

QENCODE

Video Intelligence is Now Native to the Modern Encoding Pipeline

The difference between video that just gets processed and video that is understood is likely the most expensive inefficiency in your stack.

FORMAT SUPPORT

99%

Input video formats and codecs supported

INTEGRATION

Native

TwelveLabs embedded directly in the encoding pipeline

PIPELINE

1 API

Encode and understand in a single call

RESOLUTION

Up to 16K UHD

Encoding with simultaneous AI analysis

The Stakes Just Got Higher



Every successful video platform is encoding more videos in 2026 than ever before, and the pace is only accelerating. AI-generated content alone is reshaping how much footage moves through transcoding pipelines every day.

Encoding has also gotten significantly smarter. With ML-driven compression, automatic captions and translations, smart cropping and thumbnails, and high-quality AI upscaling, the cost and quality curves of video experience and delivery have been collapsing. The output of a modern encoding pipeline is smaller, sharper, and more economical to ship than ever before.

It is also still semantically opaque. A platform can encode ten thousand hours of 4K footage overnight, and still have no idea what's in any of it until someone watches it.

That gap is currently the most expensive inefficiency in the stack. As content volumes accelerate, the cost of leaving videos “un-understood” keeps compounding. The platforms winning today make their videos searchable, analyzable, and understood immediately after they finish processing them.

Qencode has been dealing with these challenges directly, and they **finally solved them by embedding TwelveLabs directly into the encoding pipeline.**

The Biggest Gap in AI-Native Video Platforms

Before the TwelveLabs integration, Qencode had spent years building artificial intelligence into every part of their encoding stack. ML-powered encoding makes files 60% smaller at matched VMAF level, adapting settings to content type and treating sports, animation, UGC, and conference recordings as fundamentally different encoding challenges. Auto-Subtitles, speech-to-text across 70+ source languages with translation into 100+ target languages, AI-Detection, AI Upscaling, Smart Cropping, and Smart Thumbnails all run as outputs of the same job. Encoding has become smarter than ever, but until recently, one of the most fundamental intelligence features was still missing.

The encoder knew the difference between the best codec settings for a sports clip and a conference recording. It knew how to generate multilingual captions, crop intelligently for vertical formats, and surface representative thumbnails. **It did not know which scene contained the winning goal, the quote that made the keynote viral, or where to find the specific moment a viewer needs to see right now.**

The Qencode platform was AI-native at the pipeline layer, but completely clueless at the comprehension layer. When a social team needs a specific emotional reaction from a library of broadcast footage, they don't search, they scrub, clip by clip, episode by episode. When a compliance team needs to surface brand-safety violations across thousands of hours of encoded content, they can't find them without meticulously watching everything.

This is the default architecture for most video platforms operating at scale today, and it's increasingly becoming more untenable for everyone involved. As content volumes grow and AI-generated media accelerates production, the competitive advantage of solving this becomes more difficult to ignore.

Every hour a video spends waiting to be understood is an hour it can't be used. Putting more people to process this manually still adds a tremendous amount of latency, cost, and failure points that prevent content from flowing smoothly from the operations team to live on the platform. This is a foundational limitation of traditional infrastructure at scale, and it is increasingly becoming more expensive.

What Qencode Built, and Why It Changes the Equation

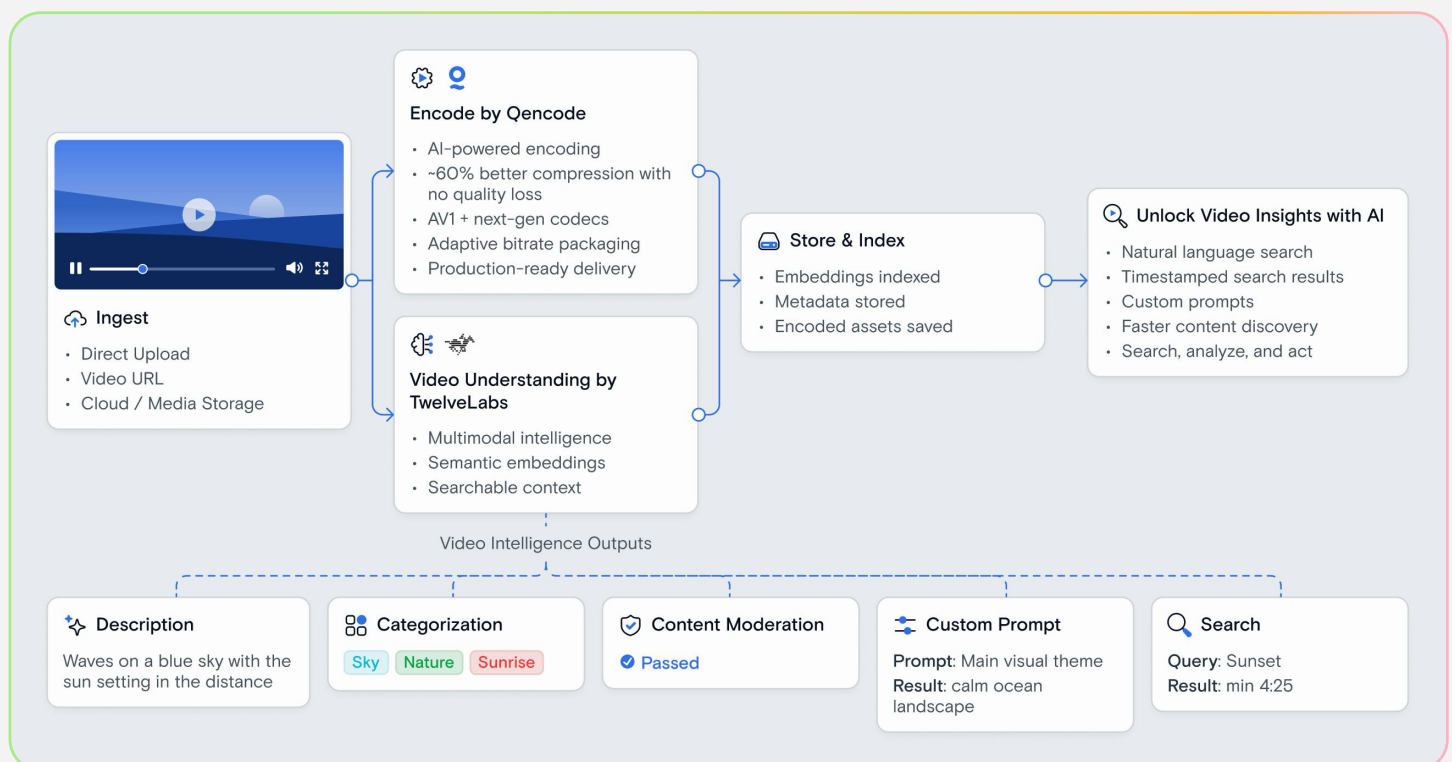
Qencode already handled some of the most challenging encoding problems: highly efficient encoding at scale, AI-powered codec optimization, adaptive bitrate packaging, and integration of most of the state-of-the-art AI features available across large video platforms today. What they added was the layer that makes all of that content intelligible, searchable, and actionable.

By embedding TwelveLabs' video understanding models directly into the Qencode platform, customers gain video intelligence as an additional output in their current encoding process.

Encode and understand, at the same time. Video enters the Qencode pipeline through the existing API with encoding, video file optimization, and adaptive packaging happening as usual. In parallel, TwelveLabs' Marengo model processes the same content and generates multimodal embeddings that capture what's visible, what's being said, what's written on screen, and how it all fits together. This is a fundamentally different approach from services that do frame-by-frame image recognition or audio transcription bolted onto a vision model. For Qencode, this was the first model that had a unified understanding of video as video.

Search in plain language. After encoding, Qencode can index every video automatically. Teams search their entire library with natural language queries, such as "highlights of the latest live performance," "outdoor scenes with crowd energy," and "any scene with a visible logo", and get back exact clips with timestamps instead of a list of files to scrub.

Intelligence without additional overhead. Since there is no separate vendor to manage, content to re-upload, or custom plumbing to bridge, it's never been easier to enrich and expand your content operations function.



What Qencode Built, and Why It Changes the Equation

THIS IS WHAT THE INTEGRATION UNLOCKS ACROSS USE CASES:

01

Media and Broadcasting

Automatically generate searchable archives, produce compliance-ready metadata, and enable editorial teams to find specific moments across thousands of hours, all inside the encoding workflow already running.

02

E-Commerce and Product Video

Detect and tag products, extract key selling moments, generate scene descriptions, turning raw product footage into structured, searchable commerce assets at the moment it's encoded.

03

Education and Enterprise

Make every training video, webinar, and corporate recording instantly searchable by topic, speaker, or concept without manual transcription or tagging workflows.

04

Security and Compliance

Encode and simultaneously analyze footage for specific events, objects, or conditions, reducing the gap between capture and actionable intelligence from days to seconds.

The TwelveLabs Advantage

Qencode evaluated the video AI landscape before committing to this integration. The decision came down to a specific technical distinction that matters for encoding platforms.

Most AI tools that claim to analyze video are image classifiers running on extracted frames, or speech-to-text models that treat audio as the primary signal. TwelveLabs' models (Marengo for multimodal embeddings and Pegasus for reasoning and generation) process the full video signal simultaneously: visual, audio, and textual, in context, across time.

For a platform serving customers who need to search for "the most heartfelt family moment" or "every scene with a product unboxing," Qencode finally had a solution that gave them exactly what they needed.

The integration was deployed rapidly, with TwelveLabs models embedded directly into Qencode's existing platform architecture. This made it super simple for anyone to add this capability to their existing pipeline without re-engineering it, or disrupting the existing customer API surface. Customers who are already encoding with Qencode can access video intelligence without changing anything else about how they work today.

"We evaluated the AI video understanding market carefully before choosing TwelveLabs. They were the only provider whose models truly **understood video** the way we always envisioned and **the way our customers actually needed**. They weren't just extracting frames and transcripts, but **indexing the full signal** in a way that was extremely useful, robust and flexible. The multi-modal combination of visual, spoken, and contextual layers absolutely blew our mind. Our imagination started running wild, and the use-cases for the diverse range of customers using our platform basically began writing themselves.

Murad Mordukhay
CEO, Qencode

Results

Qencode customers no longer run two pipelines. Encoding and video intelligence ship as one. That shift changes what's operationally possible at every scale.

- 01 **Content teams get searchable video immediately after encoding**, not after a separate ingestion process that takes hours or days.

- 02 Compliance and moderation workflows **run in parallel with delivery, not behind it.**


- 03 Editorial and social teams **find exact moments across large libraries without manual scrubbing.**

- 04 Platforms connecting AI-generated metadata to recommendation engines, ad targeting, or archive monetization do so **without building custom infrastructure to bridge their encoding stack and their AI stack.**

- 05 **As content volumes grow and AI-generated media accelerates, the integration scales with them.** The pipeline becomes more powerful as the library grows, without making it more complex to manage.

- 06 The video platforms that build this infrastructure **now are operating with a structural advantage over the ones still running parallel systems and paying for both.**

TECHNICAL ARCHITECTURE

Platform	Qencode cloud video services
AI Models	 TwelveLabs MARENGO (multimodal embeddings), PEGASUS (reasoning & generation)
Format Support	99% of input video codecs and containers
Resolution	Up to 16K UHD
Delivery	Global CDN
Integration	Native — embedded in the Qencode API surface
Pipeline Type	Unified — encode + understand in one processing job

About the Companies

TwelveLabs

TwelveLabs builds video foundation models that transform how machines understand video. Its multimodal AI platform enables enterprises to search, analyze, and generate insights from video at scale, understanding not just what's in a frame, but what's happening across an entire scene: the visuals, the audio, the text, and the context that connects them. As video volume accelerates across every industry, TwelveLabs is the infrastructure layer that makes it navigable. Based in San Francisco and Seoul. Serving customers across media, security, advertising, and enterprise software.

Qencode

Qencode is on a mission to democratize world-class video technology, making it simple for anyone to turn their idea into the next great video experience. Its full-stack video platform gives engineering teams everything they need to build and scale video products through an integrated platform with unified APIs spanning transcoding, live streaming, storage, content delivery, player, and analytics. Qencode's AI-powered optimization works seamlessly with your existing codecs, reducing file sizes by up to 60% without sacrificing quality. Based in Los Angeles, it empowers media companies to focus less on infrastructure and more on the things that truly make them great. cloud.qencode.com

Get in Touch

- twelve labs.io
- sales@twelve labs.io
- docs.twelve labs.io

