and Legal Medicine, 27(5), 1042-1053.
11. Qin, Z., Li, G., & Ye, H. (2021). Federated learning and wireless communications. IEEE Wireless Communications, 28(5), 92–98.
12. Zhang, C., Patras, P., & Haddadi, H. (2021). Deep learning in mobile and wireless networking: A survey. IEEE Communications Surveys & Tutorials, 23(2), 1153–1198.
13. Orhan, O., Swamy, V. N., Tetzlaff, T., Nassar, M., Nikopour, H., & Talwar, S. (2021). Connection management xAPP for O-RAN RIC: A graph neural network and reinforcement learning approach. 2021 20th IEEE International Conference on Machine Learning and Applications, 936–941.
14. Inala, R. (2023). AI-powered investment decision support systems: Building smart data products with embedded governance controls. Journal for ReAttach Therapy and Developmental Diversities, 6(10), 2251-2266.
15. Shin, M., Mughal, D. M., Park, S., Kim, S.-H., & Chung, M. Y. (2021). Cellular licensed band sharing technology among mobile operators: A reinforcement learning perspective. Wireless Personal Communications, 120, 27–47.

JOURNAL OF INNOVATIVE ACADEMIC RESEARCH
ISSN : 3049-2122
VOLUME: 02 ISSUE: 01
RECEIVED: JANUARY 23 REVISED: FEBRUARY 05 ACCEPTED: FEBRUARY 21 PUBLISHED: MARCH 17

16. Huang, S., Wang, S., Wang, R., Wen, M., & Huang, K. (2021). Reconfigurable intelligent surface assisted edge machine learning. IEEE Journal on Selected Areas in Communications, 39(10), 3056–3070.
17. Elhoushy, S., Ibrahim, M., & Hamouda, W. (2022). Cell-free massive MIMO: A survey. IEEE Communications Surveys & Tutorials, 24(1), 492–523.
18. Liu, F., Cui, Y., Masouros, C., Xu, J., Han, T. X., Eldar, Y. C., Buzzi, S., & Zhou, S. (2022). Integrated sensing and communications: Toward dual-functional wireless networks for 6G and beyond. IEEE Journal on Selected Areas in Communications, 40(6), 1728–1767.
19. Inala, R., & Somu, B. (2024). Agentic ai in retail banking: Redefining customer service and financial decision-making. Journal of Artificial Intelligence and Big Data Disciplines, 1(1), 1-19.
20. Zhou, W., Zhang, R., Chen, G., & Wu, W. (2022). Integrated sensing and communication waveform design: A survey. IEEE Open Journal of the Communications Society, 3, 1930–1949.
21. Masaracchia, A., Sharma, V., Canberk, B., Dobre, O. A., & Duong, T. Q. (2022). Digital twin for 6G: Taxonomy, research challenges, and the road ahead. IEEE Open Journal of the Communications Society, 3, 2137–2150.
22. Tang, F., Chen, X., Rodrigues, T. K., Zhao, M., & Kato, N. (2022). Survey on digital twin edge networks (DITEN) toward 6G. IEEE Open Journal of the Communications Society, 3, 2560–2581.
23. Kolla, S. K., & Mangalampalli, B. M. (2024). Edge-Based Deep Learning Systems for Point-of-Care Diagnostic Intelligence. Journal of Neonatal Surgery, 13(1), 2387-2399.
24. Faisal, K. M., & Choi, W. (2022). Machine learning approaches for reconfigurable intelligent surfaces: A survey. IEEE Access, 10, 27343–27367.
25. Luo, X., Chen, H. H., & Guo, Q. (2022). Semantic communications: Overview, open issues, and future research directions. IEEE Wireless Communications, 29(1), 210–219.
26. Xie, H., Qin, Z., Li, G. Y., & Juang, B.-H. F. (2022). Deep learning enabled semantic communication systems. IEEE Transactions on Signal Processing, 69, 2663–2675.
27. Gottimukkala, V. R. R. (2024). Federated Learning Approaches for Fraud Detection in International Payment Systems. https://www. jisem-journal. com/download/118_JISEM. pdf.
28. Ramezanpour, K., & Jagannath, J. (2022). Intelligent zero trust architecture for 5G/6G networks: Principles, challenges, and the role of machine learning in the context of O-RAN. Computer Networks, 217, Article 109358.
29. Hojatian, H., Nadal, J., Frigon, J.-F., & Leduc-Primeau, F. (2022). Decentralized beamforming for cell-free massive MIMO with unsupervised learning. IEEE Communications Letters, 26(5), 1042–1046.
30. Zhang, R., Zhang, Q., & Zhu, H. (2022). RIS-assisted cell-free massive MIMO systems with reflection area: AP number reduction. Physical Communication, 55, Article 101857.

JOURNAL OF INNOVATIVE ACADEMIC RESEARCH
ISSN : 3049-2122
VOLUME: 02 ISSUE: 01
RECEIVED: JANUARY 23 REVISED: FEBRUARY 05 ACCEPTED: FEBRUARY 21 PUBLISHED: MARCH 17

31. Bonati, L., Polese, M., D'Oro, S., Basagni, S., & Melodia, T. (2023). OpenRAN Gym: AI/ML development, data collection, and testing for O-RAN on PAWR platforms. Computer Networks, 220, Article 109502.
32. Inala, R. Designing Scalable Technology Architectures for Customer Data in Group Insurance and Investment Platforms.
33. Zhu, G., Lyu, Z., Jiao, X., Liu, P., Chen, M., Xu, J., Cui, S., & Zhang, P. (2023). Pushing AI to wireless network edge: An overview on integrated sensing, communication, and computation towards 6G. Science China Information Sciences, 66, Article 130301.
34. Hossain, A. R., Liu, W., Ansari, N., Kiani, A., & Saboorian, T. (2023). AI-native for 6G core network configuration. IEEE Networking Letters, 5(4), 255–259.
35. Zhang, J., Lu, W., Xing, C., Zhao, N., Al-Dhahir, N., Karagiannidis, G. K., & Yang, X. (2024). Intelligent integrated sensing and communication: A survey. Science China Information Sciences, 68, Article 131301.
36. Wang, H., Zhang, J., Nie, G., Yu, L., Yuan, Z., Li, T., Wang, J., & Liu, G. (2024). Digital twin channel for 6G: Concepts, architectures and potential applications. IEEE Communications Magazine, 62(10), 74–80.
37. Thomas, C. K., Chaccour, C., Saad, W., & Kurisummoottil Thomas, C. (2024). Causal reasoning: Charting a revolutionary course for next-generation AI-native wireless networks. IEEE Vehicular Technology Magazine, 19(2), 52–61.
38. Saad, W., Hashash, O., Thomas, C. K., Chaccour, C., Debbah, M., Mandayam, N., & Han, Z. (2024). Artificial general intelligence (AGI)-native wireless systems: A journey beyond 6G. IEEE Communications Magazine, 62(10), 44–50