



# Rethink Enterprise and Mobile Applications with Visual Intelligence



*The AT&T Visual API is a network service that brings visual analysis tools (near-copy, face recognition, quality assessment, and speech-based retrieval) to your enterprise and mobile applications. With Visual Intelligence, content is transformed into easy-to-use textual and indexed representations. You provide the content and the Visual API provides visual intelligence.*

## DATASHEET

Rethink your enterprise services and mobile applications with the Visual API, powered by the AT&T Content Analysis Engine<sup>SM</sup> and the AT&T WATSON<sup>SM</sup> speech engine.

Over the last two decades, AT&T has continued to develop world-class content analysis and computer vision technology. Video segmentation, album management, audio and video near-copy detection, face detection and recognition, natural language-based search, and quality assessment are just a handful of exposed capabilities. Each technology is packaged as a network service for easy integration into your applications through the secure and hardened Alpha API program.

Deploy your own application using visual intelligence to provide capabilities customers never imagined. Automated enterprise content management, a face-based time-card system, quality comparisons of video codecs, or a speech-driven interface to your customer support videos are all possible with AT&T's Visual API.

Network services mean your application can run on mobile devices, desktops, or back-ends. You transfer the content, and we provide the visual intelligence.

### Visual API Benefits

- No additional software to develop and deploy to utilize the functionality of the Visual API.
- As a network service, capabilities are available to mobile, desktop, enterprise, and back-end solution providers.
- The Visual API creates visual intelligence for your content, imposing no storage constraints. That means you can store, stream or playback, and re-encode your content with no changes required for your existing solutions.
- Responses from the Visual API are bandwidth-friendly. Once the content is transmitted to the Visual API, lightweight XML or JSON can be retrieved by any network connected device.

### Functionality

The Visual API is grouped into the functional resources below, but content can be analyzed by each resource independently.

- Accounts
- Namespaces
- Assets
- Video Segmentation\*
- Faces
- Near-copy\*
- Collections
- Tags
- Quality
- NLU
- Metadata Injection\*



## Visual API Functionality

### Accounts Security

Create segregated and secure account credential for your user base. As a mobile or enterprise application developer, you won't lose cycles mapping internal and external references to the Visual API.

### Namespace Partitioning

Organize content into namespaces that can be shared across users in either read-only or fully editable fashions. This functionality supports crowd-sourced projects that include experts and observers alike.

### Video Segmentation\*

Video segmentation analyzes your content for logical scene and shot partitions. This function helps to automate editing by bootstrapping segments for telepresence calls, conferences, broadcast video, and even user-generated content.

### Asset Analysis

General content analysis for basic similarity, collection creation, and general library management. If you don't need specific tools, but want the most of your content, one suite of functions does it all for you.

### Face Detection and Recognition

Detection and recognition of faces from custom or pre-created contexts. Whether your content is celebrity photos, a user's photo album, or a corporate employee roster, face-based content organization can help.

### Near-copy Detection and Recognition\*

Near-copy functionality can be used to both identify a low-quality originals (e.g., matching to an online library) and help regulate rights infringements utilizing both audio and visual cues.

## What is the AT&T Content Analysis Engine<sup>SM</sup>?

The AT&T CAE<sup>SM</sup> (Content Analysis Engine) defines a cutting-edge suite of analysis and metadata generation algorithms. These algorithms aim to automatically apply "visual intelligence" to segment content (e.g., videos into shorter segments or single frames) and represent the content with human understandable metadata. The Visual API delivers the CAE over with secure HTTP packaged resources and is deployed via the AT&T Alpha API program to accommodate large-scale applications with ease.

Building on over two decades of work in AT&T Labs - Research, the AT&T CAE<sup>SM</sup> harnesses algorithms developed and vetted by international industry and academic evaluations. Described in more depth above, these algorithms analyze low-level visual features using motion, color, texture, and object-specific cues to produce machine-readable metadata.

Many functions of the CAE have been defined to operate on a very granular level (usually single images or other and low-bandwidth resources) so that they can distribute the request efficiently and

### Collection Creation and Organization

A light-weight version of near-copy detection, similar photos or videos from different moments in time, camera angles, and illuminations can be used to recommended albums or virtual stacks.

### Tag Input and Retrieval

Label faces with human input to bootstrap automatic tagging for a wealth of applications in recognition, entertainment, and retrieval.

### Perceptual Quality Scoring

Automated perceptual quality scoring of frames, video, and regions based on leading research metrics. These scores go beyond simple energy metrics defined in international specifications to give you precise information about quality degradation as perceived by a human.

### Natural Language Understanding (NLU)

Custom speech models built for AT&T WATSON<sup>SM</sup> using your content. Define your own categories (genre, model, make, year, etc.) and combine them with speech recognition output (no transcripts required) for a precise mapping and indexing scheme.

### Metadata Injection\*

Supporting the popularity of mix-and-match systems, let the Visual API augment your existing content metadata. Injected metadata supports certain functions (e.g., NLU) without providing content for full analysis.

respond with minimal latency. Other functions that perform the analysis of long-form content (e.g., videos) may require lengthy processing times and large data transfers, but the Visual API and AT&T CAE<sup>SM</sup> schedule this analysis across a dynamically allocated array of resources to perform at scale.

In contrast to other available APIs, the AT&T Visual API is designed to providing visual intelligence and contexts that are specific to your content. This design choice means that any recognition tasks, model generation, content segmentation, and even collection formulation are driven by your content to provide the most accurate information possible.

### Learn more about the AT&T Content Analysis Engine<sup>SM</sup> at:

<http://www.research.att.com/projects/Video/>

*\* These functionalities exist in the AT&T CAE<sup>SM</sup> but their deployment into the Visual API is still ongoing. If you are interested in one of these technologies, please ask us about how to get it delivered to you more quickly!*

FOR MORE INFORMATION VISIT  
[www.research.att.com/projects/Video/](http://www.research.att.com/projects/Video/)