LIG Accuver

AI-Based No-Reference Video Quality Assessment for the 5G and 6G Era

Standardization Trends, VQML® Algorithm Overview, and Use Cases with XCAL

Doc Rev : 1.0.0

July 2026

LIG

Contents

Executive Summary

Video has become one of the clearest indicators of mobile network performance. Users judge service quality through playback continuity, visual clarity, synchronization, and responsiveness—not through radio and transport metrics alone. As operators move toward 5G Standalone and future 6G architectures, assurance must connect network KPIs with application-level Quality of Experience (QoE).

LIG Accuver's VQML® estimates human-perceived video quality directly from received RGB frames without the original source, transmission metadata, or manual subjective testing. Integrated with XCAL, it enables real-time and session-level MOS estimation in laboratory, field, device, and live-service environments.

Adopted as Model A in the approved ITU-T J.344 framework, VQML® combines content, signal, perceptual, and device-aware analysis. XCAL can correlate the predicted MOS with radio and protocol logs to help identify likely factors associated with service degradation.

1. ITU-T Video Quality Assessment Standardization Trends

1.1 Role of ITU-T SG12 in QoS and QoE Standardization

ITU-T Study Group 12 (SG12) is the ITU-T study group responsible for performance, Quality of Service (QoS), and Quality of Experience (QoE). Its work covers network performance, speech and audio quality, multimedia services, and subjective and objective quality assessment methodologies.

For video services, SG12 recommendations provide a common technical foundation for measuring how coding, transmission, capture devices, displays, and application behavior affect perceived quality. These methods are used by telecom operators, network equipment vendors, device manufacturers, service providers, regulatory authorities, and research laboratories.

The role of perceptual assessment is becoming more important as 5G SA and future 6G networks introduce network slicing, private networks, mission-critical communications, and application-specific Service-Level Agreements (SLAs). Radio metrics such as signal strength, throughput, latency, and packet loss remain essential, but they must be supplemented with measurements of the service experience delivered to the user.

1.2 J Series and P Series: Complementary Perspectives

Two ITU-T recommendation series are especially relevant to video quality assessment:

Standard Series	Primary Focus	Practical Meaning	Representative Examples
J Series	Perceptual video-signal quality for television, broadcasting, cable, and multimedia transmission	Measures visual degradation caused by coding, processing, cameras, and transmission. Includes full-reference, reduced-reference, and no-reference approaches.	J.144, J.246, J.343 series, J.344 series
P Series	Subjective and objective assessment of communication-media quality and end-user QoE	Addresses terminal- and display-based assessment, audiovisual service quality, streaming quality, and end-user perception.	P.910, P.1204 series

Table 1. Complementary roles of the ITU-T J and P series

In practical terms, the J Series focuses primarily on the perceptual integrity of the received video signal, while the P Series takes a broader view of user-perceived communication-media and service quality. Using both perspectives enables a more complete QoE assurance framework.

1.3 Full-Reference, Reduced-Reference, and No-Reference Methods

Video Quality Assessment (VQA) methods are categorized by the amount of information required from the original source signal:

Method	Required Input	Representative ITU-T Standards	Strengths	Limitations and Typical Use
Full-Reference (FR)	Complete source video and processed or received video; accurate alignment is normally required.	J.144; J.343.5–J.343.6; P.1204.4	Typically provides high precision because the complete source reference is available.	Best suited to codec development, equipment benchmarking, and controlled laboratory validation. Difficult to deploy in live or user-side monitoring.
Reduced-Reference (RR)	Received video plus selected features extracted from the source and transferred through a reference-side channel.	J.246; J.343.3–J.343.4; P.1204.4	Reduces reference-data overhead while retaining partial source information.	Requires reference-side feature extraction, synchronization, and side-channel delivery. Applicable when a controlled reference path is available.
No-Reference (NR)	Received video, metadata, or bitstream information without the complete original source. Pixel-based NR models can operate directly on decoded RGB frames.	J.343.1–J.343.2; J.344.1–J.344.2; selected P.1204 models	Most practical for field, black-box, end-user, and live-service measurement.	Technically demanding because perceptual degradation must be inferred without a complete reference. Model behavior depends on the available input type and training coverage.

Table 2. Comparison of reference requirements and deployment characteristics

FR and RR remain valuable when source-side data can be controlled. Their dependence on source access and synchronization, however, limits their scalability in live, multi-vendor mobile environments. NR assessment is therefore the practical foundation for broad field and service monitoring.

NR models also differ in their inputs. Metadata-driven models, such as P.1204.1, estimate quality from encoding and service parameters. These inputs may be unavailable in encrypted services and may not expose pixel-level visual defects. The J.344 approach addresses this limitation by predicting perceived quality directly from received RGB video without requiring the original source or transmission metadata. Advanced AI and perceptual deep learning architectures are required to overcome the computational complexity of pure RGB analysis—a challenge that LIG Accuver has successfully solved.

1.4 The Approved ITU-T J.344 Framework

The J.noref work item established a standardized approach for objective no-reference assessment of Full HD video. This work matured into the approved ITU-T J.344 recommendation family, which provides the framework and model definitions for advanced NR objective VQA.

Standard	Scope	Primary Impairment Focus
J.344	Framework and umbrella recommendation for no-reference objective VQA of Full HD video	Overall architecture, evaluation principles, and model framework
J.344.1	No-reference objective VQA model for coding artifacts in Full HD video	Coding artifacts caused by video compression, including quantization-related degradation
J.344.2	No-reference objective VQA model for coding artifacts and camera impairments in Full HD video	Coding artifacts and capture-side camera impairments, including blur, camera shake, focus errors, sensor noise, and lighting-related degradation

Table 3. Structure of the approved ITU-T J.344 recommendation family

This structure reflects real deployment conditions. Video degradation is not caused by compression alone; video conferencing, live broadcasting, surveillance, and mobile camera applications are also affected by focus instability, camera motion, low-light noise, and other capture-side impairments.

VQML® was selected as Model A within the J.344 standardization framework. It satisfies the requirements of both J.344.1 and J.344.2. According to the final standardization records, VQML® is the only no-reference model registered across both recommendations and the sole registered model in J.344.2.

1.5 Standardization and Commercial Significance

The approved J.344 family gives telecom operators, regulatory authorities, device manufacturers, and digital video service providers a consistent, repeatable, and internationally recognized framework for video quality assessment. A common benchmark reduces ambiguity when comparing networks, devices, applications, and equipment vendors.

For LIG Accuver, the J.344 standardization process strengthens the commercial credibility of VQML® and provides a common technical basis for deploying its perceptual quality measurements within XCAL-based validation workflows.

2. Algorithm Overview: LIG Accuver VQML®

2.1 What Is VQML®?

VQML® (Video Quality assessment with Machine Learning) is LIG Accuver's proprietary AI-based no-reference video quality assessment engine. It estimates human-perceived quality using decoded RGB frames from the received video and outputs a predicted Mean Opinion Score (MOS) on a 1-to-5 scale: 1 Bad, 2 Poor, 3 Fair, 4 Good, and 5 Excellent.

The approved J.344 standardization scope focuses on Full HD signals. The commercial VQML® implementation has additionally been evaluated across resolutions from 144p to 2160p (4K UHD) and across major codec formats, including H.264, H.265/HEVC, and AV1. These broader commercial validation conditions are distinct from the formal J.344 evaluation scope.

2.2 Input and Output Structure

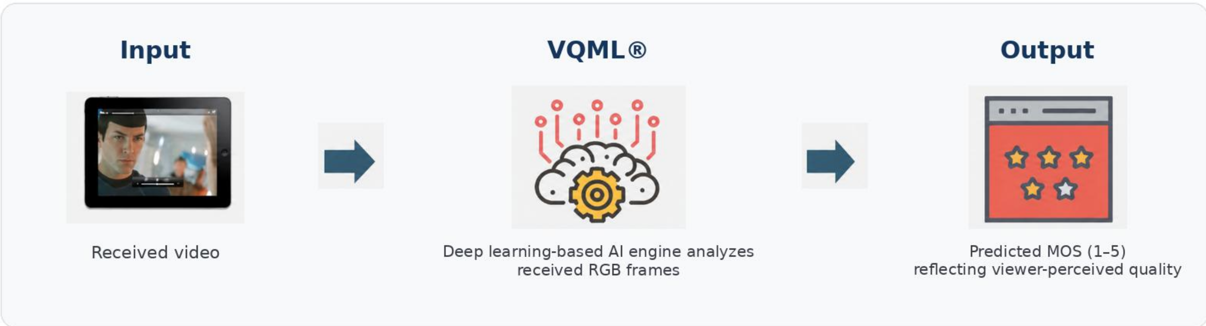


Figure 1. VQML® input, processing, and predicted MOS output

- Input: Decoded RGB frames captured from the received video through screen recording, HDMI capture, or virtual-camera interfaces.
- Processing: Deep learning modules analyze spatial artifacts, temporal consistency, content context, and device-related viewing conditions.
- Output: Time-based MOS estimates and a session-level average MOS, with supporting sub-scores for specific quality dimensions.

This streamlined input-output structure enables black-box testing of supported video, social-media, conferencing, and communication applications without requiring modification of the target application or access to its internal delivery metadata.

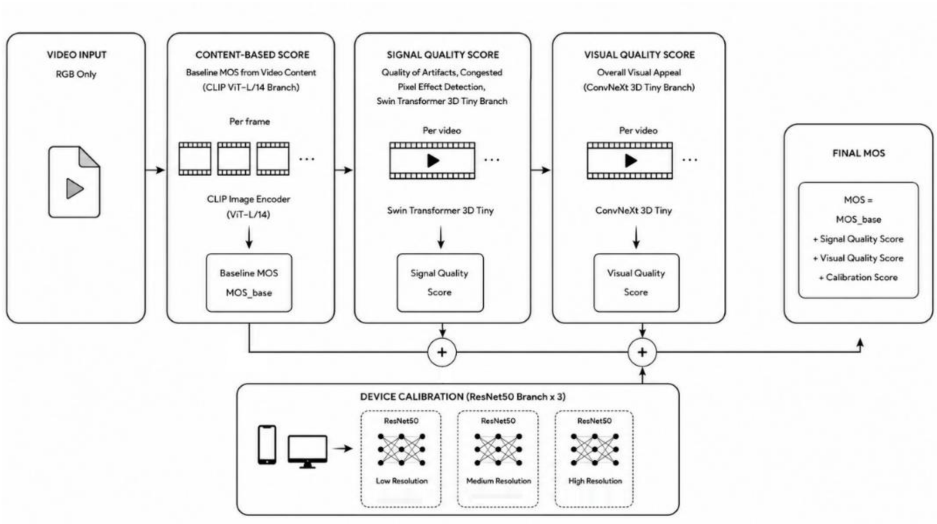
2.3 Model Architecture: Evolution from CNN + GRU to Transformer + CNN

Earlier VQA architectures commonly used Convolutional Neural Networks (CNNs) to extract localized spatial features—such as edges, textures, blockiness, and blur—from individual frames, while Gated Recurrent Units (GRUs) modeled short-term temporal changes. This structure was effective for basic coding artifacts but had limited ability to represent broader scene context, long-range temporal relationships, and semantic intent.

VQML® uses a Transformer + CNN hybrid architecture to analyze video impairments based on the semantic, technical, and aesthetic characteristics of the content. CNN-based components detect fine-grained technical artifacts, while self-attention modules interpret broader scene context and aesthetic intent, helping distinguish intentional characteristics—such as artistic blur, fast motion, or dark scenes—from actual quality degradation. The model then accounts for device-specific factors that affect perceived visual quality and synthesizes these results into a final MOS prediction.

2.4 Multi-Score VQA Architecture

Rather than producing an undifferentiated black-box score, VQML® evaluates multiple dimensions of visual quality and synthesizes them into the final predicted MOS:



.

Figure 2. Evolution from a conventional single-score architecture to the VQML® multi-score architecture

Module	Main Function	Role in MOS Prediction
Content-Based Score	Analyzes scene complexity, motion dynamics, and semantic characteristics.	Reduces false penalties by distinguishing intentional visual properties from technical distortion.
Signal Quality Score	Quantifies technical degradation associated with compression, quantization, transmission loss, and playback interruption.	Detects artifacts such as macroblocking, mosquito noise, ringing, and freezing.

Module	Main Function	Role in MOS Prediction
Visual Quality Score	Evaluates optical and capture-side impairments.	Accounts for camera defocus, motion shake, poor lighting, and sensor noise that influence real-world perception.
Device Calibration	Accounts for display resolution, pixel density, and rendering characteristics.	Adjusts the predicted quality estimate for the characteristics of the target viewing device.
Final MOS Engine	Synthesizes the branch-level scores through a learned prediction model.	Delivers a human-aligned predicted MOS on the 1-to-5 scale.

Table 4. Functions of the VQML® multi-score architecture

2.5 Performance Evaluation and Benchmark Results

VQML® was evaluated on the ITU-T standardization dataset under controlled and consistently applied test conditions. Pearson Linear Correlation Coefficient (PLCC) was used to measure how closely each objective model tracked human subjective quality scores. A value closer to 1 indicates stronger linear correlation.

VQML® achieved PLCC values of 0.8653 on the J.344.1 dataset, which focuses on coding artifacts, and 0.8077 on the J.344.2 dataset, which includes both coding artifacts and camera-related impairments. In terms of RMSE, VQML® achieved 0.4984 and 0.4870 for J.344.1 and J.344.2, respectively. VQML®'s performance level exceeds all other candidate models that were evaluated using the J.344 datasets and test conditions. VQML® was the only model whose average RMSE scored below 0.5. And furthermore, VQML® consistently resulted in high enough reliability score (measured in False Ranking Rate metric) to replace one human viewer when evaluated on each of the J.344 datasets. Together, these results demonstrate that VQML® provides stronger alignment with human subjective quality judgments, lower prediction error, and more reliable quality-ranking consistency under the evaluated conditions.

3. VQML®-Enabled QoE Assurance and XCAL Use Cases

3.1 From Network KPIs to User-Perceived QoE

In 5G SA and future 6G architectures, assurance is expanding from RF optimization to application- and SLA-level verification. RSRP, SINR, throughput, latency, and block error rates remain foundational, but they do not directly describe whether a subscriber experienced clear, stable, and responsive video.

VQML® adds an application-layer quality metric that can support SLA thresholds for premium video conferencing, cloud gaming, remote monitoring, and other real-time visual services. XCAL can evaluate predicted MOS together with initial loading behavior, adaptive-resolution changes, playback interruptions, and cellular signaling conditions.

3.2 Greater Interpretability through XCAL Integration

Conventional handcrafted NR models such as BRISQUE and NIQE rely primarily on natural-scene statistics. They can detect deviations from expected image statistics but provide limited semantic, temporal, and root-cause-level interpretability.

VQML®'s multi-score architecture provides more structured insight into the type of degradation being observed. When a signal-quality score drops because of blockiness or freezing, XCAL can align that timestamp with radio and protocol logs. This correlation helps engineers identify likely contributing factors, such as handover instability, scheduling congestion, uplink or downlink impairment, or physical-layer interference. Correlation narrows the investigation; final root-cause confirmation remains an engineering validation step.

3.3 Representative VQML® and XCAL Use Cases

1. Mobile Video Platforms and Short-Form QoE Benchmarking

With VQML® integrated into XCAL, operators can conduct automated drive tests across online video and short-form platforms such as YouTube, TikTok, Instagram, and local streaming services. Received RGB video and cellular signaling are captured under comparable mobility conditions, enabling objective comparison of predicted MOS, adaptive-resolution behavior, buffering, and visual artifacts across carrier networks.

2. Video Conferencing and Real-Time Collaboration Validation

Enterprise tools such as Zoom, Microsoft Teams, and Webex are sensitive to jitter, uplink packet loss, device processing, and camera quality. VQML® can estimate live two-way video quality without a source reference, while XCAL records the associated network conditions. Engineers can evaluate facial clarity, motion continuity, macroblocking, and session stability.

3. Mission-Critical Communications and Public-Safety Networks

VQML® has been commercially validated in disaster-network verification projects associated with Korea's public-safety communications environment. In mission-critical scenarios, the combined VQML® and XCAL workflow supports evaluation of whether live video retains sufficient visual clarity during congestion, mobility, and degraded radio conditions.

4. CCTV, Smart-City Surveillance, and Camera Benchmarking

Capture-side defects may occur before video enters the IP network. Focus drift, low-light noise, environmental vibration, and camera motion can reduce visual usability even when network KPIs are normal. The J.344.2-aligned capabilities of VQML® support scalable screening of camera feeds for both capture-side and coding-related degradation.

5. Controlled Laboratory Network Validation at Nokia Lab

At Nokia Lab, VQML® is used with XCAL's Virtual Camera function to benchmark network equipment and QoS configurations. A standardized high-definition test sequence is injected into a smartphone camera pipeline, transmitted through the test network, and decoded on the receiving device. VQML® then evaluates the received RGB stream. This method combines laboratory repeatability with realistic end-device behavior and helps isolate network-induced changes under controlled conditions.

6. 6G Immersive Media: XR, Spatial Computing, and 360-Degree Video

Immersive video introduces dynamic viewports, head-tracking, rendering latency, and spatially varying quality. These characteristics make conventional reference-based assessment operationally complex because the evaluated viewpoint changes with user movement. VQML®'s pixel-based NR architecture provides a foundation for future evaluation of viewport clarity, rendering artifacts, and perceptual consistency across AR, VR, and spatial-computing devices.

3.4 Deployment Configurations and Commercialization Roadmap

VQML® is planned for delivery in three deployment forms: GPU-server, PC, and mobile, allowing customers to select an implementation suited to their processing requirements, test scale, and operational environment. The GPU-server option is intended for centralized or high-volume processing, while optimized and lightweight PC and mobile implementations are being prepared for laboratory, field, and device-side measurement.

Within the XCAL ecosystem, VQML® will be provided through two primary measurement configurations. The first embeds VQML® into XCAL to evaluate the perceived quality of OTT and live-video services played on mobile devices. By analyzing decoded RGB video together with XCAL measurement data, this configuration supports mobile-service benchmarking, field testing, device comparison, and QoE monitoring under real network conditions.

The second configuration targets external playback devices. Video output from Apple TV, Chromecast, and comparable set-top boxes or streaming dongles can be ingested through an external HDMI video-capture device and evaluated using VQML®'s pixel-based no-reference analysis. This enables objective comparison across devices, applications, and service platforms without requiring access to the original source video or internal application metadata.

As the next stage of commercialization, LIG Accuver plans to offer VQML® as a commercially licensed software development kit (SDK). The SDK will enable customers to integrate the J.344-aligned assessment engine into their own products, validation systems, monitoring platforms, and service-assurance workflows, extending VQML® beyond the XCAL ecosystem.

4. Conclusion

Video quality assessment is becoming an essential component of network and service assurance. As 5G Standalone and future 6G networks support application-specific SLAs, private networks, mission-critical services, and immersive media, network-centric KPIs alone cannot fully represent the quality experienced by users.

VQML® addresses this gap through AI-based no-reference assessment of decoded RGB video. Its selection as Model A within the ITU-T J.344 standardization framework provides a common technical basis for Full HD no-reference video quality assessment, while the commercial implementation extends its applicability across broader resolutions, codecs, devices, and service environments.

Integrated with XCAL, VQML® connects predicted perceptual quality with radio, protocol, playback, and device-level information, enabling repeatable service comparison and more focused investigation of QoE degradation. Through future commercial SDK licensing, customers will also be able to embed the J.344-aligned VQML® engine into their own products, validation platforms, and service-assurance workflows.

This combination of international standardization, AI-based perceptual assessment, and flexible deployment positions VQML® as a scalable foundation for video QoE assurance in the 5G and 6G era.

Selected Standards and Technical References

- ITU-T J.144 — Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference.
- ITU-T J.246 — Perceptual visual quality measurement techniques for multimedia services over digital cable television networks using reduced reference.
- ITU-T J.343 series — Hybrid perceptual bitstream and pixel-based objective video quality measurement.
- ITU-T J.344, J.344.1, and J.344.2 — Approved no-reference objective video quality assessment framework and models for Full HD video.
- ITU-T P.910 — Subjective video quality assessment methods for multimedia applications.
- ITU-T P.1204 series — Video quality assessment models for streaming services.
- LIG Accuver VQML® evaluation results on the ITU-T standardization dataset and severe coding-artifact dataset (internally verified).

About LIG Accuver

LIG Accuver provides advanced test and validation solutions across 5G/6G, O-RAN, NTN, V2X, and mission-critical communications. Its portfolio supports network operators, equipment vendors, device manufacturers, automotive partners, and public-sector organizations across laboratory, field, and service-assurance environments.

For more information: sales@accuver.com | www.ligaccuver.com