

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΤΕΧΝΟΛΟΓΙΑ ΥΛΙΚΟΥ ΚΑΙ ΛΟΓΙΣΜΙΚΟΥ: ΣΧΕΔΙΑΣΗ ΚΑΙ ΑΝΑΠΤΥΞΗ
ΔΙΑΧΥΤΩΝ ΣΥΣΤΗΜΑΤΩΝ ΥΠΟΛΟΓΙΣΜΟΥ**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

HOU-CS-PGP-2026-09

«Legal and ethical challenges of Artificial Intelligence and 5G/6G communications»

ΑΓΓΕΛΟΣ ΣΙΝΑΠΗΣ

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ: ΠΕΡΙΚΛΗΣ ΧΑΤΖΗΜΙΣΙΟΣ

ΠΑΤΡΑ 2026



HELLENIC OPEN UNIVERSITY

Hardware and Software Engineering: Design and Development of
Diffuse Computing Systems

Master's Thesis

**Ethical Issues in AI-Enabled 5G and 6G Networks: A
Systematic Literature Review**

ANGELOS SINAPIS

Supervisor: Prof. Periklis Chatzimisios

Supervisor: Prof. Achilleas Kameas

Patras, September 2026

Copyright and Open Access Statement

This thesis is the intellectual property of the student (Angelos Sinapis) who prepared it. Under the open-access policy, the author/creator grants the Hellenic Open University a non-exclusive licence to reproduce, adapt, publicly lend, present to the public and digitally disseminate the work internationally, in electronic form and through any medium, for teaching and research purposes, without compensation and for the full duration of intellectual property rights. Open access to the full text for study and reading does not constitute a transfer of the author's intellectual property rights and does not permit reproduction, republication, copying, storage, sale, commercial use, transmission, distribution, publication, performance, downloading, uploading, translation or modification, in whole or in part, without the author's prior express written consent. The author retains all moral and economic rights.



Systematic Literature Review

Ethical Issues in AI-Enabled 5G and 6G Networks

ANGELOS SINAPIS

Supervisory Committee

Supervisor

Prof. Periklis Chatzimisios
Professor, International Hellenic University

Co-Supervisor

Prof. Achilleas Kameas
Professor, Hellenic Open University

Patras, September, 2026

Περίληψη

Η διπλωματική εργασία εξετάζει τα ηθικά ζητήματα που συνδέονται με την αξιοποίηση της Τεχνητής Νοημοσύνης στα δίκτυα 5G και 6G. Στόχος είναι η αποτύπωση της εξέλιξης του πεδίου από το 2020, η αναγνώριση των συχνότερων ηθικών κινδύνων, η ταξινόμηση των προτεινόμενων μεθόδων, πλαισίων και προτύπων, καθώς και ο εντοπισμός των βασικών ερευνητικών κενών. Για τον σκοπό αυτό πραγματοποιήθηκε συστηματική ανασκόπηση βιβλιογραφίας στις βάσεις IEEE Xplore, Springer, ScienceDirect και Web of Science. Μετά τον καθαρισμό, την αφαίρεση διπλοτύπων και την αξιολόγηση τίτλων και περιλήψεων, το τελικό δείγμα περιέλαβε 302 μελέτες που δημοσιεύθηκαν από τον Ιανουάριο του 2020 έως τον Ιούνιο του 2026.

Τα αποτελέσματα δείχνουν ταχεία αύξηση της ερευνητικής δραστηριότητας και σαφή μετατόπιση προς τα δίκτυα 6G και τις AI-native αρχιτεκτονικές. Οι συχνότεροι κίνδυνοι αφορούν την κυβερνοασφάλεια και τις επιθέσεις, την ιδιωτικότητα και την ακατάλληλη χρήση δεδομένων, την αναξιόπιστη συμπεριφορά των μοντέλων, την απώλεια εμπιστοσύνης, και τις αδυναμίες διακυβέρνησης. Ως βασικές λύσεις προτείνονται η μάθηση, η εξηγήσιμη ΤΝ, οι μηχανισμοί εμπιστοσύνης, το blockchain, οι τεχνολογίες προστασίας της ιδιωτικότητας και τα πρότυπα διαλειτουργικότητας. Ωστόσο, οι περισσότερες προσεγγίσεις εξετάζουν επιμέρους προβλήματα και συχνά βασίζονται σε προσομοιώσεις ή εννοιολογικά μοντέλα, χωρίς επαρκή δοκιμή σε πραγματικά τηλεπικοινωνιακά περιβάλλοντα.

Τα σημαντικότερα κενά σχετίζονται με τη ρεαλιστική επικύρωση, τη συνεχή διασφάλιση κατά τον κύκλο ζωής, τη σαφή κατανομή ευθύνης, την ανθρώπινη εποπτεία, τη δικαιοσύνη, τη συμπερίληψη και τη βιωσιμότητα. Συμπεραίνεται ότι η αξιόπιστη ΤΝ στα δίκτυα 5G/6G απαιτεί συνδυασμό τεχνικών, οργανωτικών, νομικών και ανθρωποκεντρικών μέτρων.

Λέξεις-κλειδιά: Τεχνητή Νοημοσύνη, 5G, 6G, ηθική, αξιόπιστη ΤΝ, συστηματική ανασκόπηση βιβλιογραφίας

Abstract

This thesis examines ethical issues associated with Artificial Intelligence in 5G and 6G networks. It maps research since 2020, identifies the most frequently reported risks, classifies proposed methods, frameworks and standards, and determines the main research gaps. A systematic literature review was conducted across IEEE Xplore, Springer, ScienceDirect and Web of Science. After data cleaning, duplicate removal and title-and-abstract assessment, the final corpus comprised 302 studies published between January 2020 and June 2026.

The findings show rapid growth in research activity and a clear shift towards 6G and AI- native network architectures. The most frequent risks concern cybersecurity and adversarial attacks, privacy loss, unreliable model behavior, loss of trust, weak governance and limited explainability. The main responses include federated learning, explainable AI, trust and reputation mechanisms, blockchain, privacy-enhancing technologies, robust security controls and interoperability standards. However, many approaches address only part of the assurance problem and rely on simulations or conceptual architectures rather than realistic operational evidence.

The main gaps relate to real-world validation, continuous lifecycle assurance, responsibility allocation, incident reporting, human oversight, fairness, inclusion and sustainability. The review reveals a mismatch between the rapid development of AI-native networks and the maturity of the safeguards proposed to govern them. Trustworthy AI in 5G and 6G therefore requires coordinated technical, organizational, legal and human-centered measures across design, deployment, monitoring, updating, intervention and redress.

Keywords: Artificial Intelligence, 5G, 6G, ethics, trustworthy AI, systematic literature review

Table of Contents

Περίληψη	iv
Abstract	v
Table of Contents	vi
List of Tables	xi
List of Figures	xiii
Abbreviations & Acronyms	xv
Chapter 1 Introduction	1
1.1 Background and Context.....	1
1.2 Problem Statement	1
1.3 Aim and Objectives.....	2
1.4 Research Questions	2
1.5 Scope and Delimitations	3
1.6 Expected Contribution	3
1.7 Structure of the Thesis	4
Chapter 2 Ethics	5
2.1 The Meaning and Scope of Ethics.....	5
2.2 Major Ethical Traditions	8
2.3 Applied, Digital and Technology Ethics	10
2.4 Artificial Intelligence Ethics	12
2.5 Responsible, Trustworthy and Human-Centric AI.....	13
2.6 Core Ethical Principles for AI Systems	16
2.6.1 Privacy and Data Protection.....	16
2.6.2 Fairness and Non-Discrimination	17
2.6.3 Transparency and Explainability	18

2.6.4 Accountability and Responsibility	18
2.6.5 Human Agency and Oversight	19
2.6.6 Safety, Robustness and Security	19
2.6.7 Social and Environmental Well-Being	20
2.7 Governance, Regulation and Standardization	21
2.8 Ethics in Communication Networks	24
2.9 Chapter Synthesis and Analytical Relevance.....	26
Chapter 3 Artificial Intelligence	28
3.1 Definition and Evolution of Artificial Intelligence	28
3.2 Machine Learning Paradigms	31
3.2.1 Supervised Learning	31
3.2.2 Unsupervised and Self-Supervised Learning.....	32
3.2.3 Reinforcement and Multi-Agent Learning.....	32
3.3 Deep Learning and Advanced AI Methods.....	34
3.3.1 Neural Architectures for Network Data	34
3.3.2 Generative, Neuro-Symbolic and Agentic AI.....	35
3.4 The AI System Lifecycle	36
3.5 Overview of 5G Networks	38
3.6 The Vision and Architecture of 6G	40
3.7 AI Functions in 5G and 6G Networks.....	42
3.8 AI-Enabled Network Applications.....	44
3.9 Technical Features with Ethical Implications	46
3.10 Chapter Synthesis	48
Chapter 4 Chapter 4. Methodology.....	49
4.1 Research Design and Review Scope.....	50

4.2 Research Questions and Analytical Logic	51
4.3 Scoping and Search Strategy	52
4.3.1 Data Sources and Search Output	53
4.4 Eligibility Criteria.....	53
4.4.1 Inclusion Criteria	54
4.4.2 Exclusion Criteria	55
4.5 Data Export, Cleaning and Record Management.....	55
4.6 Title and Abstract Screening	56
4.7 Cross-Database Deduplication and Master Corpus Construction	56
4.8 Final Abstract Reassessment and Study Selection.....	57
4.9 Data Extraction and Coding Framework	59
4.10 Analysis Procedures for the Four Research Questions	60
4.11 Reliability, Traceability and Quality Control	61
4.12 Methodological Limitations and Ethical Considerations	62
4.13 Chapter Synthesis	63
Chapter 5 Research Question 1: Current Status and Evolution of Research on Ethical Issues in AI-Enabled 5G/6G Networks	64
5.1 Purpose and Analytical Procedure	64
5.2 Composition of the Final RQ1 Dataset	65
5.3 Publication Growth from 2020 to June 2026	67
5.4 Distribution Across 5G and 6G Research	69
5.5 Ethical and Responsible AI Themes	71
5.6 AI Approaches and Enabling Technologies.....	74
5.7 Network Functions and Application Contexts	77
5.8 Research Focus and Maturity	80

5.9 Integrated Interpretation and Direct Answer to RQ1	82
5.10 Limitations of the RQ1 Analysis.....	84
Chapter 6 Research Question 2: Ethical Risks in AI-Enabled 5G/6G Networks	87
6.1 Purpose and Analytical Procedure	87
6.2 Operational Definition and Risk Classification	88
6.3 Overall Frequency and Ranking of Ethical Risks	90
6.4 Cybersecurity, Data Integrity and Adversarial Manipulation	94
6.5 Privacy, Surveillance and Improper Data Use	96
6.6 Reliability, Explainability and Trust	97
6.7 Governance, Fairness, Human Autonomy and Societal Impacts	99
6.8 Risk Distribution Across Network Generations, Domains and Stakeholders	102
6.9 Integrated Interpretation and Direct Answer to RQ2	105
6.10 Limitations of the RQ2 Analysis.....	107
Chapter 7 Research Question 3: Methods, Frameworks and Standards for Trustworthy AI in 5G/6G.....	109
7.1 Purpose and Analytical Procedure	110
7.2 Operational Classification of Methods, Frameworks and Standards	111
7.3 Overall Distribution and Temporal Development of Proposed Solutions	112
7.4 Federated, Distributed and Privacy-Preserving Learning	115
7.5 Explainability, Trust, Reputation and Verification	118
7.6 Security, Robustness, Blockchain and Runtime Safety	119
7.7 Standards, Governance, Compliance and Human-Centric Design	121
7.8 Cross-Lifecycle Integration, Network Generation and Evaluation Maturity	123
7.9 Integrated Interpretation and Direct Answer to RQ3	127
7.10 Limitations of the RQ3 Analysis.....	129

Chapter 8 Research Question 4: Remaining Research Gaps for Trustworthy, Accountable and Human-Centric AI in 5G/6G	132
8.1 Purpose and Analytical Procedure	133
8.2 Framework for Identifying and Classifying Research Gaps	133
8.3 Overall Distribution and Temporal Development of Research Gaps	136
8.4 Empirical Validation, Datasets and Benchmarking Gaps	139
8.5 Governance, Accountability, Standardization and Compliance Gaps	143
8.6 Human-Centric Design, Fairness, Inclusion and Stakeholder Participation Gaps ..	145
8.7 Deployment, Security, Safety and Operational Verification Gaps	147
8.8 6G-Specific, Cross-Domain and Sustainability Research Gaps	149
8.9 Integrated Interpretation and Direct Answer to RQ4	150
8.10 Limitations of the RQ4 Analysis.....	154
Chapter 9 Conclusions	156
9.1 Development and Current Status of the Field.....	156
9.2 Most Frequent Ethical Risks.....	156
9.3 Methods, Frameworks and Standards	157
9.4 Remaining Research Gaps.....	158
9.5 Contributions and Implications.....	158
9.6 Limitations and Future Research	159
9.7 Final Conclusion.....	160
References.....	161

List of Tables

Table 1-1. Research questions of the systematic literature review.	3
Table 2-1. Major ethical traditions and their relevance to AI-enabled networks.	9
Table 2-2. Relationship among applied, digital, AI and network ethics.....	12
Table 2-3. Operational distinction among responsible, trustworthy and human-centric AI....	15
Table 2-4. Core ethical principles and their network-specific implications.	21
Table 2-5. Levels of governance and assurance for AI-enabled communication networks. ...	23
Table 3-1. Core AI concepts used in this thesis.	30
Table 3-2. Learning paradigms and their use in communication networks.	33
Table 3-3. Advanced AI model families in 5G/6G research.	36
Table 3-4. Technical orientation of 5G and the emerging 6G vision.	41
Table 3-5. AI functions, methods and application contexts in advanced networks.....	44
Table 4-1. Research questions, evidence fields and analysis methods	51
Table 4-2. Search concept blocks and implementation rules	53
Table 4-3. Inclusion criteria applied during screening	54
Table 4-4. Exclusion criteria applied during screening.....	54
Table 4-5. Database-level screening output before cross-database deduplication	56
Table 4-6. Principal data-extraction and coding fields.....	60
Table 5-1. Distribution of the final RQ1 evidence set by publication period.....	66
Table 5-2. Network-generation scope of the final RQ1 dataset.....	70
Table 5-3. Ethical and responsible AI themes identified in the final corpus.	73
Table 5-4. Frequently represented AI approaches and enabling technologies.	76
Table 5-5. Network functions and application contexts represented in the final corpus.	79
Table 5-6. Abstract-level indicators of research orientation and maturity.	81
Table 6-1. Operational classification scheme used for the RQ2 risk coding.....	90
Table 6-2. Frequency and primary orientation of ethical risks in the final corpus.	91
Table 6-3. Ten most frequent risk co-occurrences.....	95
Table 6-4. Comparison of major risks in overlapping 5G-related and 6G-related studies....	103
Table 6-5. Stakeholder-oriented interpretation of the principal ethical risks.	104
Table 7-1. Operational classification used for the RQ3 coding.....	113
Table 7-2. Frequency and typical form of proposed solution families.	114

Table 7-3. Frequent combinations involving federated, privacy-preserving and traceability mechanisms.....	117
Table 7-4. Standards and regulatory families are explicitly visible in the final corpus.....	122
Table 7-5. Lifecycle position, methods and responsible actors in the RQ3 synthesis.....	125
Table 8-1. Operational category structure used for the RQ4 gap coding.	134
Table 8-2. Evidence types used to distinguish reported and review-inferred gaps.	136
Table 8-3. Frequency and interpretation of the reported RQ4 gap categories.....	140
Table 8-4. Priority research areas derived from the risk-solution-evidence comparison.	142
Table 8-5. Proposed staged research agenda for RQ4.....	153

List of Figures

Figure 2-1. Levels of ethical analysis for AI-enabled communication networks. Author's synthesis.....	7
Figure 2-2. Major ethical traditions as complementary analytical lenses. Author's synthesis.	10
Figure 2-3. Relationship between responsible, trustworthy and human-centric AI. Author's synthesis.....	15
Figure 2-4. Core ethical principles across the AI and network lifecycle. Author's synthesis.	16
Figure 2-5. Multi-level governance and assurance ecosystem for AI-enabled 5G/6G services. Author's synthesis.....	24
Figure 3-1. Relationship among the principal AI families and cross-cutting requirements.	31
Figure 3-2. Main learning paradigms and their sources of feedback.	34
Figure 3-3. AI system lifecycle with continuous monitoring, assurance and feedback.	38
Figure 3-4. Simplified AI-native 5G/6G network continuum	42
Figure 3-5. Technical features and the ethical questions they can generate.....	47
Figure 4-1. Three-phase workflow used for the systematic literature review.	49
Figure 4-2. Master-record traceability and audit structure	58
Figure 4-3. Record identification, screening, deduplication and final abstract-level selection	58
Figure 4-4. Relationship between the common coded corpus and the four RQ analyzes.....	60
Figure 5-1. Annual publication trend of the final RQ1 corpus from 2020 to June 2026.....	69
Figure 5-2. Network-generation scope by publication period.	70
Figure 5-3. Frequency of ethical and responsible AI themes in the final corpus.	74
Figure 5-4. Frequently represented specialized AI approaches and enabling technologies. ...	76
Figure 5-5. Network functions and application contexts represented in the final study set. ...	80
Figure 6-1. Annual development of the six most frequent ethical risks from 2020 to June 2026.	92
Figure 6-2. Frequency ranking of identified ethical risks in the final corpus.....	93
Figure 6-3. Most frequent co-occurrences among ethical risk categories.	99
Figure 6-4. Primary ethical risk orientation of the included studies.....	101
Figure 6-5. Frequency of major risks in overlapping 5G-related and 6G-related groups.....	102
Figure 7-1. Frequency of solution families in the 302 included studies.....	114

Figure 7-2. Annual development of the six most frequent solution families from 2020 to June 2026.....	115
Figure 7-3. Overlapping comparison of solution-family level of occurrence in 5G/B5G-related and 6G-related studies.....	124
Figure 7-4. Research focus and evaluation maturity of proposed approaches.	125
Figure 7-5. Co-occurrence of the eight leading solution families.	126
Figure 8-1. Frequency of explicit research-gap categories in the 302 included studies.	138
Figure 8-2. Temporal development of six recurrent gap categories. The 2026 values cover January to June.....	139
Figure 8-3. Evaluation maturity of the final 302-study review sample.	141
Figure 8-4. Distribution of selected gap categories by network-generation scope.....	147
Figure 8-5. Review-inferred priority matrix combining explicit gap prevalence and validation deficit.....	151

Abbreviations & Acronyms

5G	Fifth-Generation Mobile Network
6G	Sixth-Generation Mobile Network
AI	Artificial Intelligence
AI-RAN	Artificial Intelligence for Radio Access Networks
B5G	Beyond 5G
DL	Deep Learning
EU	European Union
FL	Federated Learning
GDPR	General Data Protection Regulation
IEEE	Institute of Electrical and Electronics Engineers
ITU	International Telecommunication Union
ML	Machine Learning
MNO	Mobile Network Operator
NFV	Network Functions Virtualization
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RAI	Responsible Artificial Intelligence
RAN	Radio Access Network
RQ	Research Question
SDN	Software-Defined Networking
SLR	Systematic Literature Review
XAI	Explainable Artificial Intelligence

Chapter 1 Introduction

Artificial Intelligence (AI) is becoming part of the design and operation of advanced mobile networks. It can improve adaptation, efficiency and service quality, but it also raises questions about automated decision-making, responsibility and the protection of users. Chapter 1 introduces the research problem, aim, objectives, research questions, scope and expected contribution of the systematic literature review.

1.1 Background and Context

Mobile networks have developed from largely hardware-based infrastructures into software-driven systems that combine cloud and edge computing, software-defined networking, virtualized functions and increasingly programmable radio access. Artificial intelligence and machine learning support traffic prediction, resource management, orchestration, fault detection, security monitoring and service optimization. The movement towards AI-native architecture extends these capabilities across the radio, edge and core continuum [123], [124], [141].

The transition to 6G is expected to place intelligence inside the network architecture rather than using it only as an external optimization tool. Distributed and autonomous intelligence may influence access, service quality, mobility, anomaly detection, semantic communication and resource allocation [61], [85], [302]. This autonomy creates ethical concerns because inaccurate, biased, opaque or insecure decisions can affect privacy, safety, equal treatment and human control. AI-enabled networks must be examined as socio-technical systems involving developers, vendors, operators, regulators and users [92], [145].

1.2 Problem Statement

The central problem is the absence of a consolidated account of the ethical issues associated with AI-enabled 5G and 6G networks. Literature identifies requirements involving privacy, fairness, transparency, robustness, accountability, explainability and human oversight. These requirements are expressed through different technical, legal and organizational approaches [16], [17], [58], [214]. Their application is particularly difficult when automated decisions span network domains and change after deployment.

Network intelligence may process users, devices, location and operational data across distributed edge and cloud environments. Responsibility may be divided among data providers,

model developers, vendors, operators and service providers. Relevant studies are also dispersed across engineering, cybersecurity, ethics, law and policy. Some propose technical safeguards such as federated learning, explainability and zero-trust controls, whereas others focus on governance, standards or checks [48], [85], [214], [269]. A systematic review is therefore required to organize the evidence, identify the most frequent risks, classify the proposed responses and assess where real-world evidence remains limited.

1.3 Aim and Objectives

The overall aim is to systematically identify, classify and synthesize research published since 2020 on ethical issues and trustworthy AI approaches in AI-enabled 5G and 6G networks. This review connects network uses of AI with reported risks, proposed governance responses and unresolved research needs.

The specific objectives are to:

1. Map the volume, characteristics and thematic development of the research field from 2020 onwards
2. Identify and rank the ethical risks most frequently reported in studies of AI-enabled 5G and 6G networks
3. Classify the methods, frameworks and standards proposed to support trustworthy and ethically aligned network intelligence
4. Examine how proposed responses are connected to specific risks, stakeholders and stages of the AI lifecycle
5. Identify research gaps relating to trustworthy, accountable and human-centric AI in current and future mobile networks

1.4 Research Questions

Table 1-1 presents the four research questions. Their sequence moves from mapping the field, to identifying risks, examining available responses and establishing the remaining research gaps.

Code	Research Question
RQ1	What is the present status of research concerning ethical issues for AI-enabled 5G/6G networks, and how has this particular domain progressed since 2020?
RQ2	What ethical risks are most frequently identified in the literature on AI-enabled 5G/6G networks?
RQ3	What methods, frameworks, or standards are proposed to support trustworthy and ethically aligned AI in 5G/6G?
RQ4	What research gaps remain regarding trustworthy, accountable, and human-centric AI in 5G/6G?

Table 1-1. Research questions of the systematic literature review.

1.5 Scope and Delimitations

The review covers peer-reviewed journal articles and conference papers published in English from 2020 to the final search date. The records were retrieved from IEEE Xplore, Springer, ScienceDirect and Web of Science. Eligible studies must establish a substantive connection between AI, a 5G or 6G context and an ethical risk, trustworthy AI approach, governance requirement or human-centered concern. Purely technical performance problems are included only when they are linked to effects such as unfair service, privacy loss, unsafe decisions, unclear responsibility or reduced human control. Non-AI network security studies, general AI ethics papers without a 5G or 6G application, grey literature, theses, editorials and non-English publications are excluded. These delimitations create a focused evidence base but may omit relevant industrial and policy material outside the selected academic channels.

1.6 Expected Contribution

The expected contribution is a transparent classification of ethical risks, and an account of how their emphasis varies by year, network generation, application and stakeholder. The study will also organize the methods, frameworks and standards proposed in existing research and distinguish general principles from technical measures and formal governance instruments. By connecting risks with proposed responses, the synthesis shows where the field is relatively mature. It also highlights areas where solutions remain theoretical or weakly validated. This distinction can support researchers, operators, developers, standards bodies and policymakers. The final synthesis will support a research agenda for trustworthy, accountable and human-centric AI in realistic 5G and emerging 6G environments.

1.7 Structure of the Thesis

Chapter 2 develops the ethical concepts and principles used in the assessment. Chapter 3 explains AI applications in 5G and 6G networks and the technical context of the ethical analysis. Chapter 4 presents the systematic review methodology, including searching, de-duplication, screening, quality assessment and data extraction. Chapters 5 to 8 answer RQ1 to RQ4 by examining the field, risks, proposed responses and research gaps. Chapter 9 integrates the findings, discusses limitations and presents the conclusions and priorities for future research.

Chapter 2 Ethics

Ethics provides the conceptual foundation for examining artificial intelligence in fifth generation and sixth-generation communication networks. The technical capabilities of these systems are important, but they do not determine by themselves whether a deployment is acceptable, fair or socially beneficial. AI can improve resource allocation, security, service personalization and network resilience. The same capabilities can also intensify surveillance, reduce meaningful human control, reproduce bias, conceal responsibility and distribute digital opportunities unevenly. Ethical analysis is therefore concerned with the values embedded in design choices, the consequences of automated decisions and the institutions that authorize, monitor and correct those decisions [257].

The relevance of ethics grows as AI becomes a native component of communication infrastructure. In earlier systems, AI was often added to a limited application. In emerging 5G and 6G environments, learning models may participate in network slicing, traffic classification, anomaly detection, semantic communication, edge orchestration and autonomous service management. Decisions can be taken continuously, at high speed and across organizational boundaries. Research on trustworthy network intelligence consequently treats transparency, robustness, privacy, accountability and human agency as system properties rather than optional additions [5], [61], [85], [145], [214].

This chapter develops the ethical framework used throughout the systematic literature review. It first explains the meaning and scope of ethics and introduces major ethical traditions. It then moves from applied and digital ethics to AI ethics, clarifies the relationship between responsible, trustworthy and human-centric AI, and examines seven core principles. The final sections address governance, regulation and standardization before applying the conceptual framework to communication networks. The aim is not to impose one philosophical theory on every technical problem. It establishes a structured vocabulary that supports consistent interpretation of the risks, solutions, and research gaps examined in Chapters 5 to 8.

2.1 The Meaning and Scope of Ethics

Ethics is the systematic examination of how people and institutions ought to act. It concerns reasons for action, the treatment of others, the distribution of benefits and burdens, and the types of social arrangements that can be justified. This normative character distinguishes ethics from a purely descriptive account of what organizations currently do. A widespread practice

may still be unjust, and a legally permitted practice may still be harmful. Ethics also differs from compliance. Compliance asks whether a rule has been followed, whereas ethics asks whether the rule, the decision and its consequences are defensible.

In technological settings, the object of ethical evaluation is broader than the conduct of an individual engineer. A networked AI service is produced through datasets, models, software components, organizational procedures, commercial contracts, standards and regulatory decisions. Harm can emerge even when no single actor intends it. For example, a model may be accurate on average but systematically underperform for a specific population and an operator may deploy an efficient optimization system that disadvantages low-revenue regions, or several individually reasonable components may interact in an unsafe way. Ethical responsibility must therefore be examined at individual, technical, organizational and societal levels.

A useful distinction can be made between the ethics of design, the ethics of use and the ethics of impact. The ethics of design concerns the values and assumptions built into data collection, objectives, model selection and system architecture. The ethics of use examines how a technology is deployed, by whom, under what authority and with which safeguards. The ethics of impact assesses actual and foreseeable effects on people, institutions and the environment. These dimensions are connected. Privacy-preserving architecture can be undermined by poor operational practice, while a technically reliable model may still support an unjust purpose.

Ethics also requires attention to power. AI-enabled communication networks do not merely process neutral information. They shape access to connectivity, visibility, mobility, public services and economic participation. Operators and platform providers can influence which users receive priority, which behaviors are classified as abnormal and which data are retained for future inference. The ethical question is therefore not only whether a model performs correctly, but also who defines correctness, who can contest a decision and who bears the cost of error. Social acceptance research in AI-5G systems confirms that trust and perceived risk affect adoption alongside technical usefulness [10].

Context is important, but contextual judgement should not become arbitrary. Privacy expectations vary across applications, and the level of acceptable automation in entertainment is not the same as in healthcare, transport or emergency communication. Nevertheless, recurring values such as dignity, non-discrimination, safety, accountability and respect for

autonomy provide stable points of evaluation. Ethical analysis combines these general commitments with the specific capabilities, stakeholders and harms of each application.

The analytical levels used in this thesis are summarized in Figure 2-1. The figure shows that model-level properties are only one part of the ethical assessment. A system can include explainable models but remain unaccountable if responsibilities are unclear, or it can comply with organizational procedures while creating unacceptable social exclusion. The levels should therefore be read as connected rather than independent.

Ethical analysis of AI-enabled communication networks

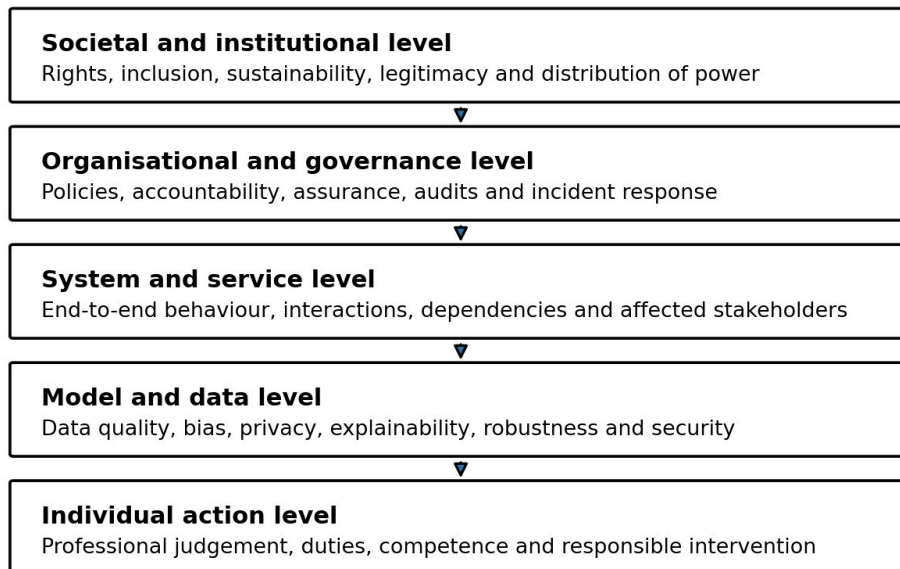


Figure 2-1. Levels of ethical analysis for AI-enabled communication networks. Author's synthesis.

2.2 Major Ethical Traditions

Major ethical traditions provide different questions for evaluating technology. Consequentialist reasoning focuses on outcomes. An AI decision is ethically preferable when it produces better consequences than the available alternatives, taking account of both benefits and harms. This approach is relevant to networks because optimization is already central to their operation. A system may compare latency, energy consumption, coverage, reliability and security. Ethical consequentialism widens the objective by including privacy loss, exclusion, social dependency and environmental costs. The main difficulty is that benefits and harms are not always commensurable, and an improvement in aggregate performance can disadvantage a smaller group.

Deontological ethics evaluates duties, rules and rights rather than only total outcomes. Some actions are unacceptable even when they appear efficient. Personal data should not be collected without adequate justification, people should not be deliberately deceived, and similarly situated users should not be treated differently for irrelevant reasons. A rights-based approach is especially important when network services support healthcare, mobility, education or public participation. It establishes limits on optimization and requires respect for privacy, autonomy, equal treatment and procedural fairness.

Virtue ethics shifts attention from isolated decisions to the character and practical judgement of responsible actors. Engineers, operators, auditors and managers must demonstrate honesty, competence, caution and willingness to correct errors. This perspective is valuable because formal rules cannot anticipate every failure in a complex adaptive network. A professional may need to question an apparently successful model, disclose a limitation or suspend an automated process before a legal obligation is triggered. Virtue ethics also reveals that responsible behavior depends on organizational culture. Individuals cannot exercise sound judgement when incentives reward speed and performance while discouraging the reporting of risk.

Theories of justice examine how opportunities, resources and burdens are distributed. In communication networks, distributive questions arise when capacity, service quality or infrastructure investment is allocated between users and regions. Procedural justice concerns whether affected parties can understand and challenge decisions. Recognition justice asks whether different needs and identities are properly represented. These questions are directly relevant to AI-driven resource allocation and personalization. Fairness assessment in

telecommunications is therefore not limited to removing statistical bias and it also concerns access, representation and the social meaning of differentiated treatment [3], [46], [65], [212],[218].

Care ethics and relational approaches focus on dependency, vulnerability and the quality of relationships. A person may be formally free to reject a service while being practically dependent on it for healthcare, employment or mobility. Care-oriented analysis asks whether the system recognizes this dependency and whether those most exposed to harm have a meaningful voice. It is particularly useful for children, older people, people with disabilities and communities with limited connectivity. Human-centric communication research similarly stresses inclusion, trust and the social conditions of adoption [65], [134], [171].

No single tradition resolves every conflict. A privacy-preserving design may reduce the data available for safety monitoring. Strong security controls may limit user autonomy, while highly personalized services may improve utility but increase surveillance. Ethical analysis is consequently pluralistic. It uses outcomes, rights, justice, professional responsibility and relationships as complementary lenses. A defensible decision should identify the competing values and explain why one balance was selected and provide mechanisms for review when the context changes.

Ethical tradition	Central question	Relevance to AI-enabled networks	Typical limitation when used alone
Consequentialism	Which option produces the best overall outcomes?	Balances service quality, safety, efficiency, access and wider harms.	Aggregate benefits can conceal burdens imposed on minorities.
Deontology and rights	Which duties and rights must be respected?	Protects privacy, dignity, consent, due process and non-discrimination.	May provide limited guidance when valid duties conflict.
Virtue ethics	What would a responsible and competent actor do?	Emphasizes professional integrity, prudence and willingness to intervene.	Depends on institutional conditions that enable good practice.
Justice theories	How should benefits, opportunities and risks be distributed?	Examines access, service quality, allocation and the digital divide.	Different concepts of justice can support different distributions.
Care and relational ethics	Who is vulnerable, dependent or insufficiently heard?	Highlights dependency, accessibility, participation and contextual harm.	Can be difficult to translate into uniform technical metrics.

Table 2-1. Major ethical traditions and their relevance to AI-enabled networks.

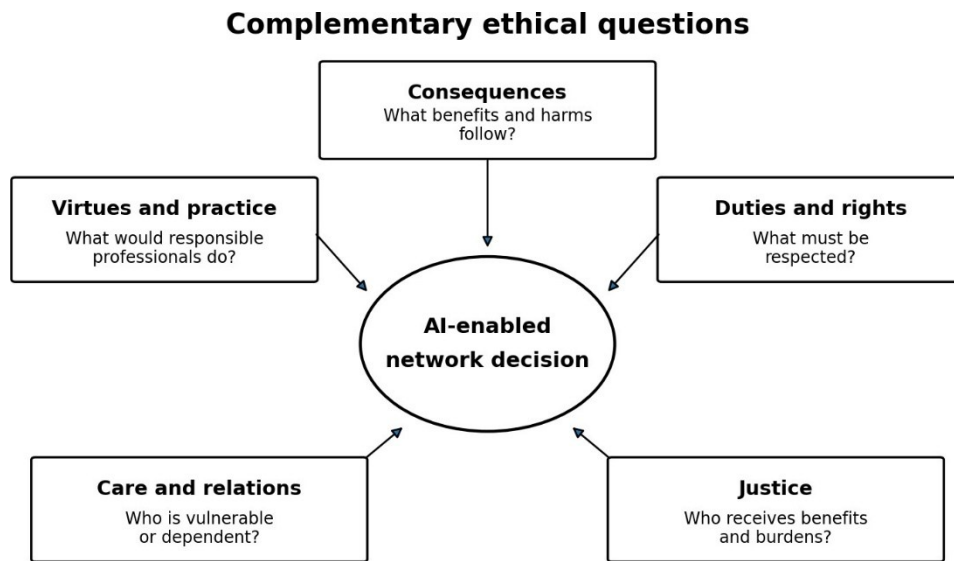


Figure 2-2. Major ethical traditions as complementary analytical lenses. Author's synthesis.

Figure 2-2 presents these traditions as a set of questions around the same AI-enabled network decision. The purpose is not to merge them into one theory, but to prevent an analysis from treating performance, rights or social distribution as the only relevant concern.

2.3 Applied, Digital and Technology Ethics

Applied ethics translates general moral reasoning into the conditions of a particular practice. Medical ethics, research ethics and engineering ethics illustrate this movement from abstract principles to specific responsibilities. The same method is needed for communication networks. Privacy, fairness and safety have different operational meanings in a clinical monitoring service, a vehicle network and a consumer entertainment platform. Applied ethics identifies the relevant stakeholders, likely harms, professional roles and institutional safeguards for each context.

Digital ethics emerged in response to the expanding role of data, platforms and automated mediation. It examines how digital systems affect identity, autonomy, participation and power. Data can be copied, combined and analyzed beyond the setting in which they were collected. Behavioral traces may reveal sensitive characteristics even when explicit identifiers have been removed. The ability to infer, predict and influence behavior therefore creates ethical concerns that are not captured by a narrow view of confidentiality. Research on technology-supported tracking during public health

emergencies demonstrates the persistent tension between collective benefit, surveillance and individual freedom [105], [240].

Digital infrastructure also changes the distribution of power. Individuals often have little knowledge of the organizations, models and data flows that shape the services they receive. Consent may be formally obtained but practically weak when participation in education, work or healthcare requires the system. These asymmetries make governance and contestability central ethical requirements. They also explain why public trust cannot be produced by communication alone. Trust must be supported by trustworthy practices, credible safeguards and evidence that complaints lead to correction [229].

Technology ethics rejects the assumption that technical artefacts are neutral tools whose moral character depends only on later use. Design establishes possibilities and constraints. Network architecture can centralize or distribute control, and a data model can make some users visible and others invisible. An optimization objective can privilege throughput over equitable access. Values are therefore expressed through technical choices, even when they are not explicitly described as ethical choices.

At the same time, ethics should not treat technology as an independent force. Outcomes are produced by socio-technical systems that include people, markets, regulations and organizational routines. The same model can support beneficial healthcare or intrusive surveillance depending on purpose, authority and safeguards. Digital health research shows the potential of data and AI to improve access and care, while studies of smart cities and metaverse environments identify concerns involving privacy, security, inequality and governance [15], [24], [134], [147], [279].

The relationship among applied ethics, digital ethics, AI ethics and network ethics is summarized in Table 2-2. These fields overlap but emphasize different objects. Applied ethics provides sector-specific judgement, digital ethics focuses on data and mediated social life, AI ethics examines learning and automated decision-making, and network ethics considers the infrastructural systems that connect users and allocate communication resources. The present thesis combines all four because AI-enabled 5G and 6G networks operate simultaneously as technical systems, social infrastructures and decision-making environments.

A central implication is that ethical assessment must follow the full chain from data collection to social impact. It should ask who defines the objective, which data makes the decision possible, how the model behaves, how the network implements the output and what remedies remain available. This expanded perspective prevents ethics from being reduced to a checklist applied after deployment.

2.4 Artificial Intelligence Ethics

AI ethics addresses the distinctive moral problems created or intensified by systems that learn from data, generate predictions and participate in decisions. Many concerns have earlier technological forms, but AI alters their scale, speed and opacity. A biased human decision may affect one case and an automated model can reproduce the same pattern across millions of interactions. A security error in a static application may remain local and an error in an autonomous network controller can propagate through interconnected services. Ethical analysis must account for these amplification effects.

Opacity is a second feature. Complex models may produce accurate output without providing reasons that a user, operator or regulator can understand. In networks, the problem is intensified by real-time control and multiple interacting models. An explanation of a single prediction may not explain the behavior of the overall service. Research on safe and explainable 6G systems therefore distinguishes model-level explanations from system-level and runtime assurance [5]. Work on explainable resource management, slicing and semantic communication similarly connects interpretability with trust, auditing and operational control [35], [85], [91], [92], [131]. Additional studies in the corpus address the same point [302].

Field	Primary object of concern	Examples in this thesis
Applied ethics	Normative problems in a defined practice or sector	Healthcare, transport, education, security and public infrastructure.
Digital ethics	Datafication, online interaction and digital power	Surveillance, profiling, identity, access, consent and platform dependence.
AI ethics	Learning, inference and automated decision-making	Bias, opacity, autonomy, robustness, accountability and human oversight.
Network ethics	Infrastructural allocation and connectivity	Slicing, edge intelligence, traffic management, service access and cross-domain control.

Table 2-2. Relationship among applied, digital, AI and network ethics.

Adaptivity and autonomy create a third challenge. AI systems may change as new data arrives, interact with other agents or pursue goals in an environment that was not fully anticipated during design. This makes one-time approval insufficient. Ethical assurance must continue during operation and must detect distribution shifts, adversarial behavior and unexpected interactions. The increasing use of agentic and autonomous network management raises additional questions about authority, escalation and the point at which human intervention remains meaningful [61], [83], [120], [155].

AI ethics also examines the relationship between prediction and decision. A model may estimate congestion, detect an intrusion or predict user demand, but an institutional process determines what happens next. Ethical responsibility cannot be assigned to the model as if it were an independent moral agent. Designers decide the objective, operators define thresholds, organizations allocate authority and regulators establish limits. Accountability must therefore connect technical evidence to human and institutional responsibility [16], [17], [221].

A lifecycle approach is needed because ethical failures can enter at several stages. The purpose may be illegitimate, the dataset unrepresentative, the model insecure, the deployment context unsuitable or the monitoring process ineffective. Correcting only the visible output can leave the source of the problem unchanged. The literature on trustworthy AI in 5G and 6G accordingly proposes frameworks that protect inputs, support explainable operation, validate outputs and monitor systems after deployment [145], [214], [269], [273].

Finally, AI ethics is not limited to preventing direct harm. It also asks what kind of technological future is being created. AI-native networks may expand access, support environmental efficiency and enable new public services. They may also normalize continuous monitoring, deepen dependence on private infrastructure and shift decisions away from affected communities. An adequate ethical framework must therefore evaluate both immediate system properties and long-term social direction.

2.5 Responsible, Trustworthy and Human-Centric AI

Responsible AI, trustworthy AI and human-centric AI are closely related, but they should not be treated as interchangeable labels. Responsible AI emphasizes the processes and actors through which AI is designed, deployed and governed. It asks who

has authority, who must document decisions, who monitors performance and who responds when harm occurs. The concept directs attention to organizational structures rather than only model characteristics.

Trustworthy AI concerns whether confidence in the system is justified. A trustworthy network intelligence system should behave reliably under expected conditions, resist relevant attacks, protect data, avoid unjustified discrimination and provide evidence that supports scrutiny. Trustworthiness is therefore not the same as user trust. People can trust an unsafe system, or distrust a well-governed one. The ethical objective is to establish properties and assurance processes that make trust reasonable [5], [131], [145], [214], [273].

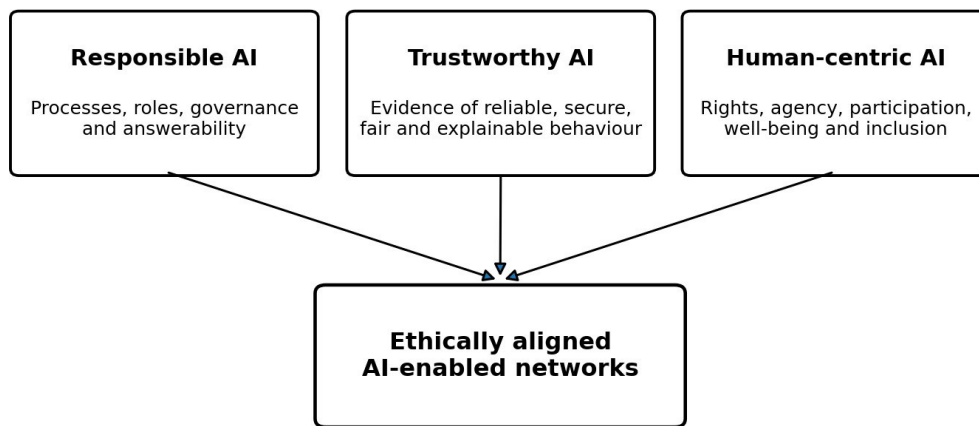
Human-centric AI begins from the position that technical performance serves human and social purposes. It prioritizes dignity, agency, accessibility, participation and well-being. Human-centric design does not mean that a person must manually approve every network action. Many 5G and 6G operations require rapid automation. It means that objectives, limits and escalation procedures remain connected to human values, and that affected people have meaningful protection and routes of contestation. Research on inclusive 6G experiences and human-centric communication makes this connection explicit [65], [171], [254].

The three approaches overlap. Responsible governance is needed to produce trustworthy properties, trustworthy evidence is needed to make human-centered commitments credible and lastly human-centered values guide what organizations should be responsible for. They can nevertheless diverge. A technically robust system may support intrusive surveillance. A participatory project may lack adequate cybersecurity. An organization may publish ethical policies without demonstrating that its models satisfy them. Ethical alignment requires three perspectives to reinforce one another.

Figure 2-3 show on the following page illustrates this overlap. The shared area represents AI-enabled networks that combine governance, demonstrable system properties and respect for human interests. The surrounding areas remain important because each approach contributes to a distinct question. Who is answerable, what evidence supports confidence, and whose rights and well-being are served.

Approach	Primary emphasis	Typical evidence or practice	Main risk if treated in isolation
Responsible AI	Who is responsible and how the system is governed.	Roles, policies, impact assessment, documentation, audit and incident response.	Good procedures may exist without adequate technical performance.
Trustworthy AI	Whether behavior merits justified confidence.	Accuracy, robustness, security, fairness, explanation, traceability and monitoring.	Technical evidence may neglect participation, rights and social purpose.
Human-centric AI	Whether human rights, agency and well-being remain central.	Participation, accessibility, meaningful oversight, contestability and inclusive design.	Broad values can remain aspirational without assurance mechanisms.

Table 2-3. Operational distinction among responsible, trustworthy and human-centric AI. Related but non-identical approaches



The approaches overlap in purpose but emphasise different governance and design questions.

Figure 2-3. Relationship between responsible, trustworthy and human-centric AI. Author's synthesis.

Table 2-3 provides the operational distinction used in this thesis. The distinction is useful for later analysis because proposed responses in the literature occur at different levels. A privacy method or robust model contributes to trustworthiness, an audit framework or incident reporting process contributes to responsibility and participatory design or accessibility requirements contribute to human-centricity. Many mature proposals combine all three [16], [17], [61], [89], [189]. Additional studies in the corpus address the same point [221].

2.6 Core Ethical Principles for AI Systems

Principle-based frameworks translate broad ethical commitments into recurring requirements. The principles used in this thesis are privacy and data protection, fairness and non-discrimination, transparency and explainability, accountability and responsibility, human agency and oversight, safety, robustness and security, and social and environmental well-being. They are consistent with the themes repeatedly identified in the reviewed telecommunications literature, although individual studies use different terms and levels of detail [5], [61], [145], [214], [221]. Additional studies in the corpus address the same point [273], [299]. Principles should not be treated as independent boxes. Privacy mechanisms can affect explainability, fairness interventions can alter performance, and stronger monitoring can support safety while increasing data collection. Ethical design requires an explicit account of these interactions. Figure 2-4 shows that each principle should be considered across the full lifecycle rather than assigned to one technical stage.

2.6.1 Privacy and Data Protection

Privacy concerns the ability of people and groups to exercise appropriate control over information, access and inference. Data protection translates part of this concern into rules and organizational duties concerning collection, purpose, retention, security and disclosure.

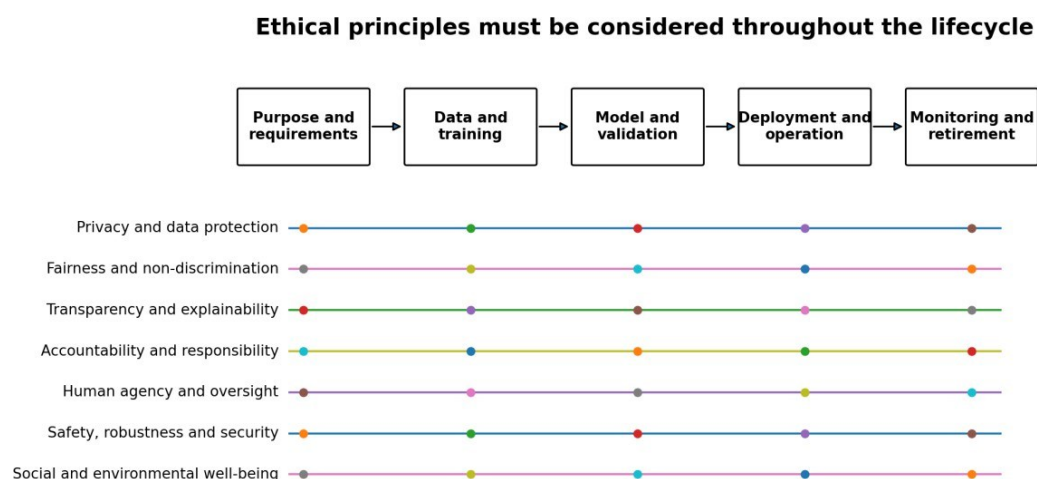


Figure 2-4. Core ethical principles across the AI and network lifecycle. Author's synthesis.

AI-enabled networks process extensive telemetry, device identifiers, traffic patterns, location traces and application data. Even when content is encrypted, metadata can reveal routines, relationships and sensitive contexts. Privacy assessment must therefore consider both direct identification and the inferences that can be produced by combining data sources.

Privacy-preserving technologies include federated learning, differential privacy, encryption, secure aggregation and distributed architecture. These approaches reduce exposures, but none creates privacy automatically. Federated learning can avoid centralizing raw data while still leaking information through gradients or model updates. Differential privacy introduces a utility trade-off, while encryption does not determine whether the purpose is legitimate. Literature consequently combines technical protection with standards, governance and auditing [43], [58], [71], [144], [159]. Additional studies in the corpus address the same point [177], [244], [265], [293]. In this thesis, privacy is treated as both a technical property and a question of purpose, power and user expectation.

2.6.2 Fairness and Non-Discrimination

Fairness concerns whether benefits, burdens and decision quality are distributed without unjustified disadvantages. In AI systems, bias can enter through historical data, sampling, labels, objectives, proxies and deployment conditions. A model may have similar overall accuracy across groups while producing unequally false-positive rates, or it may allocate resources efficiently while systematically giving poorer service to less profitable areas. Fairness therefore requires clarity about the affected population, the relevant comparison and the type of harm being prevented.

Communication networks create distinctive fairness questions. Scheduling and slicing algorithms allocate scarce resources, coverage decisions affect regional opportunity, personalized services can create differentiated quality, and automated security controls can classify legitimate behavior as suspicious. Statistical metrics are useful but cannot select the morally appropriate definition of fairness. The choice requires contextual and institutional judgement. Current research proposes fairness-aware scheduling, explanation-guided fair federated learning and standard-compliant auditing, showing a movement from general concern toward operational assessment [3], [46], [212], [218].

2.6.3 Transparency and Explainability

Transparency is the availability of relevant information about a system, including its purpose, data sources, limitations, governance and decision process. Explainability is a more specific capacity to provide reasons or representations that help a stakeholder understand an output or behavior. The two concepts serve different audiences. A network engineer may require diagnostic details, an auditor may require reproducible evidence, and an affected user may require a clear explanation of consequences and available remedies.

Explainability is valuable because opaque models can undermine error detection, accountability and justified trust. It is not sufficient by itself. An explanation may be technically faithful but incomprehensible, or understandable but misleading. It can also reveal sensitive information or create latency costs. In real-time networks, explanation must be designed with operational purpose in mind. Studies apply explainable AI to intrusion detection, channel estimation, slicing, resource management and semantic communication, while broader work calls for system-level and runtime explanation [1], [5], [6], [8], [35].

Additional studies in the corpus address the same point [85], [91], [92], [131], [163].

Related studies provide further evidence on the same issue [212], [221], [302], [182].

2.6.4 Accountability and Responsibility

Accountability requires actors to explain and justify decisions, accept scrutiny and respond to failure. Responsibility identifies the duties attached to roles before and after an event. In an AI-enabled network, these roles may be distributed among model developers, equipment vendors, operators, cloud providers, application owners and public authorities. The presence of many contributors can create a responsibility gap if each actor controls only one component and no institution remain answerable for the overall service.

Accountability therefore depends on traceability, documentation, auditability, monitoring and incident response. Technical logs must be connected to organizational authority, and escalation procedures must specify when automated operation is restricted or stopped. Regulatory readiness and AI incident reporting are emerging as distinct concerns in telecommunications

because conventional cybersecurity reporting does not capture bias, unsafe behavior or governance failure [16], [17]. Proposed approaches include blockchain- backed traceability, trustworthy analytics, accountable network processes and end-to-end trust frameworks [89], [145], [155], [189], [214]. Additional studies in the corpus address the same point [221], [269].

2.6.5 Human Agency and Oversight

Human agency concerns the capacity of people to form intentions, make informed choices and avoid being reduced to passive objects of automated control. Oversight is the practical arrangement through which humans supervise, challenge or interrupt an AI process. Meaningful oversight is not achieved simply by placing a person in the loop. The person must have sufficient information, competence, authority and time to intervene. Otherwise, oversight can become symbolic and can even transfer blame to an operator who cannot realistically alter the outcome.

Communication networks complicate oversight because many decisions occur in milliseconds. A suitable design may therefore use several levels such as human-defined objectives and constraints before operation, automated monitoring during operation, and human review or appeal after significant decisions. High-risk applications such as connected vehicles, health monitoring and public surveillance require additional safeguards because error or misuse can affect bodily safety and fundamental freedoms [7], [105], [127], [240]. Human-centered research also Emphasizes acceptance, inclusion and participation, showing that oversight should involve affected communities as well as technical experts [10], [65], [113], [134], [166]. Additional studies in the corpus address the same point [171], [258].

2.6.6 Safety, Robustness and Security

Safety is the prevention of unacceptable harm during intended use and reasonably foreseeable misuse. Robustness is the ability to maintain acceptable behavior under variation, uncertainty or disturbance. Security concerns protection against malicious action, while reliability and resilience concern continuity and recovery. These concepts overlap but should not be collapsed. A model can be secure against attack but unsafe because of design error, or robust under normal variation but vulnerable to an adversarial input.

AI-native networks face distribution shifts, poisoning, evasion attacks, compromised agents, faulty sensors and interactions among autonomous controllers. Safety must therefore be considered at model, system and runtime levels. Confidence thresholds, fail-safe controls, causal analysis, explainable detection and continuous monitoring can support intervention, but each introduces new assumptions and operational costs. Literature increasingly combines explainability with robust intrusion detection, trustworthy scheduling, runtime enforcement and resilient orchestration [1], [5], [11], [12], [39]. Additional studies in the corpus address the same point [75], [77], [83], [96], [120]. Related studies provide further evidence on the same issue [125], [131], [145], [179], [184]. Further studies in the corpus support this interpretation [221], [269], [273], [253].

2.6.7 Social and Environmental Well-Being

Ethical AI should contribute to social well-being rather than only avoid individual harm. Relevant concerns include universal access, accessibility, labor effects, surveillance, concentration of power and the digital divide. 6G visions often associate ubiquitous intelligence with healthcare, smart cities and inclusive services, but these benefits depend on affordability, infrastructure and local capacity. A technologically advanced service can widen inequality when it is available only to well-connected populations or when automated personalization excludes users who do not fit expected patterns [2], [65], [134], [171].

Environmental well-being concerns energy use, hardware production, resource consumption and the wider effects of infrastructure. AI can improve energy efficiency and service placement, yet larger models, dense edge deployment and expanding traffic can increase total consumption. Sustainability therefore requires lifecycle assessment and attention to rebound effects, not only a reduction in energy per operation. Research on sustainable 6G architectures, smart cities and trustworthy edge intelligence links environmental efficiency with security, accessibility and long-term social benefit [22], [24], [33], [99], [147]. Additional studies in the corpus address the same point [226], [258], [273], [250], [288].

Principle	Network-specific ethical question	Illustrative mechanisms	Representative references
Privacy and data protection	Are data collection and inference proportionate, secure and compatible with user expectations?	Federated learning, differential privacy, encryption, secure aggregation, retention controls.	[4], [58], [71], [244], [265]
Fairness and non-discrimination	Does allocation or classification create unjustified disadvantage across users, regions or groups?	Bias testing, fairness metrics, constrained optimization, certification and appeal.	[3], [46], [212], [218]
Transparency and explainability	Can relevant stakeholders understand purpose, limitations and important decisions?	XAI, model cards, traceable logs, interpretable policies and stakeholder-specific explanations.	[5], [92], [131], [221], [302]
Accountability and responsibility	Who can justify, correct and accept responsibility for system behavior?	Audit trails, role allocation, impact assessment, incident reporting and governance.	[16], [17], [189], [214], [221]
Human agency and oversight	Can people set limits, intervene and contest significant outcomes?	Human-in/on-the-loop control, escalation, participation, review and remedy.	[7], [10], [65], [127], [171]
Safety, robustness and security	Does the service remain acceptably safe under error, change and attack?	Adversarial testing, runtime monitoring, fail-safe control, redundancy and recovery.	[1], [5], [120], [145], [269]
Social and environmental well-being	Does the network support inclusion and sustainability across its lifecycle?	Accessibility, universal service, energy-aware orchestration, lifecycle assessment and public value criteria.	[2], [22], [24], [258], [273]

Table 2-4. Core ethical principles and their network-specific implications.

Table 2-4 consolidates the seven principles and shows how they can be translated into network-specific questions. The mechanisms are illustrative rather than exhaustive. Their ethical adequacy depends on context, evidence and the way they interact with other principles.

2.7 Governance, Regulation and Standardization

Ethical principles require governance if they are to influence practice. Governance is the set of roles, procedures, incentives and controls through which decisions are made and reviewed. It begins before model development, when an organization defines the purpose and risk appetite of a service, and continues through procurement, deployment, monitoring, incident response and retirement. A governance system should specify who owns each decision, what evidence is required and which conditions trigger escalation.

Technical governance includes testing, documentation, access control, monitoring and assurance. Organizational governance includes leadership responsibility, impact assessment, professional competence and independent review. Sector governance includes standards, interoperability profiles and certification. Public governance includes law, regulatory supervision, public participation and access to remedy. These levels should support one another. Technical controls without organizational responsibility can be ignored, while regulation without measurable requirements can be difficult to enforce.

Risk-based governance is especially relevant to AI-enabled networks because applications differ greatly in potential harm. A model that recommends entertainment content does not require the same safeguards as a controller for emergency communication or autonomous transport. Risk classification should consider purpose, affected population, scale, reversibility and dependency. It should also be updated when an initially limited system becomes integrated into a wider service.

Regulation establishes mandatory limits and duties. Telecommunications AI combines several legal domains, including communications law, cybersecurity, data protection, consumer protection and emerging AI regulation. Cross-jurisdictional readiness is uneven, and existing rules may not recognize incidents involving bias, unsafe learning or unpredictable model behavior. Recent work proposes regulatory readiness assessment and dedicated AI incident reporting to address this gap [16], [17]. Interdisciplinary analysis of connected vehicles also demonstrates how AI regulation and future 6G standards can interact in one application [7], [236].

Standardization can make broad principles operational. Standards define terminology, testing procedures, interoperability requirements and evidence formats. They can reduce uncertainty and support comparison across vendors. Current proposals address fairness assessment, privacy-preserving federated learning, AI-native architectures, wearable health systems and standards-driven governance [3], [58], [95], [123], [165]. Additional studies in the corpus address the same point [218]. Standards are particularly important in telecommunications because components from different organizations must operate together and because network functions can cross national boundaries.

Standards also have limits. Conformance can become a minimum target rather than evidence of ethical adequacy. A standard may lag technology, focus on easily measured properties or reflect the interests of actors with greater resources. Certification therefore should not replace contextual impact assessment, continuous monitoring or stakeholder participation. A system that passes a laboratory test can still fail after deployment when users, data or operating conditions change.

Assurance connects principles, evidence and governance. It asks whether claims such as safe, fair or trustworthy are supported by appropriate tests and records. Trustworthy analytics in zero-touch management, fairness certification and standards-compliant architectures show how assurance is entering the telecommunications literature [189], [218], [268]. Figure 2-5 presents this process as a set of nested levels around the AI-enabled service. The figure Emphasizes that no single technical mechanism can carry the full burden of ethical governance [291].

Table 2-5 on the previous page summarizes the actors, instruments and evidence associated with each level. The table also supports the analysis in later chapters, where methods, frameworks and standards are evaluated not only by their stated goal but also by the governance level at which they operate.

Governance level	Main actors	Typical instruments	Evidence of effective implementation
Technical	Developers, data scientists, network engineers	Testing, privacy controls, explanations, monitoring, access control and fail-safe mechanisms.	Validated metrics, logs, red-team results, incident records and reproducible tests.
Organizational	Operators, vendors, management, compliance and audit functions	Policies, role allocation, impact assessment, procurement rules and incident response.	Clear ownership, documented decisions, independent review and corrective action.
Sector and standards	Standards bodies, industry alliances, certification organizations	Interoperability profiles, assurance requirements, rating schemes and certification.	Comparable assessments, conformance evidence and update procedures.
Regulatory and societal	Legislators, regulators, courts, civil society and affected communities	Law, regulatory guidance, public consultation, enforcement and access to remedy.	Rights protection, transparency, sanctions, participation and cross-border coordination.

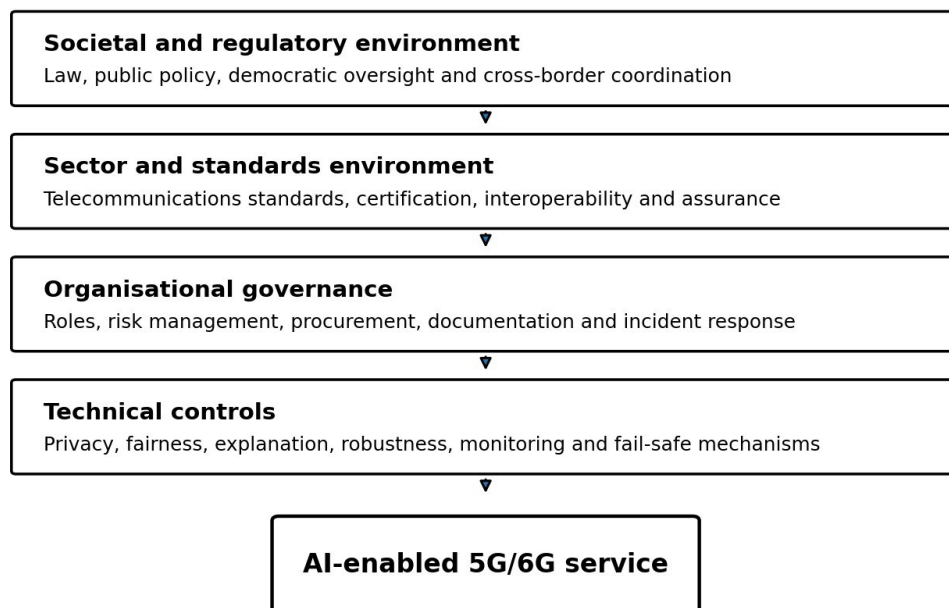
Table 2-5. Levels of governance and assurance for AI-enabled communication networks.

2.8 Ethics in Communication Networks

Communication networks are ethically significant because they are infrastructures through which other activities become possible. They mediate healthcare, mobility, education, commerce, emergency response and social interaction. Failures can therefore spread beyond the network itself. A discriminatory allocation policy can affect access to a public service, a privacy breach can reveal location and relationships, and an unreliable controller can disrupt several dependent systems. Ethical evaluation must account for this infrastructural reach.

5G and 6G architectures introduce several features that intensify the ethical challenge. Network slicing allows differentiated virtual services over shared infrastructure. Edge intelligence moves processing closer to users and devices. Open radio access networks distribute functions across vendors.

Multi-level governance and assurance



Governance layers jointly define authority, evidence, safeguards and remedies.

Figure 2-5. Multi-level governance and assurance ecosystem for AI-enabled 5G/6G services. Author's synthesis.

Federated learning coordinates models across locations, while digital twins and semantic communication create new forms of representation and automation. Each feature offers technical benefits and creates new questions about data control, responsibility, interoperability and oversight [43], [48], [58], [123], [244]. Additional studies in the corpus address the same point [302], [251].

Network slicing illustrates the interaction of efficiency and justice. AI can allocate resources dynamically according to demand and service-level objectives. This can improve reliability, but it can also prioritize commercially valuable services or reproduce historical patterns of underinvestment. Fair and explainable slicing research responds by combining performance objectives with fairness constraints and explanations [78], [115], [125], [158], [188]. Additional studies in the corpus address the same point [212]. The ethical issue is not whether differentiated service is always wrong, but whether the criteria are legitimate, transparent and open to review.

Edge and federated intelligence illustrate the interaction of privacy and accountability. Keeping data near the source may reduce central exposure, but distributed learning increases the number of actors and components involved. Model updates can leak information, malicious participants can poison training and responsibility may become fragmented. Privacy-preserving mechanisms therefore need authentication, robustness, traceability, and governance [43], [48], [58], [71], [159]. Additional studies in the corpus address the same point [177], [265], [293]. A claim that data never leaves the device is not a complete ethical justification [34], [224], [246].

Autonomous and zero-touch management illustrates the interaction of speed and human control. AI can detect faults, allocate resources and recover services faster than manual operation. Yet the same speed can make errors difficult to observe before they propagate. Explainability, causal analysis, confidence measures and runtime safety controls are proposed to make automated management more accountable [35], [120], [155], [163], [169]. Additional studies in the corpus address the same point [189], [221]. Human oversight must be designed at the correct level by setting objectives and limits, supervising patterns of behavior and retaining authority over high-impact exceptions.

Security illustrates the interaction of protection and surveillance. Networks require monitoring to identify attacks, fraud and abnormal traffic. AI improves detection but may also classify legitimate users incorrectly or justify extensive data collection. Security decisions can restrict access, reroute traffic or trigger investigation. Explainable intrusion detection and transparent assurance help operators review these outcomes, but they do not remove the need for proportionality, data minimization and remedy [1], [4], [39], [76], [101]. Additional studies in the corpus address the same point [179], [184].

The network context also expands the stakeholder set. Relevant actors include users, non- users affected indirectly, operators, vendors, cloud providers, application developers, public authorities, standards organizations and communities. Their interests are not identical. Operators may value efficiency and resilience, users may prioritize privacy and affordability, regulators may require accountability, and communities may focus on coverage and environmental impact. Ethical governance must make these conflicts visible rather than allowing one optimization objective to decide them implicitly [210].

For these reasons, ethics in communication networks should be embedded by design and maintained through operation. Ethical-by-design does not mean that values can be solved once in architecture. It means that purpose, risk, evidence and responsibility are considered from the beginning, documented in technical and organizational decisions, monitored after deployment and revised when harms or conditions change. Trustworthy frameworks for B5G and 6G increasingly adopt this lifecycle and multi-stakeholder view [61], [127], [145], [214], [258]. Additional studies in the corpus address the same point [269], [273].

2.9 Chapter Synthesis and Analytical Relevance

This chapter has established ethics as a multi-level examination of actions, systems, institutions and social outcomes. Ethical analysis is not equivalent to legal compliance, technical reliability or public acceptance, although each can provide relevant evidence. A system may be lawful but unjust, reliable but intrusive, or popular but unsafe. The evaluation must consider purpose, process, distribution, rights, consequences and the availability of remedy.

The major ethical traditions provide complementary questions. Consequential reasoning examines benefits and harms. Rights-based reasoning establishes limits, virtue ethics focuses on professional judgement, justice examines distribution and procedure and care ethics highlights dependency and vulnerability. The thesis adopts a pluralistic approach because AI-enabled networks create conflicts that cannot be resolved by one metric. The ethical quality of a decision depends on how openly these conflicts are recognized and justified. Responsible, trustworthy and human-centric AI provide three connected perspectives. Responsible AI identifies actors and governance processes. Trustworthy AI requires evidence that systems behave in a reliable, secure, fair and explainable manner. Human-centric AI ensures that rights, agency, participation, and well-being guide the system's purpose. Later chapters use these distinctions to avoid treating every technical protection as a complete ethical solution.

Seven principles form the coding and interpretation framework. Privacy and data protection, fairness and non-discrimination, transparency and explainability, accountability and responsibility, human agency and oversight, safety, robustness and security, and social and environmental well-being. The principles are applied across the lifecycle and are interpreted in relation to the network context. Their interactions are as important as their individual presence. A privacy mechanism can reduce transparency, a security control can limit autonomy, and an efficiency improvement can create unequal access [62], [156], [178].

The framework supports the four research questions in different ways. RQ1 uses the principles to map the thematic development of the field. RQ2 treats a risk as a plausible pathway through which one or more principles may be violated. RQ3 distinguishes technical methods, governance frameworks and standards according to the principle and lifecycle stage they address. RQ4 identifies gaps where important risks lack mature responses, where evidence is limited or where human and institutional dimensions are underdeveloped.

This distinction between risk, response and gap is essential. Privacy is a principle, unauthorized inference is a risk, differential privacy is a possible method, a privacy governance framework is an organizational response and the absence of real-world validation is a research gap. Similar distinctions apply to fairness, explanation, accountability and safety. Keeping the categories separate makes the later synthesis clearer and prevents the frequency of a term from being mistaken for the effectiveness of a solution.

The chapter also establishes why communication networks require a dedicated ethical analysis. They are pervasive infrastructures, operate in real time, connect multiple organizations and increasingly delegate control to adaptive AI. Their effects extend to sectors that depend on connectivity. Ethical assurance must therefore be end-to-end, cross-organizational and continuous. It must combine technical evidence with governance, standards, regulation and participation [88], [152].

On this basis, the empirical chapters can evaluate the literature without reducing ethics to cybersecurity or using responsible AI as a broad slogan. Chapter 5 examines how the field has developed, Chapter 6 analyzes the risks that dominate literature. Chapter 7 evaluates the proposed methods, frameworks and standards, and Chapter 8 identifies the remaining gaps. The ethical framework developed here provides the common language that connects those analyzes.

Chapter 3 Artificial Intelligence

Artificial intelligence is becoming a structural component of modern communication networks rather than a separate software feature. In fifth-generation networks, learning methods already support traffic prediction, resource allocation, anomaly detection, mobility management and service orchestration. The move toward sixth-generation systems extends this role. AI is increasingly expected to participate in the design, operation and continuous adaptation of the network itself. This transition changes the technical meaning of network intelligence because decisions are distributed across cloud, edge, radio and device environments and must often be produced under strict latency, reliability and energy constraints [41], [82], [124], [228], [239].

The term artificial intelligence covers several families of methods with different assumptions and operational properties. Rule-based reasoning, statistical machine learning, deep neural networks, reinforcement learning, federated learning, generative models and agentic systems do not solve problems in the same way. They require different forms of data, supervision, computation and evaluation. These differences are important for the present thesis because ethical risks and governance needs depend partly on how a model learns, where it is deployed and how directly it affects network users and services [54], [55], [141].

This chapter provides the technical foundation for the systematic literature review. It defines the main AI concepts used in the selected studies, explains the learning paradigms and advanced model families, describes the AI lifecycle and outlines the evolution from 5G automation to AI-native 6G architectures. It then examines the network functions and application domains in which AI is used and identifies the technical properties that later become relevant to privacy, fairness, transparency, accountability, safety and human oversight. The aim is not to evaluate the findings of the review in advance, but to establish a consistent vocabulary for Chapters 5 to 8 [61], [85], [185], [191].

3.1 Definition and Evolution of Artificial Intelligence

Artificial intelligence can be understood as the design of computational systems that perform tasks associated with perception, prediction, reasoning, decision-making, language use or adaptive control. This definition is deliberately broad. It includes systems that follow explicitly encoded rules as well as systems that infer patterns from data. In communication networks, AI may classify traffic, predict channel conditions, detect abnormal behavior, allocate radio resources or coordinate autonomous control loops. The common element is that the system

transforms observations into outputs that support or directly execute a decision [41], [141], [185].

A distinction should be made between artificial intelligence, machine learning and deep learning. Artificial intelligence is the wider field. Machine learning is a major subfield in which behavior is learned from examples, observations or interaction rather than specified entirely through fixed rules. Deep learning is a further subfield that uses multilayer neural networks to learn complex representations. Generative AI uses learned distributions to produce new content, plans, code, signals or explanations, while agentic AI combines models with memory, tools and goal-directed action. These categories overlap, but they should not be treated as synonyms [54], [56], [63].

The historical development of AI has moved through alternating periods of symbolic reasoning and data-driven learning. Early systems relied heavily on logic, search and expert rules. Their behavior was often easier to trace, but their performance depended on manually represented knowledge and was difficult to scale in uncertain environments. The growth of digital data, computational power and statistical learning shifted attention toward models that infer decision boundaries from examples. Neural methods then expanded this shift by learning internal representations directly from high-dimensional data [41], [176].

Communication networks became an important field for applied AI because they continuously generate structured measurements and require repeated decisions. Traffic records, radio measurements, mobility traces, alarms, topology information and service-level indicators provide data for prediction and control. At the same time, the network environment is non-stationary. User demand changes, radio channels vary, devices join and leave, attacks evolve and service objectives conflict. This makes network intelligence a continuous learning problem rather than a one-time classification exercise [55], [82], [101].

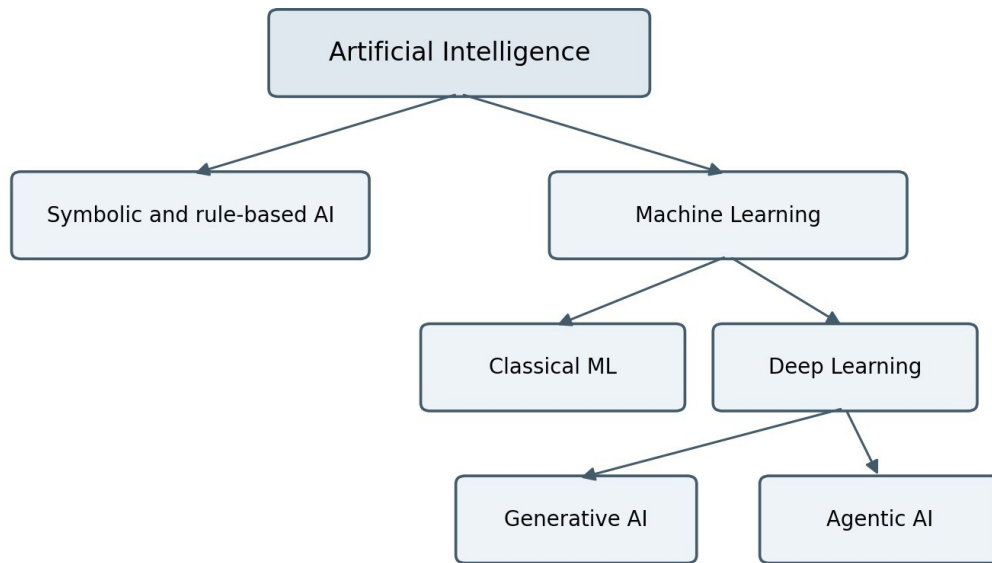
The role of AI has also changed across network generations. In earlier deployments, AI was usually placed around the network as an optimization or analytics tool. In 5G, software-defined networking, network function virtualization, edge computing and network slicing created more programmable environments in which learning could influence operational decisions. In 6G research, AI is increasingly described as native to architecture, meaning that data collection, model training, inference, orchestration and assurance are considered part of the network design itself [104], [123], [124].

AI-native does not mean that every function should be replaced by a learned model. Deterministic protocols remain preferable when requirements are stable, formally specified and safety critical. AI is valuable when the relation between inputs and desired outputs is complex, uncertain or dynamic. Sound architecture therefore combines learned components with conventional control, verification, fallback procedures and policy constraints. This hybrid view is reflected in work on trustworthy orchestration, runtime safety and neuro- symbolic control [12], [120], [155].

The conceptual hierarchy used in this thesis is summarized in Table 3-1. Figure 3-1, presented on the following page, illustrates the relationship among the main AI families without implying that development follows a single linear path. The figure also shows that explainability, robustness and governance are cross-cutting requirements rather than separate model classes [85], [92], [214].

Concept	Operational meaning	Typical network role	Key distinction
Artificial intelligence	Wider field of computational perception, prediction, reasoning and adaptive action	Network analytics, optimization and control	Includes learned and rule-based systems
Machine learning	Models learn patterns from data or interaction	Traffic prediction, classification and allocation	Behavior is inferred rather than fully specified
Deep learning	Multilayer neural models learn complex representations	Signal, sequence, graph and multimodal processing	Higher representation capacity and computational cost
Generative AI	Models produce new content, representations or plans	Synthetic data, semantic communication and assistants	Generation rather than only prediction
Agentic AI	Goal-directed systems combine models, memory, tools and action	Intent translation and autonomous orchestration	Acts through multi-step workflows

Table 3-1. Core AI concepts used in this thesis.



Cross-cutting properties: explainability • robustness • privacy • assurance • governance

Figure 3-1. Relationship among the principal AI families and cross-cutting requirements.

3.2 Machine Learning Paradigms

Machine learning methods are commonly organized according to the form of supervision available during learning. The main paradigms are supervised, unsupervised, self-supervised and reinforcement learning. Semi-supervised and multi-task approaches combine elements of these categories. In network research, the choice of paradigm is influenced by the availability of labelled events, the cost of experimentation, the speed at which conditions change and the degree to which actions affect future observations [176], [244].

Supervised learning uses examples paired with target labels or values. Classification models may distinguish normal from malicious traffic, identify service classes or recognize device behavior. Regression models may estimate channel quality, latency, energy demand or traffic volume. Supervised methods are attractive because their performance can be measured against known targets. Their main limitation is dependence on representative labelled data, which are often expensive to obtain and may become outdated when network conditions change [1], [39], [101].

3.2.1 Supervised Learning

A supervised learning pipeline begins with feature construction or representation learning, followed by training, validation and testing. In classical machine learning, features may be engineered from packet statistics, radio indicators or topology measures. Deep models can learn features directly from sequences, images, graphs or raw signals. Regardless of model

type, the training data define the empirical world that the model can recognize. Rare attacks, unusual mobility patterns and minority user contexts may therefore be poorly represented [44], [96], [179].

Supervised models are widely used in intrusion detection, channel estimation, quality prediction and resource management. Their operational value depends on more than average accuracy. False positives can create service interruptions, while false negatives can leave attacks undetected. Calibration, class-specific recall, uncertainty and robustness under distribution shift are therefore important evaluation dimensions. Studies on trustworthy channel estimation and explainable intrusion detection illustrate the movement from single accuracy measures toward richer evidence about model behavior [1], [77], [179], [175].

3.2.2 Unsupervised and Self-Supervised Learning

Unsupervised learning identifies structure without predefined labels. Clustering, dimensionality reduction, density estimation and autoencoders can reveal traffic patterns, group users, detect outliers or compress network state. These methods are useful when labelled anomalies are scarce or when the purpose is exploratory. However, an unusual pattern is not automatically harmful, and a cluster does not automatically correspond to a meaningful operational category. Human interpretation and contextual validation remain necessary [44], [94].

Self-supervised learning constructs learning signals from the data itself. A model may predict masked elements, reconstruct corrupted observations or compare different views of the same event. This allows large volumes of unlabelled network data to contribute to representation learning before a smaller labelled dataset is used for a specific task. Self-supervised intrusion detection and semantic representation methods are particularly relevant to evolving network environments, although their learned representations may still encode hidden correlations and operational bias [106], [153].

3.2.3 Reinforcement and Multi-Agent Learning

Reinforcement learning addresses sequential decisions. An agent observes a state, selects an action and receives a reward that reflects the result. Through repeated interaction, it learns a policy intended to maximize cumulative reward. This structure is suitable for scheduling, routing, slicing, power control, handover and orchestration because current actions change future network conditions. The challenge is that the reward function compresses several

objectives into one signal and may not capture every operational or social constraint [78], [115], [188].

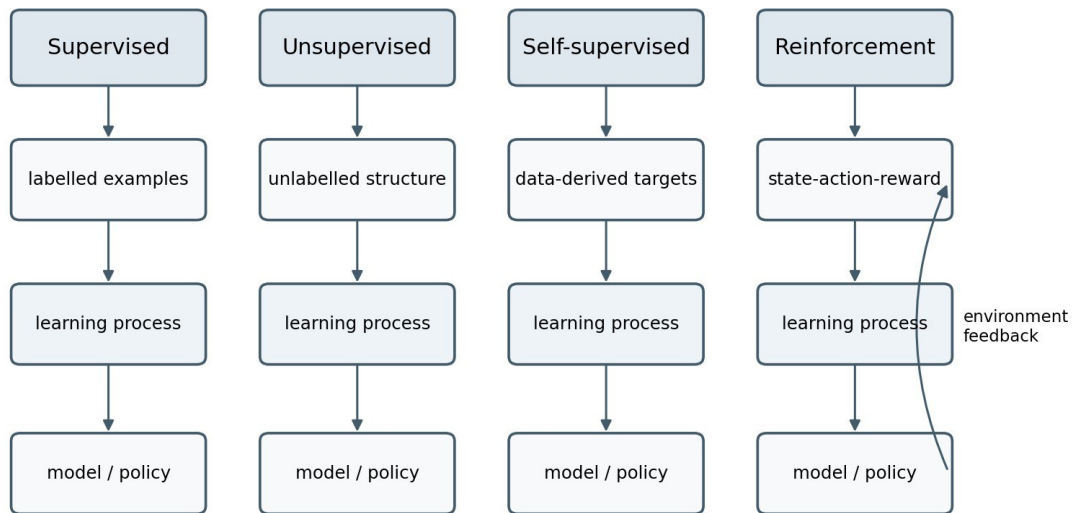
Deep reinforcement learning combines reinforcement learning with neural function approximation. It can manage large state and action spaces, but it can also be unstable, data intensive and difficult to interpret. In live networks, unrestricted exploration is unacceptable because unsafe actions can interrupt services. Training is therefore commonly performed in simulation or digital twins, with constrained policies, offline data, safety filters or fallback controllers used before deployment [121], [132], [155].

Multi-agent reinforcement learning distributes decisions among several agents, such as base stations, slices, edge nodes or autonomous services. Cooperation can improve scalability and responsiveness, but local objectives may conflict and the behavior of one agent changes the environment faced by others. Credit assignment, communication overhead, convergence and accountability become more difficult. Agentic and multi-agent proposals for 6G therefore increasingly combine policy learning with intent constraints, shared knowledge and supervisory mechanisms [57], [126], [159].

Table 3-2 compares the learning paradigms according to their data requirements; typical network uses and principal technical limitations. Figure 3-2, shown on the following page, presents the feedback structure of the main paradigms and highlights the different sources from which a model receives information during learning

Paradigm	Learning signal	Example uses	Principal limitations
Supervised learning	Labelled examples or target values	Intrusion detection, channel and traffic prediction	Labelling cost, imbalance and distribution shift
Unsupervised learning	Structure in unlabelled data	Clustering, anomaly discovery and compression	Outputs require contextual interpretation
Self-supervised learning	Targets generated from the data itself	Representation learning from telemetry or signals	Pre-training objective may not match deployment
Reinforcement learning	Reward from sequential interaction	Scheduling, slicing, routing and orchestration	Unsafe exploration and reward misspecification
Multi-agent learning	Local and shared rewards among interacting agents	Distributed RAN and edge co-ordination	Non-stationarity, coordination and credit assignment

Table 3-2. Learning paradigms and their use in communication networks.



Different learning signals produce different evidence, risks and validation requirements.

Figure 3-2. Main learning paradigms and their sources of feedback.

3.3 Deep Learning and Advanced AI Methods

Deep learning uses layered neural networks to learn representations at multiple levels of abstraction. Earlier layers may identify local or simple patterns, while later layers combine them into task-specific structures. This capability is valuable when manual feature engineering is difficult, as in raw radio signals, long traffic sequences, high-dimensional telemetry and multimodal sensor data. Its cost is greater dependence on data, computation and model selection [54], [96], [141].

3.3.1 Neural Architectures for Network Data

Convolutional neural networks learn local patterns through shared filters. They are best known for image processing, but one-dimensional and two-dimensional convolutions can also process radio measurements, spectrograms, traffic windows and spatial network maps. Their parameter sharing improves efficiency, yet their assumptions about locality may not fully represent long-range dependencies or changing network topology [9], [109].

Recurrent neural networks and gated variants process ordered sequences by maintaining a hidden state. They have been applied to traffic forecasting, mobility prediction and anomaly detection. Their sequential computation can become inefficient for long records, and information may be lost over extended dependencies. Attention mechanisms and transformers

address part of this limitation by allowing the model to weigh relationships across a sequence more directly [72], [100].

Transformers have become important because they can model long-range dependencies and support pre-training on large datasets. In network settings, they can analyze logs, telemetry, protocol sequences and natural-language intents. Their scale creates practical concerns such as training and inference may require substantial computation, and a model trained on broad data may generate plausible but incorrect outputs. For this reason, transformer-based network systems require grounding, constrained tool use and domain-specific validation [51], [63], [186].

Graph neural networks represent entities as nodes and their relationships as edges. They are well suited to network topology, interference relations, mobility graphs, service dependencies and attack propagation. A graph model can aggregate information from neighboring nodes and learn structural patterns that ordinary tabular models may miss. Its performance can nevertheless depend on the quality of the graph definition, and dynamic graphs introduce additional complexity [39], [148].

3.3.2 Generative, Neuro-Symbolic and Agentic AI

Generative AI learns patterns that allow it to create new data or structured outputs. Generative adversarial networks, variational models, diffusion models and large language models can support synthetic data generation, anomaly simulation, semantic communication, code generation and operational assistance. Synthetic data may reduce access barriers or improve rare-event coverage, but it can also reproduce errors, expose memorized information or create unrealistic confidence if the generation process is not validated [54], [89], [153].

Large language models extend generative AI into natural-language reasoning and tool-mediated workflows. They can translate operator intent, summarize alarms, propose configurations, retrieve documentation and coordinate specialized models. In beyond-5G and 6G research, they are being connected to intent-based networking and zero-touch management. Their value lies in flexible interaction, while their main weaknesses include hallucination, prompt sensitivity, uncertain provenance and difficulty guaranteeing deterministic behavior [51], [63], [126].

Neuro-symbolic AI combines learned representations with explicit rules, knowledge graphs or logical constraints. The objective is to retain the adaptability of learning while improving consistency, traceability and reasoning. In network slicing and autonomous control, symbolic constraints can encode service-level limits, safety rules or policy obligations that a purely statistical model might violate. The main challenge is integrating continuous data-driven models with discrete knowledge and maintaining the rule base as the network changes [12], [148], [169].

Agentic AI adds goal-directed planning, memory and action to generative or predictive models. An agent may decompose a task, call monitoring tools, request information, execute a network action and evaluate the result. Multiple agents may specialize in security, optimization, compliance or service management. This architecture can improve modularity, but it also creates new failure paths through tool misuse, conflicting agents and cascading actions. Proposals for AGI-driven and agentic RAN automation therefore place increasing emphasis on permissions, observability and coordination [56], [57], [83], [276].

Advanced models should be selected according to the structure of the task rather than current popularity. A small interpretable model may be more suitable than a large neural model when data are limited and decisions require clear justification. Conversely, high-dimensional or multimodal environments may justify more complex architecture. Table 3-3 summarizes the main model families, the network data they process and their typical strengths and limitations [85], [273].

Model family	Data structure	Network examples	Technical concerns
Convolutional networks	Local spatial or temporal patterns	Signal analysis and anomaly detection	Limited long-range context
Recurrent models	Ordered sequences	Traffic and mobility forecasting	Long-term dependencies and sequential cost
Transformers and LLMs	Long sequences and language	Intent handling, logs and operational assistants	Scale, hallucination and grounding
Graph neural networks	Nodes and relations	Topology, interference and attack propagation	Graph quality and dynamic structure
Neuro-symbolic systems	Learned representations plus rules	Policy-constrained slicing and reasoning	Integration and rule maintenance
Agentic systems	Models, memory, tools and actions	Autonomous service management	Tool misuse and cascading actions

Table 3-3. Advanced AI model families in 5G/6G research.

3.4 The AI System Lifecycle

An AI system is not only a trained model. It is a lifecycle that begins with problem definition and continues through data acquisition, development, validation, deployment, monitoring,

updating and retirement. Each stage introduces technical decisions that affect later performance. A well-performing model can still fail if the objective is poorly specified, the data pipeline changes, the deployment environment differs from testing or operators cannot recognize when the model is no longer reliable [45], [214], [221].

Problem formulation determines what the system is expected to optimize. Network objectives are often expressed through latency, throughput, energy, reliability, cost or security. These metrics can conflict. A model trained to maximize aggregate throughput may reduce service quality for low-volume users, and a security detector tuned for sensitivity may generate operationally unacceptable false alarms. Multi-objective design and explicit constraints are therefore part of the technical lifecycle, not later additions [46], [65], [131].

Data preparation includes collection, cleaning, synchronization, labelling, sampling and feature construction. Network data are distributed across operators, vendors, devices and services, and they may be protected by privacy or commercial restrictions. Measurement errors, missing records, changing protocols and imbalanced events can distort learning. Documentation of data sources and transformations is needed so that later results can be interpreted and reproduced [50], [58], [244].

Model development requires separation between training, validation and test data, together with procedures that prevent leakage. Hyperparameter selection, architecture search and repeated experimentation can indirectly overfit the validation set. Network research also depends heavily on simulators, which may simplify radio propagation, user behavior or attack conditions. A strong evaluation therefore combines controlled experiments with sensitivity analysis and, where possible, testbed or field evidence [45], [120], [189].

Deployment moves the model into a socio-technical environment. Inference may occur in a central cloud, at the network edge, inside a radio controller or on a device. These locations differ in latency, computational capacity, energy use, attack exposure and access to data. Model compression, quantization and hardware acceleration can support low-latency operation, but they may change accuracy and uncertainty. Deployment decisions must therefore be included in validation [108], [157], [181].

Monitoring is necessary because network conditions and model inputs change. Data drift occurs when the distribution of observations changes, while concept drift occurs when the relationship between observations and desired outputs changes. Monitoring should track

performance, uncertainty, resource use, policy violations and unusual actions. It should also define thresholds for human review, rollback or retraining. Runtime safety controls and assurance frameworks are intended to make these responses systematic [120], [189], [221].

Updating a model creates versioning and accountability requirements. A new model may improve average performance but behave differently for specific services or regions.

Operators need records of training data, configuration, evaluation, approvals and deployment history. In distributed learning, the update process also includes participant selection, aggregation and protection against malicious or low-quality contributions. Lifecycle governance therefore depends on technical traceability [48], [89], [293].

Figure 3-3, presented on the following page, shows the lifecycle as a monitored loop rather than a one-way development process. The feedback arrows indicate that operational evidence can lead to revised objectives, new data or model replacement. This structure is used later in the thesis when methods and research gaps are mapped to specific lifecycle stages.

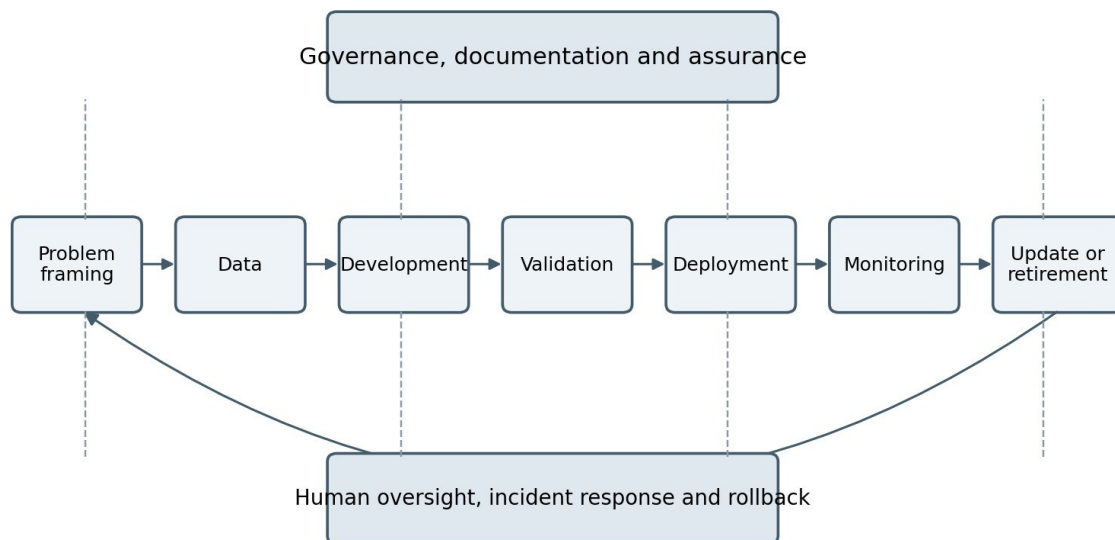


Figure 3-3. AI system lifecycle with continuous monitoring, assurance and feedback.

3.5 Overview of 5G Networks

Fifth-generation mobile networks were designed to support enhanced mobile broadband, massive machine-type communication and ultra-reliable low-latency services. These service classes expanded the range of devices and applications connected to mobile infrastructure. The architecture also became more software driven through virtualised network functions, software-defined control and cloud-native deployment. This programmability created

conditions in which AI could be integrated into management and optimization [41], [104], [176].

Network slicing is a central 5G concept. A physical infrastructure can support several logical networks with different performance objectives. AI can forecast slice demand, allocate resources, detect violations and support admission control. Because slices share infrastructure, the model must manage trade-offs among efficiency, isolation and service guarantees. Explainable resource reservation and slicing studies demonstrate why network decisions increasingly require both predictive performance and operational justification [158], [188].

Multi-access edge computing places computation closer to users and devices. Edge deployment reduces latency and can limit the transfer of raw data to central clouds. It also creates heterogeneous execution environments with different hardware, ownership and security conditions. Edge intelligence therefore involves placement decisions of which model should run, where it should run, what data it can access and how updates should be coordinated [45], [144], [273], [152].

The radio access network is another major area of AI use. Models can support channel estimation, beam management, interference control, mobility prediction and traffic steering. Open and programmable RAN architectures increase the opportunity for third-party applications and intelligent controllers. At the same time, they broaden the attack surface and make assurance across components more difficult [77], [94], [149].

5G security applications use AI to analyze high-volume and rapidly changing traffic. Intrusion detection, fraud detection and analyses can identify patterns that fixed rules miss. However, AI systems are themselves vulnerable to poisoning, evasion and model extraction. Surveys of AI-assisted 5G security therefore treat AI as both a defense mechanism and a new protected asset [1], [101], [150].

The 5G experience provides the operational foundation for 6G. It shows that AI can improve automation but also that integration requires data access, orchestration, monitoring and cross-vendor interoperability. Many 6G proposals extend mechanisms already visible in advanced 5G, including distributed learning, zero-touch management, digital twins and intent-based control. The distinction is one of architectural depth and scale rather than a complete break [104], [123], [124].

3.6 The Vision and Architecture of 6G

Sixth-generation networks are commonly presented as an evolution toward ubiquitous, intelligent and context-aware connectivity. Proposed capabilities include tighter integration of terrestrial and non-terrestrial networks, sensing and communication, extreme edge intelligence, semantic communication and support for highly immersive or autonomous services. Although standards and deployment remain under development, the literature consistently treats AI as a principal enabling layer [2], [21], [55].

The phrase AI-native architecture expresses a stronger form of integration than AI-assisted operation. In an AI-assisted network, models optimize selected functions. In an AI-native network, the architecture is designed to produce data for learning, distribute models, support inference and evaluate intelligent decisions across domains. This requires shared interfaces, model management, telemetry, assurance and coordination between communication and computing resources [61], [123], [124].

6G is expected to combine cloud, edge and device intelligence. Training may occur centrally for large models, collaboratively across several domains or locally for privacy and responsiveness. Inference may move dynamically according to latency, energy, connectivity and confidentiality requirements. This continuum is more flexible than a fixed cloud-edge division, but it makes placement and responsibility more complex [110], [157], [273].

Semantic communication is one example of the change in design philosophy. Conventional systems aim to reproduce transmitted symbols accurately. Semantic approaches attempt to preserve task-relevant meaning, allowing the communication process to adapt to the purpose of the receiver. AI is used to learn representations, select relevant information and reconstruct meaning. The performance of such systems depends on shared models and context, which creates new requirements for robustness and explainability [97], [153], [302].

Integrated sensing and communication allow network infrastructure to support both data transmission and environmental perception. AI can combine radio, visual, spatial and device information to infer location, movement or physical state. This supports transport, industrial and health applications, but it also increases the sensitivity of the data collected. The technical architecture must therefore separate necessary sensing from unrestricted inference and provide mechanisms for access control and audit [18], [127], [128], [289].

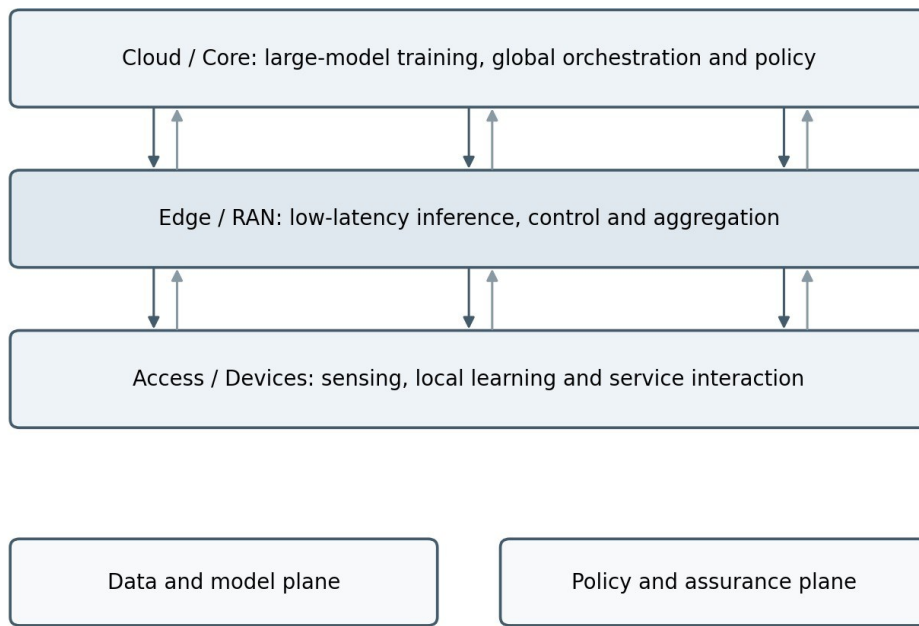
Non-terrestrial networks extend connectivity through satellites, high-altitude platforms and aerial systems. AI can support routing, resource management, handover and adaptation across variable links. These environments have limited energy, intermittent connectivity and large-scale coordination challenges. Federated and distributed learning can reduce central data transfer, although communication efficiency and participant reliability become central design issues [32], [110], [160], [68], [283].

Sustainability is also part of the 6G vision. More intelligence may improve energy allocation, sleep modes and infrastructure use, but training and operating large models can consume significant resources. Architecture design must consider the energy cost of learning alongside the energy savings created by optimization. Sustainable 6G research therefore connects model efficiency, hardware, network utilization and wider environmental objectives [2], [22], [33].

Table 3-4 compares the technical orientation of 5G and the emerging 6G vision. Figure 3-4, shown on the following page, presents a simplified AI-native architecture in which data, models, policies and assurance circulate across devices, edge nodes, the RAN and cloud or core platforms. The figure Emphasizes that intelligence is distributed and that governance interfaces must follow the same distribution [61], [124], [269].

Dimension	5G orientation	Emerging 6G orientation	Implication for AI
Architecture	Virtualized and software- defined functions	AI-native, distributed and cross-domain architecture	Model management becomes architectural
Service model	eMBB, mMTC and URLLC	Immersive, sensing, autonomous and semantic services	More diverse objectives and data
Computing	Cloud and edge integration	Cloud-edge-device intelligence continuum	Dynamic training and inference placement
Automation	AI-assisted optimization and orchestration	More autonomous and intent-driven control	Stronger runtime assurance requirements
Connectivity	Primarily terrestrial mobile infrastructure	Terrestrial and non-terrestrial integration	Heterogeneous participants and constraints
Evaluation	Performance and service-level metrics	Performance, intelligence, trust and sustainability	Broader evidence and lifecycle monitoring

Table 3-4. Technical orientation of 5G and the emerging 6G vision.



AI-native operation links telemetry, models, actions and evidence across the network continuum.

Figure 3-4. Simplified AI-native 5G/6G network continuum

3.7 AI Functions in 5G and 6G Networks

AI supports network functions at several layers. At the physical and radio layers, models estimate channels, select beams, allocate spectrum and adapt transmission parameters. These tasks operate under tight timing constraints and are often evaluated through throughput, error rate and reliability. Uncertainty-aware and explainable channel estimation research shows that even low-level functions can require evidence about confidence and failure conditions [8], [77], [153], [300].

At the access and mobility layers, AI predicts movement, supports handover and balances load among cells or access technologies. Mobility models can reduce interruption and improve resource planning, but they depend on location and behavioral data. Errors may affect users unevenly, particularly in dense, rural or rapidly changing environments. Robust evaluation should therefore include different mobility patterns rather than one average scenario [170], [181].

Resource allocation and scheduling are major application areas because networks continuously distribute limited capacity. Supervised prediction can estimate demand, while reinforcement learning can select actions over time. In sliced networks, a model may decide how much

bandwidth, computing power or priority each service receives. Trustworthy resource management combines performance with explanation, constraints and fairness analysis [46], [115], [131].

Network management and orchestration use AI to coordinate virtual functions, edge services and end-to-end service objectives. Intent-based networking allows operators to express desired outcomes at a higher level, after which an automated system translates them into policies and configurations. Large language models and agentic reinforcement learning are being explored for this translation, but execution must remain constrained by verifiable policies and runtime controls [51], [120], [126].

Security functions include intrusion detection, threat classification, anomaly analysis and adaptive defense. AI can integrate signals from several parts of the network and identify subtle patterns. Explainable systems are intended to help analysts determine why an alert was generated and whether a response is proportionate. The strongest designs combine detection accuracy with uncertainty, traceability and resistance to adversarial manipulation [1], [39], [44].

Federated learning enables several nodes to train a shared model without sending all raw data to one location. It is useful for cross-device, cross-edge and cross-operator settings. The server or coordinator aggregates local updates, but the process must address statistical heterogeneity, unreliable participants, communication cost and attacks on updates. Secure aggregation, differential privacy and reputation mechanisms are common supporting techniques [48], [50], [71], [204].

Digital twins provide simulated or continuously updated representations of network components, services or environments. They can generate training scenarios, test policies and support predictive maintenance. Their usefulness depends on fidelity and a twin that omits important behaviors can produce misleading confidence. Standardized models and validation against operational evidence are therefore necessary [59], [121], [168].

Zero-touch and autonomous networking bring these functions together in closed loops. Monitoring identifies the current state, analytics predict outcomes, a controller selects an action and feedback updates the system. The objective is rapid adaptation with limited manual intervention. Trustworthy autonomy requires boundaries, escalation paths and the ability to explain or reverse high-impact actions [104], [155], [189].

The relationship among network functions, AI methods and operational outputs is summarized in Table 3-5. The table is not intended as a one-to-one mapping. The same function may use several methods, and the same method may support several layers. This many-to-many structure explains why the later SLR chapters code both the AI approach and the network context of each study.

Network function	Common AI approaches	Operational output	Illustrative application contexts
Radio and channel management	Deep learning, GNNs and uncertainty models	Channel estimates, beam and power decisions	Mobile broadband, industrial and aerial links
Mobility and access	Sequence models and reinforcement learning	Handover and traffic steering	Vehicles, public networks and robotics
Slicing and resource allocation	Prediction, DRL and explainable optimization	Admission, scheduling and resource shares	Healthcare, smart cities and private networks
Security and resilience	Supervised, self-supervised and federated learning	Alerts, classification and adaptive response	IoT, transport, industry and critical services
Orchestration and zero-touch control	LLMs, agents, digital twins and neuro-symbolic AI	Policies, configurations and closed-loop actions	End-to-end services and autonomous operations

Table 3-5. AI functions, methods and application contexts in advanced networks.

3.8 AI-Enabled Network Applications

AI-enabled networks support application domains in which communication, computation and decision-making are closely connected. Healthcare is a prominent example. 5G and 6G can connect wearable sensors, imaging systems, remote monitoring and clinical services, while AI analyzes data and supports prediction. The technical requirements include low latency, reliability, interoperability and privacy-preserving learning across institutions [15], [18], [107], [290].

Connected and autonomous transport combines vehicle sensors, roadside infrastructure, edge computing and network coordination. AI supports perception, route planning, collision risk analysis and cooperative decision-making. Communication delays or model errors can have physical consequences, and operating conditions vary sharply across weather, geography and traffic density. Vehicle applications therefore require resilience, fallback behavior and continuous validation [13], [100], [128], [256], [187].

Smart cities use AI to coordinate transport, energy, public safety, environmental monitoring and municipal services. The network integrates data from heterogeneous public and private systems. This creates opportunities for efficient service delivery, but it also creates large data linkages and dependencies on platform providers. Technical designs must support

interoperability and selective data sharing rather than unrestricted centralization [24], [99], [118], [70], [183].

Industrial and Industry 5.0 applications use private networks, edge intelligence, digital twins and federated learning for predictive maintenance, quality control and adaptive production. Reliability and timing are important because model outputs may influence physical processes. Industrial systems also contain legacy equipment and proprietary data, making migration and cross-vendor integration significant challenges [23], [109], [168], [262], [295].

Immersive applications and the metaverse require high data rates, low latency and continuous interaction among users, devices and virtual environments. AI generates content, personalizes experiences and coordinates digital representations. These applications can amplify identity, privacy and safety concerns because behavioral and biometric information may be collected at high frequency. Network and AI security must therefore be treated as a combined problem [14], [147], [197].

Robotics and autonomous systems use communication networks to share perception, coordinate tasks and access remote intelligence. 6G research considers distributed robots, drones and embodied AI that operate across physical and digital environments. Their actions make uncertainty and oversight especially important, because a communication failure or incorrect inference can be translated into physical movement [127], [142], [166].

Agriculture and environmental monitoring illustrate how AI-enabled wireless sensing can support disease detection, resource management and remote operation. These applications often operate in areas with limited connectivity, power and labelled data. Model design must therefore address lightweight inference, coverage gaps and transfer across locations. Benefits also depend on whether systems are affordable and usable for the intended communities [130], [147].

Across domains, the network is not a neutral transport layer. Latency, coverage, resource allocation and data placement influence which AI services are available and how reliably they operate. Application assessment must therefore examine the complete service chain from sensing to inference and action. This system's view is essential for interpreting the ethical risks identified in Chapter 6 and the proposed responses examined in Chapter 7 [65], [171], [258].

3.9 Technical Features with Ethical Implications

The ethical significance of AI-enabled networks arises partly from technical design choices. Data-intensive learning can improve prediction while increasing the amount of information collected or inferred. Distributed learning can reduce raw data transfer while creating new leakage paths through model updates. Low-latency automation can improve service quality while reducing the time available for human review. These are not simple oppositions, but trade-offs that must be evaluated in context [50], [71], [244].

Model opacity is a central example. Deep models may achieve strong performance without producing reasons that operators or affected people can understand. Explainable AI methods can identify influential features, examples or decision paths, but explanations differ in fidelity and audience. A technical explanation that is useful to a radio engineer may be unsuitable for a regulator or user. Explainability must therefore be designed as an interface between the model and specific stakeholders [6], [85], [92].

Data heterogeneity creates both performance and fairness concerns. Devices, regions and services generate different data distributions. A federated model may perform well overall while providing weaker results for participants with less data or different conditions. Participant selection and aggregation rules can reinforce these differences. Fair federated learning and fairness-aware scheduling attempt to make distributional effects visible and controllable [46], [212], [218].

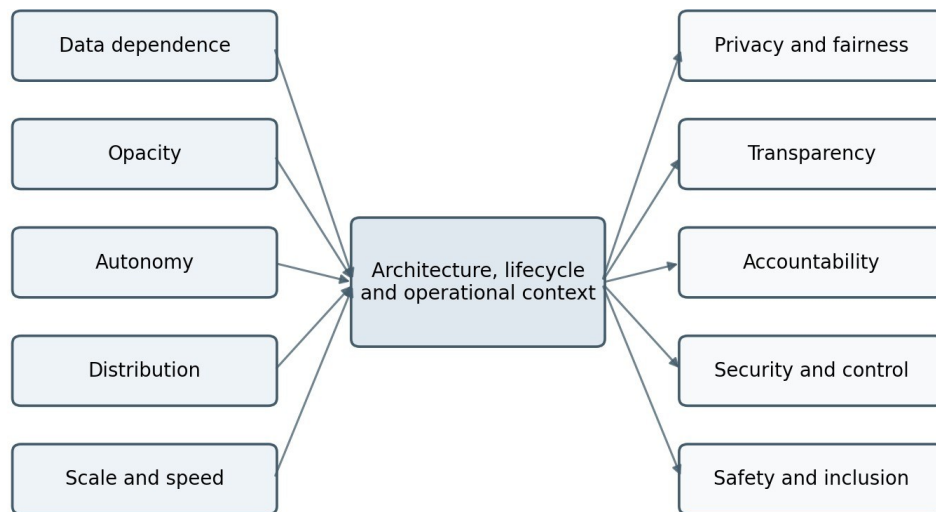
Autonomy creates accountability challenges because decisions are produced through interacting models, controllers and organizations. When an automated slice manager changes resources, the outcome may depend on telemetry, a prediction model, a policy engine and vendor-specific orchestration. Logs alone do not establish responsibility unless the system records why a decision was permitted and which actor was authorized to define the policy [120], [169], [221].

Robustness and safety are also linked to architecture. A model can fail because inputs are noisy, an attacker manipulates observations, the environment changes or several agents interact unexpectedly. Runtime monitors, uncertainty estimation, constrained actions and fallback controllers reduce but do not eliminate this risk. Safety should therefore be treated as a property of the complete control loop rather than of the predictive model alone [5], [75], [145].

Privacy mechanisms can alter utility and explainability. Differential privacy adds noise, encryption limits inspection and secure aggregation hide individual contributions. These mechanisms may be necessary, but their parameters determine the balance between protection and performance. A claim that a system is privacy preserving is incomplete without information about the threat model, leakage channels and operational cost [58], [71], [293].

Standardization and interoperability influence whether technical safeguards can be applied consistently. Shared interfaces for model description, testing, incident reporting and lifecycle evidence can reduce ambiguity across vendors. However, standardization must remain responsive to new model types and distributed architectures. AI-native 6G research therefore links technical architecture with assurance profiles, governance and cross-domain trust [16], [17], [123].

Figure 3-5, presented on the following page, summarizes the connection between technical features and ethical questions. The figure does not suggest that a feature is inherently ethical or unethical. It shows the mechanisms through which data dependence, opacity, autonomy,



Technical features shape ethical questions, but context determines the direction and severity of impact.

Figure 3-5. Technical features and the ethical questions they can generate.

distribution and scale can create risks or support safeguards. This mapping provides the technical bridge between the present chapter and the ethical framework established in Chapter 2 [127], [214], [269].

3.10 Chapter Synthesis

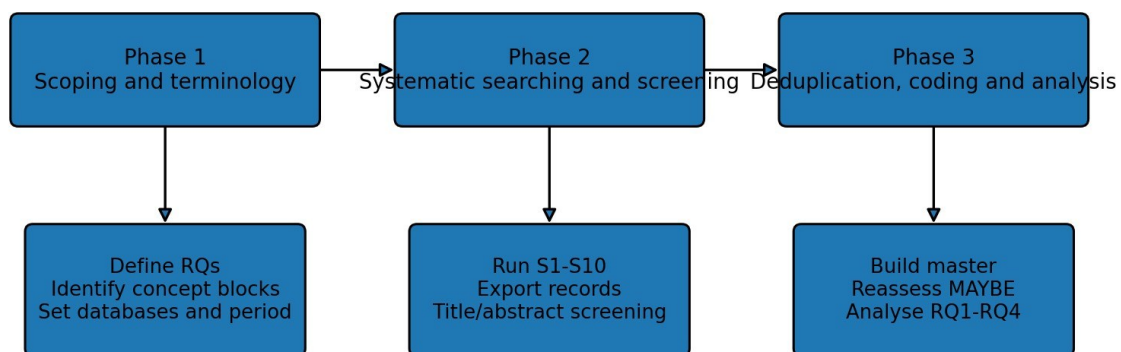
This chapter has presented AI as a group of learning and reasoning approaches embedded across the communication-network lifecycle. The progression from 5G to 6G moves from programmable networks with AI-supported functions toward distributed, AI-native architectures that connect data, models, policies and operational actions. This technical foundation guides the SLR coding of AI approach, network function, application context, evaluation environment and lifecycle coverage. Chapter 4 explains how these relationships were analyzed systematically [45], [61], [124], [214].

Chapter 4 Chapter 4. Methodology

This chapter describes the design and execution of the systematic literature review used to examine ethical issues in artificial-intelligence-enabled 5G and 6G networks. The procedure was developed to support transparent study identification, consistent eligibility decisions, traceable record management and reproducible analysis across the four research questions. The final analytical corpus covers publications from January 2020 to June 2026 and is based on records retrieved from IEEE Xplore, Springer, ScienceDirect and Web of Science.

The methodology combines structured database searching, title-and-abstract screening, cross-database deduplication, metadata consolidation and multi-label thematic coding. It also reflects the practical stage achieved by the project. The evidence used for final study selection and for the analyzes in Chapters 5 to 8 was the title, abstract, keywords and bibliographic metadata. A complete full-text screening was not undertaken for this version of the review. The final corpus should therefore be understood as a rigorously reassessed abstract-level corpus rather than a full-text-validated evidence set.

The review was organized around three connected functions. The first function defined the scope and search concepts. The second function identified and screened potentially eligible records. The third function transformed the retained records into a structured dataset for quantitative description and qualitative synthesis. Figure 4-1, presented on the following page, summarizes this sequence. The workflow was designed to preserve a direct link between every final record and its original database, search set, source file and row location.



Traceability is preserved from every final study to its database, query, file and source row.

Figure 4-1. Three-phase workflow used for the systematic literature review.

4.1 Research Design and Review Scope

A systematic literature review was selected because the topic crosses several technical and normative areas. Relevant studies may appear in telecommunications, artificial intelligence, cybersecurity, distributed computing, law, governance, digital health, transport or smart-city research. A narrow narrative search would risk overrepresenting familiar venues or selected ethical concepts. The systematic approach instead applies predefined questions, databases, search concepts, eligibility rules and analytical categories to all retrieved records.

The review period begins in 2020. This starting point captures the stage at which 5G deployment, early 6G research and responsible AI discussions increasingly began to overlap. It also creates a consistent time window for examining the progression of the field in RQ1. The search cutoff was June 2026. Publications assigned to 2026 therefore represent only the first half of that year and are interpreted as a partial annual period in the results chapters.

The unit of analysis is the individual scientific publication. Journal articles, conference papers, scientific review papers and other accepted scholarly types were considered where their metadata established a peer-reviewed research context. The same publication could be retrieved through several databases or several search sets. Such repeated occurrences were not treated as independent evidence. They were consolidated into one master record before the final analysis.

The review focuses on the combined relationship among three dimensions. The first is an identifiable AI component, including machine learning, deep learning, federated learning, explainable AI, generative models or other AI-enabled systems. The second is a clear connection to 5G, 6G, beyond-5G or a genuinely related next-generation telecommunications environment. The third is an ethical or responsible AI dimension, such as privacy, fairness, explainability, accountability, trustworthiness, governance, regulation, standards, human oversight or social impact. The presence of isolated keywords was not considered sufficient. The relationship among these dimensions had to be central enough to support at least one research question.

4.2 Research Questions and Analytical Logic

The four research questions were defined before the final analysis. They form a sequential analytical structure. RQ1 maps the size, development and thematic composition of the field. RQ2 identifies the ethical risks that occur most frequently. RQ3 examines the methods, frameworks and standards proposed to address those risks. RQ4 synthesizes the remaining gaps by comparing the state of the field, the identified risks and the available responses.

The analytical design is mixed. Quantitative description is used for publication counts, percentages, annual trends, network generations, thematic frequencies and category comparisons. Qualitative synthesis is used to explain the meaning of the categories, the contexts in which they appear and the relationships among risks, safeguards and research gaps. Frequency is treated as evidence of attention in the literature, not as direct evidence of severity, effectiveness or scientific quality.

Table 4-1 links each research question to its principal evidence and analysis. The table also shows why a single extraction sheet would have been insufficient. RQ2, RQ3 and RQ4 can contain several relevant items within the same study. The analytical dataset therefore supports one-to-many relations between a study and its risks, solutions or gaps

RQ	Primary evidence	Main analysis	Output
RQ1	Year, network scope, AI approach, theme, domain and study orientation	Counts, percentages, trends and cross-tabulation	State and progression of the field
RQ2	Normalized ethical risks, context, affected stakeholder and consequence	Multi-label coding, frequency ranking and co-occurrence	Most frequently identified ethical risks
RQ3	Methods, frameworks, standards, targeted risk and maturity	Classification, risk-solution mapping and maturity comparison	Available approaches for trustworthy AI
RQ4	Reported limitations, future work and cross-RQ imbalances	Thematic synthesis and reported/inferred gap separation	Research gaps and future agenda

Table 4-1. Research questions, evidence fields and analysis methods

4.3 Scoping and Search Strategy

The search strategy was developed through an initial scoping stage followed by systematic database searches. The scoping stage was used to identify the terminology applied across AI, mobile networks and responsible technology. It also helped separate broad ethical language from operational concepts commonly used in engineering studies. The final strategy used ten predefined search sets, labelled S1 to S10, across the four selected databases.

The ten search sets combined three concept blocks. The AI block covered artificial intelligence and related learning methods. The network block covered 5G, 6G and next- generation wireless or mobile systems. The ethical block varied across concepts such as ethics, privacy, trustworthiness, governance, accountability, fairness, explainability and human-centered design. The search sets were intentionally overlapping. This increased recall because the same study could describe its contribution through different ethical terms. Overlap was handled later through DOI-based deduplication.

Exact syntax had to be adapted to the search interface of each database. The underlying concepts were kept stable, but field restrictions, filters and export formats differed. For example, ScienceDirect searches were implemented through its broad article-term field to reproduce the scale identified during scoping, whereas the other databases used the closest available title, abstract, keyword or general search functions. The temporal filter was set to 2020 through June 2026 and English-language scholarly publications were retained for screening.

Two confirmed examples illustrate the query logic. Search set S5 used AI AND 5G AND Governance, while S7 used AI AND 5G AND Accountability. The remaining sets followed the same principle by varying the network generation and ethical concept. Table 4-2 records the concept blocks and implementation rules rather than presenting database-specific syntax as if it were identical across platforms.

Concept block	Operational content	Implementation rule
AI and learning	Artificial intelligence, machine learning, deep learning and related AI-enabled learning methods.	Combined with a network block and an ethical or responsible- AI block in each search set.
Network context	5G, 6G, B5G and next-generation wireless or mobile systems directly related to telecommunications.	Network-generation terms were varied across the ten predefined search sets.
Ethical and responsible-AI concepts	Ethics, privacy, trustworthiness, governance, accountability, fairness, explainability and human-centered design.	Ethical terms were varied across S1-S10 to broaden recall while preserving the same conceptual scope.
Database implementation	IEEE Xplore, Springer, ScienceDirect and Web of Science.	Search syntax and available fields were adapted to each database. the time window was 2020-June 2026 and English scholarly publications were retained.
Confirmed query examples	S5: AI AND 5G AND Governance S7: AI AND 5G AND Accountability.	The remaining search sets followed the same concept-block principle by varying network-generation and ethical terms.

Table 4-2. Search concept blocks and implementation rules

4.3.1 Data Sources and Search Output

The systematic searches produced 17,946 source records across 80 exported Excel files. ScienceDirect contributed 10,932 records, Springer 5,640, Web of Science 766 and IEEE Xplore 608. The differences reflect database coverage, query behavior and the number of publication-type exports rather than the importance of each database. Springer and ScienceDirect results were divided into several source categories, including research articles, review articles, conference papers and chapters, while the IEEE and Web of Science exports followed their database-specific record structures.

4.4 Eligibility Criteria

Eligibility criteria were defined before the final master corpus was produced. They were applied to the title, abstract, keywords and bibliographic metadata. A record was marked YES only when all inclusion criteria were satisfied and none of the exclusion criteria applied. A record was marked NO when at least one exclusion criterion was clear. MAYBE was reserved for genuine uncertainty where the available abstract-level information suggested relevance but did not establish the combined AI, network and ethical relationship with sufficient confidence.

Code	Inclusion criterion
I1	Publication dated from 2020 onward.
I2	Publication written in English.
I3	Peer-reviewed journal article, conference paper, scientific review or other accepted scholarly type.
I4	Retrieved from IEEE Xplore, Springer, ScienceDirect or Web of Science.
I5	Substantive reference to AI, ML, DL or an identifiable AI-enabled system.
I6	Substantive connection to 5G, 6G, beyond-5G or a genuinely related communications network.
I7	Substantive ethical or responsible-AI issue, principle, governance mechanism, standard or framework.
I8	Relevance to at least one of the four research questions.
I9	Sufficient title, abstract, keyword or metadata evidence for abstract-level assessment.

Table 4-3. Inclusion criteria applied during screening

Code	Exclusion criterion
E1	Publication dated before 2020.
E2	Publication not written in English.
E3	Duplicate record.
E4	Unaccepted or non-peer-reviewed material.
E5	AI discussed without a substantive 5G, 6G or related telecommunications connection.
E6	5G, 6G or network context discussed without a substantive AI component.
E7	Ethics, governance or responsible-AI terminology mentioned only superficially.
E8	Purely technical network-performance or security study without a meaningful responsible-AI dimension.
E9	Insufficient abstract or bibliographic evidence to assess relevance.

Table 4-4. Exclusion criteria applied during screening

Tables 4-3 and 4-4 present the inclusion and exclusion criteria used in screening. Although the original protocol included a possible tenth exclusion rule concerning research-question relevance, that issue was already covered positively by I8. E10 was therefore not applied as an independent exclusion rule in the operational screening.

4.4.1 Inclusion Criteria

The inclusion criteria ensured temporal, linguistic, publication-type, database and conceptual relevance. I9 was operationalized at abstract level in this version of the review. A record needed sufficient title, abstract, keyword or metadata information to support a reasoned decision. Missing or highly incomplete evidence was not treated as uncertainty when no meaningful assessment could be made, it was excluded under E9.

4.4.2 Exclusion Criteria

The exclusion criteria functioned as explicit reasons for rejection. Several could apply to the same record, but the screening justification reported only the criteria supported by the available information. E7 and E8 were especially important. E7 removed studies where ethical or trustworthy-AI language was superficial. E8 removed studies centered only on performance, latency, routing, handover, resource allocation, attack detection or energy optimization when no broader responsible-AI dimension was evident.

4.5 Data Export, Cleaning and Record Management

Records were exported from the databases in the formats supported by each platform. ScienceDirect results were commonly exported in batches through RIS, imported into Zotero and then converted to CSV so that abstracts and citation metadata could be retained. Other databases supplied spreadsheets or bibliographic exports directly. The exported files were cleaned without deleting irrelevant studies. Screening status was stored in dedicated columns so that rejected records remained auditable.

Every record screened received one of three status values YES, NO or MAYBE. YES, justifications briefly identify AI, network and ethical relationships. NO justifications recorded the applicable exclusion criterion. MAYBE justifications described the unresolved point without assigning an unsupported inclusion or exclusion code. Short and consistent justifications were used to make rechecking later easier.

The screening stage generated 675 YES occurrences and 294 MAYBE occurrences, giving 969 eligible occurrences before cross-database deduplication. These are occurrences rather than unique studies. A publication retrieved through several searches contributes several occurrences at this stage but only one record to the master corpus. Table 4-5 summarizes the database-level outputs and the number of eligible occurrences entering consolidation.

Database	Source records	YES	MAYBE	Eligible occurrences	Unique master records
Web of Science	766	103	34	137	73
Springer	5,640	90	84	174	30
ScienceDirect	10,932	304	170	474	171
IEEE Xplore	608	178	6	184	111
TOTAL	17,946	675	294	969	344

Table 4-5. Database-level screening output before cross-database deduplication

4.6 Title and Abstract Screening

The initial relevance assessment was conducted using titles and abstracts, with keywords and publication metadata used as supporting information. The screening question was not whether a paper contained individual search terms, but whether the available evidence connected AI, an advanced communications environment and an ethical or responsible-AI issue in a way that could contribute to at least one RQ.

The decision rule was conservative. Clear exclusion took precedence over ambiguity. MAYBE was used only where a relevant relationship could plausibly exist but could not be confirmed. Records with no usable abstract were excluded when the remaining metadata did not support evaluation. During the initial source-file screening, duplicate checks were limited to repetitions inside the same file. Cross-database and cross-query duplication was handled separately during master construction.

4.7 Cross-Database Deduplication and Master Corpus Construction

The 969 eligible occurrences were merged into a master dataset. DOI was used as the primary identity key because it is more stable than database-specific record identifiers. DOI strings were normalized by removing resolver prefixes, DOI URLs, spaces, differences in capitalization and trailing punctuation. The audit found 344 distinct DOI values both before and after normalization, showing that the exported DOI strings were already largely consistent.

A DOI appearing more than once formed a duplicate group. There were 208 duplicate groups and 136 DOI values that occurred only once. In total, 625 repeated occurrences were removed, reducing the 969 eligible occurrences to 344 unique master records. No case was found in which the same DOI was linked to genuinely different titles. This supported the validity of DOI-based consolidation.

Deduplication did not simply discard all but the first row. The most complete occurrence was selected as the base record, and missing metadata were filled with other occurrences in the same DOI group. The master retained all source databases, search sets, source categories, file names and original row locations. When duplicate occurrences had conflicting YES and MAYBE decisions, YES was given priority because at least one source record contained sufficient abstract-level evidence for inclusion. Twenty-three mixed-status groups were resolved in this way.

Figure 4-2, presented on the following page, shows the traceability structure created during consolidation. The structure separates the canonical study record from its multiple source occurrences. This is important because database overlaps are analytically useful. It reveals how broadly a study was discoverable without allowing repeated retrieval to inflate publication or theme frequencies.

4.8 Final Abstract Reassessment and Study Selection

After deduplication, the master dataset contained 237 YES records and 107 MAYBE records. All existing YES decisions were retained. The 107 MAYBE records were then reassessed using their titles and abstracts. The purpose of this second pass was to remove uncertainty before the RQ analyzes and to establish one final status for every unique record.

The reassessment retained records where the abstract demonstrated a substantive combined link among AI, 5G or 6G and a relevant ethical concern. It excluded records where the network link, AI link or ethical dimension remained secondary or insufficiently developed. Sixty-five MAYBE records were promoted to YES and 42 were reclassified as NO. The final master therefore contained 302 YES records, 42 NO records and no remaining MAYBE records.

Figure 4-3, presented on the following page, reports the complete record flow. The diagram is PRISMA-like in its documentation of identification, screening, deduplication and inclusion, but it is not presented as a completed full-text PRISMA process. The final 302 studies form the abstract-screened analytical corpus used in Chapters 5 to 8.

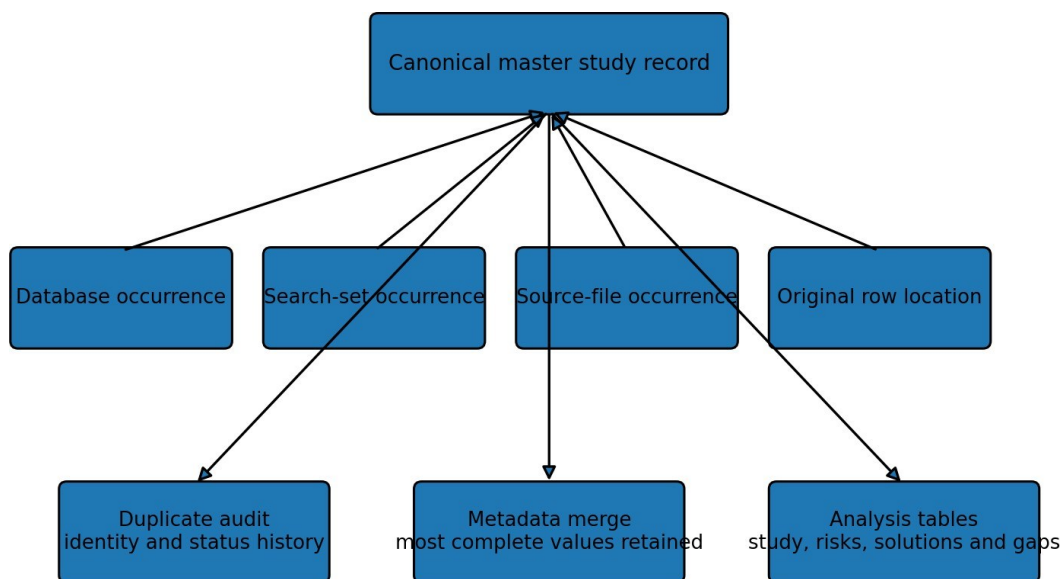


Figure 4-2. Master-record traceability and audit structure

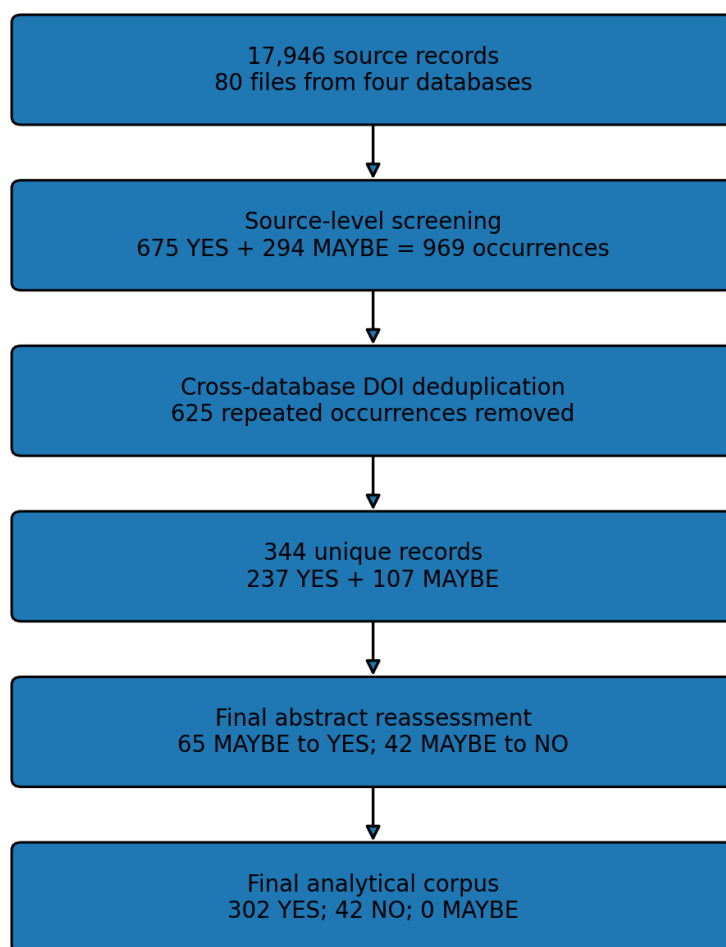


Figure 4-3. Record identification, screening, deduplication and final abstract-level selection

4.9 Data Extraction and Coding Framework

A structured coding framework was created after final study selection. The master metadata supplied the study identifier, title, year, publication venue, database provenance and abstract. Additional analytical variables were derived from the title, abstract and keywords. Coding was multi-label because one publication could address several ethical themes, risks, technical approaches or proposed safeguards.

The second group supported RQ2. Ethical risks were recorded using the wording of the study and then mapped to normalized categories. Similar terms were merged only when they described the same underlying concern. For frequency analysis, a risk category was counted once per study even when the abstract repeated the term. A study could contribute to several different risk categories.

The third group supported RQ3. Proposed responses were classified as methods, frameworks or standards and then assigned to solution families such as federated learning, privacy-enhancing technologies, explainability, blockchain traceability, trust management, governance, runtime assurance or standardization. The coding also considered whether a proposal appeared conceptual, simulation-based, prototype-oriented or connected to a practical implementation. Responsible network-AI research often combines technical and organizational mechanisms, as illustrated by work on trustworthy frameworks, privacy standards and network assurance [58], [145], [214], [269].

The fourth group supported RQ4. Explicit limitations and future research directions were recorded when visible in the abstract. Review-inferred gaps were identified by comparing the distributions found in RQ1, the risks found in RQ2, and the maturity of responses found in RQ3. The inferred gap was not based on low frequency alone. It required supporting evidence such as repeated acknowledgement of a problem, limited empirical evaluation, weak standardization, missing stakeholder coverage or a substantial risk-solution imbalance.

Table 4-6 presented on the following page summarizes the data-extraction fields and coding rules. Figure 4-4, presented on the following page, shows how the same coded corpus supports four distinct analytical outputs while preserving a common study identifier.

Data group	Core fields	Coding rule
Study identity	Study ID, DOI, title, year, venue and provenance	One canonical row per unique DOI
Research context	5G/6G scope, AI approach, network function, application and study orientation	Multi-label where several contexts are explicit
Ethical risks	Original term, normalized category, stakeholder and consequence	Each risk counted once per study
Responses	Method, framework or standard, targeted risk, lifecycle stage and maturity	Exact name retained and broader family assigned
Research gaps	Reported limitation, future direction, gap inferred and supporting evidence	Reported and review-inferred gaps kept separate
Traceability	Database, search set, source file and source row	All source occurrences retained in audit fields

Table 4-6. Principal data-extraction and coding fields

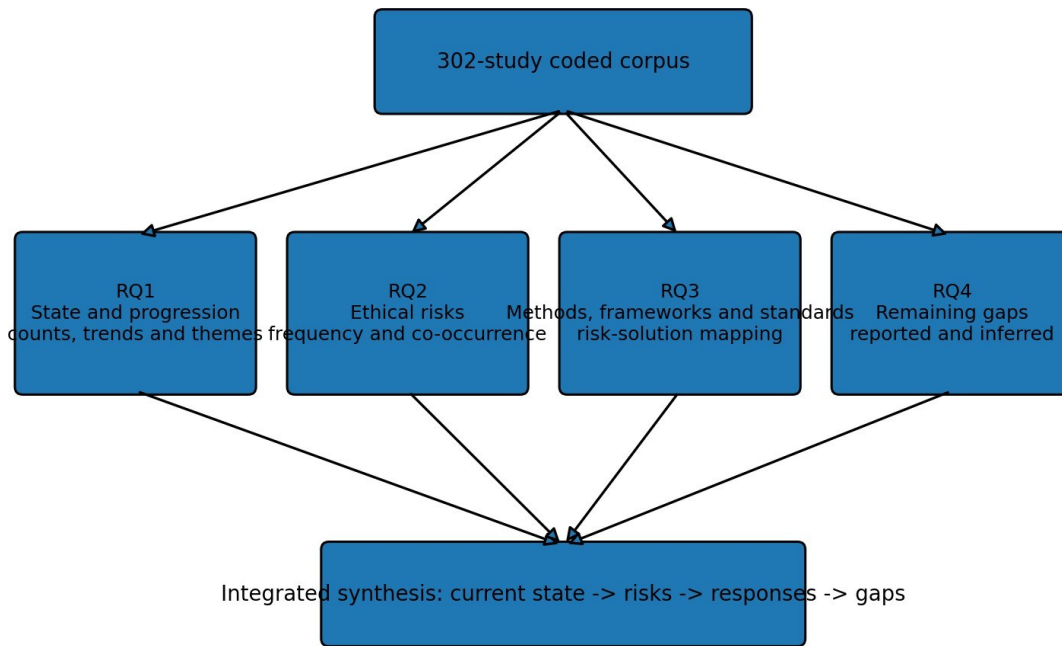


Figure 4-4. Relationship between the common coded corpus and the four RQ analyzes

4.10 Analysis Procedures for the Four Research Questions

RQ1 was analyzed through descriptive statistics and thematic mapping. Publications were counted by year, network generation, AI approach, application domain, ethical themes and research orientation. Cross-tabulation was used to examine changes over time and differences between 5G and 6G research. The partial nature of 2026 was stated explicitly. Growth in publication numbers was interpreted as growth in research activity, not automatically as methodological maturity.

RQ2 combined qualitative coding with frequency analysis. Ethical-risk terms were normalized into categories and ranked according to the number of unique studies in which they appeared. Co-occurrence analysis examined which risks were discussed together. Comparisons by network scope, year, application context and stakeholder group helped distinguish a generally

frequent risk from one concentrated in a specific setting. Frequency was not treated as a measure of harm severity.

RQ3 classified proposed responses by form, solution family, targeted risk, lifecycle position, stakeholder responsibility and level of evaluation. Frequencies showed which approaches received the most attention, while qualitative synthesis explained their intended role. Risk-to-solution matrices were used to examine whether common risks were matched by corresponding methods, governance frameworks or standards. Frequent mention was not interpreted as proof of effectiveness.

RQ4 was analyzed last because it depends on the first three results. Author-reported gaps visible in abstracts were separated from review-inferred gaps. The synthesis compared research concentration, ethical-risk prevalence, solution coverage and evaluation maturity. Gaps were grouped into conceptual, methodological, empirical, governance, standardization, human-centered, dataset, benchmark and deployment categories. The final output was a prioritized research agenda rather than a simple list of future-work statements.

This analytical sequence follows the development of responsible network intelligence from description to problem identification, response and remaining need. It is compatible with the way the wider literature separates current state, trustworthy-AI properties and unresolved challenges [85], [141], [185], [302].

4.11 Reliability, Traceability and Quality Control

Several controls were applied to reduce processing errors. All 80 source files were independently reopened during the audit. Counts of YES and MAYBE occurrences were recalculated from the source data. Exact and normalized DOI counts were compared. Duplicate-group sizes were examined, and titles associated with the same DOI were checked for inconsistency. No false DOI merge or same-file eligible duplication was identified.

Traceability was maintained at record level. Each master row contains the source database, search set, publication category, file name and original source location. A separate duplicate-audit records the identity key, final status, number of occurrences, original statuses and all source locations. A file-summary sheet records processing counts for every export. An independent-recheck sheet documents the main consistency tests, while the abstract-recheck sheet records the final decisions for the 107 MAYBE studies.

A formal study-quality score was not applied. The corpus contains diverse publication types and the evidence available for the present analysis is mainly abstract level. Assigning a detailed methodological quality score without full-text information would create false precision. Quality control therefore focuses on eligibility consistency, metadata completeness, duplicate integrity, traceability and cautious interpretation.

4.12 Methodological Limitations and Ethical Considerations

The main methodological limitation is the absence of full-text screening. Relevant details may be missing from abstracts, while some abstracts may overstate the practical contribution of a paper. This limitation is particularly important for governance mechanisms, empirical validation, stakeholder involvement and research gaps. The results therefore avoid claims that require evidence unavailable in the abstract-level corpus.

Database and query effects are also unavoidable. Search interfaces differ, database coverage changes over time and newly indexed publications can alter result counts. The use of ten overlapping searches and four databases reduce dependence on one terminology or source, but it does not guarantee complete coverage. English-language restriction may exclude relevant work published in other languages.

The review analyzes published information and does not involve human participants, private datasets or personal data collection. Ethical responsibility therefore concerns accurate representation of sources, transparent decision rules and avoidance of unsupported claims. Automated processing was used for record management, normalization and structured counting, but relevance decisions and category interpretations were tied to the available textual evidence. The methodology does not claim full human reading of every publication when only the abstract was available.

These limitations do not remove the value of the review. They define the level at which their findings should be interpreted. The corpus provides a broad, traceable map of how ethical and trustworthy AI is represented in 5G and 6G research up to June 2026. The findings are therefore interpreted within the final abstract-level protocol and studies using full-text screening, dual-reviewer agreement testing or detailed methodological quality appraisal may provide additional evidence beyond this scope.

4.13 Chapter Synthesis

The review identified 17,946 source records from four academic databases and retained 969 YES or MAYBE occurrences after source-level title-and-abstract screening. DOI-based consolidation removed 625 repeated occurrences and produced 344 unique records. Final reassessment of the 107 MAYBE records yielded an abstract-screened analytical corpus of 302 YES studies and 42 excluded studies.

The methodology connects every final record to its original source and applies a common coding framework across the four research questions. Quantitative description measures the development and distribution of literature, while qualitative synthesis explains risks, proposed responses and remaining gaps. The next four chapters apply this framework to RQ1, RQ2, RQ3 and RQ4 respectively.

Chapter 5 Research Question 1: Current Status and Evolution of Research on Ethical Issues in AI-Enabled 5G/6G Networks

This chapter addresses the first research question. What is the present status of research concerning ethical issues for AI-enabled 5G/6G networks, and how has this domain progressed since 2020? The analysis uses the final Master dataset. After deduplication, 344 unique records remained. The second title-and-abstract assessment retained 302 studies as YES and excluded 42 as NO, leaving no MAYBE records. The period extends from January 2020 to June 2026, so the value of 2026 is treated as a partial-year indicator.

RQ1 maps the field before the later chapters examine individual risks, proposed solutions and research gaps. It analyzes publication growth, the movement from 5G toward 6G, the ethical themes receiving attention, the technologies and application settings involved, and abstract-level indicators of research maturity. This distinction prevents publication frequency from being interpreted as evidence of risk severity or solution effectiveness

5.1 Purpose and Analytical Procedure

RQ1 was analyzed through descriptive evidence mapping and structured thematic coding. This combination was selected because the question concerns both the scale of the field and the nature of its development. Publication counts show whether activity is increasing, but they do not explain what researchers are studying. Thematic mapping provides that second layer. Each unique YES record was treated as one study, regardless of how many databases indexed it. The variables examined were publication year, network scope, ethical themes, AI approaches, enabling technologies, network functions, application contexts and documented indicators of research focus.

Network scope was coded into four mutually exclusive categories such as 5G only, 6G only, both 5G and 6G, and broader next-generation networks. The broader category was used for B5G, O-RAN, non-terrestrial, vehicular, edge or future-wireless research when the abstract established direct relevance but did not support a precise generation label. This avoids assigning 6G merely because technology is commonly associated with it. It also recognizes continuity across generations, since trustworthy-AI mechanisms developed in 5G or B5G are often presented as foundations for 6G [43], [85], [104].

The thematic, technological and application analyses used multi-label coding. A study could contribute to several categories, but each category was counted only once per study. For example, a federated-learning architecture could be coded under privacy, trustworthiness, edge intelligence and blockchain when these elements were substantive. Consequently, multi-label percentages do not sum to 100%. Closely related terms were consolidated through a controlled vocabulary, such as interpretability and XAI under explainability, and confidentiality and data minimization under privacy.

Coding was deliberately conservative and based on titles, abstracts and available author keywords. A category was assigned only when the metadata established a substantive connection and not a passing mention. Reviews, frameworks, simulations, prototypes, testbeds and case studies were identified through explicit descriptors. These indicators provide a structured view of progression from problem definition toward evaluation. but they are not a formal study-quality score and cannot replace full-text analysis.

This procedure also supports reproducibility. The coding categories were defined before interpreting the final distributions, and the same decision rule was applied to every included record. Counts were based on studies rather than occurrences of keywords, preventing a long abstract from receiving greater weight than a concise one. The resulting map is therefore suitable for describing the structure of the corpus, while the original titles, abstracts and screening justifications remain available for checking individual assignments and documenting the final coding decisions.

5.2 Composition of the Final RQ1 Dataset

The final corpus contains 302 unique studies. Titles and abstracts were available for all records, while author keywords were available for 224. The studies were identified through IEEE Xplore, ScienceDirect, Web of Science and Springer. ScienceDirect contributed to the identification of 136 included studies, IEEE Xplore to 108, Web of Science to 73 and Springer to 26. The values overlap because 41 studies appeared in more than one database, confirming that no single source covered the full intersection of telecommunications, AI, ethics and application research.

Publication period	Studies	Share of corpus
2020	6	2.0%
2021	9	3.0%
2022	20	6.6%
2023	17	5.6%
2024	38	12.6%
2025	101	33.4%
2026, January-June	111	36.8%
Total	302	100.0%

Table 5-1. Distribution of the final RQ1 evidence set by publication period.

The corpus increased after 65 of the 107 records initially labelled MAYBE were retained during the second title-and-abstract assessment. They were included when the relationship between AI, a 5G/6G or directly relevant future-network setting, and a responsible- technology concern became sufficiently clear. The remaining 42 were excluded because one of these links was absent or peripheral. All counts and percentages in this chapter were recalculated from the 302 final YES records.

The distribution in Table 5-1 is strongly concentrated in the latest period. The 212 studies published in 2025, and January-June 2026 represent 70.2% of the review sample. This pattern provides clear evidence of rapid expansion. In parallel, much of the research is recent and uses terminology that is still developing. Many proposals have also not been replicated or tested over long operating periods. The field is therefore best described as rapidly developing and not yet fully established.

The disciplinary spread of the records is also relevant to RQ1. Telecommunications and networking venues dominate, but the corpus includes computer science, IoT, health technology, transport, law and policy research. This distribution explains why the same concept may be treated differently across studies. Trust may be a measurable property of a protocol, an organizational relationship among network actors or a social judgement made by users. The mapping preserves these differences rather than limiting the corpus to one disciplinary interpretation [252].

5.3 Publication Growth from 2020 to June 2026

The temporal pattern shows a transition from a small conceptual literature to a major research stream. Six studies were identified in 2020 and nine in 2021. The core questions were already present. Explainability was framed as a basis for human-machine trust in 6G [92], trustworthy deep learning was linked to quality-of-trust indicators for mass autonomy [145], and machine learning was presented as both a privacy safeguard and a source of new privacy risks [244]. Early telecommunications research also recognized that AI changes operational and societal relationships, not only network performance [41].

The corpus rose to 20 studies in 2022. Research moved from broad concern toward explicit frameworks and design requirements. Trustworthy-AI frameworks for B5G/6G addressed protected inputs, explainable operation and reliable outputs [214]. Native explainable and robust AI was proposed as a property of 6G architecture [85], while ethically responsible autonomous systems were treated as a network requirement [285]. Application-oriented work connected explainability with slicing [188] and privacy-preserving computation with 5G-based IoT [144], [220].

The 2023 count decreased slightly to 17, but the research became more specific. Explanation-guided reinforcement learning was applied to 6G RAN slicing [78], SIX-Trust addressed secure future-network architecture [269], and 6GENABLERS proposed pervasive trust across functions and stakeholders [151]. Federated explainable models were developed for beyond-5G/6G [43], while Open RAN research highlighted the security and governance complexity created by programmable, multi-vendor control [149], [213].

A second expansion began in 2024, when 38 studies were identified. Responsible-AI mechanisms were applied across the network stack, and they were, explainable and robust resource management [131], fair federated RAN slicing [212], XAI for channel estimation [8], assurance across the advanced 5G edge continuum [45], and accountable and resilient AI-assisted networks [221]. The field was no longer centered only on broad principles but also, it increasingly examined how safeguards affect specific operational decisions [237], [263].

The strongest growth occurred in 2025 and the first half of 2026. The 101 studies published in 2025 represent 33.4% of the corpus, and 111 were already indexed between January and June 2026. The 2025 literature consolidated surveys, standards and combined safeguards, including XAI for 5G-and-beyond security [232], explainable 6G communications and slicing [243], federated cybersecurity [48], privacy standards for digital-twin environments [58] and standardization of AI-native architectures [123], [106], [161].

The 2026 records extend the field toward agentic, continuously acting and institutionally consequential systems. Research examines agentic AI as a new attack surface [83], runtime verification for AI-native O-RAN control [120], neuro-symbolic knowledge planes for verifiable automation [148], legal preparedness across jurisdictions [16] and AI incident reporting in telecommunications policy [17]. Human-centered work considers embodied AI and extended agency [166] as well as ethical and privacy issues in 6G-enabled robotics [127], [236], [276], [287].

The progression can therefore be summarized in four stages and can be presented as foundational requirements, dedicated trustworthy-AI frameworks, integration into network functions and applications and governance of AI-native, agentic and pervasive systems. Figure 5-1 on the following page presents the annual publication trend. It should be interpreted together with this qualitative progression, because increasing output does not necessarily imply equivalent growth in validation quality or governance capacity.

The sharp increase after 2024 should therefore be read as the convergence of several development such as the maturation of 5G automation, the design of AI-native 6G, the wider availability of federated and explainable methods, and growing regulatory attention to high-risk AI. It does not demonstrate that every ethical problem is receiving equivalent attention. The later annual cohorts are larger and technologically broader, but they continue to concentrate on properties that can be measured within simulations or datasets. The temporal trend is consequently evidence of expansion and diversification, not proof that governance and social validation have advanced at the same rate.

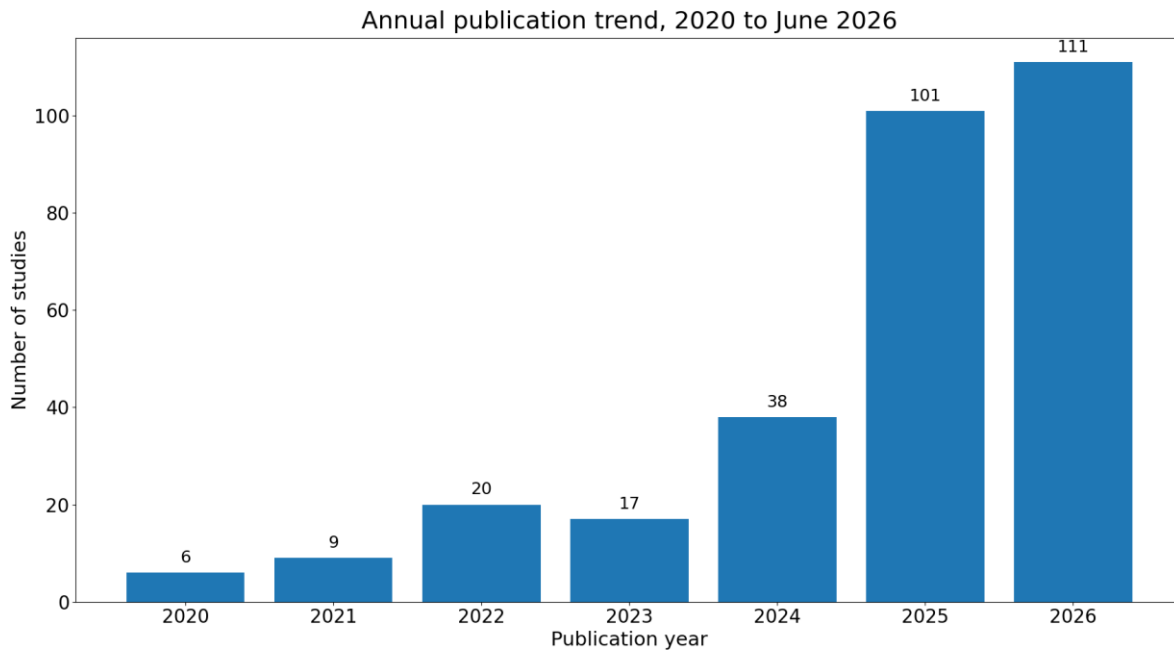


Figure 5-1. Annual publication trend of the final RQ1 corpus from 2020 to June 2026

5.4 Distribution Across 5G and 6G Research

The network-generation distribution confirms that 6G is now the principal focus of the field. A total of 129 studies, or 42.7%, focused exclusively on 6G, and 74 addressed both generations. Together, 203 studies, equal to 67.2% of the corpus, contained an explicit 6G dimension. Sixty-six studies focused only on 5G, while 33 examined broader next-generation settings. The result reflects an expectation that AI-native, highly autonomous 6G systems will intensify existing privacy, security and accountability concerns.

The transition is evolutionary and does not represent a complete replacement of 5G. Zero-trust self-driven networking [104], federated explainable learning [43] and native explainable and robust AI [85] show how mechanisms developed in 5G or B5G are extended toward 6G. Research limited to 5G remains important in slicing, edge computing, industrial IoT and security. From 2023 onward, 6G-only studies became the largest group and reached 56 records during January-June 2026.

The ethical emphasis also broadens in 6G-oriented research. Human-centered communication is directly linked with social, ethical and policy challenges [171]. Embodied AI raises questions about dependency and extended agency [166]. Empirical work on AI and 5G acceptance also shows that deployment is shaped by user perception [10]. These studies move beyond

Network scope	Studies	Share of corpus
6G only	129	42.7%
5G and 6G	74	24.5%
5G only	66	21.9%
Broader next-generation networks	33	10.9%
Total	302	100.0%

Table 5-2. Network-generation scope of the final RQ1 dataset

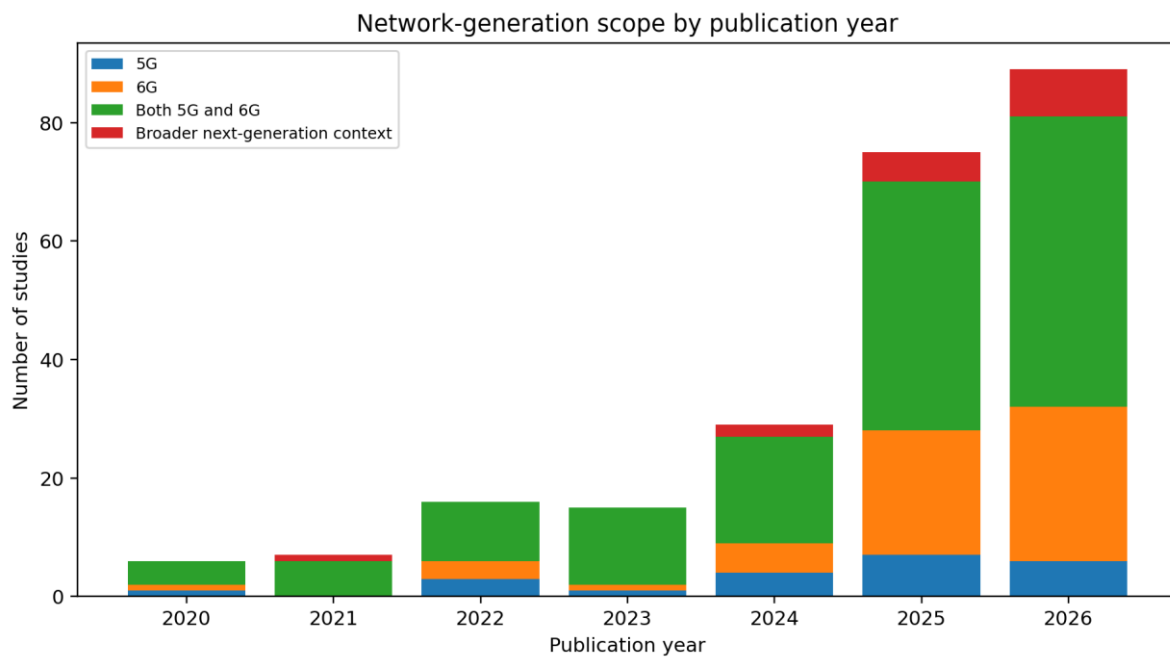


Figure 5-2. Network-generation scope by publication period.

protection of data toward autonomy, participation, legitimacy and the distribution of responsibility.

The broader next-generation category prevents overstatement by retaining relevant O-RAN, edge, non-terrestrial and future-wireless studies that do not justify a precise generation label. Table 5-2 gives the final distribution, while Figure 5-2 on the following page shows how the balance changes across the synthesis period.

The comparison also suggests a difference in research posture. Much 5G-oriented work evaluates responsible AI mechanisms against existing architectures and operational requirements. A larger share of 6G work is forward-looking, asking which safeguards should be designed before widespread deployment. This pre-deployment character allows broader

ethical questions to enter the evidence base, but it also means that some claims are based on expected architecture instead of observed systems. The study combines evidence from current 5G practice with forward-looking 6G design research.

5.5 Ethical and Responsible AI Themes

The thematic mapping shows that technical trustworthiness dominates the field. Security and resilience appeared in 209 studies, safety, reliability and robustness in 166, privacy and data protection in 163, and trust and trustworthiness in 152. These categories overlap because future-network studies commonly combine safeguards. Federated learning may reduce raw-data movement, blockchain may support traceability, anomaly detection may protect operation and XAI may improve inspection. Responsible AI is therefore usually treated as a system of interdependent properties [34], [227], [298].

Security, resilience and safety form the largest cluster because AI has a dual role. It supports adaptive defense and automation but introduces model poisoning, adversarial manipulation and autonomous attack paths. Open RAN adds programmable, multi-vendor interfaces [149], zero-trust approaches require continuous verification [87], [104], and agentic AI creates new vulnerabilities through reasoning, memory and tool use [83]. Runtime safety control [120] and trustworthy resource management [131] show that responsible operation must address both malicious attacks and unintended model failure [297].

Privacy and data protection appear in 54.0% of the corpus. The issue was central from the beginning, when machine learning in 6G was described as simultaneously protective and risky [244]. Privacy-preserving computation in 5G-based IoT [144], federated learning for heterogeneous 6G data [50] and privacy standards for federated digital twins [58] demonstrate the movement from general concern to architectural and standards-oriented responses. The persistence of privacy reflects the dependence of future networks on mobility, radio, health, industrial and behavioral data [119], [140], [251].

Trustworthiness operates as an umbrella concept. It may refer to trust in a model, reputation among devices, confidence in network outcomes or public trust in the system. Quality-of-trust indicators seek measurable assurance [145], global reputation frameworks address distributed actors [40], and trustworthy edge-intelligence surveys combine security, reliability, transparency and sustainability [273]. This breadth is useful, but it also creates inconsistent terminology and makes direct comparison difficult [113], [216], [274].

Governance, ethics and regulation appeared in 129 studies, while accountability and responsibility appeared in 85. Cross-jurisdictional regulatory readiness [16], telecommunications incident reporting [17] and techno-social analysis of future services [258] show that model safeguards must be accompanied by institutional rules. Accountable and resilient network research [221] and blockchain-backed traceability [89] provide technical support, but technical records alone do not determine who is legally or organizationally responsible for a harmful decision.

Explainability and transparency appeared in 115 studies. The field progressed from general arguments about trust [92] to methods for RAN slicing [78], channel estimation [8], 5G performance analysis [6] and network security [232]. Surveys now consolidate XAI across 6G communications and use cases [243], [267]. Explainability is becoming an operational requirement, although explanation quality, intended audience and the relationship between technical interpretability and meaningful accountability remain inconsistent.

Sustainability appeared in 63 studies. Sustainable 6G architecture [22], incentives for green infrastructure [33] and trustworthy edge intelligence [273] connect environmental responsibility with network design. By contrast, inclusion and social impact appeared in 48 studies, fairness in 32 and explicit human-centricity or oversight in 17. Fair scheduling [46], fairness-assessment standards [3] and inclusive 6G-experience frameworks [65] demonstrate that these issues are present but substantially less developed than technical assurance [64], [289].

The central thematic finding is therefore an imbalance in the form of ethics being operationalized. Privacy, security, robustness, trust and explainability receive extensive technical attention. Rights, accessibility, non-discrimination, participation and meaningful human control are less visible. Table 5-3 summarizes the categories. Figure 5-3 on the following page makes the concentration in technical trustworthiness and the limited representation of human and distributional themes more apparent

Theme	Studies	Share of corpus
Security and resilience	209	69.2%
Safety, reliability and robustness	166	55.0%
Privacy and data protection	163	54.0%
Trust and trustworthiness	152	50.3%
Governance, ethics and regulation	129	42.7%
Explainability and transparency	115	38.1%
Accountability and responsibility	85	28.1%
Sustainability and environmental responsibility	63	20.9%
Inclusion, accessibility and social impact	48	15.9%
Fairness, bias and non-discrimination	32	10.6%
Human-centricity and oversight	17	5.6%

Table 5-3. Ethical and responsible AI themes identified in the final corpus.

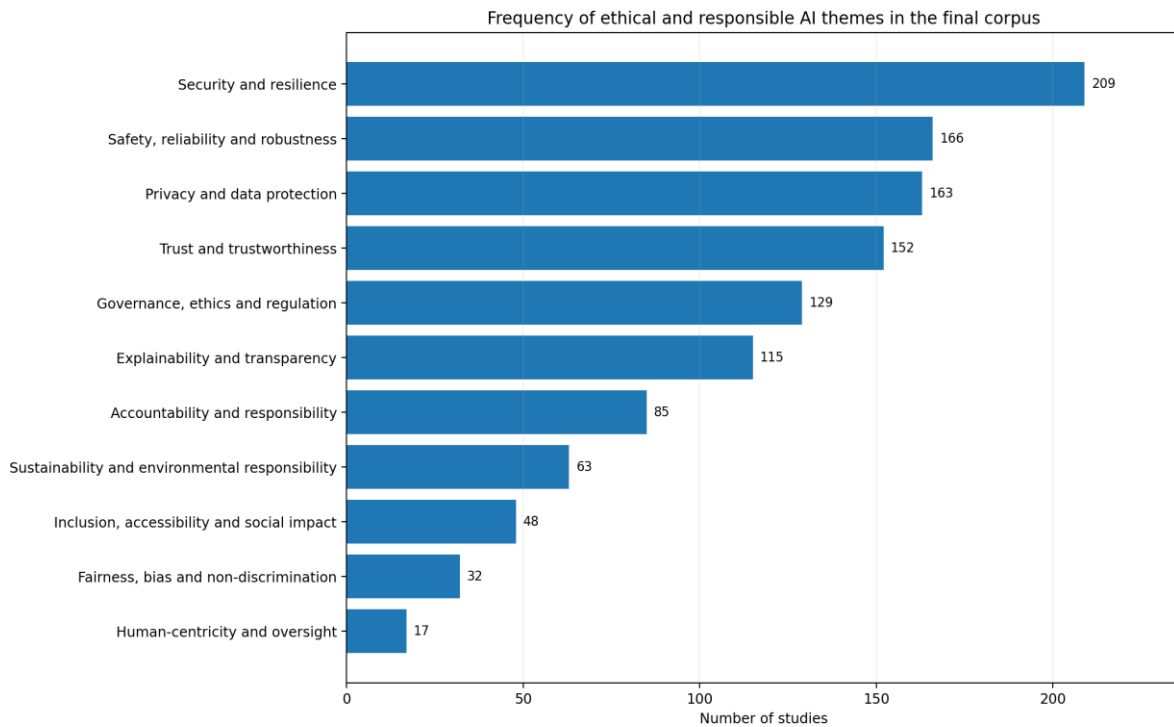


Figure 5-3. Frequency of ethical and responsible AI themes in the final corpus.

The themes should not be interpreted independently. Privacy mechanisms can reduce access to data needed for auditing, stronger security monitoring can increase surveillance, explainability can expose information useful to attackers, and highly robust automation can reduce opportunities for human intervention. Sustainability may also conflict with computationally expensive assurance methods. Literature increasingly recognizes these trade-offs, but most studies still optimize a limited set of properties. A mature responsible-AI approach will require explicit treatment of conflicts among values rather than the assumption that every safeguard can be maximized simultaneously [156], [172], [180].

5.6 AI Approaches and Enabling Technologies

General AI or machine-learning terminology appeared in 276 studies. Federated learning was the most frequent specialized approach, appearing in 98. Its prominence reflects the distributed architecture of 5G/6G and the need to train models across devices, edge nodes and organizations without centralizing all raw data. Surveys connect it with cybersecurity and trustworthiness [48], while privacy-preserving designs address heterogeneous 6G data [50]. However, federated learning can still leak information, distribute bias or admit poisoned

participants, so it is often combined with privacy, reputation and audit mechanisms [248], [280].

Explainable AI appeared in 93 studies and deep learning in 73. Neural models support complex decisions but reduce visibility into why a decision was made. XAI is therefore used for trust, debugging, auditability and operator understanding. The progression is visible from early 6G explainability [92] to resource management [131], slicing [78], channel estimation

[8] and interpretable 5G analysis [6]. Literature increasingly treats explanation as part of network operation rather than an optional post-processing step.

Edge intelligence appeared in 72 studies and blockchain in 59. Edge processing can reduce latency and data transfer but fragments control across devices, operators and clouds. Trustworthy edge-intelligence research therefore combines security, reliability, transparency and sustainability [273]. Distributed ledgers are used for identity, reputation, integrity and traceability, often alongside federated learning [89]. Their ethical value is not automatic, and an immutable record can document a decision without establishing that the decision was fair, necessary or lawful [173], [249], [277].

Digital twins appeared in 41 studies, generative and agentic AI in 36, and reinforcement learning in 35. Digital twins create privacy, fidelity and accountability questions because decisions depend on a virtual representation. Standard-oriented work addresses federated digital twins [58] and sensitive wearable environments [165]. Reinforcement learning supports autonomous control [78], while agentic systems introduce longer chains of action and new security risks [83]. Neuro-symbolic and knowledge-driven approaches remain smaller but are proposed to make automation more verifiable [148], [42], [60], [186].

The broader pattern is one of layered responsible-AI architectures rather than a single dominant solution. Federated or edge learning limits data movement, blockchain or reputation supports coordination, XAI exposes model behavior and symbolic or runtime controls constrain actions. Table 5-4 gives the frequencies. Figure 5-4 is presented on the following page [201].

AI approach or enabling technology	Studies	Share of corpus
Machine learning and AI-based methods	276	91.4%
Federated learning	98	32.5%
Explainable AI	93	30.8%
Deep learning and neural networks	73	24.2%
Edge intelligence and MEC	72	23.8%
Blockchain and distributed ledger	59	19.5%
Digital twins	41	13.6%
Generative and agentic AI	36	11.9%
Reinforcement learning	35	11.6%
Graph-based and knowledge-driven AI	13	4.3%

Table 5-4. Frequently represented AI approaches and enabling technologies.

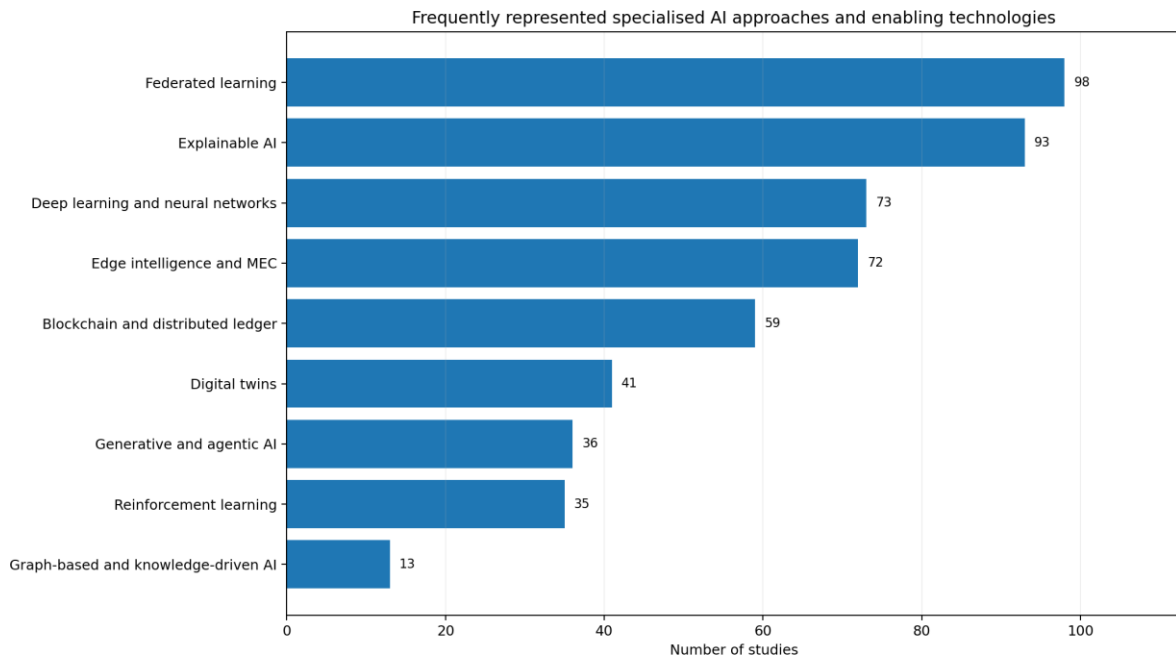


Figure 5-4. Frequently represented specialized AI approaches and enabling technologies.

The technology distribution also shows that responsible AI is becoming a property of composite systems. No single method addresses the full set of ethical concerns. Federated learning may support privacy, XAI may support inspection, blockchain may support traceability and runtime control may support safety, yet the combination introduces additional complexity and energy cost. Evaluation must therefore move from isolated algorithmic components toward end-to-end architecture. This is especially important for 6G, where several learning agents may interact across radio, edge, cloud and application layers [174], [193], [195].

5.7 Network Functions and Application Contexts

IoT and industrial IoT were the largest context, appearing in 104 studies, followed by resource allocation and network management in 96. This shows that ethical AI is connected directly to routing, scheduling, slice assignment, device coordination and traffic management. Trustworthy traffic scheduling [96], explainable resource management [131] and explanation-guided slicing [78] illustrate how reliability and transparency are being integrated into operational decisions that affect service quality and access [175].

Edge-cloud and distributed computing appeared in 71 studies. The edge can improve latency and reduce data exposure, but it divides responsibility among devices, edge operators and cloud services. Assurance across the advanced 5G edge continuum demonstrates that trust must be maintained across several domains [45]. The frequent combination of edge intelligence, federated learning and blockchain reflects attempts to coordinate local processing, distributed verification and shared records [37], [187], [233].

Healthcare, wearables and body networks appeared in 63 studies. These settings involve highly personal data and safety-critical services. Digital-health research Emphasizes the potential of data and AI [15], while 6G healthcare surveys identify privacy, security and implementation challenges [18]. Ethical analysis of 6G robotics [127] and standards-driven governance for wearable digital twins [165] show that network reliability, consent, model governance and institutional accountability must be considered together [73], [247], [290].

Cybersecurity and anomaly detection appeared in 61 studies, and vehicular, UAV and autonomous mobility in 54. In both areas, false decisions can create immediate safety effects. Connected-vehicle fatigue recognition has been examined in relation to 6G and the EU AI Act [7]. Smart cities and public infrastructure, represented in 39 studies, add collective concerns because sensing and automated decisions may affect people who did not actively choose to participate [24], [99].

Lower network layers are also represented. Spectrum and physical-layer optimization appeared in 36 studies, RAN and O-RAN in 28, slicing in 21, non-terrestrial networks in 13 and semantic communication in eight. XAI for channel estimation [8], O-RAN security [149], runtime safety in AI-native control [120] and explainable semantic communication

[301] show that ethical issues extend from physical reliability to meaning, governance and cross-vendor accountability. Table 5-5 on the following page summarizes the context, and Figure 5-5 appears on the following page [208].

The ethical significance of an application depends on both the network decision and the people affected. A minor optimization error in a non-critical service differs from an opaque decision affecting emergency communication, clinical monitoring or autonomous mobility. The present mapping does not rank these consequences, but it shows where such questions are likely to arise. The application distribution provides a contextual basis for RQ2. The next analysis separates how often a risk appears from the severity and type of harm described in the underlying studies .

Network function or application context	Studies	Share of corpus
Internet of Things and industrial IoT	104	34.4%
Resource allocation and network management	96	31.8%
Edge-cloud and distributed computing	71	23.5%
Healthcare, wearables and body networks	63	20.9%
Cybersecurity, intrusion and anomaly detection	61	20.2%
Vehicular, UAV and autonomous mobility	54	17.9%
Smart cities, homes and public infrastructure	39	12.9%
Spectrum, channel and physical-layer optimization	36	11.9%
Radio access networks and O-RAN	28	9.3%
Network slicing and service orchestration	21	7.0%
Non-terrestrial and satellite networks	13	4.3%
Semantic and goal-oriented communications	8	2.6%

Table 5-5. Network functions and application contexts represented in the final corpus.

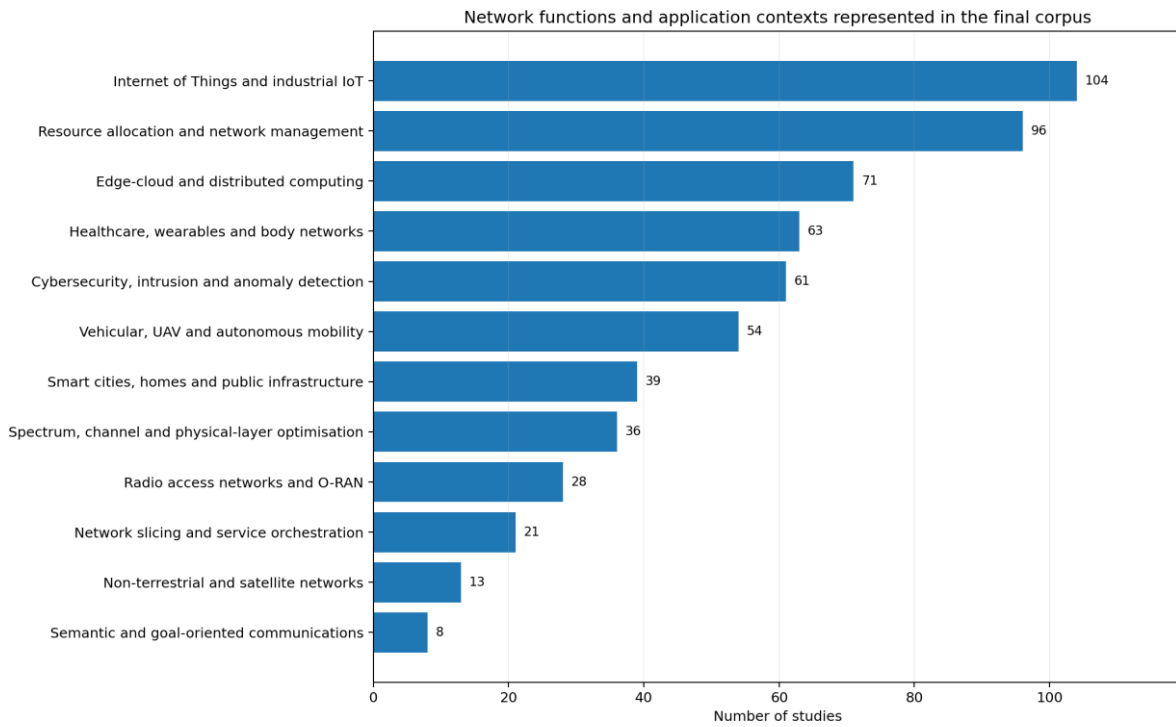


Figure 5-5. Network functions and application contexts represented in the final study set.

5.8 Research Focus and Maturity

The abstract-level indicators describe substantial conceptual and experimental activity but limited large-scale operational validation. Eighty studies, or 26.5%, were reviews or surveys. Surveys on AI for 6G [55], federated cybersecurity [48], XAI for 5G-and-beyond security [232] and explainable 6G use cases [267] consolidate a rapidly changing field. Their prevalence is useful but also indicates that terminology, boundaries and research agendas are still being defined.

Sixty-three titles explicitly proposed a framework, architecture, model, scheme or mechanism. This demonstrates a strong design orientation, ranging from trustworthy-AI frameworks [214] and pervasive-trust architectures [151] to privacy standards [58] and AI- native standardization [123]. Among non-review studies, 122 contained explicit simulation or quantitative evaluation. Examples include explanation-guided slicing [78], XAI channel estimation [8] and trustworthy traffic scheduling [96]. Responsible-AI proposals are therefore increasingly tested under network constraints [74].

Only 13 studies explicitly referred to a prototype, proof-of-concept or testbed. A privacy-preserving B5G pandemic network was presented as a proof-of-concept [177], trustworthy analytics were examined through a 5G security case study [189], and runtime safety

mechanisms represent a more operational direction [120]. Forty-two studies contained broader real-world, field or case evidence, but this should not be read as long-term commercial deployment [197].

Only 23 titles had a primary policy, legal, ethical or standardization focus. Governance is discussed in many abstracts, but institutional analysis remains less common than technical evaluation. Regulatory readiness [16], incident reporting [17], green-network incentives [33] and techno-social analysis [258] are important exceptions. Technical maturity is therefore advancing faster than organizational and societal maturity.

The field can be described as partially mature. Many studies evaluate algorithms and architectures, but common benchmarks, longitudinal validation, cross-operator testing and human-centered assessment remain limited. Accuracy, latency and robustness are easier to measure than fairness, accountability or user understanding. Table 5-6 summarizes these non-exclusive indicators.

Non-exclusive orientation indicator	Studies	Share of corpus
Review or survey	80	26.5%
Framework, architecture, model, scheme or mechanism in title	63	20.9%
Non-review study with simulation or quantitative evaluation	122	40.4%
Prototype, proof-of-concept or testbed	13	4.3%
Real-world, field or case-study evidence	42	13.9%
Policy, legal, ethical or standardization focus in title	23	7.6%

Table 5-6. Abstract-level indicators of research orientation and maturity.

Maturity should also be assessed across several dimensions. Technical maturity concerns whether a method performs reliably under realistic network constraints. Operational maturity concerns whether it can be integrated, monitored and maintained across vendors and domains. Governance maturity concerns audit rights, responsibility, incident response and compliance. Social maturity concerns whether affected users can understand that contest or influence decisions and whether benefits are distributed fairly. The review sample shows strongest progress in the first dimension, emerging progress in the second and third, and limited evidence in the fourth.

5.9 Integrated Interpretation and Direct Answer to RQ1

The present status of research on ethical issues in AI-enabled 5G/6G networks is that of a rapidly expanding, multidisciplinary and increasingly 6G-oriented field. The final corpus contains 302 studies from 2020 to June 2026. Output increased from six studies in 2020 to 101 in 2025, while 111 were already indexed during the first half of 2026. The growth is accompanied by diversification from privacy, trust and explainability toward federated learning, O-RAN, slicing, edge intelligence, healthcare, mobility, digital twins, semantic communication, agentic AI and regulatory readiness [47], [192], [284].

The development can be understood through four stages. The earliest work established the foundational problem and learned network control reduces direct visibility and introduces privacy and trust requirements [92], [145], [244]. The next stage proposed dedicated responsible and trustworthy-AI frameworks [85], [214], [285]. A third stage integrated safeguards into specific functions, including slicing, channel estimation and resource management [8], [78], [131], [188]. The latest stage concerns AI-native and agentic systems whose actions require runtime control, verifiable reasoning and legal governance [16], [83], [120], [148], [198].

The movement toward 6G is pronounced but does not erase 5G. Two thirds of the corpus contain an explicit 6G dimension, yet many 6G mechanisms are designed or evaluated in 5G and B5G settings. Existing concerns are therefore accumulated rather than replaced. Privacy, security and explainability remain, while denser sensing, distributed intelligence and autonomous control add questions of cognitive sovereignty, human dependency and responsibility across organizations and jurisdictions.

The thematic distribution reveals both strength and imbalance. Security, safety, privacy, trust and explainability are extensively represented and supported by substantial technical work. Governance and accountability are growing [16], [17], [58], [221]. Fairness, accessibility, explicit human oversight and broader social impacts remain less frequent. The field is strongest in technical trustworthiness and less developed in empirical analysis of how network decisions affect rights, participation, unequal outcomes and human agency [129], [215].

Responsible AI is also becoming architectural. Federated learning, XAI, edge intelligence, blockchain, digital twins and runtime controls are frequently combined to manage a distributed trust problem involving data, models, infrastructure and several organizations. The largest application groups, including IoT, resource management, healthcare, cybersecurity and mobility, show that ethical issues arise where network decisions have direct operational or human consequences. Even lower-layer functions affect the reliability and availability of higher-level services [117], [139], [203].

RQ1 can therefore be answered directly as follows. Since 2020, the domain has progressed from a small foundational literature into a large and technically productive field centered increasingly on 6G and AI-native operation. Research has moved from identifying concerns to designing and evaluating safeguards for specific functions and applications. However, it is not fully mature. Simulation and quantitative assessment are common, but prototypes, long-term deployments, comparable governance criteria and human-centered validation remain limited. The next stage requires technical assurance to be matched by institutional accountability and evidence about societal outcomes.

A further implication is that ethical and technical research are becoming less separable. In AI-native networks, privacy, explanation, safety and accountability influence architecture, interfaces, data flows and control loops. They cannot be added only after a system has been optimized for performance. The evidence therefore supports an ethics-by-design interpretation of future-network development. However, the current literature applies this principle unevenly because some studies integrate safeguards into system design, while others use responsible-AI terminology primarily to describe a performance-enhancing security or privacy mechanism [206].

The progression also changes the location of ethical responsibility. In early studies, responsibility is often discussed at the level of the individual model or application. In later AI-

native architectures, responsibility is distributed across data providers, model developers, equipment vendors, operators, orchestration platforms and application owners. This distribution makes technical assurance insufficient. A reliable model may still operate within an unclear governance structure. The current state of the field therefore requires mechanisms that connect model evidence with organizational roles, audit access, escalation procedures and remedies for affected stakeholders.

A second integrated finding concerns the relationship between anticipation and evidence. Much 6G research is necessarily forward-looking because the complete ecosystem is not yet deployed. Anticipatory analysis is valuable because it allows safeguards to influence standards and architecture before harmful practices become embedded. At the same time, anticipated use cases can overstate technological certainty or social benefit. The strongest future research will combine prospective ethics with evidence from 5G, B5G, testbeds and sector-specific deployments, using current systems to test assumptions about future autonomy and human impact.

5.10 Limitations of the RQ1 Analysis

The first limitation concerns the evidence level. Coding was based on titles, abstracts and available author keywords for the 302 YES studies. These sources support consistent mapping of a large corpus, but they cannot show every method or the depth of ethical analysis in the full text. Study may use trust, fairness or human centrality only as motivation, while a relevant discussion may be absent from the abstract. The frequencies should therefore be interpreted as research-orientation indicators rather than a final measurement of conceptual depth or study quality.

The second limitation is temporal. The 2026 value covers January to June only and is not directly equivalent to a complete annual count. Early-access dates and indexing schedules may also differ across databases. The publication year recorded in the final Master was retained for consistency, but the value was not annualized or used to forecast the end of 2026. Future studies can repeat the same search and screening procedure for the remaining months before comparing complete years.

The third limitation concerns coverage. The four databases provide broad and overlapping access, but no search source is complete. Relevant standards, regulatory material, institutional reports, preprints or studies indexed elsewhere may be absent. Search terminology also shapes retrieval. Some publications may examine responsible network intelligence without using explicit ethics vocabulary, while others may use terms such as trustworthy or human-centric only as broad framing. The observed corpus is therefore a transparent but bounded representation of the field.

The fourth limitation is conceptual overlap. Security, safety, resilience, reliability and trust are not defined uniformly, and governance or accountability may refer to different levels. Multi-label coding preserves this overlap but does not eliminate ambiguity. The controlled vocabulary was applied consistently, and the analysis retains the original terms and avoids merging specific risks, solutions or gaps when the abstract-level evidence is insufficient. The broader next-generation category similarly reduces misclassification but includes heterogeneous architectures.

The fifth limitation is interpretive. Frequency measures research attention, not ethical severity or effectiveness. A common risk is not necessarily the most harmful, and a frequent method is not necessarily the most successful. Technical performance may be validated while accountability, fairness or user impact remains untested. RQ2, RQ3 and RQ4 are needed to distinguish attention from consequence, proposal from effectiveness and limited coverage from a defensible research gap.

Finally, the field is changing quickly. Agentic AI, large models, digital twins and AI-native RAN control may alter the thematic balance within a short period. This chapter provides a reproducible snapshot up to June 2026, not a permanent description. Its contribution is to establish a final corpus, explicit coding rules and quantified trends. Future studies may extend the time horizon and test whether emerging concepts become durable research directions, while the present chapter provides the final synthesis for the review period ending in June 2026.

The analysis also does not weight studies by methodological quality, citation impact, venue or sample size. Each included study contributes equally to the descriptive counts because RQ1 maps the field rather than estimating an effect. This choice prevents highly cited surveys from dominating the results, but it means that an early-stage proposal and a thoroughly validated study can contribute to the same category. Quality appraisal and evidence strength should

therefore inform the interpretation of specific claims in the later chapters, even though they are not used to adjust the RQ1 frequencies.

A further limitation is that the analysis does not distinguish whether an ethical theme is the primary research contribution or a secondary supporting concern in every record. Multi-label coding captures breadth, but a study focused mainly on intrusion detection may contribute to privacy or explainability only as an additional design feature. The thematic counts therefore indicate presence, not centrality. The final interpretation distinguishes presence from centrality and avoids treating secondary design features as dominant research objectives

The chapter also does not compare geographical, institutional or sectoral representation. Legal readiness, infrastructure ownership and acceptable levels of monitoring differ across jurisdictions and application domains. A study set dominated by particular research regions or technical sectors may not represent the priorities of lower-resource environments, smaller operators or communities with limited access to advanced connectivity. Future analysis should examine where studies originate, which stakeholders are represented and whether proposed governance approaches can transfer across different regulatory and socioeconomic settings.

Chapter 6 Research Question 2: Ethical Risks in AI-Enabled 5G/6G Networks

This chapter addresses the second research question. What ethical risks are most frequently identified in the literature on AI-enabled 5G/6G networks? The analysis uses the same final corpus of 302 studies examined in Chapter 5 and retains the temporal boundary from January 2020 to June 2026. Whereas Chapter 5 mapped broad research themes and the development of the field, Chapter 6 converts those themes into explicit risk statements. The purpose is not merely to identify topics such as privacy or trust, but to clarify the form of harm, vulnerability or institutional deficit associated with each topic.

Literature is dominated by risks that can be translated into technical properties such as cybersecurity compromise, privacy loss, unreliable behavior, loss of trust and limited explainability. Governance and regulatory inadequacy also receive substantial attention, while fairness, exclusion, surveillance and loss of human control remain much less frequent. The field therefore tends to operate ethics through security, model performance and data protection, while rights, power, participation and unequal social outcomes are treated less consistently [215].

6.1 Purpose and Analytical Procedure

RQ2 requires both qualitative interpretation and quantitative counting. A purely qualitative discussion could describe many relevant risks but would not answer which risks are identified most frequently. A simple keyword count would also be insufficient because different articles use different terms for related concerns and because repeated use of a word within one abstract should not increase the number of studies supporting a risk. The adopted procedure therefore combined structured thematic coding with study-level frequency analysis. Each risk was counted at most once per study, even when the title, abstract or keywords repeated the same concept several times.

The analytical unit was included in the unique study. Coding used the title, abstract and available author keywords in the final Master Excel. The screening justification was not treated as independent evidence because it had been written during the selection process rather than by the authors of the studies. A risk category was assigned when the available metadata explicitly identified a threat, harmful outcome, vulnerability, ethical deficit or need for protection. For example, a study was coded for privacy when it discussed privacy,

confidentiality, personal-data protection, data leakage, consent, anonymity or a privacy-preserving mechanism. It was coded for opacity when it referred to explainability, interpretability, transparency or black-box behavior. This explicit-risk layer is narrower than the thematic mapping used in Chapter 5 and therefore produces slightly different counts for some categories.

The coding was multi-label because ethical risks rarely occur independently. A federated-learning study may simultaneously address privacy, poisoning, reliability and trust. A network-slicing study may link explainability with safety and accountability. Multi-label coding preserves these relationships rather than forcing each article into one exclusive class. At the same time, a primary-risk orientation was assigned to identify the issue receiving the strongest emphasis in the title and abstract. The primary classification is used only to show the center of attention of each study, and it does not remove secondary risks from the frequency analysis [90], [225], [241].

Three additional comparisons were made. First, the development of the six most frequent risks was examined by publication year. Second, risk prevalence was compared between 5G-related and 6G-related studies. These two groups are non-exclusive because studies covering both generations contribute to both. Third, co-occurrence was calculated for every pair of risk categories to identify recurring combinations. Percentages in the main ranking use the 302 included studies as the denominator. Because the categories are multi-label, the percentages do not come to 100%.

6.2 Operational Definition and Risk Classification

An ethical risk was defined as a plausible way in which AI-enabled 5G/6G systems may cause harm, distribute benefits or burdens unfairly, reduce individual or collective control, weaken institutional responsibility, or fail to satisfy justified expectations of safety, privacy and trust. This definition treats the network as a socio-technical system. Harm may originate from an algorithm, training data, a malicious actor, an organizational decision, an absence of standards or the interaction between these elements. The definition is therefore broader than cybersecurity, although cybersecurity remains its most frequent component.

Thirteen categories were retained. The taxonomy separates risks that are often merged in broad trustworthy-AI discussions. Privacy is separated from surveillance because data protection can fail without continuous monitoring, while surveillance can be ethically problematic even

when data are securely stored. Trust is separated from reliability because a system can be technically reliable but still lack legitimacy or public confidence. Accountability is separated from general governance because rules may exist without a clear allocation of responsibility when an automated decision causes harm. Fairness is separated from social exclusion because biased model outputs and unequal access to infrastructure are related but not identical problems.

The categories were applied conservatively. Generic references to a technical environment were not coded as environmental risk. A claim of high accuracy was not coded as reliability unless the abstract explicitly discussed robustness, safety, faults or uncertainty. Similarly, the word autonomous did not automatically indicate loss of human control. Human-control coding requires an explicit reference to human-centricity, human judgment, agency, oversight or meaningful control. These decisions reduce false positives and make the frequency ranking easier to interpret.

A final coding principle concerned the relationship between risks and protective mechanisms. A privacy-preserving or security method was treated as evidence that the corresponding risk was recognized, even when the abstract focused mainly on the solution. This is necessary because technical research frequently expresses risk indirectly through the objective of the proposed mechanism. Differential privacy, zero-trust networking, explainable intrusion detection and fairness auditing all imply an underlying privacy, security, opacity or discrimination risk [1], [3], [71], [104]. However, the existence of a solution was not interpreted as proof that the risk had been eliminated.

Risk category	Operational coding meaning
Cybersecurity compromise and adversarial manipulation	Attacks, malicious model behavior, intrusion, evasion, spoofing, jamming, Byzantine behavior or other vulnerabilities affecting AI-enabled network operation.
Privacy loss and improper data use	Unauthorized disclosure, inference or secondary use of personal or sensitive data, weak consent, confidentiality or anonymity.
Unsafe, unreliable or non-robust AI behavior	Failure under faults, uncertainty, changing conditions or critical operation, insufficient robustness, reliability or safety.
Loss of trust and confidence	Insufficient grounds for stakeholders to rely on models, data, devices, services or institutions, weak reputation or assurance.
Governance, legal and regulatory inadequacy	Missing, fragmented or unclear rules, standards, compliance procedures, policy responsibilities or legal readiness.
Opacity and limited explainability	Black-box decisions, limited interpretability or transparency, and insufficient information for scrutiny or contestation.
Environmental and resource burden	Energy, carbon, infrastructure or lifecycle impacts associated with AI computation and future-network deployment.
Accountability and responsibility gaps	Unclear answerability, liability or responsibility for decisions, failures, incidents and remedies.
Bias and discriminatory outcomes	Systematic unfairness or unequal model outcomes affecting users, traffic, services or groups.
Data integrity, quality and poisoning	Corrupted, manipulated, heterogeneous, low-quality or poorly governed data and model updates.
Social exclusion, accessibility and unequal benefit	Digital divide, unequal access, affordability barriers or uneven distribution of network benefits and burdens.
Loss of human control and insufficient oversight	Weak human judgment, agency, intervention, meaningful control or oversight over automated action.
Surveillance and intrusive monitoring	Pervasive tracking, profiling or monitoring that restrict privacy, autonomy or acceptable participation.

Table 6-1. Operational classification scheme used for the RQ2 risk coding.

6.3 Overall Frequency and Ranking of Ethical Risks

The ranking identifies cybersecurity compromise and adversarial manipulation as the most frequent risk, appearing in 192 studies, or 63.6% of the corpus. Privacy loss and improper data use follow with 163 studies, equal to 54.0%. Unsafe, unreliable or non-robust behavior appears in 159 studies, while loss of trust and confidence appears in 153. Governance, legal and regulatory inadequacy is identified in 134 studies, and opacity or limited explainability in 116. These six categories form the dominant risk structure of the field [133].

The remaining categories are considerably less frequent. Environmental and resource burden appears in 62 studies. Accountability and responsibility gaps appear explicitly in 34, bias and discriminatory outcomes in 30, and data integrity, quality and poisoning in 28. Social exclusion and unequal benefit also appear in 28 studies. Loss of human control is identified in 15, while surveillance and intrusive monitoring appear in 12. The low value does not mean that these risks are unimportant. They indicate that they are less often made explicit in titles, abstracts and keywords and are therefore less central to the visible research agenda.

Ethical risk category	Studies	Share of corpus	Primary risk
Cybersecurity compromise and adversarial manipulation	192	63.6%	85
Privacy loss and improper data use	163	54.0%	49
Unsafe, unreliable or non-robust AI behavior	159	52.6%	20
Loss of trust and confidence	153	50.7%	36
Governance, legal and regulatory inadequacy	134	44.4%	37
Opacity and limited explainability	116	38.4%	48
Environmental and resource burden	62	20.5%	14
Accountability and responsibility gaps	34	11.3%	1
Bias and discriminatory outcomes	30	9.9%	4
Data integrity, quality and poisoning	28	9.3%	0
Social exclusion, accessibility and unequal benefit	28	9.3%	2
Loss of human control and insufficient oversight	15	5.0%	4
Surveillance and intrusive monitoring	12	4.0%	2

Table 6-2. Frequency and primary orientation of ethical risks in the final corpus.

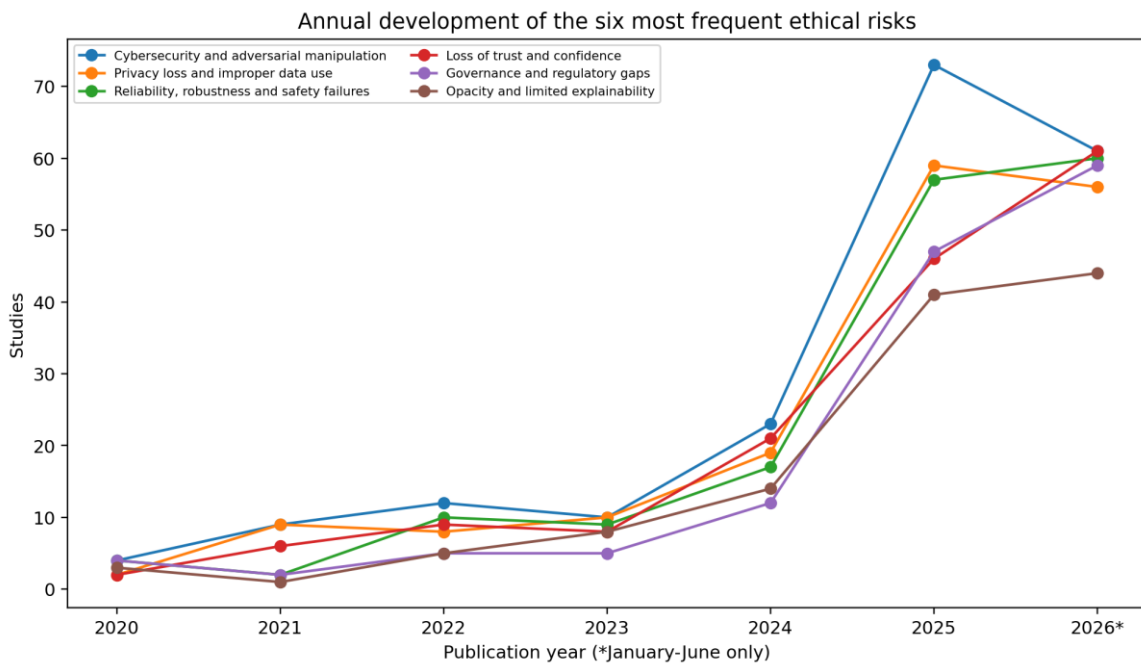


Figure 6-1. Annual development of the six most frequent ethical risks from 2020 to June 2026.

The difference between total and primary frequency provides another perspective. Cybersecurity is the primary risk in 85 studies, privacy in 49, opacity in 48, governance in 37 and trust in 36. Reliability is primary in only 20 studies despite appearing as a secondary concern in 159. This suggests that reliability is frequently treated as a supporting requirement of secure, trustworthy or explainable systems rather than as the sole organizing problem. Environmental risk is primary in 14 studies. Fairness, human control, exclusion and surveillance together account for only a small share of primary orientations [261].

The annual development of the six dominant categories is shown in Figure 6-1 on the following page. The figure is used from January to June only for 2026, so the final point must not be interpreted as a complete annual total.

All six leading risks grow sharply after 2024, but the pattern is not identical. Security rises most strongly in 2025 and remains the largest category in the first half of 2026. Governance and trust show especially rapid growth in 2026, reflecting the expansion of AI-native, agentic and standards-oriented research. Explainability also grows steadily from early conceptual work on human-machine trust [92] to application-specific studies in channel estimation, slicing, semantic communication and intrusion detection [8], [78], [184], [302]. The temporal pattern

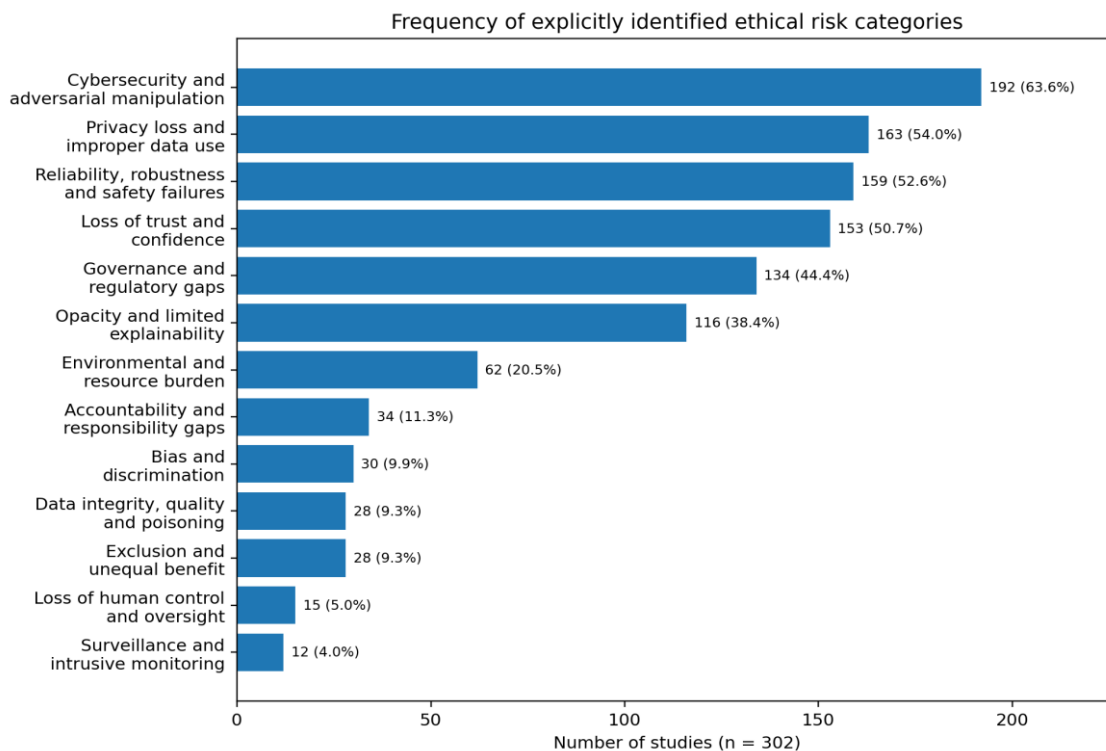


Figure 6-2. Frequency ranking of identified ethical risks in the final corpus.

indicates that the field is not moving away from technical risks but instead, it is adding governance and verification concerns around progressively autonomous systems.

Figure 6-2 presents the complete frequency ranking and is shown on the following page. It makes the steep decline after the six dominant categories visually clear.

The concentration of attention in the leading categories produces a two-tier risk landscape. The first tier consists of security, privacy, reliability, trust, governance and explainability. These risks are discussed across many network functions and application domains. The second tier consists of environmental burden, accountability, fairness, data quality, exclusion, human control and surveillance. These concerns are often present in specialized studies or policy-oriented discussions but do not yet have comparable coverage across the corpus. The imbalance becomes a central finding of RQ2 because it identifies not only what literature recognizes, but also what it tends to leave at the margins [270].

6.4 Cybersecurity, Data Integrity and Adversarial Manipulation

Cybersecurity compromise is the most frequently identified ethical risk because AI is simultaneously a defensive tool and a new attack surface. AI supports intrusion detection, adaptive access control, anomaly recognition and threat intelligence, but models can be poisoned, evaded, manipulated or used to automate attacks. The ethical significance extends beyond technical loss. Successful attacks can interrupt essential services, expose personal data, distort automated resource decisions and transfer risk to users who cannot evaluate or control the underlying systems. Explainable intrusion-detection frameworks [1], responsible AI-driven attack detection for 6G-IoT [11] and interpretable anomaly detection for O-RAN [44], [184] all respond to this dual role [223].

The attack surface becomes broader as control functions move toward open, programmable and distributed architectures. O-RAN introduces intelligent controllers and third-party applications, edge intelligence distributes models across devices, and network slicing creates multiple logical services over shared infrastructure. Open RAN security research identifies new interfaces and trust boundaries [149], while zero-trust self-driven networking treats every component and interaction as potentially untrusted [104]. In these settings, an incorrect or malicious AI decision can propagate through orchestration, scheduling or policy enforcement rather than remaining isolated within one application.

Agentic and generative AI intensify this problem because the model may plan actions, use tools and interact with other agents. The 2026 literature describes agentic AI as a distinct source of attack vectors and vulnerabilities [83]. Runtime safety mechanisms for AI-native O-RAN attempt to verify intended actions and impose fail-safe behavior before a harmful control command reaches the network [120]. These studies show a shift from protecting static prediction models to supervising systems that act continuously and adapt their strategies. The corresponding ethical risk is a loss of dependable control over critical infrastructure.

Data integrity and poisoning form a related but more specific risk. Federated learning reduces centralized data collection, yet it allows compromised or low-quality participants to influence the global model. Byzantine-robust orchestration [67], blockchain-assisted federated learning [136] and trustworthiness analysis under heterogeneous data [143] address different parts of this problem. Data heterogeneity can also create uneven performance without deliberate attack, while poisoning introduces intentional manipulation. Both conditions can produce unreliable

outcomes and unfairly affect groups or locations whose data differ from the dominant training distribution [13], [112], [164].

The security category co-occurs with privacy in 134 studies, reliability in 99 and trust in 98. These combinations show that security is rarely a separate ethical objective. Strong monitoring may protect the network but increase surveillance. Encryption and privacy mechanisms may complicate inspection and accountability. Aggressive anomaly detection may improve resilience while reducing availability for atypical users. The most frequent co- occurrences are listed in Table 6-3 [209], [233].

Risk pair	Studies	Share of corpus
Cybersecurity compromise and adversarial manipulation + Privacy loss and improper data use	134	44.4%
Cybersecurity compromise and adversarial manipulation + Unsafe, unreliable or non-robust AI behavior	99	32.8%
Cybersecurity compromise and adversarial manipulation + Loss of trust and confidence	98	32.5%
Loss of trust and confidence + Unsafe, unreliable or non-robust AI behavior	87	28.8%
Loss of trust and confidence + Opacity and limited explainability	86	28.5%
Cybersecurity compromise and adversarial manipulation + Governance, legal and regulatory inadequacy	82	27.2%
Privacy loss and improper data use + Unsafe, unreliable or non-robust AI behavior	79	26.2%
Opacity and limited explainability + Unsafe, unreliable or non-robust AI behavior	73	24.2%
Loss of trust and confidence + Privacy loss and improper data use	69	22.8%
Governance, legal and regulatory inadequacy + Privacy loss and improper data use	67	22.2%

Table 6-3. Ten most frequent risk co-occurrences.

The practical implication is that security mechanisms should be evaluated as socio-technical interventions. Their assessment should include who is monitored, who can challenge a decision, what happens after a false alarm, how model updates are governed and which actor carries responsibility for failure. Trustworthy analytics in ETSI zero-touch management [189], accountable and resilient AI-assisted networks [221] and secure future-network architectures [269] move in this direction by linking protection with assurance and responsibility. Nevertheless, the corpus remains much stronger in attack detection than in the analysis of downstream harms caused by defensive automation [146].

6.5 Privacy, Surveillance and Improper Data Use

Privacy loss is the second most frequent risk and appears in 163 studies. The prominence of privacy follows directly from the distributed and data-intensive character of AI-enabled networks. 5G/6G applications collect location, mobility, health, industrial, behavioral and communication data at high frequency. Edge processing can keep data close to the source, but it also multiplies the number of devices and intermediaries that handle sensitive information. The early 6G privacy literature already described machine learning as both a means of protection and a source of leakage [244], [206].

Federated learning is the most common privacy-oriented response. It avoids transferring raw data to a central server, yet the corpus repeatedly notes that decentralized training does not guarantee privacy. Model updates can reveal information, malicious clients can infer properties of local datasets, and heterogeneous participation can weaken both protection and performance. Differentially private federated learning [71], privacy-preserving training under data heterogeneity [50] and standards for federated learning in digital-twin environments [58] demonstrate attempts to formalize these protections. Privacy-preserving computation in 5G-based IoT [144] and vehicular networks [86] extends the same concern to latency-sensitive distributed applications [53], [98], [170].

Privacy risk is especially acute in healthcare, wearables and connected mobility. Health data can reveal disease, behavior and identity, while mobility traces expose routines and relationships. Privacy-aware wearable edge intelligence [159], UAV and V2X protection [172], pandemic-response networks [177] and digital-twin security [301] illustrate the breadth of the issue. The potential harm is not limited to disclosure. Inaccurate profiles, unauthorized

secondary use and the inability to withdraw data can affect insurance, employment, access to services and personal autonomy [93], [217].

The corpus distinguishes privacy from security in principle, but the two categories are strongly coupled. Security controls protect confidentiality and integrity, while privacy asks whether data should be collected, linked or retained in the first place. A perfectly secure system can still be privacy-invasive if it continuously monitors individuals without meaningful consent. This distinction is visible in work on AI-supported radar surveillance [81] and mass-surveillance healthcare frameworks [105]. Surveillance appears explicitly in only 12 studies, which suggests that the ethical implications of pervasive sensing are less developed than the technical mechanisms for securing the resulting data.

Consent is difficult in future-network environments because data flows are continuous, distributed and often inferred rather than directly provided. Users may accept a service without understanding that network intelligence can derive location, activity or health attributes. Devices may communicate on behalf of users, and digital twins may preserve information after the original context has changed. The ethical risk therefore includes a loss of contextual control. Privacy standards and governance must address the purpose of processing, data minimization, retention, inference and cross-domain reuse, rather than focusing only on encryption.

The risk is therefore best understood as improper control over data and inferences, not only as data theft. Literature recognizes leakage, confidentiality and privacy-preserving computation well, but pays less attention to collective privacy, group profiling and the cumulative effects of connected sensing. It also offers limited evidence about remedies after misuse occurs. These gaps help explain why privacy is frequent as a technical risk while surveillance, consent and redress remain comparatively marginal ethical topics.

6.6 Reliability, Explainability and Trust

Reliability, explainability and trust form a connected group. Unsafe or non-robust behavior appears in 159 studies, loss of trust in 153 and opacity in 116. A system cannot be reasonably trusted when it behaves unpredictably or cannot provide an intelligible account of its decisions. Nevertheless, the concepts remain distinct, and reliability concerns consistent performance, explainability concerns intelligible reasoning, and trust concerns justified stakeholder confidence.

Reliability becomes ethically significant when AI controls resources, mobility, health services or emergency communications. A small prediction error may be tolerable in a recommendation system but unacceptable in mission-critical network slicing or autonomous mobility. Safe and explainable telecommunications research [5], neuro-symbolic slice optimization [12] and trustworthy traffic scheduling [96] reflect this concern. Quality-of-trust indicators for 6G-enabled mass autonomy attempt to make trustworthiness measurable [145], while runtime safety control in O-RAN introduces intervention before unsafe actions are executed [120], [31], [116].

Opacity creates a second layer of risk. AI decisions can affect resource access, traffic classification, slice admission and security response, yet deep models often provide little explanation. Early work positioned explainability as a basis for trust between humans and machines [92]. Later studies apply XAI to RAN slicing [78], 5G performance analysis [6], channel estimation [8], intent-based networking [51], semantic communication [153], autonomous systems [116] and network-security decisions [232]. The field has therefore moved from general motivation to task-specific explanation methods.

Nevertheless, the presence of an explanation method does not guarantee meaningful transparency. Post-hoc feature attribution may show which input variables influenced a prediction without explaining whether the decision is normatively acceptable. Operators, developers, regulators and end users also require different levels of detail. A technically precise explanation may be unusable for a person affected by service denial, while a simplified explanation may be insufficient for audit. Research connecting causal reasoning with stakeholder trust [35] and neuro-symbolic approaches with verifiable automation [148],

[169] responds to this limitation by seeking explanations linked more closely to system logic.

Trust is the broadest of the three risks. It may refer to trust in a model, reputation among devices, confidence in data, trust between network stakeholders or public acceptance of the service. Global reputation management [40], zero-trust architectures [87], trustworthy, federated learning [48], trust-aware slicing [125] and assurance in advanced 5G edge systems [45] illustrate these different meanings. The breadth of the term can be useful, but it can also obscure what evidence is required. Trust should be justified by security, reliability, transparency and governance rather than treated as a property produced by naming a framework trustworthy [122], [139], [199].

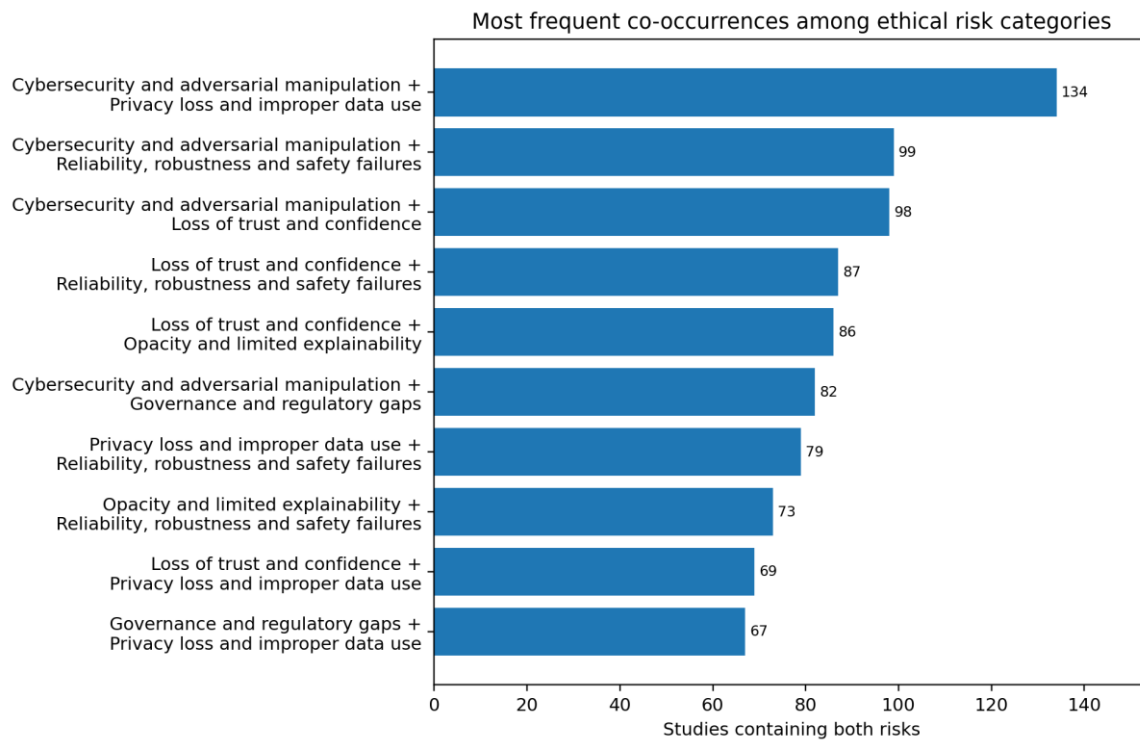


Figure 6-3. Most frequent co-occurrences among ethical risk categories.

Figure 6-3, presents the most frequent pairs of risks and demonstrates how strongly trust, reliability and opacity relate to security and privacy.

The co-occurrence pattern supports a layered interpretation. Security and privacy protect infrastructure and data, reliability supports dependable operation, explainability enables scrutiny, and governance assigns requirements and responses. Trust may emerge when these conditions are supported by evidence. Surveys of explainable 6G use cases [267], trustworthy edge intelligence [273] and accountable networks [221] increasingly recognize this structure, although empirical studies still evaluate individual technical metrics more often than stakeholder trust in the complete system [9], [25], [259].

6.7 Governance, Fairness, Human Autonomy and Societal Impacts

Governance, legal and regulatory inadequacy is the fifth most frequent risk, appearing in 134 studies. This category covers uncertainty about applicable rules, fragmented responsibilities, weak standards, cross-jurisdictional inconsistency and the absence of procedures for audit, certification or incident reporting. As AI becomes embedded in telecommunications

infrastructure, governance cannot be added after deployment. Decisions about data access, model updates, acceptable risk and emergency intervention are part of network design.

Recent literature moves from broad ethical principles toward sector-specific governance. The AI Regulatory Readiness Index evaluates legal preparedness across telecommunications jurisdictions [16], while incident-reporting research examines how AI failures can be incorporated into telecommunications law and policy [17]. Standards-oriented work addresses AI-native architectures [123], privacy-preserving federated learning [58], wearable health systems [95] and compliance in digital-twin applications [190]. These studies indicate growing recognition that voluntary technical safeguards are insufficient when systems affect safety, rights and critical services.

Accountability is less frequently explicit than governance. Only 34 studies use language directly associated with responsibility, liability or answerability. This difference suggests that the reviewed work more often proposes rules and frameworks than it specifies who must justify a decision or compensate for harm. Traceable and auditable federated AI [89], accountable and resilient networks [221] and runtime verification [120] offer mechanisms that can support accountability. However, a log or explanation does not allocate responsibility among operators, vendors, model developers, data providers and application owners.

Fairness and discrimination appear in 30 studies. The issue arises in resource allocation, packet scheduling, federated learning and service access. Fairness-aware packet scheduling [46], explanation-guided fair federated learning for RAN slicing [212] and standard-compliant fairness auditing [3], [218] show that the concept can be defined technically. Yet fairness metrics depend on which groups, outcomes and trade-offs are selected. A mathematically balanced allocation may still disadvantage users with poor coverage, limited devices or atypical traffic. The ethical risk therefore includes both biased decisions and narrow definitions of who counts in the evaluation.

Social exclusion and unequal benefit appear in 28 studies, while explicit loss of human control appears in 15. Research on inclusive user experience [65], human-centric communication [171], augmented AI with human judgment [220] and embodied AI [166] examines agency and dependency. However, many technical papers treat a human-centered focus as an aspiration. They do not always specify participation, accessibility, ways to challenge decisions or protections for people who do not want continuous connectivity.

Environmental and resource burden appears in 62 studies. Sustainable architecture [22], green-network incentives [33], smart-city research [24], [99] and trustworthy edge intelligence [273] connect ethical responsibility with energy and infrastructure. Efficiency claims alone are insufficient and assessment should consider lifecycle impacts, rebound effects, device replacement and whether costs are shifted to communities with limited decision-making power.

Figure 6-4, compares primary-risk orientations. It demonstrates that fairness, human control, exclusion and surveillance are rarely the central organizing concern of a study, even when they may appear as secondary issues.

The primary-risk distribution confirms that the field is strongest where ethical requirements can be represented as technical objectives. Security, privacy and explainability together constitute the primary focus of 182 studies. By comparison, only 12 studies are primarily oriented toward fairness, human control, exclusion, surveillance or explicit accountability. This does not invalidate technical work. It shows that a more balanced research agenda would need stronger collaboration with law, social science, human-computer interaction and environmental assessment [88].

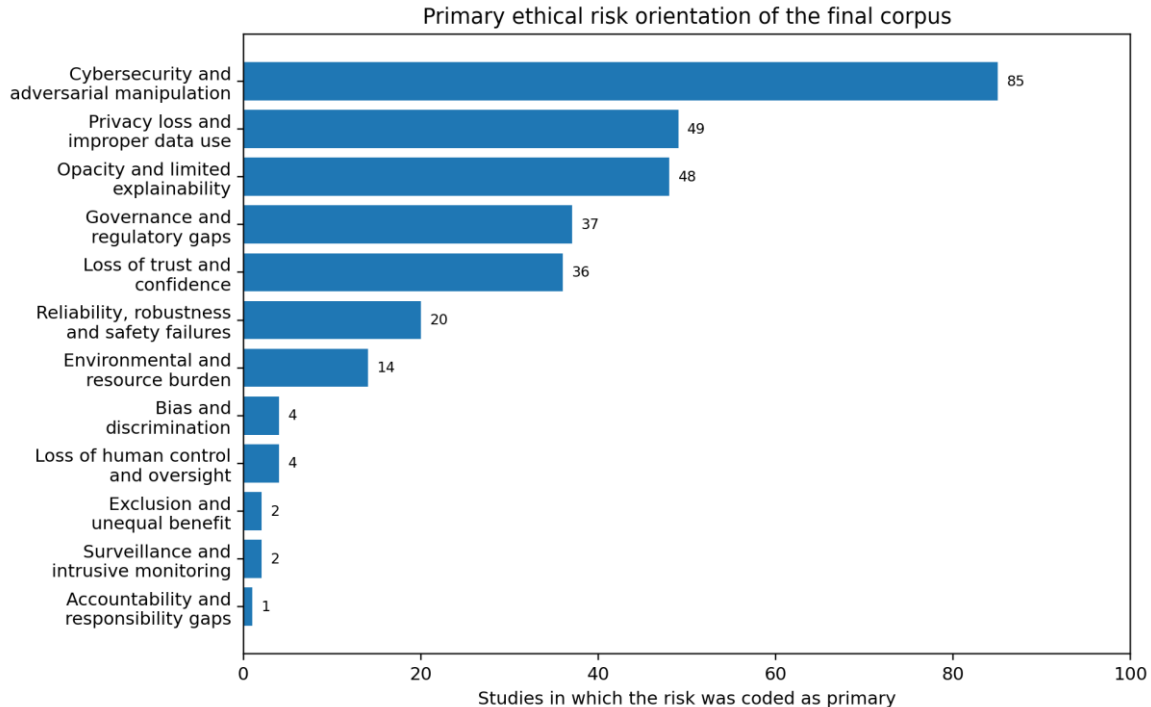


Figure 6-4. Primary ethical risk orientation of the included studies.

6.8 Risk Distribution Across Network Generations, Domains and Stakeholders

Risk prevalence differs between 5G-related and 6G-related research, although the groups overlap. Using the same network-generation classification applied in Chapter 5, the final corpus contains 140 studies with an explicit 5G or B5G connection and 203 with an explicit 6G connection. 74 address both, while 33 use broader next-generation framing. Figure 6-5 on the following page compares the prevalence of the principal risks in the two non-exclusive groups.

Security remains the largest risk in both groups, appearing in 67.1% of 5G-related studies and 62.6% of 6G-related studies. Privacy is also similar, at 53.6% and 53.7%, respectively. Unsafe, unreliable or non-robust AI behavior appears in 50.0% of 5G-related studies and 54.2% of 6G-related studies. Trust is more visible in the 6G-related group, appearing in 57.6% compared with 47.1% of 5G-related studies. Governance appears in 47.9% of 5G-related studies and 45.3% of 6G-related studies. Opacity and limited explainability appear in 37.9% of both groups. These overlapping percentages describe the visibility of risks within each network-generation group and do not assign studies to mutually exclusive categories.

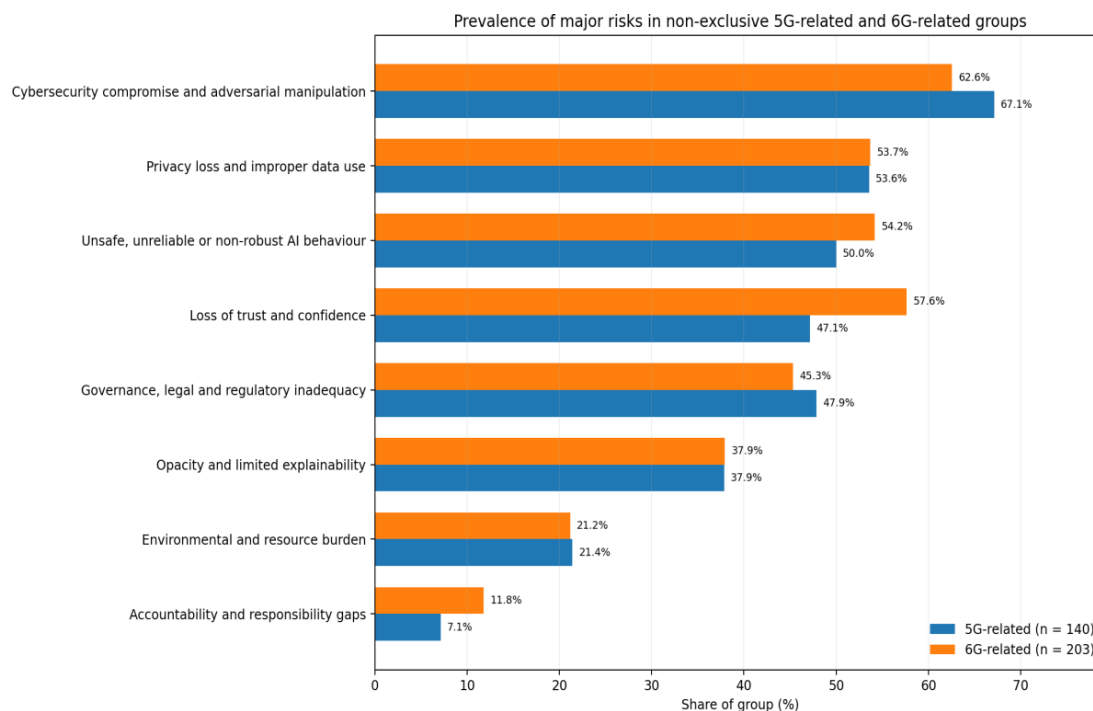


Figure 6-5. Frequency of major risks in overlapping 5G-related and 6G-related groups.

Risk category	5G-related studies (n = 140)	6G-related studies (n = 203)
Cybersecurity compromise and adversarial manipulation	94 (67.1%)	127 (62.6%)
Privacy loss and improper data use	75 (53.6%)	109 (53.7%)
Unsafe, unreliable or non-robust AI behaviour	70 (50.0%)	110 (54.2%)
Loss of trust and confidence	66 (47.1%)	117 (57.6%)
Governance, legal and regulatory inadequacy	67 (47.9%)	92 (45.3%)
Opacity and limited explainability	53 (37.9%)	77 (37.9%)
Environmental and resource burden	30 (21.4%)	43 (21.2%)
Accountability and responsibility gaps	10 (7.1%)	24 (11.8%)

Table 6-4. Comparison of major risks in overlapping 5G-related and 6G-related studies.

Application context also shapes risk. In IoT, industrial and edge-intelligence studies, security and privacy dominate because large numbers of heterogeneous devices exchange sensitive data across distributed infrastructure. In healthcare and wearables, privacy, safety and governance become closely linked because incorrect or unauthorized processing can affect physical well-being and highly sensitive information [18], [20], [107], [137]. Vehicular and UAV research combines security, privacy and reliability because communication failures or manipulated decisions can have immediate physical consequences [36], [100], [128], [160], [117].

Network management, RAN and slicing research gives greater attention to trust and explainability. These systems allocate spectrum, admit services and change policies dynamically, so operators need evidence about why an AI controller selected an action. Trustworthy resource management [131], explainable slicing [78], [188], zero-touch management [163] and causal orchestration [132] illustrate this orientation. Smart-city and public-infrastructure studies add questions of surveillance, inclusion and public legitimacy because network intelligence interacts with urban services and shared spaces [99], [118], [147].

The affected stakeholder is not always the actor operating the AI model. End users and data subjects face privacy loss, exclusion and unexplained service decisions. Network operators face reliability, security and compliance obligations. Developers and vendors influence model design, documentation and update mechanisms. Application providers determine how network

intelligence affects healthcare, transport or employment. Regulators and public authorities define acceptable risk and remedies. Communities and future generations may bear environmental and social costs even when they are not direct service users [52], [60], [287].

The cross-domain comparison therefore supports risk-based contextualization rather than one universal checklist. Common principles such as privacy, robustness and accountability remain relevant, but their practical meaning depends on the network function, the application and the stakeholder. This finding prepares the transition to Chapter 7, where methods, frameworks and standards are examined according to the risks they address and the stage of the lifecycle in which they operate.

Stakeholder group	Principal risk exposure	Relevant assurance need
End users and data subjects	Privacy, surveillance, unexplained decisions, exclusion, discrimination and loss of control.	Access to explanation, consent, contestation, remedy and proportional protection.
Network operators	Security compromise, unsafe automation, service disruption and compliance failure.	Monitoring, runtime intervention, audit evidence, incident response and accountable operation.
AI developers and model providers	Bias, opacity, poor robustness, unsafe updates and undocumented limitations.	Testing, documentation, traceability, lifecycle monitoring and disclosure of model limits.
Infrastructure and device vendors	Vulnerable components, insecure interfaces, data leakage and interoperability failures.	Secure design, update support, assurance of components and supply-chain accountability.
Application and service providers	Misuse of network intelligence in health, mobility, finance, employment or public services.	Purpose limitation, domain validation, user protection and meaningful human review.
Regulators and public authorities	Fragmented governance, unclear liability, weak standards and unequal social impact.	Rules, certification, oversight, reporting, enforcement and protection of public interests.
Communities and future generations	Environmental burden, unequal infrastructure access and concentration of power.	Inclusive planning, lifecycle assessment and fair distribution of benefits and costs.

Table 6-5. Stakeholder-oriented interpretation of the principal ethical risks.

6.9 Integrated Interpretation and Direct Answer to RQ2

The direct answer to RQ2 is that literature most frequently identifies cybersecurity compromise, followed by privacy loss, unreliable or unsafe behavior, loss of trust, governance and regulatory inadequacy, and opacity. These six risks appear in more than one third of the 302 studies. Security in 192, privacy in 163, reliability in 159, trust in 153, governance in 134 and opacity in 116. Environmental burden is established but secondary, while accountability, fairness, data integrity, exclusion, human control and surveillance receive much less explicit attention [202].

The ranking indicates that ethical risk is primarily understood through the language of technical trustworthiness. A system is considered ethically problematic when it can be attacked, leaks data, fails under changing conditions or produces decisions that cannot be explained. This framing is appropriate for telecommunications because network failures can propagate quickly and affect critical services. It also supports measurable evaluation. Attack- detection rates, privacy budgets, robustness tests and explanation fidelity can be incorporated into experiments. The strong presence of these categories is therefore both substantively justified and methodologically convenient.

At the same time, the corpus shows that technical properties are interdependent. Security co-occurs with privacy in 134 studies and with reliability in 99. Trust co-occurs with reliability in 87 and with opacity in 86. These combinations imply that no single metric can establish ethical acceptability. A secure model may be opaque, a privacy-preserving model may be unreliable, and an explainable model may still be unfair. Trust is best interpreted as a conclusion supported by evidence across several dimensions rather than as a separate feature that can be added through one mechanism [234], [275].

The risk landscape also changes as research moves toward 6G. In 5G, much work applies AI to identifiable functions such as intrusion detection, slicing and resource management. In 6G, the network is increasingly described as AI-native, agentic and autonomous. Ethical concern therefore expands from the correctness of individual models to the governance of continuously acting systems. Runtime intervention, cognitive sovereignty, agent security and cross-stakeholder trust become more prominent [61], [83], [120], [148]. The 6G literature anticipates systems that not only predict but also plan, negotiate and execute actions across administrative domains [231].

A second conclusion concerns the difference between visibility and importance. Fairness, human oversight, social exclusion and surveillance have low frequencies, but these categories address fundamental questions about rights and power. Their limited presence should not be interpreted as evidence of low risk. It more likely reflects the composition of the literature and the difficulty of representing these issues through conventional networking experiments. Fairness-auditing studies [3], [218], human-centric communication research [171] and embodied-AI analysis [166] demonstrate that these topics can be investigated, but they remain specialized rather than mainstream.

The corpus also reveals a governance gap between proposing principles and assigning responsibility. Governance and regulation are mentioned in 134 studies, yet explicit accountability appears in only 34. Many articles call for standards, compliance or trustworthy frameworks without specifying which actor must explain failure, preserve evidence, compensate affected parties or stop an unsafe model. This distinction matters in multi-vendor and cross-domain networks. Responsibility may be distributed among infrastructure providers, model developers, data owners, application providers and public authorities, creating the possibility that every actor controls only part of the system while no actor accepts the whole consequence.

The risk categories can be arranged into four layers. The first layer concerns protection of infrastructure and data such as security, privacy and data integrity. The second concerns dependable model behavior such as reliability, robustness and safety. The third concerns epistemic and institutional assurance such as explainability, trust, governance and accountability. The fourth concerns societal distribution and human values such as fairness, inclusion, autonomy, surveillance and environmental impact. Literature is strongest on the first two layers, increasingly active in the third, and comparatively weak in the fourth. This layered model provides a concise synthesis of the field and avoids treating all ethical terms as interchangeable [66], [81].

In summary, the corpus concentrates on risks that can be translated into technical controls. Cybersecurity is the most frequent, followed by privacy, reliability, trust, governance and explainability. The main imbalance is the uneven depth of analysis and technical harms are investigated through models and experiments, while institutional and societal harms are more often acknowledged briefly or addressed by specialized publications. RQ2 is therefore answered by both the frequency ranking and this imbalance [49], [62], [216].

6.10 Limitations of the RQ2 Analysis

The first limitation is the use of titles, abstracts and author keywords rather than complete full-text coding. The final Master provides a structured abstract-level synthesis, but abstracts compress complex studies and may omit secondary ethical discussions, detailed limitations or stakeholder consequences. A study may contain a substantial accountability or fairness section that is not visible in its abstract. Conversely, an abstract may use terms such as trustworthy or responsible without defining them in the full text. The results should therefore be interpreted as the visible risk profile of the indexed corpus, not as an exhaustive inventory of every argument in all 302 publications.

The second limitation concerns terminology. Ethical concepts overlap and are used inconsistently. Trust can refer to interpersonal confidence, device reputation, model performance or cryptographic assurance. Safety, robustness, resilience and reliability may be distinguished carefully in one study and used almost interchangeably in another. Governance may include accountability, while other studies treat accountability as a separate principle. The codebook reduces this variation by applying stable categories, but every normalization decision simplifies the authors' original language. The separate counts should therefore be read together with the qualitative explanations and co-occurrence analysis.

Third, the coding is multi-label, and frequency does not measure severity. Broad terms such as security and privacy are easier to identify than risks expressed indirectly. Human autonomy may be embedded in design choices without explicit oversight language, and exclusion may appear through coverage or affordability. Conservative coding reduces unsupported classification but may undercount implicit risks.

Fourth, the corpus is shaped by the search strategy, databases and inclusion criteria. IEEE Xplore and technical publishers contribute a large share of the records, which favors engineering studies and explains the prominence of security, privacy and performance-related concerns. Legal, social-science and ethics research may use different terminology or discuss telecommunications indirectly and may therefore be underrepresented. The inclusion of Web of Science, ScienceDirect and Springer broadens disciplinary coverage, but the final dataset remains centered on technical research about AI and future networks.

Fifth, the period ending in June 2026 contains a partial final year. The corpus already includes 111 studies from the first six months of 2026 but indexing and publication schedules may

change the final annual distribution. The rapid growth of agentic AI, standards and governance research means that both counts, and terminology are likely to evolve. The temporal figures should therefore be treated as a snapshot taken at the search date. They are useful for showing the direction of development but not for predicting the complete 2026 output.

Sixth, 5G-related and 6G-related groups are non-exclusive. Studies addressing both generations contribute to each group because the purpose is to compare risk prevalence in literature connected with each generation. This approach reflects the continuity between B5G and 6G, but it means that the group percentages are not independent samples. Differences should be interpreted descriptively rather than as statistically significant contrasts. Some records also use broader terms such as next-generation networks, edge intelligence or O-RAN without making a precise generation claim, and these studies may still influence the overall ranking.

Eighth, the analysis does not evaluate the effectiveness of solutions. A study was coded for privacy when it proposed a privacy-preserving method and for explainability when it used XAI, but Chapter 6 does not determine whether the method provides adequate protection, meaningful explanation or acceptable trade-offs. Frequency can reveal where research effort is concentrated, not whether a risk has been solved. That assessment belongs to RQ3, which examines methods, frameworks and standards, and to RQ4, which evaluates remaining gaps and the absence of validation.

Finally, taxonomy supports synthesis rather than legal adjudication or deployment-specific risk calculation. It does not estimate probability, impact or residual risk and cannot resolve every value conflict. Stronger monitoring may improve security while increasing surveillance, and stricter privacy may reduce data available for fairness auditing. Such interactions require contextual decisions and stakeholder participation. The analysis nevertheless establishes which risks dominate, co-occur and remain underrepresented.

Chapter 7 Research Question 3: Methods, Frameworks and Standards for Trustworthy AI in 5G/6G

This chapter addresses the third research question. What methods, frameworks, or standards are proposed to support trustworthy and ethically aligned AI in 5G/6G? The analysis uses the 302 studies classified as YES in the final Master Excel and covers publications from January 2020 to June 2026. Whereas Chapter 6 identified the risks discussed in the literature, the present chapter examines how researchers attempt to manage those risks. The focus is not limited to individual algorithms. It includes technical mechanisms, system-level architectures, governance arrangements, standardization pathways and compliance instruments that translate broad principles such as privacy, transparency, robustness, fairness and accountability into operational requirements [129], [208], [291].

Literature does not present one uniform solution. Instead, it proposes a layered collection of measures that operate at different points of the AI and network lifecycle. Federated learning limits centralized data transfer, explainable AI supports inspection of model behavior, privacy-enhancing technologies protect data and model updates, blockchain supports provenance and auditability, robust learning and runtime controls reduce unsafe operation, while standards and governance frameworks define responsibilities and common interfaces. These measures are often combined because a single method rarely addresses all relevant risks. A privacy-preserving model may still be opaque, a transparent model may remain vulnerable to poisoning, and a secure architecture may lack a clear accountability process [98], [164], [282].

The chapter first explains the analytical procedure and the operational taxonomy. It then presents the overall frequency and temporal development of solution families. Separate sections examine federated and privacy-preserving learning, explainability and trust, security and runtime assurance, and the emerging standards and governance landscape. The final sections compare 5G and 6G orientations, evaluate the maturity of the proposed approaches, provide a direct answer to RQ3 and explain the limitations of an abstract-based synthesis [242], [268], [275].

7.1 Purpose and Analytical Procedure

RQ3 requires classification as well as interpretation. A simple list of named tools would not be sufficient because the same term may describe different levels of intervention. For example, federated learning is an algorithmic training paradigm, but a federated-learning paper may also propose architecture, a trust-management framework and a standards-aligned governance process. Conversely, a document described as a framework may consist mainly of one technical model. The analysis therefore distinguishes the functional role of a proposal rather than relying only on the label used by the authors.

The unit of analysis was the unique included study. Coding used the title, abstract and available author keywords in the final Master Excel. Each record was examined for three non-exclusive forms of intervention. The first was a technical method or mechanism, including algorithms, protocols, privacy controls, learning schemes and verification techniques. The second was a framework, architecture or platform that combines several components or defines how actors and processes interact. The third was a standard, governance or policy instrument, including interoperability specifications, compliance models, certification approaches, regulatory frameworks and sectoral guidance. A study could be coded in more than one form because many recent proposals combine technical and institutional elements.

A second coding layer identified fourteen solution families. These included federated and distributed learning, explainability, trust and assurance, blockchain and traceability, standards and interoperability, privacy-enhancing technologies, security and adversarial defense, governance and compliance, digital twins, human-centric design, runtime safety, neuro-symbolic and causal AI, fairness mechanisms, and sustainability-oriented measures. Coding was conservative. A general statement that a network should be secure was not treated as a security method unless the abstract identified a mechanism, framework or operational approach. Similarly, fairness was coded as a solution only when the study discussed assessment, auditing, certification, fairness-aware learning or bias mitigation [238], [245], [277].

The coding was multi-label because trustworthy AI depends on combinations of controls. Co-occurrence was therefore treated as an analytical result rather than as duplication. The analysis measured how often solution families appeared together, how their use changed by year, and whether their prevalence differed between studies explicitly connected with 5G/B5G and

those explicitly connected with 6G. Network-generation groups are non-exclusive, since studies that discuss both generations contribute to both groups. This reflects the continuity between deployed 5G systems, B5G evolution and emerging 6G architectures.

The analysis also examined lifecycle location and evaluation maturity. Lifecycle coding identified whether a proposal acted during data collection, model training, inference, network deployment and orchestration, monitoring and auditing, or incident response and adaptation. Maturity coding distinguished reviews and conceptual syntheses, design proposals, simulation or experimental evaluations, prototypes and testbeds, applied empirical or case-based assessments, and descriptive discussions. These categories do not rank scientific quality. They indicate how far literature moves from conceptual recommendation toward operational validation.

7.2 Operational Classification of Methods, Frameworks and Standards

The distinction between a method, a framework and a standard is central to the interpretation of RQ3. A method is an operational technique that changes how data are processed, how a model is trained, how a decision is explained or how an attack is detected. Differential privacy, secure aggregation, explanation-guided learning and Byzantine-robust aggregation are examples. A framework connects several methods and specifies components, data flows, control loops, stakeholder roles or assurance stages. The trustworthy AI/ML framework proposed for B5G/6G, for example, treats input protection, explainable operation and dependable outputs as connected requirements rather than isolated model properties [214]. A standard or governance instrument goes further by defining common obligations, interfaces, evidence, terminology or conformance expectations across organizations [191], [241], [253].

Literature increasingly combines the three levels. Federated learning can be placed inside an explainable B5G/6G architecture so that local data remains distributed and model decisions can be inspected [43]. A standards-aligned digital-twin framework can combine differential privacy, homomorphic encryption, secure multi-party computation and blockchain-based logging [58]. Fairness can be operationalized through an auditing and certification tool that maps technical metrics to national telecommunications requirements [218].

Fourteen families as the Table 7-1 shows in the next page , were retained because they represented recurring and analytically distinct responses. Federated learning was separated from privacy-enhancing technologies because decentralized training does not guarantee privacy itself. Explainability was separated from trust because explanations are one source of trust, but trust may also arise from reputation, verification, reliability or institutional assurance. Blockchain was separated from governance because an immutable ledger can support evidence without deciding which behavior is legitimate. Standards were separated from governance because interoperability specifications and certification criteria have a different role from organizational accountability and legal readiness [26], [200], [229].

7.3 Overall Distribution and Temporal Development of Proposed Solutions

Federated and distributed learning is the most frequent solution family, appearing in 99 studies, or 32.8% of the corpus. Explainability and interpretability follow closely with 98 studies, or 32.5%. Trust, reputation and assurance appear in 68 studies, blockchain and traceability in 63, and standards and interoperability in 62. Privacy-enhancing technologies and explicit security or adversarial-defense mechanisms each appear in 51 studies. Governance and compliance are identified in 36, while digital-twin or simulation-based assurance appears in 34. Human-centric design is present in 15, runtime safety in 11, neuro- symbolic or causal assurance in 10, and explicit fairness and sustainability mechanisms in four studies each [69], [76], [173].

The frequency ranking is presented in Figure 7-1 on the following page. Because coding is multi-label, the values do not come to 302. The purpose of the figure is to show the relative visibility of solution families, not to allocate every study to one exclusive category.

The ranking shows two dominant methods and a broad middle tier. Federated learning and explainability are almost equally frequent, but they address different parts of the problem. Federated learning changes where and how data are used, while XAI changes what evidence is available about model behavior. Trust, blockchain and standards form the middle tier and frequently connect the two dominant families with assurance, provenance and interoperability. The lower frequencies of fairness, safety and sustainability indicate that these concerns are less often converted into explicit operational mechanisms [138], [250], [278].

Solution family	Operational meaning	Typical mechanisms	Primary ethical purpose
Federated and distributed learning	Collaborative model training without routine centralization of raw data	Federated averaging, clustered FL, split or distributed learning	Data locality, privacy and cross-domain collaboration
Explainability and interpretability	Methods that reveal or constrain the basis of AI decisions	SHAP, LIME, interpretable models, explanation-guided learning	Transparency, operator understanding and contestability
Privacy-enhancing technologies	Formal or cryptographic controls on data, updates and inference	Differential privacy, secure aggregation, homomorphic encryption, ZKPs	Privacy, confidentiality and data minimization
Security and robust learning	Controls against malicious inputs, clients, models and network behavior	IDS, adversarial defense, Byzantine robustness, zero trust	Integrity, resilience and safe continuity
Trust, reputation and assurance	Evidence and mechanisms for justified confidence in actors and models	Reputation, verification, uncertainty, auditability, QoT	Trustworthiness and accountability
Blockchain and traceability	Distributed evidence and rule enforcement	DLT, smart contracts, provenance and immutable logs	Auditability, coordination and non-repudiation
Standards and interoperability	Common requirements, interfaces, evidence and conformance processes	3GPP, ETSI, O-RAN, ITU, ISO/IEC, IEEE and national standards	Consistency, interoperability and certification
Governance and compliance	Organizational and legal control of AI across its lifecycle	Risk management, incident reporting, sovereignty and regulatory readiness	Responsibility, oversight and remedy
Digital twins and simulation-based assurance	Virtual or mirrored environments used to test AI-enabled network behavior	Network digital twins, simulation, scenario testing and what-if analysis	Pre-deployment assurance, monitoring and safer validation
Runtime safety and control	Controls that constrain or verify actions during operation	Runtime assurance, predictive verification, fail-safe control and rollback	Safety, reliability and human intervention
Neuro-symbolic, causal and verifiable AI	Methods that combine learned models with structured reasoning or causal evidence	Knowledge graphs, causal reasoning, symbolic constraints and formal verification	Explainability, verifiability and accountable reasoning
Human-centric, inclusive and participatory design	Methods that incorporate stakeholder needs, participation and oversight	Human-in-the-loop design, participatory evaluation, acceptance studies and accessibility	Human autonomy, inclusion and social legitimacy
Fairness and bias mitigation	Methods that measure or reduce unequal model or service outcomes	Fairness metrics, bias auditing, fairness-aware learning and certification	Fairness, non-discrimination and equitable treatment
Sustainability and resource efficiency	Methods that reduce or govern computational and environmental burden	Energy-aware learning, resource-efficient models and sustainability-aware optimization	Environmental responsibility and sustainable operation

Table 7-1. Operational classification used for the RQ3 coding.

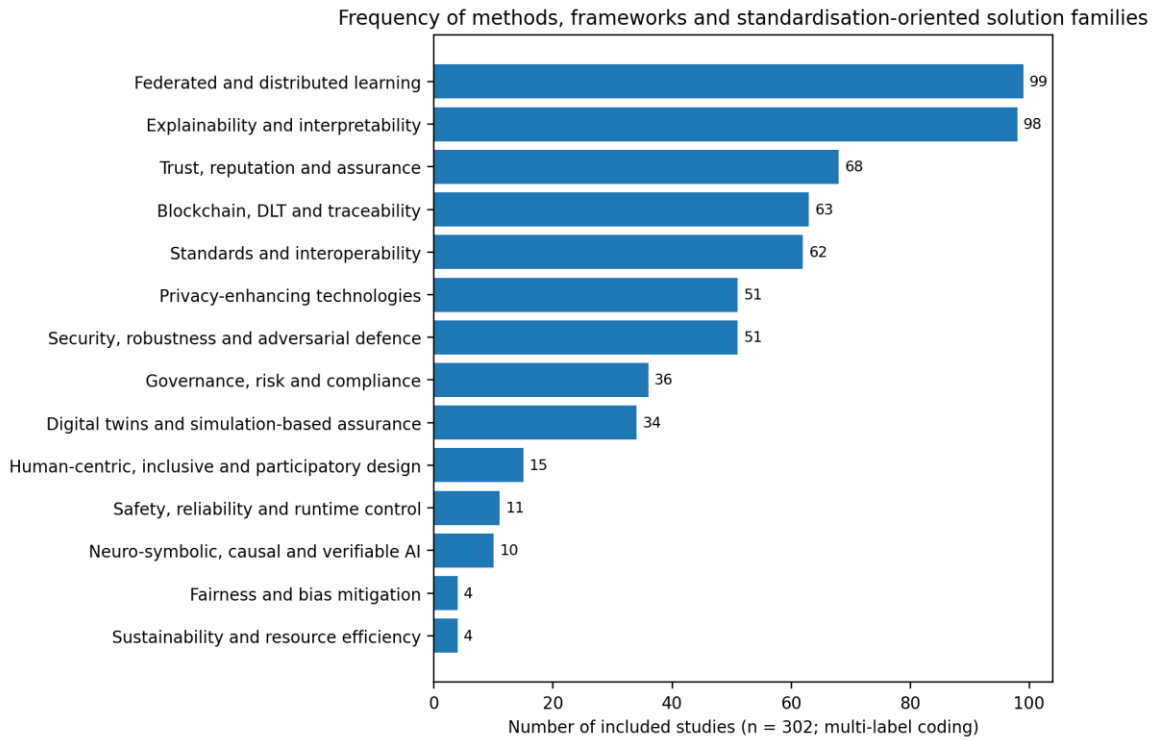


Figure 7-1. Frequency of solution families in the 302 included studies.

Solution family	Studies	Percentage of corpus	Typical form
Federated and distributed learning	99	32.8%	Technical method and architecture
Explainability and interpretability	98	32.5%	Technical method and assurance
Trust, reputation and assurance	68	22.5%	Framework and monitoring mechanism
Blockchain, DLT and traceability	63	20.9%	Technical infrastructure and audit mechanism
Standards and interoperability	62	20.5%	Standard, reference architecture and conformance
Privacy-enhancing technologies	51	16.9%	Technical and cryptographic method
Security, robustness and adversarial defense	51	16.9%	Technical method and security architecture
Governance, risk and compliance	36	11.9%	Governance or policy framework
Digital twins and simulation-based assurance	34	11.3%	Architecture and evaluation environment
Human-centric, inclusive and participatory design	15	5.0%	Design and adoption framework
Safety, reliability and runtime control	11	3.6%	Runtime mechanism and assurance framework
Neuro-symbolic, causal and verifiable AI	10	3.3%	Model and verification method
Fairness and bias mitigation	4	1.3%	Audit, learning constraint and certification
Sustainability and resource efficiency	4	1.3%	Architecture and incentive framework

Table 7-2. Frequency and typical form of proposed solution families.

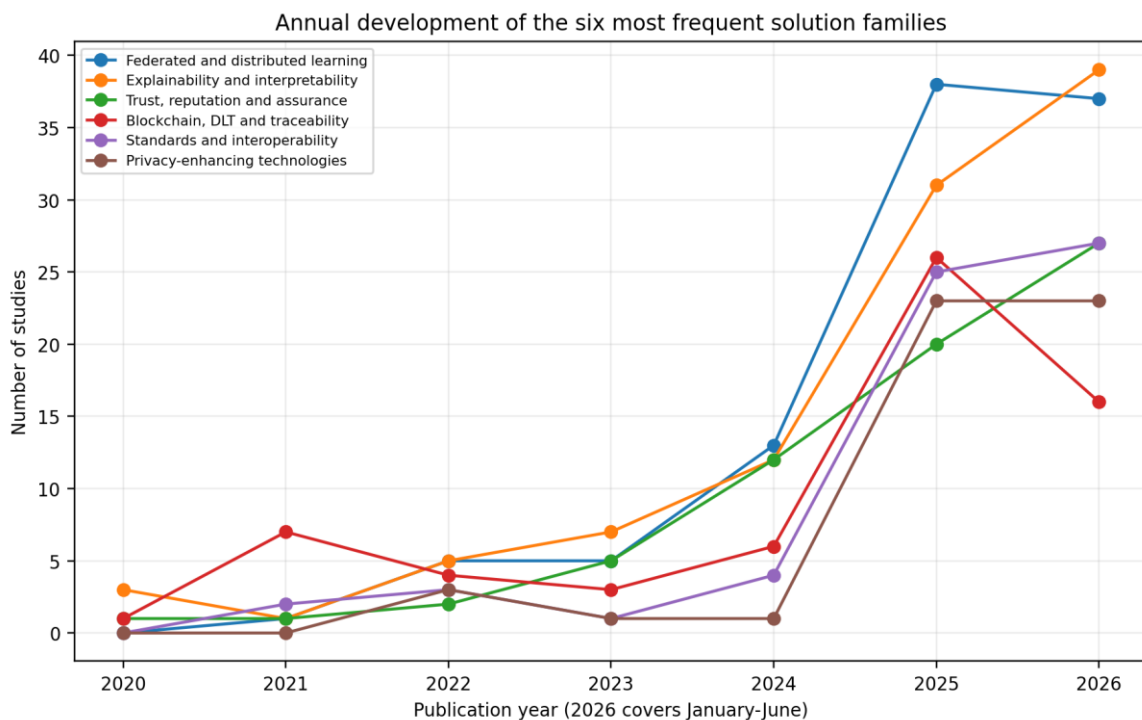


Figure 7-2. Annual development of the six most frequent solution families from 2020 to June 2026.

The temporal development of the six leading families is shown in Figure 7-2 on the following page. The 2026 values cover January to June and must not be interpreted as complete annual totals.

The early period was shaped by explainability, privacy surveys and initial trustworthy-AI concepts [92], [145], [244]. The number of concrete architectures grows after 2023, and the largest expansion occurs in 2025 and the first half of 2026. Federated learning, XAI and blockchain grow rapidly in 2025, while standards, governance, runtime safety and sovereign or agentic control gain visibility in 2026 [58], [61], [120], [123], [148]. Additional studies in the corpus address the same point [165]. This progression indicates a movement from isolated model properties toward lifecycle and architecture-level assurance [84], [154], [231].

7.4 Federated, Distributed and Privacy-Preserving Learning

Federated and distributed learning forms the largest technical response in the corpus. Its appeal is closely connected with the structure of 5G/6G systems. Data are generated across devices, vehicles, healthcare sensors, industrial equipment and edge nodes, while legal, commercial and practical constraints limit centralized transfer. Federated learning allows local participants to train a shared model without routinely sending raw records to one repository. Early B5G work

used this approach for privacy-preserving pandemic response and distributed edge intelligence [177], [265]. Later studies applied it to network slicing, anomaly detection, healthcare, tactile IoT, smart cities, digital twins and multimodal wearable intelligence [28], [29], [43], [159], [47].

However, the literature does not treat federated learning as sufficient by itself. Forty-two studies combine federated learning with explicit privacy-enhancing technologies, making this the most frequent solution co-occurrence in the corpus. Model updates can leak information, gradients can reveal local examples, and malicious clients can poison the global model. Consequently, proposals add differential privacy, secure aggregation, homomorphic encryption, secure multi-party computation, authenticated updates, reputation scoring or trusted execution. Differentially private FL frameworks for 6G attempt to calibrate noise while maintaining useful model performance [50], [71], [108]. Standard-aligned digital-twin work combines several cryptographic controls rather than depending on decentralization alone [58], [225], [280].

The relationship between privacy and utility is one of the most persistent design tensions. Differential privacy can provide a measurable privacy budget, but stronger noise may reduce detection accuracy or destabilize learning under heterogeneous data. Encryption can protect updates in transit and during aggregation, but increases computation, latency and energy consumption. Secure multi-party computation reduces exposure to the aggregator but may be difficult in highly mobile environments. These trade-offs are especially important in ultra- low-latency control, where a mathematically strong protection mechanism may be operationally unsuitable if it delays decisions or prevents timely adaptation.

Federated learning also changes the trust model. Instead of trusting one central repository, the system must trust participant admission, update validation, aggregation rules and the process used to exclude malicious or low-quality clients. Research on trustworthy, federated intrusion detection shows that non-identically distributed data can reduce reliability even without an active attacker [48], [143]. Byzantine-robust orchestration and authenticated zero-knowledge proofs attempt to verify contributions without revealing private information [67], [293]. These mechanisms convert trust from an informal expectation into checks on identity, update integrity and aggregation behavior [28], [53], [222].

Blockchain and distributed ledgers are frequently added to this architecture. Thirty-two studies combine blockchain with federated learning, and eighteen combine blockchain with privacy-enhancing technologies. The ledger is used to preserve model-update provenance, coordinate decentralized participation, record access decisions and support smart-contract enforcement. Blockchain-assisted approaches appear in vehicular networks, O-RAN, IoMT and edge computing [86], [114], [136], [167]. Auditable federated GAN architectures use blockchain-backed traceability and smart contracts to connect technical learning with compliance evidence [89], [36], [135], [281].

Overall, federated learning is the dominant method because it fits the distributed topology of future networks and offers a practical response to data-locality concerns. Its ethical effectiveness is strongest when it is combined with privacy accounting, robust aggregation, provenance, explanation and governance. Literature increasingly recognizes this requirement, but full integration remains less common than the use of FL as a general privacy label.

Frequent combination	Studies	Interpretation
Federated learning + privacy technologies	42	Data locality is supplemented with formal or cryptographic protection.
Federated learning + blockchain/DLT	32	Distributed training is combined with provenance, coordination and audit logs.
Federated learning + security/robustness	32	Aggregation and client participation are protected against malicious updates.
Explainability + federated learning	27	Distributed models are made inspectable for operators and stakeholders.
Federated learning + trust/assurance	20	Participant, model and aggregation confidence are explicitly managed.
Federated learning + standards	19	Architectures address interoperability, interfaces or conformance requirements.
Blockchain + privacy technologies	18	Traceability is combined with encryption, access control or privacy-preserving computation.
Privacy technologies + security/robustness	18	Confidentiality and integrity controls are treated as complementary.

Table 7-3. Frequent combinations involving federated, privacy-preserving and traceability mechanisms.

7.5 Explainability, Trust, Reputation and Verification

Explainability is the second most frequent solution family, appearing in 98 studies. It entered the 6G discussion early as a means of rebuilding trust when network optimization moves from interpretable mathematical models toward data-driven deep learning [92]. Work on mass autonomy extended this idea by proposing quality-of-trust indicators for different human stakeholders, including users, engineers and legal actors [145]. Later research developed application-specific explanation methods for channel estimation, intrusion detection, path-loss prediction, slicing, semantic communication and zero-touch management [6], [8], [31], [74], [94]. Additional studies in the corpus address the same point [158], [188], [302].

Literature uses explainability in three main ways. The first is post-hoc interpretation, where tools such as SHAP, LIME or feature attribution explain a prediction after a model has been trained. The second is intrinsically interpretable design, where the model structure is intended to remain understandable. The third is explanation-guided learning, where explanations influence training or control rather than serving only as a report. Explanation-guided deep reinforcement learning for RAN slicing and fair federated learning illustrate the third approach [78], [212]. This is important because explanations become part of system control and service-level assurance rather than optional visualization.

Trust, reputation and assurance appear in 68 studies and overlap with explainability in 33. Trust is broader than model transparency. It includes confidence in data, devices, participants, network operators, automated agents and institutional processes. SIX-Trust treats future-network trust as an architecture that must cover security, privacy and multi-domain interaction [269]. Global reputation-management frameworks assess entities across distributed 6G environments [40]. Trustworthy AI/ML frameworks protect model inputs, support explainable operation and verify outputs [214]. These proposals recognize that trust must be supported by evidence at several layers.

Assurance mechanisms attempt to convert trust into inspectable claims. They include verification, validation, uncertainty estimation, audit logs, model cards, quality-of-trust indicators and continuous monitoring. Trustworthy resource-management research combines robust AI with explanations so that network decisions can be evaluated under changing conditions [115], [131]. Accountable and resilient network processes integrate threat detection,

explanation and recovery [221]. ETSI ZSM case studies connect trustworthy analytics with standardized management architecture and zero-touch orchestration [189], [272].

The corpus also moves toward causal and neuro-symbolic assurance. Causal learning is used to distinguish correlation from mechanisms and to improve stakeholder trust [35], [132]. Neuro-symbolic approaches combine neural pattern recognition with explicit rules, knowledge graphs or logical constraints [12], [169]. A proposed neuro-symbolic knowledge plane for 6G uses symbolic constraints to verify actions produced by large action models [148]. These approaches are less frequent than post-hoc XAI, but they directly address the need for verifiable reasoning in agentic and autonomous networks.

Explainability does not automatically create accountability. An operator may understand why a model selected an action but still lack authority to stop it, preserve evidence or assign responsibility. Assurance is therefore stronger when explanations relate to thresholds, approval rules, runtime controls and audit processes. Literature supports progression from visibility to actionable assurance, and the path is first reveal relevant model behavior, then test it against requirements, record the evidence and trigger corrective action when the requirement is not met.

7.6 Security, Robustness, Blockchain and Runtime Safety

Security and adversarial-defense mechanisms appear explicitly in 51 studies, while blockchain and traceability appear in 63. The security literature is more selective here than in Chapter 6 because general discussion of cybersecurity was not counted as a solution. Included records identify operational mechanisms such as intrusion detection, anomaly detection, adversarial defense, Byzantine-robust learning, zero-trust architecture and secure orchestration. These methods are important because AI-enabled networks are both targets of attack and active decision makers. A compromised model can misclassify traffic, allocate resources incorrectly or conceal malicious behavior [23], [38].

Explainable intrusion-detection frameworks are a major pattern. They combine deep learning or graph models with SHAP, LIME or interpretable autoencoders so that operators can examine why traffic was classified as malicious [1], [39], [44], [80]. Trustworthy, federated IDS proposals add distributed learning, privacy protection and robustness under heterogeneous data [48], [90], [143]. Other work combines transfer learning, post-hoc retraining or feedback loops to improve performance after explanations reveal unstable features [80], [223]. These

approaches illustrate how XAI can be used for model maintenance rather than only for presentation [140], [174].

Robustness methods address adversarial examples, poisoning, Byzantine clients, distribution shift and unknown attacks. A model that performs well under normal test data may fail when network traffic changes or when an attacker manipulates updates. Byzantine-robust federated orchestration attempts to identify and limit harmful contributions [67]. Trustworthy stacked-hybrid models and robust explainable IDS frameworks combine multiple detectors or attention mechanisms to reduce dependence on one model [11], [179]. The ethical value lies in reducing preventable failure, but robustness claims remain conditional on the threat model and evaluation data [42], [182], [237].

Blockchain-based traceability supports security by recording events, model updates and access decisions. It also contributes to accountability when records are accessible to authorized auditors and linked with clear responsibility. Blockchain-supported federated learning is proposed for trustworthy IoMT image analysis, O-RAN privacy protection and secure edge intelligence [114], [136], [265]. Smart-contract governance can enforce participation rules and preserve an immutable compliance trail [89]. However, the corpus rarely examines who controls the ledger, how incorrect records are corrected or how data-protection rights apply to immutable evidence [194], [235], [298].

Runtime safety is a smaller but strategically important family. Only eleven studies present explicit safety, fail-safe, fault-tolerant or runtime-verification mechanisms. A runtime safety copilot for AI-native O-RAN predicts unsafe control actions and enforces fail-safe constraints before they affect the near-real-time RIC loop [120]. Resilient and explainable AI work connects reliability with understandable failure handling [75]. Accountable network processes combine pre-deployment testing, monitoring, incident detection and recovery [221]. These mechanisms are especially relevant when AI controls mission-critical slices, connected vehicles, healthcare or industrial systems [138], [219], [264].

Neuro-symbolic and causal methods reinforce this trend. Causal graphs support fault diagnosis in cloud-native 5G base stations [72], while causal explainable reinforcement learning supports scalable resource orchestration [132]. Neuro-symbolic network-slicing and knowledge-plane architectures aim to combine learning flexibility with explicit constraints [12], [148]. These

approaches can make decisions more verifiable, but their symbolic rules and knowledge graphs must remain complete, current and free from hidden bias.

7.7 Standards, Governance, Compliance and Human-Centric Design

Standards and interoperability appear in 62 studies, while explicit governance, risk and compliance approaches appear in 36. This is a substantial increase in the 2025-2026 literature. The annual count of standards-oriented studies rises from four in 2024 to twenty-five in 2025 and twenty-seven in the first half of 2026. Governance and compliance show a similar increase. The change reflects the transition from general 6G vision papers toward concrete questions about AI-native architecture, cross-domain operation, certification and legal responsibility. The landscape of standards is fragmented across telecommunications, AI, cybersecurity and sectoral regulation. Thirteen studies explicitly mention 3GPP and thirteen refer to O-RAN specifications. Ten refers to ETSI or ZSM, eight to ITU or IMT-2030, eight to ISO/IEC standards and five to IEEE standards or the IEEE 7000 series. Four explicitly connect the EU AI Act or GDPR with AI-enabled network applications, three refer to NIST-related risk guidance, and two discuss TEC or national telecommunications standards. These counts indicate visibility in abstracts, and they do not show formal conformance or adoption.

Telecommunications standards provide architecture, interfaces and operational roles. Standards-oriented work on AI-native 5G/6G proposes enriched KPIs, semantic-communication protocols, privacy taxonomies and interoperability requirements that extend current 3GPP, ITU-T and O-RAN activities [123]. ETSI ZSM offers an architecture for automated management and provides a context for trustworthy analytics [189]. REASON and related AI-native architectures show how multi-access networks can expose intelligence across domains [124]. Network digital-twin research proposes reference functional architectures to reduce conceptual and implementation fragmentation [294], [295].

AI and ethics standards provide a different layer. They define processes for risk management, fairness, transparency, data governance and human oversight. Ethical and privacy analysis of 6G robotics connects the EU AI Act, GDPR and IEEE 7000 series with autonomous systems [127]. Standards-driven wearable digital twins combine security controls with AI governance in a health-monitoring context [165]. Fairness

assessment is moving toward auditable and certifiable tools. The Nishpaksh framework operationalizes TEC requirements and connects technical bias metrics with telecommunications certification [218], while related work develops broader fairness-assessment standards [3], [79], [226], [279].

Regulatory frameworks also address institutional preparedness. The AI Regulatory Readiness Index evaluates whether national telecommunications laws contain substantive obligations, operational oversight and enforcement capacity for AI in critical infrastructure [16]. AI incident-reporting proposals argue that telecommunications law must recognize algorithmic incidents that are not captured by conventional cybersecurity or data-breach categories [17]. Connected-vehicle research assesses drowsiness and fatigue systems in relation to the EU AI Act [7]. These studies shift attention from voluntary technical design to duties of reporting, documentation and remedy [88], [230].

Standard or regulatory family	Explicitly identified studies	Role in trustworthy AI
3GPP	13	Network architecture, interfaces, management and service requirements
O-RAN specifications	13	Open RAN control, RIC applications, interfaces and multi-vendor operation
ETSI and ZSM	10	Zero-touch management, orchestration and trustworthy analytics
ITU and IMT-2030	8	6G vision, requirements, terminology and global alignment
ISO/IEC standards	8	AI risk, security, governance, quality and management processes
IEEE standards and IEEE 7000 series	5	Ethical system design, transparency and human-centered processes
EU AI Act and GDPR	4	Legal risk classification, data protection, documentation and human oversight
NIST AI RMF and related guidance	3	Risk identification, measurement, management and governance
TEC and national telecom standards	2	National certification, fairness assessment and sector-specific conformity

Table 7-4. Standards and regulatory families are explicitly visible in the final corpus.

Governance frameworks are strongest when they assign lifecycle responsibility. Organizational Sovereign AI proposes end-to-end operator control over data, models, infrastructure and compliance in AI-native 6G [61]. Accountable and resilient AI processes define stages for prevention, detection and response [221]. Auditable federated architectures use blockchain logs and smart contracts to support traceability [89]. The earlier trustworthy AI/ML framework for B5G/6G Emphasizes protection of input, explainable operation and proper output [214].

Together, these proposals show a movement toward governance by design rather than governance as an external review after deployment

7.8 Cross-Lifecycle Integration, Network Generation and Evaluation

Maturity

The distribution of solution families differs between 5G/B5G and 6G, but the differences are evolutionary rather than absolute. The final evidence set contains 140 studies with an explicit 5G/B5G connection and 203 with an explicit 6G connection but the groups overlap. Federated learning appears in 25.7% of the 5G/B5G group and 36.0% of the 6G group. Explainability is similar at 30.7% and 31.5%. Trust and verification are more visible in 6G-related work, appearing in 26.6% compared with 20.7% of 5G/B5G studies. Governance is also more common in 6G, while blockchain and standards have comparable proportional visibility in both groups.

Figure 7-3 on the following page compares the overlapping 5G/B5G and 6G groups. Compared with task-focused 5G research, 6G places greater emphasis on native control, governance, runtime safety, standards and sovereign control [291].

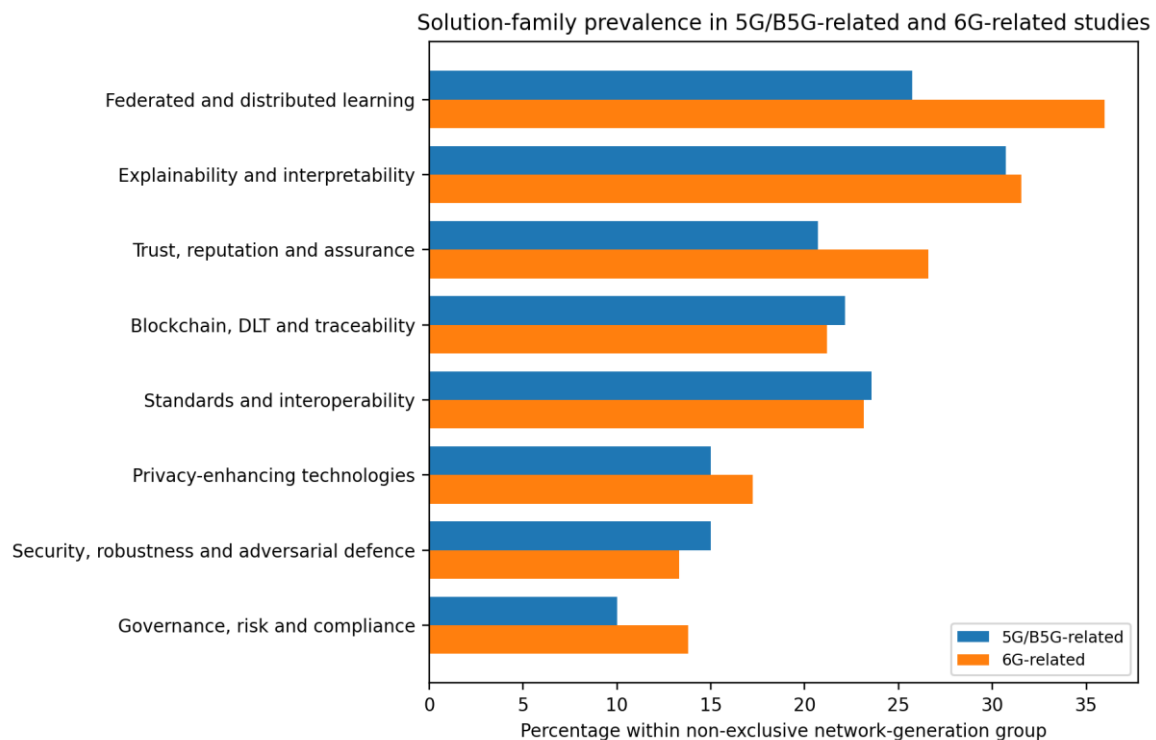


Figure 7-3. Overlapping comparison of solution-family level of occurrence in 5G/B5G-related and 6G-related studies.

Lifecycle analysis shows that trustworthy-AI proposals concentrate on monitoring, inference, deployment and training. Monitoring, auditing and verification are present in 150 studies, inference and automated decision-making in 146, deployment and orchestration in 135, and model training in 124. Incident response, adaptation and updating appear in 70, while data collection and preparation appear in 43. The distribution is summarized in Table 7-5. Comparatively low attention to data preparation suggests that data history, representativeness and consent are often treated as assumptions instead of continuously governed processes.

Evaluation maturity is presented in Figure 7-4 on the following page. Reviews or conceptual syntheses and applied empirical or case-based assessments are the two largest groups, with 84 and 85 studies respectively. Fifty-three are conceptual or design proposals, 46 report simulation or experimental evaluation, 21 describe a prototype, testbed or proof of concept, and 13 are primarily descriptive. This distribution shows that the field contains substantial applied work, but the evidence is heterogeneous and often evaluates technical metrics, while ethical outcomes receive less attention.

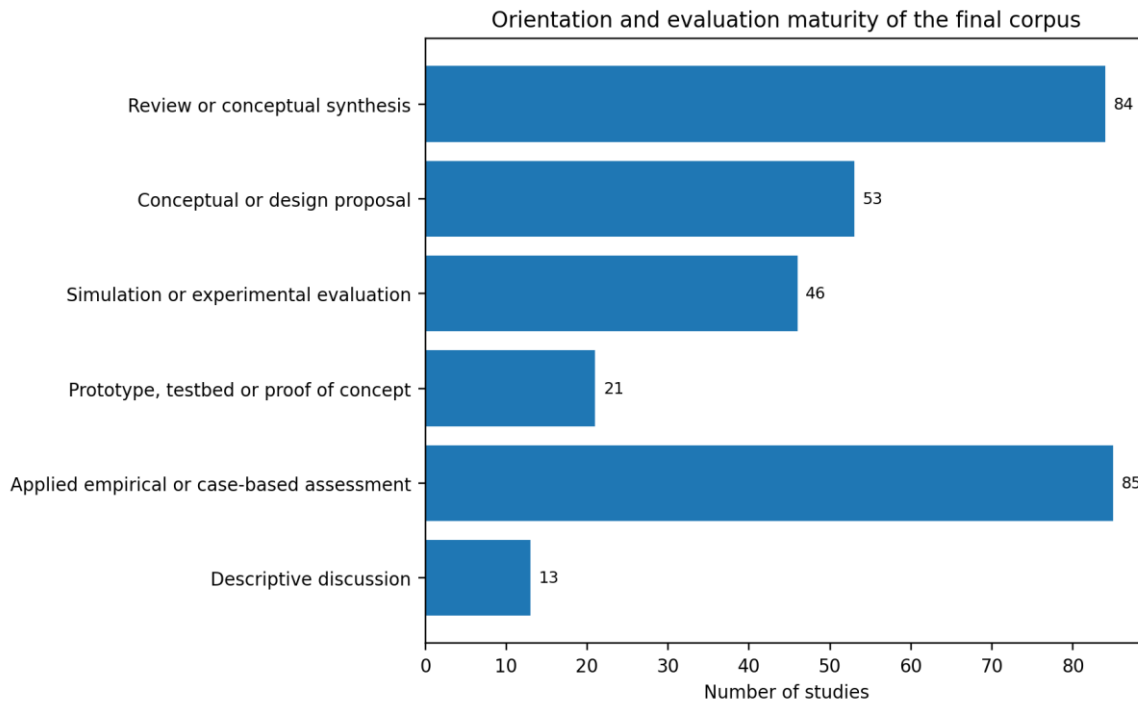


Figure 7-4. Research focus and evaluation maturity of proposed approaches.

Lifecycle stage	Studies	Common methods or evidence	Principal responsible actors
Monitoring, auditing and assurance	150	Audit logs, verification, drift monitoring, conformance and explanation review	Operator, auditor, regulator and standards body
Inference and automated decision-making	146	XAI, uncertainty estimation, causal reasoning and safety constraints	Network controller, application provider and human supervisor
Deployment, orchestration and network control	135	Zero trust, policy enforcement, runtime assurance and fail-safe control	Network operator, vendor and service orchestrator
Model training and collaborative learning	124	Federated learning, privacy accounting, robust aggregation and bias checks	Model developer, FL coordinator and participating node
Incident response, adaptation and updating	70	Incident reporting, rollback, retraining, recovery and stakeholder notification	Operator, vendor, incident team and regulator
Data collection and preparation	43	Data minimization, provenance, consent, quality and access control	Application owner, data controller and data provider

Table 7-5. Lifecycle position, methods and responsible actors in the RQ3 synthesis.

Co-occurrence provides a clearer view of integration. Figure 7-5, maps the eight leading solution families. The strongest pair is federated learning with privacy-enhancing technologies, appearing in 42 studies. Explainability and trust co-occur in 33. Federated learning appears with blockchain in 32 and with security or adversarial defense in 32. Explainability appears with federated learning in 27 and with security in 24. Governance and standards co-occur in 14. The matrix confirms that trustworthy AI is increasingly constructed through combinations, although the most developed combinations remain technical [205], [260], [266].

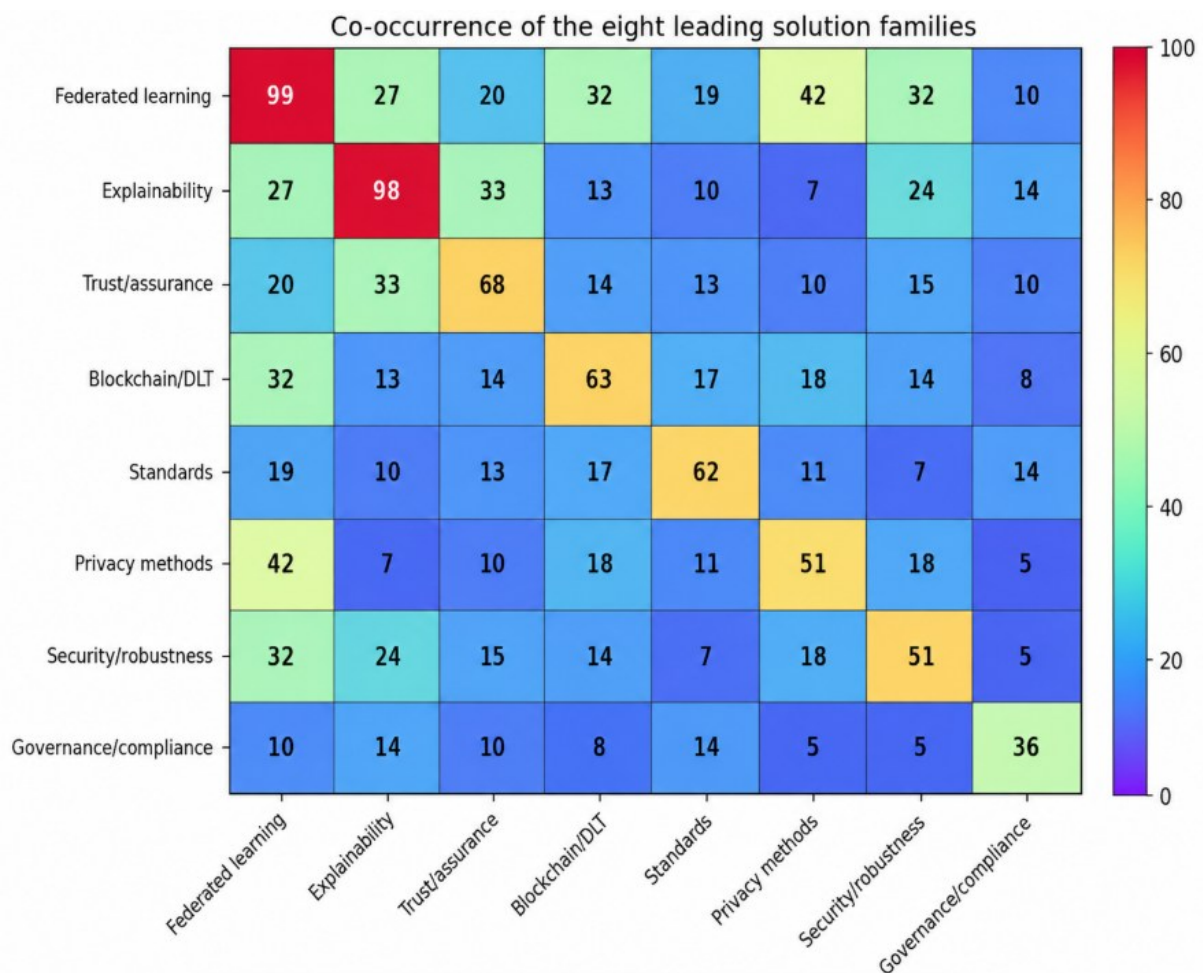


Figure 7-5. Co-occurrence of the eight leading solution families.

Stakeholder responsibility differs by lifecycle stage. Developers select models, privacy mechanisms and explanation methods. Data providers and application owners determine lawful data use and representativeness. Network operators control deployment, monitoring, access and incident response. Vendors and standards bodies define interfaces, oversight evidence and update procedures. Regulators and auditors assess compliance, while affected users require understandable information, ability to challenge decisions and remedy. A solution is incomplete when its technical component is specified but the responsible actor is not.

Table 7-5 links lifecycle stages with common methods and responsible stakeholders. The table demonstrates why a trustworthy-AI program cannot be reduced to one model test. Evidence must be accumulated from data preparation through retirement. For adaptive and agentic systems, the process must repeat after updates, changes in context or new incidents.

7.9 Integrated Interpretation and Direct Answer to RQ3

The direct answer to RQ3 is that the literature proposes a layered set of technical methods, system frameworks and standardization or governance instruments rather than one dominant trustworthy-AI solution. The most frequent technical approaches are federated and distributed learning, identified in 99 studies, and explainability or interpretability, identified in 98. These are followed by trust, reputation and assurance mechanisms in 68 studies, blockchain and traceability in 63, standards and interoperability in 62, and explicit privacy-enhancing and security or adversarial-defense methods in 51 each. Governance and compliance frameworks appear in 36 studies, while digital twins provide a test and assurance environment in 34. Human-centric design, runtime safety, causal or neuro-symbolic assurance, fairness and sustainability are visible but much less frequent [49], [64], [292].

The results support a three-layer interpretation. The first layer consists of protective and model-level methods. Federated learning, differential privacy, secure aggregation, encryption, robust aggregation, adversarial defense and explainability change how models are trained, protected and interpreted. These methods address immediate risks such as data exposure, poisoning, opacity and unstable behavior. Their strength is operational specificity and they can be implemented and measured. Their limitation is narrow scope. A privacy-preserving algorithm does not assign responsibility, and an explanation tool does not guarantee that a decision is lawful or contestable [207], [211].

The second layer consists of assurance frameworks and architectures. These combine several methods with data flows, lifecycle stages and stakeholder roles. Trustworthy B5G/6G frameworks connect input protection, explainable processing and dependable output [214]. Standards-aligned digital-twin and federated-learning frameworks combine privacy controls, blockchain logging and interoperability [58]. Runtime-safety architectures add predictive verification and fail-safe enforcement to AI-RAN control loops [120]. Sovereign-AI and neuro-symbolic knowledge-plane proposals extend assurance to model ownership, distributed reasoning and verifiable agentic action [61], [148]. This layer is necessary because the ethical performance of one component depends on the surrounding system [199], [284], [297].

The third layer consists of standards, governance and policy instruments. Telecommunications specifications define interfaces and operational roles, AI standards define risk and assurance processes, and law establishes duties, oversight and remedy. Standards-oriented work calls for alignment among 3GPP, ITU, ETSI, O-RAN, ISO/IEC, IEEE and sectoral regulation [123], [127], [165], [189], [294]. Regulatory-readiness and incident-reporting approaches address legal preparedness and accountability [16], [17]. Fairness auditing and certification demonstrate how a broad ethical principle can become a measurable conformance process [3], [218].

The co-occurrence results show that the field already recognizes the need for combinations. Federated learning co-occurs with privacy-enhancing technologies in 42 studies, with blockchain in 32 and with security or adversarial defense in 32. Explainability co-occurs with trust and assurance in 33, with federated learning in 27 and with security in 24. Governance co-occurs with standards in 14. These combinations form the beginnings of trustworthy-AI stacks, but integration is uneven. The most common stacks protect data and explain models. Fewer include fairness, human oversight, incident response and environmental constraints [161], [180], [217].

A mature implementation should therefore begin with a risk and context assessment rather than selecting a fashionable tool. For a healthcare or wearable application, the stack may require data minimization, local or federated training, differential privacy, secure aggregation, clinical safety constraints, explanation for professionals and patients, and documented human override. For autonomous RAN control, it may require robust reinforcement learning, causal explanations, runtime verification, fail-safe action, authenticated logging and operator accountability. For cross-domain digital twins, it may require common semantics, model

provenance, access control, interoperability testing and continuous compliance monitoring [27], [37], [286].

At the same time, the low frequency of explicit fairness, human-centric and sustainability mechanisms is a major imbalance. Fairness assessment and certification appear in only a small number of studies [3], [212], [218]. Human participation and inclusive adoption are discussed more often as goals than as implemented processes [65], [171]. Sustainability is prominent in broad 6G visions but rarely operationalized as an ethical constraint on AI models and infrastructure [33]. This imbalance means that literature supports trustworthy operation mainly through technical security, privacy and transparency, while distributional justice and social control remain less developed [195], [259], [288].

In conclusion, RQ3 is answered by a portfolio of mutually dependent interventions. Federated learning and privacy technologies reduce centralized exposure, XAI and causal methods support understanding, robust learning, zero trust and runtime safety constrain harmful behavior, blockchain and provenance preserve evidence, trust and assurance frameworks integrate controls, standards create common requirements and governance assigns responsibility across the lifecycle. No single intervention is sufficient. Ethically aligned AI in 5G/6G requires a layered architecture in which technical mechanisms produce evidence, assurance frameworks evaluate that evidence and standards, or governance processes determine action and accountability [255], [263], [296].

7.10 Limitations of the RQ3 Analysis

The first limitation is that coding was based on titles, abstracts and author keywords rather than complete full-text extraction. Abstracts usually describe the main technical contribution, but they may omit secondary governance processes, implementation details, evaluation metrics or the exact standard used. A paper may present a substantial privacy or accountability mechanism in the full text without naming it in the abstract. Conversely, an abstract may describe a framework as trustworthy without providing sufficient operational evidence. The counts therefore represent visible solution claims in the available metadata rather than a definitive inventory of every mechanism.

The second limitation concerns terminology. Method, model, framework, architecture, system and platform are used inconsistently. Some authors call a single classifier a framework, while others reserve the term for multi-component governance structures. The analysis classified

proposals by function, but borderline cases remain. The same problem affects standards. An article may say that a design is standards-aligned without demonstrating conformance or may cite 3GPP and ITU only as background. Explicit mentions were counted conservatively, yet the analysis cannot confirm whether a proposal implements a binding specification, adapts voluntary guidance or merely discusses future standardization.

Third, keyword frequency does not measure effectiveness. Federated learning and explainability are frequent because they are established research areas with accessible experimental methods. Their prevalence does not prove that they offer the strongest ethical protection. Differential privacy may have a formal guarantee but impose unacceptable utility loss. An explanation method may be technically faithful but unusable for the affected stakeholder. Blockchain may preserve evidence but create governance, scalability or data-protection problems. Full evaluation would require study-specific analysis of assumptions, metrics, threat models, datasets, baselines and limitations [19], [102], [119].

Fourth, the coding is multi-label and categories overlap. Privacy-enhancing technologies often appear inside federated-learning frameworks, while blockchain may be used for security, trust, governance or interoperability. Explainability is both a method and an assurance mechanism. Standards can define architecture, and digital twins can function as application platforms or validation environments. Co-occurrence captures integration but cannot determine which component is central or whether the combination is coherent. The final synthesis therefore treats co-occurrence as evidence of integration without assuming which component is central when that role is not explicit in the available metadata [162], [271], [274].

Seventh, the study period ends in June 2026. The final year is partial, and publication or indexing delays may change the apparent trend. The rapid emergence of agentic AI, large telecom models, sovereign AI and AI-RAN safety means that categories are evolving quickly. New standards and regulatory guidance may also alter how researchers describe compliance. Temporal results should therefore be interpreted as the state of the corpus at the search cut-off, not as a complete account of 2026 [198].

Eighth, the corpus is shaped by database coverage and screening criteria. Engineering databases and technical publishers contribute a large proportion of the studies, which favors algorithmic methods, communication architectures and measurable performance. Legal scholarships, organizational governance research and participatory design may use

terminology that does not explicitly mention AI-enabled 5G/6G and may therefore be underrepresented. This helps explain the high visibility of federated learning, XAI and security compared with fairness, labor impact, social inclusion and democratic oversight.

Tenth, the analysis does not calculate comparative effect sizes or rank individual methods. The studies address different network functions, datasets, threats and objectives. Accuracy, latency, privacy budget, explanation fidelity, robustness and compliance are not directly comparable across the corpus. A method that is suitable for intrusion detection may be inappropriate for health monitoring or autonomous control. RQ3 therefore answers what categories of solutions are proposed and how they are combined, not which single solution is universally superior [196], [246].

These limitations define the scope of the final RQ3 synthesis rather than reducing its analytical value. The chapter reports the methods, frameworks and standards visible in the final abstract-level corpus and interprets their maturity and effectiveness cautiously. The resulting evidence matrix is used directly in Chapter 8 to connect ethical risks, proposed responses and remaining research gaps.

Chapter 8 Research Question 4: Remaining Research Gaps for Trustworthy, Accountable and Human-Centric AI in 5G/6G

This chapter addresses the fourth research question. What research gaps remain regarding trustworthy, accountable, and human-centric AI in 5G/6G? The analysis uses the same final corpus of 302 studies examined in Chapters 5, 6 and 7 and covers publications from January 2020 to June 2026. RQ4 is different from the preceding questions because it does not seek only to describe what the literature contains. It must identify what is still absent, weakly tested, fragmented, or insufficiently connected to operational telecommunications practice.

The answer is developed through two complementary forms of evidence. The first consists of gaps explicitly signaled by the authors through terms such as challenge, limitation, open issue, future direction, lack, unresolved problem, or need for further work. The second consists of review-inferred gaps produced by comparing the status of the field in Chapter 5, the risk structure in Chapter 6, and the available methods, frameworks and standards in Chapter 7. This distinction is necessary because a low publication count alone does not prove that a research gap exists. A gap becomes more credible when limited coverage is combined with a substantial ethical risk, weak empirical validation, incomplete standards, or an absence of responsible stakeholders and lifecycle controls.

The chapter therefore treats research gaps as evidence-based mismatches. Some mismatches occur between high-risk applications and the maturity of available safeguards. Others occur between technical methods and governance arrangements, between general AI principles and telecommunications-specific requirements, or between the rapid growth of 6G research and the limited availability of real operational evidence. The chapter first explains the analytical procedure and gap taxonomy. It then presents the distribution of reported gaps, examines empirical, governance, human-centered and operational weaknesses, and concludes with an integrated research agenda and the limitations of the RQ4 analysis.

8.1 Purpose and Analytical Procedure

RQ4 was analyzed after RQ1, RQ2 and RQ3 because the identification of gaps depends on an understanding of the field. Chapter 5 showed that literature is expanding rapidly and is increasingly oriented toward 6G, but that much of it is recent. Chapter 6 showed that cybersecurity, privacy, reliability, trust, governance and opacity form the dominant risk structure. Chapter 7 showed that federated learning and explainability are the most frequent technical responses, while explicit fairness, sustainability, runtime safety and human-centered design are much less common. These findings create the basis for asking whether the available responses are proportionate to the risks and whether they have been validated in sufficiently realistic settings [205], [227], [261].

The unit of analysis remained unique, including study. Coding used the title, abstract, keywords and screening justification stored in the final Master Excel. A study was coded as containing an explicit gap signal when a challenge-related expression occurred in the same sentence as a gap category. This sentence-level rule was used to reduce the risk of counting an article merely because it mentioned a challenge in one part of the abstract and an unrelated ethical term elsewhere. Coding was multi-label because one study could report several unresolved issues, such as privacy, scalability, interoperability and accountability.

8.2 Framework for Identifying and Classifying Research Gaps

Fourteen gap categories were retained. They cover 6G-specific architectural integration, security and adversarial resilience, privacy and data protection, scalability and resource efficiency, safety and runtime assurance, sustainability, governance and accountability, explainability, standards and interoperability, empirical validation, fairness and inclusion, human-centered participation, data and benchmarking, and cross-domain generalization. The categories were selected because they recur across the corpus and because they connect directly with trustworthy, accountable or human-centric AI. The operational definitions are presented in Table 8-1 [111], [150], [196].

Gap category	Operational coding focus	Primary RQ4 dimension
6G architecture and AI-native control	Challenges involving AI-native RAN, slicing, semantic communication, zero-touch or agentic control	Trustworthy and accountable
Security and adversarial resilience	Attacks, poisoning, intrusion, malicious clients, zero trust and defense limitations	Trustworthy
Privacy and data protection	Leakage, consent, confidentiality, model-update privacy and cross-border data use	Trustworthy and human-centric
Scalability and resource efficiency	Latency, overhead, complexity, energy and deployment-scale trade-offs	Trustworthy
Safety and runtime assurance	Reliability, robustness, uncertainty, fail-safe action, verification and rollback	Trustworthy and accountable
Sustainability and environmental assessment	Energy use, computational cost, carbon footprint, rebound effects and lifecycle resource burden	Trustworthy and human-centric
Governance and accountability	Regulation, audit, liability, responsibility, incident reporting and compliance	Accountable
Explainability and transparency	Opaque decisions, limited interpretability, explanation quality, operator understanding and contestability	Trustworthy and accountable
Standards and interoperability	Common interfaces, certification, profiles, metrics and cross-vendor evidence	Accountable
Empirical validation and real-world deployment	Lack of representative testbeds, field trials, multi-operator evidence and long-duration deployment validation	Trustworthy and accountable
Fairness and inclusion	Bias, non-discrimination, digital divide, unequal access and distribution of benefits	Human-centric
Human-centered design and stakeholder participation	User and operator involvement, acceptance, meaningful oversight, contestability and redress	Human-centric
Data and benchmarking	Representative datasets, provenance, non-IID data, common scenarios and reproducibility	Trustworthy
Generalization and transferability	Cross-operator, cross-domain, cross-jurisdiction and distribution-shift evidence	All three dimensions

Table 8-1. Operational category structure used for the RQ4 gap coding.

A second analytical layer assessed evaluation maturity. Each study was classified as conceptual, survey or architecture-level simulation, or dataset-based, or prototype, testbed or real-world validation. This classification does not measure scientific quality. It indicates how close the reported evidence is to operational deployment. A technically sophisticated simulation can provide valuable evidence, but it cannot fully reproduce the organizational, regulatory, behavioral and adversarial conditions of a live 5G/6G environment. The maturity layer was therefore used only to identify where repeated conceptual proposals are not yet matched by realistic evaluation.

The final synthesis uses cautious priority logic. Explicit gap prevalence was combined with the share of conceptual evidence in each category. High prevalence with a large validation deficit was interpreted as a strong research priority. Low prevalence was interpreted more carefully. It may indicate a neglected issue, but it may also reflect terminology differences or the limited information contained in abstracts. For this reason, fairness, stakeholder participation and cross-domain generalization are discussed as underrepresented but important areas rather than automatically ranked above more frequently documented operational problems.

The first distinction is between an author-reported gap and a review-inferred gap. An author-reported gap is directly visible in the source record. Examples include a stated lack of common standards, limited datasets, insufficient explainability, untested scalability, or a need for future regulation. A review-inferred gap is identified by comparing several parts of the SLR. For example, privacy appears to be a major risk in Chapter 6 and federated learning appears as a common response in Chapter 7. A remaining gap is inferred when the literature gives limited attention to leakage from model updates, realistic adversaries, user consent, cross-border data governance, or deployment accountability despite the frequency of privacy-preserving claims [210].

A second distinction concerns the level at which the gap occurs. Technical gaps relate to model robustness, explainability, data quality, communication overhead, security or runtime control. Socio-technical gaps arise when a technically effective system lacks stakeholder participation, understandable recourse, organizational responsibility or assessment of unequal effects. Governance gaps concern legal readiness, auditability, incident reporting, liability, procurement, certification and coordination between telecommunications and AI standards. These levels are interdependent. Technical control may fail in practice when no organization is responsible for monitoring it, while a governance requirement may remain symbolic when it cannot be translated into measurable network or model behavior [133].

The third distinction concerns lifecycle location. Data-stage gaps include consent, representativeness, provenance, non-IID distributions and secure sharing. Model-stage gaps include bias mitigation, uncertainty, explainability, adversarial learning and privacy leakage. Network-stage gaps include resource allocation, slicing, O-RAN control, semantic communication and cross-domain orchestration. Deployment-stage gaps include human oversight, monitoring, rollback, incident reporting and redress. The lifecycle perspective prevents the review from treating ethical alignment as a property that can be added after model

Evidence type	Definition	How it is used in RQ4	Main caution
Author-reported gap	A challenge, limitation, open issue or future need stated in the source record.	Supports frequency counts and category mapping.	Abstracts may omit detailed limitations.
Review-inferred coverage gap	A major risk or lifecycle stage receives limited solution coverage.	Compares Chapter 6 risk frequencies with Chapter 7 solution frequencies.	Categories are related but not perfectly equivalent.
Review-inferred validation gap	A proposed safeguard is mainly conceptual or simulation based.	Uses maturity coding and the priority matrix.	Maturity indicators are abstract-level proxies.
Review-inferred accountability gap	A technical action lacks a clearly responsible actor, audit process or redress path.	Connects methods with governance and stakeholder roles.	Responsibility may be described only in the full text.
Review-inferred human-centered gap	User effects, participation, fairness or acceptance are weakly represented.	Identifies underdeveloped socio-technical evidence.	Low frequency is not by itself proof of low importance.

Table 8-2. Evidence types used to distinguish reported and review-inferred gaps.

training. It also reveals that many proposed safeguards operate at one stage while risks propagate across several stages [29], [84].

This framework supports a structured rather than impressionistic answer. It prevents the chapter from presenting a long list of future-work statements without explaining their relationships. It also allows explicit statements from surveys on 6G, healthcare, metaverse systems, federated learning and explainable networking to relate to narrower technical studies and regulatory work [14], [18], [19], [21], [55]. Additional studies in the corpus address the same point [85], [232], [244], [267]. The result is a gap map that links unresolved problems to their technical, organizational and human consequences.

8.3 Overall Distribution and Temporal Development of Research Gaps

The abstract-level coding identified fourteen recurring gap categories. The most frequent category concerned 6G-specific architecture and AI-native control, with 103 studies containing an explicit challenge signal. Security and adversarial resilience followed with 93 studies, privacy and data protection with 77, and scalability, latency or resource efficiency with 72. Safety, reliability and runtime assurance appeared in 67 studies, while sustainability and environmental assessment appeared in 59. These values are multi-label and therefore do not sum to 302 [192], [202], [239].

Governance, regulation and accountability were explicitly connected with unresolved issues in 51 studies. Explainability and transparency appeared in 44, standards and interoperability in 38, and empirical validation or real-world deployment in 33. The least frequent categories were

fairness and inclusion with 17 studies, human-centered design and stakeholder participation with 16, data and benchmarking with 13, and cross-domain generalization with

10. The frequency ranking is presented in Figure 8-1 on the following page. The lower courts should not be read as evidence that these issues are unimportant. In several cases, they indicate that literature is dominated by network and model performance while social evaluation is less developed.

The ranking shows that the field still frames many ethical problems through engineering constraints. Security, privacy, robustness, latency and scalability receive substantial attention because they can be connected to measurable system behavior. Governance and human-centered concerns are less easily reduced to one technical metric and are therefore more fragmented. The imbalance is visible when Chapter 6 and Chapter 7 are compared. Governance-related risks appeared in 134 studies, but governance and compliance solutions appeared in only 36. Human-centered solutions appeared in 15 studies, and explicit fairness mechanisms in four. This mismatch supports a review-inferred gap in translating broad ethical concerns into operational controls and responsibilities [79], [245].

The temporal pattern is also important. Most reported gaps have increased sharply in 2025 and the first half of 2026 because the corpus itself is concentrated in these years. The annual development of six recurrent categories is shown in Figure 8-2 on the following page. Security, privacy and scalability grow with the expansion of AI-native networking, federated learning, semantic communication and autonomous control. Governance and standards also rise, but their growth occurs later and remains smaller in absolute terms. The 2026 values cover only January to June and must not be interpreted as a complete annual total.

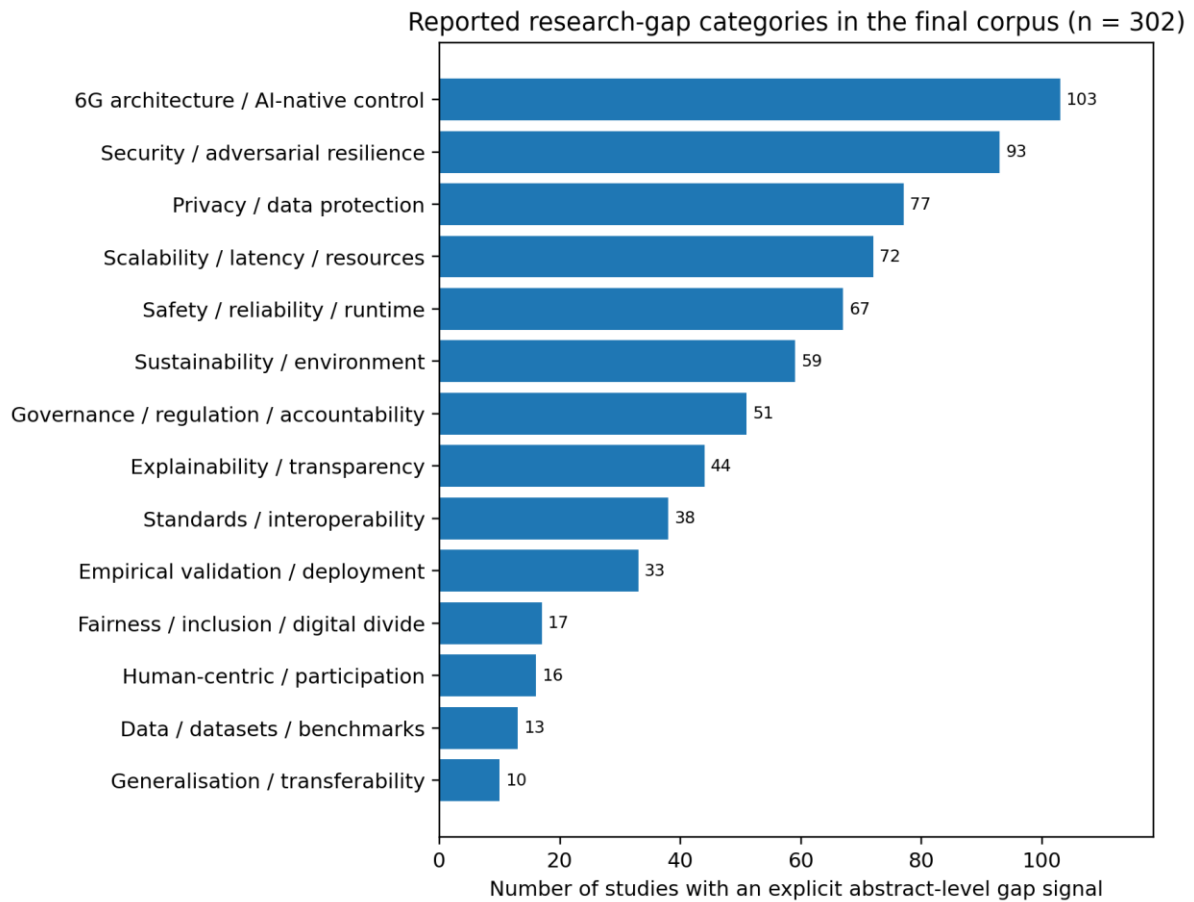


Figure 8-1. Frequency of explicit research-gap categories in the 302 included studies.

The overall status is therefore not a simple movement from problems to solutions. The literature is creating a more detailed understanding of the problem space, while the technical system is simultaneously becoming more complex. RQ4 captures this moving boundary. Some gaps may narrow as standards and testbeds develop, but new gaps emerge when 6G architectures distribute intelligence across devices, edges, clouds, network controllers and autonomous agents. The main implication is that future research must evaluate combinations of safeguards rather than isolated methods.

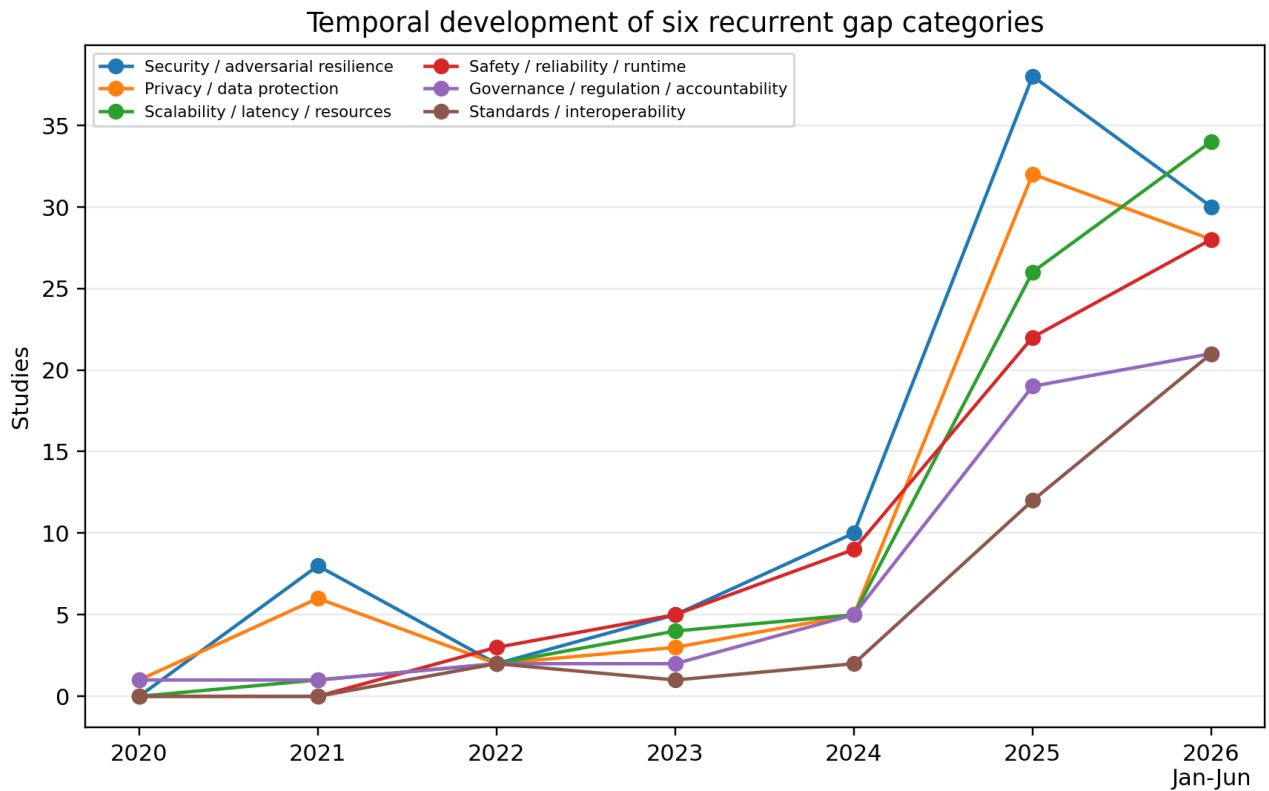


Figure 8-2. Temporal development of six recurrent gap categories. The 2026 values cover January to June.

8.4 Empirical Validation, Datasets and Benchmarking Gaps

The strongest cross-cutting weakness is the limited amount of operational evidence. Of the 302 studies, 168 were classified as conceptual, survey or architecture-level, 113 contained simulation or dataset-based evaluation, and only 21 showed clear indicators of prototype, testbed or real-world validation. Figure 8-4, presented on the following page, summarizes this distribution. The result does not mean that 168 studies are weak. Surveys and architecture are necessary in an emerging field. However, it shows that claims about trustworthiness frequently precede evidence from realistic multi-operator, multi-domain or long-duration settings.

Gap category	Studies	Corpus %	Interpretive meaning
6G architecture / AI-native control	103	34.1%	Rapid architectural change creates unresolved authority, integration and assurance questions.
Security / adversarial resilience	93	30.8%	Known defenses remain incomplete against adaptive and cross-layer attacks.
Privacy / data protection	77	25.5%	Distributed learning reduces centralization but does not remove leakage, consent or governance risk.
Scalability / latency / resources	72	23.8%	Safeguards must remain feasible under network timing and resource constraints.
Safety / reliability / runtime	67	22.2%	Static accuracy does not establish safe behavior during failure or change.
Sustainability / environment	59	19.5%	Energy and lifecycle effects are recognized but seldom integrated with assurance.
Governance / regulation / accountability	51	16.9%	Responsibility and compliance mechanisms lag risk recognition.
Explainability / transparency	44	14.6%	Explanation methods require operational, human and lifecycle validation.
Standards / interoperability	38	12.6%	Common metrics, interfaces and assurance artifacts remain incomplete.
Empirical validation / deployment	33	10.9%	Many claims have not progressed to representative testbeds or field evidence.
Fairness / inclusion / digital divide	17	5.6%	Telecommunications-specific fairness and access effects are underdeveloped.
Human-centric / participation	16	5.3%	Affected users and operators are rarely included in requirements and evaluation.
Data / datasets / benchmarks	13	4.3%	Shared and ethically governed benchmarks are limited.
Generalization / transferability	10	3.3%	Evidence across operators, domains and jurisdictions remains scarce.

Table 8-3. Frequency and interpretation of the reported RQ4 gap categories.

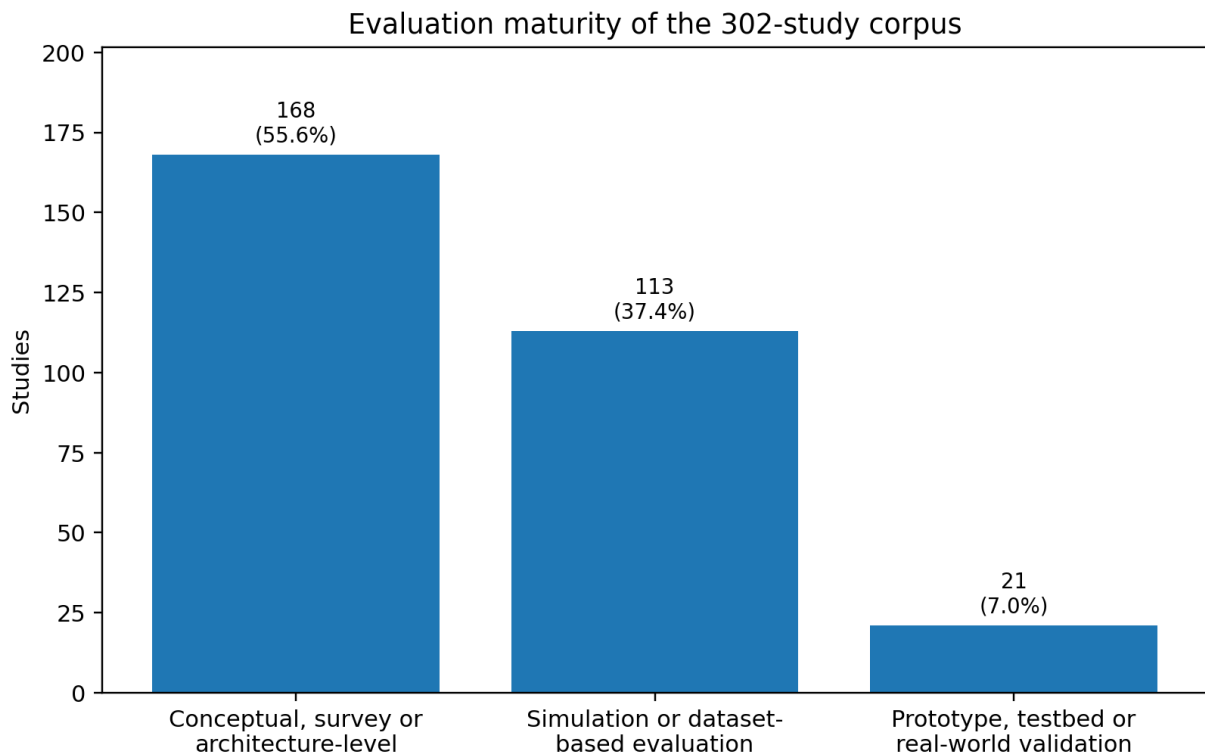


Figure 8-3. Evaluation maturity of the final 302-study review sample.

The gap is especially visible in explainable and autonomous network control. Many studies propose interpretable channel estimation, network slicing, intrusion detection or resource orchestration [6], [8], [39], [78], [94]. Additional studies in the corpus address the same points [125], [131], [132], [158]. Evaluation commonly focuses on prediction accuracy, throughput, latency or an explanation score. Less attention is given to whether operators understand the explanation under time pressure, whether explanations remain stable after distribution shift, whether an explanation supports accountability, or whether a control action can be safely reversed. Future experiments should include operational users, failure scenarios and decision-quality outcomes rather than treating explanation generation as the final objective.

Priority area	Gap evidence	Recommended next research step	Expected assurance output
Realistic validation	168 conceptual/survey studies and only 21 with clear prototype, testbed or real-world indicators.	Create shared validation ladders and multi-operator testbeds.	Reproducible evidence linked to specific ethical claims.
Lifecycle accountability	Governance risks are frequent, but operational governance solutions are fewer.	Map responsibility, logging, incident reporting, redress and recertification.	End-to-end accountability and audit trail.
Human-centered assurance	Few studies test participation, acceptance, fairness or meaningful oversight.	Use mixed methods and domain-specific stakeholder evaluation.	Evidence of usability, equity and appropriate reliance.
Assurance interoperability	Methods and standards are fragmented across layers and vendors.	Define common assurance artifacts and interfaces.	Portable evidence across models, networks and organizations.
Continuous and agentic assurance	Adaptive systems change after initial validation.	Develop runtime constraints, change triggers, rollback and update provenance.	Assurance that remains valid during adaptation.
Sustainable and inclusive 6G	Environmental and distributional effects are rarely assessed together.	Combine lifecycle energy, access, affordability and benefit metrics.	Socio-environmental assurance case.

Table 8-4. Priority research areas derived from the risk-solution-evidence comparison.

Dataset limitations form a related gap. Security and privacy studies often rely on established intrusion-detection datasets or synthetically generated traffic. These resources support comparison, but they may not reflect encrypted traffic, changing attacker behavior, novel 6G protocols, heterogeneous edge devices or operator-specific configurations. Federated systems add non-IID data, unequal client participation and variable hardware. Studies addressing data heterogeneity, differential privacy and federated intrusion detection demonstrate progress [50], [71], [103], [143], [159], [177], but the corpus contains limited evidence of shared, representative and ethically governed benchmarks for AI-native 5G/6G [248], [271], [278].

A benchmark gap is not solved simply by publishing more data. Telecommunications datasets may contain personal, commercial or security-sensitive information. Future benchmark design must therefore include provenance, access controls, privacy risk, documentation of affected groups, task definitions and reproducible evaluation protocols. Synthetic data and digital twins may help, but they introduce model assumptions that need validation against physical networks. Standardized scenario descriptions would also allow researchers to compare results across different simulators, traffic models and network topologies. Surveys of KPIs, tools and AI models provide a starting point, but common trustworthy-AI benchmarks remain incomplete [111], [294], [224].

Longitudinal evaluation is also rare. Trustworthy behavior must persist after software updates, network reconfiguration, new devices, changing traffic and adversarial adaptation. A model that is robust in a static experiment may become unsafe after concept drift. This issue is particularly important for zero-touch management, O-RAN control loops and self-learning security systems [104], [120], [155], [189]. Future work should report monitoring periods, drift detection, rollback procedures, incident thresholds and post-deployment update rules. Without these elements, evaluation measures the initial model rather than the continuing socio-technical service.

The immediate research need is a shared validation ladder. Early work can use simulation, but later stages should progress through reproducible datasets, hardware-in-the-loop evaluation, federated testbeds, operator trials and monitored deployment. Each stage should specify what ethical claim is being tested. Privacy should be assessed through leakage and attack models, fairness through subgroup outcomes, explainability through user comprehension and decision support, accountability through traceability and incident reconstruction, and safety through constrained failure experiments. This would convert trustworthy-AI language into falsifiable evidence [26], [193], [234].

8.5 Governance, Accountability, Standardization and Compliance Gaps

Governance is one of the clearest mismatches between risk recognition and solution maturity. Chapter 6 identified governance, legal and regulatory inadequacy in 134 studies, whereas Chapter 7 identified governance and compliance mechanisms in 36. In the RQ4 coding, 51 studies explicitly connected governance, regulation or accountability with unresolved issues, and 38 did the same for standards and interoperability. These results show that the literature recognizes institutional problems but provides fewer tested mechanisms for assigning responsibility and demonstrating compliance.

The first gap concerns fragmented regulatory readiness. Telecommunications networks are governed through technical standards, spectrum rules, cybersecurity duties, consumer law, sector regulation and contractual relationships. AI regulation adds requirements related to risk management, transparency, data governance and human oversight. The interaction between these regimes is not yet consistently mapped for AI-native networks. Work on regulatory readiness and AI incident reporting shows the importance of this interface [16], [17], while studies connecting 6G applications with the EU AI Act illustrate domain-specific implications

[7], [127], [165]. Future research must clarify how obligations are distributed among equipment vendors, model providers, operators, service providers, data controllers and end users [230].

A second gap concerns accountability across distributed control. In federated, edge and O-RAN systems, no single actor necessarily observes the entire decision chain. A harmful allocation or security action may result from local data, a shared model, an orchestration policy, a third-party application and an automated controller. Traceability mechanisms, blockchain records and zero-touch assurance can preserve events [40], [89], [189], [221], but an event log does not itself determine responsibility. Research is needed on causal attribution, evidentiary quality, authority boundaries and procedures for contesting or reversing automated actions [201], [209], [260].

Incident reporting is a practical part of accountability that remains underdeveloped. Telecommunications outages and cybersecurity incidents already have reporting processes, but AI incidents may involve silent performance degradation, discriminatory service, unsafe optimization or explanation failure. These events do not always fit existing categories. The literature should develop common incident taxonomies, severity levels and minimum evidence requirements. It should also examine when reporting must cross organizational or national boundaries and how commercially sensitive network information can be protected while still supporting oversight [17].

Standardization presents a related challenge. The corpus contains work on AI-native architectures, semantic communication, privacy standards, wearable-health profiles and fairness certification [3], [58], [95], [123], [218]. However, many proposals remain domain-specific or are not connected to a complete lifecycle. Future standards should define measurable properties, interfaces and assurance artefacts rather than only high-level principles. A privacy standard, for example, should clarify threat models, leakage measures, data-retention rules and acceptable trade-offs. An explainability standard should distinguish developer debugging, operator decision support, regulatory evidence and user communication because these purposes require different explanations [20], [162], [252].

Interoperability is both technical and institutional. Models and assurance evidence may move across vendors, operators, cloud providers and jurisdictions. Common interfaces are needed for model cards, data documentation, audit logs, uncertainty estimates, explanation metadata and policy constraints. Standards-driven architectures can support this exchange, but they must

avoid creating a compliance checklist that hides unresolved risk. Independent testing, version control and change management are necessary so that certification remains valid after model or network updates.

The governance gap therefore cannot be closed by one regulation or one standard. It requires an assurance ecosystem that connects technical testing, organizational responsibility, incident reporting, independent audit and user redress. Research should evaluate this ecosystem through realistic case studies rather than discussing each part in isolation.

8.6 Human-Centric Design, Fairness, Inclusion and Stakeholder Participation Gaps

Human-centric AI is the least developed major dimension of the review. Only 16 studies contained an explicit abstract-level gap signal related to human-centered design or stakeholder participation, and 17 contained one related to fairness, inclusion or the digital divide. Chapter 7 found 15 human-centered solution studies and only four explicit fairness mechanisms. These low counts are important because 5G/6G systems affect access to healthcare, transport, education, employment, public services and communication. A technically trustworthy network can still produce unequal or unacceptable outcomes when user needs and social conditions are excluded from design [69].

The first gap concerns participation. Most technical studies define objectives such as throughput, latency, energy use or detection accuracy before the affected stakeholders are considered. Future research should involve operators, engineers, regulators, application providers and users when defining acceptable trade-offs. In healthcare, connected vehicles and public-safety systems, affected groups may have different priorities regarding privacy, speed, explanation and human control [7], [18], [107], [137]. Participatory methods can identify harms that are not visible in performance metrics, including loss of dignity, surveillance concerns, exclusion or reduced ability to challenge automated decisions.

Technology acceptance is also understudied. One empirical study directly examines acceptance of AI and 5G systems [10], while market-driven and user-experience frameworks address trust, personalization and inclusion [65]. These studies remain exceptions. Trust should not be treated only as a technical reputation score. Human trust depends on experience, perceived fairness, transparency, institutional legitimacy and the ability to obtain help when a service fails. Future

work should distinguish appropriate reliance from blind trust and should test how explanations, warnings and human override affect behavior.

Fairness requires telecommunications-specific definitions. General AI fairness metrics do not automatically translate to network allocation. Relevant groups may be defined by location, income, disability, service type, device capability, mobility or exposure to poor coverage. Fair packet scheduling and fairness auditing provide useful directions [3], [46], [212], [218], but future studies should connect statistical fairness with service quality and real social impact. They should also examine whether improvements for one group shift costs to another group or reduce network resilience.

The digital divide is likely to change rather than disappear in 6G. Advanced services may benefit regions and users with modern devices, dense infrastructure and high-quality data, while rural or resource-constrained settings remain less represented. Work on sustainable and human-centered 6G recognizes this concern [2], [171], and application reviews highlight opportunities and barriers in smart homes, education and healthcare [107], [134], [254]. Future research should compare deployment scenarios, affordability, accessibility, digital literacy and language support. Inclusion should be assessed as a system requirement rather than a later social benefit [73], [247].

Human oversight is another unresolved area. Many papers call for human-in-the-loop or human-centric control, but they rarely specify who the human is, what information is available, how quickly intervention must occur, or what happens when the operator disagrees with an AI recommendation. In high-speed networks, continuous manual approval is impossible. Research is needed on meaningful oversight models, escalation rules, supervisory interfaces and workload. Explanations must be matched to the user's role such as a radio engineer, security analyst, regulator and consumer require different levels of detail.

The human-centric research gap is therefore methodological as well as thematic. The field needs mixed-method studies, controlled user experiments, interviews, participatory design and social-impact assessment alongside simulations. These methods would not replace engineering evaluation. They would test whether technical safeguards are understandable, usable and socially acceptable.

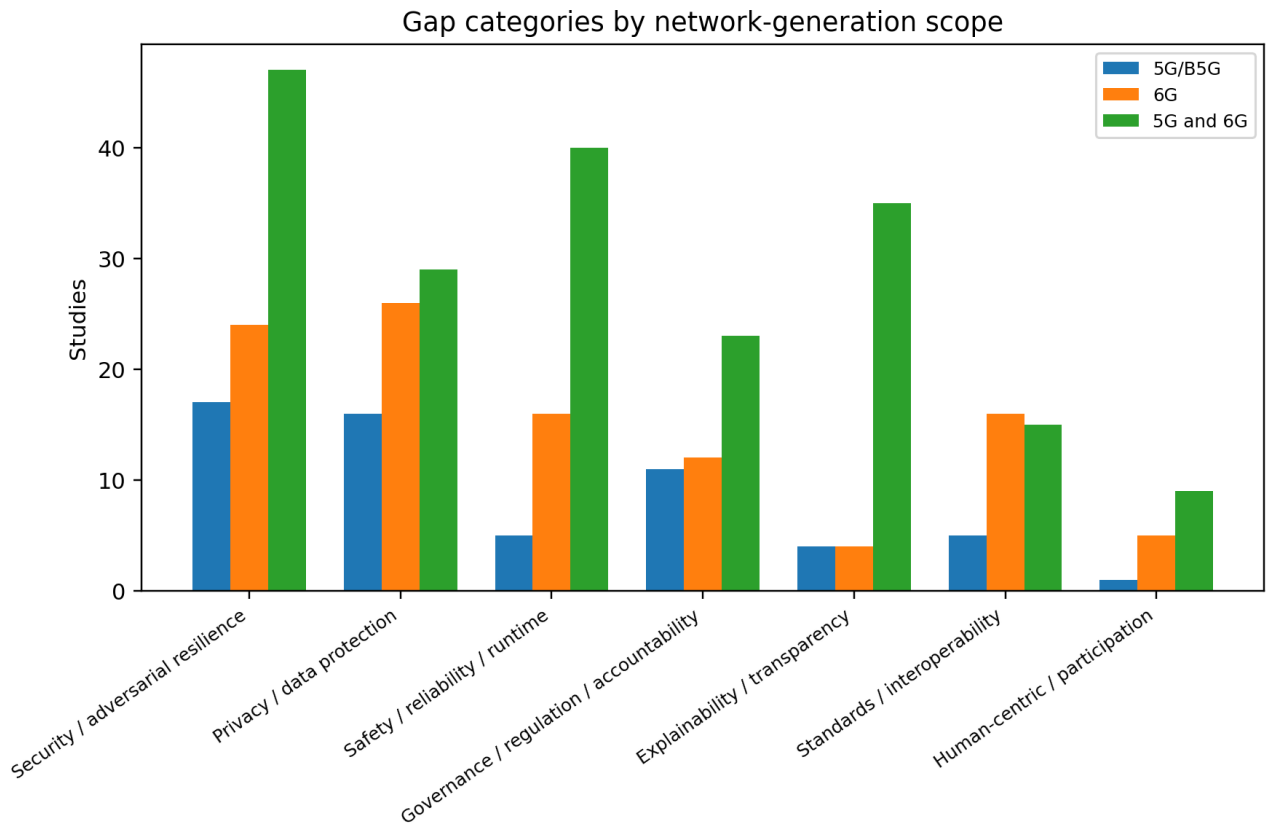


Figure 8-4. Distribution of selected gap categories by network-generation scope.

The network-generation distribution of selected gap categories is presented in Figure 8-3 on the following page. The overlapping 5G-and-6G group is shown separately because many studies discuss the transition between generations rather than one network in isolation.

8.7 Deployment, Security, Safety and Operational Verification Gaps

Security and adversarial resilience form the second most frequent reported gap category, appearing in 93 studies. Safety, reliability and runtime assurance appear in 67, while scalability and resource efficiency appear in 72. These categories are closely related. A defense that adds excessive latency may be unusable in mission-critical communication, and an efficient autonomous controller may become unsafe when attacked or exposed to unfamiliar conditions. Future research must therefore evaluate joint security, safety and performance properties [256].

Adversarial evaluation remains incomplete. The corpus includes intrusion detection, poisoning defense, privacy-preserving federated learning, zero-trust networking and robust orchestration [1], [11], [39], [67], [77]. Additional studies in the corpus address the same point [102], [103],

[179], [293]. Many evaluations use known attack types and static datasets. AI-native 6G introduces adaptive attackers who can manipulate data, model updates, intent messages, explanations or agent interactions. Research should include compound attacks, attacker adaptation, insider threats and attacks that exploit the coordination between network layers. Red-team exercises and shared threat models would improve comparability [66], [211], [242].

Runtime assurance is an emerging but small solution family. Work on fail-safe O-RAN control, trustworthy orchestration and safe explainable systems demonstrates possible architectures [5], [75], [120], [155]. The gap is to define enforceable safety envelopes for systems that learn and act at different timescales. A runtime monitor must decide when to block, degrade or reverse an AI action without creating instability. Future work should test these controls during congestion, handover, cyberattack, component failure and conflicting intents. It should also report false alarms and the cost of unnecessary intervention.

Uncertainty remains uneven. Accuracy alone does not show whether a model knows when it is outside its competence. Trustworthy channel estimation and resource management studies increasingly include uncertainty or robust learning [77], [96], [131], but calibrated confidence, out-of-distribution detection and uncertainty propagation are not consistently evaluated across the corpus. This is a major gap for autonomous decisions because uncertainty at one component can spread through slicing, routing or orchestration. Future systems should communicate uncertainty in machine-readable form and connect it to safe fallback behavior.

Scalability creates another deployment gap. Federated learning, explainability, blockchain and audit logging all consume communication, computation and storage resources. The costs are often evaluated separately from the ethical benefit. Stronger privacy can reduce accuracy, detailed explanations can increase latency, and distributed ledgers can increase energy use. Studies on privacy-preserving learning and explainable slicing identify these trade-offs [50], [71], [154], [159], [213]. Additional studies in the corpus address the same point [265]. Future research should use multi-objective evaluation and report the conditions under which a safeguard becomes impractical [27], [135], [207].

Operational monitoring is also fragmented. Network assurance traditionally focuses on availability and performance, while AI assurance requires drift, bias, uncertainty, explanation and data-quality indicators. These signals should be integrated without overwhelming operators. Research is needed on dashboards, alert prioritization, causal diagnosis and

automated evidence collection. Digital twins can support controlled testing and predictive maintenance [112], [121], [146], [168], [294], but their fidelity and update processes must be assessed.

The operational assurance gap is ultimately an integration problem. Security testing, safety constraints, performance engineering and governance evidence must be combined into one deployment process. Research that evaluates only one property under ideal conditions cannot establish that an AI-enabled network is trustworthy in practice.

8.8 6G-Specific, Cross-Domain and Sustainability Research Gaps

The corpus is increasingly 6G-oriented, and 203 studies contain an explicit 6G dimension. This focus is anticipatory because 6G standards, architectures and business models are still developing. The RQ4 coding identified 103 studies that connected 6G architecture or AI- native control with explicit challenges. This is the largest reported gap category, but it is broad. It includes unresolved questions about semantic communication, integrated sensing, non-terrestrial networks, O-RAN, edge intelligence, digital twins, agentic control and distributed autonomy [54], [55], [82], [91], [110]. Additional studies in the corpus address the same point [124], [148], [302].

A central 6G-specific gap is the boundary of autonomous authority. AI may optimize radio resources, configure slices, coordinate agents and manage services with limited human intervention. It remains unclear which decisions can be delegated, which require approval, and which must be constrained by non-negotiable policies. Symbiotic and agentic network proposals expose this issue [56], [57], [61]. Future research should define authority models, policy conflict resolution, secure tool use and emergency shutdown. These mechanisms must operate at machine speed while remaining auditable by humans.

Semantic communication creates a different set of gaps. When AI determines which meaning is relevant and which information can be omitted, communication quality is no longer measured only through bit accuracy. Trustworthiness depends on semantic fidelity, task relevance, bias and the consequences of missing information. Explainable semantic communication research is emerging [153], [302], but common metrics and assurance methods are limited. Future studies should test semantic errors in safety-critical tasks and examine who defines the intended meaning [97].

Non-terrestrial, aerial and integrated networks create cross-layer and cross-jurisdictional problems. UAVs, satellites, edge devices and terrestrial operators may share data and control. Security, privacy and accountability can change as information moves between infrastructures and legal environments [30], [32], [110], [160], [178]. Research should examine identity, trust transfer, incident coordination and continuity of assurance across these boundaries. It should also include intermittent connectivity and resource-constrained devices rather than assuming stable cloud access [219], [300].

Cross-domain transfer remains weakly studied. A model designed for one operator, city or application may not be transferred to another. Healthcare, vehicles, smart grids, agriculture and industrial systems have different risk thresholds and affected stakeholders [10], [18], [107], [130], [137]. Additional studies in the corpus address the same point [142]. Future research should compare domains using common assurance concepts while preserving context-specific requirements. Transfer learning and federated learning may support adaptation, but they do not remove the need to test local fairness, safety and regulation [52], [68], [283].

Sustainability is frequently mentioned but rarely integrated with ethical assurance. Fifty-nine studies explicitly connected sustainability or environmental concerns with unresolved issues, while Chapter 7 identified only four explicit sustainability mechanisms. Energy-efficient AI, green networking and sustainable architectures are active topics [2], [22], [24], [33], [99], [108], but future work should measure the full lifecycle cost of model training, inference, communication, hardware and replacement. It should also examine rebound effects, where efficiency enables more intensive use and increases total resource consumption [262].

The broader 6G gap is a lack of stable objects of evaluation. Architectures, interfaces and standards continue to change, so methods can become obsolete before large trials are completed. Research should use modular assurance claims that can be updated as components change. Reference architectures and interoperability profiles can provide continuity [95], [123], [294], but they must be validated against evolving standards and real deployment constraints.

8.9 Integrated Interpretation and Direct Answer to RQ4

The direct answer to RQ4 is that the field still lacks an integrated and empirically validated pathway from responsible-AI principles to operational 5G/6G assurance. The literature has identified many risks and proposed many methods, but the relationship

between risk, control, evidence and responsible actors remains incomplete. The largest gaps are not isolated technical problems. They are coordination problems across data, models, network functions, organizations, standards and users.

The priority is realistic validation. More than half of the corpus is conceptual, survey or architecture-level, while only 21 studies show clear prototype, testbed or real-world indicators. Future work should establish shared validation ladders, benchmarks and multi- operator testbeds. Claims should be linked to measurable evidence and failure scenarios. This is especially important for AI-native control, security, privacy, safety and governance, which combines high reported gap prevalence with substantial validation deficits. The priority matrix is presented in Figure 8 [25].

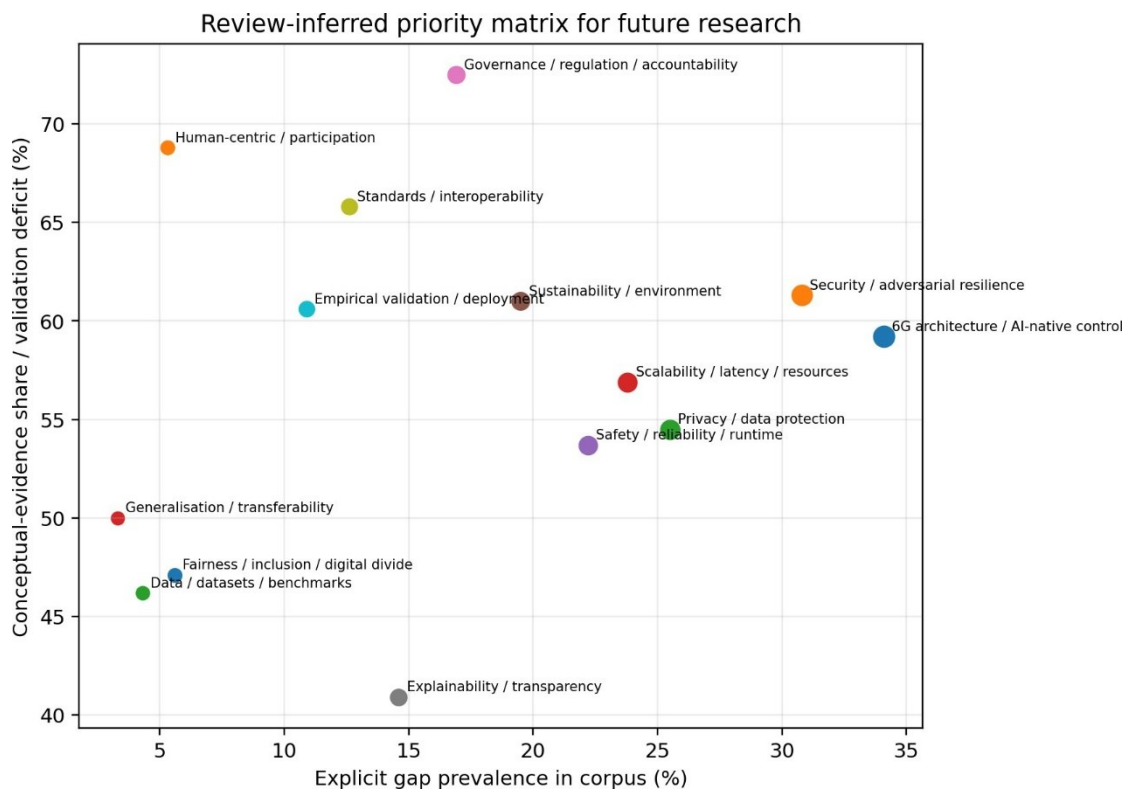


Figure 8-5. Review-inferred priority matrix combining explicit gap prevalence and validation deficit.

The second priority is lifecycle accountability. The field needs mechanisms that identify who is responsible for data, models, policies, decisions, monitoring and incident response. Logging and traceability are useful, but they must support causal reconstruction, legal duties and redress. Regulatory-readiness indices, incident-reporting proposals, standard-compliant fairness auditing and accountable network case studies provide initial directions [16], [17], [218], [221]. Future work should test these mechanisms in distributed operator and vendor ecosystems.

The third priority is human-centered evaluation. User acceptance, meaningful oversight, fairness, inclusion and contestability remain less developed than technical security or privacy. Research should include affected groups and operational personnel when defining requirements and evaluating systems. It should measure comprehension, workload, appropriate reliance, service equity and access to redress. Human-centered communication and market-driven adoption studies show that this agenda is beginning to emerge [10], [65], [171], but it is not yet a routine part of network-AI evaluation.

The fourth priority is assurance interoperability. Technical controls are currently distributed across federated learning, XAI, blockchain, privacy mechanisms, runtime monitors and standards. Future architecture should allow these controls to exchange evidence. A network controller should be able to use data provenance, uncertainty, policy constraints, explanation metadata and audit records in a consistent way. Privacy standards, AI-native architecture standards and network-digital-twin reference models illustrate pieces of this ecosystem [58], [123], [294]. The research gap is to integrate them without excessive latency, complexity or energy use [238], [281], [292].

The fifth priority is continuous assurance for adaptive and agentic systems. Static certification is insufficient when models learn, environments change and agents access new tools. Research should define triggers for re-evaluation, safe update procedures, model and prompt provenance, runtime policy enforcement, rollback and post-incident learning. Fail- safe O-RAN control and zero-touch assurance provide useful foundations [120], [155], [189], but broader evidence is needed across applications and operators.

The sixth priority is context-sensitive 6G governance. 6G research should avoid if one trustworthy-AI framework will fit every domain. Healthcare, vehicles, industrial systems, education and public services have different consequences and oversight needs.

Time horizon	Research agenda	Illustrative research question
Short term	Standardized reporting, dataset documentation and evaluation scenarios.	Which minimum evidence should accompany a trustworthy AI claim for a network function?
Short term	Create common attack, failure and distribution-shift scenarios.	Does the safeguard remain effective under adaptive and compound threats?
Medium term	Build federated, cross-operator testbeds with governance and human-oversight roles.	How can evidence be exchanged while protecting commercial and personal data?
Medium term	Evaluate fairness, acceptance, workload and contestability in domain cases.	Can affected users understand, challenge and obtain correction of AI-enabled network decisions?
Medium term	Integrate technical standards with AI incident reporting and audit procedures.	Which actor is responsible at each point of a distributed decision chain?
Long term	Develop continuous assurance for AI-native and agentic 6G control.	How can certification remain valid when models, prompts, tools and policies change?
Long term	Establish lifecycle sustainability and inclusion benchmarks.	Do efficiency gains reduce total environmental and social cost across the service lifecycle?

Table 8-5. Proposed staged research agenda for RQ4.

A common assurance core can support interoperability, but domain-specific thresholds, stakeholders and redress processes remain necessary. Comparative cross-domain studies should test what can be standardized and what must remain contextual [93], [194].

Overall, the remaining research gaps can be summarized through a risk-solution-evidence-responsibility chain. A trustworthy 5G/6G system requires a clearly defined risk, a control that addresses that risk, evidence that the control works under realistic conditions, and an actor responsible for maintaining it. Current studies often cover two or three links, but fewer cover all four. Future research should therefore move from isolated algorithms and broad ethical principles toward end-to-end assurance cases that remain valid throughout deployment and change

8.10 Limitations of the RQ4 Analysis

The first limitation is that the gap analysis is based on titles, abstracts, keywords and screening justifications rather than complete full-text coding. Abstracts frequently summarize contributions and major challenges, but they may omit detailed limitations, negative results and future-work statements. The reported counts therefore represent visible abstract-level gap signals. They should not be interpreted as a complete inventory of every limitation described in the 302 studies.

The second limitation concerns keyword dependence. Different research communities use different terminology for similar problems. Accountability may be discussed as responsibility, auditability, liability or assurance. Human-centered design may appear as user acceptance, social impact or participatory design. The taxonomy attempted to group these terms, but some records may have been missed or assigned to a neighboring category. Multi-label coding reduces forced classification, yet it also means that categories overlap and percentages do not sum to 100%.

The third limitation is the distinction between reported and inferred gaps. Explicit gap signals are easier to reproduce because they rely on visible language. Review-inferred gaps require interpretation across RQ1, RQ2 and RQ3. For example, the difference between 134 governance-risk studies and 36 governance-solution studies suggests a gap, but the two categories are not perfectly equivalent. Some governance risks may be addressed indirectly through technical controls, and some solution studies may not use governance terminology. The inferred findings should therefore be treated as structured analytical conclusions rather than exact measurements.

The fourth limitation concerns evaluation maturity. The three maturity categories were inferred from abstract-level indicators such as simulation, dataset, prototype, testbed and real-world validation. A paper may contain extensive experiments that are not described in the abstract, or it may use word deployment without reporting operational evidence. The result is best used to show the broad balance between conceptual and applied research, not to rank individual studies by quality.

The fifth limitation concerns the temporal boundary. The search covers January 2020 to June 2026. The 2026 total is a partial-year value, and some early-access publications may have indexing dates that differ from final publication dates. The field is changing rapidly,

particularly in agentic AI, AI-RAN, semantic communication and regulation. New standards, laws and trials may alter the identified priorities after the search date.

The sixth limitation is corpus scope. The SLR focuses on AI-enabled 5G/6G and ethical or responsible-AI concerns. Relevant evidence may exist in adjacent fields such as cloud governance, safety-critical AI, general telecommunications regulation, human-computer interaction and cybersecurity assurance without explicitly mentioning 5G or 6G. Excluding these studies preserves the review boundary but may understate transferable methods and established governance practices [122].

The seventh limitation is that publication frequency does not equal severity or importance. A rarely discussed issue can be critical, while a frequent issue may have received substantial attention. The priority analysis therefore combines prevalence with evidence maturity and with the risk-solution comparisons from earlier chapters. Even so, final research priorities require expert judgement, stakeholder consultation and domain-specific impact assessment.

Despite these limitations, the RQ4 analysis provides a consistent map of the most visible unresolved issues in the final corpus. It also makes the basis of inference explicit. Within the stated abstract-level protocol, this final synthesis records visible limitations, future directions, affected lifecycle stages and the distinction between author-reported and review-inferred gaps. This preserves traceability while keeping the analytical scope consistent with the final corpus used throughout the thesis.

Chapter 9 Conclusions

This chapter synthesizes the findings of the systematic literature review on ethical issues in AI-enabled 5G and 6G networks. The review covered publications from 2020 to June 2026 and used a final corpus of 302 studies selected through the title-and-abstract protocol described in Chapter 4. The four research questions were examined as a connected sequence of the development of the field, the most frequent risks, the proposed methods and standards, and the remaining research gaps.

9.1 Development and Current Status of the Field

The first conclusion is that the field has expanded rapidly and has become increasingly centered on 6G. Only 52 studies in the corpus were published between 2020 and 2023, compared with 38 in 2024, 101 in 2025 and 111 during the first six months of 2026. Although the 2026 total is partial and may include early-access publications, the overall pattern shows a clear rise in academic attention. Ethical questions are no longer treated only as secondary concerns in network security. They now accompany research on AI-native architectures, autonomous control, semantic communication, network slicing, edge intelligence and distributed learning.

Studies focused only on 6G represented 42.7% of the corpus, while 24.5% addressed both 5G and 6G. This reflects a shift from using AI as a tool for individual network tasks towards treating AI as part of the network architecture itself. Early studies connected explainability and quality-of-trust with human confidence in autonomous 6G systems [92], [145]. Later work extended the discussion to robust resource management, sovereign AI, agentic control and semantic communication [61], [131], [302]. The answer to RQ1 is therefore that the field is growing quickly, moving towards 6G and adopting a broader socio-technical understanding of trustworthiness. However, its maturity remains uneven because simulations and conceptual architecture are much more common than testbeds and field evidence.

9.2 Most Frequent Ethical Risks

The answer to RQ2 is that cybersecurity compromise and adversarial manipulation are the most frequently identified risks, appearing in 63.6% of the corpus. Privacy loss and improper data use followed at 54.0%, unsafe or unreliable AI behavior at 52.6%, and loss of trust at 50.7%.

Governance and regulatory inadequacy appeared in 44.4% of studies, while opacity and limited explainability appeared in 38.4%. Environmental and resource burden was less frequent at 20.5%, but the lower count indicates limited integration into responsible-AI analysis rather than low importance [266].

These risks are strongly connected. Security and privacy were coded together in 44.4% of the studies, while security also frequently co-occurred with unreliability, loss of trust and governance weaknesses. An attack can expose personal data, change automated decisions, interrupt critical services and create uncertainty about responsibility. Federated learning illustrates the same interdependence and keeping raw data local can reduce centralization, but it does not automatically prevent model-update leakage, malicious clients or unaccountable use [48], [244], [293]. Ethical risk in 5G/6G should therefore be treated as a chain of technical, organizational and human consequences rather than as isolated categories [204], [286].

9.3 Methods, Frameworks and Standards

RQ3 showed that the literature offers many safeguards, but no single method covers the complete assurance problem. Federated and distributed learning was the most frequent solution family, appearing in 99 studies, followed by explainability and interpretability in 98. Trust, reputation and assurance mechanisms appeared in 68 studies, blockchain and distributed ledgers in 63, and standards or interoperability approaches in 62. Privacy-enhancing technologies and security or robust-learning controls each appeared in 51 studies [249], [255], [296].

The main conclusion is that trustworthy network AI requires layered assurance. Federated learning supports data locality, cryptographic and privacy methods protect data and updates, explainable AI supports inspection, robust learning and zero-trust controls protect integrity, blockchain can support provenance, and runtime controls can preserve intervention and fail-safe behavior [120]. Governance frameworks and standards are needed to define roles, evidence and conformance procedures [123], [189], [214]. The strongest proposals connect prevention, transparency, monitoring and accountability. Privacy standards for federated systems also show that technical safeguards must be linked to common interfaces and auditable processes [58], [30], [200], [282].

9.4 Remaining Research Gaps

The answer to RQ4 is that the largest gaps concern AI-native 6G control, realistic validation and lifecycle accountability. The most frequent gap category was 6G architecture and AI-native control, identified in 103 studies. Security and adversarial resilience gaps appeared in 93 studies, privacy gaps in 77, scalability and resource constraints in 72, safety or runtime assurance gaps in 67, sustainability in 59, and governance or accountability gaps in 51 [222], [264], [299].

The central weakness is the distance between proposed safeguards and operational evidence. The analysis identified 168 conceptual or survey-oriented studies but only 21 records with clear prototype, testbed or real-world indicators. Controlled evaluation cannot fully represent multi-operator networks, changing data, strategic attackers, legal conflict or human workload. A second weakness is incomplete lifecycle accountability. Many studies focus on training or inference, while fewer describe incident reporting, rollback, recertification and remedy. Research on regulatory readiness and AI incident reporting shows why telecommunications governance must identify responsible actors before deployment and provide procedures for responding to harm [16], [17]. A third weakness is limited human-centered evidence. Fairness, acceptance, contestability and meaningful oversight are often recognized but rarely evaluated with users or operators [127], [171].

9.5 Contributions and Implications

The main contribution of this SLR is the integration of four forms of evidence that are often examined separately. It maps the growth of the field, ranks its risks, classifies proposed responses and connects them to remaining gaps. This sequence prevents frequency from being confused with effectiveness. A risk may be widely recognized but poorly addressed, a method may be frequently proposed but weakly validated, and a standard may exist without operational monitoring or a clear route to redress.

For researchers, the findings support reproducible, cross-layer and lifecycle evaluation. For operators, they emphasize the need to combine cybersecurity monitoring with model governance, explanation, incident response and human intervention. For vendors and developers, they highlight documentation, update provenance and transparent limitations. For

regulators and standards bodies, they support common assurance artefacts, interoperable audit evidence and clear responsibility maps. Emerging approaches such as sovereign AI, secure trust frameworks and network digital twins may support this development, but they also require governance of their own [61], [269], [294], [38], [59].

9.6 Limitations and Future Research

The conclusions should be interpreted in relation to review design. The final corpus was created through title-and-abstract assessment rather than complete full-text screening of every article. Abstracts may omit secondary risks, standards, limitations and implementation details. The searches were also limited to English-language publications from IEEE Xplore, Springer, ScienceDirect and Web of Science. Relevant standards documents, policy reports and work in other languages may not be represented. The period ends in June 2026, and the 2026 count is not directly comparable with complete years. In addition, the risk and solution categories were non-exclusive, so their percentages did not come to 100%.

Future research should create shared datasets, attack scenarios, benchmarks and validation ladders for trustworthy network AI. Multi-operator testbeds should evaluate safeguards under drift, failure, malicious participation and resource constraints. Responsibility should be mapped across data providers, model developers, vendors, operators, service providers and public authorities. Explainability should be assessed through real decisions, because an explanation has practical value only when it can be understood, challenged and used for corrective action. Continuous assurance is also required for AI-native and agentic networks, including runtime constraints, update provenance, incident triggers and safe rollback [85], [120]. Finally, sustainability and inclusion should become measurable assurance requirements through lifecycle energy, affordability, accessibility and distribution-of-benefit indicators [203], [228], [270].

9.7 Final Conclusion

The overall conclusion is that ethical alignment in AI-enabled 5G and 6G networks cannot be achieved through a single algorithm, security control or policy statement. It requires coordinated technical, organizational, legal and human safeguards across the full lifecycle. Literature has made substantial progress in recognizing risks and proposing methods, but the evidence remains fragmented and many solutions have not been validated under realistic conditions. As AI moves from supporting isolated tasks to directing AI-native network behavior, trustworthiness must become part of architecture, operation and governance from the beginning. The central challenge is not merely to create more intelligent networks, but to ensure that their intelligence remains secure, understandable, contestable, governable and aligned with the people and communities that depend on it.

References

- [1] A. Siam, S. Shohan, N. Faruqui, M. Alazab. (2025). Explainable Intrusion Detection and Prevention Framework for Next-Generation Networks. 2025 IEEE International Conference on Signal Processing, Information, Communication and Systems (SPICSCON). <https://doi.org/10.1109/spicscon69221.2025.11503950>
- [2] Agarwal, A. Nanda. (2024). Beyond 5G: 6G Wireless Systems for Sustainable Earth and Humanity with Artificial Intelligence. 2024 IEEE International Conference on Blockchain and Distributed Systems Security (ICBDS). <https://doi.org/10.1109/icbds61829.2024.10837145>
- [3] Agarwal, M. Kumar, M. J. Nene. (2025). Enhancements for Developing a Comprehensive AI Fairness Assessment Standard. 2025 17th International Conference on COMmunication Systems and NETworks (COMSNETS). <https://doi.org/10.1109/comsnets63942.2025.10885551>
- [4] Bandi, S. Yalamarthi. (2022). Towards Artificial Intelligence Empowered Security and Privacy Issues in 6G Communications. 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS). <https://doi.org/10.1109/icscds53736.2022.9760857>
- [5] Do Nascimento Júnior, A. Palaios, K. Raizer, S. K. Mohalik, R. Inam. (2025). Safe and Explainable AI in 6G Telecommunication Systems: The Way Forward. 2025 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA). <https://doi.org/10.1109/cogsima64436.2025.11079488>
- [6] F. Şahin, Y. Güven, S. T. Başaran, T. Kumbasar. (2025). XAI in Wireless Communications: A Case Study on Interpretable 5G Performance Analysis. 2025 IEEE 36th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC). <https://doi.org/10.1109/pimrc62392.2025.11275455>
- [7] Ferreira, B. O. Özparlak, H. Aden, I. Symeonidis, M. Çetİn, E. Kanaki, P. Chatzimisios. (2025). Drowsiness and Fatigue Recognition Systems for Connected Vehicles, 6G and the EU AI Act. 2025 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit). <https://doi.org/10.1109/eucnc/6gsummit63408.2025.11037012>
- [8] K. Gizzini, Y. Medjahdi, M. B. Mabrouk. (2024). GRACE: Gradient-based XAI Scheme for Channel Estimation in Wireless Communications. 2024 IEEE International Mediterranean Conference on Communications and Networking (MeditCom). <https://doi.org/10.1109/meditcom61057.2024.10621232>
- [9] K. Sangaiah, J. Anandakrishnan, S. Kumar, G. -B. Bian, S. A. AlQahtani, D. Draheim. (2026). Point-KAN: Leveraging Trustworthy AI for Reliable 3-D Point Cloud Completion With Kolmogorov-Arnold Networks for 6G-IoT Applications. IEEE Internet of Things Journal. <https://doi.org/10.1109/jiot.2025.3576434>
- [10] Mittal, S. Chowdhury, M. P. M. Raj, J. Kaur, P. K. Singh. (2026). Technology Acceptance of AI and 5G Systems an Empirical Social Science Study. 2026 5th International Conference on Innovative Practices in Technology and Management (ICIPTM). <https://doi.org/10.1109/iciptm69057.2026.11465771>
- [11] Sharma, S. Rani. (2026). Enhancing 6G-IoT Network Security: A Trustworthy and Responsible AI-Driven Stacked-Hybrid Model for Attack Detection. IEEE Internet of Things Journal. <https://doi.org/10.1109/jiot.2025.3566403>

- [12] Sharma, V. Babu, V. Khirasaria, P. D. (2025). Neuro-Symbolic AI for Real-Time, Explainable 6G Network Slice Optimization in Mission-Critical Environments. 2025 16th International Conference on Information and Communication Technology Convergence (ICTC). <https://doi.org/10.1109/ictc66702.2025.11388529>
- [13] Abdullahi, Ahmed Danladi, Bahrami, Erfan, Dargahi, Tooska, Al-Khalidi, Mohammed, Hammoudeh, Mohammad. (2025). Robust trust management in Intelligent Transportation System: A machine learning approach. International Journal of Critical Infrastructure Protection. <https://doi.org/10.1016/j.ijcip.2025.100812>
- [14] Adil, Muhammad, Song, Houbing, Khan, Muhammad Khurram, Farouk, Ahmed, Jin, Zhanpeng. (2024). 5G/6G-enabled metaverse technologies: Taxonomy, applications, and open security challenges with future research directions. Journal of Network and Computer Applications. <https://doi.org/10.1016/j.jnca.2024.103828>
- [15] Aerts, Ann, Bogdan-Martin, Doreen. (2021). Leveraging data and AI to deliver on the promise of digital health. International Journal of Medical Informatics. <https://doi.org/10.1016/j.ijmedinf.2021.104456>
- [16] Agarwal, Avinash, Agarwal, Peeyush, Nene, Manisha J.. (2026). The AI Regulatory Readiness Index ARRI: Assessing Cross-jurisdictional legal preparedness for AI in telecommunications. Computer Law & Security Review. <https://doi.org/10.1016/j.clsr.2026.106340>
- [17] Agarwal, Avinash, Nene, Manisha J.. (2026). Incorporating AI incident reporting into telecommunications law and policy: Insights from India. Computer Law & Security Review. <https://doi.org/10.1016/j.clsr.2026.106263>
- [18] Ahad, Abdul, Jiangbina, Zheng, Tahir, Mohammad, Shayea, Ibraheem, Sheikh, Muhammad Aman, Rasheed, Faizan. (2024). 6G and intelligent healthcare: Taxonomy, technologies, open issues and future research directions. Internet of Things. <https://doi.org/10.1016/j.iot.2024.101068>
- [19] Ahmed, Imran, Ahmad, Misbah, Jeon, Gwanggil. (2025). Federated Learning in Convergence ICT: A Systematic Review on Recent Advancements, Challenges, and Future Directions. Computers, Materials and Continua. <https://doi.org/10.32604/cmc.2025.068319>
- [20] Ait Lahcen, Abdellatif, Rajendran, Jegan, Slaughter, Gymama. (2026). Advances and challenges in digitally connected point-of-care biosensing. Biosensors and Bioelectronics: X. <https://doi.org/10.1016/j.biosx.2025.100728>
- [21] Akbar, Muhammad Sajjad, Hussain, Zawar, Ikram, Muhammad, Sheng, Quan Z., Mukhopadhyay, Subhas Chandra. (2025). On challenges of sixth-generation (6G) wireless networks: A comprehensive survey of requirements, applications, and security issues. Journal of Network and Computer Applications. <https://doi.org/10.1016/j.jnca.2024.104040>
- [22] Akgül, Özgür Umut, Varvara, Antonio, de la Oliva, Antonio, Charatsaris, Panagiotis, Diamanti, Maria, Burguera, Pere Garau, Ericson, Mårten, Wänstedt, Stefan, Ziółkowski, Marcin, Tarasiuk, Halina, Hellaoui, Hamed, Papavassiliou, Symeon, Tsekenis, Vasileios, Barmponakis, Sokratis, Demestichas, Panagiotis, Khorsandi, Bahare M., Harkous, Hasanin. (2025). Sustainable 6G architecture: An organic evolution of 5G networks. Computer Networks. <https://doi.org/10.1016/j.comnet.2025.111629>

- [23] Alatawi, Mohammed Naif. (2025). EdgeGuard-IoT: 6G-Enabled Edge Intelligence for Secure Federated Learning and Adaptive Anomaly Detection in Industry 5.0. *Computers, Materials and Continua*. <https://doi.org/10.32604/cmc.2025.066606>
- [24] Aldehim, Ghadah, Khan, Sunawar, Shahzad, Tariq, Khan, Muhammad Amir, Ghadi, Yazeed Yasin, Jiang, Weiwei, Mazhar, Tehseen, Hamam, Habib. (2025). Balancing sustainability and security: A review of 5G and IoT in smart cities. *Digital Communications and Networks*. <https://doi.org/10.1016/j.dcan.2025.06.007>
- [25] Ali, Jehad, Kumar Singh, Sushil, Jiang, Weiwei, Alenezi, Abdulmajeed M., Islam, Muhammad, Ibrahim Daradkeh, Yousef, Mehmood, Asif. (2025). A deep dive into cybersecurity solutions for AI-driven IoT-enabled smart cities in advanced communication networks. *Computer Communications*. <https://doi.org/10.1016/j.comcom.2024.108000>
- [26] Almadhor, Ahmad, Alqahtani, Abdullah, Al Hejaili, Abdullah, Ayari, Mohamed, Zaidi, Monji Mohamed, Kryvinska, Natalia, Gadekallu, Thippa Reddy, Abbas, Sidra. (2025). 6G-Enabled Federated Intelligence and Transparent Framework for Aerial Scene Classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 18, 25630–25639. <https://doi.org/10.1109/jstars.2025.3617077>
- [27] Alnajar, Omar, Barnawi, Ahmed. (2024). TactiFlex: A Federated learning-enhanced in-content aware resource allocation flexible architecture for Tactile IoT in 6G networks. *Engineering Applications of Artificial Intelligence*. <https://doi.org/10.1016/j.engappai.2024.108934>
- [28] Alnajar, Omar, Barnawi, Ahmed. (2025). A Novel Clustered Distributed Federated Learning Architecture for Tactile Internet of Things Applications in 6G Environment. *CMES - Computer Modeling in Engineering and Sciences*. <https://doi.org/10.32604/cmcs.2025.065833>
- [29] Alosban, Nujud, Alharbi, Abeer A.K.. (2025). NeuroCivitas: A Federated Deep Learning Model for Adaptive Urban Intelligence in 6G Cognitive Cities. *Computers, Materials and Continua*. <https://doi.org/10.32604/cmc.2025.067523>
- [30] Alsamhi, Saeed Hamood, Srivastava, Sumit, Rashid, Mamoon, Alhabeeb, Amnna, Kumar, Santosh, Rajput, Navin Singh, Hawbani, Ammar, Zhao, Liang, Al-qaness, Mohammed A.A., Curry, Edward. (2025). Skyward secure: Advancing drone data-sharing in 6G with decentralized dataspace and supported technologies. *Journal of Parallel and Distributed Computing*. <https://doi.org/10.1016/j.jpdc.2025.105040>
- [31] Ameer, Mongi Ben, Chebil, Jalel, Habaei, Mohamed Hadi, Tahar, Jamel Bel Hadj, Islam, Md Rafiqul, Sheikh, Abdul Manan. (2026). Explainable AI with EDA for V2I path loss prediction. *SCIENTIFIC REPORTS*, 16(1), 4954. <https://doi.org/10.1038/s41598-026-34987-8>
- [32] Amodu, Oluwatosin Ahmed, Hanapi, Zurina Mohd, Mahmood, Raja Azlina Raja, Saif, Faten A., Althumali, Huda, Jarray, Chedia, Bakar, Umar Ali, Adam, Mohammed Sani. (2026). Machine Learning-Enabled NTN-Assisted IoT: Mapping the Security Landscape. *Computers, Materials and Continua*. <https://doi.org/10.32604/cmc.2026.074678>
- [33] Andrés Azcoitia, Santiago, Illescas Cabiró, Sergio, Frías, Zoraida. (2026). A framework to incentivise green networks and infrastructure in the 6G mobile ecosystem. *Telecommunications Policy*. <https://doi.org/10.1016/j.telpol.2026.103221>

- [34] Anjum, Naveed, Latif, Zohaib, Chen, Hongsong. (2025). Security and privacy of industrial big data: Motivation, opportunities, and challenges. *Journal of Network and Computer Applications*. <https://doi.org/10.1016/j.jnca.2025.104130>
- [35] Arana-Catania, Miguel, Sonee, Amir, Khan, Abdul-Manan, Fatehi, Kavan, Tang, Yun, Jin, Bailu, Soligo, Anna, Boyle, David, Calinescu, Radu, Yadav, Poonam, Ahmadi, Hamed, Tsourdos, Antonios, Guo, Weisi, Russo, Alessandra. (2025). Explainable Reinforcement and Causal Learning for Improving Trust to 6G Stakeholders. *IEEE Open Journal of the Communications Society*, 6, 4101–4125. <https://doi.org/10.1109/ojcoms.2025.3563415>
- [36] Arash HeidariSeyed Hamed RastegarAhmad Khonsari. (2026). Artificial Intelligence-driven privacy preservation in the internet of vehicles: a comprehensive systematic literature review. *Journal of Big Data*. <https://doi.org/10.1186/s40537-025-01360-x>
- [37] Asim, Muhammad, Ateya, Abdelhamied A., Wani, Mudassir Ahmad, Ali, Gauhar, ElAffendi, Mohammed, El-Latif, Ahmed A. Abd, Siyal, Reshma. (2026). A Comprehensive Survey on Blockchain-Enabled Techniques and Federated Learning for Secure 5G/6G Networks: Challenges, Opportunities, and Future Directions. *Computers, Materials and Continua*. <https://doi.org/10.32604/cmc.2025.070684>
- [38] Aslam, Anjum Mohd, Chaudhary, Rajat, Bhardwaj, Aditya, Kumar, Neeraj, Buyya, Rajkumar. (2025). Digital twins-enabled game theoretical models and techniques for metaverse Connected and Autonomous Vehicles: A survey. *Journal of Network and Computer Applications*. <https://doi.org/10.1016/j.jnca.2025.104138>
- [39] Azeez, Sarmad Dheyaa, Ilyas, Muhammad, Ahmed, Saadaldeen Rashid. (2026). Graph Neural Networks with Multi-Head Attention and SHAP-Based Explainability for Robust, Interpretable, and High-Throughput Intrusion Detection in 5G-Enabled Software Defined Networks. *Computers, Materials and Continua*. <https://doi.org/10.32604/cmc.2026.074930>
- [40] Bac Trinh-NguyenSara BerriSin G. TeoTram Truong-HuuArsenia Chorti. (2026). A Framework for Global Trust and Reputation Management in 6G Networks. *Machine Learning for Networking*. https://doi.org/10.1007/978-3-032-00552-6_4
- [41] Balmer, Roberto E., Levin, Stanford L., Schmidt, Stephen. (2020). Artificial Intelligence Applications in Telecommunications and other network industries. *Artificial intelligence, economy and society*. <https://doi.org/10.1016/j.telpol.2020.101977>
- [42] Banad, Yaser M., Sharif, Sarah S., Rezaei, Zahra. (2025). Artificial intelligence and machine learning for smart grids: from foundational paradigms to emerging technologies with digital twin and large language model-driven intelligence. *Energy Conversion and Management: X*. <https://doi.org/10.1016/j.ecmx.2025.101329>
- [43] Barcena, Jose Luis Corcuera, Ducange, Pietro, Marcelloni, Francesco, Nardini, Giovanni, Noferi, Alessandro, Renda, Alessandro, Ruffini, Fabrizio, Schiavo, Alessio, Stea, Giovanni, Viridis, Antonio. (2023). Enabling federated learning of explainable AI models within beyond-5G/6G networks. *Computer Communications*, 210, 356–375. <https://doi.org/10.1016/j.comcom.2023.07.039>
- [44] Basaran, Osman Tugay, Dressler, Falko. (2025). XAIomaly: Explainable and interpretable Deep Contractive Autoencoder for O-RAN traffic anomaly detection. *Computer Networks*. <https://doi.org/10.1016/j.comnet.2025.111145>

- [45] Berto, Filippo, Ardagna, Claudio A., Banzi, Massimo, Anisetti, Marco. (2024). Assurance in Advanced 5G Edge Continuum. IEEE ACCESS, 12, 178659–178671. <https://doi.org/10.1109/access.2024.3503437>
- [46] Bhandari, Sovit, Ranjan, Navin, Kim, Yeong-Chan, Khan, Pervez, Kim, Hoon. (2022). A Fairness of Data Combination in Wireless Packet Scheduling. SENSORS, 22(4), 1658. <https://doi.org/10.3390/s22041658>
- [47] Bin CaoJianwei ZhaoXin LiuYun Li. (2024). Adaptive 5G-and-beyond network-enabled interpretable federated learning enhanced by neuroevolution. Science China Information Sciences. <https://doi.org/10.1007/s11432-023-4011-4>
- [48] Blika, Afroditi, Pamos, Stefanos, Doukas, George, Lamprou, Vangelis, Pelekis, Sotiris, Kontoulis, Michael, Ntanos, Christos, Askounis, Dimitris. (2025). Federated Learning for Enhanced Cybersecurity and Trustworthiness in 5G and 6G Networks: A Comprehensive Survey. IEEE Open Journal of the Communications Society, 6, 3094–3130. <https://doi.org/10.1109/ojcoms.2024.3449563>
- [49] Thacker, A. K. Pandey. (2023). AI System Security and Privacy Risks in Sixth-Generation (6G) Mobile Cloud Using High-Performance Computing Implementation. 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE). <https://doi.org/10.1109/icacite57410.2023.10183326>
- [50] Zhang, X. Ren, W. Zhang, Y. Yuan, Z. Xiong, C. Li, L. Zhu. (2025). Privacy-Preserving Federated Learning for Data Heterogeneity in 6G Mobile Networks. IEEE Network. <https://doi.org/10.1109/mnet.2024.3510427>
- [51] Caglar TuncKaustubh Joshi. (2026). Next-Generation Intent-Based Networking in B5G with LLMs and Explainable AI. GenAI and LLMs for Beyond 5G Networks. https://doi.org/10.1007/978-3-032-06418-9_12
- [52] Cai, Taoyang, Ma, Jingfei, Ge-Zhang, Shangjie. (2025). Smart cities and smart health: Innovations in home medical devices for efficient healthcare delivery. Results in Engineering. <https://doi.org/10.1016/j.rineng.2025.105739>
- [53] Catak, Ferhat Ozgur, Kuzlu, Murat, Seo, Jungwon, Cali, Umit. (2026). Quantum key distribution secured federated learning for AI-driven physical layer tasks in next generation networks. Physical Communication. <https://doi.org/10.1016/j.phycom.2026.103186>
- [54] Celik, Abdulkadir, Eltawil, Ahmed M.. (2024). At the Dawn of Generative AI Era: A Tutorial-cum-Survey on New Frontiers in 6G Wireless Intelligence. IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY, 5, 2433–2489. <https://doi.org/10.1109/ojcoms.2024.3362271>
- [55] Chatzieleftheriou, Constantina, Liotou, Eirini. (2026). A Survey on AI for 6G: Challenges and Opportunities. IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY, 7, 3189–3221. <https://doi.org/10.1109/ojcoms.2026.3677293>
- [56] Chatzistefanidis, Ilias, Nikaein, Navid. (2025). Symbiotic agents: A novel paradigm for trustworthy AGI-driven networks. Computer Networks, 273, 111749. <https://doi.org/10.1016/j.comnet.2025.111749>
- [57] Chatzistefanidis, Ilias, Nikaein, Navid, Leone, Andrea, Maatouk, Ali, Tassioulas, Leandros, Morabito, Roberto, Pitsiorlas, Ioannis, Kountouris, Marios. (2026). AGORAN: An agentic open marketplace for 6G RAN automation. Computer Networks. <https://doi.org/10.1016/j.comnet.2025.111927>

- [58] Chen, Keyin, Liu, Yuanlin. (2025). Toward Privacy-Preserving AI Standards for Federated Learning in 6G-Enabled Digital Twin Environments. IEEE Communications Standards Magazine. <https://doi.org/10.1109/mcomstd.2025.3623142>
- [59] Chen, Qian, Wang, Xiangping, Liu, Lijin. (2026). NEXT-GENERATION AI-DRIVEN DIGITAL TWIN FRAMEWORK FOR MARKET VOLATILITY WARNING AND RISK DETECTION IN GLOBAL LOGISTICS SUPPLY CHAINS. IEEE COMMUNICATIONS STANDARDS MAGAZINE. <https://doi.org/10.1109/mcomstd.2025.3641867>
- [60] Cheng, Weibin, Lian, Wanmin, Tian, Junzhang. (2022). Building the hospital intelligent twins for all-scenario intelligence health care. DIGITAL HEALTH, 8, 20552076221107896. <https://doi.org/10.1177/20552076221107894>
- [61] Chetty, Swarna Bindu, Grace, David, Saunders, Simon, Harris, Paul, Tsiropoulou, Eirini Eleni, Quek, Tony, Ahmadi, Hamed. (2026). SovAI for 6G: Towards the Future of AI-Native Networks. IEEE Communications Standards Magazine. <https://doi.org/10.1109/mcomstd.2026.3689995>
- [62] Chithaluru, Premkumar, Singh, Aman, Dhatteval, Jagjit Singh, Sodhro, Ali Hassan, Albahar, Marwan Ali, Jurcut, Anca, Alkhayyat, Ahmed. (2023). An Optimized Privacy Information Exchange Schema for Explainable AI Empowered WiMAX-based IoT networks. Future Generation Computer Systems. <https://doi.org/10.1016/j.future.2023.06.003>
- [63] Christo Kurisummoottil ThomasOmar HashashWalid Saad. (2026). Reasoning Foundations for Generative AI-Based Wireless Networks. GenAI and LLMs for Beyond 5G Networks. https://doi.org/10.1007/978-3-032-06418-9_7
- [64] Corazza, Giovanni Emanuele, Rinaldi, Ilaria, Coletto, Irene, Mirabile, Mario, Mazzarella, Sara, Parasecolo, Francesca, Baiocchi, Andrea, Colonnese, Stefania, Panarese, Paola, Yildiz, Mert, Costa, Cristina, De Minico, Giovanna, Abbondante, Fulvia, De Tullio, Maria Francesca, Girardi, Maria Chiara, Parisi, Stefania, Serafini, Stefania, Tinnirello, Ilenia. (2026). A glimpse on the futures of telecommunication networks: From market, technology and regulation trends to strategic foresight. Journal of Open Innovation: Technology, Market, and Complexity. <https://doi.org/10.1016/j.joitmc.2026.100735>
- [65] D. Mitka, R. Fonseca, P. S. R. Henrique, P. Valter, R. Prasad. (2025). Designing for 6G Adoption: A Market-Driven Framework for Trustworthy, Personalization, and Inclusive Next Generation of User Experiences. 2025 28th International Symposium on Wireless Personal Multimedia Communications (WPMC). <https://doi.org/10.1109/wpmc67460.2025.11351138>
- [66] Dai, Xianlong, Cheng, Guang, Deng, Li, Yu, Ziyang, Chen, Zihan, Zhang, Xuman, Duan, Bingjie, Li, Yuheng. (2025). ITD: A novel measure for detecting portable hotspot devices in modern cyberspace. Computer Networks. <https://doi.org/10.1016/j.comnet.2025.111549>
- [67] Dang, Quang-Vinh, Vo, Thi-Bich-Diem, Nguyen, Ngoc-Son-An. (2026). FORTRESS-FL: Byzantine-robust and privacy-preserving federated orchestration for next-generation networks. Array. <https://doi.org/10.1016/j.array.2026.100680>
- [68] Das, Sree Krishna, Benkhelifa, Fatma, Sun, Yao, Abumarshoud, Hanaa, Abbasi, Qammer H., Imran, Muhammad Ali, Mohjazi, Lina. (2023). Comprehensive review on ML-based RIS-enhanced IoT systems: basics, research progress and future challenges. Computer Networks. <https://doi.org/10.1016/j.comnet.2023.109581>

- [69] Deng, Xiaoheng, Wang, Leilei, Gui, Jinsong, Jiang, Ping, Chen, Xuechen, Zeng, Feng, Wan, Shaohua. (2023). A review of 6G autonomous intelligent transportation systems: Mechanisms, applications and challenges. *Journal of Systems Architecture*. <https://doi.org/10.1016/j.sysarc.2023.102929>
- [70] Dritsas, Elias, Trigka, Maria. (2025). Big data and Internet of Things applications in smart cities: Recent advances, challenges, and critical issues. *Internet of Things*. <https://doi.org/10.1016/j.iot.2025.101770>
- [71] Du, Miao, Yang, Peng, Liu, Yinqiu, He, Xiaoming, Chen, Mingkai. (2025). DP-Fed6G: An adaptive differential privacy-empowered federated learning framework for 6G networks. *Digital Communications and Networks*. <https://doi.org/10.1016/j.dcan.2025.07.006>
- [72] Dui, Hongyan, Zhai, Jiabao, Xia, Wanyun, Xing, Liudong, Shao, Haidong, Wang, Ning. (2026). CausaLM-Net: An LLM-guided causal graph and state-space learning framework for fault diagnosis in cloud native 5G base stations. *Expert Systems with Applications*. <https://doi.org/10.1016/j.eswa.2026.131961>
- [73] Dunn, Patrick, Charbonneau, Rachael, Ozemek, Cemal, Arena, Ross, Wisløff, Ulrik. (2026). Digital care in your pocket: The role of technology in exercise, fitness, prevention, and rehabilitation. *Progress in Cardiovascular Diseases*. <https://doi.org/10.1016/j.pcad.2026.04.002>
- [74] G. De Lope, J. Zolfaghari-Bengar, R. Rony, A. Mahadevan, N. Kourtellis. (2026). X-CE: Scenario-Aware Explainable Deep Learning for MIMO Channel Estimation. *IEEE INFOCOM 2026 - IEEE Conference on Computer Communications*. <https://doi.org/10.1109/infocom59046.2026.11571707>
- [75] Egidia CirilloAlberto MoccardiMarcello PelosiFlora Amato. (2026). Resilient and Explainable AI for Reliable Systems. *Advanced Information Networking and Applications*. https://doi.org/10.1007/978-3-032-23335-6_18
- [76] Elahi, Haroon, Wang, Guojun, Xu, Yang, Castiglione, Aniello, Yan, Qiben, Shehzad, Muhammad Naeem. (2021). On the Characterization and Risk Assessment of AI-Powered Mobile Cloud Applications. *Computer Standards & Interfaces*. <https://doi.org/10.1016/j.csi.2021.103538>
- [77] O. Catak, U. Cali, M. Kuzlu, S. Sarp. (2023). Uncertainty Aware Deep Learning Model for Secure and Trustworthy Channel Estimation in 5G Networks. *2023 12th Mediterranean Conference on Embedded Computing (MECO)*. <https://doi.org/10.1109/meco58584.2023.10155011>
- [78] Rezazadeh, H. Chergui, J. Mangues-Bafalluy. (2023). Explanation-Guided Deep Reinforcement Learning for Trustworthy 6G RAN Slicing. *2023 IEEE International Conference on Communications Workshops (ICC Workshops)*. <https://doi.org/10.1109/iccworkshops57953.2023.10283684>
- [79] Fabian Chukwudi OgenyiChinyere Nneoma UgwuVal Hyginus Udoka EzeOkechukwu Paul-Chima UgwuJovita Nnenna UgwuMichael Ben OkonKelechi John Ukagwu. (2026). Transforming digital health using the internet of things for personalized interoperable and secure healthcare systems. *Discover Health Systems*. <https://doi.org/10.1007/s44250-026-00359-2>
- [80] Farooqui, Muhammad Saad, Khattak, Aizaz Ahmad, Awaji, Bakri Hossain, Alturki, Nazik, Alnazzawi, Noha, Hanif, Muhammad, Khan, Muhammad Shahbaz. (2025).

- AutoSHARC: Feedback Driven Explainable Intrusion Detection with SHAP-Guided Post-Hoc Retraining for QoS Sensitive IoT Networks. CMES - Computer Modeling in Engineering and Sciences, 145(3), 4395–4439. <https://doi.org/10.32604/cmes.2025.072023>
- [81] Fatma DönmezThomas Max PatalasMonika Eigenstetter. (2025). Ensuring Acceptance of Surveillance Through AI-Supported Radar Technology. Artificial Intelligence in HCI. https://doi.org/10.1007/978-3-031-93412-4_9
- [82] Feng, Chenyuan, Yang, Howard H., Guo, Kun, Xia, Wenchao, Liu, Chenxi, Quek, Tony Q.S.. (2026). AI-RAN: The pathway to future wireless networks. Journal of Information and Intelligence. <https://doi.org/10.1016/j.jiixd.2025.12.013>
- [83] Feng, Hailin, Gadekallu, Thippa Reddy, Xia, Yuchao, Zhao, Yongkang, Wen, Zhihao, Cai, Jijing, Bhattacharya, Pronaya, Fang, Kai, Liyanage, Madhusanka. (2026). Agentic AI Security in 6G Networks: A Survey of Emerging Attack Vectors, Vulnerabilities, and Defenses. IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY, 7, 6334–6368. <https://doi.org/10.1109/ojcoms.2026.3700939>
- [84] FengLan Wang. (2024). The Application of Intelligent Information Systems Driven by 6G Big Data in Product Sales Traceability. Wireless Personal Communications. <https://doi.org/10.1007/s11277-024-11210-3>
- [85] Fiandrino, Claudio, Attanasio, Giulia, Fiore, Marco, Widmer, Joerg. (2022). Toward native explainable and robust AI in 6G networks: Current state, challenges and road ahead. Computer Communications. <https://doi.org/10.1016/j.comcom.2022.06.036>
- [86] Firdaus, Muhammad, Larasati, Harashta Tatimma, Rhee, Kyung-Hyune. (2023). A Blockchain-Assisted Distributed Edge Intelligence for Privacy-Preserving Vehicular Networks. Computers, Materials and Continua. <https://doi.org/10.32604/cmc.2023.039487>
- [87] M. KØien. (2024). The Road to a Trustworthy 6G On the Need for a “Zero Trust 6G” Paradigm. Journal of Mobile Multimedia. <https://doi.org/10.13052/jmm1550-4646.2013>
- [88] G. R, U. P. Sah, B. Banerjee, M. V. Kumar. (2025). Enhancement of Employee Engagement with Ethical AI Enabled 6G Wireless Networks. Journal of Mobile Multimedia. <https://doi.org/10.13052/jmm1550-4646.21348>
- [89] G. Sajiv, N. Meenakshisundaram. (2025). Explainable and Compliant AI for 6G Communication Systems: Auditable Federated GANs with Blockchain-Backed Traceability and Smart Contract Governance. 2025 6th International Conference on Electronics and Sustainable Communication Systems (ICESC). <https://doi.org/10.1109/icesc65114.2025.11212412>
- [90] Garroppo, Rosario G., Giardina, Pietro Giuseppe, Landi, Giada, Ruta, Marco. (2025). Trustworthy AI and Federated Learning for Intrusion Detection in 6G-Connected Smart Buildings. FUTURE INTERNET, 17(5), 191. <https://doi.org/10.3390/fi17050191>
- [91] Guo, Shuaishuai, Zhong, Yutong, Feng, Zhenyu, Kang, Shengqi, Chen, Jichao. (2026). Towards transparent 6G AI-RAN: A survey on explainable deep reinforcement learning for intelligent network slicing. Journal of Information and Intelligence. <https://doi.org/10.1016/j.jiixd.2025.12.005>
- [92] Guo, Weisi. (2020). Explainable Artificial Intelligence for 6G: Improving Trust between Human and Machine. IEEE COMMUNICATIONS MAGAZINE, 58(6), 39–45. <https://doi.org/10.1109/mcom.001.2000050>

- [93] G. Khalid, H. H. Abbas, M. J. Hussein, M. J. Abu-AlShaeer, S. K. Mukhlif, D. Khlaponin. (2024). Unveiling the Next Frontier: The Future of Connected Technologies. 2024 36th Conference of Open Innovations Association (FRUCT). <https://doi.org/10.23919/fruct64283.2024.10749851>
- [94] H. U. Rashid, S. H. Jeong. (2025). Dynamic 5G Network Slice Management using Unsupervised Modeling and Explainable AI. 2025 Sixteenth International Conference on Ubiquitous and Future Networks (ICUFN). <https://doi.org/10.1109/icufn65838.2025.11169967>
- [95] H. You, Y. Jia, Y. Wang. (2026). Standards and Interoperability Profiles for 6G Wearable Health Monitoring: A Reference Architecture for Intelligent Sports Applications. IEEE Communications Standards Magazine. <https://doi.org/10.1109/mcomstd.2026.3675652>
- [96] H. Zheng, H. Zhang, M. Wu, L. Zhang. (2026). Trustworthy Deep Learning for Large-Scale Traffic Scheduling in 6G-IoT. IEEE Internet of Things Journal. <https://doi.org/10.1109/jiot.2025.3624239>
- [97] Han, Kaiyang, Jia, Xiaoqiang, Lin, Yangfei, Yoshinaga, Tsutomu, Li, Yalong, Wu, Jiale. (2025). Entropy-Bottleneck-Based Privacy Protection Mechanism for Semantic Communication. Computers, Materials and Continua. <https://doi.org/10.32604/cmc.2025.061563>
- [98] Hasan, Mohammad Kamrul, Habib, A.K.M. Ahasan, Islam, Shayla, Safie, Nurhizam, Ghazal, Taher M., Khan, Muhammad Attique, Alzahrani, Ahmed Ibrahim, Alalwan, Nasser, Kadry, Seifedine, Masood, Anum. (2024). Federated learning enables 6 G communication technology: Requirements, applications, and integrated with intelligence framework. Alexandria Engineering Journal. <https://doi.org/10.1016/j.aej.2024.02.044>
- [99] Hassan, Ali H., Ahmed, Elsadig Musa, Hussien, Jamal M., Sulaiman, Riza bin, Abdulhak, Mansoor, Kahtan, Hasan. (2025). A cyber physical sustainable smart city framework toward society 5.0: Explainable AI for enhanced SDGs monitoring. Research in Globalization. <https://doi.org/10.1016/j.resglo.2025.100275>
- [100] Hassija, Vikas, Majumder, Tamonash, Roy, Debangshu, Piyush, Raja, Chamola, Vinay. (2026). The role of large language models (LLMs) in enhancing intelligent transportation systems: A survey. Vehicular Communications. <https://doi.org/10.1016/j.vehcom.2025.100996>
- [101] He, Haoxin, Fei, Shufan, Yan, Zheng. (2026). Advancing 5G Security and Privacy with AI: A Survey. ACM COMPUTING SURVEYS, 58(2), 47. <https://doi.org/10.1145/3744555>
- [102] Heidari, Arash, Khonsari, Ahmad, Rastegar, Seyed Hamed. (2026). An energy-efficient privacy-preserving framework for intrusion detection in the internet of vehicles. Computers and Electrical Engineering. <https://doi.org/10.1016/j.compeleceng.2026.111003>
- [103] Heidari, Arash, Rastegar, Seyed Hamed, Khonsari, Ahmad. (2026). FedIoV: A secure and adaptive federated framework for real-time intrusion detection in vehicular networks. Future Generation Computer Systems. <https://doi.org/10.1016/j.future.2026.108448>
- [104] Hireche, Othmane, Benzaïd, Chafika, Taleb, Tarik. (2022). Deep data plane programming and AI for zero-trust self-driven networking in beyond 5G. Computer Networks. <https://doi.org/10.1016/j.comnet.2021.108668>

- [105] Hossain, M. Shamim, Muhammad, Ghulam, Guizani, Nadra. (2020). Explainable AI and Mass Surveillance System-Based Healthcare Framework to Combat COVID-19 Like Pandemics. *IEEE Network*, 34(4), 126–132. <https://doi.org/10.1109/mnet.011.2000458>
- [106] Hossain, Shajjad, Senouci, Sidi-Mohammed, Brik, Bouziane, Boualouache, Abdelwahab. (2025). A privacy-preserving Self-Supervised Learning-based intrusion detection system for 5G-V2X networks. *Ad Hoc Networks*. <https://doi.org/10.1016/j.adhoc.2024.103674>
- [107] Hosseini, Mohsen Masoumian, Masoumian Hosseini, Seyedeh Toktam, Haghighi, Elnaz, Qayumi, Karim, Ebrahimipour, Hossein, Pourabbasi, Ata, Koohepaei, Alireza, Alizadeh, Maryam, Shafiei, Zahra. (2025). The revolutionary impact of 6G technology on empowering health and building a smart society: A scoping review. *Computers in Biology and Medicine*. <https://doi.org/10.1016/j.compbiomed.2025.110496>
- [108] Hou, Chen, Huang, Tao, Huang, Qingyu, Yang, Xu, Wang, Xiaoding, Hu, Jia, Ali Khowaja, Sunder, Dev, Kapal. (2026). Adaptive Gradient Methods for Differentially Private TinyML in 6G. *IEEE WIRELESS COMMUNICATIONS*, 33(2), 58–64. <https://doi.org/10.1109/mwc.2025.3640093>
- [109] Huang, Lijia, Jing, Hengchang. (2026). Boosting Trustworthiness in 6G-Enabled Smart Industries 5.0 Through Explainable Artificial Intelligence on Industrial Machine Anomaly Data. *IEEE Internet of Things Journal*, 13(5), 8182–8189. <https://doi.org/10.1109/jiot.2025.3615253>
- [110] Ibrahim, Reham Wafae, Rodrigues, Tiago Koketsu, Kato, Nei, Ariyoshi, Masayuki, Hasegawa, Yohei. (2026). A Survey on AI-Empowered High-Frequency and FSO-Enabled Mobile Edge Computing in Non-Terrestrial Networks for 6G. *IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING*, 12, 9504–9528. <https://doi.org/10.1109/tccn.2026.3706515>
- [111] Islam, Md. Monirul, Hasan, Kamrul, Jeong, Seong Ho. (2026). Performance evaluation and optimization for 6G networks: A survey of KPIs, Tools, and AI models. *ICT Express*. <https://doi.org/10.1016/j.icte.2025.12.012>
- [112] Issa, Wael, Moustafa, Nour, Turnbull, Benjamin, Choo, Kim-Kwang Raymond. (2025). DT-BFL: Digital Twins for Blockchain-enabled Federated Learning in Internet of Things networks. *Ad Hoc Networks*. <https://doi.org/10.1016/j.adhoc.2025.103934>
- [113] Li, H. Ding, H. Li, Y. Xiao. (2025). Public Sentiment Analysis Proposal for Beyond-5G Deployment and Reconstructing Confidence in 6G Wireless Systems with Liquid Crystal Reconfigurability and Ethical AI. 2025 12th International Conference on Social Networks Analysis, Management and Security (SNAMS). <https://doi.org/10.1109/snams67467.2025.11391010>
- [114] Mu, Q. Liu, Z. Wang, T. Yuan, H. Sun, P. Yu. (2024). Leveraging Blockchain Based Federated Learning for Trustworthy 6G IoMT Image Analysis. *IEEE Wireless Communications*. <https://doi.org/10.1109/mwc.013.2300518>
- [115] J. Pérez-Romero, O. Sallent, I. Vilà, E. Kartsakli, Ö. F. Tuna, S. K. Mohalik, X. Tao. (2024). Towards Trustworthy Reinforcement Learning-based Resource Management in Beyond 5G. 2024 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit). <https://doi.org/10.1109/eucnc/6gsummit60053.2024.10597001>

- [116] J. T. Rajesh, L. K. R. Bhavanam, M. M. S. A. Varma, C. Rajesh, M. S. Khan, S. S. Kinthada. (2026). Optimizing Explainable AI for 5G-Enabled Autonomous Systems: Enhancing Performance and Transparency. 2026 5th International Conference on Sentiment Analysis and Deep Learning (ICSADL). <https://doi.org/10.1109/icsadl67539.2026.11452051>
- [117] Jagatheesaperumal, Senthil Kumar, Mishra, Preeti, Moustafa, Nour, Chauhan, Rahul. (2022). A holistic survey on the use of emerging technologies to provision secure healthcare solutions. Computers and Electrical Engineering. <https://doi.org/10.1016/j.compeleceng.2022.107691>
- [118] Javed, Abdul Rehman, Shahzad, Faisal, Rehman, Saif ur, Zikria, Yousaf Bin, Razzak, Imran, Jalil, Zunera, Xu, Guandong. (2022). Future smart cities: requirements, emerging technologies, applications, challenges, and future aspects. Cities. <https://doi.org/10.1016/j.cities.2022.103794>
- [119] Javeed, Danish, Saeed, Muhammad Shahid, Ahmad, Ijaz, Adil, Muhammad, Kumar, Prabhat, Islam, A.K.M. Najmul. (2024). Quantum-empowered federated learning and 6G wireless networks for IoT security: Concept, challenges and future directions. Future Generation Computer Systems. <https://doi.org/10.1016/j.future.2024.06.023>
- [120] Jawad, Md Asef, Munna, Mohammad Mehedi Hasan, Kabir, Acramul Haque, Antu, Nahid Hossain, Tulona, Reefka Fabliha. (2026). A Runtime Safety Copilot for AI-Native O-RAN: Predictive Verification and Fail-Safe Enforcement in Near-RT RIC Control Loops. IEEE ACCESS, 14, 63106–63120. <https://doi.org/10.1109/access.2026.3686132>
- [121] Ji, Yangmin, Wang, Fan. (2025). Digital Twin-Based Predictive Maintenance in 6G Smart Stadiums Using Standardized AI-Driven Models. IEEE COMMUNICATIONS STANDARDS MAGAZINE. <https://doi.org/10.1109/mcomstd.2025.3610349>
- [122] Ahmad Awan, I. Ud Din, A. Almogren, M. Adnan, A. Altameem, I. Syed, S. Ahmad. (2026). TANF—Trustworthy Adaptive Neural Framework for Reliable and Scalable 6G Internet of Things. IEEE Internet of Things Journal. <https://doi.org/10.1109/jiot.2025.3554707>
- [123] K. Jain, P. Krishnan, P. Pachiyannan, L. Jaganathan, M. A. Khan, Y. Li. (2025). Toward Smart 5G and 6G: Standardization of AI-Native Network Architectures and Semantic Communication Protocols. IEEE Communications Standards Magazine. <https://doi.org/10.1109/mcomstd.2025.3618114>
- [124] K. Katsaros, I. Mavromatis, K. Antonakoglou, S. Ghosh, D. Kaleshi, T. Mahmoodi, H. Asgari, A. Karousos, I. Tavakkolnia, H. Safi, H. Hass, C. Vrontos, A. Emami, J. Marcelo Parra-Ullauri, S. Moazzeni, D. Simeonidou. (2024). AI-Native Multi-Access Future Networks—The REASON Architecture. IEEE Access. <https://doi.org/10.1109/access.2024.3507186>
- [125] K. Zhang, B. Zheng, J. Xue, Y. Zhou. (2026). Explainable and Trust-Aware AI-Driven Network Slicing Framework for 6G IoT Using Deep Learning. IEEE Internet of Things Journal. <https://doi.org/10.1109/jiot.2025.3619970>
- [126] Kamarajapandian, P., Nithya, S., Syed Haroon, A., Bharathi, V., Kaliappan, A., Meenachi, L.. (2026). Intent-based networking using LLM-driven agentic reinforcement learning for zero-touch 5G/6G and IoT systems. Computer Communications. <https://doi.org/10.1016/j.comcom.2026.108602>

- [127] Kanaki, Eirini, Hessami, Ali, Goncalves, Paulo, Chatzimisios, Periklis. (2026). Ethical and Privacy Issues for AI in 6G Empowering Robotics. *IEEE Communications Standards Magazine*, 10(2), 296–304. <https://doi.org/10.1109/mcomstd.2026.3658562>
- [128] Karabulut, Muhammet Ali, Shah, A. F. M. Shahan, Pathan, Al-Sakib Khan, Bradford, Phillip G.. (2026). IoT-Driven Intelligent Transportation System in the Era of 6G and AI: A Review. *Computers, Materials and Continua*. <https://doi.org/10.32604/cmc.2026.077625>
- [129] Karim, Muhammad Monjurul, Shi, Yan, Zhang, Shucheng, Wang, Bingzhang, Nasri, Mehrdad, Wang, Yin Hai. (2025). Large language models and their applications in roadway safety and mobility enhancement: A comprehensive review. *Artificial Intelligence for Transportation*. <https://doi.org/10.1016/j.ait.2025.100004>
- [130] Kayombo Sakachiva Jackson Phiri Mayumbu Nyirenda. (2026). A systematic review of intelligent wireless sensing architectures for AI-enabled maize disease monitoring in 5G and 6G networks. *Discover Networks*. <https://doi.org/10.1007/s44354-026-00034-1>
- [131] Khan, Nasir, Coleri, Sinem, Abdallah, Asmaa, Celik, Abdulkadir, Eltawil, Ahmed M.. (2024). Explainable and Robust Artificial Intelligence for Trustworthy Resource Management in 6G Networks. *IEEE Communications Magazine*, 62(4), 50–56. <https://doi.org/10.1109/mcom.001.2300172>
- [132] Khan, Salman, Loh, Woong-Kee, Ullah, Hamid, Khan, Javed Ali, Mylonas, Alexios. (2026). Causality aware explainable deep reinforcement learning with adaptive attention mechanisms for scalable resource orchestration in 6G wireless networks. *Computer Networks*. <https://doi.org/10.1016/j.comnet.2026.112074>
- [133] Kokkinis, Dimitris, Trachanas, Georgios P., Ioannou, Nikos, Katsianis, Dimitris, Varoutas, Dimitris. (2026). Assessing the 6G readiness across the European Union: A multi-criteria decision-making approach under incomplete information. *Telecommunications Policy*. <https://doi.org/10.1016/j.telpol.2026.103250>
- [134] Koukaras, Christos, Stavrinides, Stavros G., Hatzikraniotis, Euripides, Mitsiaki, Maria, Koukaras, Paraskevas, Tjortjis, Christos. (2026). Navigating the Future of Education: A Review on Telecommunications and AI Technologies, Ethical Implications, and Equity Challenges. *TELECOM*, 7(1), 2. <https://doi.org/10.3390/telecom7010002>
- [135] Kudakwashe Nyandoro Duy Hoang Pham Yong Han Ahn. (2026). Exploring the Synergy Between Blockchain Technology and 6G Networks: A Bibliometric and Systematic Literature Study. *Sustainable and Smart Cities: Ecological Design, Technological Innovation, and Climate Resilience*. https://doi.org/10.1007/978-3-032-00545-8_10
- [136] Kulkarni, Raghavendra, kumar Kondeti, Venkata Satya Suresh, Pillai, Binu Sudhakaran, Rajendran, Surendran. (2025). Blockchain-Assisted Improved Cryptographic Privacy-Preserving FL Model with Consensus Algorithm for ORAN. *Computers, Materials and Continua*. <https://doi.org/10.32604/cmc.2025.069835>
- [137] Kumar, Abhishek, Yadav, Jagat Pal, Maheshwari, Shubhrat, Singh, Aditya, Srivastava, Vineet, Khalilullah, Habibullah, Verma, Amita. (2025). Revolutionizing healthcare with 5 G and AI: Integrating emerging technologies for personalized care and cancer management. *Intelligent Hospital*. <https://doi.org/10.1016/j.inhs.2025.100005>
- [138] Kumar, Narendra, Kumar, Krishna, Aeron, Anurag, Verre, Filippo. (2025). Blockchain technology in supply chain management: Innovations, applications, and challenges. *Telematics and Informatics Reports*. <https://doi.org/10.1016/j.teler.2025.100204>

- [139] Kumar, Sudhakar, Singh, Sunil K., Kumar, Rakesh, Subba, Chandra Kumari, Chui, Kwok Tai, Gupta, Brij B.. (2025). Neoteric Threat Intelligence Ensuring Digital Sovereignty and Trust through ML-Infused Proactive Defense Analytics for NEXT-G and Beyond Ecosystems. International Conference on Digital Sovereignty (ICDS). <https://doi.org/10.1016/j.procs.2025.02.062>
- [140] Kumari, Aparna, Patadia, Divya, Tanwar, Sudeep, Pau, Giovanni, Alqahtani, Faye, Tolba, Amr. (2025). CNN-based cancer prediction scheme using 5G-assisted federated learning for healthcare Industry 5.0. Alexandria Engineering Journal. <https://doi.org/10.1016/j.aej.2025.04.043>
- [141] Lall, Shruti, Nyathi, Thambo, Pillay, Nelishia. (2026). Computational intelligence for 6G networks: A survey of machine learning and optimization for enabling technologies and architectures. Journal of Network and Computer Applications. <https://doi.org/10.1016/j.jnca.2026.104533>
- [142] Lamichhane, Badri Raj, Aueawattthanaphisut, Aueaphum, Srijuntongsiri, Gun, Horanont, Teerayut. (2026). Enhancing autonomous vehicle resilience: Roadside sensor networks for robust perception and decision-making in challenging environments. Array. <https://doi.org/10.1016/j.array.2026.100784>
- [143] Lamprou, Vangelis, Doukas, George, Ntanos, Christos, Askounis, Dimitris. (2025). On the trustworthiness of federated learning models for 5G network intrusion detection under heterogeneous data. Computer Networks. <https://doi.org/10.1016/j.comnet.2025.111598>
- [144] Le, Tan, Shetty, Sachin. (2022). Artificial intelligence-aided privacy preserving trustworthy computation and communication in 5G-based IoT networks. Ad Hoc Networks, 126, 102752. <https://doi.org/10.1016/j.adhoc.2021.102752>
- [145] Li, Chen, Guo, Weisi, Sun, Schyler Chengyao, Al-Rubaye, Saba, Tsourdos, Antonios. (2020). Trustworthy Deep Learning in 6G-Enabled Mass Autonomy: From Concept to Quality-of-Trust Key Performance Indicators. IEEE Vehicular Technology Magazine, 15(4), 112–121. <https://doi.org/10.1109/mvt.2020.3017181>
- [146] Li, Qinling. (2026). AI and Digital Twin-Enabled Zero-Touch 6G Architecture for Fraud-Resistant Financial Transaction Governance. IEEE Communications Standards Magazine. <https://doi.org/10.1109/mcomstd.2026.3654884>
- [147] Lifelo, Zita, Ding, Jianguo, Ning, Huansheng, Dhelim, Sahraoui. (2024). Artificial Intelligence-Enabled Metaverse for Sustainable Smart Cities: Technologies, Applications, Challenges, and Future Directions. ELECTRONICS, 13(24), 4874. <https://doi.org/10.3390/electronics13244874>
- [148] Lin, Jiaqi, Liu, Qi, Zhou, Zibo, Zhong, Xudong, Qian, Qijie, Qin, Zhen, Peng, Mugen, Ren, Baoquan. (2026). Cognitive sovereignty in 6G: architecting a Neuro-Symbolic knowledge plane for verifiable agentic automation. Digital Communications and Networks. <https://doi.org/10.1016/j.dcan.2026.04.004>
- [149] Liyanage, Madhusanka, Braeken, An, Shahabuddin, Shahriar, Ranaweera, Pasika. (2023). Open RAN security: Challenges and opportunities. Journal of Network and Computer Applications. <https://doi.org/10.1016/j.jnca.2023.103621>
- [150] -D. Nguyen, V. H. La, R. Cavalli, E. M. de Oca. (2022). Towards improving explainability, resilience and performance of cybersecurity analysis of 5G/IoT networks (work-in-progress paper). 2022 IEEE International Conference on Software Testing,

- Verification and Validation Workshops (ICSTW).
<https://doi.org/10.1109/icstw55395.2022.00016>
- [151] Asad, A. Fernández-Fernández, H. Chergui, M. Compastié, M. Montagud, S. Fernández, S. Siddiqui. (2023). 6GENABLERS: A Holistic Approach to Establish Pervasive Trust in 6G Networks. 2023 IEEE 28th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD).
<https://doi.org/10.1109/camad59638.2023.10478383>
- [152] M. AyyaduraiK. VijayK. LokeshP. Nandha KumarC. Hari Sainath. (2026). Cloud-to-Thing Continuum: Connecting IoT Devices and Cloud Infrastructures. Learning-Driven Data Fabrics for Sustainability. https://doi.org/10.1007/978-3-032-09090-4_2
- [153] M. Chen, Y. Gu, X. He, S. Wang, F. Huang, L. Wang. (2026). Explainable Artificial Intelligence Enhance Image Semantic Communication System in 6G-IoT. IEEE Internet of Things Journal. <https://doi.org/10.1109/jiot.2025.3606530>
- [154] M. Chiarani, S. Roy, C. Verikoukis, F. Granelli. (2025). XAI-Driven Client Selection for Federated Learning in Scalable 6G Network Slicing. ICC 2025 - IEEE International Conference on Communications. <https://doi.org/10.1109/icc52391.2025.11161532>
- [155] M. Dua, T. Kundu, A. Gupta. (2026). Trustworthy Autonomous RAN Orchestration: An Architectural Framework for Zero-Trust Intent-Based Control with Explainable AI. 2026 35th Wireless and Optical Communications Conference (WOCC).
<https://doi.org/10.1109/wocc69802.2026.11556178>
- [156] M. Fakhrul Safitra, D. Prastika Putri, M. Lubis, H. Fakhrurroja, F. Saidi Lubis, V. Satria Praditha. (2023). Intelligent Computing in Communication Networks: Challenges, Solutions, and Research Prospects. 2023 IEEE International Conference on Computing (ICOCO). <https://doi.org/10.1109/icoco59262.2023.10397643>
- [157] M. K. Nallakaruppan, R. Kumar Dhanaraj, S. Shukla, S. Krishnamoorthi, R. Kumar Kaushal, M. Kumar Goyal, S. Basheer, M. Tabrez Quasim. (2026). A Federated Autoencoder Framework With Explainable AI for Intelligent 6G-IoT Infrastructure Optimization. IEEE Internet of Things Journal. <https://doi.org/10.1109/jiot.2025.3606232>
- [158] M. Maduranga, H. Lakmal, U. Oruthota, V. Tilwari, D. Sharma, S. Gayan. (2025). 5G Network Slicing Prediction Using Explainable Artificial Intelligence(X-AI). 2025 10th International Conference on Signal Processing and Communication (ICSC).
<https://doi.org/10.1109/icsc64553.2025.10968200>
- [159] M. Teraiya, M. Shukla, S. B. Khan. (2026). A Multi-Agent Federated Learning Framework for Privacy-Aware Multimodal Wearable Edge Intelligence in 6G Networks. IEEE Open Journal of the Communications Society. <https://doi.org/10.1109/ojcoms.2026.3701757>
- [160] Mahbub, Mobasshir, Saym, Mir Md., Jahan, Sarwar, Paul, Anup Kumar, Vahid, Alireza, Hosseinalipour, Seyyedali, Barua, Bobby, Yeh, Hen-Geul, Shubair, Raed M., Taleb, Tarik, Kaushik, Aryan, Alsharif, Mohammed H., Islam, M. Shariful, Mahmud, Russel Reza, Niyato, Dusit. (2025). UAV-assisted wireless communications in the 6G-and-beyond era: An extensive survey on characteristics, standardization and regulations, enabling technologies, challenges, and future directions. Vehicular Communications. <https://doi.org/10.1016/j.vehcom.2025.100977>

- [161] Maroua, Drissi. (2024). A state-of-the-art on federated learning for vehicular communications. *Vehicular Communications*.
<https://doi.org/10.1016/j.vehcom.2023.100709>
- [162] Mbunge, Elliot, Jiyane, Sipho'esihle, Muchemwa, Benhildah. (2022). Towards emotive sensory Web in virtual health care: Trends, technologies, challenges and ethical issues. *Sensors International*. <https://doi.org/10.1016/j.sintl.2021.100134>
- [163] Mekrache, Abdelkader, Mekki, Mohamed, Ksentini, Adlen, Brik, Bouziane, Verikoukis, Christos. (2025). On Combining XAI and LLMs for Trustworthy Zero-Touch Network and Service Management in 6G. *IEEE Communications Magazine*, 63(4), 154–160.
<https://doi.org/10.1109/mcom.002.2400276>
- [164] Miglani, Arzoo, Kumar, Neeraj. (2021). Blockchain management and machine learning adaptation for IoT environment in 5G and beyond networks: A systematic review. *Computer Communications*. <https://doi.org/10.1016/j.comcom.2021.07.009>
- [165] Min, Min, Xia, Liping. (2026). Standards-Driven Security and AI Governance for 6G Edge Wearable Digital Twins for Mental-Health Monitoring in University Campuses. *IEEE Communications Standards Magazine*.
<https://doi.org/10.1109/mcomstd.2026.3677758>
- [166] Mirko FarinaAndrea Lavazza. (2026). 6G and Embodied AI: Benefits and Risks of Increasingly Extended Minds. *Neuroethics*. <https://doi.org/10.1007/s12152-026-09632-w>
- [167] Moreira, Rodrigo, Villaça, Rodolfo S., Ribeiro, Moisés R.N., Martins, Joberto S.B., Corrêa, João Henrique, Carvalho, Tereza C., de Oliveira Silva, Flávio. (2025). An intelligent native network slicing security architecture empowered by federated learning. *Future Generation Computer Systems*. <https://doi.org/10.1016/j.future.2024.107537>
- [168] Morgan, Ernest Fiko, Ali, Mohd. Hasan. (2025). Digital Twin-Driven Cybersecurity for 5G/6G-Enabled Electric Vehicle Charging Infrastructure: A Review. *ENERGIES*, 18(22), 6048. <https://doi.org/10.3390/en18226048>
- [169] Munir, Md. Shirajum, Kim, Ki Tae, Adhikary, Apurba, Saad, Walid, Shetty, Sachin, Park, Seong-Bae, Hong, Choong Seon. (2023). Neuro-Symbolic Explainable Artificial Intelligence Twin for Zero-Touch IoE in Wireless Network. *IEEE Internet of Things Journal*, 10(24), 22451–22468. <https://doi.org/10.1109/jiot.2023.3303713>
- [170] Munir, Muhammad Umair, Ahmedy, Ismail, Azzuhri, Saaidal Razalli, Raj, Bryan, Noor, Rafidah Md, Ali, Ihsan. (2026). A systematic review of handover optimization techniques in 5G self-organizing networks. *ICT Express*. <https://doi.org/10.1016/j.icte.2026.03.020>
- [171] Gorowara, D. Singh, A. KUMAR, S. Sharma, P. Singh. (2026). 6G Networks and Human-Centric Communication Social, Ethical, and Policy Challenges. 2026 5th International Conference on Innovative Practices in Technology and Management (ICIPTM). <https://doi.org/10.1109/iciptm69057.2026.11465447>
- [172] N. K. Bairwa, G. Jain, S. Umrao, P. Ponde. (2024). Impact of AI Driven Predictive Privacy Mechanisms in UAV Integrated V2X 6G Networks. 2024 International Conference on Emerging Technologies and Innovation for Sustainability (EmergIN). <https://doi.org/10.1109/emergin63207.2024.10961785>
- [173] N. Mazhar, S. Asad Ullah, S. Basheer, H. Jung, M. S. J. Solaija, A. Mahmood, M. Gidlund, S. Ali Hassan. (2026). Toward Trustworthy and Fresh Data Delivery in 6G IoT: A DRL-Aided Cognitive NOMA and Backscatter Framework. *IEEE Internet of Things Journal*. <https://doi.org/10.1109/jiot.2025.3611868>

- [174] N. Meenakshisundaram, G. Sajiv. (2025). Deploying Secure and Interpretable Medical AI Models on Edge Devices: A 6G-Enabled FL-GAN Framework. 2025 6th International Conference on Electronics and Sustainable Communication Systems (ICESC). <https://doi.org/10.1109/icesc65114.2025.11212235>
- [175] Narteni, Sara, Muselli, Marco, Dabbene, Fabrizio, Mongelli, Maurizio. (2023). Trustworthy artificial intelligence classification-based equivalent bandwidth control. Computer Communications. <https://doi.org/10.1016/j.comcom.2023.07.005>
- [176] Nassef, Omar, Sun, Wenting, Purmehdi, Hakimeh, Tatipamula, Mallik, Mahmoodi, Toktam. (2022). A survey: Distributed Machine Learning for 5G and beyond. Computer Networks. <https://doi.org/10.1016/j.comnet.2022.108820>
- [177] Nasser, Nidal, Fadlullah, Zubair Md, Fouda, Mostafa M., Ali, Asmaa, Imran, Muhammad. (2022). A lightweight federated learning based privacy preserving B5G pandemic response network using unmanned aerial vehicles: A proof-of-concept. Computer Networks. <https://doi.org/10.1016/j.comnet.2021.108672>
- [178] Nawaz, Muhammad Waqas, Zhang, Wanquan, Flynn, David, Zhang, Lei, Swash, Rafiq, Abbasi, Qammer H., Imran, Muhammad Ali, Popoola, Olaoluwa. (2024). 6G edge-networks and multi-UAV knowledge fusion for urban autonomous vehicles. Physical Communication. <https://doi.org/10.1016/j.phycom.2024.102479>
- [179] Nezhadsistani, Nasim, Tajgardan, Mohsen, Rabbani, Mahdi, Niu, Weijie, Stiller, Burkhard. (2026). Design and evaluation of a robust and explainable intrusion detection framework for 5G/B5G networks. Computer Networks. <https://doi.org/10.1016/j.comnet.2026.112100>
- [180] Noor, Kinzah, Imoize, Agbotiname Lucky, Adedosu Adelabu, Michael, Lee, Cheng-Chi. (2025). A Comprehensive Survey on AI-Assisted Multiple Access Enablers for 6G and beyond Wireless Networks. CMES - Computer Modeling in Engineering and Sciences. <https://doi.org/10.32604/cmcs.2025.073200>
- [181] Noorani, Naserali, Hosseini Seno, Seyed Amin. (2026). Intelligent fluid computing framework for XAI-enhanced task offloading and resource management in vehicular networks. Vehicular Communications. <https://doi.org/10.1016/j.vehcom.2026.101019>
- [182] Ekpo, V. Casola, P. Asuquo, B. Agbor. (2026). An Explainable Deep Learning-based Intrusion Detection Framework for Smart Mobility. 2026 IEEE 4th International Conference on Mobility, Operations, Services and Technologies (MOST). <https://doi.org/10.1109/most69733.2026.00025>
- [183] O. J. Adeleke, K. D. Jovanovich, S. Ogunbunmi, O. Samuel, T. O. Kehinde. (2025). Comprehensive Exploration of Smart Cities: A Systematic Review of Benefits, Challenges, and Future Directions in Telecommunications and Urban Development. IEEE Sensors Reviews. <https://doi.org/10.1109/sr.2025.3569239>
- [184] O. T. Basaran, F. Dressler. (2024). XAIomaly: Explainable, Interpretable and Trustworthy AI for xURLLC in 6G Open-RAN. 2024 3rd International Conference on 6G Networking (6GNet). <https://doi.org/10.1109/6gnet63182.2024.10765734>
- [185] Ogenyi, Fabian Chukwudi, Ugwu, Chinyere Nneoma, Ugwu, Okechukwu Paul-Chima. (2025). A comprehensive review of AI-native 6G: integrating semantic communications, reconfigurable intelligent surfaces, and edge intelligence for next-generation connectivity. FRONTIERS IN COMMUNICATIONS AND NETWORKS, 6, 1655410. <https://doi.org/10.3389/frcmn.2025.1655410>

- [186] Omar I. Alsaleh. (2026). Large language models for DAMA-aligned data management in telecommunication systems: a review of OSS/BSS and 5G/6G applications. *Telecommunication Systems*. <https://doi.org/10.1007/s11235-026-01462-8>
- [187] Ovidiu Vermesan, Reiner John Patrick, Pyper Gerardo, Daalderop Meghashyam, Ashwathnarayan Roy, Bahr Tore, Karlsen Hans-Erik, Sand. (2021). Internet of Vehicles – System of Systems Distributed Intelligence for Mobility Applications. *Intelligent Technologies for Internet of Vehicles*. https://doi.org/10.1007/978-3-030-76493-7_4
- [188] Barnard, I. Macaluso, N. Marchetti, L. A. DaSilva. (2022). Resource Reservation in Sliced Networks: An Explainable Artificial Intelligence (XAI) Approach. *ICC 2022 - IEEE International Conference on Communications*. <https://doi.org/10.1109/icc45855.2022.9838766>
- [189] P. Radoglou-Grammatikis, E. Katsaros, E. Syrmos, G. P. Katsikas, D. Klonidis, V. Argyriou, T. Lagkas, P. Sarigiannidis. (2025). Trustworthy Analytics in ETSI ZSM: A 5G Security Case Study. *IEEE Open Journal of the Communications Society*. <https://doi.org/10.1109/ojcoms.2024.3505555>
- [190] Pan, Xuanchen. (2025). Detection and Compliance in Financial Management via 6G-Enabled Standardized AI-Powered Digital Twins. *IEEE Communications Standards Magazine*. <https://doi.org/10.1109/mcomstd.2025.3639269>
- [191] Parra-Ullauri, Juan, Zhou, Xueqing, Moazzeni, Shadi, Hussain, Rasheed, Vasilakos, Xenofon, Wu, Yulei, Baby, Renjith, Hassan Mahmud, M. M., Incorvaia, Gabriele, Hond, Darryl, Asgari, Hamid, Tassi, Andrea, Warren, Daniel, Simeonidou, Dimitra. (2025). Lifecycle Management of Trustworthy AI Models in 6G Networks: the Reason Approach. *IEEE Wireless Communications*, 32(2), 42–51. <https://doi.org/10.1109/mwc.001.2400288>
- [192] Paul, Anal, Singh, Keshav. (2026). Trustworthy AI for 6G-IoV: A Privacy-Preserved Distributed Multiagent Federated DRL for Dynamic Electric Vehicle Charging and Task Offloading. *IEEE Internet of Things Journal*, 13(5), 7785–7800. <https://doi.org/10.1109/jiot.2025.3572893>
- [193] Prabha, M., Nandhini, S., Dayanidhy, M., Pradeep, R.. (2025). Edge-AI integrated secure wireless IoT architecture for real time healthcare monitoring and federated anomaly detection. *SCIENTIFIC REPORTS*, 16(1), 574. <https://doi.org/10.1038/s41598-025-30150-x>
- [194] Priyanka Mishra, Ghanshyam Singh. (2025). Innovative Solutions for Smart Cities with Internet of Things. *Sustainable Smart Cities 2.0*. https://doi.org/10.1007/978-3-032-01102-2_3
- [195] Priyanka Mishra, Ghanshyam Singh. (2025). Smart Cities with 6G-Internet of Things. *Sustainable Smart Cities 2.0*. https://doi.org/10.1007/978-3-032-01102-2_6
- [196] Prodomo, Vittorio, González, Roberto, Sperli, Giancarlo, Romano, Simon Pietro. (2026). Towards the integration of Privacy-Preserving technologies in future mobile networking. *Computer Communications*. <https://doi.org/10.1016/j.comcom.2026.108445>
- [197] Qayyum, Adnan, Butt, Muhammad Atif, Ali, Hassan, Usman, Muhammad, Halabi, Osama, Al-Fuqaha, Ala, Abbasi, Qammer H., Imran, Muhammad Ali, Qadir, Junaid. (2024). Secure and Trustworthy Artificial Intelligence-extended Reality (AI-XR) for Metaverses. *ACM COMPUTING SURVEYS*, 56(7), 170. <https://doi.org/10.1145/3614426>

- [198] Qimei CuiXiaohu YouNi WeiGuoshun NanXuefei ZhangJianhua ZhangXinchen LyuMing AiXiaofeng TaoZhiyong FengPing ZhangQingqing WuMeixia TaoYongming HuangChongwen HuangGuangyi LiuChenghui PengZhiwen PanTao SunDusit NiyatoTao ChenMuhammad Khurram KhanAbbas JamalipourMohsen GuizaniChau Yuen. (2025). Overview of AI and communication for 6G network: fundamentals, challenges, and future research opportunities. *Science China Information Sciences*. <https://doi.org/10.1007/s11432-024-4337-1>
- [199] Chaparadza, M. Elkotob, B. Radier, T. B. Meriem, E. Hammad, T. Choi. (2022). Autonomic/Autonomous Networking (AN): Federations & Governance of ANs, Progress & Open Challenges in Standards: Analysis Report. 2022 IEEE Future Networks World Forum (FNWF). <https://doi.org/10.1109/fnwf55208.2022.00039>
- [200] Chaparadza, M. Elkotob, B. Radier, T. Ben Meriem, T. Choi, R. B. Bohn, A. Battou, T. Zhang. (2022). The Era of Knowledge Plane (KP) Platforms Driven Networking—Anchor for Federation of Autonomic/Autonomous Networks (ANs) Across Industry Sectors. 2022 13th International Conference on Information and Communication Technology Convergence (ICTC). <https://doi.org/10.1109/ictc55196.2022.9952650>
- [201] R. Zhang, S. Wu, C. Jiang, N. Gao, X. Qiu, W. Zhang. (2024). Trustworthy and Scalable Federated Edge Learning for Future Integrated Positioning, Communication, and Computing System: Attacks and Defenses. *IEEE Internet of Things Journal*. <https://doi.org/10.1109/jiot.2023.3330166>
- [202] Rafique, Wajid, Qadir, Junaid. (2024). Internet of everything meets the metaverse: Bridging physical and virtual worlds with blockchain. *Computer Science Review*. <https://doi.org/10.1016/j.cosrev.2024.100678>
- [203] Rago, Arcangela, Fascista, Alessio, Piro, Giuseppe, Boggia, Gennaro, Grieco, Luigi Alfredo. (2026). Internet of Medical Things Through Explainable AI at the Edge: A Practical Case Study. *IEEE INTERNET OF THINGS MAGAZINE*. <https://doi.org/10.1109/miot.2026.3673130>
- [204] Rahman, Anichur, Hasan, Kamrul, Kundu, Dipanjali, Islam, Md. Jahidul, Debnath, Tanoy, Band, Shahab S., Kumar, Neeraj. (2023). On the ICN-IoT with federated learning integration of communication: Concepts, security-privacy issues, applications, and future perspectives. *Future Generation Computer Systems*. <https://doi.org/10.1016/j.future.2022.08.004>
- [205] Rahman, Md Abdur, Hossain, M. Shamim, Showail, Ahmad J., Alrajeh, Nabil A., Alhamid, Mohammed F.. (2021). A secure, private, and explainable IoHT framework to support sustainable health monitoring in a smart city. *Sustainable Cities and Society*. <https://doi.org/10.1016/j.scs.2021.103083>
- [206] Rahman, Md Mijanur, Khatun, Fatema, Sami, Sadia Islam, Uzzaman, Ashik. (2022). The evolving roles and impacts of 5G enabled technologies in healthcare: The world epidemic COVID-19 issues. *Array*. <https://doi.org/10.1016/j.array.2022.100178>
- [207] Rasheed, Iftikhar, Mostafa, Hala. (2025). Federated learning-based anomaly detection for zero-day attack prevention in 6G network slices. *Computer Networks*. <https://doi.org/10.1016/j.comnet.2025.111376>
- [208] Rashid, Harun Ur, Jeong, Seong Ho. (2025). AI empowered 6G technologies and network layers: Recent trends, opportunities, and challenges. *Expert Systems with Applications*, 267, 125985. <https://doi.org/10.1016/j.eswa.2024.125985>

- [209] Razafimanjato, Mahalinoro, Saad, Malik Muhammad, Kim, Dongkyun. (2025). Blockchain-based trust management systems in the Internet of Vehicles: A comprehensive survey. *ICT Express*. <https://doi.org/10.1016/j.ict.2025.09.012>
- [210] Reis, Manuel J. C. S., Gupta, Nishu. (2025). Federated Learning for Vision-Based Applications in 6G Networks: A Simulation-Based Performance Study. *CMES - Computer Modeling in Engineering and Sciences*. <https://doi.org/10.32604/cmes.2025.073366>
- [211] Rezaei, Hadiseh, Taheri, Rahim, Shojafar, Mohammad. (2025). FedLLMGuard: A federated large language model for anomaly detection in 5G networks. *Computer Networks*. <https://doi.org/10.1016/j.comnet.2025.111473>
- [212] Roy, Swastika, Chergui, Hatim, Verikoukis, Christos. (2024). Explanation-Guided Fair Federated Learning for Transparent 6G RAN Slicing. *IEEE Transactions on Cognitive Communications and Networking*, 10(6), 2269–2281. <https://doi.org/10.1109/tccn.2024.3400524>
- [213] Roy, Swastika, Chergui, Hatim, Verikoukis, Christos. (2024). Toward Bridging the FL Performance-Explainability Tradeoff: A Trustworthy 6G RAN Slicing Use-Case. *IEEE Transactions on Vehicular Technology*, 73(7), 10529–10538. <https://doi.org/10.1109/tvt.2024.3364363>
- [214] Barmounakis, P. Demestichas. (2022). Framework for Trustworthy AI/ML in B5G/6G. 2022 1st International Conference on 6G Networking (6GNet). <https://doi.org/10.1109/6gnet54646.2022.9830408>
- [215] S. Caner-Yildirim, I. Corradini, V. Dimitrova, V. Loscri, M. Harizaj. (2025). Gender and Intersectional Bias in Cybersecurity for Next Generation Wireless Networks. 2025 IEEE Conference on Communications and Network Security (CNS). <https://doi.org/10.1109/cns66487.2025.11194920>
- [216] S. H. Patel. (2026). Robust and Explainable RIC-Based AI Controllers for 6G RAN Automation. SoutheastCon 2026. <https://doi.org/10.1109/southeastcon63549.2026.11476692>
- [217] S. Hashini, M. V. S. L. Tejasri, R. Dhivya Bharathi, P. Harini, B. Yuvasri, P. J. Beslin Pajila. (2025). Mobile Computing and Intelligence: Integrating AI to Revolutionize Real-Time Applications and Future Innovations. 2025 6th International Conference on Electronics and Sustainable Communication Systems (ICESC). <https://doi.org/10.1109/icesc65114.2025.11212332>
- [218] S. Prakash, R. Prasad, A. Agarwal. (2026). Nishpaksh: TEC Standard-Compliant Framework for Fairness Auditing and Certification of AI Models. 2026 18th International Conference on COMMunication Systems and NETWORKS (COMSNETS). <https://doi.org/10.1109/comsnets67989.2026.11418125>
- [219] S. S. Shinde. (2025). Revolutionizing Electronics and Telecommunication: The Impact of AI and Machine Learning. 2025 1st International Conference on AIML-Applications for Engineering & Technology (ICAET). <https://doi.org/10.1109/icaet63349.2025.10932259>
- [220] S. Sinha, J. W. Lambrechts, A. Bimana, A. Ashipala. (2025). Augmented Artificial Intelligence in 5G, 6G, and Beyond: A Quantum Leap. *Computer*, 58(1), 24–32. <https://doi.org/10.1109/mc.2024.3445532>

- [221] S. Wang, C. Sandeepa, T. Senevirathna, B. Siniarski, M. -D. Nguyen, S. Marchal, M. Liyanage. (2024). Towards Accountable and Resilient AI-Assisted Networks: Case Studies and Future Challenges. 2024 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit). <https://doi.org/10.1109/eucnc/6gsummit60053.2024.10597060>
- [222] Saeedi Taleghani, Elmira, Lucila Sandoval Orozco, Ana, Javier García Villalba, Luis. (2026). FBBTrust: Decentralized trust modeling in 6G via fuzzy inference, BiLSTM, and blockchain. Computer Networks. <https://doi.org/10.1016/j.comnet.2025.111868>
- [223] Saheed, Yakub Kayode, Chukwuere, Joshua Ebere. (2024). XAIEnsembleTL-IoV: A new eXplainable Artificial Intelligence ensemble transfer learning for zero-day botnet attack detection in the Internet of Vehicles. Results in Engineering. <https://doi.org/10.1016/j.rineng.2024.103171>
- [224] Sai, Akshita Maradapu Vera Venkata, Wang, Chenyu, Cai, Zhipeng, Li, Yingshu. (2024). Navigating the Digital Twin Network landscape: A survey on architecture, applications, privacy and security. High-Confidence Computing. <https://doi.org/10.1016/j.hcc.2024.100269>
- [225] Sajid, Asher, Lipman, Justin, Abolhasan, Mehran. (2026). Towards SMPC-enabled O-RAN: A survey with deployment-oriented insights. Computer Networks. <https://doi.org/10.1016/j.comnet.2025.111816>
- [226] Samruddhi Phalak. (2025). Smart Governance for Urban Digital Twin (DT) Technology: Implementation, Challenges, and Opportunities. Environmentally Sustainable and Socially Resilient Urban Development and Management. https://doi.org/10.1007/978-3-031-92850-5_18
- [227] Sandeepa, Chamara, Zeydan, Engin, Samarasinghe, Tharaka, Liyanage, Madhusanka. (2025). Federated Learning for 6G Networks: Navigating Privacy Benefits and Challenges. IEEE Open Journal of the Communications Society, 6, 90–129. <https://doi.org/10.1109/ojcoms.2024.3513832>
- [228] Saqib, Sheikh Muhammad, Mazhar, Tehseen, Tariq, Muhammad Usman, Shahzad, Tariq, AlAlwan, Asem Ibrahim, Hamam, Habib. (2026). Agentic latency control and cooperative vehicle coordination in 5G-MEC: A prescriptive and explainable AI framework. Computer Communications. <https://doi.org/10.1016/j.comcom.2026.108433>
- [229] Sathya, K., Magesh, G.. (2026). Explainable retinal deep learning for cardiovascular risk stratification: a multiple modality analysis framework with vascular-centric interpretability and robustness. BMC MEDICAL INFORMATICS AND DECISION MAKING, 26(1), 235. <https://doi.org/10.1186/s12911-026-03516-y>
- [230] Schuh, Günther, Stroh, Max-Ferdinand, Johanning, Lara. (2023). Framework To Design Compliance Rules For Digital Technologies In Manufacturing Companies. The 33rd CIRP Design Conference. <https://doi.org/10.1016/j.procir.2023.04.005>
- [231] Senbagavalli, M., Debnath, Saswati, Vaneeta, M., Arumugam, Thamodharan, Beena, B. M., Prajapat, Sunil. (2026). Adaptive Semantic Integrity and Privacy Shielding in Agentic AI Networks for Trustworthy 6G Cognitive Communication System. IEEE Communications Standards Magazine. <https://doi.org/10.1109/mcomstd.2026.3661577>
- [232] Senevirathna, Thulitha, La, Vinh Hoa, Marcha, Samuel, Siniarski, Bartlomiej, Liyanage, Madhusanka, Wang, Shen. (2025). A Survey on XAI for 5G and Beyond Security: Technical Aspects, Challenges and Research Directions. IEEE COMMUNICATIONS

SURVEYS AND TUTORIALS, 27(2), 941–973.
<https://doi.org/10.1109/comst.2024.3437248>

- [233] Shafee, Ahmed, Hasan, S.R., Awaad, Tasneem A.. (2025). Privacy and security vulnerabilities in edge intelligence: An analysis and countermeasures. Computers and Electrical Engineering. <https://doi.org/10.1016/j.compeleceng.2025.110146>
- [234] Shahrour, Gheida, Junejo, Aisha, Al-Said Ahmad, Amro. (2026). Security and privacy challenges in healthcare IoT systems with cloud, edge, and fog computing: A systematic literature review. Computer Science Review. <https://doi.org/10.1016/j.cosrev.2026.101005>
- [235] Sharma, Parjanay, Jain, Siddhant, Gupta, Shashank, Chamola, Vinay. (2021). Role of machine learning and deep learning in securing 5G-driven industrial IoT applications. Ad Hoc Networks. <https://doi.org/10.1016/j.adhoc.2021.102685>
- [236] Sheth, Karan, Patel, Keyur, Shah, Het, Tanwar, Sudeep, Gupta, Rajesh, Kumar, Neeraj. (2020). A taxonomy of AI techniques for 6G communication networks. Computer Communications. <https://doi.org/10.1016/j.comcom.2020.07.035>
- [237] Shome, Subhankar, Das, Suman, Das, Saumya, Pal, Debashish. (2025). An extensive review of THz communication in 6G: Facilitating technologies with edge computing and native AI. Franklin Open. <https://doi.org/10.1016/j.fraope.2025.100434>
- [238] Singh, Rohit, Bodile, Roshan, Kaushik, Aryan, Dolas, Amit, Kaur, Amandeep, Chatzimisios, Periklis. (2026). Integrated AI and 6G Driven e-Health: A Trifecta for Smart Healthcare. IEEE COMMUNICATIONS STANDARDS MAGAZINE, 10(2), 341–350. <https://doi.org/10.1109/mcomstd.2025.3645228>
- [239] Sinha, Shivangi, Singh, Manya, Chowdhury, Arpita Roy, Venkatesan, RajeshKumar, Rajamani, Vetriveeran, Udayakumar, Ashwini, Selvaprabhu, Poongundran. (2026). Intelligent V2X communications in the 6G era: Handover, sustainability and resource allocation for next-generation ITS. PHYSICAL COMMUNICATION, 78, 103232. <https://doi.org/10.1016/j.phycom.2026.103232>
- [240] Smidt, Hermanus J, Jokonya, Osden. (2021). The challenge of privacy and security when using technology to track people in times of COVID-19 pandemic. CENTERIS 2020 - International Conference on ENTERprise Information Systems / ProjMAN 2020 - International Conference on Project MANagement / HCist 2020 - International Conference on Health and Social Care Information Systems and Technologies 2020, CENTERIS/ProjMAN/HCist 2020. <https://doi.org/10.1016/j.procs.2021.01.281>
- [241] Soto, Paola, Camelo, Miguel, García-Avilés, Ginés, Municio, Esteban, Gramaglia, Marco, Kosmatos, Evangelos, Slamnik-Kriještorac, Nina, De Vleeschauwer, Danny, Bazco-Nogueras, Antonio, Fuentes, Lidia, Ballesteros, Joaquin, Lutu, Andra, Cominardi, Luca, Paez, Ivan, Alcalá-Marín, Sergi, Chatzieftheriou, Livia Elena, García-Saavedra, Andrés, Fiore, Marco. (2024). Designing the Network Intelligence Stratum for 6G networks. Computer Networks. <https://doi.org/10.1016/j.comnet.2024.110780>
- [242] Soyi JungSoo Hyun ParkSeok Bin SonHaemin LeeJoongheon Kim. (2024). Network Security and Trustworthiness. Fundamentals of 6G Communications and Networking. https://doi.org/10.1007/978-3-031-37920-8_30
- [243] Sun, Haochen, Liu, Yifan, Al-Tahmeesschi, Ahmed, Nag, Avishek, Soleimanpour, Mohadeseh, Canberk, Berk, Arslan, Huseyin, Ahmadi, Hamed. (2025). Advancing 6G: Survey for Explainable AI on Communications and Network Slicing. IEEE Open Journal

- of the Communications Society, 6, 1372–1412.
<https://doi.org/10.1109/ojcoms.2025.3534626>
- [244] Sun, Yuanyuan, Liu, Jiajia, Wang, Jiadai, Cao, Yurui, Kato, Nei. (2020). When Machine Learning Meets Privacy in 6G: A Survey. *IEEE COMMUNICATIONS SURVEYS AND TUTORIALS*, 22(4), 2694–2724. <https://doi.org/10.1109/comst.2020.3011561>
- [245] Lili, Z. Jia, D. Ying, W. Nan. (2024). Research on risk identification and industrial governance of digital education products based on data annotation technology. *China Communications*. <https://doi.org/10.23919/jcc.fa.2023-0708.202403>
- [246] T. Yang, J. Wang. (2026). BHMFL: Autonomous Federated Learning for Privacy-Aware Health Monitoring via Self-Balancing Data Synthesis in Smart 6G-HIoT Systems. *IEEE Internet of Things Journal*. <https://doi.org/10.1109/jiot.2025.3611403>
- [247] Taleb, Houssein, Andrieux, Guillaume, Khalife, Daniele, Ajami, Alain, Nasser, Abbass. (2025). Exploring healthcare in the 6G and AI era: Opportunities and challenges. *Internet of Things*, 34, 101744. <https://doi.org/10.1016/j.iot.2025.101744>
- [248] Tang, Ning, Luo, Wang, Wang, Yiwei, Feng, Bao, Yang, Shuang, Xu, Jiangtao, Zhu, Daohua, Huang, Zhechen, Liang, Wei. (2025). FedEPC: An Efficient and Privacy-Enhancing Clustering Federated Learning Method for Sensing-Computing Fusion Scenarios. *Computers, Materials and Continua*. <https://doi.org/10.32604/cmc.2025.066241>
- [249] Telagam, Nagarjuna, Kandasamy, Nehru, Manoharan, Arun Kumar, Anandhi, Palani, Atchudan, Raji. (2025). Beyond 5G: Exploring key enabling technologies, use cases, and future prospects of 6 G communication. *Nano Communication Networks*. <https://doi.org/10.1016/j.nancom.2024.100560>
- [250] Toumi, Nassima, Caceres, Joaquin, Lamela, Fernando, Rodriguez, Santiago, Ayimba, Constantine, Tzanettis, Ioannis, Zafeiropoulos, Anastasios, Karacay, Leyli, Lamprousi, Vasiliki, Karaca, Mehmet, Ruggeri, Franco, Pavon, Ignacio Labrador. (2026). Toward an Intelligent Energy Efficient and Trustworthy 6G M&O Architecture. *IEEE Communications Magazine*, 64(2). <https://doi.org/10.1109/mcom.001.2500296>
- [251] Tulaskar, Dhiraj P., Sindhu, Battina, Chakole, Nitin, Parteki, Rina, Leema, A. Anny, Balakrishnan, P., Avthanka, Ankita, Girhe, Rangnath, Kulkarni, Madhusudan B., Bhaiyya, Manish. (2026). AI and ML empowering 5G and shaping the 6G future: Models, metrics, architectures, and applications. *ICT Express*. <https://doi.org/10.1016/j.icte.2025.11.024>
- [252] S. S, N. K. V, D. R, T. B, A. T P, N. M. (2025). Internet of Things in the Era of Smart Applications: Emerging Trends and Future Challenges. 2025 5th International Conference on Evolutionary Computing and Mobile Sustainable Networks (ICECMSN). <https://doi.org/10.1109/icecmsn68058.2025.11383314>
- [253] Uccello, Federica, Nadjm-Tehrani, Simin. (2026). Feature attribution in 5G intrusion detection: A statistical vs. logic-based comparison. *Journal of Information Security and Applications*. <https://doi.org/10.1016/j.jisa.2026.104448>
- [254] Ullah, Arif, Fawad, Nadeem, Aamir, Arif, Muhammad, Bashir, Muhammad Mehran, Choi, Wooyeol. (2025). 6G Internet-of-Things assisted smart homes and buildings: Enabling technologies, opportunities and challenges. *Internet of Things*. <https://doi.org/10.1016/j.iot.2025.101658>

- [255] Ullah, Inam, Arishi, Ali, Singh, Sushil Kumar, Alharbi, Faisal, Ibrahim, Anwar Hassan, Islam, Muhammad, Daradkeh, Yousef Ibrahim, Choi, Chang. (2025). Autonomous network management for 6G communication: A comprehensive survey. Digital Communications and Networks. <https://doi.org/10.1016/j.dcan.2025.07.001>
- [256] Ullah, Inam, Haider, Zeeshan Ali, Su, Xin, Choi, Chang, Arishi, Ali, Bilal, Hazrat. (2026). Vehicular communications for autonomous and intelligent transportation systems: a comprehensive survey. Digital Communications and Networks. <https://doi.org/10.1016/j.dcan.2026.06.003>
- [257] Usman, Yusuf, Oladipupo, Habeeb, During, Adegboyega D., Akl, Robert, Chataut, Robin. (2025). AI, ML, and LLM Integration in 5G/6G Networks: A Comprehensive Survey of Architectures, Challenges, and Future Directions. IEEE Access, 13, 168914–168950. <https://doi.org/10.1109/access.2025.3608736>
- [258] Sridhar, A. Mishra, A. Iqubbal. (2025). Analysis of Techno-Social and Regulatory Challenges in Future 6G Networks and Services. IEEE Transactions on Technology and Society. <https://doi.org/10.1109/tts.2025.3621860>
- [259] V. Ziegler, P. Schneider, H. Viswanathan, M. Montag, S. Kanugovi, A. Rezaki. (2021). Security and Trust in the 6G Era. IEEE Access. <https://doi.org/10.1109/access.2021.3120143>
- [260] Valdez-Solis, Wilson, Parra-Ullauri, Juan, Simeonidou, Dimitra, Kertesz, Attila. (2026). Privacy and security for 6G networks via Fog/Edge Computing, Blockchain, and Federated Learning: A survey and taxonomy. Journal of Network and Computer Applications. <https://doi.org/10.1016/j.jnca.2026.104510>
- [261] Venkatesan, P., Kumaratharan, N.. (2026). Reinforcement learning-based dynamic spectrum allocation for efficient cognitive radio network management. Computer Networks. <https://doi.org/10.1016/j.comnet.2026.112217>
- [262] Verma, Ashwin, Prasad, Vivek Kumar, Kumari, Aparna, Bhattacharya, Pronaya, Srivastava, Gautam, Fang, Kai, Wang, Wei, Gadekallu, Thippa Reddy. (2025). Industry 6.0: Vision, technical landscape, and opportunities. Alexandria Engineering Journal. <https://doi.org/10.1016/j.aej.2025.08.040>
- [263] VijayK. SebasthiraniJ. JeyamaniM. GokulS. ArunkumarAmal Megha John. (2024). Review of Recent Research and Future Scope of Explainable Artificial Intelligence in Wireless Communication Networks. ICT: Smart Systems and Technologies. https://doi.org/10.1007/978-981-99-9489-2_19
- [264] Song, X. Wang, S. Gao, Z. Liu. (2026). Trustworthy Smart Athletic Performance Enhancement Through 6G Internet of Things: Ultrareliable Low-Latency Athletic Intelligence, Computing, and Control Framework. IEEE Internet of Things Journal. <https://doi.org/10.1109/jiot.2025.3619063>
- [265] Wan, Yichen, Qu, Youyang, Gao, Longxiang, Xiang, Yong. (2022). Privacy-preserving blockchain-enabled federated learning for B5G-Driven edge computing. Computer Networks. <https://doi.org/10.1016/j.comnet.2021.108671>
- [266] Wang, Li, Cheng, Xiu-Ming. (2025). End-to-End Federated Intelligence for Secure and Standardized Digital Twin Orchestration in 6G Smart Cities. IEEE Communications Standards Magazine. <https://doi.org/10.1109/mcomstd.2025.3626809>
- [267] Wang, Shen, Qureshi, M. Atif, Miralles-Pechuan, Luis, Huynh-The, Thien, Gadekallu, Thippa Reddy, Liyanage, Madhusanka. (2024). Explainable AI for 6G Use Cases:

- Technical Aspects and Research Challenges. *IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY*, 5, 2490–2540. <https://doi.org/10.1109/ojcoms.2024.3386872>
- [268] Wang, Wei. (2026). Trust-Centric and Standards-Compliant Zero-Touch 6G Architecture for Global English Learning Applications. *IEEE Communications Standards Magazine*. <https://doi.org/10.1109/mcomstd.2026.3657845>
- [269] Wang, Yiyang, Kang, Xin, Li, Tiejian, Wang, Haiguang, Chu, Cheng-Kang, Lei, Zhongding. (2023). SIX-Trust for 6G: Toward a Secure and Trustworthy Future Network. *IEEE Access*, 11, 107657–107668. <https://doi.org/10.1109/access.2023.3321114>
- [270] Watfa, Mohamed K.. (2026). Energy-constrained interpretable reinforcement learning using evolutionary symbolic policies for V2X traffic control. *Applied Soft Computing*. <https://doi.org/10.1016/j.asoc.2026.115566>
- [271] Wu, Junjie, Fung, Benjamin C. M., Stakhanova, Natalia, Khan, Faiyaz, Yu, Hanbo. (2026). Securing automotive data flow: A survey of telematics security across intra-vehicle, V2X, and cloud layers. *Vehicular Communications*. <https://doi.org/10.1016/j.vehcom.2026.101024>
- [272] Luo, H. Guo, R. Ambrocio. (2026). Toward Trustworthy Symbolic Music Generation in 6G-Enabled IoT Environment via Automatic Neural Architecture Design. *IEEE Internet of Things Journal*. <https://doi.org/10.1109/jiot.2025.3613577>
- [273] X. Wang, B. Wang, Y. Wu, Z. Ning, S. Guo, F. R. Yu. (2025). A Survey on Trustworthy Edge Intelligence: From Security and Reliability to Transparency and Sustainability. *IEEE Communications Surveys & Tutorials*. <https://doi.org/10.1109/comst.2024.3446585>
- [274] Xianfa ZhangLina ChenJing Wang. (2025). The research landscape and evolutionary trends of robot-assisted orthopedic telesurgery: a bibliometric and visualized analysis. *Journal of Robotic Surgery*. <https://doi.org/10.1007/s11701-025-02913-1>
- [275] Xiao, Bingnan, Yu, Xichen, Ni, Wei, Wang, Xin, Poor, H. Vincent. (2025). Over-the-air federated learning: Status quo, open challenges, and future directions. *Fundamental Research*. <https://doi.org/10.1016/j.fmre.2024.01.011>
- [276] Xiaodong DuanZhenglei HuangShiyu LiangShaowen ZhengLu LuTao Sun. (2025). AI-agent communication network for 6G: vision, architecture, and key technologies. *Frontiers of Information Technology & Electronic Engineering*. <https://doi.org/10.1631/fitee.2500582>
- [277] Xiaolong XuFan WuMuhammad BilalXiaoyu XiaWanchun DouLina YaoWeiyi Zhong. (2024). XRL-SHAP-Cache: an explainable reinforcement learning approach for intelligent edge service caching in content delivery networks. *Science China Information Sciences*. <https://doi.org/10.1007/s11432-023-3987-y>
- [278] Xu, Han, Nanda, Priyadarsi, Liang, Jie. (2024). Reciprocal Federated Learning Framework: Balancing incentives for model and data owners. *Future Generation Computer Systems*. <https://doi.org/10.1016/j.future.2024.06.055>
- [279] Xu, Peijun, Liu, Miao, Liu, Maoyun, Shen, Ai. (2025). Artificial intelligence in surgical oncology: A comprehensive review from preoperative planning to postoperative care. *Intelligent Oncology*. <https://doi.org/10.1016/j.intonc.2025.09.001>

- [280] Xu, Qichao, Su, Zhou, Li, Ruidong. (2022). Security and Privacy in Artificial Intelligence-Enabled 6G. *IEEE NETWORK*, 36(5), 188–196. <https://doi.org/10.1109/mnet.117.2100730>
- [281] Chen, J. Zhang, W. Dai, J. Han, M. Chen, Z. Liu, A. Farouk. (2026). An Entropy-Based Privacy-Preserving Federated Deep Reinforcement Learning Framework for Task Offloading in Vehicular Edge Computing Networks. *IEEE Internet of Things Journal*. <https://doi.org/10.1109/jiot.2025.3602798>
- [282] Y. Rumes, T. T. Senevirathna, P. Porambage, M. Liyanage, M. Ylianttila. (2023). Comprehensive Analysis Over Centralized and Federated Learning-Based Anomaly Detection in Networks with Explainable AI (XAI). *ICC 2023 - IEEE International Conference on Communications*. <https://doi.org/10.1109/icc45041.2023.10278845>
- [283] Y. Sun, B. D. Parameshachari, H. Wang. (2026). Trustworthy AI-Driven 6G-IoT Architecture for Remote Healthcare: Reliable Resource Orchestration and Adaptive Network Intelligence. *IEEE Internet of Things Journal*. <https://doi.org/10.1109/jiot.2025.3583340>
- [284] Y. Wang, B. -G. Kim, X. Li. (2026). Large Model-Driven Trustworthy Child Psychoeducation Mechanism for 6G Internet of Everything. *IEEE Internet of Things Journal*. <https://doi.org/10.1109/jiot.2025.3590172>
- [285] Y. Wu. (2022). Ethically Responsible and Trustworthy Autonomous Systems for 6G. *IEEE Network*. <https://doi.org/10.1109/mnet.005.2100711>
- [286] Y. Wu, Y. Wang. (2026). Impact-Safe Martial Arts Training via Secure Digital Twin Standardization and 6G Edge-Enabled Collaborative Computing. *IEEE Communications Standards Magazine*. <https://doi.org/10.1109/mcomstd.2026.3677775>
- [287] Yang, Mengmeng, Qu, Youyang, Ranbaduge, Thilina, Thapa, Chandra, Sultan, Nazatul Haque, Ding, Ming, Suzuki, Hajime, Ni, Wei, Abuadba, Sharif, Smith, David, Tyler, Paul, Pieprzyk, Josef, Rakotoarivelo, Thierry, Guan, Xinlong, Mrabet, Sirine. (2026). From 5G to 6G: A Survey on Security, Privacy, and Standardization Pathways. *ACM COMPUTING SURVEYS*, 58(8), 193. <https://doi.org/10.1145/3785467>
- [288] Yang, Zhaohui, Xu, Wei, Liang, Le, Cui, Yuanhao, Qin, Zhijin, Debbah, Merouane. (2025). On privacy, security, and trustworthiness in distributed wireless large AI models. *Science China Information Sciences*, 68(7), 170301. <https://doi.org/10.1007/s11432-024-4465-3>
- [289] Yin, Hao, You, Xiaohu, Huang, Yongming, Xu, Wei, Zhang, Jiao, Pan, Zhengang, Liu, Guangyi, Wei, Kejun, Wu, Gang. (2026). Future mobile information networks: Fundamental theories, network architectures, key technologies, and key devices. *Fundamental Research*. <https://doi.org/10.1016/j.fmre.2025.12.033>
- [290] Younesi, Abolfazl, Oustad, Elyas, Ansari, Mohsen, Fahringer, Thomas, Buyya, Rajkumar. (2026). HealthCare 5.0: An industry 5.0 perspective for next-generation medical systems with synergistic integration of IoT, AI, and 6G. *Internet of Things*, 35, 101815. <https://doi.org/10.1016/j.iot.2025.101815>
- [291] Yousef SanjalaweSalam FraihatSalam Al-E'mariSharif Naser Makhadmeh. (2026). Bridging the gap: A comprehensive survey on AI-driven digital twin networks for future wireless systems. *Journal of King Saud University Computer and Information Sciences*, 38(3), 113. <https://doi.org/10.1007/s44443-026-00522-y>

- [292] Cao, B. Zhang, S. Zong, Y. Yu. (2026). Computer Vision-Based Beamforming Optimization With Trustworthy AI for 6G-IoT Services. *IEEE Internet of Things Journal*. <https://doi.org/10.1109/jiot.2025.3616189>
- [293] Z. Li, X. Chen, Q. Sun, Y. Lu, D. Li, J. Zhang, B. Ai. (2026). Trustworthy Federated Learning with Authenticated ZKPs in Mobile Edge Intelligence. *IEEE Transactions on Mobile Computing*. <https://doi.org/10.1109/tmc.2026.3670154>
- [294] Zaki-Hindi, Ayat, Soto, Paola, Castellanos, German, Pritom, Touhid Hossain, Volpe, Gaetano, Zellagui, Iskander, Hensel, Burkhard, Jimenez, Julian, Marvulli, Michele, Santos, Luis, Sayin, Ayse, Sottet, Jean-Sebastien, Ali, Wasim, El-Korbi, Ines, Turcanu, Ion, Belogaev, Andrey, Duarte, Andre, Kelly, Ultan, Kumar, Sumit, Myagkov, Georgy, Parker, Stephen, Franke, Mario, Hossain, Shajjad, Sanyal, Rajarshi, Shafi, Nida, Sommer, Christoph, Camelo, Miguel, Baudouin, Julien, Decorme, Regis, Fanti, Maria Pia, Fuladi, Ramin, Murphy, Chris, Pereira, Ana, Senouci, Sidi Mohammed, Pryor, Simon, Faye, Sebastien. (2026). A Reference Functional Architecture for Network Digital Twins in 6G Systems. *IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY*, 7, 2068–2101. <https://doi.org/10.1109/ojcoms.2026.3668035>
- [295] Zeb, Shah, Mahmood, Aamir, Khowaja, Sunder Ali, Dev, Kapal, Hassan, Syed Ali, Gidlund, Mikael, Bellavista, Paolo. (2024). Towards defining industry 5.0 vision with intelligent and softwarized wireless network architectures and services: A survey. *Journal of Network and Computer Applications*. <https://doi.org/10.1016/j.jnca.2023.103796>
- [296] Zhang, Rui, Song, Mingcong, Li, Tao, Yu, Zhibin, Dai, Yuting, Liu, Xiaoguang, Wang, Gang. (2021). Democratic learning: hardware/software co-design for lightweight blockchain-secured on-device machine learning. *Journal of Systems Architecture*. <https://doi.org/10.1016/j.sysarc.2021.102205>
- [297] Zhang, Ruichen, Liu, Guangyuan, Liu, Yinqiu, Zhao, Changyuan, Wang, Jiacheng, Xu, Yunting, Niyato, Dusit, Kang, Jiawen, Li, Yonghui, Mao, Shiwen, Sun, Sumei, Shen, Xuemin, Kim, Dong In. (2026). Toward Edge General Intelligence With Agentic AI and Agentification: Concepts, Technologies, and Future Directions. *IEEE COMMUNICATIONS SURVEYS AND TUTORIALS*, 28, 4285–4318. <https://doi.org/10.1109/comst.2026.3651702>
- [298] Zheng, Wei, Anwlkom, Tomley, Liu, Ye. (2026). TRUST-Med6G: A Secure and Trustworthy Medical Data Privacy Protection Mechanism in 6G-IoT. *IEEE Internet of Things Journal*, 13(5), 8248–8260. <https://doi.org/10.1109/jiot.2025.3622761>
- [299] Zhou, Junxi, Zhang, Qingbo, Zhang, Ming, Aikebaier, Subinuer, Zhu, Xiao, Zhou, Xiaorong. (2026). Flexible wearable medical devices: from material innovations and data processing to intelligent healthcare applications. *Journal of Advanced Research*. <https://doi.org/10.1016/j.jare.2025.08.036>
- [300] Zhou, Li, Yin, Hao, Zhao, Haitao, Wei, Jibo, Hu, Dewen, Leung, Victor C.M.. (2026). A comprehensive survey of artificial intelligence applications in UAV-enabled wireless networks. *Digital Communications and Networks*. <https://doi.org/10.1016/j.dcan.2024.11.005>
- [301] Zhu, Xufei. (2025). Research on Privacy Protection of Digital Twin Intelligence Based on Big Data in 5G System Security. *IEEE Communications Standards Magazine*. <https://doi.org/10.1109/mcomstd.2025.3622196>
- [302] Zia, Muhammad Furqan, Ouameur, Messaoud Ahmed, Bagaa, Miloud, Massicotte, Daniel, Ksentini, Adlen. (2026). A survey on explainable AI for semantic communication:

Architecture, challenges, and future opportunities. COMPUTER NETWORKS, 280, 112142. <https://doi.org/10.1016/j.comnet.2026.112142>