ITUPublications

International Telecommunication Union
Standardization Sector

ITU-T Technical Report

(12/2025)

XSTR.trust-metaverse

Technical challenges to achieving trustworthy metaverses

Technical Report ITU-T XSTR.trust-metaverse

Technical challenges to achieving trustworthy metaverses

Summary

The metaverse is an integrative ecosystem of virtual worlds, where participating entities may have one or more identities. Its essential enablers are cutting-edge technologies including artificial intelligence (AI), Web 3.0, blockchain, augmented reality (AR), virtual reality (VR) and the Internet of things (IoT). When all these advanced technologies are applied and used in some scenarios, it will bring a number of concerns, such as concerns over safety, security and privacy. In the metaverse, all these concerns and even other unexpected problems make trustworthiness one of the key issues for the metaverse and its development.

This Technical Report describes metaverses that are reliable, responsible and can be trusted/trustworthy, i.e., trustworthy metaverses. It presents an overview of the metaverse environment and the need for, features of, and technical challenges to, trustworthy metaverses.

Keywords

Artificial intelligence, augmented reality, blockchain, metaverse, trustworthy, virtual reality, web 3.0.

Note

This is an informative ITU-T publication. Mandatory provisions, such as those found in ITU-T Recommendations, are outside the scope of this publication. This publication should only be referenced bibliographically in ITU-T Recommendations.

Table of Contents

Technical Report ITU-T XSTR.trust-metaverse

Technical challenges to achieving trustworthy metaverses

1 Scope

This Technical Report focuses primarily on metaverses that are reliable, responsible and can be trusted, i.e., trustworthy metaverses.

The scope of this Technical Report includes:

– An overview of the metaverse environment and the need for trustworthy metaverses.

– Features of trustworthy metaverses.

– Technical challenges for trustworthy metaverses.

2 References

[ITU-T Y.3052] Recommendation ITU-T Y.3052 (2017), *Overview of trust provisioning in information and communication technology infrastructures and services.*

[ITU-T Y.4000] Recommendation ITU-T Y.4000/Y.2060 (2012), *Overview of the Internet of things.*

3 Definitions

3.1 Terms defined elsewhere

This Technical Report uses the following terms defined elsewhere:

3.1.1 application [b-ITU-T Y.2091]: A structured set of capabilities, which provide value-added functionality supported by one or more services, which may be supported by an API interface.

3.1.2 avatar [b-ISO/IEC 23005-4]: Entity that can be used as a (visual) representation of the user inside the virtual environments.

3.1.3 blockchain [b-ITU-T X.1400]: A type of distributed ledger which is composed of digitally recorded data arranged as a successively growing chain of blocks with each block cryptographically linked and hardened against tampering and revision.

3.1.4 decentralized system [b-ITU-T X.1400]: Distributed system wherein control is distributed among the persons or organizations participating in the operation of system.

3.1.5 distributed ledger [b-ITU-T X.1400]: A type of ledger that is shared, replicated, and synchronized in a distributed and decentralized manner.

3.1.6 distributed ledger technology (DLT) [b-ITU-T X.1400]: Technology that enables the operation and use of distributed ledgers.

3.1.7 internet of things (IoT) [ITU-T Y.4000]: A global infrastructure for the information society, enabling advanced services by interconnecting (physical and virtual) things based on, existing and evolving, interoperable information and communication technologies.

NOTE 1 – Through the exploitation of identification, data capture, processing and communication capabilities, the IoT makes full use of things to offer services to all kinds of applications, whilst ensuring that security and privacy requirements are fulfilled.

NOTE 2 – From a broader perspective, the IoT can be perceived as a vision with technological and societal implications.

3.1.8 ledger [b-ITU-T X.1400]: Information store that keeps final and definitive (immutable) records of transactions.

3.1.9 metaverse [b-ITU-T Y.4238]: A collective virtual environment where physical and virtual worlds converge, that enables users to interact with shared digital spaces, objects and services.

NOTE – A metaverse can be virtual, augmented, representative of or associated with the physical world.

3.1.10 thing [ITU-T Y.4000]: With regard to the Internet of things, this is an object of the physical world (physical things) or of the information world (virtual things), which is capable of being identified and integrated into communication networks.

3.1.11 trust [ITU-T Y.3052]: The measurable belief and/or confidence which represents accumulated value from history and the expecting value for future.

3.1.12 trustworthiness [b-ISO/IEC 22989]: ability to meet stakeholder expectations in a verifiable way.

NOTE 1 – Depending on the context or sector, and also on the specific product or service, data and technology used, different characteristics apply and need verification to ensure stakeholders' expectations are met.

NOTE 2 – Characteristics of trustworthiness include, for instance, reliability, availability, resilience, security, privacy, safety, accountability, transparency, integrity, authenticity, quality and usability.

NOTE 3 – Trustworthiness is an attribute that can be applied to services, products, technology, data and information as well as, in the context of governance, to organizations.

3.2 Terms defined in this Technical Report

None.

4 Abbreviations and acronyms

This Technical Report uses the following abbreviations and acronyms:

AI	Artificial Intelligence
AIGC	Artificial Intelligence Generated Content
AR	Augmented Reality
IoT	Internet of Things
LCM	Lifecycle Management
NLP	Natural Language Processing
PII	Personally Identifiable Information
QoE	Quality of Experience
QoS	Quality of Service
VR	Virtual Reality

5 Conventions

None.

6 An overview of the metaverse environment

6.1 Background of the metaverse

Defined as a collective virtual environment that enables users to interact with shared digital spaces, objects and services [b-ITU-T Y.4238], the metaverse is a space where participating entities may

have one or more identities. Its essential enablers are cutting-edge technologies including artificial intelligence (AI), Web 3.0, blockchain, augmented reality (AR), virtual reality (VR) and the Internet of things (IoT). A virtual representation of the real world, everything in the metaverse is digitalized and virtualized, and "created" by companies, organizations, persons and even by the metaverse itself.

However, there are also potential risks. When all these advanced technologies are applied and used in the metaverse, they will bring a number of concerns, such as concerns over safety, security and privacy. In the metaverse, all these concerns and other unexpected problems make trustworthiness one of the key issues for the metaverse and its development.

The trustworthy metaverse is a metaverse that is reliable, responsible and can be trusted. Its core concepts, technical features, trustworthy factors, key roles and technical challenges need to be figured out and clarified as a foundation toward a reference model to further investigate the standardization landscape of trustworthy metaverses. This Technical Report aims to present key concepts, features and technical challenges in this regard.

6.2 Need for trustworthy metaverses

ITU's trust-related standardization activities have addressed the complex challenges associated with emerging technologies and digital ecosystems, promoting a trustworthy environment for users, organizations and entities participating in the evolving digital landscape. To this end, [ITU-T Y.3052] defines trust as the measurable belief and/or confidence which represents accumulated value from history and the expecting value for the future.

To address the risks identified in the metaverse, it is necessary to consider the following aspects:

– Interactions considering the cyber-physical relationships between humans and things.
– Reliable data processing and management for monitoring, analytics, prediction and decision-making.
– Transparent sharing and exchange of digital resources, including their identification.
– Safety guarantee for the digital assets which can be exchanged as currency in the virtual world(s) and directly related to the properties in the real world.
– Secure and correct operations with autonomous decision-making.
– Measurable indicators and evaluation methodology for different levels of trust.

As a decision-making behaviour, trust is affected by past experience and associated predictions for the future. Historically, the study of trust in systems has been a topic of psychology [b-ITU-T TR Trust]; however, with the development of intelligent technologies, several unique changes and challenges arise; the interaction between a user and a system is becoming very important. In the meantime, trust itself is a complexity-reduction mechanism whose importance increases the less the technology(-ies) is known/understood.

When it comes to trust-related topics, the following issues and concerns should be considered:

– Is this metaverse trustworthy or not? How can trustworthiness be measured and what is the benchmark of being trustworthy?
– If it is somewhat trustworthy, how much it can be trusted?
– If it is not trustworthy, how it can be improved and optimized in order to be trusted?
– What are the standardized methods and parameters to measure the trustworthiness of the metaverse?

Above all, trust is an important topic and concern for the users, vendors and supervisors in the metaverse. The trustworthy metaverse is one of the essential and urgent topics for commercial usage of the metaverse and, with this Technical Report, the core concepts of the trustworthy metaverse will be discussed and defined and a series of pre-standard topics will be discussed and studied.

7 Features of the trustworthy metaverse

7.1 Technical features of the trustworthy metaverse

Considering that the metaverse is a mirror of the real world, the important issue of trust is even more important in the metaverse.

When it comes to trust topic for the metaverse, the following aspects should be considered:

– **Mirroring the real world**: In the metaverse, one of its attractive features is that the metaverse includes the virtualized real world. In order to mirror the real world into the metaverse, it is necessary to use digital twin technologies.

– **Interaction between real world and metaverse virtual worlds**: As the metaverse can contain more than one "world", the languages in the metaverse are diverse. Natural language processing (NLP) is crucial in the metaverse virtual worlds to achieving free and smooth interaction.

– **Digital copyright**: The virtualized worlds are full of virtual and digital lives, mirroring and virtual entities, digital images, music and so on, each of which may have its copyright or relevant legal identity. With so many copyright and legal issues to deal with, AI technology will play an important role in fulfilling legal requirements.

– **Content creation**: All the contents in the metaverse are virtualized and digitalized by programs or algorithms. As the contents evolve to become richer and richer, it gets harder and harder for them to be designed, operated, managed and maintained by engineers alone; hence, again, AI technology is the way to the continuous development and evolution of the metaverse, e.g., the creation of avatar(s).

– **Avatars**: Avatars are entities that can be used as (visual) representations of the user inside the virtual environments [b-ISO/IEC 23005-4].

– **Digital identity**: Digital identity is defined in [b-ITU-T X.1252] as a digital representation of the information known about a resource, specific individual, group or organization. In the metaverse, digital identity is a representative identity, such as an avatar, created upon validated user/entity login authentication and authorization.

– **Digital asset**: In the document, a digital asset is a digital representation of value recorded on a cryptographically secured distributed ledger or similar technology, and it is supposed to be capable of being exchanged and traded in a digital world like metaverse.

7.2 Trust considerations for the metaverse

To make trust in the metaverse more intuitive and acceptable to humans in the real world and not just to specialists, the following aspects should be considered for trust in the metaverse:

– **Strict quality of experience (QoE) requirements**: In the metaverse, all interactions and objects are virtualized and enabled by AI, computational power and relevant technologies. It is required that the metaverse should achieve immersive experience and real-time interactions and should be capable of being accessed anytime and anywhere; these are strict QoE requirements.

– **Compliance**: As virtual worlds in which people can play different roles in different virtual scenarios and many interactions will happen in the metaverse, it is important and urgent to study, discuss and ensure that all services, technologies and interactions are compliant with relevant regulations.

– **Accountability**: Service providers or vendors of the metaverse should take responsibility for the autonomous actions executed by AI and relevant technologies.

– **Equitability**: In the metaverse virtual worlds, intended or unintended bias(es) or unfairness(es) should be avoided, because the bias or unfairness would inadvertently cause harm, damage and loss.
– **Safety, data security and privacy**: Safety, data security and privacy issues should be ensured in the trustworthy metaverse.

8 Technical challenges for the trustworthy metaverse

8.1 Proof of identity

In the metaverse, proof of the user's identity and the ownership of digital assets is the key to maintaining the sustainable and healthy development of the metaverse. Without digital identity, the infrastructure of the metaverse would be vulnerable. In the event of a cyberattack, weak digital identities, such as usernames and passwords, will be stolen or used for other fraudulent activities.

Digital identities in the metaverse face the following challenges:

– **Personally identifiable information (PII)**: Digital identity information and sensitive (personally identifiable) data, such as names, e-mail addresses and phone numbers, need to be protected. These data may be obtained illegally and abused. Attackers can obtain these sensitive data through various means, such as network sniffing and man-in-the-middle attacks, for malicious activities such as identity theft and fraud.
– **Malicious software and attacks**: As more users and metaverses join, the network must be able to handle large-scale data exchanges and transactions while maintaining low latency and high throughput.
– **Identity interoperability**: In the metaverse, each user maps multiple identities in the metaverse platforms. Identification and interoperability between multiple identities can pose challenges. Other proof of identity methods could be included (e.g., multi-factor authentication).

8.2 Network connection

The metaverse integrates VR, AR, AI, blockchain and other technologies. An efficient and reliable network is a key enabler for the metaverse. In the metaverse, a lot of computation and data transmission is required. If the network is not reliable, it can lead to data loss, inaccurate computation results and poor user experience.

The trust challenges for network connection include:

– **Interoperability**: Different metaverses have different data formats, interaction methods and economic models. There is a lack of common standards and protocols to ensure that data and assets can be shared across metaverses.
– **Scalability**: As more users and metaverses join, the network must be able to handle large-scale data exchanges and transactions while maintaining low latency and high throughput.

8.3 AI technology

AI technology can be applied to a wide variety of scenarios, such as content generation, character modelling, speech recognition and user behaviour analysis in the metaverse. The metaverse must support the trusted AI operation in the data, modelling, analysis, prediction and decision-making process. If the AI modelling or operation process cannot be trusted, it will lead to the entire metaverse not being trusted, with serious consequences.

The trust challenges for AI include:

– **Lack of transparency**: The decision-making process of AI is often not transparent, which makes it difficult for people to understand and trust the decisions made by AI, and puts AI at risk of abuse.
– **Data supply**: The foundation of AI computing is data and AI computing in the metaverse requires access to a large amount of data. As a virtual space, information and property in the metaverse are easily stolen and attacked in the process of AI computing.

8.4 Blockchain technology

Blockchain is one of the key foundational technologies of the metaverse and its security, reliability, decentralization and interoperability provide a secure, trustworthy and transparent environment for the metaverse. However, blockchain technology faces a number of security challenges:

– **Growing data volume challenge**: With the growth of a blockchain, the volume of blockchain data stored by nodes is getting larger and larger, and the burden of its storage and computation is getting heavier and heavier, which will bring great difficulties to the operation of metaverse clients.
– **Low blockchain application efficiency**: Blockchain transactions require multiple confirmations, each of which creates a delay. Such efficiency does not meet the real-time requirements of the metaverse.

8.5 Trust lifecycle management

Lifecycle management (LCM) of trust refers to a complete process of managing for trust throughout the lifecycle of a trustworthy metaverse. The general process of trust LCM is illustrated in Figure 1.

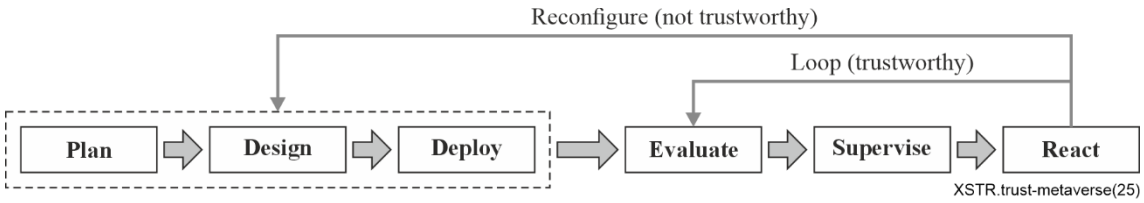

Figure 1 – General process of trust LCM

Based on Figure 1, the trust LCM can be applied in the lifecycle of the metaverse to make the metaverse trustworthy. The following steps should be considered, even though not all steps are necessary in a lifecycle, i.e., the steps in a lifecycle depend on specific conditions or scenarios:

– **Plan**: At the beginning of the trust lifecycle for the trustworthy metaverse, it is recommended to make plans of trust during the whole lifecycle before the relevant metaverse works.
– **Design**: After planning, the design of the trustworthy metaverse is essential and necessary. It is recommended that, during the design, all the elements, processes, factors and so on be designed to be trustworthy from the very beginning or designed with the fundamental principle to be trustworthy.
– **Evaluate**: In order to be objective, measurable and quantifiable, it is recommended that the trustworthiness of the relevant system in the lifecycle of the metaverse be computed or evaluated so that the operators, users, governors and supervisors can understand the levels of trustworthiness in order to make decisions, actions, judgements and authorizations.
– **Supervise**: In the trust lifecycle for the trustworthy metaverse, after the computation/evaluation, it is recommended to monitor the fluctuation of trustworthiness in order to ensure the metaverse is trustworthy throughout its lifecycle.
– **React**: After the evaluation/assessment of trustworthiness, it is recommended that the metaverse system make appropriate reactions based on the trustworthiness

evaluation/assessment results, i.e., to continue working if trustworthy or to reconfigure the system if not trustworthy.

– **Loop**: To ensure continuous and sustainable trustworthiness, a lifecycle should be followed by another lifecycle in the trustworthy metaverse.

– **Reconfigure**: If the metaverse system is not trustworthy enough, i.e., the evaluation results are not ideal enough to be trusted, it is recommended to reconfigure the metaverse system to make it trustworthy.

8.6 Trust indicators and measurement for computational trust

8.6.1 Indicators of computational trust

In order to make trust in the trustworthy metaverse computable, measurable and quantifiable, i.e., to make trust itself more objective and quantitative for the trustworthy metaverse, it is proposed to use trust indicators. Based on these indicators, the degree or level of trust can be computed to represent trustworthiness directly and objectively. Table 1 lists trust indicators and factors for computational trust of the trustworthy metaverse. In the table, a factor is a sub-attribute that contributes to a trust indicator.

Table 1 – Trust indicators for computational trust of the trustworthy metaverse

Indicators	Factors
Accuracy	QoE
	QoS
	Timeliness
	Resource
Stability	Interruption
	Accident
	Maturity
	Variability
Controllability	Predictability
	Supervision
	Compliance
	Taken over
Resilience	Backup
	Reset
Adaptability	Flexibility
	Adjustment
Security	Privacy
	Asset safety

8.6.2 Computation and measurement of trustworthiness

To make trust itself computable and measurable for the metaverse, it is proposed to study the general method and process for computation and measurement of trustworthiness. The measurement of trustworthiness should be conducted in the commercial environment or the environment that is mirrored by the commercial one(s). During the evaluation/assessment of trustworthiness, depicted in Figure 2, all the above indicators and related factors in clause 8.6.1 should be taken into consideration objectively.

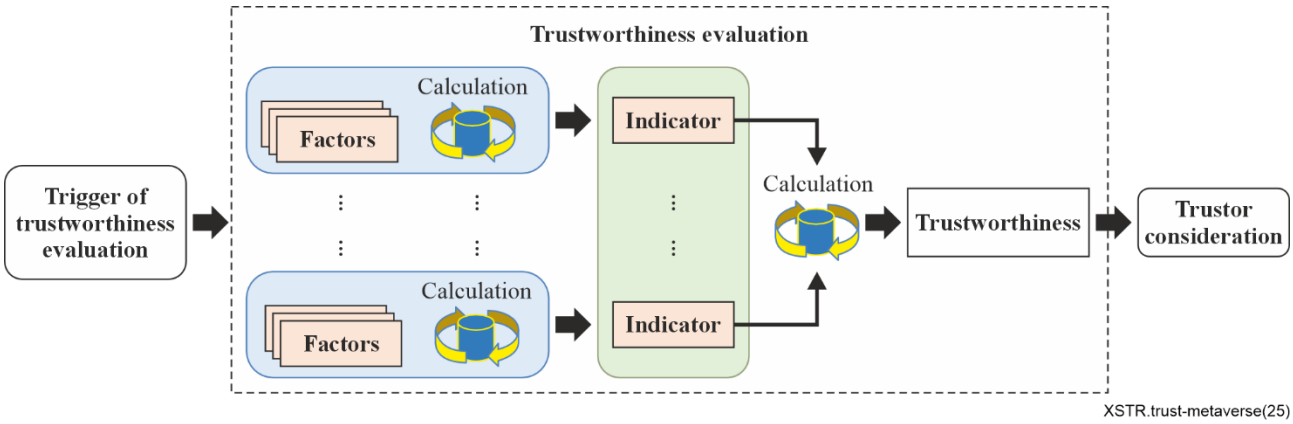

Figure 2 – General process of trustworthiness evaluation for the metaverse

As Figure 2 illustrates, trustworthiness can be evaluated/assessed/measured quantitatively and objectively. The following aspects are essential to trustworthiness evaluation for the metaverse:

– **Environment of trustworthiness evaluation**: Trustworthiness of metaverse evaluation can take place in commercial networks; it also can take place in a test or simulation environment that is mirrored by the commercial network.

– **Trigger of trustworthiness evaluation**: Trustworthiness evaluation of the metaverse can be triggered by the trustor, as well as by the trustee. The trigger includes the following situations:

  • Orders before/at the start-up of some metaverse systems.

  • Configured orders, including periodic orders and aperiodic orders at specific time points.

  • Temporary orders at random time points.

The trustor can trigger the trustworthiness evaluation of the metaverse by using some original and configured/standardized input/order for different scenarios; the trustee can also trigger the trustworthiness evaluation by itself in order to gain the trust of the trustor if necessary.

– **Indicators of trustworthiness evaluation**: Metrics associated with indicators are parameters that are specified to make trust/trustworthiness measurable and quantifiable for the metaverse (system).

  NOTE 1 – Indicator(s) in the same trustworthiness evaluation should be unified with the same unit and in a unified way.

– **Factors of trustworthiness evaluation**: A series of related factors should be defined for each indicator. These factors, once assessed or evaluated, serve as calculation inputs for their corresponding indicators.

  NOTE 2 – Factor(s) in the same trustworthiness evaluation should be unified in the same unit with the same unified way; the unit and unified way of sub-metrics and metrics should remain consistent throughout the same trustworthiness evaluation.

– **Results of trustworthiness evaluation**: Trustworthiness evaluation results should be handed over to the trustor, in order to take into consideration, and make decisions or cast judgement on, the following authorizations or progress.

8.7 Trusted artificial intelligence generated content (AIGC) technologies

AIGC technology is a crucial, powerful and productive enabler to the metaverse system(s), in which most of the (virtual) contents can be generated manually or automatically. Due to the need for a large number of contents in the virtual world of the metaverse, AIGC technologies are supposed to be applied more widely, extensively and substantially and AICG is the main resource of continuous generative contents and creativities in the metaverse.

To ensure the trustworthiness of AIGC, the following aspects should be considered:

– **Trusted contents**: The contents, especially the generated contents, should be trustworthy enough, i.e., the contents should meet the trust-related requirements.

– **Trusted AI technologies**: The AI technologies for content generation should be trustworthy; in the commercial environment, the AI technologies that are applied in some metaverse systems should be verified/certified in order to achieve the related requirement of trustworthiness.

– **Content verification**: The generated contents should be verified in order to be trustworthy, so that the users or the trustors can carry out the relevant interactions or authorizations.

– **Content traceability**: All the generated content should be traceable with some original mark(s) or information.

– **Content security**: The generated content should meet relevant security requirements.

8.8 Trusted data

As the important input and fertilizer for the metaverse, data is the key enabler and fuel for most processes of the metaverse. Trustworthy data and datasets are needed to make the metaverse trusted and trustworthy.

To ensure trustworthiness of data in the metaverse, the following detailed and concrete aspects are suggested to be considered:

– **Compliance**: All the data of the trustworthy metaverse, including the input data and the output data, should be compliant with specific rules and related specifications and even laws within the specific metaverse (system).

– **Traceability**: The data of trustworthy metaverses should be traceable and trusted data or datasets may be configured with an identity or mark to enable traceability.

– **Privacy**: All the trusted data, including the input data and the output data, should meet the requirement of user privacy protection.

8.9 Trusted digital asset

A digital asset is a digital representation of value, recorded on a cryptographically secured distributed ledger or similar technology, and is supposed to be capable of being exchanged and traded in a digital world like the metaverse without an intermediary. In the trustworthy metaverse, the digital assets should be trustworthy enough for exchanges and trades in the digital world as the metaverse.

To ensure trustworthiness of a digital asset in the metaverse, the following essential aspects should be taken into consideration:

– **Decentralization**: With numerous entities of a metaverse distributed in the digital world, each entity has highly autonomous characteristics. The entities can freely connect to each other and form new connection units. Each entity can become a periodic centre but does not have mandatory central control functions. Interactions among entities create a nonlinear causal relationship throughout the metaverse. Thus, "decentralization" should be considered one of the essential properties of a trusted digital asset.

– **Encryption**: Encryption is the process of converting data into a message that no one can understand without the correct key through cryptographic arithmetic. A trusted digital asset should be encrypted all the time in the digital world.

– **Traceable**: All the traces of the digital asset should be recorded, so that they can be traced back if necessary.

– **Immutable**: Based on the traceability, all the recorded traces should be immutable, i.e., they cannot be falsified.

– **Privacy**: All the private information, especially the assets, should be well protected.
– **Security**: The physical security, network security, data encryption, identity authentication and so on should be ensured, so that exchanges are protected from attack and illegal access.
– **Trusted exchange/payment**: All exchanges or payments should be trustworthy enough whether in peer-to-peer, real-time or offline scenarios.

Bibliography

[b-ITU-T X.1400] Recommendation ITU-T X.1400 (2026), *Terms and definitions for distributed ledger technology.*

[b-ITU-T Y.2091] Recommendation ITU-T Y.2091 (2011), *Terms and definitions for next generation networks.*

[b-ITU-T Y.4238] Recommendation ITU-T Y.4238 (2025), *Requirements for integrating virtual and physical worlds through digital twins in the metaverse.*

[b-ITU-T TR Trust] ITU-T Technical Report – *Trust in ICT (2017).*

[b-ISO/IEC 22989] ISO/IEC 22989:2022, *Information technology – Artificial intelligence – Artificial intelligence concepts and terminology*.

[b-ISO/IEC 23005-4] ISO/IEC 23005-4:2018, *Information technology – Media context and control – Part 4: Virtual world object characteristics*