

Estudio del estado del arte del etiquetado automático de contenidos audiovisuales

Mayo 2022

Este estudio ha sido realizado con financiación del Ministerio de Asuntos Económicos y Transformación Digital, a través de la Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales (SETELECO), que no comparte necesariamente los contenidos expresados en el mismo. Dichos contenidos son responsabilidad exclusiva de sus autores. Se permite su copia y distribución por cualquier medio siempre que se mantenga el reconocimiento de sus autores, no se haga uso comercial de las obras y no se realice ninguna modificación de las mismas.

Índice

1.PREIDENTIFICACIÓN DE TECNOLOGÍAS	3
1.1.Overview	3
1.2.Fichas de las soluciones tecnológicas identificadas	5
2.ANÁLISIS DE TECNOLOGÍAS POTENCIALES	25
2.1.Análisis general del stack tecnológico	25
2.1.1. Capa de etiquetado según tipo de fuente	28
2.1.2. Capa de moderación / calificación	30
2.2.Análisis de modelos potenciales de soluciones tecnológicas	31
2.2.1. Modelos preentrenados (taxonomía propia)	31
2.2.2. Modelos con taxonomía ad-hoc	34
3.RECOMENDACIONES	37
3.1.Modelo de desarrollo de una solución común auto-corregulada	37
3.1.1. Modelo tecnológico	38
3.1.2. Modelo organizativo	40
3.1.3. Instrumentos de contratación pública	42
3.2.Modelo de autorregulado de certificación de soluciones	44

1. Preidentificación de tecnologías

1.1. Overview

El etiquetado automático de contenidos audiovisuales es una actividad en auge impulsada por el desarrollo de modelos de Inteligencia Artificial que permiten automatizar una buena parte de este proceso.

Los metadatos generados son potencialmente utilizados para multitud de análisis, que van desde la identificación de contenido inadecuado hasta la alimentación de sistemas recomendadores para la mejora de la experiencia del usuario en plataformas audiovisuales.

La exploración del mercado de soluciones de etiquetado automático de contenidos audiovisuales deja de manifiesto la proliferación de soluciones de moderación de contenidos generados por el usuario. Estos escenarios son muy exigentes en términos de tiempos de respuesta para no penalizar la experiencia de los usuarios cuando contribuyen en una plataforma social.

Esta exigencia de eficacia y eficiencia ha impulsado el desarrollo y la mejora de modelos análisis y aprendizaje basados en Inteligencia Artificial, contribuyendo a madurar las tecnologías de etiquetados, consiguiendo niveles muy precisos de identificación para determinados elementos.

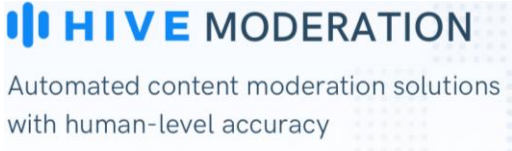

Salvo para entornos de vigilancia, donde la imagen es la fuente esencial de información potencialmente analizable, los análisis de texto y de video son los que mayor peso tienen en las soluciones analizadas. El análisis del audio añade una capa adicional de comprensión del contexto, pero si se disponen de los diálogos transcritos, no es absolutamente indispensable alcanzar este nivel para identificar contenido sensible con precisión.

Para realizar la identificación de las principales tecnologías y modelos implicados en el proceso de etiquetado automático de contenidos audiovisuales, atendiendo a la metodología propuesta, se procedió a hacer una búsqueda por keywords en Crunchbase para identificar las soluciones más importantes presentes y funcionando en el mercado. De este modo, la investigación parte de soluciones que tienen una penetración real en el mercado.

A continuación, se muestra la lista de las 20 empresas mejor valoradas en virtud de su CBRank y en el siguiente epígrafe incluimos una ficha resumen de cada una de ellas.

Compañía	País	Fuentes			Funding	Web
		Texto	Audio	Video		
Hive Moderation	USA	Texto	Audio	Video	\$ 120.700.000	https://hivemoderation.com/
AssemblyAI	USA	Texto	Audio		\$ 33.100.000	https://www.assemblyai.com/
Clarifai	USA	Texto	Audio	Video	\$ 100.000.000	https://www.clarifai.com/services/data-labeling
Bodyguard	Francia	Texto			\$ 9.900.000	https://www.bodyguard.ai/
Spectrum Labs	USA	Texto	Audio		\$ 45.000.000	https://www.spectrumlabsai.com/
Active Fence	Israel	Texto	Audio	Video	\$ 100.000.000	https://www.activefence.com/solutions/harmful-content-detection/
Two Hat (Microsoft)	Canadá	Texto	Audio	Video	\$ 10.700.000	https://www.twohat.com/solutions/content-moderation-platform/
Besedo	Suecia	Texto	Audio	Video	\$ 3.300.000	https://besedo.com/product/customized-ai-moderation-models/
Lucid	USA	Texto	Audio	Video	\$ 2.800.000	https://lucidvideo.ai/automated-video-content-tagging/
Filestack	USA	Texto	Audio	Video	\$ 1.800.000	https://www.filestack.com/products/intelligence/
Cinfo	España			Video	\$ 7.100.000	https://www.cinfo.es/
Appen	Australia			Video	\$ 587.450.000	https://appen.com/blog/content-moderation/
Valossa	Finlandia	Texto		Video	\$ 2.700.000	https://valossa.com/video-moderation-metadata/
Utopia Analytics	Finlandia	Texto			\$ 1.200.000	https://utopiaanalytics.com/utopia-ai-moderator/
Sensifai	Bélgica	Texto	Audio	Video	\$ 1.500.000	https://sensifai.com/
Smart Moderation	USA	Texto			\$ 239.500	https://smartmoderation.com/
DragonflAI	UK	Texto	Audio	Video	\$ 114.700	https://www.dragonflai.co/
Anolytics	India	Texto	Audio	Video	No publicado	https://www.anolytics.ai/
WebPurify	USA	Texto	Audio	Video	No publicado	https://www.webpurify.com/video-moderation/
Fassoo	Alemania			Video	No publicado	https://www.fassoo.com/

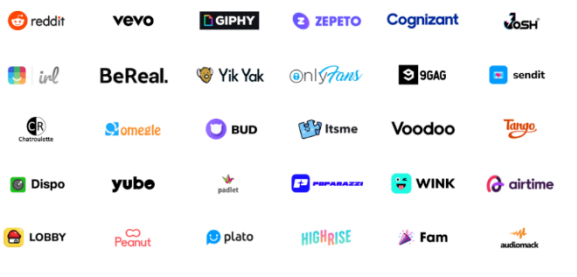
1.2. Fichas de las soluciones tecnológicas identificadas

Nombre	Hive Moderation			
País	USA			
Funding	\$120.700.000			
Fuentes	Video	Texto	Audio	
Website	hivemoderation.com			

Solución completa para la moderación de contenidos generados por el usuario. Soporta elementos visuales (imágenes y video), texto y audio. Disponen de modelos muy entrenados para la detección de contenido sensible y ofrecen acceso a sus modelos vía API, que reciben contenido y devuelve la metadata de los elementos detectados. Los videos son analizados frame a frame. Soporta una gran variedad de idiomas. Su modelo de pricing está basado en cuatro componentes:

- Suscripción al dashboard (con un fee mensual)
- Coste unitario por elemento que se envíe analizar:
 - Visual Moderation API: \$3.00 / 1000 imágenes o frames
 - Text Moderation API: \$0.50 / 1000 mensajes (hasta 1024 caracteres por mensaje)
 - Audio Moderation API: \$1.50 / hora de audio

We power 150+ leading platforms including:

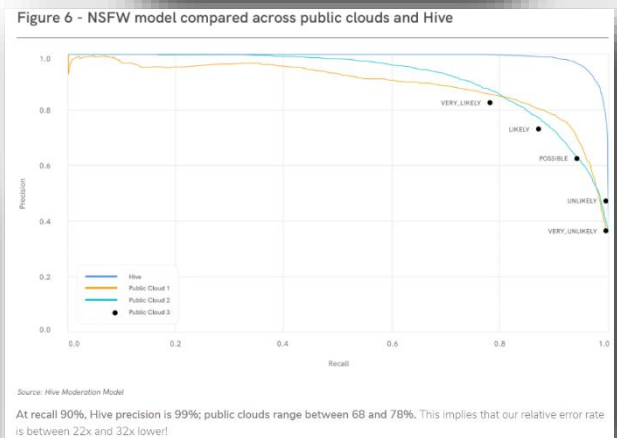


Our solution processes all media types and returns tagged metadata in real time

IMAGES VIDEOS GIFS
WEBP AUDIO LIVE STREAMS

One API to handle all content types

- VISUAL MODERATION**
Real-time moderation of visual content
- TEXT MODERATION**
Comprehensive moderation of textual strings and text in media
- AUDIO MODERATION**
Real-time audio transcription and moderation
- MODERATION DASHBOARD**
All-in-one content moderation tool



Nombre	AssemblyAI		
País	USA		
Funding	\$30.100.000		
Fuentes			Audio
Website	assemblyai.com		



AssemblyAI

AssemblyAI es especialista en el análisis del audio. Su foco se centra en la moderación de contenidos generados por el usuario.

Soporta español, aunque no están disponibles aun todas las características.

Realizan tanto la transcripción del audio a texto como análisis más complejos del audio, incluyendo Sentiment Analysis y moderación de contenido sensible.

Como todas las soluciones que se ocupan de este aspecto, su velocidad de ejecución es muy alta para ser aplicada a conversaciones en tiempo real.

Su modelo de precio se tarifica en segundos de audio analizados.

Core Transcription

\$0.00025 per second

Automatically convert audio/video files, and live audio streams, into accurate transcriptions with a simple API. Powered by cutting edge research into Automatic Speech Recognition.

- ✓ Asynchronous transcription
- ✓ Real-time streaming transcription
- ✓ Process multiple files/streams in parallel
- ✓ State-of-the-Art accuracy
- ✓ Billed per second transcribed

[Get an API token](#)

Audio Intelligence


\$0.000167 per second

Leverage cutting edge AI research to build powerful applications and features with higher ROI. Audio Intelligence features analyze the speech in your audio data to provide you with unparalleled insights about your data.

- ✓ Summarization
- ✓ Sentiment Analysis
- ✓ Entity Detection
- ✓ PII Redaction
- ✓ Content Moderation

[Get an API token](#)

AssemblyAI's Speech-to-Text APIs are trusted by companies of every size - from startups to Fortune 500. [Read more](#)



Content Moderation

With **Content Safety Detection**, AssemblyAI can detect if any of the following sensitive content is spoken in your audio/video files, and pinpoint exactly when and what was spoken:

LABEL	DESCRIPTION	MODEL OUTPUT	SUPPORTED BY SEVERITY SCORE
Accidents	Any man-made incident that happens unexpectedly and results in damage, injury, or death.	accidents	Yes
Alcohol	Content that discusses any alcoholic beverage or its consumption.	alcohol	Yes
Company Financials	Content that discusses any sensitive company financial information.	financials	No
Crime Violence	Content that discusses any type of criminal activity or extreme violence that is criminal in nature.	crime_violence	Yes
Drugs	Content that discusses illegal drugs or their usage.	drugs	Yes
Gambling	Includes gambling on casino-based games such as poker,	gambling	Yes

WILLEM DAFOE 1s
Hi. I'm Willem Dafoe, and this is the timeline of my career.

[TOPIC Careers](#)
[TOPIC Pop Culture](#)
[ENTITY Willem Dafoe](#)
[SUMMARY Timeline of Willem Dafoe's career](#)

WILLEM DAFOE 6s
I never really decided I wanted to be an actor. I always kind of thought I'd end up doing something differently. And then enough time passed where I thought, I guess I'm an actor. I've been doing this for a while.

[TOPIC Theater](#)
[TOPIC Movies](#)
[SENTIMENT Neutral](#)

WILLEM DAFOE 17s
I went to school for a little while in an acting program. But I dropped out very quickly, and I was reading about people like the performance group, the Manhattan Project, Grotowski, Robert Wilson. So I went to Mecca, which was New York.

[TOPIC Indie and Arthouse Movies](#)
[TOPIC The Manhattan Project](#)
[SENTIMENT Neutral](#)
[ENTITY Grotowski](#)



[ENTITY Robert Wilson](#)
[SUMMARY I went to Mecca, which was New York](#)

WILLEM DAFOE 32s
There was part of me that thought I would try to be a traditional actor. But I kept on finding myself going downtown, going to loft performances, sing, dance. And I got involved with a company called The Wooster Group that I ended up being with for 27 years. And that was my day to day life. My first IMDb credit is Heaven's Gate.

[TOPIC Musicals](#)
[TOPIC Theater Venues and Events](#)
[SENTIMENT Positive](#)

[SUMMARY I got involved with a company called The Wooster Group](#)

Estudio del estado del arte del etiquetado automático de contenidos audiovisuales

Nombre	Clarifai			
País	USA			
Funding	\$100.000.000			
Fuentes	Video	Texto	Audio	
Website	clarifai.com			

Clarifai forma parte de la familia de soluciones de moderación de contenidos que analiza video, texto y audio.

Está orientada sobre todo al contenido generado por el usuario, por lo que la precisión y la rapidez de detección son críticos.

Además, ofrece un interesante servicio de creación de datasets para el entrenamiento de la IA con objeto de reducir costes y escalar más rápidamente proyectos de modelado IA.

BENEFITS

AI-powered content moderation can be used to protect your brand, gain insights into customer sentiment, make it safer for customers to engage with your content, and ensure legal and brand compliance. Using AI, all of this can be done on a large scale and across multiple channels — with greater speed and improved accuracy.

AI-automation filters unwanted content in images, video, and text 100x faster and with greater accuracy.

- ✓ Reduce the harmful effects on individual moderators of seeing inappropriate images.
- ✓ Improve review times by 98% and scale quickly to keep up with increased content volume.
- ✓ Reduce time spent posting content by 95%.
- ✓ Scale campaigns confidently without worry of negative effects on your brand.

Screen content 100x faster with AI-automated content moderation.

Whether you're looking to filter out NSFW images, screen for copy infringement, or flag profane videos, Clarifai's automated content moderation solutions filter out unwanted content 100x faster than human moderators and with greater accuracy.

98%

Reduction in time spent to post user generated content

100x

Faster than human moderators with AI-automation

90%

Reduction in manual work using AI driven moderation

Protect your online customers and brand with AI-automated content moderation

Protect online communities
Recognize images containing illegal content. Detect toxic, obscene, racist, or threatening language. Protect your online community from trust and safety risks.

Find off-brand or poor quality content
Identify content that poses a threat of copyright infringement or doesn't meet brand guidelines. Screen for low quality images and logos and outdated content.



Filter profanity and toxic sentiment in text
Review text content faster. Monitor product or service reviews, customer chat logs, and social media posts to identify and remove content that could impact your brand and offend your customers.

Moderate embedded text in images
Combine computer vision to classify images, OCR to extract image text, and NLP to classify text in order to detect toxic, offensive, and suggestive content in social posts.

Data labeling services to reduce overhead and scale AI modeling projects faster.

High quality labeled data is the most important factor in building effective deep learning models. When it comes to data labeling services, we offer something that few others in AI can—**Security, Accuracy and Experience.**

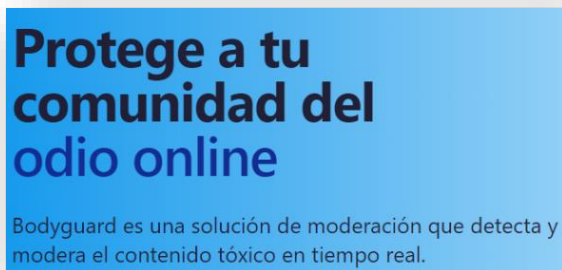
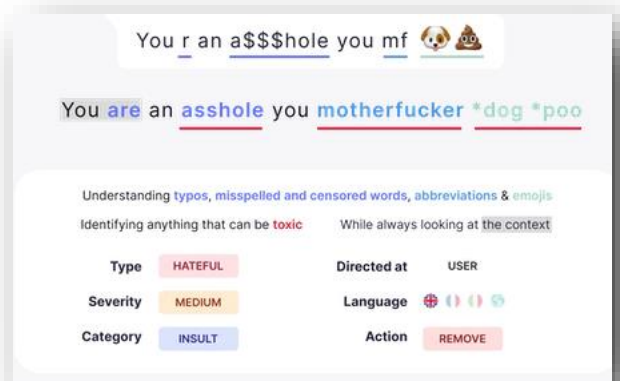
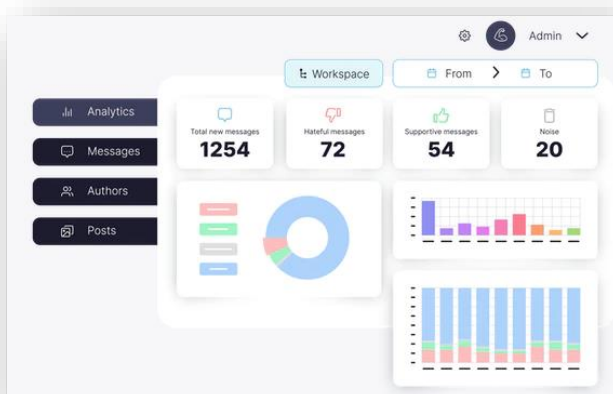
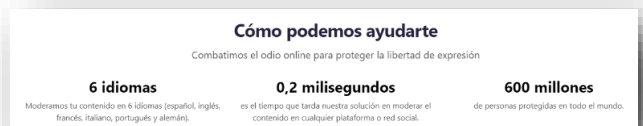
Outsource your data labeling tasks—images, video and documents—to a team of specialized data annotators and researchers. We combine cutting edge **AI-automation** on a secure **AI platform** to deliver you the highest quality training datasets you need for the most complex use cases.



Nombre	Bodyguard		
País	Francia		
Funding	\$9.900.000		
Fuentes		Texto	
Website	bodyguard.ai		

Bodyguard es una solución de moderación para contenido generado por el usuario, especialmente comentarios y conversaciones en comunidades.

Únicamente trabaja con textos, por lo que su orientación está fundamentalmente en evitar comportamientos abusivos, ilegales o impropios en comunidades de usuarios.

Es un servicio gratuito para particulares de cara a identificar y eliminar comentarios abusivos en las redes de sus clientes. Y es un servicio de pago para las plataformas sociales, segmento en el que se basa su modelo de negocio. Es tecnología europea.

Nombre	Spectrum Labs			
País	USA			
Funding	\$45.000.000			
Fuentes		Texto	Audio	
Website	spectrumlabsai.com			

Spectrum Labs también es una solución de moderación para contenido generado por el usuario. En este caso se centra en contenido de texto y en audio, concretamente en el análisis de la voz.

Detecta comportamientos no adecuados en las interacciones de los usuarios. Su IA genera metadata a partir de los contenidos generados por el usuario y utiliza esta información para reconocer estas actitudes no permitidas.

Es capaz de reconocer discursos de odio & radicalización, contenido sexual, spam, comportamiento autolesivo, abuso de menores, amenazas, violencia y socialización del sexo, entre otros.


Detect & Stop Disruptive Behaviors in Your Community

Toxic user behavior in online communities is a massive, accelerating problem. Hate speech, bullying and spam drive users away and erode your brand value. Underage users and sex solicitation create legal risks.


Spectrum Labs detects harmful behaviors in text and voice content across languages. Our Contextual AI finds behaviors that other solutions miss. We help you take automated, effective action to protect your users and build a brighter, growing community.

Industries


Every community has different content moderation policies and needs. See how we've helped Trust & Safety professionals in your industry.




Social Network



Dating



Gaming





Marketplaces

Sophisticated Behavior Detection

Some toxic behaviors are easy to find, but others depend on context. Our Contextual AI looks at metadata from user activity on your platform like conversations, when and where they happen, and past user behaviors. This reveals patterns to find behaviors that other solutions miss.

- Hate Speech & Radicalization
- Sexual Content
- Bullying & Harassment
- CSAM Grooming
- Scams/Spam

<p>Other companies</p> <ul style="list-style-type: none"> _____ Poor coverage of complex changing behaviors _____ Not configurable to meet unique needs _____ Slow and expensive to add new languages _____ Not scalable to handle increasing volume _____ Lack important components to achieve results 	<p>Spectrum Labs</p> <ul style="list-style-type: none"> ✓ Cutting-edge Contextual AI behavior detection _____ ✓ Patented AI multi-language approach _____ ✓ Automated, effective user- and content-level action _____ ✓ Analytics for visibility & dedicated customer success _____ ✓ Active community of Trust & Safety professionals
--	--

Nombre	ActiveFence			
País	Israel 			
Funding	\$100.000.000			
Fuentes	Video	Texto	Audio	
Website	activefence.com			

ActiveFence es una empresa israelí de inteligencia que presta servicio a organizaciones públicas y privadas de todo el mundo.

Entre sus servicios cuenta con detección de contenidos potencialmente dañinos. Se adapta a cualquier lenguaje y es capaz de procesar prácticamente cualquier tipo de fuente. Entre sus clientes audiovisuales más destacados se encuentra Vimeo.

Como parte de sus servicios de inteligencia, esta compañía hace rastreo activo de la web y de la web profunda para detectar actividad delictiva o potencialmente peligrosa para la seguridad o contra la legalidad.

Protecting from Multiple Abuses

Eliminate blindspots of unwanted activities across multiple languages, content formats, and abuse areas to give your users a safe and trusted experience.

- TERROR & EXTREMISM
- HATE SPEECH
- DISINFORMATION & MISINFORMATION
- CHILD SAFETY
- HUMAN TRAFFICKING
- FRAUD & SCAMS
- ILLEGAL GOODS
- COPYRIGHT INFRINGEMENT

PROACTIVE DETECTION

Across every corner of the web

Our proactive content detection capabilities allow us to monitor and analyze malicious content and chatter from the clear, deep, and dark web so you can know about and act on threats before they affect your platform or your users.

- ALTERNATIVE SOCIAL NETWORKS
- MESSAGING APPS
- DARK WEB
- INVITE-ONLY FORUMS
- PASTEBINS
- FILE SHARING SERVICES

Multi-Language & Regional Support

Combining AI-based technology and regional expertise, ActiveFence detects malicious content across multiple languages and dialects to provide Trust & Safety teams with the full context.



Monitor Across Content Formats

ActiveFence monitors and detects malicious behavior and harmful content no matter what the format, including video, audio, image, text, binaries, and live streaming.

Trust & Safety at Scale

ActiveFence automatically collects data from millions of sources and applies contextual AI to power trust and safety operations of any size.

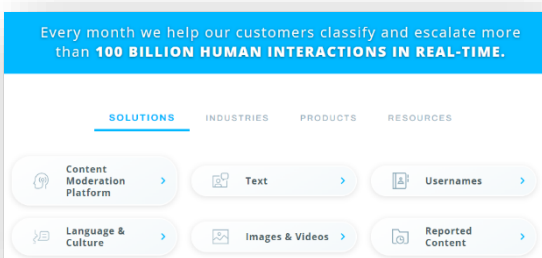
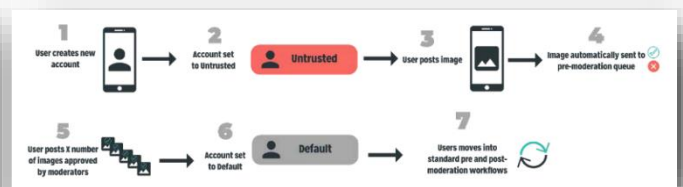
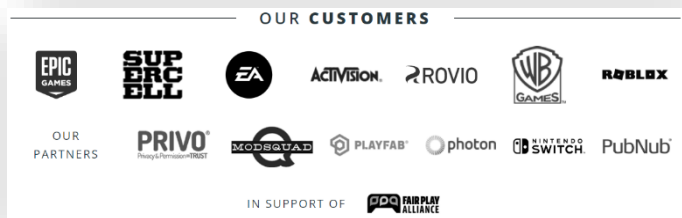
COLLECTING	ANALYZING	PROTECTING
10M+	130M+	3B+
SOURCES	SIGNALS DAILY	USERS

Nombre	Two Hat Security (Microsoft)			
País	Canadá 			
Funding	\$10.700.000			
Fuentes	Video	Texto	Audio	
Website	twohat.com			

Two Hat es una empresa canadiense de moderación de contenido adquirida por Microsoft en octubre de 2021.

Tiene una fuerte presencia en la industria del Gaming, con clientes tan relevantes como EA, Epic Games, Roblox o WB Games, entre otros. También prestan servicio en otro tipo de plataformas, como redes sociales, plataformas infantiles y comunidades online.

Su principal foco es la moderación de contenidos generados por los usuarios en tiempo real, ya sean texto, audio o video, con el objetivo común de este tipo de soluciones: detectar el contenido abusivo y potencialmente dañino para el usuario, especialmente para los menores de edad.





As technology evolves, more and more images and videos are uploaded and shared online. While most are innocent, some contain offensive or unwanted content. Disturbingly, sometimes that includes child abuse.

Unwanted exposure to violent, abusive or pornographic images has a profoundly negative effect on users, brands, and moderators. The problem is too big for human moderators alone to solve.

With our cutting-edge image scanning technology, social networks can **detect and remove dangerous and illegal content from their platforms in real-time**. Use our one-of-a-kind ensemble model to automatically identify **pornography, extremism, gore, weapons, and drugs** and drastically reduce your users' and moderators' exposure to NSFW content.

The experts on our Client Success team will work with you to build optimal workflows and leverage triaging techniques to ensure you get the most out of our image and video moderation solution. With up to 98.97% accuracy, you can rest easy knowing that your users, your brand, and your moderators are protected from undue exposure to potentially damaging content.

And through our work with law enforcement, we now provide **child sexual abuse material (CSAM) detection for social networks**. Find out how our [collaboration with Canadian law enforcement led to CEASE.ai](#).

Nombre	Besedo		
País	Suecia		
Funding	\$3.300.000		
Fuentes	Video	Texto	
Website	besedo.com		

Besedo es una solución de moderación que combina la IA con moderadores humanos, que tienen en esta compañía mucho peso, no sólo para el entrenamiento de datasets, sino para la ejecución diaria de actividades de moderación.

Su foco está en el contenido generado por el usuario en comunidades y plataformas sociales. Trabajan sobre todo con texto, utilizando reglas y listas negras de palabras para llevar a cabo una primera etapa de filtrado automático.

En cuanto al análisis visual, hacen análisis de imágenes estáticas. Para los videos utilizan un muestreo de 1 frame por segundo para eficientar el proceso de análisis.

Su plataforma está adaptada para la moderación manual. Están abiertos a trabajar con taxonomía del cliente, utilizando su plantilla de moderación para entrenar a la IA de cara a general los datasets inicial para llevar el nivel automático de filtrado a más de un 90%.

Discovery and consultancy
Discuss your needs and challenges with a solutions designer. We will provide you with ideas to streamline and improve your policies and implement exactly what you need to keep your user base happy and your site safe.

Send data through API
The Implo API is built on a RESTful architecture which allows easy integration.

Content is reviewed 24/7
We review your content according to your policies with a mix of filter automation, AI moderation, and highly trained human experts.

Quality assurance and monthly reports
We deliver a report with actionable insights so you can understand content threat areas and user behavior better.

ebay **anibis.ch** **Kaidee** **subtto** **connected2.me**

This is where we find it
To handle the complex nature of user generated content you need a multi-pronged content moderation strategy.

1-2-1 messages	✓	Ads/listings	✓
Comments/forums	✓	Profiles	✓
Reviews	✓	Ratings	✓
Images	✓	Video	✓

Fraud
2% of all user generated content is fraudulent. Protect your users against romance scams, pyramid schemes or "too good to be true" offers.

De-duplication
75% of users encounter duplicate content when browsing online marketplaces. Improve user experience and onsite SEO with de-duplication tech.

Image quality
100% of users state that image quality is the main reason for preferring one site over another. Remove stock photos, blurry images, and other low-quality pictures.



Nudity
Keep your site family-friendly by rejecting inappropriate pictures and preventing unsolicited offers before they go live.

Cyberbullying
13% of people have stopped using an online service after observing harassment of others. Stop both targeted harassment, hate speech, and social bullying.

Counterfeits
Counterfeits are consistently one of the top 3 items we reject. Improve user experience and avoid lawsuits by removing fake products from your site.

Illegal Items
Comply with ever more stringent government regulations by removing illegal items like drugs, weapons or culturally sensitive content.

Categorization
73% would not return to a site where they encountered irrelevant items. Increase conversion and reduce churn by ensuring correct categorization.


Nombre	Lucid		
País	USA		
Funding	\$2.800.000		
Fuentes	Video		
Website	lucidvideo.ai		

Lucid video es una solución de etiquetado automático de contenidos audiovisuales que va orientada a maximizar el rendimiento de los videos que publican los creadores de contenido.

El uso de tecnología de etiquetado automático y generación de metadata por IA está focalizado en este caso en ofrecer servicios de valor añadido a la experiencia del consumidor para adaptarla y ajustarla a sus hábitos de consumo: videos cortos (trailers, previews, highlights), recomendación de otros contenidos relacionados.

Apply Best-In-Class Metadata Extraction

Once added to the Lucid portal via simple upload or MRSS feed function, your content will be cataloged with detailed metadata such as actors, emotions, scene locations, activities, brands, products, objects, and more. Our AI assistant does all the work, saving you a significant amount of manual tagging time or money by not outsourcing tagging to a third-party service provider.



Reporter On Location
Automatically Generated Tags

- Reporter
- Office Building
- Outdoors
- Vaccine
- Coronavirus
- Microphone
- Road
- Eyeglasses
- Mask
- Facemask

Basic	Pro	Enterprise
\$59 /mo	\$499 /mo	Custom
Free Trial	Limited Free Trial	Negotiable Trial Period
<ul style="list-style-type: none"> 1 user account 10 videos 3 previews per video 15s & 30s preview lengths 3 or 5 segments per preview 	<ul style="list-style-type: none"> 5 user accounts 100 videos 5 previews per video 15s, 30s, 40s & 60s preview lengths 3, 5, 10, or 15 segments per preview 	<ul style="list-style-type: none"> Team accounts 1000+ videos 10+ previews per video Custom preview length Custom segment count per video
<p>Features</p> <ul style="list-style-type: none"> Free 7 day trial No watermarks on videos Download or embed preview videos Fast, responsive video player Web SDK Video transcoding & adaptive streaming High performance video CDN 	<p>Features</p> <ul style="list-style-type: none"> Free 7 day trial of Basic plan No watermarks on videos Download or embed preview videos Fast, responsive video player Web SDK Video transcoding & adaptive streaming High performance video CDN 	<p>Features</p> <ul style="list-style-type: none"> Sales-lead trial period Negotiable trial period User profile/genome generation Web SDK & integration Video transcoding and adaptive streaming MRSS feed integration

Powerful Artificial Intelligence Helps You



- Automatically edit short videos that people want to watch
- Automatically tag & moderate videos around-the-clock
- Automatically recommend your videos that hook viewers
- Automatically view user insights with video analytics

Short Videos, Highlights & Previews Created By AI

We understand the pain of producing and editing short videos to improve click through and conversion rates. **What if there was a way to have a superhuman editing team at your command?**

Our AI assistant automates tedious editing tasks, allowing one-click creation, downloading, sharing, or embedding of short preview videos in place of still thumbnails. **Work that once took you days can now be scaled to minutes.**

We don't stop there. As more people view, click, and spend time on your videos, **we continue to improve our editing.**

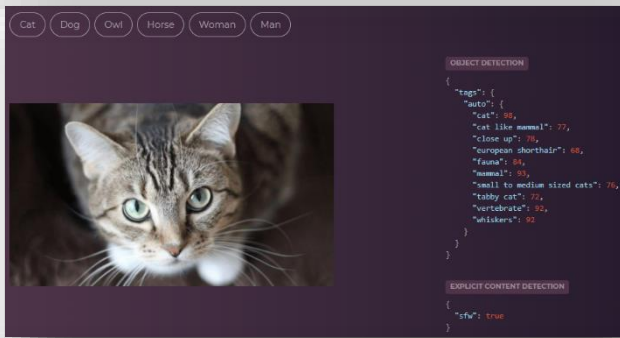
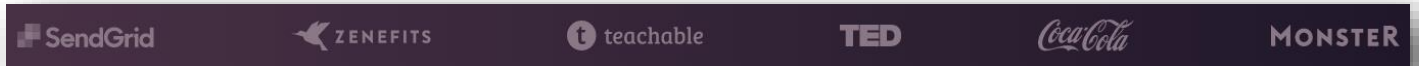
Nombre	Filestack			
País	USA			
Funding	\$1.800.000			
Fuentes	Video			
Website	filestack.com			


El core de los servicios de filestack están relacionados con la subida de contenidos y conversión entre diferentes formatos.

Esta solución ofrece entre sus servicios la generación de metadata a partir de contenidos visuales incluyendo la detección de contenidos explícitos.

Como dato interesante adicional, también proveen de servicios de creación de modelos de IA ad-hoc para el cliente.


Entre sus clientes más relevante está Coca Cola, TED y Monster.






OBJECT RECOGNITION

How valuable would it be to "see" what is in every image? With our image auto-tagging service, you can easily abstract what objects appear in your uploaded images. People, animals, transportation, and thousands of other categories are at your disposal.




EXPLICIT CONTENT DETECTION

Moderate your images to ensure you are only showing content that comply with your rules and boundaries. By implementing "safe for work" and a "not safe for work" score, our image tagging service can help you automate content moderation so you can focus on your core business.




OBJECT DETECTION

Sometimes knowing which objects are in your photos isn't enough - you need to know where they are, too. Filestack Object Detection uses state-of-the-art neural networks to detect and locate common objects in photos.




OCR (OPTICAL CHARACTER RECOGNITION)

OCR extracts text from images so you don't have to. Whether you are transcribing old documents, or grabbing license plate numbers from security footage, Filestack OCR reliably extracts printed and handwritten data from your images.






COPYRIGHT

Now detect images that are copyright protected. Upload a single image or millions of images, a single API call will display the copyright status of any image. This is especially useful for businesses that rely on user generated content and require their uploads to be free of copyrights.



VIDEO INTELLIGENCE

SEO can amplify how users search for and find your website. Give your site the boost it needs with automatic video tagging when you apply Filestack Video Intelligence. With our custom machine learning models, long form video tags are applied to your video uploads to optimize your site's performance with SEO front and center.

Nombre	Cinfo			
País	España			
Funding	\$7.100.000			
Fuentes	Video			
Website	cinfo.es			

cinfo es una empresa española especializada en servicios de video. Entre sus clientes audiovisuales se encuentran grandes plataformas como rtve, Televisa, Rakuten y Sony.

Tienen servicios que cubren un gran rango de necesidades distintas relacionadas con video: cloud para eventos multicámara, donde el usuario puede elegir la cámara que quiere ver (con gran foro en eventos deportivos y culturales); servicios para educación remota, servicios de videovigilancia, contenido inmersivo y hasta una plataforma OTT propia.

Entre sus servicios cuenta con la generación automática de metadata para contenidos audiovisuales.

OUR TECHNOLOGIES

cinfo is a company that specializes in video technologies and everything you need for an Internet TV to work properly. From the generation and administration of metadata, to the complete management of video, interfaces and user apps, websites that respond to subscriptions, transport and distribution of video, encodings and encryption.

In the latest versions of our products we incorporate the concept of automatic television generated and governed by artificial intelligence. This opens up the possibility of generating high quality video and television for small sports, shops, brands, sponsors, concert halls and even for industry. To find out all these possibilities just [contact us](#).





cinfo is a company co-financed by Xunta de Galicia according to the Resolution of "Axencia Galega de Innovación" (Principia Program) of June 8, 2017. Also, we have staff co-funded by the Xunta de Galicia according to Resolution of "Galician Innovation Invention" (Senior Talent Program) of September 19, 2018.

In 2017 we were beneficiaries of a grant financed by the Ministry of Economy, Industry and Competitiveness, whose purpose is to support aies to improve the competitiveness of SMEs. Our project, Collaborative Artificial Intelligence for Connected Industry 4.0, aims to explore the common and specific needs for the development of new IA algorithms for mass data processing in industrial environments, especially in the manufacturing industry.

cinfo is a company co-financed by Xunta de Galicia according to the Resolution of "Axencia Galega de Innovación" (Principia Program) of June 8, 2017. Also, we have staff co-funded by the Xunta de Galicia according to Resolution of "Galician Innovation Invention" (Senior Talent Program) of September 19, 2018.

In 2017 we were beneficiaries of a grant financed by the Ministry of Economy, Industry and Competitiveness, whose purpose is to support aies to improve the competitiveness of SMEs. Our project, Collaborative Artificial Intelligence for Connected Industry 4.0, aims to explore the common and specific needs for the development of new IA algorithms for mass data processing in industrial environments, especially in the manufacturing industry.




Nombre	Appen			
País	Australia			
Funding	\$ 587.450.000			
Fuentes	Video	Texto	Audio	
Website	appen.com			

appen es una compañía australiana especializada en la creación y licenciamiento de datasets para el entrenamiento de IA. Disponen de datasets para el entrenamiento de algoritmos de generación de metadata y moderación de contenidos tanto video, audio como texto. Estos datasets entrenado son claves para que sus clientes puedan escalar sus modelos mucho más rápidamente.

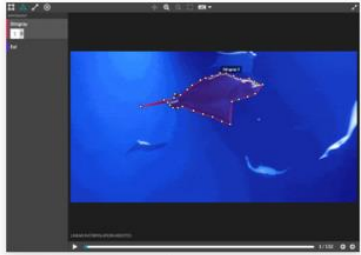
Es una compañía plenamente asentada en el mercado. Entre sus principales clientes están Google, Amazon, Microsoft y Adobe, entre otros. appen cotiza en bolsa, la cifra de funding es una estimación de su valor bursátil a fecha de cierre del presente estudio.

Trusted by global leaders to power their mission-critical AI with data for over 25 years



Video

Collect, classify, transcribe or annotate videos to assist your models to see and interpret the world around them. Our annotation tooling includes special transcription and time stamping tools, object tracking (with additional Speed Labeling capabilities), object detection and time stamping as well as ontology attribute annotation. Our Machine Learning assistance and custom built tools give you the flexibility to easily collect accurate video annotations at scale.




Our video object tracking with Speed Labeling eliminates the need to annotate every single frame and expedites the annotation process.

Text

- Natural language processing (NLP):** To understand human language, computers rely on **NLP**. They may use techniques like keyword filtering to identify unfavorable language for removal.
- Sentiment analysis:** Context matters on the internet and sentiment analysis helps computers identify tones, such as sarcasm or anger.
- Knowledge bases:** Computers can rely on databases of known information to make predictions on which articles are likely fake news or identify common scams.

Audio

Collect, classify, transcribe or annotate audio data for your NLP projects. Our Audio Annotation tool is twice as fast as traditional annotation tools. Segment audio into layers, speakers and timestamps for your Audio Speech Recognition and other audio models. Generate high-quality audio transcripts rapidly with acoustic tags in a variety of languages, leveraging NLP to improve transcription quality and efficiency. Our all-in-one audio tooling has been purpose built to deliver clear and crisp audio annotations and transcriptions to train your models.



Our audio annotation tool enables you to automatically segment audio files, and time stamp and transcribe quickly efficiently, allowing for accurate annotations at scale.

Image and Video

- Object detection:** Image analysis can identify target objects, such as nudity, in images and videos that don't meet platform standards.
- Scene understanding:** Computers are learning to understand the context of what's happening in a scene, driving more accurate decision-making.

How Does Content Moderation Work?


The content queues and escalation rules for ML-based review systems will vary by company but generally will include AI moderation at either step one, step two, or both:

- Pre-moderation.** AI moderates user content before posting. Content categorized as not harmful is then made visible to users. Content deemed to have a high probability of being harmful or **not business-friendly is removed**, if the AI model has low confidence in its predictions, it will flag the content for human review.
- Post-moderation.** Users report harmful content, which AI or a human then reviews. If the AI does the review, it will follow the same workflow described in step one, automatically deleting any content determined to be harmful.



Depending on the type of media, AI uses a variety of ML techniques to make content predictions.

Text

Collect, classify and annotate text to enhance your NLP model's understanding of nuanced human speech. Speed Labeling capabilities include built-in multi-language tokenizers to assist human annotation efforts. Target entity extraction and span labeling with options to bring your model outputs to accelerate contributor annotations. We can also help with text evaluation and post-editing data generated from your NLP models. Our specialized tools and linguistic experts will deliver the high-quality training data you need to build your NLP model for any chosen market.



Our purpose built text annotation tools make it easy to annotate text in detail, allowing your models to be trained to understand text and gain valuable insights.

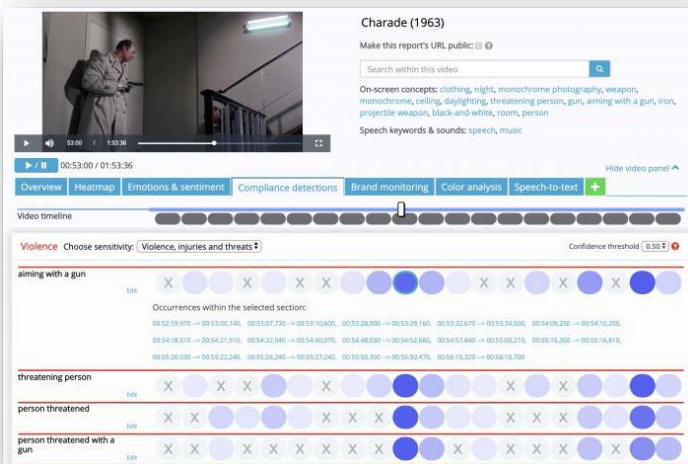
Nombre	Valossa			
País	Finlandia			
Funding	\$2.700.000			
Fuentes	Video	Texto	Audio	
Website	valossa.com			

Valossa es una solución completa de etiquetado automático de contenidos audiovisuales y moderación. Dispone de taxonomía propia para identificar contenidos sensibles con algoritmos preentrenados que ofrecen una alta precisión.

También ofrecen servicios de creación de modelos adaptados al cliente partiendo de sus datasets entrenados, para incrementar la velocidad de escalado de estas soluciones.

En la generación de metadatos incorporan un mapeado con la taxonomía IAB, un consorcio internacional para la publicidad en medios digitales que trata de estandarizar conceptos para facilitar el escalado de soluciones de publicidad digital.



Actualmente ofrecen servicios a plataformas OTT de creación de previews y recomendaciones automáticas.



Industry leading taxonomy for sensitive content.

Valossa AI sees and hears what's inside videos to detect explicit content, sensuality, violence, injuries and threats, accidents and disasters, substance use, guns and weapons. It also understands spoken bad language, as well as sensitive content topics like social issues, crime, death, hateful language and more.

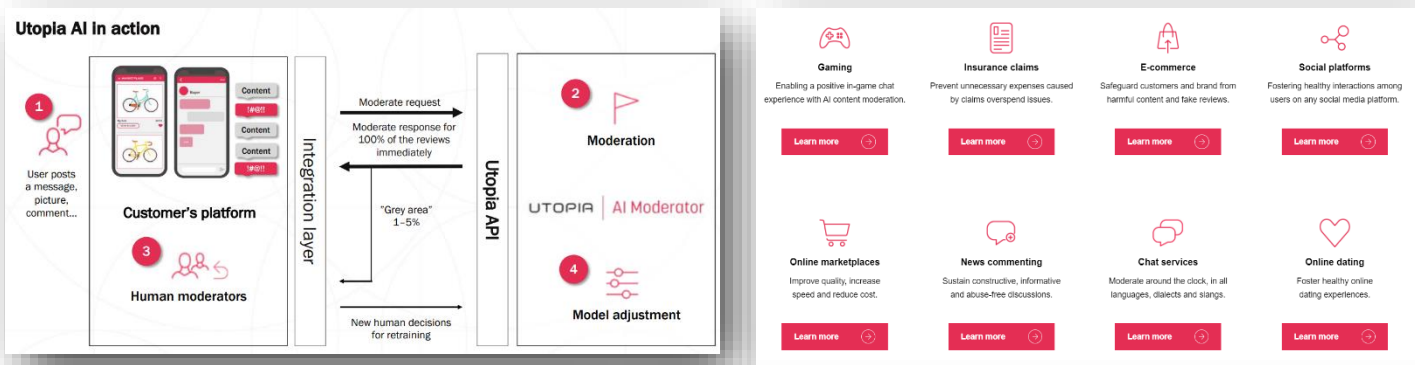
[Learn more](#)

Nombre	Utopía Analytics		
País	Finlandia		
Funding	\$1.200.000		
Fuentes		Texto	
Website	utopiaanalytics.com		

Utopia, es una herramienta basada en Inteligencia Artificial (AI) que automatiza la moderación de contenido (texto e imagen no estructurados), en cualquier idioma. Su enfoque no se basa en utilizar listas de palabras o procesamiento de lenguaje natural - NLP, sino que sus modelos aprenden a imitar las decisiones humanas, según las políticas del cliente y las decisiones anteriores del moderador.

La compañía nace de la creación de esta tecnología llamada “Algoritmos de lenguaje independiente para análisis de datos” por parte de su fundadora y CEO Mari-Sanna Paukkeri.

En su información comercial declaran los siguientes resultados en base a la experiencia con sus clientes: Integración en 2 semanas; reducción de tiempos de manipulación: 90% - 99%; ahorro de costes: 20% -50%; aumento de la eficiencia \geq 30%; reducción de moderación manual \geq 80%.




Utopia AI Moderator

Utopia AI Moderator is a fully automated moderation tool that protects your online community and your brand from abusive user-generated content, fraud, cyberbullies and spam. It learns from the publishing decisions your human moderators made previously, works in real time, with a higher degree of accuracy when compared to humans.

Utopia AI Moderator understands the context, works in any language, and is especially good in informal language, slang or dialect. Our tool increases the quality of published content. It removes publishing delays and provides reliable, consistent curation around the clock, allowing human moderators to focus on moderation policy management and only the most difficult cases.

Utopia AI Moderator is ready for production use in only two weeks, and it moderates 100% of your incoming content. It stays up-to-date by learning as it works.



Utopia AI Moderator solution is listed by the **kidSAFE Seal Program**. Click on the seal to learn more.

Tailor-made off-the-shelf tool

To implement Utopia AI Moderator takes only 2 weeks. Utopia AI Moderator is a SaaS solution and your IT team will be able to integrate your database to the Utopia API in few days.

Your rules, our tools

Utopia AI Moderator doesn't define your moderation policy. You do. It learns from your human moderation decisions to meet your specific requirements. If you want to change your moderation policy, that's also very easy: Just few clicks and Utopia AI learns automatically.

Functional features

Utopia AI Moderator can define the reason why a particular content was flagged. It can also handle content in other languages, process links, email addresses, phone numbers, passwords, usernames, identify various contents occurring in spam messages, duplicates, near duplicates, atypical content, and all kinds of meta information, user behaviour, bans, robots and emotions.

Social media language



Processing informal language is natural for Utopia AI Moderator. It understands the semantic meaning regardless of spelling errors, informal grammar use or unusual meanings of words. It is the only true language-agnostic moderation tool in the world, and was originally developed in one of the most complex language in the world, Finnish. As a result it can easily be deployed into any language or dialect.

Image and video moderation

As with text, so can images and videos also be moderated automatically according to your own publishing policy. Adding an image and video moderation feature increases the overall quality of your user-generated content.

Utopia AI

The unique text analytics AI technology behind Utopia AI Moderator is developed by our top scientists. It doesn't require Natural Language Processing (NLP) components. Utopia AI creates the big picture, and because the approach is unbiased, it can be used as a basis for open decisions.

Nombre	Sensifai			
País	Bélgica			
Funding	\$1.500.000			
Fuentes	Video	Texto	Audio	
Website	sensifai.com			

La tecnología de sensifai está especializada en reconocimiento y etiquetado de video, audio y texto.

Estas capacidades de identificación de elementos (objetos, personas, paisajes, logos, escenas, atributos, etc) las aplica paquetizándolas en diferentes servicios que van desde el reconocimiento de desnudez y la identificación de contenidos de abuso de menores hasta aplicada a servicios de videovigilancia para reacción temprana ante determinados hechos.

También aplica la tecnología implementando un sistema de detección de colisiones automático para la industria Auto.

Content Compliance	Video Metadata Recognition	Speech Recognition
Find and tag nudity, violence, smoking, drinking, and swearwords in videos and images with Sensifai.	Recognize and tag thousands of concepts, attributes objects, actions, and sports in your videos and images.	Add automatic subtitling to your video or audio files and Make them searchable.



NSFW Recognition and Content Compliance

Sensifai monitors all user-generated images/videos on your website in realtime and removes content with nudity, gore, substance abuse and swear words by its not-safe-for-work (NSFW) image recognition, video AI, and speech recognition system. Sensifai offers the world's most accurate video recognition for content compliance to moderate inappropriate scenes and checks all your multimedia content planned for broadcasting to make sure they comply with regulations.

Due to the ever-increasing volume of user-generated video, image, and speech in social media and video sharing websites, it's a must to increase the content compliance efforts to be sure those contents are adapted for a specific occasion or for a group of viewers. Sensifai NSFW recognition system yields a world-class service that can recognize a wide variety of NSFW concepts, such as sexual visual, nudity, violence, and substance abuse and keeps your contents safe by inspecting them and alarming in the case of inappropriate pieces whether visuals or sounds.

Solution

Sensifai video AI offers a solution that screens video streams in a speed of 1X to 100X faster than real-time and has the ability to recognize what's inside an image, audio or video file and then summarize its improper content. In particular, Sensifai NSFW recognition system offers the following features:

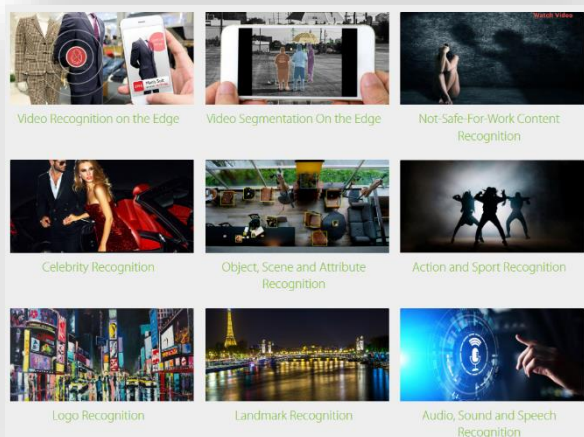
- ◆ Nudity recognition
- ◆ Gore recognition
- ◆ Drug recognition
- ◆ Drinking recognition
- ◆ Swear words recognition (bad language detection).



This technology generates time-based metadata that allows compliance teams to work with episodes instead of still frames using its generic speech recognition, video AI, image recognition systems.

WORLD-CLASS Performance

Benchmarking on different standard databases shows that Sensifai NSFW recognition system has the highest recognition accuracy compared to other existing solutions in the market and performs much faster than them.

We have collaborated with a variety of media companies during the development of the tool to satisfy their business needs. Our tool's unique value proposition is the ability to interpret sounds, speech, and visual content simultaneously and interactively for monitoring their image, speech and video content.



Nombre	Smart Moderation		
País	USA		
Funding	\$239.500		
Fuentes		Texto	
Website	smartmoderation.ai		

Smart Moderation se centra en la moderación de comentarios en Social Media. Como parte de esta familia de soluciones, realiza el etiquetado de contenidos de texto en tiempo real para no degradar la experiencia del usuario.






Ofrecen servicio a agencias y a marcas para proteger su presencia en redes de comentarios ilícitos, inapropiados o spam.






Se basa en tecnologías de Natural Language Processing (NLP) y Machine Learning con modelos preentrenados para el reconocimiento de comentarios nocivos.



Best Comment Moderation App For Social Media

We deliver complete protection for your Social Media and Websites through detection of illegal and inappropriate comments, spam, trolls and erasing them in real time.

BASIC	PRO	ENTERPRISE
Yearly Price \$79 /mo	Yearly Price \$199 /mo	Yearly & Monthly Price Custom
Monthly Price \$99	Monthly Price \$249	
GET STARTED	GET STARTED	CONTACT US

				
CONNECT social profiles with a single click. No complex integration required!	MONITORING all comments in real time. No reason to worry about missing a comment!	FILTERING illegal comments, profanity and abusive language, spam, trolls, etc. No profanity blacklist needed!	AUTO-HIDE inappropriate comments within the minute that they are posted. Get complete online safety!	PERSONALIZE by autonomous learning according to your moderating needs and preferences.

 Reduce Cost Save up to 80% of Time and Money	 Increase Accuracy Increase moderation success rate up to 96%	 Protect Reputation Keep members safe from the social media crisis 24/7/365
 Increase Happiness Make community managers happy by doing dirtiest footwork	 Make Money Improve online conversations to affect purchase decisions	

Nombre	DragonflAI			
País	UK			
Funding	\$114.700			
Fuentes	Video	Texto	Audio	
Website	dragonflai.co			

DragonflAI es una solución de moderación de contenidos generados por el usuario directamente en dispositivos para plataformas sociales. Su foco principal está en la detección de desnudez.

El algoritmo de detección se integra en la app de la plataforma cliente y permite realizar la moderación en cada subida de contenido por parte del usuario. Esto hace que la capacidad computacional del análisis de contenidos quede distribuida por los dispositivos de los usuarios de la plataforma en lugar de ser ejecutada vía servidor, evitando innecesarias recepciones de contenido no permitido.

Plug & play Serverless Moderation

Moderate images and videos, completely on-device before they can be uploaded to your platform

- ✓ Eliminate server costs & liability
- ✓ Up & running in an afternoon

Moderation Designed for Growing Platforms

A library dedicated to reducing platform liability and the cost of moderation.

01 / Install

With simple documentation and examples, get moderation for your app running in minutes.

02 / Push

Update the app to begin moderating offline instantly, without any backend infrastructure required.

03 / Grow

Focus on growth and community management rather than retroactively moderating content.

<18 mb

Algorithm Size

Lightweight algorithm designed to minimise the size of your app.

97.9 %


Nudity Detection F1 Score

Tested on a wide variety of images, DragonflAI will continue to improve.

~60 ms


Analysis Time

Running on Samsung Galaxy S9+ installed within the DragonflAI app.




Installed in Minutes

With simple documentation and examples, add moderation to your app in minutes




Instant Moderation

Results in less than 100ms based on your community guidelines




Reduce Liability

Prevent indecent content, including CSAM, from being uploaded




Forget Server Uptime

Within-app moderation eliminates server costs



Infinitely Scalable

Each new user brings their own computing power, their phones



Unlimited User Uploads

Upload volume has no impact on moderation costs, it is all completed on device

Nombre	Analytics			
País	India 			
Funding	No publicado			
Fuentes	Video	Texto	Audio	
Website	analytics.ai			

Analytics es una compañía que ofrece servicios de etiquetado de datos para generar datasets de entrenamiento para modelos de IA.



Esta compañía es interesante porque ilustra una parte importante de la cadena de valor de la puesta en marcha de cualquier solución tecnológica basada en Machine Learning, como es la generación de los datasets de entrenamiento para entrenar a los algoritmos de IA.

No es casualidad que esta compañía esté basada en India, ya que ofrece una solución low-cost para realizar esta intensa labor de generación de los datos de entrenamiento. Cubren un gran rango de industrias y casos de uso.



Entre sus clientes, destacan Unilever y Caltech.

Helping 100+ Companies Achieve AI Excellence

















QUALITY WITH ACCURACY  
Get the best-in-class quality services with highest accuracy level delivering an excellence in image annotation through multiple stages of auditing and reviewing of labeled data.



FULLY SCALABLE SERVICE
Working with hundreds of workforce to annotate pictures as per the demand providing a completely scalable solution with turnaround time to meet the different client needs.

SECURITY WITH PRIVACY  
We are certified with SOC 2 TYPE 1 Company for maintaining the high standards of data security with privacy while working with our clients to ensure their confidentiality.

COST-EFFECTIVE PRICING
Image annotation outsourcing to us means our clients get a cost-effective data labeling service helping them to minimize the cost of their project with best efficiency.

 Self Driving Used for computerized vision learning for autonomous vehicle driving	 Health Care The diseases diagnosis and analysis through annotated medical imaging	 AI in Retail Used for commerce and retail supermarkets through vision based inspection	 Autonomous Flying Mainly used for aerial view capturing and object recognition through drones
 Robotics Used as robotic arms for material handling at warehouse or assembly plants	 Security Mainly used to keep monitoring the objects through automated security cameras	 Satellite Imagery Used for satellite images to get the view for urban management and smart cities	 Agriculture Used as robotics in agriculture for soil monitoring, crop harvesting or yield analysis

 Polygon Annotation Drawing precisely with contours in polygon shapes over the products to detect the objects and localization of videos or images.	 Polylines Annotation Applied into lane detection on roads for autonomous cars and other vehicles to detect the lane and move into the right direction.
 Semantic Segmentation It is the process of image annotation by partitioning the images into multiple segments for autonomous vehicles and healthcare.	 Bounding Box A popular method of classifying objects in images or videos for retail, robotics, autonomous vehicles, Machine learning needs.
 Landmark Annotation Plotting a sequence of points correctly to determine the shape of natural objects like facial features, emotions and expressions	 3D Cuboid Annotation Capture the object in 3D cuboid to get the more in-depth dimension view and build a more comprehensive model for detailed info.


Nombre	WebPurify			
País	USA			
Funding	No publicado			
Fuentes	Video	Texto	Audio	
Website	webpurify.com			

WebPurify ofrece una solución de moderación de contenido en formato video que combina la tecnología de IA que ellos llaman Automated Intelligent Moderator (AMI) con la moderación manual para dar respuesta a los casos dudosos y refinar el modelo acorde a las necesidades del cliente.

Su tecnología de IA está preentrenada para identificar contenido sensible, entre el que se encuentran las siguientes categorías: desnudez, drogas, género, odio, gestos ofensivos, menores, armas y alcohol.

AI + Live Video Moderation.

To deliver the best results, we offer a hybrid solution that leverages our AI technology to immediately reject obvious violations and flag potential issues for escalation to live moderators for further investigation.



Fast & Accurate

Automated results are returned within minutes, while videos reviewed by our live teams are returned within an hour or less, 24x7, 365 days a year. (For even faster results, we recommend a custom solution.)

Highly Scalable













Between our live moderation teams and our AI technology, there's no limit to the amount of video you can send to WebPurify. Our solutions easily scale to meet any capacity.

Easy to Integrate

Our video moderation technology is easy to implement for anyone with experience using HTTP web services. All it takes is a few lines of code to be up and running in no time.

AI-based video moderation.

Our Automated Intelligent Moderation (AIM) API can detect, score, and remove unwanted video content within minutes across these 12 integrated categories.

-  Nudity
-  Minors
-  Drugs
-  Weapons
-  Gender
-  Blurriness
-  Hate
-  Alcohol
-  Offensive Gestures
-  Embedded Text
-  Celebrities
-  Sunglasses

Live video moderation.



Let our expert team review your video and audio to capture nuances—such as sexually suggestive behavior vs. nudity, or violence and hate vs. weapons—in as little as an hour. They can also reject broken videos and flag languages other than English. Pricing starts at just 15¢ per minute of video content moderated.

- ⊗ Nudity or partial nudity
- ⊗ Sexually suggestive behavior
- ⊗ Hate or hate crimes
- ⊗ Violence
- ⊗ Offensive gestures or language (English Only)
- ⊗ Drugs, drug paraphernalia, or drug use
- ⊗ Blank or broken videos

Other Language Label: If a video complies with our moderation criteria but contains languages other than English, we will label it as "Other Language."



CASE STUDY

Protecting Pringles' User-Generated #TasteTracks Videos

When Pringles launched its Mr P chatbot, fans were invited to create and share their very own #TasteTracks music videos via Facebook Messenger for a chance to win a nifty vacation.

These user-generated videos were then automatically assembled into a tasty mashup... but not before thousands of them were instantly moderated by WebPurify to make sure nothing too dodgy or inflammatory got through.

Nombre	Fasso			
País	Alemania			
Funding	No Publicado			
Fuentes	Video			
Website	fassoo.com			

Fasso es otra solución de etiquetado automático de contenidos de video con algoritmos preentrenados capaces de identificar contenido sensible.

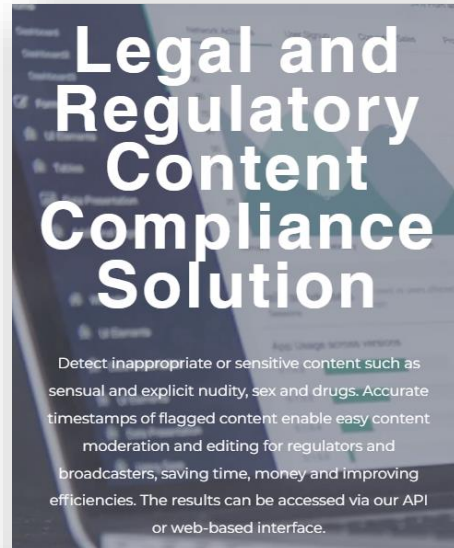
Este es un ejemplo interesante de solución desarrollada en el ámbito europeo, co-financiada por la Unión Europea que tiene como uno de sus cuatro servicios específicos relacionados con el etiquetado la detección de contenido sensible bajo el nombre de “Legal and Regulatory Content Compliance Solution”.

Dispone de modelos preentrenados con los que es capaz de reconocer no sólo elementos relacionados con contenido sensible, sino también elementos relacionados con eventos deportivos y monitorización de marcas.



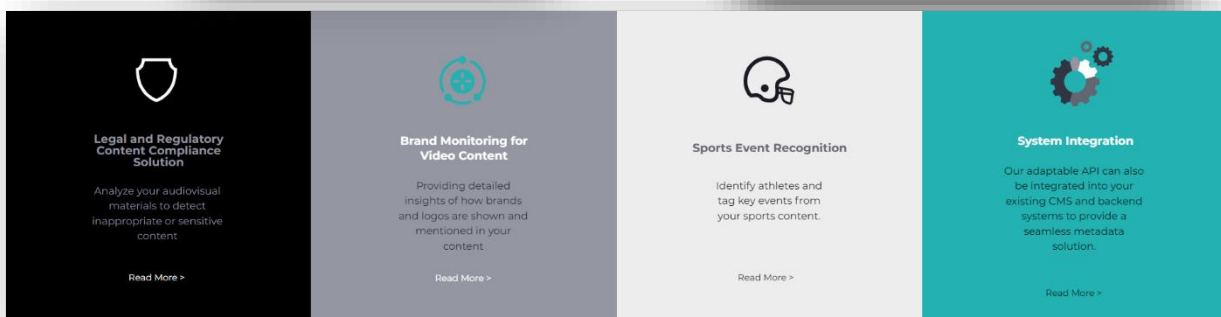
What can Fasso do for you?

- Brand Monitoring**
Recognize thousands of brands and logos to provide automated logo detection and visibility analysis that provides a rich set of statistics.
- Sports Events Recognition**
Recognize the faces, actions, events and context that are most relevant for you, especially when read in conjunction with sponsors and brands.
- System Integration**
Our adaptable API can also be integrated into your existing CMS and backend systems to provide a seamless metadata solution.
- Sensitive content filtering**
Find problematic scenes and decide quickly how to deal with them. Search by type or for a target audience based on rating or legal jurisdiction.

Legal and Regulatory Content Compliance Solution

Detect inappropriate or sensitive content such as sensual and explicit nudity, sex and drugs. Accurate timestamps of flagged content enable easy content moderation and editing for regulators and broadcasters, saving time, money and improving efficiencies. The results can be accessed via our API or web-based interface.



- Legal and Regulatory Content Compliance Solution**
Analyze your audiovisual materials to detect inappropriate or sensitive content.
[Read More >](#)
- Brand Monitoring for Video Content**
Providing detailed insights of how brands and logos are shown and mentioned in your content.
[Read More >](#)
- Sports Event Recognition**
Identify athletes and tag key events from your sports content.
[Read More >](#)
- System Integration**
Our adaptable API can also be integrated into your existing CMS and backend systems to provide a seamless metadata solution.
[Read More >](#)

2. Análisis de tecnologías potenciales

2.1. Análisis general del stack tecnológico

Para facilitar la comprensión de los distintos elementos tecnológicos relevantes en el ámbito del etiquetado automático de contenidos audiovisuales vamos a ir construyendo de forma incremental el stack tecnológico arquetípico y las diferentes alternativas de cada uno de sus elementos.

Una de las conclusiones más relevantes de este estudio es que, en cuanto al esquema tecnológico utilizado, no hay grandes sorpresas respecto a las premisas de las que se partía inicialmente. En general, en su parte más básica, los sistemas de etiquetado se basan en algoritmos de Inteligencia Artificial que tienen fundamentalmente dos componentes:

1. **Algoritmos de procesamiento de las fuentes de información:** que hace capaz al sistema de leer la información que se le inyecta en función de su formato, bien sea texto, audio o video.
2. **Modelo de aprendizaje automático:** en el que el sistema se apoya para tomar decisiones respecto a la información que está procesando, en base a los elementos que está entrenado para identificar.

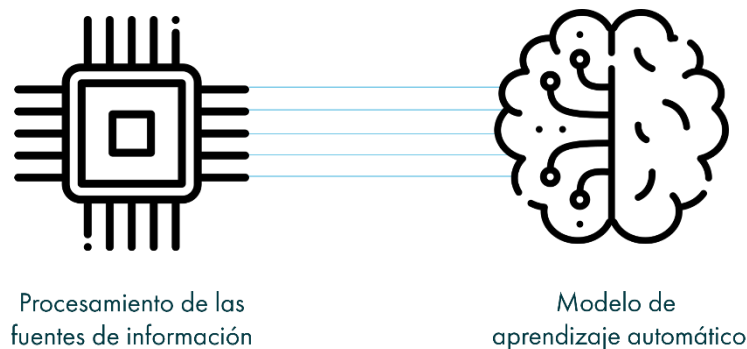


Figura 1: Sistema elemental · Etiquetado

La salida arquetípica de este tipo de sistemas es un conjunto de tags asociado a un modelo probabilístico que indica el grado de confianza con el que el algoritmo le asigna una etiqueta determinada, como muestra la siguiente ilustración:



Figura 2: Modelo probabilístico. Fuente: Valossa

Y esto nos lleva a identificar explícitamente otro de los elementos clave del sistema:

- **Taxonomía:** Este elemento no es puramente tecnológico, pero tiene implicaciones enormes a todos los niveles porque impacta en la necesidad de precisión. Se trata de la concreción de los elementos que se desean identificar en cada pieza audiovisual. Dicho de otro modo, las etiquetas que el sistema ha de poder asignar de forma adecuada con el objetivo de identificar todos los elementos claves para poder, en última instancia, calificar la pieza y enriquecer la información que se ofrece al usuario sobre ella. En la taxonomía usada podemos distinguir dos capas:
 - **Descriptiva:** En esta capa lo relevante es añadir las etiquetas que describen las escenas, sin ningún tipo de juicio de valor a nivel de calificación.
 - **Prescriptiva:** Se trata del mapeado de etiquetas clave que implican una calificación de edad determinada (típicamente relacionadas con elementos potencialmente dañinos para el público infantil, como es la violencia extrema, el sexo explícito, abusos verbales, etc.)

Para que este sistema logre una precisión aceptable, es necesario incluir en el esquema dos elementos adicionales cuya ponderación y relevancia en el esquema tecnológico es inversamente proporcional, como explicamos a continuación:

- **Datasets de entrenamiento:** Se trata del conjunto de base de datos de piezas audiovisuales ya etiquetadas correctamente a partir de las cuáles, el sistema va a aprender a reconocer los diferentes elementos y etiquetarlos adecuadamente. Cuanto mayor es el conjunto de datos de partida, mejor es la precisión del modelo de forma automática. Típicamente al inicio de la implementación de un modelo de estas características, no se dispone del volumen suficiente para que el modelo sea razonablemente autónomo. Por esta razón, es indispensable añadir el siguiente elemento al esquema:

- **Refinamiento humano:** Esta es una capa de supervisión humana de las decisiones que el algoritmo va tomando. Como decíamos, es inversamente proporcional al dataset de entrenamiento en el sentido de que cuanto mayor es la base de datos de aprendizaje, más precisión tiene la IA y es necesaria menor implicación del humano. Por tanto, dependiendo de la madurez del modelo, la implicación humana tiene un rol diferente:
 - a. **Entrenamiento:** donde el humano etiqueta de inicio cada pieza audiovisual para construir el dataset inicial que se le pasará al algoritmo para comenzar a operar.

 - b. **Corrección y enriquecimiento:** La capacidad de detección automática depende de la tecnología. Existen elementos que son muy complicados de identificar, como puede ser el lenguaje irónico. Para cubrir este gap y extender la potencia de la labor del etiquetado, el humano en esta etapa tiene una misión de refinar el modelo para aumentar su precisión.

 - c. **Supervisión:** Cuando el modelo es razonablemente autónomo, la participación humana se puede limitar a tareas de supervisión en los casos en los que el algoritmo no pueda ofrecer un grado de confianza suficiente para alguno de los elementos claves.

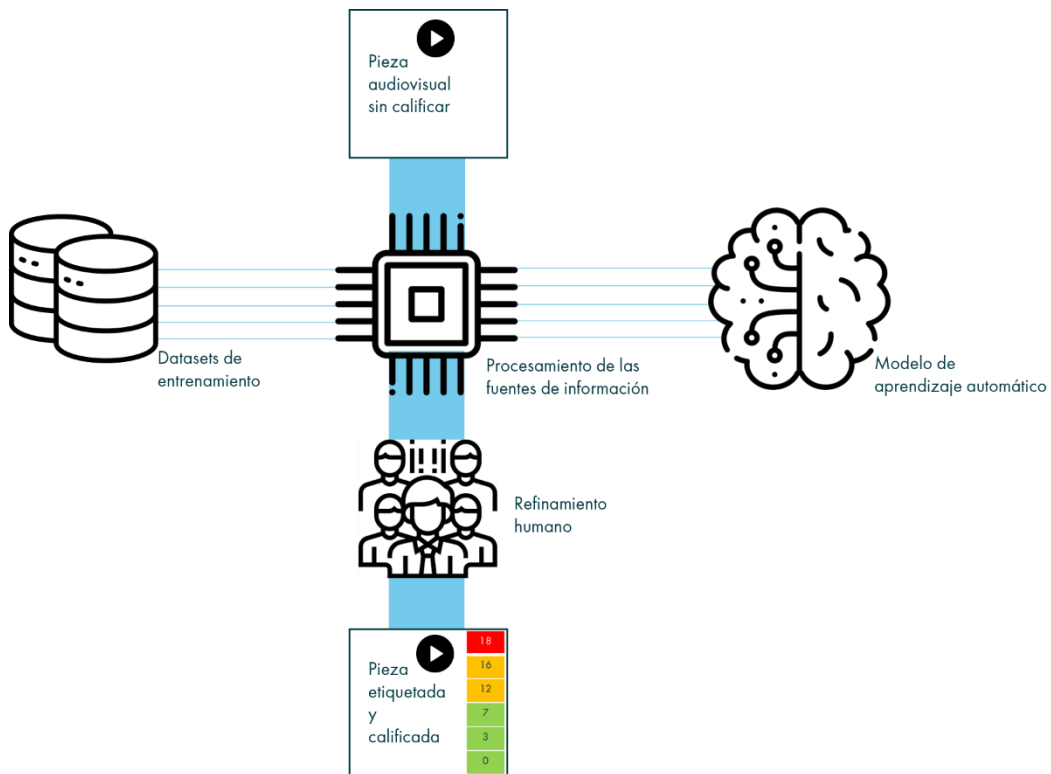


Figura 3: Stack tecnológico básico completo · Etiquetado

2.1.1. Capa de etiquetado según tipo de fuente

Los productos audiovisuales cuentan fundamentalmente con tres tipos de fuentes de información: texto, audio y video. Cada uno de ellos aporta una parte del análisis completo. Si bien con el análisis de las transcripciones de un producto audiovisual se puede alcanzar un nivel de calificación razonable, existen otros elementos, especialmente visuales que están fuera de este análisis y que podrían no tener un reflejo en la transcripción (por ejemplo, una conversación para todos los públicos entre personas desnudas en un ambiente visualmente sexual).

Análisis de Texto

El análisis de texto en una pieza audiovisual se aplica fundamentalmente a la transcripción de los diálogos y algunos elementos adicionales potencialmente subtítulos. Generalmente cada pieza se entrega con su transcripción, por lo que de entrada ya se cuenta con esta fuente preparada para su análisis sin ningún paso previo. En el caso en que esta fuente de texto no estuviera disponible, es necesario aplicar tecnología de audio para la transcripción. Los enfoques tecnológicos más relevantes identificados son:

- **Natural Language Processing (NLP):** Los algoritmos NLP organizan la información recibida en elementos más pequeños y manejables y los analiza para entender la

relación entre los diferentes elementos. En su inmensa mayoría están basados en reglas que permiten taggear el contenido textual para extraer la información relevante. Existen herramientas tanto de código abierto como propietario para realizar este tipo de análisis a partir del texto (NLTK, Gensim, SpaCy). Estas soluciones típicamente mantienen una lista negra de conceptos que no están permitidos.

- **Language Independent Text Mining:** Estos modelos no usan un enfoque basado en listas de palabras, sino que aprenden a imitar las decisiones humanas por contexto mediante la ingestión de frases completas taggeadas. En base a este dataset inicial, entienden el contexto semántico de una oración completa, en lugar de verificar palabras independientemente, reglas definidas o filtros generales. Estos modelos infieren el lenguaje, las políticas y las reglas de las decisiones humanas de moderación tomadas anteriormente (dataset inicial).

Análisis de Video

El análisis de video está dominado por algoritmos de visión computerizada (Computer Vision). Su misión es identificar elementos para cuyo reconocimiento han sido entrenados. Van desde objetos (coche, casa, pistola,...), seres vivos (animales, plantas, personas...), personas concretas (reconocimiento facial) hasta actitudes (persona apuntando con un arma, sexo entre adultos,...). Se puede entrenar a los algoritmos para que reconozcan los elementos clave que ayudan a tomar las decisiones oportunas en base a lo que ocurre en imagen.

Generalmente el análisis se lleva a cabo frame a frame. Algunas soluciones realizan un muestreo para reducir costes, por ejemplo, 1 frame por segundo.

Análisis de Audio

Si tenemos en cuenta que los diálogos se analizan a partir del texto, el análisis de audio no es una pieza absolutamente fundamental, pero sí que añade información de contexto. Concretamente son dos los ámbitos en los que el análisis de audio puede añadir valor:

- **Análisis emocional de la voz:** Este tipo de análisis ofrece un nivel adicional de información contenida no en el texto, sino en la forma en la que está expresado a través de la voz. Éste es un análisis complejo y en la mayoría de los casos no se realiza, ya que generalmente el estado emocional se infiere del texto y de la imagen de una forma muy precisa.
- **Sound Event Analysis:** Este nivel de análisis aporta información adicional a la comprensión de los contenidos derivadas de todos los eventos de audio que no son texto (por ejemplo: disparos, gemidos, gritos). De forma combinada con el análisis del resto de fuentes añade información de comprensión del contexto.

2.1.2. Capa de moderación / calificación

En la lista de soluciones identificadas se distinguen soluciones que ofrecen etiquetado automático y otras soluciones que abordan el proceso denominado “de moderación”.

El alcance de ambos enfoques está íntimamente relacionado con los dos niveles taxonómicos identificados previamente:

- **Nivel descriptivo:** existe una asociación entre los hechos detectados en la pieza y las etiquetas que los describen. En este nivel no existe ninguna asociación con rangos de edad. Está enfocado a describir de la forma más precisa una determinada pieza audiovisual. En ella se basan las soluciones
- **Nivel prescriptivo:** existe una asociación entre etiquetas descriptivas claves y los diferentes públicos objetivos para los que se considera aptas o no aptas. Este es el nivel del juicio de valor respecto a qué contenidos son adecuados o no para los determinados rangos de edad. Es un mapeo entre descriptores y rangos de edad que indica la política vigente en un momento determinado del tiempo.

En base a estos dos niveles, se identifican en el mercado fundamentalmente dos niveles de alcance de los servicios de etiquetado automático:

- **Etiquetado automático:** en virtud de este modelo, el proveedor recibe la pieza audiovisual, la analiza y la etiqueta, por lo que el output de este proceso es la metadata con todas las etiquetas de la pieza organizada en el tiempo asociada a un modelo probabilístico que indica la probabilidad en la que una etiqueta es cierta en base a lo que sus algoritmos ya saben y son capaces de reconocer. En otras palabras, ofrecen etiquetado descriptivo.
- **Moderación:** Las soluciones de moderación no sólo hacen el etiquetado automático, sino que aplican una capa de comprensión para emitir un juicio de valor sobre la pieza analizada, típicamente si es apta o no apta para un determinado nicho objetivo de usuarios atendiendo a las reglas prescriptivas para las que haya sido entrenado. En otras palabras, ofrecen no sólo etiquetado descriptivo, sino también el resultado aplicar las reglas del nivel prescriptivo. Son especialmente en entornos que operan en tiempo real con contenido generado por el usuario.

2.2. Análisis de modelos potenciales de soluciones tecnológicas

Con independencia del tipo de fuente, la taxonomía de etiquetado determina la complejidad del sistema orientado a etiquetar un contenido audiovisual. Como ya se ha explicado, está íntimamente relacionada con el esfuerzo de entrenamiento que es necesario para que la IA sea capaz de funcionar de forma autónoma con un coste en tiempo y recursos de computación que lo haga viable.

Hay que tener en cuenta que las tecnologías de reconocimiento de elementos y etiquetado automáticos existen, están en el mercado y son capaces de obtener buen rendimiento. Pero todo ello viene determinado por la sencillez o complejidad de matices que tenga la taxonomía de etiquetas que se quiera aplicar. No es lo mismo reconocer una persona con un arma en actitud amenazadora que identificar una connotación sexual en un chiste basado en ironía fina.

La precisión o ambigüedad de la taxonomía que se pretende aplicar determinará en gran medida el éxito o fracaso de la aplicación de cualquier tecnología de etiquetado automático.

Como se ha visto en la primera parte del presente estudio, las soluciones de etiquetado automático están compuesta por un stack tecnológico complejo con un número elevado de componentes distintos.

La primera conclusión obtenida del análisis de mercado es que para cada uno de estos múltiples elementos existe una gran variedad de opciones.

La forma en la que se mezclan y se implementa la solución final está dentro de un rasgo inmensamente grande de combinaciones de tecnologías.

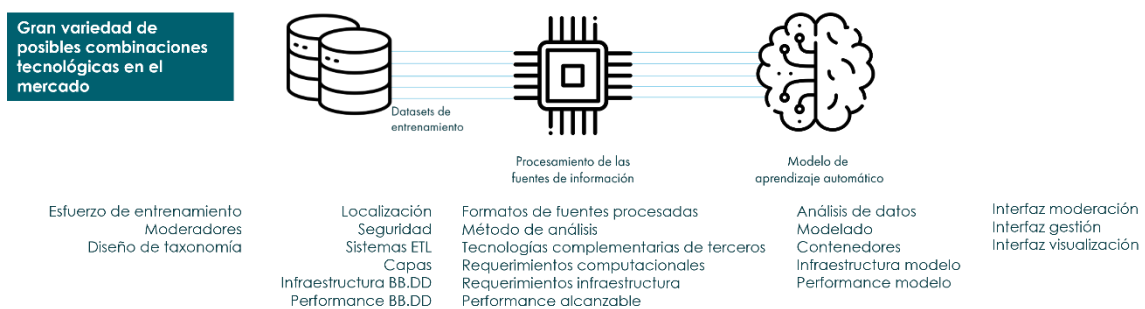


Figura 4: Elementos tecnológicos con gran variedad de posibilidades en el mercado

Ante esta inmensa diversidad de potenciales soluciones, para añadir el máximo valor posible al presente estudio, el análisis tecnológico se lleva a cabo contraponiendo dos modelos tecnológicos identificados en el mercado que determinan las implicaciones más relevantes en términos de costes y esfuerzo de entrenamiento. Representan dos extremos en cuanto al tipo de solución tecnológica que se puede aplicar.

2.2.1. Modelos preentrenados (taxonomía propia)

Este tipo de soluciones ofrecen un nivel de precisión al inicio muy alto debido a que, al usar una taxonomía propia, disponen en general de grandes cantidades de datos de entrenamiento previo y por tanto sus modelos están muy entrenados para el reconocimiento de elementos específicos.

Sus mayores ventajas son:

- **Precisión:** debido a lo afinado que están sus modelos de IA debido al intenso entrenamiento.
- **Reducido time to market:** el esfuerzo y el tiempo necesario para disponer de un algoritmo suficientemente entrenado en producción es mucho menor que con modelos que parten de cero.

Su mayor desventaja es que utiliza una taxonomía propia, por lo que, si la clave no está en el reconocimiento del elemento en cuestión sino en generar una etiqueta concreta, cualquier sistema deberá adaptarse o mapearse a su taxonomía para poder realizar el etiquetado final.

Costes asociados

Costes de setup

Estos algoritmos están ya preparados para operar desde prácticamente el minuto 0. Sus modelos están preentrenados para reconocer elementos específicos.

El principal foco de coste viene relacionado con el delta entre el modelo los elementos actualmente reconocibles por los modelos del proveedor y la taxonomía concreta que el cliente busque.

Dicho esto, se ha de indicar que la detección de contenido sensible es un problema que es abordado directamente y con amplitud por numerosos proveedores, por lo que el potencial delta estaría compuesto por el afinado de las etiquetas y la inclusión de matices que no estén actualmente soportados.

Costes de operación

Infraestructura: La mayoría de las soluciones se ofrecen vía API con una infraestructura Cloud. Algunas de ellas ofrecen una implantación *On Premises* en la infraestructura del cliente. En este caso hay que añadir a la estructura de costes una partida de mantenimiento de esta infraestructura.

Consumo: típicamente el pricing de estas soluciones se tarifica a partir de la cuantificación del flujo de información que se requiere procesar, bien sean elementos

conversacionales (frases), frames u horas de audio / video. La mayoría de las soluciones ofrecen descuentos a escala ofreciendo descuentos asegurando volumen.

Precisión y refinado humano: La presencia de moderadores humanos incrementa el coste tanto de setup como de operación de forma significativa. En este caso, dado que los modelos están preentrenados, esta necesidad y por tanto este coste es dramáticamente menor que en el modelo con taxonomía ad-hoc. Su participación se circunscribe al solventado de casos dudosos y al entrenamiento en la identificación de aquellos elementos que la tecnología no soportara originariamente.

Facilidad de uso

Estos servicios se consumen típicamente vía API. El análisis de la pieza audiovisual se puede llevar a cabo estando alojada en una ubicación remota, lo que incide positivamente en la seguridad relacionada con la preservación y salvaguarda de la propiedad intelectual de la pieza.

La mayoría de las plataformas en esta categoría ofrecen una interfaz para moderadores humanos de cara a facilitar la labor de refinado. En cualquiera de los casos vía API el cliente puede implementar o integrar a su propia plataforma de moderación.

El problema que puede surgir aquí es que no haya un matching claro entre el modelo provisto por el proveedor y la taxonomía que se quiere utilizar para el etiquetado. En este caso no se dispone de tanta flexibilidad como en un modelo con taxonomía ad-hoc, por lo que las posibilidades de adaptación, sin impacto significativo en los costes, es limitada.

Propiedad del modelo

La IP de los modelos preentrenados es del proveedor. Lo que típicamente se incluye en contratos donde el cliente requiere una adaptación / ampliación de la capacidad del modelo es que este nuevo modelo ad-hoc basado en el preentrenado sea de uso exclusivo del cliente.

Grado de consecución de los objetivos de SETELECO

Las soluciones con modelos preentrenados enfocadas específicamente a la detección de contenido sensible para la protección del menor cubren razonablemente los objetivos de SETELECO en términos de etiquetado automático de contenidos prácticamente desde el minuto 0, por lo que el tiempo para adquirir la capacidad de detección con calidad de producción es mínimo.

Obviamente, como todo sistema, se requiere una adaptación y un refinado para alcanzar cotas aún más altas, pero la capacidad actual de la tecnología para llevar a cabo esta labor y hacerlo de forma automática es muy alta con un esfuerzo de entrenamiento circunscrito a casos dudosos y a refinado.

Momento 0	1	2	3	4	5	6	7	8	9	10
Esfuerzo de entrenamiento · refinado										
Detección de contenido sensible · momento 0										
Automatización										
Tiempo necesario para calidad de producción										

TRL y obsolescencia

Las soluciones en esta categoría están en TRL9, están prestando servicio en el mercado cubriendo las necesidades específicas que se buscan.

No se prevé una obsolescencia cercana. Al contrario, los modelos tienen cada vez más precisión y lo que ya tiende a reducirse es la necesidad de moderación humana.

	1	2	3	4	5	6	7	8	9
TRL									

2.2.2. Modelos con taxonomía ad-hoc

Este tipo de soluciones están basadas en algoritmos de reconocimiento de elementos con una alta capacidad de aprendizaje de modo que son adaptables a cualquier taxonomía requerida definida por el cliente (siempre dentro de los límites que la tecnología ofrece en el momento actual).

Su mayor ventaja es la adaptabilidad y flexibilidad que ofrecen para crear una solución a medida.

Su mayor desventaja es que **requiere un considerable esfuerzo de entrenamiento**, y esto ha de ser tenido en cuenta en términos de coste en tiempo y recursos necesarios.

Costes asociados

Costes de setup

Esta fase es la que tiene mayor peso en los costes debido a la necesidad de estos modelos de ser entrenados de forma exhaustiva para lograr una precisión a nivel de producción.

Determinación de la taxonomía: Es vital conocer qué se quiere identificar en la pieza audiovisual y cómo se va a etiquetar. Hay que tener en cuenta un principio mencionado varias ocasiones durante el presente estudio, y es que la complejidad de los elementos que se quieran identificar y por ende de la taxonomía empleada está íntimamente relacionados con el esfuerzo de entrenamiento necesario para disponer de un algoritmo razonablemente autónomo en producción. Por tanto, esta taxonomía debería de ser la mínima imprescindible para cumplir con los objetivos perseguidos.

Generación de datasets iniciales: Toda cifra necesaria para alcanzar este hito es orientativa, pero identificamos cierto consenso en los 100.000 elementos analizados, entendiendo como elementos frases y frames. Los datasets pueden ser adquiridos o licenciados.

Moderación humana: La alternativa más razonable en este contexto es general el dataset inicial a partir de la labor de moderación humana.

Costes de operación

Infraestructura: La mayoría de las soluciones se ofrecen vía API con una infraestructura Cloud del proveedor. Algunas de ellas ofrecen una implantación *On Premises* en la infraestructura del cliente. En este caso hay que añadir a la estructura de costes una partida de mantenimiento de esta infraestructura.

Consumo: típicamente el pricing de estas soluciones se tarifica a partir de la cuantificación del flujo de información que se requiere procesar, bien sean elementos conversacionales (frases), frames u horas de audio / video. La mayoría de las soluciones ofrecen descuentos a escala asegurando volumen.

Precisión y refinamiento humano: La participación humana no sólo es clave en la elaboración de los datasets iniciales, sino también a la hora de afinar y refinar el modelo, resolviendo casos los casos dudosos y aportando de este modo feedback al modelo.

Facilidad de uso

El uso es muy similar al caso anterior, con la salvedad de que, en este caso, se dispone de un modelo específicamente adaptado al sistema de etiquetado que se requiere. Esta opción añade mayor flexibilidad y control sobre la metadata producida. Pero añade una necesaria participación humana que será intensiva hasta que el modelo adquiera suficiente precisión como para funcionar en producción de forma autónoma.

Propiedad del modelo

La IP de los modelos de IA en general son del proveedor salvo desarrollos desde 0. Como medida de protección secundaria caso de que la propiedad no sea transferida, es que este nuevo modelo ad-hoc sea de uso exclusivo del cliente.

Grado de consecución de los objetivos de SETELECO

El esfuerzo de entrenamiento de estas soluciones es muy alto. Se requiere de un tiempo y esfuerzo inicial de entrenamiento para que el modelo sea capaz de etiquetar con calidad de producción.

Momento 0

	1	2	3	4	5	6	7	8	9	10
Esfuerzo de entrenamiento										
Detección de contenido sensible · momento 0										
Automatización										
Tiempo necesario para calidad de producción										

Una vez el algoritmo está suficientemente entrenado, cubre los objetivos de SETELECO en este ámbito con la solvencia del caso anterior. La diferencia, como se indica previamente, radica en el esfuerzo y tiempo necesario para llegar a este mismo punto.

Algoritmo entrenado

	1	2	3	4	5	6	7	8	9	10
Esfuerzo de entrenamiento										
Detección de contenido sensible · producción										
Automatización										

TRL y obsolescencia

Como en el modelo anterior, las soluciones en esta categoría se enmarcan en TRL9, están prestando servicio en el mercado cubriendo las necesidades específicas que se buscan.

No se prevé una obsolescencia cercana. Al contrario, los modelos tienen cada vez más precisión y lo que ya tiende a reducirse es la necesidad de moderación humana.

	1	2	3	4	5	6	7	8	9
TRL									

3. Recomendaciones

Las recomendaciones del presente estudio se organizan en dos epígrafes. Cada uno de ellos está dedicado a describir un modelo alternativo para la creación del sistema de etiquetado común.

1. El primero de ellos es un modelo basado en la creación de un sistema de etiquetado automático común. En este ámbito se plasman recomendaciones a nivel tecnológico, organizativo y de instrumentos de contratación para llevarlos a cabo.
2. El segundo es un modelo basado en la certificación de proveedores de sistemas de etiquetado automático, que exploramos desde el punto de vista organizativo.

3.1. Modelo de desarrollo de una solución común auto-corregulada

Como resultado del análisis se aprecia que las tecnologías de etiquetado automático de contenidos audiovisuales ya existen en el mercado. Actualmente, como se ha visto, existen numerosas soluciones que están prestando servicios de etiquetado y moderación a empresas de todos los tamaños en prácticamente cualquier geografía e idioma.

Las premisas de las que se parten son:

- Prioridad de la detección del hecho sobre el uso de una taxonomía de etiquetas específica.
- Evitar la hiperdependencia de una determinada tecnología o proveedor que genere un riesgo de mantenimiento del sistema.

En la siguiente ilustración se representa el modelo explicado a continuación:

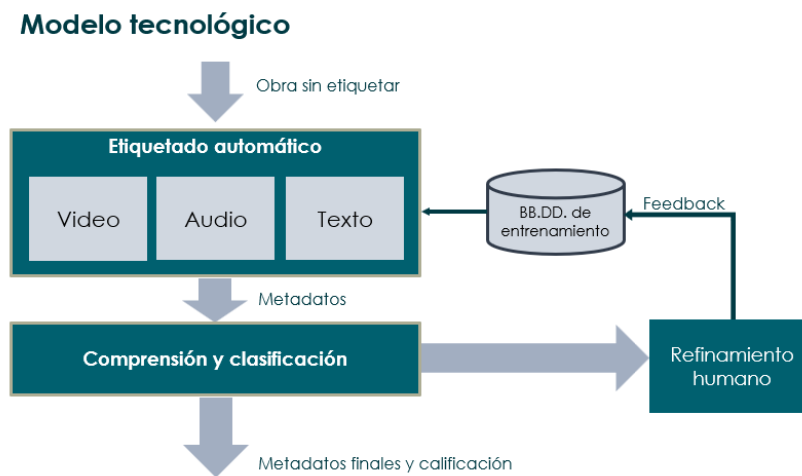


Figura 5: Modelo tecnológico de solución común de etiquetado automático de contenidos audiovisuales

3.1.1. Modelo tecnológico

El modelo tecnológico que se recomienda se compone de 2 capas:

CAPA 1 (Etiquetado)

La primera capa es la de análisis de fuentes de información y etiquetado automático. Típicamente se compondrá de un algoritmo de procesamiento de fuente por cada tipo de fuente (video, texto, audio) junto con un modelo de aprendizaje automático.

Para esta capa se recomienda modelos preentrenados que permitan tener una alta capacidad de reconocimiento desde el momento 0. Para ello hay soluciones que permiten enriquecer el modelo o incluso generar uno ad-hoc a partir de los modelos preentrenados. Estas opciones son aceleradores que facilitan enormemente la implantación de un sistema de estas características.

Es fundamental que el foco se centre en analizar soluciones con modelos que ya sean capaces de detectar los hechos claves que deriven en una calificación de edad concreta. Estos hechos generalmente son denominados por las soluciones como “Compliance detection” o “harmful content detection”. En otras palabras, se recomienda priorizar la capacidad de detectar estos hechos al uso de una taxonomía de etiquetas específica.

Acorde a análisis del stack tecnológico realizado previamente, en esta capa se distinguen tres flujos de análisis de información:

Análisis de texto

El análisis de texto es generalmente la primera etapa del análisis y la que típicamente tiene menor coste. Un buen análisis de texto debería identificar una gran parte de los elementos de riesgo.

Existen herramientas tanto de código abierto como propietario para realizar este tipo de análisis a partir del texto (NLTK, Gensim, SpaCy). Estas soluciones típicamente mantienen una lista negra de conceptos que no están permitidos.

Durante la exploración tecnológica se detectan algoritmos especialmente efectivos a la hora de interpretar el texto con su contexto que no dependen de listas de palabras. Un ejemplo ilustrativo es la solución propietaria de Utopía Analytics. Su tecnología viene respaldada por un reconocido trabajo de investigación europeo y el prestigio internacional de la doctora que lo lidera.

Análisis de video

El análisis del video es la segunda etapa del análisis de la obra audiovisual. Reconocer objetos visuales a través de algoritmos de visión computerizada tiene un coste de entrenamiento significativo. Entender el contexto, incluso disponer de los textos, ayuda a relacionar esos objetos con lo que está sucediendo en la escena.

Por esta razón, las soluciones de moderación cuentan en su mayoría con algoritmo intensamente pre-entrenados. Esto les confiere la capacidad de tener una alta precisión para los elementos automáticamente identificables desde el momento inicial de aplicación.

Proveedores como Active Fence o Hive Moderation que tienen taxonomía propia están abiertos a trabajar con taxonomía del cliente e incluso hacer la transición entre la identificación de elementos propia hacia la del cliente entrenando nuevos modelos en base a los ya existentes. Esta es una solución intermedia muy interesante de explorar para reducir los tiempos de puesta en marcha de la solución con calidad de producción.

Análisis de audio

Como se ha señalado previamente en el estudio, el análisis de audio es complementario. No es estrictamente indispensable para obtener unos metadatos precisos, pero sí añade una capa adicional de información de contexto de una escena.

Dado que los diálogos se procesan a través del análisis de texto, recomendamos que esta capa esté primariamente orientada a identificar elementos sonoros no verbales.

Es posible también aplicar algoritmos de análisis de voz para identificar emociones, pero este análisis de intención suele estar suficientemente bien cubierto mediante el análisis del texto.

CAPA 2 (Comprensión y calificación)

La segunda capa es de agregación, que recibe los resultados de la capa anterior, consolida el etiquetado (si es necesario se hace mapping con las etiquetas detectadas y las etiquetas finales) y califica el contenido audiovisual.

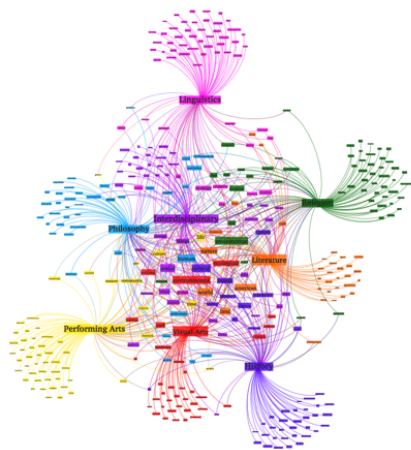
Esta capa permite realizar una abstracción respecto de dos aspectos:

1. **Abstracción de la tecnología de etiquetado:** esta capa tiene sentido con independencia de las tecnologías que se utilicen para etiquetar. Habilita la posibilidad de utilizar sistemas de etiquetado distintos o de cambiar de proveedor sin

que el resultado final deba verse dramáticamente afectado.

2. **Abstracción de la taxonomía de etiquetado:** en esta capa se puede implementar un mapeado de etiquetas habilitando utilizar una taxonomía propia mapeada con la de la capa anterior. De este modo, con independencia del proveedor tecnológico y su taxonomía, la salida del sistema podría seguir siendo la misma.
3. **Adición de capacidades adicionales de agregación y comprensión:** Para este análisis agregado recomendamos la utilización de algoritmos de **Topic Modeling** que permite la organización de etiquetas por conceptos y analizar la cercanía entre ellos. De este modo se puede determinar con mayor precisión por ejemplo si un desnudo tiene un propósito médico o pornográfico.

Análisis basado en campos semánticos (Topic Modeling)



- Permite analizar los hechos por campos semánticos
- Permite analizar la proximidad de los diferentes tópicos de cara a comprender mejor el contenido
(Por ejemplo, la proximidad entre desnudos y médico permite inferir con mayor probabilidad que es a pesar de que se muestran desnudos, no es un contenido sexual potencialmente hiriente)

Figura 6: Topic Modeling

3.1.2. Modelo organizativo

En la actualidad, la evolución tecnológica tan brutal que está viviendo el sector exige una vigilancia constante de elementos que aparecen, mutan y evolucionan a un ritmo vertiginoso. Los elementos principales que están en juego a nuestro juicio son:

- **Tecnología:** Es evidente, y en el presente estudio queda de manifiesto, que la evolución tecnológica permite posibilidades que hace años eran impensables (como el propio etiquetado automático de contenidos audiovisuales). La precisión de estos sistemas tiende a incrementarse, aunque como señalamos a continuación, también evoluciona la tipología de contenidos, por lo que la necesidad de adaptación a las

nuevas realidades es continua.

- **Formatos y plataformas:** La industria audiovisual ha aumentado significativamente los espacios donde se consume. Mientras los formatos clásicos como las películas, series, documentales y programas de TV/radio se siguen produciendo, con la llegada de esos nuevos espacios, aparecen nuevos formatos que no existían hace pocas fechas (YouTube, TikTok, podcasts, etc.).
- **Contenidos:** El mayor cambio en las últimas décadas relacionadas con los contenidos audiovisuales ha sido la llegada del usuario como creador de contenidos. La cantidad de contenido audiovisual generado por el usuario aumenta cada día a un ritmo exponencial. Este contenido va asociado con la actualidad y con códigos de comunicación que evolucionan. En consecuencia, también cambia cómo se manifiestan los elementos que son potencialmente sensibles para el espectador y, en consecuencia, cómo se identifican, cómo se sitúan en el contexto y cómo se califican.
- **Regulación:** En consecuencia, los reguladores tratan de dar respuesta a esta realidad cambiante. Prueba de ello es la Ley de Servicios Digitales ("Digital Services Act", DSA) que está actualmente en tramitación europea.

Teniendo en cuenta este contexto se recomienda la existencia de un **organismo coordinador** conectado con reguladores y plataformas que de forma continua realice esta vigilancia y sea proactiva a la hora de proponer adaptaciones del sistema que permita incrementar su eficacia y eficiencia.

Necesidad de organismo / gestor



Figura 7: Modelo organizativo

Es importante que exista una **fuerte conexión con los reguladores**, de modo que se efficienten los procesos de detección de incumplimientos para que se actúe con celeridad. También es importante desde el punto de vista propositivo de adaptaciones y cambios regulatorios a tenor de los cambios que identifique en los elementos claves antes citados.

Es fundamental también el **trabajo conjunto con las plataformas audiovisuales** para establecer una taxonomía común de etiquetado de contenidos. Proveen de un sistema que ayude a realizar este etiquetado tiene todo el sentido en un **entorno de auto-corregulación**, donde las plataformas están implicadas en:

- La participación en la elaboración de la taxonomía de etiquetado.
- La realización del etiquetado automático acorde a la regulación y/o el uso de una herramienta común para etiquetar de forma única las obras.
- Aportar feedback sobre el proceso de etiquetado para contribuir en su refinamiento y mejora.

Sistema de auto-corregulación (ilustrativo)

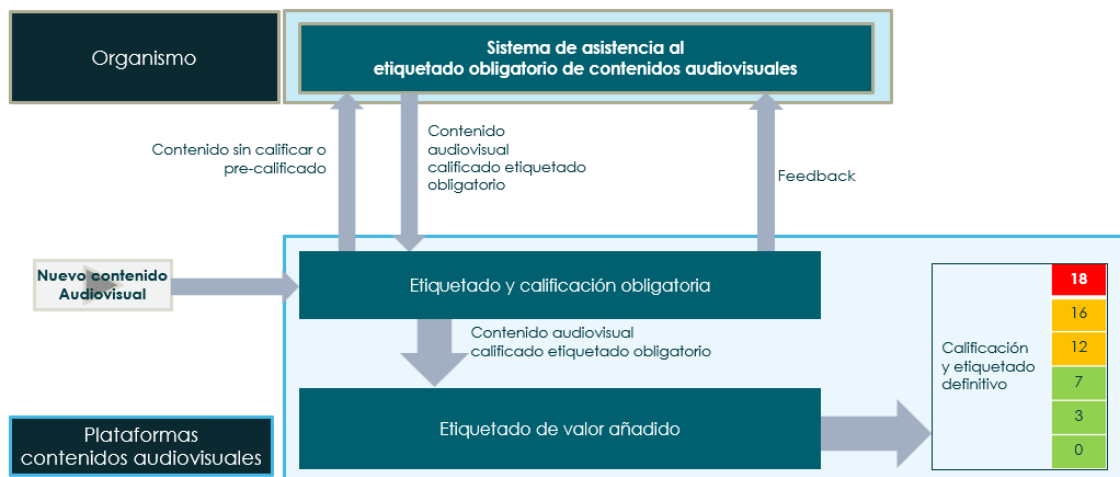


Figura 8: Sistema de correulación con sistema común de etiquetado automático

En este escenario, las plataformas dispondrían de dos niveles de etiquetado:

- **Identificación de etiquetas comunes relacionados con la obra y su calificación única.**

- Metadatos propios de la plataforma, que forman parte de su valor diferencial y que alimentan sus sistemas de recomendación y enriquecimiento de la experiencia de usuario.

Pese a que las plataformas sin base en el territorio español pueden no estar obligadas a cumplir con la regulación de etiquetado, es altamente aconsejable que participen del modelo por la extensión de uso que tienen en la actualidad.

3.1.3. Instrumentos de contratación pública

Como conclusiones de este estudio se destacan dos hechos ya señalados en el análisis tecnológico:

1. Existe en el mercado tecnología de etiquetado automático de contenidos audiovisuales especialmente orientadas a la detección de contenido sensible para el menor suficientemente maduras como para tener experiencia de años dando este tipo de servicios.
2. Las soluciones de etiquetado automático están compuestas por un stack tecnológico complejo con un número elevado de componentes distintos. Para cada uno de estos múltiples elementos existe una gran variedad de opciones. La forma en la que se mezclan y se implementa la solución final está dentro de un rasgo inmensamente grande de combinaciones de tecnologías.

Asociación para la Innovación (no recomendado)

En el aspecto de los instrumentos de contratación pública, al inicio del proyecto se planteaba la posibilidad de optar por una **Asociación para la Innovación** de cara a explorar distintas alternativas para crear una nueva tecnología. La base fundamental para la utilización de este instrumento es que para la necesidad que se quiere cubrir no exista actualmente soluciones tecnológicas que la satisfagan. Como sea probado, esta precondition no se cumple en este caso.

Contratación de Servicios con Consulta Preliminar

Por tanto, el instrumento indicado es la **Contratación de Servicios**. El matiz importante que se recomienda a partir de las conclusiones del presente estudio es utilizar el instrumento de **Consulta Preliminar al Mercado (CPM)**. Este instrumento permite explorar el mercado para obtener propuestas tecnológicas definidas que contemplen todo el stack tecnológico y que ayuden a concretar las prescripciones técnicas de la futura licitación en términos de infraestructura, stack tecnológico, equipo y costes.

Con las conclusiones obtenidas de la CPM, se procede a la **redacción del pliego** para la contratación.

Este proyecto es complejo, **con muchas componentes tecnológicas y organizativas**. Recomendamos la creación de una **oficina técnica** que gestione los procesos y las personas implicados que abarcan no sólo cuestiones tecnológicas, sino también legales, organizativas y metodológicas.

Figura 1: Sistema elemental · Etiquetado

Instrumentos de contratación · Roadmap recomendado


- 1   Consulta Preliminar al Mercado (CPM)
- 2   Contratación: Elaboración del pliego
- 3   Oficina técnica

Figura 9: Roadmap recomendado

3.2. Modelo de autorregulado de certificación de soluciones

El modelo alternativo al desarrollo de una solución específica de asistencia al etiquetado automático de contenido audiovisual es un modelo de certificación de soluciones que cuente con la participación de las plataformas audiovisuales para corregular el sistema.

El modelo organizativo recomendado en este ámbito es muy similar al caso del desarrollo de la solución. Se recomienda la existencia de un **organismo coordinador / supervisor** conectado con reguladores y plataformas que de forma continua realice esta vigilancia y sea proactiva a la hora de proponer adaptaciones del sistema que permita incrementar su eficacia y eficiencia.

Necesidad de organismo / gestor



Figura 10: Modelo organizacional

Es importante que exista una **fuerte conexión con los reguladores**, de modo que se efficienten los procesos de detección de incumplimientos para que se actúe con celeridad. También es importante desde el punto de vista propositivo de adaptaciones y cambios regulatorios a tenor de los cambios que identifique en los elementos claves antes citados.

Es fundamental también el **trabajo conjunto con las plataformas audiovisuales** para establecer una taxonomía común de etiquetado de contenidos. Proveer de un sistema que ayude a realizar este etiquetado tiene todo el sentido en un **entorno de auto-corregulación**, donde las plataformas están implicadas en:

- La participación en la elaboración de una taxonomía de etiquetado común. Sobre esta cuestión actualmente hay esfuerzos en la industria por determinar una suerte de estándar en cuanto al reconocimiento y etiquetado de contenidos. Un ejemplo de estos esfuerzos es la taxonomía IAB (<https://iabtechlab.com/standards/content-taxonomy/>) impulsada por las empresas de marketing digital que necesitan entender de forma automática qué ocurre para explotar esa información.
- La realización del etiquetado automático acorde a la regulación y/o el uso de una herramienta común para etiquetar de forma única las obras
- Aportar feedback sobre el proceso de etiquetado para contribuir en su refinamiento y mejora.

A imagen de lo que sucede en otros países del entorno europeo, poniendo como ejemplo concreto Alemania, existen diferentes organismos no gubernamentales que participan del sistema de autorregulación (por ejemplo, FSM). En este país existe varias organizaciones de autocorregulación dependiendo asociadas a actividades concretas de la industria

audiovisual.

Sistema de auto-corregulación (ilustrativo)



Figura 11: Sistema de correulación · proveedores certificados

En este escenario, como en el anterior, se recomienda que las plataformas puedan disponer de dos niveles de etiquetado:

- Identificación de etiquetas comunes relacionados con la obra y su calificación única, independiente de la plataforma.
- Metadatos propios de la plataforma, que forman parte de su valor diferencial y que alimentan sus sistemas de recomendación y enriquecimiento de la experiencia de usuario.

A diferencia del modelo anterior, aquí el organismo supervisor no tiene relación directa con el sistema de etiquetado concreto utilizado por las plataformas. Las plataformas pueden hacerlo internamente o utilizar proveedores a tal efecto.

Lo que puede hacer este organismo, junto con los organismos de autorregulación de la industria, es **crear un sistema de certificación de proveedores** que facilite la labor de contratación al asegurar que el proveedor cumple con las características que la regulación exige para el etiquetado y calificación por rangos de edades.

Pese a que las plataformas sin base en el territorio español pueden no estar obligadas a cumplir con la regulación de etiquetado, es altamente aconsejable que participen del modelo por la extensión de uso que tienen en la actualidad.