


# RESÚMENES TEXTUALES DE DOCUMENTOS AUDIOVISUALES A TRAVÉS DE IA PARA DEPARTAMENTOS DE INFORMACIÓN EN TELEVISIÓN

Textual summaries of audiovisual document using AI for television new department

Jorge Caldera Serrano

Universidad de Extremadura, Departamento de Información y Comunicación  
Badajoz, España  
jcalser@unex.es

<https://orcid.org/0000-0002-4049-1057> 

A lista completa com informações dos autores está no final do artigo 

## RESUMO

**Objetivo:** analizar las posibilidades documentales que tiene la confección de resúmenes automáticos por medio de los desarrollos algorítmicos de la Inteligencia Artificial para la automatización de la extracción de resúmenes textuales de documentales audiovisuales televisivos en el marco de la gestión de la información audiovisual en los Media/Digital Asset Management.

**Método:** Tras una búsqueda de información en bases de datos sobre resúmenes documentales, Inteligencia Artificial y su desarrollo en la confección de resúmenes, y una búsqueda por las principales organizaciones que desarrollan algoritmos de IA, se lleva a cabo una valoración y selección de ellos por medio de una exhaustiva revisión bibliográfica.

**Resultado:** Se exponen herramientas de IA y el valor que puede tener la integración de dicha automatización en la gestión documental en las televisiones facilitando el acceso a los contenidos de forma más ágil y rápida, tanto por parte de los profesionales de la documentación como por los periodistas de la cadena.

**Conclusiones:** existen herramientas para la generación de resúmenes textuales de documentos audiovisuales, no obstante, aún se están desarrollando algoritmos para la mejora del tratamiento para lograr así optimar resultados.

**PALAVRAS-CHAVE:** Resumen documental. Documentación audiovisual. Información televisiva. Automatización de procesos. Acceso a la información. Inteligencia Artificial.

## ABSTRACT

**Objective:** to analyze the documentary possibilities of the preparation of automatic summaries through the algorithmic developments of Artificial Intelligence for the automation of the extraction of textual summaries of television audiovisual documentaries within the framework of the management of audiovisual information in the Media/ Digital Asset Management.

**Methods:** After a search for information in databases on documentary summaries, Artificial Intelligence and its development in the preparation of summaries, and a search for the main organizations that develop AI algorithms, an assessment and selection of them is carried out by through an exhaustive bibliographic review.

**Results:** AI tools and the value that the integration of said automation can have in document management in television networks are presented, facilitating access to content in a more agile and rapid way, both by documentation professionals and by network journalists.

**Conclusions:** there are tools for generating textual summaries of audiovisual documents, however algorithms are still being developed to improve treatment to achieve optimal results.

**KEYWORDS:** Documentary summary. Audiovisual documentation. Television information. Process automation. Access to information. Artificial Intelligence.

## 1 INTRODUCCIÓN

La evolución tecnológica es un proceso fascinante que ha transformado de manera radical y fundamental la forma en que nos relacionamos, vivimos, trabajamos, educamos y comunicamos. Desde la perspectiva de las Tecnologías de la Información y la Comunicación, el primer gran desarrollo fue el telégrafo, inventado en el siglo XIX, y posteriormente el teléfono en 1876 por Alexander Graham Bell, que catapultaron el

desarrollo tecnológico. El hito fundamental, casi fundacional, de una nueva era con la supremacía del Homo Videns, fue el surgimiento de la televisión en 1925 con la primera retransmisión de la BBC. Este desarrollo marcó un antes y un después en la manera en que consumimos información y entretenimiento.

El desarrollo de la computadora fue también un momento disruptivo. Alan Turing y John von Neumann sentaron las bases de todo el desarrollo posterior. El avance tecnológico en la era digital ha traído consigo una serie de dispositivos electrónicos con capacidades y posibilidades cada vez mayores. Ordenadores, tabletas, relojes inteligentes y otros periféricos se han convertido en parte integral de nuestra vida cotidiana. Además, estas innovaciones están estrechamente relacionadas con nuevas realidades como la realidad aumentada y la realidad virtual.

En los últimos años, hemos sido testigos de un nuevo hito tecnológico que podría equipararse a la invención de la rueda, la imprenta o la revolución industrial. Aunque inmerso en la era digital, este avance está transformando la forma en que nos relacionamos con nuestro entorno, tanto a nivel personal como global. La inteligencia artificial (IA), aunque aún en estado embrionario en algunos casos, nos sorprende con su potencial para automatizar tareas que antes eran exclusivas del ser humano. Desde asistentes virtuales hasta algoritmos de aprendizaje automático, la IA está revolucionando nuestras rutinas diarias (Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura, 2023). Especialmente destacable es la IA generativa, que altera nuestro ecosistema y promete ser una tecnología inmersiva en nuestro entorno vital.

La IA, transversal a todas las disciplinas, también se está incorporando en los quehaceres de las cadenas de televisión, y evidentemente también se prevé que la automatización de procesos y procedimientos llegue a los servicios de documentación audiovisual de las cadenas televisivas.

En este trabajo se pretende mostrar cómo se pueden llevar a cabo resúmenes de documentos audiovisuales que agilicen y faciliten el acceso a los contenidos televisivos utilizando algoritmos propios de la IA, automatizando el proceso de extracción de contenido y aportando herramientas para la recuperación documental.

## **2 OBJETOS, OBJETIVOS Y METODOLOGÍA**

El objetivo del trabajo es mostrar las opciones existentes con IA para realizar resúmenes de documentos audiovisuales orientados a los servicios de documentación de

las empresas televisivas. Los objetos analizados, por tanto, son múltiples, aunque relacionados. Por un lado, tenemos a la gestión de la documentación audiovisual, los Digital Asset Management de las cadenas de televisión, y la integración de la IA para automatizar uno de los procesos que se realizan.

La metodología empleada en esta investigación se enmarca dentro de los enfoques prospectivos y especulativos característicos de las Ciencias Sociales, utilizando principalmente métodos descriptivos y observacionales. Estos enfoques permiten explorar escenarios futuros y desarrollar hipótesis basadas en observaciones de tendencias actuales y datos existentes. Inicialmente, se realizó una exhaustiva revisión bibliográfica que abarcó tanto los temas de estudio propuestos como la literatura relacionada con la inteligencia artificial (IA). Esta revisión incluyó una variedad de fuentes académicas y científicas, así como publicaciones de investigación de referencia en el campo. Sin embargo, dada la naturaleza emergente de muchos de los conceptos y la lenta difusión científica en algunos casos, se recurrió también a fuentes web, en particular a sitios de desarrolladores de IA, para obtener información actualizada. Esto fue especialmente útil para rastrear contenidos relevantes para el análisis de los temas de estudio y para identificar las tendencias emergentes en algoritmos de IA.

Después de analizar una amplia gama de fuentes, se llevó a cabo una selección operativa del material básico, identificando los recursos de referencia más pertinentes para la investigación. Este proceso incluyó la selección de artículos, informes técnicos y estudios de caso relevantes, que proporcionaron una base sólida para el análisis posterior. A continuación, se emplearon técnicas de observación y análisis para examinar la realidad actual de la IA, especialmente en el contexto de su aplicación en cadenas televisivas y otros medios de comunicación. Mediante este análisis, se realizaron inferencias y descripciones sobre la capacidad potencial de los profesionales del sector para desempeñar roles más actuales a tenor de los requerimientos actuales y de las posibilidades tecnológicas.

Es importante subrayar que este trabajo tiene un carácter exploratorio, lo que significa que se centra en identificar posibilidades y escenarios futuros sin pretender proporcionar respuestas definitivas o soluciones cerradas. La implementación y prueba de las acciones sugeridas, tales como la adopción de tecnologías específicas de IA o la formación de personal en estas nuevas competencias, deberán ser llevadas a cabo directamente por las empresas televisivas. Estas organizaciones tendrán que evaluar la viabilidad práctica de las propuestas y adaptarlas a sus necesidades y recursos específicos,

considerando también el rápido avance tecnológico y las posibles implicaciones éticas y sociales del uso de IA en los medios de comunicación.

### 3 IA EN SERVICIOS DE DOCUMENTACIÓN TELEVISIVOS

La inteligencia artificial (IA) está desencadenando una profunda transformación en diversos aspectos de nuestras vidas, con un impacto que promete ser aún más significativo en el futuro. Definir la IA no es una tarea sencilla, como han señalado diversos autores a lo largo del tiempo (Mishkoff, 1988; Rouhiainen, 2018). Esta dificultad radica en la complejidad y la variedad de factores que influyen en el desarrollo de la IA, lo que complica la búsqueda de una definición que sea universalmente aceptada.

Amador Hidalgo (1996) propone una clasificación y diferenciación de la IA basada en su objeto y objetivo, aunque esta clasificación puede ser más aplicable a definiciones anteriores al siglo XXI. Por otro lado, Barrera Arestegui (2012) ofrece un análisis histórico de las definiciones de IA, abordándolas desde una perspectiva tanto tecnológica como filosófica. La revisión sistemática presentada por Gallegos *et al.* (2014) ofrece un recorrido didáctico por las diversas concepciones históricas de la IA.

En cuanto a enfoques más contemporáneos, varios estudios (Giletta *et al.*, 2020; Canavilhas, 2022; Ventura-Pocino, 2021; Pihlajarinne; Alén-Savikko, 2022) examinan el concepto de IA desde perspectivas sociales, económicas, ideológicas y éticas, ofreciendo una visión multidimensional de la tecnología. Rouhiainen (2018) proporciona una definición concisa de la IA, respaldada por otros expertos (Crawford, 2021), que subraya la capacidad de las computadoras para realizar tareas que tradicionalmente requerían inteligencia humana, utilizando algoritmos complejos y aprendizaje autónomo a partir de grandes volúmenes de datos.

Armstrong (2016) identifica varias áreas clave en el desarrollo de la IA, incluyendo el reconocimiento de imágenes, algoritmos comerciales, aplicaciones en el ámbito sanitario, predicción industrial, ciberseguridad y marketing personalizado en redes sociales. Para mantenerse al tanto de las tendencias en el desarrollo de la IA, se recomienda consultar a los principales desarrolladores y fuentes especializadas en tecnología.

Es importante señalar que, aunque se están realizando esfuerzos significativos para mejorar la gestión de la información en los medios de comunicación, especialmente en televisión (Bazán-Gil, 2023), todavía estamos lejos de alcanzar el pleno potencial que la tecnología puede ofrecer en este ámbito. Existen estudios que abordan la IA desde la

perspectiva de los medios de comunicación, tanto en su capacidad para generar información (Calvo-Rubio; Ufarte-Ruiz, 2021; Rojas-Torrijos, 2021; Brennen; Howard; Nielsen, 2002) como en diferentes etapas de la cadena de producción informativa (Diakopoulos, 2019; Túnnez-López, Fieiras-Ceide; Vaz-Álvarez, 2021). Las primeras aplicaciones conocidas de IA en el periodismo datan de 2015, cuando *Los Ángeles Times* y luego *Associated Press* comenzaron a utilizarla para redactar artículos automáticamente (Sánchez-García, *et al.*, 2023). Sin embargo, es evidente que, pese a la disponibilidad de herramientas y aplicaciones, la implementación completa de la IA en los medios de comunicación sigue siendo limitada (Bruccoleri *et al.*, 2022; AI4Media, 2023), ya sea por la falta de recursos para su adopción o porque el desarrollo necesario aún no se ha alcanzado.

En el contexto de los sistemas documentales televisivos, algunas de las soluciones más avanzadas incluyen *Microsoft Azure Video Indexer*, que ofrece transcripciones y realiza reconocimiento facial y de objetos; *Vilynx* y *Telescope*, que están diseñadas específicamente para televisión; y *Google Cloud Video Intelligence API*, *IBM Watson Media*, *Amazon Rekognition Video*, *VidMOB* y *Wipster*, que etiquetan objetos, escenas, entre otros, generando descriptores.

#### **4 RESUMEN DE DOCUMENTOS AUDIOVISUALES CON IA: METODOLOGÍA, USOS Y APLICACIONES**

Los resúmenes documentales juegan un papel crucial en el ámbito de las ciencias de la documentación, al facilitar el acceso a información clave de manera rápida y eficiente. Estos resúmenes, que son representaciones abreviadas del contenido de documentos más extensos, permiten a los profesionales de la información y a los usuarios en general identificar rápidamente la relevancia de un documento para sus necesidades específicas. De acuerdo con Rowley & Farrow (2017), los resúmenes documentales ayudan a ahorrar tiempo y esfuerzo, permitiendo a los usuarios concentrarse en los textos -o imágenes en movimiento- que realmente necesitan leer o visualizar con mayor detalle o en su totalidad.

En un contexto de creciente sobrecarga de información, donde la cantidad de publicaciones científicas y técnicas se incrementa exponencialmente, la función de los resúmenes documentales se vuelve aún más esencial. Según Lancaster (2003), los resúmenes no solo proporcionan una visión general del contenido del documento, sino que también ayudan a la indexación y recuperación de información en sistemas de bases de datos y motores de búsqueda. Esto es particularmente importante en disciplinas

académicas y científicas, donde la rápida obtención de información relevante puede marcar la diferencia en la realización de investigaciones efectivas y actualizadas.

Además, los resúmenes documentales desempeñan un papel fundamental en la transferencia de conocimiento entre diferentes disciplinas. Tal y como destaca Borko (1968), un resumen bien redactado puede hacer que la información contenida en un documento sea accesible no solo para expertos en un campo específico, sino también para profesionales de otras áreas. Esto fomenta la interdisciplinariedad y la colaboración, permitiendo a los investigadores y profesionales aplicar conocimientos de un campo a otro, lo que puede conducir a innovaciones y nuevas soluciones a problemas complejos.

La inteligencia artificial (IA) ha demostrado ser una herramienta revolucionaria para la generación de resúmenes de documentos audiovisuales. Esta capacidad se basa en técnicas avanzadas de procesamiento del lenguaje natural (NLP, por sus siglas en inglés) y visión por computadora, permitiendo la condensación de grandes volúmenes de contenido audiovisual en resúmenes textuales precisos y útiles.

#### **4.1 Tecnología que soporta la generación de resúmenes**

La creación de resúmenes textuales de documentos audiovisuales utilizando inteligencia artificial (IA) es un proceso que combina tecnologías avanzadas de reconocimiento de voz, procesamiento de lenguaje natural (NLP, por sus siglas en inglés) y modelos de generación de lenguaje. A continuación, se detallan los pasos clave de este proceso:

El primer paso en la creación de resúmenes textuales a partir de contenido audiovisual es la transcripción del audio a texto, utilizando sistemas de reconocimiento automático del habla (ASR). Estos sistemas capturan el audio del vídeo y lo convierten en texto escrito, extrayendo tanto diálogos como otros elementos verbales importantes. Este proceso es crucial porque proporciona la base textual necesaria para los pasos subsecuentes de procesamiento y resumen (Park *et al.*, 2019).

Una vez obtenida la transcripción, se emplean técnicas de procesamiento de lenguaje natural (NLP) para analizar y refinar el texto. Este análisis incluye varios subprocesos, como la extracción de entidades, donde se identifican nombres de personas, lugares, organizaciones, etc.; la identificación de temas, que clasifica y agrupa partes del texto según su contenido temático; y la detección de sentimientos, que evalúa el tono emocional del discurso. Estas técnicas permiten una comprensión más profunda del



contenido y ayudan a identificar las partes más relevantes del texto para incluir en el resumen (Liu *et al.*, 2021).

Puede también integrarse información textual procedente de la información audiovisual vía la captación de texto presente en la imagen por medio del reconocimiento óptico de caracteres (OCR). Estos textos pueden proceder tanto de subtítulos como de información integrada en la imagen (Smith, 2007).

El reconocimiento biométrico, que incluye tecnologías procedentes del reconocimiento facial y de la identificación de personas por el reconocimiento de voz, puede ayudar de manera significativa para la generación de resúmenes automáticos, identificando claramente a las personas físicas y jurídicas presentes en la pieza audiovisual. En el marco de esta tecnología pueden detectarse y extraerse, asociado a los personajes, una gama amplia de emociones a través de la expresión facial y del análisis de la voz, que aportan información del contexto que ayuda al mejor entendimiento de la acción y, por lo tanto, a la confección de un mejor resumen (Poria *et al.*, 2017).

Destacado es la potencial integración del contexto audiovisual, lo que implica el uso de técnicas de visión por computadora para interpretar y describir el contenido, pudiendo detectar y clasificar objetos, escenas, actividades y movimientos (identificar si es una imagen deportiva, de una manifestación, de una entrevista, etc.), proporciona contexto para la comprensión del material y facilita el resumen (Simonyan; Zisserman, 2015).

La detección de cambio en la escena (segmentación de imágenes) se realiza atendiendo a cambios en la imagen, tal como textura, colores, etc., de tal manera que sea capaz de discriminar cuándo de una acción se pasa a una declaración, por ejemplo, de tal manera que se puede segmentar el contenido en unidades más manejables que puedan tener, posteriormente, una única representación semántica (Yeung *et al.*, 2016).

Igualmente, puede utilizarse para la confección de resúmenes documentales los metadatos asociados al vídeo, tales como fecha, ubicación, participantes, autores, etc. (Caldera-Serrano; Freire-Andino, 2016), siempre que esta información sea pertinente para la descripción del contenido.

Puede integrarse en el programa para la confección de resúmenes la detección y etiquetado de contenido inapropiado o potencialmente ofensivo. Cuestión importante para incluir y alertar en el resumen sobre imágenes potencialmente inadecuada para la audiencia o la detección de uso inadecuado del lenguaje en la banda de audio.

Para la confección de los resúmenes se utilizan dos enfoques: el extractivo, en el que se extraen el texto directamente del original, seleccionadas atendiendo a métricos

como frecuencia de palabras claves, importancia en el contexto y su relación con los temas identificados (Gambhir; Gupta, 2017); o los abstractivos, en el que se genera nuevas frases, extraídas de modelos de IA avanzados como BERT y GPT (Devlin *et al.*, 2019).

## 4.2 Aplicaciones para la realización de resúmenes

Presentamos a continuación las que estimamos las diez principales herramientas para la realización de resúmenes, con validez para llevar a cabo la condensación de documentos audiovisuales televisivos. Se ha realizado una valoración subjetiva puntuando sobre 10 la adaptabilidad a la realización de resúmenes de documentos audiovisuales televisivos.

### *Google Cloud Video Intelligence*

Uso: Aplicación que realiza resúmenes y analiza contenidos visuales y sonoros.

Ventajas: Alta precisión en los resultados e integración con otros servicios de Google Cloud que mejora resultados.

Desventajas: Requiere experiencia técnica para su configuración e integración en un sistema de gestión de información. Coste elevado para su uso.

Web: [*Google Cloud Video Intelligence*] <https://cloud.google.com/video-intelligence>

Valoración: 9. Ofrece un conjunto completo de herramientas, incluyendo OCR para texto y subtítulos, reconocimiento de personas, detección de contexto audiovisual, y alertas para contenido inapropiado. Su integración con otros servicios de Google lo hace altamente versátil para profesionales que buscan análisis detallados y precisos.

### *IBM Watson Media*

Uso: Análisis de contenido de vídeo, transcripción y reconocimiento de objetos.

Ventajas: Personalización y robustas opciones de análisis.

Desventajas: Complejidad en la configuración inicial; precios altos para características avanzadas.

Web: [*IBM Watson Media*] <https://www.ibm.com/watson/media/>

Valoración: 8.5. Similar a Google, IBM Watson Media ofrece capacidades avanzadas como OCR, reconocimiento de personas y emociones, y análisis de contexto. Su personalización y robustez son excelentes, pero puede ser complejo y caro, lo que puede limitar su accesibilidad para algunos usuarios.



### *Microsoft Azure Video Indexer*

Uso: Transcripción de audio a texto, y una vez realizado, pueden confeccionarse los resúmenes. Etiquetado de escenas y análisis de contenido.

Ventajas: Amplia integración con Azure y capacidades avanzadas.

Desventajas: Curva de aprendizaje para usuarios no técnicos; costes altos asociados a un uso intensivo.

Web: [*Microsoft Azure Video Indexer*] <https://azure.microsoft.com/en-us/services/media-services/video-indexer/>

Valoración: 8. Proporciona funcionalidades avanzadas de análisis de vídeo, incluyendo OCR, reconocimiento de personas y metadatos asociados. Es una opción potente para quienes ya están integrados en el ecosistema Azure, aunque la curva de aprendizaje y los costes asociados a su implementación son altos.

### *Otter.ai*

Uso: Transcripción automática de reuniones, entrevistas y otros eventos hablados.

Ventajas: Fácil de usar, buena precisión y potencialidad para el intercambio y la colaboración.

Desventajas: Limitaciones en la personalización de transcripciones; depende de la calidad del audio.

Web: [*Otter.ai*] <https://otter.ai>

Valoración: 7. Ofrece herramientas sólidas para transcripción y edición, con buena precisión. Sin embargo, como Otter.ai, carece de funciones avanzadas como reconocimiento biométrico y análisis de contenido audiovisual, lo que limita su utilidad para resúmenes más complejos.

### *Trint*

Uso: Transcripción automática y edición de contenido de audio y vídeo.

Ventajas: Interfaz intuitiva, herramientas de edición colaborativa.

Desventajas: Precio por suscripción; variabilidad en la precisión dependiendo del acento y la claridad del audio.

Web: [*Trint*] <https://trint.com>

Valoración: 7.5. Ofrece herramientas sólidas para transcripción y edición, con buena precisión. Sin embargo, como Otter.ai, carece de funciones avanzadas como

reconocimiento biométrico y análisis de contenido audiovisual, lo que limita su utilidad para resúmenes más complejos

### *Descript*

Uso: Transcripción, edición de audio y vídeo, y generación de resúmenes.

Ventajas: Herramientas integradas, fácil colaboración.

Desventajas: Puede ser costoso para equipos grandes; algunas características avanzadas pueden tener una curva de aprendizaje alto.

Web: [*Descript*] <https://www.descript.com>

Valoración: 8. Combina transcripción con edición de audio y vídeo, facilitando la colaboración. Carece de funciones avanzadas de análisis de contenido, su facilidad de uso y funcionalidades integradas lo hacen una opción robusta.

### *Sonix*

Uso: Transcripción automática y traducción de archivos de audio y vídeo.

Ventajas: Alta precisión, múltiples idiomas.

Desventajas: Suscripción mensual; calidad de transcripción puede variar con acentos y ruido de fondo.

Web: [*Sonix*] <https://sonix.ai>

Valoración: 7.5. Es similar a *Trint* y *Otter.ai* en sus capacidades de transcripción, con soporte multilingüe. Sin embargo, su enfoque es más limitado en términos de análisis y no ofrece herramientas avanzadas para el análisis de contenido audiovisual.

### *Rev.ai*

Uso: API de transcripción automática para archivos de audio y vídeo.

Ventajas: Precisión en la transcripción y fácil integración con otras aplicaciones.

Desventajas: Puede ser costoso para volúmenes con muchos datos y peso informático; requiere habilidades técnicas para integración.

Web: [*Rev.ai*] <https://www.rev.ai>

Valoración: 8.5. Alta precisión en transcripciones y fácil integración a través de APIs, pero como las anteriores, carece de análisis de contexto audiovisual y funciones avanzadas de reconocimiento

### *Kapwing*

Uso: Edición de vídeo en línea con capacidades de subtulado y transcripción.

Ventajas: Herramientas de edición simples, accesibles.

Desventajas: Limitaciones en funciones avanzadas; depende de una buena conexión a internet.

Web: [*Kapwing*] <https://www.kapwing.com>

Valoración: 6.5. Principalmente una herramienta de edición de vídeo en línea, con capacidades básicas de subtulado y transcripción. No ofrece análisis avanzados ni capacidades de reconocimiento, lo que limita su validez para resúmenes complejos.

### *Happy Scribe*

Uso: Transcripción y subtulado de vídeos y audios automáticamente.

Ventajas: Multilingüe, interfaz amigable.

Desventajas: Puede no manejar bien acentos fuertes; necesita suscripción para características completas.

Web: [*Happy Scribe*] <https://www.happyscribe.com>

Valoración: 7. Similar a otras herramientas de transcripción, es útil para subtulado y transcripción en múltiples idiomas. Sin embargo, carece de análisis avanzado y herramientas de reconocimiento, lo que limita su utilidad en contextos más exigentes.

La validez para realizar resúmenes sobre material audiovisual varía según factores como la precisión de transcripción, la facilidad de uso, las capacidades de personalización y las características avanzadas como el análisis de sentimientos o el etiquetado de escenas. Se debe considerar estos factores y las necesidades específicas de su aplicación para seleccionar la herramienta más adecuada.

Para realizar la valoración numérica se han tenido en cuenta el rango de funcionalidades, analizando la presencia de opciones como OCR, reconocimiento biométrico, detección de emociones y contexto; la precisión en la transcripción y análisis; la facilidad de uso; la integración con otros sistemas y la escalabilidad especialmente para grandes empresas audiovisuales; y el coste y la accesibilidad.

A continuación, se presenta una tabla que indica las funcionalidades específicas de las diez aplicaciones mencionadas, incluyendo si cuentan con OCR para texto y subtítulos, reconocimiento biométrico para personas y emociones, detección de contexto audiovisual, detección de cambio de escenas, detección de metadatos asociados al vídeo, y alertas de imágenes y palabras inapropiadas.

Tabla 1: Aplicaciones IA para confección resumen audiovisuales y características

<i>Aplicación</i>	1	2	3	4	5	6	7	8	9
Google Cloud Video <i>Intelligence</i>	✓	✓	✓	✗	✓	✓	✓	✓	✓
IBM Watson Media	✓	✓	✓	✓	✓	✓	✓	✓	✓
Microsoft Azure Video <i>Indexer</i>	✓	✓	✓	✗	✓	✓	✓	✓	✓
Otter.ai	✗	✗	✗	✗	✗	✗	✗	✗	✗
Trint	✗	✗	✗	✗	✗	✗	✗	✗	✗
Descript	✗	✗	✗	✗	✗	✗	✗	✗	✗
Sonix	✗	✗	✗	✗	✗	✗	✗	✗	✗
Rev.ai	✗	✗	✗	✗	✗	✗	✗	✗	✗
Kapwing	✗	✗	✗	✗	✗	✗	✗	✗	✗
Happy Scribe	✗	✗	✗	✗	✗	✗	✗	✗	✗

Fuente: elaboración propia

Explicación de tabla:

1.- OCR Texto y 2 OCR subtítulo: reconocimiento de textos y subtítulos en los vídeos, extrayendo dicha información para su transcripción y análisis.

3.- Reconocimiento biométrico y 4 reconocimiento de emociones: capacidad de reconocimiento de personas y emociones, utilizando para ello tecnología biométrica.

5.- Contexto audiovisual, análisis de contenido para entender el contexto general, detectando acciones, secuencias y escenas (identificar entrevistas, deportes, etc.)

6.- Detección cambio de escenas. Identificación de transiciones entre diferentes escenas y secuencias.

7.- Detección y uso de metadatos asociados. Detección de información asociada al vídeo vía metadatos y que puede aportar información relevante.

8.- Detección imágenes inapropiadas y 9 detección de audios inapropiados. Capacidad del sistema de identificar contenido visual o verbal que se entiende inadecuado u ofensivo.

## 5 CONCLUSIONES

La creación de resúmenes audiovisuales a partir de documentos audiovisuales es una tecnología emergente que combina técnicas de procesamiento de vídeo y NLP. Las aplicaciones más avanzadas son *Google Cloud Video Intelligence*, *IBM Watson Media* y *Microsoft Azure Video Indexer* que ofrecen un conjunto más completo de características avanzadas para el análisis de contenido audiovisual. Las demás herramientas, como *Otter.ai*, *Trint* y *Sonix* se centran principalmente en la transcripción y edición de audio y vídeo, sin las capacidades avanzadas de análisis contextual o reconocimiento biométrico.

A pesar de los avances significativos, la creación de resúmenes de documentos audiovisuales con IA aún enfrenta varios desafíos. La precisión del reconocimiento

automático del habla puede verse afectada por la calidad del audio, la presencia de múltiples hablantes y el uso de jergas o acentos (Amodei *et al.*, 2016). Además, la generación de resúmenes que mantengan la coherencia y el contexto del contenido original sigue siendo un desafío complejo.

El futuro de esta tecnología apunta a la mejora continua de los algoritmos de NLP y visión por computadora. El desarrollo de modelos más avanzados y precisos, junto con el aumento de la capacidad de procesamiento de datos, permitirá la creación de resúmenes más precisos y útiles. La integración de tecnologías emergentes como la realidad aumentada y la realidad virtual también promete transformar la manera en que consumimos y procesamos información audiovisual (Liu *et al.*, 2021).

La creación de resúmenes se dirige hacia una mayor integración de capacidades multimodales. Esto implica que las herramientas emergentes no solo procesarán textos escritos, sino que también serán capaces de interpretar y resumir información proveniente de imágenes, videos y otros formatos multimedia. Esta capacidad permitirá generar resúmenes más completos y enriquecidos, fusionando diversas fuentes de datos en una única síntesis coherente. La IA podrá analizar simultáneamente contenido visual y textual para proporcionar una visión más amplia y detallada de cualquier material documental.

Además, las herramientas de IA podrán personalizar los resúmenes según el perfil de los usuarios, como documentalistas audiovisuales y periodistas. Para los documentalistas, la IA podrá generar resúmenes ajustados a los requerimientos específicos de catalogación y archivo audiovisual, facilitando el procesamiento de grandes volúmenes de contenido en diferentes formatos. Para los periodistas, la IA podrá adaptar los resúmenes en relación a su estilo de trabajo, considerando sus usos frecuentes y el historial de búsqueda, de manera que reciban información sintetizada que se ajuste a sus intereses y necesidades informativas de manera rápida y precisa.

Esta evolución permitirá que los documentalistas audiovisuales, especialmente en el ámbito televisivo, puedan concentrarse en labores más críticas como la verificación y depuración de información. En un entorno mediático donde los *deepfakes* y las *fake news* son cada vez más frecuentes, la IA asumirá las tareas de resumen, que tradicionalmente han requerido mucho tiempo y habilidades especializadas. De este modo, los documentalistas podrán enfocar sus esfuerzos en la validación de contenidos, asegurando la calidad y veracidad de la información que se transmite al público.

## RECONOCIMIENTO

This research was supported by the Junta de Extremadura and FEDER Funds “A way to make Europe” [QUINARI Research Group SEJ013 GR18130].

## REFERÊNCIAS

AI4MEDIA. **The AI4Media project**. Unión Europea, 2023. Sitio web: ai4media.eu. Disponible en: <https://www.ai4media.eu>. Acceso en: 28 out. 2024.

AMADOR HIDALGO, L. **Inteligencia artificial y sistemas expertos**. Córdoba: Universidad de Córdoba, 1996. Disponible en: <https://helvia.uco.es/handle/10396/6938>. Acceso en: 28 out. 2024.

AMODEI, D.; ANANTHANARAYANAN, S.; ANUBHAI, R.; BAI, J.; BATTENBERG, E.; CASE, C.; CASPER, J.; CATANZARO, B.; CHENG, Q.; CHEN, G.; CHRZANOWKI, M.; COATES, A.; DIAMOS, G.; ELSEEN, E.; ENGEL, J.; FANG, W.; FAN, L.; FOUNGNER, C.; HANNUN, A.; JUN, B.; LeGRESLEY, P.; LI, X.; NARANG, S.; NG, A.; OZAI, S.; PRENGER, R.; SATHEESH, S.; SEETAPUN, D.; SENGUPTA, S.; WANG, Z.; WANG, Y.; XIAO, B.; ZHANG, C.; ZHU, Z. Deep speech 2: End-to-end speech recognition in English and Mandarin. In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING, 33., 2016, New York. **Anales** [...]. New York: JMLR, 2016. Disponible en: <https://doi.org/10.48550/arXiv.1512.02595>. Acceso en: 28 out. 2024.

ARSTRONG, M. **The future of A.I.** New York: 18 nov. 2016. Sitio web: statista.com. Disponible en: <https://www.statista.com/chart/6810/the-future-of-ai>. Acceso en: 28 out. 2024.

BARRERA ARRESTEGUI, L. Fundamentos históricos y filosóficos de la inteligencia artificial. *UCV-HACER. Revista de Investigación y Cultura, Chiclayo*, v. 1, n. 1, p. 87-92, 2012. Disponible en: <https://www.redalyc.org/pdf/5217/521752338014.pdf>. Acceso en: 28 out. 2024.

BAZÁN-GIL, V. Artificial intelligence applications in media archives. **Profesional de la información**, [S. l.], v. 32, n. 5, 2023. Disponible en: <https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/87392>. Acceso en: 28 oct. 2024.

BORKO, H. Information science: What is it?. **American Documentation**, New York, v. 19, n. 1, p. 3-5, 1968. Disponible en: <https://doi.org/10.1002/asi.5090190103>. Acceso en: 28 out. 2024.

BRENNEN, J. S.; HOWARD, P. N.; NIELSEN, R. K. What to expect when you're expecting robots: Futures, expectations, and pseudo-artificial general intelligence in UK news. **Journalism**, [S. l.], v. 23, n. 1, p. 22-38, 2022. Disponible en: <https://doi.org/10.1177/1464884920947535>. Acceso en: 28 out. 2024.

BRUCCOLERI, A.; IACOVIELLO, R.; MESSINA, A.; METTA, S.; MONTAGNUELO, M.; NEGRO, F. **AI in vision**: High quality video production & content automation. RAI. Turim:



Centre for Research, Technological Innovation and Experimentation, 2022.. Disponible en: <https://www.ai4media.eu/whitepapers/ai-in-vision-high-quality-video-production-content-automation>. Acceso en: 28 out. 2024.

CALVO-RUBIO, L.M.; UFARTE-RUIZ, M.J. Artificial intelligence and journalism: Systematic review of scientific production in Web of Science and Scopus (2008-2019). **Communication & Society**, Navarra, v. 34, n. 2, p. 159-176, 2021. Disponible en: <https://hdl.handle.net/10171/62242>. Acceso en: 28 out. 2024.

CALDERA-SERRANO, J; FREIRE-ANDINO, R.O. Los metadatos asociados a la información audiovisual televisiva por agentes externos al servicio de documentación: validez, uso y posibilidades. **Biblios**, Lima, v. 62, p. 63-75, 2016. Disponible en: <https://www.redalyc.org/pdf/161/16144742005.pdf>. Acceso en: 28 out. 2024.

CANAVILHAS, J. Inteligencia artificial aplicada al periodismo: estudio de caso del proyecto "A European Perspective" (UER). **Revista Latina de Comunicación Social**, Madrid, v. 80, p. 1–13, 2022. Disponible en: <https://nuevaepoca.revistalatinacs.org/index.php/revista/article/view/1671>. Acceso en: 28 out. 2024.

CRAWFORD, K. **The atlas of AI: Power, politics, and the planetary costs of artificial intelligence**. New haven: Yale University Press, 2021.

DEVLIN, J.; CHANG, M. W.; LEE, K.; TOUTONOVA, K. BERT: Pre-training of deep bidirectional transformers for language understanding. *In*: CONFERENCE OF THE NORTH AMERICAN CHAPTER OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 2019, Minneapolis. **Anales** [...]. Minneapolis: ACL, 2019. p. 4171-4186.

DIAKOPOULOS, N. **Automating the news: How algorithms are rewriting the media**. Cambridge: Harvard University Press, 2019. Disponible en: <https://doi.org/10.4159/9780674239302>. Acceso en: 28 out. 2024.

GAMBHIR, M.; GUPTA, V. Recent automatic text summarization techniques: A survey. **Artificial Intelligence Review**, [S.l.], v.47, n. 1, p. 1-66, 2017. Disponible en: <https://doi.org/10.1007/s10462-016-9475-9>. Acceso en: 28 out. 2024.

GILLETA, M.; GIORDANO, A.; MERCAÚ, N.; ORDEN, P.; VILLARREAL, V. Inteligencia artificial: Definiciones en disputa. **Sociales Investiga**, Córdoba, v. 9, p. 20-33, 2020. Disponible en: <https://socialesinvestiga.unvm.edu.ar/ojs/index.php/socialesinvestiga/article/view/320>. Acceso en: 28 out. 2024.

GALLEGOS, J. C. P. *et al.* (orgs.) **Inteligencia Artificial** [S.l.]: Iniciativa Latinoamericana de Libros de Texto Abiertos (LATIn), 2014. Disponible en: <https://rephip.unr.edu.ar/server/api/core/bitstreams/bb5e5b0c-01b6-482c-a3a4-a469f994c92b/content>. Acceso en: 28 out. 2024.

LANCASTER, F. W. **Indexing and abstracting in theory and practice**. Cambridge: Facet Publishing, 2003.

LIU, Y.; OTT, M.; GOYAL, N.; DU, J.; JOSHI, M.; CHEN, D.; LEVE, O.; LEWIS, M.; ZETTLEMOYER, L.; STOYANOV, V. RoBERTa: A robustly optimized BERT pretraining approach. **arXiv preprint**:1907.11692, 2021. Disponible en: <https://arxiv.org/abs/1907.11692>. Acceso en: 28 out. 2024.

MISKOFF, H. C. **A fondo**: Inteligencia artificial. [S.l.]: Anaya Multimedia, 1988.

ORGANIZACIÓN DE LAS NACIONES UNIDAS PARA LA EDUCACIÓN, LA CIENCIA Y LA CULTURA **Inteligencia artificial en la educación**. Paris: UNESCO, 2023. Disponible en: <https://www.unesco.org/es/digital-education/artificial-intelligence>. Acceso en: 28 out. 2024.

PARK, D. S.; CHAN, W.; ZHANG, Y.; CHIU, C. C.; ZOPH, B.; CUBUK, E. D.; LE, Q. V. SpecAugment: A simple data augmentation method for automatic speech recognition. **arXiv**: 1904.08779, 2019. Disponible en: <https://doi.org/10.48550/arXiv.1904.08779>. Acceso en: 28 out. 2024.

PIHLAJARINNE, T.; ALÉN-SAVIKKO, A. (Eds.). **Artificial intelligence and the media**: Reconsidering rights and responsibilities. London: Edward Elgar Publishing, 2022. Disponible en: <https://www.elgaronline.com/view/edcoll/9781839109966/9781839109966.xml>. Acceso en: 28 out. 2024.

PORIA, S.; CAMBRIA, E.; HAZARIKA, D.; KWOK, K. A review of affective computing: From unimodal analysis to multimodal fusion. **Information Fusion**, [S.l.], v. 37, p. 98-125, 2017. Disponible en: [https://www.researchgate.net/publication/313322315\\_A\\_Review\\_of\\_Affective\\_Computing\\_From\\_Unimodal\\_Analysis\\_to\\_Multimodal\\_Fusion](https://www.researchgate.net/publication/313322315_A_Review_of_Affective_Computing_From_Unimodal_Analysis_to_Multimodal_Fusion). Acceso en: 28 out. 2024.

ROJAS-TORRIJOS, J.-L. Semi-automated journalism: Reinforcing ethics to make the most of artificial intelligence for writing news. In: Luengo, M.; Herrera-Damas, S. (Eds.). **News media innovation reconsidered**: Ethics and values in a creative reconstruction of journalism (pp. 124-137). [S.l.]: Wiley-Blackwell, 2021. Disponible en: <https://doi.org/10.1002/9781119706519.ch8>. Acceso en: 28 out. 2024.

ROUHIAINEN, L. **Inteligencia artificial**: 101 cosas que debes saber hoy sobre nuestro futuro. [S.l.]: Planeta de libros, 2018.

ROWLEY, J.; FARROW, J. **Organizing knowledge**: An introduction to managing access to information. [S.l.]: Routledge, 2017.

SÁNCHEZ-GARCÍA, P.; MERAYO-ÁLVAREZ, N.; CALVO-BARBERO, C.; DÍEZ-GRACIA, A. Desarrollo tecnológico español de la inteligencia artificial aplicada al periodismo: Empresas y herramientas de documentación, producción y distribución de información. **El Profesional de la Información**, [S.l.] v. 32, n. 2, e320208, 2023. Disponible en: <https://doi.org/10.3145/epi.2023.mar.08>. Acceso en: 28 out. 2024.

SIMONYAN, K.; ZISSERMAN, A. (2015). Very deep convolutional networks for large-scale image recognition. **arXiv**:1409.1556, 2015. Disponible en: <https://doi.org/10.48550/arXiv.1409.1556>. Acceso en: 28 out. 2024.



SMITH, R. An overview of the Tesseract OCR engine. *In: INTERNATIONAL CONFERENCE ON DOCUMENT ANALYSIS AND RECOGNITION, 9., 2007, Curitiba. Anales [...]. Curitiba: IEEE, 2007. p. 629-633. Disponível em: [10.1109/ICDAR.2007.4376991](https://doi.org/10.1109/ICDAR.2007.4376991). Acesso em: 28 out. 2024.*

TÚÑEZ-LÓPEZ, J.M.; FIERAS-CEIDE, C.; VAZ-ÁLVAREZ, M. Impact of artificial intelligence on journalism: Transformations in the company, products, contents and professional profile. **Communication & Society**, *Navarra*, v. 34, n. 1, p. 177-193, 2021. Disponível em: <https://doi.org/10.15581/003.34.1.177-193>. Acesso em: 28 out. 2024.

VENTURA-POCINO, P. **Algorithms in the newsrooms**: Challenges and recommendations for artificial intelligence with the ethical values of journalism. Barcelona: Catalan Press Council, 2021.

YEUNG, S.; RASTEGARI, M.; STARK, M.; DARRELL, T. End-to-end learning of action detection from frame glimpses in videos. arXiv:1511.06984, 2016. Disponível em: <http://vision.stanford.edu/pdf/yeung2016cvpr.pdf>. Acesso em: 28 out. 2024.

## NOTAS

### AGRADECIMENTOS

This research was supported by the Junta de Extremadura and FEDER Funds “A way to make Europe” [QUINARI Research Group SEJ013 GR18130].

### CONTRIBUIÇÃO DE AUTORIA

**Concepção e elaboração do manuscrito:** J. Caldera-Serrano

**Coleta de dados:** J. Caldera-Serrano

**Análise de dados:** J. Caldera-Serrano

**Discussão dos resultados:** J. Caldera-Serrano

**Revisão e aprovação:** J. Caldera-Serrano

### CONJUNTO DE DADOS DE PESQUISA

#### FINANCIAMENTO

Não se aplica.

### CONSENTIMENTO DE USO DE IMAGEM

Não se aplica.

### APROVAÇÃO DE COMITÊ DE ÉTICA EM PESQUISA

Não se aplica.

### CONFLITO DE INTERESSES

Não se aplica.

### LICENÇA DE USO

Os autores cedem à **Encontros Bibli** os direitos exclusivos de primeira publicação, com o trabalho simultaneamente licenciado sob a [Licença Creative Commons Attribution \(CC BY\) 4.0](https://creativecommons.org/licenses/by/4.0/) International. Esta licença permite que **terceiros** remixem, adaptem e criem a partir do trabalho publicado, atribuindo o devido crédito de autoria e publicação inicial neste periódico. Os **autores** têm autorização para assumir contratos adicionais separadamente, para distribuição não exclusiva da versão do trabalho publicada neste periódico (ex.: publicar em repositório institucional, em site

pessoal, publicar uma tradução, ou como capítulo de livro), com reconhecimento de autoria e publicação inicial neste periódico.

### **PUBLISHER**

Universidade Federal de Santa Catarina. Programa de Pós-graduação em Ciência da Informação. Publicação no [Portal de Periódicos UFSC](#). As ideias expressadas neste artigo são de responsabilidade de seus autores, não representando, necessariamente, a opinião dos editores ou da universidade.

### **EDITORES**

Edgar Bisset Alvarez, Ana Clara Cândido, Patrícia Neubert, Genilson Geraldo, Jônatas Edison da Silva, Mayara Madeira Trevisol, Edna Karina da Silva Lira e Luan Soares Silva.

### **HISTÓRICO**

Recebido em: 19-08-2024 – Aprovado em: 11-11-2024 – Publicado em: 29-12-2024

