Governing trustworthy AI in critical communications infrastructure: an ethical and regulatory framework for 5G/6G Open RAN and network automation

Sushanta Paul

Abstract

Purpose – *This paper aims to examine how trustworthy artificial intelligence (AI) principles can be operationalised where AI is becoming an operating control layer in fifth-generation (5G) and emerging sixth-generation (6G) communications infrastructure, and to propose a governance framework bridging horizontal AI rules and telecom-specific assurance.*

Design/methodology/approach – *It uses a structured documentary review and cross-jurisdictional gap analysis of AI governance instruments, Open Radio Access Network (Open RAN) assurance literature, telecom-security standards, and policy developments in the United States, the European Union, the United Kingdom, Japan and Singapore, without primary empirical data.*

Findings – *The telecom standards stack is mature on architecture and baseline security but weaker on explainability, post-deployment AI assurance, model provenance and human-override regimes, while horizontal AI laws lack telecom-specific operational depth. The paper develops a risk taxonomy, a cross-jurisdictional gap map and TRACE-5G/6G (Tier, Register, Assure, Control, Evaluate).*

Research limitations/implications – *The framework is analytical rather than empirically validated because the study draws on documentary sources rather than primary field data or a formal systematic-review protocol; future work should test it against operational deployments.*

Practical implications – *It offers regulators, operators, integrators and standards bodies an analytical synthesis, a gap map, a mapping table and directions for future testing.*

Originality/value – *The paper organises dispersed trustworthy-AI obligations into a telecom-tailored lifecycle checklist and provides a transparent cross-jurisdictional gap map as an analytical tool for future profiling exercises.*

Keywords *Trustworthy AI, AI governance, Algorithmic accountability, Human oversight, Open RAN, Telecommunications regulation, 5G/6G, Critical infrastructure*

Paper type *Conceptual paper*

Sushanta Paul is based at National Board of Revenue (NBR), Government of Bangladesh, Dhaka, Bangladesh.

Received 25 May 2026
Revised 19 June 2026
Accepted 6 July 2026

Competing interests: The author has no competing interests to declare that are relevant to the content of this article.

Introduction

Telecommunications infrastructure is undergoing a transformation that is ethical as much as technical. In fifth-generation (5G) and emerging sixth-generation (6G) environments, network functions are increasingly virtualised, software-driven and orchestrated through artificial intelligence (AI)-enabled control loops that allocate spectrum, prioritise traffic, route around disturbance and detect anomalies, increasingly without contemporaneous human approval. The International Telecommunication Union Radiocommunication Sector (ITU-R) framework for IMT-2030 anticipates AI/machine learning (ML) as integral to the IMT-2030 architecture, including an integrated AI-and-communication usage scenario (ITU-R, 2023), and the Global Coalition on Telecommunications (GCOT) 6G principles call for resilience-by-design and

safe failover (GCOT, 2026). Policy institutions no longer treat this as a purely technical matter: the US National Telecommunications and Information Administration (NTIA) calls secure telecommunications vital to competitiveness and national security and has opened Executive Branch policy for 6G to public input (NTIA, 2024).

The problem is not absent standards but dispersed ones, spread across overlapping layers. The NIST AI Risk Management Framework (AI RMF 1.0) is built around the GOVERN, MAP, MEASURE and MANAGE functions and trustworthiness characteristics such as validity, security, transparency, explainability, privacy and fairness (NIST, 2023); NIST CSF 2.0 adds governance and supply-chain risk management (NIST, 2024a); 3GPP TS 33.501 defines the 5G security architecture (3GPP, 2025) and TR 33.877 studies AI/ML security in NG-RAN (3GPP, 2023); the ENISA 5G controls matrix consolidates 144 high-level and 399 detailed controls (ENISA, 2023); and the EU AI Act adds horizontal duties for high-risk systems (European Union, 2024). No single instrument consolidates these into a telecom-specific, lifecycle-organised form for AI-enabled network automation and Open Radio Access Network (Open RAN).

This dispersion directly affects network users. The Body of European Regulators for Electronic Communications (BEREC) has documented AI use across network operations and warned of explainability, liability, privacy and cybersecurity implications (BEREC, 2023). In high-consequence settings – emergency communications, public safety, national-scale carrier networks – an opaque or poorly governed model produces harms that are not merely technical: degraded service, disputes no human can adjudicate and outcomes no party owns.

The US national-interest implications are direct. NTIA's 6G and Open RAN work emphasises trusted suppliers, interoperability and specialised use cases including defence scenarios (NTIA, 2024; NTIA, 2025; NTIA and ITS, 2025), while CISA's Joint Cyber Defense Collaborative AI playbook shows US institutions increasingly expect AI vulnerabilities to be shared in collective defence (CISA, 2025).

This study therefore asks: what risk-based framework can align trustworthy AI principles, cybersecurity requirements and telecom-specific operational realities for Open RAN and AI-enabled 5G/6G automation in a way that is auditable, accountable and bounded by human authority? It answers through a structured documentary synthesis and proposes a governance model for regulators, operators, integrators and standards communities.

Literature review

Trustworthy AI: ethical foundations and governance frameworks

The trustworthy AI literature has converged on a common normative core: AI should be valid and reliable, safe, secure and resilient, accountable and transparent, explainable, privacy-enhanced and fair with harmful bias managed (NIST, 2023). The NIST AI RMF organises practice around GOVERN, MAP, MEASURE and MANAGE (NIST, 2023). Its 2026 concept note for a critical-infrastructure profile confirms that high-stakes environments require lifecycle trustworthiness across information technology, operational technology and industrial systems (NIST, 2026) – logic that transfers directly to carrier networks.

The horizontal regulatory layer is similarly maturing. The EU AI Act establishes risk-based duties for high-risk AI – lifecycle risk management, human oversight, robustness, cybersecurity, and transparency – and classifies AI used as a safety component in the management and operation of critical digital infrastructure as high-risk (European Union, 2024). ISO/IEC 23894 and 42001 add international risk-management and management-system layers that turn voluntary principles into auditable processes (ISO/IEC, 2023a; ISO/IEC, 2023b), while NIST CSF 2.0 emphasises governance and supply-chain risk management and notes that cybersecurity and privacy considerations apply across the AI

lifecycle (NIST, 2024a). A telecom-relevant baseline is emerging: ETSI's baseline cybersecurity requirements for AI models and systems define lifecycle security principles for developers, integrators and operators across the AI supply chain (ETSI, 2025).

General AI ethics scholarship adds a relevant caution. The OECD Recommendation and the EU High-Level Expert Group locate trustworthy AI in human-centred values, transparency, robustness, fairness and accountability (OECD, 2024; HLEG, 2019). Floridi *et al.* (2018) framed it around beneficence, non-maleficence, autonomy, justice and explicability, and Jobin *et al.* (2019) show global guidelines converge on core principles but diverge on implementation. Mittelstadt's (2019) critique was central – principles cannot guarantee ethical AI unless translated into institutions and enforceable practices – and Selbst *et al.* (2019) warned that abstraction can hide the social conditions that produce harm. ML assurance, likewise, must be generated as evidence across the model lifecycle – data management, verification, deployment and monitoring – not asserted at a single checkpoint (Ashmore *et al.*, 2021).

This study is situated within three regulatory-governance traditions. Firstly, algorithmic regulation describes systems that govern through continuous data-driven monitoring and automated or recommended intervention (Yeung, 2018); AI-enabled network automation is such a system, making its governance a regulatory and not only a security question. Secondly, responsible-innovation scholarship argues that emerging technologies should be governed through anticipation, reflexivity, inclusion and responsiveness (Stilgoe *et al.*, 2013), embedding public values into technical design rather than appending them afterwards. Thirdly, telecommunication is infrastructure for other infrastructures, so its interdependencies propagate failures across sectors (Rinaldi *et al.*, 2001); this raises the public-interest stakes above those of a single sector.

AI-security ethics also matters directly. NIST's adversarial ML taxonomy shows that attacks span lifecycle stages and include evasion, poisoning, privacy compromise and model manipulation (NIST, 2025). Each category has an ethical valence: poisoning and evasion degrade fairness and safety; privacy compromise violates data-subject rights; model manipulation undermines the conditions of meaningful explanation. For telecom specifically, supervising AI-enabled control loops requires more than generic cybersecurity controls; it requires defences against data poisoning, model tampering, and environment-induced drift, visible to the institutions accountable to users. Recent experimental work on the RAN Intelligent Controller makes these threats concrete: Sapavath *et al.* (2023) show that manipulating data in a shared database RAN Intelligent Controller RIC inside the near-real-time RAN Intelligent Controller (RIC) can materially degrade the accuracy of an ML-based near-real-time RAN application xApp and, through it, system performance, while Lacava *et al.* (2025) demonstrate backdoor data-poisoning of deep-reinforcement-learning xApps that, once triggered on a live O-RAN deployment, produced average network-performance degradation of roughly 87%. Such RIC-specific evidence is precisely what a telecom assurance regime must anticipate.

Open RAN, 5G/6G automation and assurance

Open RAN scholarship presents a double-edged architecture. Liyanage *et al.* (2023) argue that disaggregation, softwarisation, AI and virtualisation create flexibility and innovation but expand the attack surface and complicate security and privacy, while Yungaicela-Naula *et al.* (2024) show that AI/ML in RIC environments can both improve operations and intensify misconfiguration risk, especially where near-real-time RAN applications and non-real-time RAN applications (rApps) interact or policies conflict.

Official documents reinforce this conclusion. The NTIA Open RAN Security Report observes that Open RAN introduces additional security dynamics operators must manage, and that, lacking a standard way of sourcing and deploying Open RAN, mobile network operators

(MNOs) must define and enforce security roles throughout the lifecycle (NTIA, 2023). Japan's MIC reached a parallel conclusion in the Quad Open RAN Security Report (MIC, 2023).

Industry standards bodies are moving in the same direction. The O-RAN Alliance Security Work Group (WG11) 2025 update reports published security requirements, tests, threat modelling, progress towards zero-trust maturity, and a security-assurance programme based on O-RAN Security Assurance Specifications (SCAS) aligned with the GSMA NESAS framework, and identifies 6G priorities including AI security, continuous monitoring, and post-quantum cryptography (O-RAN Alliance, 2025). The 2024 zero-trust white paper aligns with NIST SP 800-207, the CISA Zero Trust Maturity Model, and US mobile-industry guidance (O-RAN Alliance, 2024). This zero-trust maturation between 2024 and 2025 bears directly on two TRACE functions: WG11's security-assurance and conformance artefacts inform the pre-deployment evidence required at Assure, while its runtime zero-trust constraints and continuous-monitoring direction inform the enforcement and override regime at Control. These strengthen security and interface controls without resolving the ethical questions of explainability, retraining and model provenance in automated operations.

Standards and policy landscape

At the global standards level, ITU Telecommunication Standardization Sector (ITU-T) Recommendation Y.3172 provides one of the earliest formal architectures for ML in future networks, including an ML pipeline and ML management and orchestration functions (ITU-T, 2019). This early architecture remains a reference point for AI-driven network management, and is complemented by ETSI's zero-touch network and service management (ZSM) work, whose reference architecture targets closed-loop, largely autonomous management and orchestration across multi-vendor domains (ETSI GS ZSM 002, 2019); together they show that closed-loop automation has had standardised architectural form for several years, sharpening rather than softening the governance question. ITU-R Recommendation M.2160 projects this logic into IMT-2030 by envisioning an AI-native air interface with sensing integration, distributed ML and AI-compute orchestration (ITU-R, 2023). This confirms that AI is not a peripheral enhancement to future networks but part of the base architecture, making the governance question more urgent.

The policy layer is also thickening. GCOT's 2025 principles on AI in telecommunications call for transparency, explainability, human oversight, redress, privacy, fairness, and security-by-design (GCOT, 2025), and its 2026 6G principles treat 6G as critical national infrastructure, demanding resilience-by-design, safe failover, resilient AI, telecom-AI incident sharing and multi-vendor Open RAN support (GCOT, 2026). Japan's 2026 AI Guidelines for Business retain a soft-law, risk-based lifecycle approach (METI, 2026), and Singapore's IMDA frameworks emphasise accountability and human responsibility (IMDA, 2026).

Convergence does not equal closure. BEREC has warned that telecom regulators may need roles in AI Act implementation, competition oversight, explainability and user rights as AI becomes embedded in electronic communications networks and services (ECN/S) (BEREC, 2023). Industry white papers from major vendors describe AI-native, intent-based or zero-touch network futures (Ericsson, 2025; Roeland *et al.*, 2026), but those documents are strongest on capability architecture and weakest on public-interest guardrails. The standards stack is advancing towards autonomous networks faster than the governance stack is advancing towards auditable, telecom-specific assurance. The ethical risk is that capability outruns accountability.

Methods

Design

This study used a structured documentary review combined with cross-jurisdictional ethical and regulatory gap-mapping. Document analysis is an established qualitative method for

systematically examining institutional, regulatory and technical texts to derive themes and evidence (Bowen, 2009), and suits a field governed primarily by standards, legal instruments and official guidance. The design adopts transparent reporting principles – explicit search logic, stated inclusion and exclusion criteria, source categorisation, and reproducible synthesis – but is a structured documentary review rather than a fully audited database systematic review, because the corpus mixes legal instruments, official standards, white papers, government advisories and academic literature.

Search strategy and data sources

The source base combined four categories. The first was official standards and government or regulatory documents: the NIST AI RMF, CSF and adversarial-ML guidance, NTIA 6G and Open RAN reports, the CISA JCDC AI playbook, ITU-R M.2160, ITU-T Y.3172, 3GPP TS 33.501 and TR 33.877, the ENISA 5G Security Controls Matrix, the EU AI Act and GCOT documents. The second was peer-reviewed literature, especially studies of Open RAN security, O-RAN misconfiguration, and machine-learning assurance. The third was industry and standards-adjacent white papers from the O-RAN Alliance and Ericsson, used cautiously as supplementary. The fourth was regional-regulator material, including BEREC, METI, IMDA and MIC. All cited sources appear in the reference list; documentary sources were reviewed up to 10 May 2026. The five jurisdictions were selected to span contrasting governance models rather than for exhaustiveness: the US represents a risk-based, agency-led approach; the European Union a binding, rights- and risk-based statutory regime; the UK, through the GCOT, a pro-innovation, sectoral approach; and Japan and Singapore soft-law, lifecycle-based models. A trans-jurisdictional standards-and-industry column captures the SDO layer that operates across all blocs.

Search logic combined the following concept clusters: trustworthy AI; AI governance and risk management; Open RAN security; O-RAN misconfiguration; 5G automation; 6G AI-native; RAN Intelligent Controller; xApps and rApps; zero trust; critical infrastructure; explainability; human oversight; algorithmic accountability.

Inclusion and exclusion criteria

Sources were included if they were official standards, regulations or government documents relevant to AI-enabled telecom infrastructure; peer-reviewed studies materially addressing Open RAN security, AI/ML assurance or autonomous network management; or industry documents adding architecture or assurance context unavailable elsewhere. English-language sources were prioritised. Sources were excluded if duplicative, purely promotional, unrelated to telecom infrastructure or too remote from the focus on ethical governance and cybersecurity. Particular weight was given to documents addressing at least one of five dimensions: lifecycle AI governance, telecom cybersecurity, Open RAN assurance, operational automation and autonomy and critical-infrastructure or public-interest implications.

Analytical approach

The analysis proceeded in three steps: sources were coded into thematic clusters (trustworthy-AI principles, cybersecurity controls, Open RAN architecture and assurance, automation and autonomy, and regulatory obligations); risks were mapped into a taxonomy spanning technical, organisational, legal and systemic dimensions, attentive to ethical implications such as opacity, bias and accountability gaps; and a comparative gap map was constructed across jurisdictions and standards bodies to locate where governance coverage was comprehensive, partial or absent. Operational definitions of the four maturity levels, the cell-assignment decision rule and a per-cell justification appear in the Supplementary material A. The reviewed corpus comprised the 46 sources listed in the

references – 25 official standards and government or regulatory documents, 13 peer-reviewed studies, 4 industry and standards-adjacent white papers, and 4 regional-regulator texts – retained after applying the inclusion and exclusion criteria stated above to the candidate sources identified through the issuing bodies' repositories and the Scopus, IEEE Xplore and Google Scholar searches. Finally, the gap map is an analytical instrument, not formal legal advice; it characterises the relative maturity of governance coverage rather than assessing enforceability jurisdiction by jurisdiction.

Results

Thematic synthesis

The documentary record indicates a transitional governance phase. In governance terms, however, the sector has not produced a unified framework equivalent to the NIST AI RMF for general AI or ENISA for broad 5G controls. The ethical implication is that AI capability is being deployed into an accountability architecture still under construction.

Official actors already recognise AI-enabled telecom as a strategic governance domain whose ethical stakes – public safety, fair access, redress and democratic accountability – are commensurate with its technical complexity.

The literature also converges on a core institutional finding: Open RAN redistributes accountability. The NTIA Open RAN Security Report stresses that there is no standardised sourcing or integration model and that operators must define and enforce roles across the lifecycle, concluding that Open RAN can be as secure as traditional RAN if stakeholders adopt sound practices, while warning that AI/ML poisoning remains unresolved (NTIA, 2023).

Telecom AI governance is not only about preventing control-plane failure but about preserving user rights, accountability, dispute resolution and public trust – the conditions under which AI-mediated communications are ethically legitimate.

Risk taxonomy

The taxonomy in Table 1 synthesises NIST AI RMF trustworthiness characteristics, the NIST adversarial ML taxonomy, the EU AI Act and peer-reviewed Open RAN literature.

Cross-jurisdictional and standards gap map

The gap map in Figure 1 compares the governance landscape across jurisdictions and standards ecosystems. "Strong" denotes explicit, materially relevant coverage; "Partial" adjacent or general coverage; "Emerging" active but soft-law or developmental coverage; and "Gap" materially underdeveloped coverage for telecom AI operations.

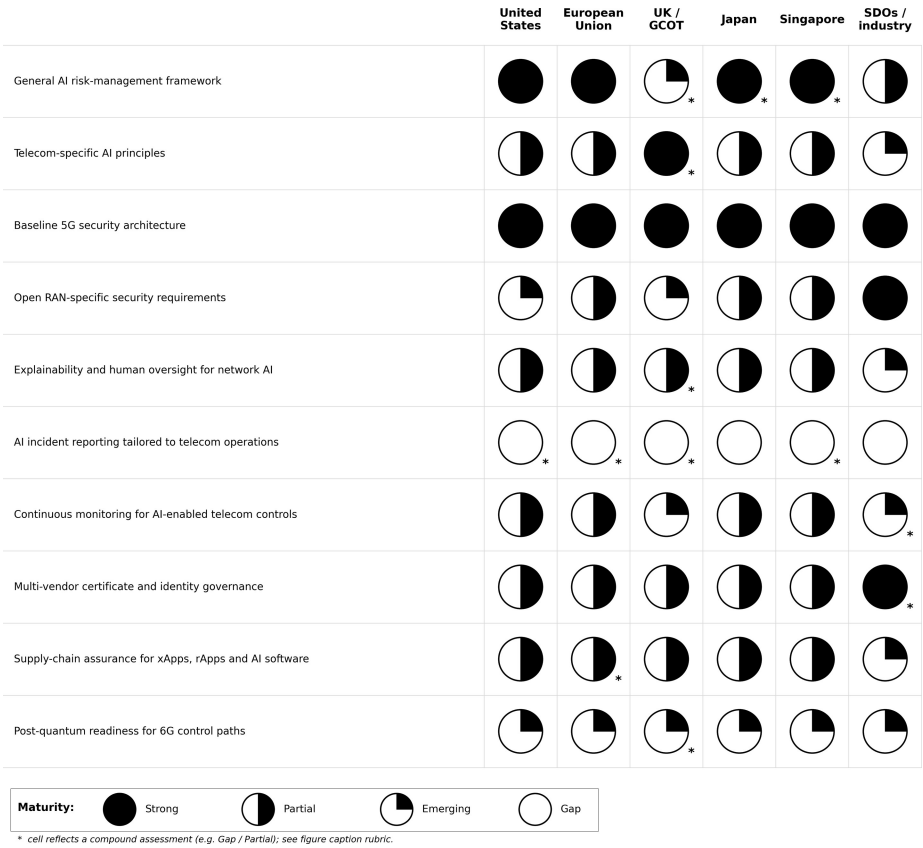
The least mature domains across all blocs are post-deployment AI assurance, model provenance, explainability for operational decision loops and telecom-specific incident reporting. The standards ecosystem is ahead of regulation on runtime technical controls, while regulation is ahead on formal accountability and legal consequence; neither yet delivers a complete assurance chain for AI-native telecom infrastructure. That gap is precisely where ethical harms accumulate – unaccounted for, unaudited, and unredressed. It also bears on how the principles are weighted: in critical-infrastructure settings, safety, security and resilience typically carry higher regulatory priority than fairness, because failures in the control plane translate directly into service-continuity and public-safety consequences. The gap map and taxonomy therefore treat the trustworthy-AI principles as differentially weighted in this sector rather than of equal force, while still requiring that fairness and other principles be addressed.

Table 1 Risk taxonomy for trustworthy AI in Open RAN and 5G/6G automation

Risk domain	*Telecom manifestation*	*Governance failure mode*	*Illustrative controls*	*Evidence basis*
Model integrity and adversarial ML	Poisoned training data, model tampering, evasion against anomaly or traffic models, unsafe retraining	No model provenance, weak testing, evaluation, verification and validation (TEVV); no drift thresholds; no signed model release process	Model registry, signed models, training-data lineage, red-teaming, rollback authority	Official documents, peer-reviewed literature
Policy conflict and misconfiguration	Conflicting xApp or rApp actions, unsafe optimisation, oscillation across RIC loops	No policy hierarchy, unclear change authority, no conflict-resolution workflow	Policy arbitration, staged deployment, simulation or digital-twin validation, human approval gates	Official documents, peer-reviewed literature
Interface and protocol exposure	A1, E2, O1, O2 and Open Fronthaul interface exposure; certificate sprawl in multi-vendor systems	Inconsistent Transport Layer Security (TLS) and mutual TLS (mTLS), poor certificate lifecycle governance, partial interface hardening	Mandatory Transport Layer Security (TLS) and mTLS 1.3, certificate management, interface conformance tests, zero-trust architecture	Official documents
Supply-chain and provenance risk	Third-party xApps and rApps, cloud dependencies, opaque software stacks, concentration risk	Incomplete software bill of materials (SBOM) or cryptographic bill of materials (CBOM); absent model cards, datasheets or model-registry records; no supplier assurance; weak contract controls	SBOM, CBOM, supplier security clauses, Open Test and Integration Centre (OTIC), O-RAN Security Assurance Specifications (SCAS) evidence aligned with GSMA NESAS, procurement assurance	Official documents, industry
Explainability and accountability risk	Opaque routing, prioritisation, self-healing, customer-facing AI, automated incident response	No audit trail, no explanation layer, unclear redress path	Logging, explanation records, operator dashboards, challenge-and-override procedures	Official documents, regulation
Privacy and data-governance risk	Telemetry, location, subscriber behaviour, edge inferencing, shared data for training	Excessive data collection, unclear secondary use, weak access control	Data minimisation, role-based access, privacy-preserving learning, retention rules	Official documents, regulation
Resilience and fail-safe risk	Cascading errors, update failures, weakness in Global Navigation Satellite System and Position-Navigation-Timing (GNSS/PNT) signals, disaster recovery failure, false base-station exposure	No safe-failover design, no fallback mode, untested restoration logic	Kill-switches, alternative-access failover, traffic criticality tiers, recovery metrics	Official documents
Organisational and legal accountability	Multi-vendor blame shifting, unclear operator and integrator roles, inadequate regulator visibility	Diffuse responsibility, contract ambiguity, missing incident obligations	Responsible, accountable, consulted, informed (RACI) model, accountable executive, audit rights, mandatory incident classification and reporting	Official documents, regulatory text

Note(s): TEVV = testing, evaluation, verification and validation; SBOM = software bill of materials; CBOM = cryptographic bill of materials; SCAS = security assurance specifications; NESAS = network equipment security assurance scheme; RACI = responsible, accountable, consulted, informed
Source(s): Author's own work

Figure 1 Cross-jurisdictional and standards gap map for telecom AI governance. Maturity: full = strong, half = partial, quarter = emerging, empty = gap; an asterisk marks a compound assessment. Coding follows the rubric in Supplementary material A and the sources in the reference list and is analytical, not a legal-enforceability assessment. SDOs = standards development organisations



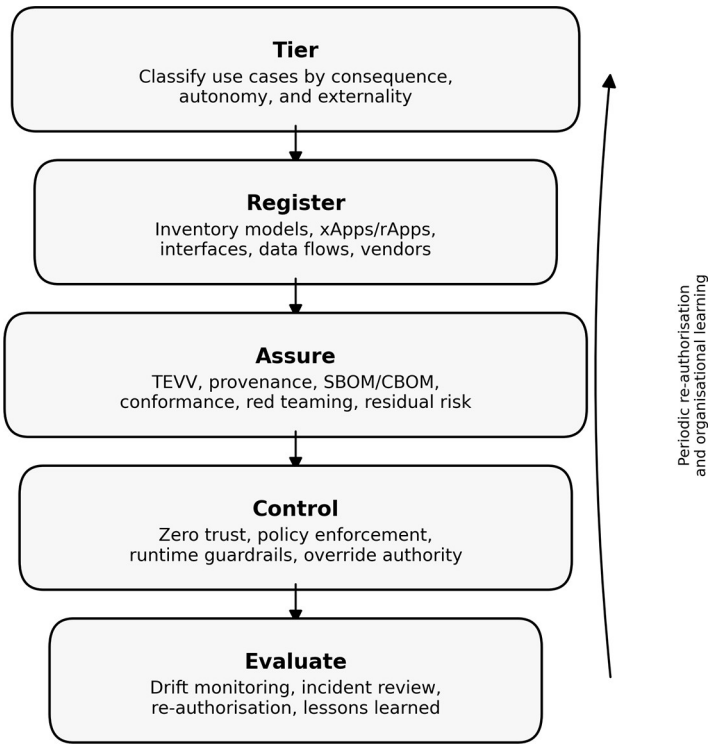
Source: Author's own work

The TRACE-5G/6G governance framework

The documentary synthesis supports a five-part governance model applicable by regulators and operators without waiting for a single universal telecom AI law. Figure 2 illustrates it as a closed-loop assurance cycle, organised around five sequential and recursive functions. The framework's contribution is not another principle list but an organising device: it arranges dispersed and overlapping obligations into five lifecycle control points – risk classification, inventory and provenance, pre-deployment assurance, runtime constraint and post-deployment reauthorisation – that existing instruments address only in dispersed or non-telecom-specific form, as Table 2 sets out.

Tier. Use cases are classified by consequence, autonomy and external dependency: traffic steering in a low-consequence enterprise slice should not be governed like emergency-service prioritisation, autonomous failover, or infrastructure safety functions, following the contextual trustworthiness logic of NIST AI RMF and the risk-based EU AI Act (NIST, 2023; European Union, 2024); emerging explainability standards such as IEEE 2894 provide an architectural reference for the explanation surfaces this requires (IEEE, 2024). Because near-real-time control loops operate under sub-millisecond to low-tens-of-milliseconds budgets, explanation here cannot rely on heavyweight

Figure 2 The TRACE-5G/6G governance framework. Sequential functions of Tier, Register, Assure, Control, and Evaluate close into a feedback loop, with the Evaluate function returning to Tier for periodic reauthorisation. TEVV = testing, evaluation, verification and validation; SBOM/CBOM = software/cryptographic bill of materials; xApps/rApps = near-real-time/non-real-time RAN applications



Source: Author's own work

Table 2 Mapping of TRACE-5G/6G functions to existing instruments and what each consolidates for the telecom domain

TRACE function	*Core requirement*	*Closest coverage in existing instruments*	*What TRACE consolidates for telecom*
Tier	Classify AI use cases by consequence, autonomy and externality	NIST AI RMF (GOVERN, MAP); EU AI Act risk-based classification	Consequence- and autonomy-based tiering specific to network-control AI
Register	Inventory models, xApps and rApps, interfaces, data flows and vendors	NIST AI RMF (MAP); EU AI Act record-keeping and technical documentation	A single register linking AI models to O-RAN interfaces, suppliers and fallback modes
Assure	Pre-deployment TEVV, provenance, SBOM and CBOM, model cards, conformance and residual-risk evidence	NIST AI RMF (MEASURE); EU AI Act conformity assessment; O-RAN SCAS aligned with GSMA NESAS; ETSI TS 104 223	AI-model assurance and O-RAN security assurance combined in a single pre-deployment gate
Control	Zero-trust policy enforcement, runtime guardrails and override authority	NIST AI RMF (MANAGE); EU AI Act human oversight; O-RAN zero-trust architecture	Runtime control bound to specific O-RAN interfaces, with locatable override authority
Evaluate	Drift monitoring, incident review, reauthorisation and lessons learned	NIST AI RMF (MANAGE); EU AI Act post-market monitoring; GCOT oversight principles	Telecom-specific post-deployment reauthorisation and incident-feedback loop

Note(s): TRACE introduces no new control primitives. Its contribution is to organise requirements that the NIST AI RMF, the EU AI Act, and O-RAN/SCAS already address, in dispersed or non-telecom-specific form, into a single telecom-specific lifecycle sequence; the gap map (Figure 1) and Supplementary material A identify where telecom-specific coverage is currently partial or absent
Source(s): Author's own work

post-hoc methods; it calls for lightweight, telecom-appropriate explainability whose computational cost is compatible with operational latency, with fuller explanation reserved for non-real-time and offline review.

Register. Operators maintain a live inventory of models, training inputs, interfaces, certificates, xApps and rApps, suppliers and fallback modes – addressing the visibility and role-allocation problems identified by NTIA and O-RAN WG11 (NTIA, 2023; O-RAN Alliance, 2025).

Assure. Pre-deployment evidence is required: security test results, interoperability validation, robustness testing, SBOM and CBOM records, model cards, bias testing where applicable and operator sign-off for risk thresholds – resonating with ENISA's matrix logic, O-RAN SCAS aligned with GSMA NESAS and NTIA testbed findings (ENISA, 2023; NTIA and ITS, 2025; O-RAN Alliance, 2025).

Control. Runtime governance embeds human oversight, explanation surfaces, policy arbitration, zero-trust constraints and emergency override. GCOT principles call for failover and kill-switch scenarios (GCOT, 2025), and the EU AI Act requires human oversight commensurate with risk (European Union, 2024).

Evaluate. Once-and-done certification yields to continuous assurance: drift detection, incident reporting, post-change revalidation and periodic reauthorisation – the least mature area today and the most important for AI-native networks (O-RAN Alliance, 2025).

Discussion and recommendations

Ethical and institutional implications

Telecom AI assurance is best treated as a socio-technical control regime, not a stack of isolated technical safeguards. Engineering standards alone cannot resolve who may authorise retraining, how conflicting model outputs are adjudicated, when self-healing should yield to human intervention, or how operators demonstrate compliance in multi-vendor environments; horizontal AI laws alone cannot specify the latency, interoperability, resilience and safety requirements of carrier-grade systems. What is needed is an organising layer – governance technically literate enough for telecom and ethically robust enough to discharge the duties expected of critical infrastructure (GCOT, 2026; NIST, 2023; BEREC, 2023; GCOT, 2025). The framework foregrounds, at telecom-specific decision points, four ethical commitments that existing instruments articulate but rarely tie to network-control operations: that AI-mediated outcomes affecting users are explainable to them; that authority over autonomous behaviour is locatable in identifiable actors; that high-consequence decisions remain reversible by human action; and that harms from AI-enabled control loops are remediable through accessible redress. Because telecommunications is infrastructure for other infrastructures, these commitments carry beyond the network itself: automated transport systems, energy-grid synchronisation and other smart-city services increasingly depend on AI-governed telecom control, so a governance failure in the control plane can propagate into the physical systems that rely on it.

For regulators, existing AI and cyber instruments should be profiled rather than replaced, as the forthcoming NIST critical-infrastructure AI profile suggests (NIST, 2026).

For operators, the decisive issue is accountability architecture. Open RAN and AI-native operations do not remove accountability from operators; they intensify it. Boards should treat AI-enabled RAN and orchestration as enterprise-risk subjects, not R&D pilots.

For standards bodies, the next frontier is not more interface detail but operational assurance: model provenance, auditable explanation records, policy-conflict handling, retraining authority, assurance levels for autonomous actions and machine-speed incident sharing.

Comparison with existing instruments

Table 2 maps each TRACE-5G/6G function to its closest coverage in existing instruments – the NIST AI Risk Management Framework, the EU AI Act's high-risk duties, and O-RAN and ETSI assurance work – and identifies what each function consolidates into a telecom-specific, operational form rather than what it invents.

Research priorities

Two lines of work would materially advance the agenda, combining ethical urgency with empirical tractability. Telecom-AI explainability research should establish what level and form of explanation are required for safe operator and user understanding of automated network decisions. 6G-readiness profiling should establish which TRACE-5G/6G obligations ought to harden into binding requirements before 6G deployment.

Indicative directions for future operationalisation and testing

The following directions for future operationalisation focus on the near term; they are hypothetical and require pilot validation rather than being prescriptive, and each step should be calibrated to jurisdiction-specific capacity and the sequencing of binding rules. This near-term horizon (roughly 0–12 months) lays foundations: use-case tiering, an inventory baseline, a drafted telecom-AI governance profile, assurance-evidence scoping, pilot provenance controls, initial runtime-monitoring metrics and incident-reporting templates.

Limitations

This article privileges authoritative, high-confidence documents, but five limitations remain. First, it is a documentary analysis without primary field data or operator interviews. Second, several industry sources used for architectural context are vendor white papers rather than peer-reviewed or official texts, and are treated as supplementary. Third, the pace of change in 6G, agentic AI and telecom automation means some recommendations – especially on runtime assurance and resilient AI – will require iterative refinement as standards mature; emerging profiles for generative and agentic systems, such as the NIST generative-AI profile (NIST, 2024b), indicate the direction this refinement is likely to take. Fourth, the review does not report the duplicate-removal and screening counts standard in a fully audited bibliometric review. Fifth, the gap-map coding was performed by a single coder without inter-coder reliability testing, so the ratings should be read as a transparent analytical judgement rather than a consensus-coded classification.

Conclusion

The hardest ethical question in AI-enabled 5G/6G is not whether AI should be used in telecom networks, but how to ensure AI-enabled control remains explainable, auditable, resilient, fair and bounded by accountable human and institutional authority. Standards bodies have advanced cyber baselines, interface security and architectural clarity; policy-makers have advanced AI risk management, human oversight and critical-infrastructure framing. What is less developed is a consolidated, telecom-specific way to convert trustworthy AI commitments into verifiable institutional behaviour. The gaps the analysis isolates map directly onto the framework's functions: the absence of telecom-specific post-deployment assurance and incident reporting is the work of Evaluate; reliance on general-purpose AI rules without telecom operational depth is what Tier and Assure are meant to localise; and the developmental state of post-quantum-cryptography readiness for 6G control paths is a forward obligation that Register and Evaluate must carry as standards mature.

TRACE-5G/6G is offered as such an organising layer, not as a replacement for NIST, ENISA, 3GPP, O-RAN or statutory regulation, and as a heuristic for future empirical or multi-coder validation rather than a settled solution. It maps available governance principles to telecom-specific decision points: which systems deserve higher oversight, what must be inventoried, what counts as adequate assurance, how runtime autonomy is constrained, and when reauthorisation is required. Whether such a framework supports improvements in cybersecurity, ethical accountability, procurement quality, regulatory coherence and public trust is a question for future empirical work to test, not a property demonstrated here. A central tension runs through any such regime: the transparency, auditability and reporting that accountability demands can compete with the ultra-low-latency, real-time behaviour carrier networks require, so the governance challenge is not to maximise oversight in the abstract but to design assurance that discharges accountability without compromising network performance – favouring lightweight runtime explanation and continuous monitoring at machine speed, with heavier evidence-generation shifted to non-real-time and offline review. On that basis, the near-term priorities are concrete: standards bodies should converge ITU-T, ETSI ZSM, O-RAN and 3GPP work towards auditable runtime assurance and machine-speed incident sharing; industry should treat model provenance, signed releases and lightweight explanation as procurement-grade requirements for xApps and rApps; and regulators should profile existing AI and cyber instruments into a single telecom-AI supervisory map rather than legislate anew, hardening selected obligations into binding form ahead of 6G.

Trustworthy AI governance for telecom therefore belongs within both AI ethics and critical-infrastructure policy. Future networks will not only carry traffic but allocate resources, prioritise services and interact with physical infrastructure in real time, and the ethical governance of those functions must be as mature as the engineering that enables them.

Acknowledgements

The author is employed by the National Board of Revenue (NBR), Government of Bangladesh. This research was conducted in a personal capacity, and the views expressed are the author's own and do not represent the official position of the NBR.

Funding

No funding was received for conducting this study.

Ethics statement

Not applicable. This study is a structured documentary review of public sources and did not involve human or animal subjects.

Data availability

No new datasets were generated or analysed during the current study. All sources used in the documentary review are publicly available official documents, peer-reviewed publications or open industry white papers, accessible through the URLs and DOIs listed in the reference list.

Code availability

Not applicable. No code was generated during the current study.

Consent to participate/Consent to publish

Not applicable. No human participants were involved and no identifiable individual data are reported.

Use of artificial intelligence tools

During the preparation of this manuscript, the author used Claude Opus 4.7 (Anthropic) and ChatGPT (OpenAI, GPT-5.5) only for language editing, copy-editing, formatting-consistency checks and non-substantive structural review of the author's own pre-existing draft text. These tools were not used to generate new scholarly content, data, sources, citations, analysis, findings or the TRACE-5G/6G framework. The author reviewed and verified all edited text, checked all sources and citations, and takes full responsibility for the integrity, originality and validity of the manuscript. No AI tool is listed as an author.

Author contributions

Sushanta Paul: Conceptualisation, methodology, investigation, formal analysis, framework design, writing – original draft, writing – review and editing, and visualisation. The author read and approved the final manuscript.

References

3GPP (2023), "TR 33.877: study on the security aspects of artificial intelligence (AI) and machine learning (ML) for the next generation radio access network (NG-RAN)", Release 18, 3GPP, Sophia Antipolis.

3GPP (2025), "TS 33.501: security architecture and procedures for 5G system", Release 18, 3GPP, Sophia Antipolis.

Ashmore, R., Calinescu, R. and Paterson, C. (2021), "Assuring the machine learning lifecycle: desiderata, methods, and challenges", *ACM Computing Surveys*, Vol. 54 No. 5, p. 111, doi: 10.1145/3453444.

BEREC (2023), "Report on the impact of artificial intelligence (AI) solutions in the telecommunications sector on regulation, BoR (23) 93", BEREC, Riga, available at: www.berec.europa.eu/system/files/2023-06/BoR%20%2823%29%2093%20BEREC%20Report%20on%20impact%20of%20AI%20solutions%20in%20telecom%20sector%20on%20regulation.pdf (accessed 9 May 2026).

Bowen, G.A. (2009), "Document analysis as a qualitative research method", *Qualitative Research Journal*, Vol. 9 No. 2, pp. 27-40, doi: 10.3316/QRJ0902027.

CISA (2025), "JCDC AI cybersecurity collaboration playbook", CISA, Washington, DC, available at: www.cisa.gov/sites/default/files/2025-01/JCDC%20AI%20Playbook.pdf (accessed 9 May 2026).

ENISA (2023), "5G security controls matrix", ENISA, Athens, available at: www.enisa.europa.eu/sites/default/files/all_files/ENISA%205G%20Security%20Controls%20Matrix%20paper.pdf (accessed 9 May 2026).

Ericsson (2025), "AI agents in the telecommunication network architecture", Ericsson, Stockholm, available at: www.ericsson.com/en/reports-and-papers/white-papers/ai-agents-and-network-architecture (accessed 9 May 2026).

ETSI (2025), "ETSI TS 104 223: securing artificial intelligence (SAI); baseline cyber security requirements for AI models and systems", European Telecommunications Standards Institute, Sophia Antipolis, available at: www.etsi.org/deliver/etsi_ts/104200_104299/104223/01.01.01_60/ts_104223v010101p.pdf (accessed 9 May 2026).

ETSI GS ZSM 002 (2019), "Zero-touch network and service management (ZSM); reference architecture, V1.1.1", European Telecommunications Standards Institute, Sophia Antipolis, available at: www.etsi.org/deliver/etsi_gs/ZSM/001_099/002/01.01.01_60/gs_ZSM002v010101p.pdf (accessed 10 May 2026).

European Union (2024), "Regulation (EU) 2024/1689 of the European parliament and of the council of 13 June 2024 laying down harmonised rules on artificial intelligence (artificial intelligence act)", *Official Journal of the European Union, L 2024/1689*, available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ%3AL_202401689 (accessed 9 May 2026).

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P. and Vayena, E. (2018), "AI4People – an ethical framework for a good AI society: opportunities, risks, principles, and recommendations", *Minds and Machines*, Vol. 28 No. 4, pp. 689-707, doi: 10.1007/s11023-018-9482-5.

GCOT (2025), "Principles on AI adoption in the telecommunications industry", UK Government, London, available at: www.gov.uk/government/publications/global-coalition-on-telecommunications-principles-on-ai-adoption-in-the-telecommunications-industry (accessed 9 May 2026).

GCOT (2026), "Security and resilience principles for 6G", UK Government, London, available at: www.gov.uk/government/publications/global-coalition-on-telecoms-security-and-resilience-principles-for-6g (accessed 9 May 2026).

HLEG (2019), "Ethics guidelines for trustworthy AI", European Commission, Brussels, available at: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (accessed 9 May 2026).

IEEE (2024), "IEEE 2894-2024: IEEE guide for an architectural framework and application guidance for explainable artificial intelligence", Institute of Electrical and Electronics Engineers, Piscataway, NJ, doi: 10.1109/IEEESTD.2024.10659410.

IMDA (2026), "Model AI governance framework for agentic AI", IMDA, Singapore, available at: www.imda.gov.sg/-/media/imda/files/about/emerging-tech-and-research/artificial-intelligence/mgf-for-agentic-ai.pdf (accessed 10 May 2026).

ISO/IEC (2023a), *ISO/IEC 23894:2023, Information Technology — Artificial Intelligence — Guidance on Risk Management*, ISO, Geneva.

ISO/IEC (2023b), *ISO/IEC 42001:2023, Information Technology — Artificial Intelligence — Management System*, ISO, Geneva.

ITU-R (2023), "Recommendation ITU-R M.2160-0: framework and overall objectives of the future development of IMT for 2030 and Beyond", ITU, Geneva, available at: www.itu.int/rec/R-REC-M.2160-0-202311-I/en (accessed 9 May 2026).

ITU-T (2019), "Recommendation ITU-T Y.3172: architectural framework for machine learning in future networks including IMT-2020", ITU, Geneva, available at: www.itu.int/rec/T-REC-Y.3172-201906-I (accessed 9 May 2026).

Jobin, A., Ienca, M. and Vayena, E. (2019), "The global landscape of AI ethics guidelines", *Nature Machine Intelligence*, Vol. 1 No. 9, pp. 389-399, doi: 10.1038/s42256-019-0088-2.

Lacava, A., Maxenti, S., Bonati, L., D'Oro, S., Oprea, A., Melodia, T. and Restuccia, F. (2025), "How to poison an xApp: dissecting backdoor attacks to deep reinforcement learning in open radio access networks", *Computer Networks*, Vol. 273, p. 111727, doi: 10.1016/j.comnet.2025.111727.

Liyanage, M., Braeken, A., Shahabuddin, S. and Ranaweera, P. (2023), "Open RAN security: challenges and opportunities", *Journal of Network and Computer Applications*, Vol. 214, p. 103621, doi: 10.1016/j.jnca.2023.103621.

METI (2026), "AI guidelines for business, version 1.2", METI, Tokyo, available at: www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20260331_12.pdf (accessed 10 May 2026).

MIC (2023), "Open RAN security report", Ministry of Internal Affairs and Communications, Tokyo, available at: www.soumu.go.jp/main_sosiki/joho_tsusin/eng/pressrelease/2023/5/20_1.html (accessed 9 May 2026).

Mittelstadt, B. (2019), "Principles alone cannot guarantee ethical AI", *Nature Machine Intelligence*, Vol. 1 No. 11, pp. 501-507, doi: 10.1038/s42256-019-0114-4.

NIST (2023), "Artificial intelligence risk management framework (AI RMF 1.0), NIST AI 100-1", NIST, Gaithersburg, doi: 10.6028/NIST.AI.100-1.

NIST (2024a), "The NIST cybersecurity framework (CSF) 2.0, NIST CSWP 29", NIST, Gaithersburg, doi: 10.6028/NIST.CSWP.29.

NIST (2024b), "Artificial intelligence risk management framework: generative artificial intelligence profile, NIST AI 600-1", NIST, Gaithersburg, doi: 10.6028/NIST.AI.600-1.

NIST (2025), "Adversarial machine learning: a taxonomy and terminology of attacks and mitigations, NIST AI 100-2e2025", NIST, Gaithersburg (2025 revision of NIST AI 100-2; the taxonomy was first published in 2024), doi: 10.6028/NIST.AI.100-2e2025.

NIST (2026), "Draft concept note: artificial intelligence risk management framework – trustworthy AI in critical infrastructure profile", NIST, Gaithersburg, available at: www.nist.gov/programs-projects/concept-note-ai-rmf-profile-trustworthy-ai-critical-infrastructure (accessed 9 May 2026).

NTIA (2023), "Open RAN security report", NTIA, Washington, DC, available at: www.ntia.gov/sites/default/files/publications/open_ran_security_report_full_report_0.pdf (accessed 9 May 2026).

NTIA (2024), "Advancement of 6G telecommunications technology: request for comment, federal register notice", NTIA, Washington, DC, available at: www.ntia.gov/federal-register-notice/2024/advancement-6g-telecommunications-technology (accessed 9 May 2026).

NTIA (2025), "Plotting the path to 6G and supporting the next generation of wireless", NTIA, Washington, DC, available at: www.ntia.gov/blog/2025/plotting-path-6g-and-supporting-next-generation-wireless (accessed 9 May 2026).

NTIA and ITS (2025), "Test plans, results, and lessons learned about open RAN integration during the NTIA 5G challenges, SP-25-578", ITS, Boulder, available at: https://its.ntia.gov/publications/download/SP-25-578.pdf (accessed 9 May 2026).

OECD (2024), "Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449", OECD, Paris, available at: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 (accessed 9 May 2026).

O-RAN Alliance (2024), "Zero trust architecture for secure O-RAN, white paper", O-RAN Alliance, Bonn, available at: https://mediastorage.o-ran.org/white-papers/O-RAN.WG11.ZTA%20for%20Secure%20O-RAN%20White%20Paper-2024-05.pdf (accessed 9 May 2026).

O-RAN Alliance (2025), "O-RAN alliance security update 2025", O-RAN Alliance, Bonn, available at: www.o-ran.org/blog/o-ran-alliance-security-update-2025 (accessed 9 May 2026).

Rinaldi, S.M., Peerenboom, J.P. and Kelly, T.K. (2001), "Identifying, understanding, and analyzing critical infrastructure interdependencies", *IEEE Control Systems Magazine*, Vol. 21 No. 6, pp. 11-25, doi: 10.1109/37.969131.

Roeland, D., Moradi, F., Lokhanko, A., Más, I. and Terrill, S. (2026), "Autonomy by design – enabling self-managing networks across the life cycle", Ericsson Technology Review, available at: www.ericsson.com/en/reports-and-papers/ericsson-technology-review/articles/autonomy-by-design (accessed 9 May 2026).

Sapavath, N.N., Kim, B., Chowdhury, K. and Shah, V.K. (2023), "Experimental study of adversarial attacks on ML-based xApps in O-RAN", *GLOBECOM 2023 – 2023 IEEE Global Communications Conference, IEEE*, *Kuala Lumpur*, pp. 6352-6357, doi: 10.1109/GLOBECOM54140.2023.10437125.

Selbst, A.D., Boyd, D., Friedler, S.A., Venkatasubramanian, S. and Vertesi, J. (2019), "Fairness and abstraction in sociotechnical systems", *Proceedings of the Conference on Fairness, Accountability, and Transparency*, *ACM*, New York, NY, pp. 59-68, doi: 10.1145/3287560.3287598.

Stilgoe, J., Owen, R. and Macnaghten, P. (2013), "Developing a framework for responsible innovation", *Research Policy*, Vol. 42 No. 9, pp. 1568-1580, doi: 10.1016/j.respol.2013.05.008.

Yeung, K. (2018), "Algorithmic regulation: a critical interrogation", *Regulation & Governance*, Vol. 12 No. 4, pp. 505-523, doi: 10.1111/rego.12158.

Yungaicela-Naula, N.M., Sharma, V. and Scott-Hayward, S. (2024), "Misconfiguration in O-RAN: analysis of the impact of AI/ML", *Computer Networks*, Vol. 247, p. 110455, doi: 10.1016/j.comnet.2024.110455.

Supplementary material

The supplementary material for the article can be found online.

Corresponding author

Sushanta Paul can be contacted at: sushanta.researcher@gmail.com