Técnicas Forenses para la Detección de Contenido Generado por Inteligencia Artificial

Dr. Marcelo Temperini

POSGRADO ESPECIALIZACIÓN EN INFORMÁTICA FORENSE

Abril 2025

AsegurarTe
IT Security & Compliance

UNIVERSIDAD FASTA
FACULTAD DE INGENIERÍA

Índice

Resumen

La actual revolución de las tecnologías de la información y las comunicaciones ha impulsado el desarrollo de la inteligencia artificial, permitiendo la generación de contenidos sintéticos de una manera sin precedentes. El alto nivel de disponibilidad y de calidad de estas técnicas de manipulación de contenido (a través de una aplicación o servicios web) permiten que el contenido creado resulte, a los ojos de una persona no entrenada, imposible de distinguir si es real o es sintético, poniendo en riesgo la veracidad de la información y en consecuencia, la confianza jurídica sobre estos potenciales elementos probatorios.

Este contexto plantea un desafío significativo para el ámbito de la informática forense, a fin de poder estandarizar algún procedimiento que permita evaluar los contenidos sospechados de manipulación por inteligencia artificial.

En el presente trabajo, a partir del estudio comparativo de algunas de las herramientas disponibles, se desarrolla una propuesta de una primera guía forense de buenas prácticas que permita analizar casos donde exista sospecha que imágenes o videos que serán tenidos en cuenta como evidencia digital, podrían haber sido manipulados por inteligencia artificial.

Palabras claves:

Inteligencia Artificial; Evidencia Digital, Deep Fake, Técnicas Forenses, contenidos sintéticos.

1. Introducción

En el contexto actual de evolución tecnológica, la generación de imágenes y videos sintéticos mediante inteligencia artificial (IA), como los denominados deepfakes, plantea un desafío significativo para el ámbito forense. La accesibilidad creciente a herramientas avanzadas de IA ha permitido que incluso individuos con conocimientos técnicos limitados puedan crear contenido altamente realista, dificultando la distinción entre una imagen auténtica y una generada artificialmente. Esta situación compromete la veracidad de la información y la confianza en los medios digitales, con implicaciones directas en el análisis de evidencia digital y la administración de justicia.

Este fenómeno se convierte en un riesgo crítico cuando los contenidos sintéticos son aceptados como evidencia digital auténtica, lo que podría afectar la imparcialidad y confiabilidad del sistema judicial. Desde una perspectiva aplicada, este trabajo tiene como objetivo identificar y validar técnicas forenses que permitan detectar y analizar imágenes y videos generados por IA. Se busca realizar una revisión de herramientas y metodologías aplicables en el ámbito pericial, facilitando su uso en la generación de dictámenes forenses, investigaciones judiciales y otros contextos de seguridad digital. Para ello, se analizó patrones característicos presentes en contenido sintético, se evaluó las herramientas tecnológicas disponibles y se diseñará una guía forense replicable para el análisis de este tipo de evidencia digital.

Como parte fundamental del estudio, se llevará a cabo un análisis comparativo mediante la aplicación de pruebas de concepto con diferentes técnicas forenses, con el propósito de

evaluar sus fortalezas y limitaciones. Estas pruebas se realizarán utilizando un dataset propio compuesto por imágenes sintéticas y reales, lo que permitirá generar métricas estadísticas para determinar la eficacia de cada herramienta forense analizada.

La investigación adoptó una metodología mixta (cualitativa y cuantitativa), combinando revisión bibliográfica, experimentación práctica y diseño de buenas prácticas. Una de las principales contribuciones del estudio es la elaboración de una guía forense estandarizada, que brinde a los peritos disponer de un marco metodológico confiable para el análisis de contenido digital sospechoso.

Con este enfoque integral, el proyecto pretende fortalecer las capacidades forenses en la identificación y análisis de contenido sintético, proporcionando herramientas concretas para la protección de la integridad digital. Además, busca generar conciencia sobre el impacto ético y legal del uso de IA en la manipulación de información, promoviendo buenas prácticas en el manejo de evidencia digital.

A través de esta iniciativa, se espera contribuir al ámbito forense, al sistema judicial y a la sociedad en general, dotando a los profesionales de una guía a tener en cuenta para enfrentar los desafíos emergentes de la presentación de contenidos sintéticos en el ámbito informático pericial forense.

2. Objetivos del trabajo

2.1. Objetivo general:

- Desarrollar un guía forense de buenas prácticas que permita analizar casos donde exista sospecha que imágenes y videos que serán tenidos en cuenta como evidencia digital, podrían haber sido manipulado por inteligencia artificial.

2.2. Objetivos específicos:

- Identificar patrones característicos en imágenes y videos generados por inteligencia artificial que permitan diferenciarlos de contenido genuino.
- Identificar y evaluar técnicas y herramientas existentes para la detección de contenido generado por inteligencia artificial, a partir de los datasets previamente creados.
- Evaluar resultados de herramientas tecnológicas actuales para la detección de contenido manipulado mediante inteligencia artificial.
- Generar avances en el campo de las técnicas forenses destinadas a la detección de imágenes y videos modificados por inteligencia artificial.

2.3. Aspectos jurídicos

El Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo[1], conocido como el Reglamento de Inteligencia Artificial de la UE (Ley de IA), aprobado formalmente el 13 de junio

[1] Texto completo disponible en Boletín Oficial del Estado Español: https://www.boe.es/doue/2016/119/L00001-00088.pdf

de 2024 y publicado en el Diario Oficial de la Unión Europea, brinda ciertas reglas que entendemos como importantes para destacar en el objeto de estudio de este trabajo.

1. **Obligación de transparencia en contenidos generados por IA:** De acuerdo al Art. 50 y 52 del Reglamento 2024/1689, se exige que los usuarios de sistemas de IA que generen o manipulen contenidos de imagen, audio o video informen claramente que el contenido ha sido generado artificialmente. Esta obligación impulsa la necesidad de desarrollar (por parte de las empresas dedicadas a la publicación de contenidos) técnicas automáticas y manuales para detectar cuándo una imagen ha sido cerrada o alterada por IA, ya que la obligación de transparencia puede requerir verificación, monitoreo o auditoría. A través de la misma, la norma busca proteger al público frente a manipulación, desinformación o fraudes, lo que convierte la detección de contenidos sintéticos en una obligación derivada del cumplimiento regulatorio.
2. **Protección de derechos fundamentales:** De acuerdo a los Considerando 5 y 30 del Reglamento 2024/1689, se establece que todo desarrollo y uso de IA debe respetar los derechos fundamentales reconocidos por la Carta de Derechos Fundamentales de la UE. Entendemos que la generación de imágenes creadas o alteradas por IA puede atentar contra derechos como la privacidad, el honor, la libertad de expresión y la dignidad humana, por lo que la detección proactiva de contenido manipulado es una medida de cumplimiento con los principios fundacionales de la UE.
3. **Prevención del riesgo de manipulación o desinformación:** De acuerdo a los Considerando 60 y 61 del Reglamento 2024/1689, se reconocen los riesgos asociados al uso de IA para manipular la opinión pública, engañar a usuarios o generar contenidos sintéticos no identificados. En este aspecto, se legitima el desarrollo de tecnologías de detección de imágenes falsas como una medida preventiva y de gestión de riesgos, especialmente para proteger procesos democráticos, electorales y la confianza pública.

2.4. Aplicación práctica del conocimiento

A continuación, se desarrolla brevemente algunas de las aplicaciones prácticas donde podría aplicarse el conocimiento informático forense que proponemos a través del presente trabajo:

1. **Ámbito Judicial:** La posibilidad de detectar contenido visual modificado por IA resulta crítica para garantizar procesos justos y preservar la integridad probatoria. Una imagen, captura de pantalla o video falsificado podría ser presentado como prueba en una causa judicial, poniendo en riesgo la correcta valoración del hecho. La informática forense debería actuar como barrera técnica y científica contra la manipulación, permitiendo validar la autenticidad, integridad y origen temporal de los contenidos digitales, lo cual es fundamental para la tutela efectiva de derechos y la confiabilidad del sistema judicial.
2. **Medios de comunicación:** La manipulación digital permite la creación de noticias falsas altamente verosímiles, capaces de influir en la opinión pública, alterar procesos electorales, dañar reputaciones o manipular mercados. La verificación forense de la autenticidad de imágenes y videos difundidos masivamente se convierte así en una

herramienta indispensable para preservar la integridad del ecosistema informativo, reforzar la confianza social y proteger el derecho ciudadano a una información veraz[2].

3. **Sector Financiero y de Seguros:** La informática forense aplicada a la detección de imágenes generadas o modificadas por IA cumple un rol clave en la prevención del fraude. El uso de fotos alteradas digitalmente para simular daños en vehículos, bienes o viviendas, permite nuevos tipos de estafas que buscan obtener compensaciones indebidas. La capacidad técnica para autenticar la fuente, la integridad y el historial de una imagen permite a las aseguradoras evaluar la veracidad de las reclamaciones con mayor certeza, reducir riesgos económicos y evitar prácticas abusivas que impactan en la prima del resto de los asegurados.
4. **Protección de la privacidad y el honor:** La proliferación de imágenes falsas generadas por IA, como las que involucran desnudos falsos, montajes comprometedores o suplantación de identidad, representa una amenaza directa a la privacidad, el honor y la dignidad de las personas. La informática forense permitiría identificar indicios técnicos de manipulación y rastrear el origen del contenido alterado, lo que resulta crucial para actuar con rapidez frente a ataques a la reputación digital. Esta capacidad técnica fortalece el ejercicio del derecho a la privacidad[3] y a la protección de datos personales (sobre todo a la imagen y voz), pilares esenciales en la era digital para frenar el daño viral y facilitar mecanismos de reparación legal.

3. Tipos de IA generativas

En esta primer parte del trabajo, se aborda una breve descripción de las tecnologías de Inteligencia Artificial generativas más difundidas, y por lo tanto, que más probabilidades tengamos de identificar en el marco de un análisis digital forense.

Entre estas las tecnologías de generación de imágenes y videos sintéticos, podemos encontrar:

3.1. Redes Generativas Antagónicas (GANs)

Las Redes Generativas Antagónicas (GANs) fueron introducidas por Ian Goodfellow (Goodfellow et al. 2014) y han revolucionado la generación de contenido visual sintético. Estas redes están compuestas por dos modelos de redes neuronales que interactúan en un proceso de entrenamiento adversarial. Por un lado, un modelo generador, donde su objetivo es crear datos sintéticos (imágenes, videos, entre otros) que sean indistinguibles de los datos reales. Por otro, el modelo discriminador, que se encarga de evaluar los datos y determinar si son reales o

[2] Constitución Nacional Argentina, Art. 42: *"Los consumidores y usuarios de bienes y servicios tienen derecho, en la relación de consumo, a la protección de su salud, seguridad e intereses económicos; a una información adecuada y veraz; a la libertad de elección, y a condiciones de trato equitativo y digno. [...]"*

[3] Si bien la Constitución Argentina no menciona la palabra "privacidad," sí se refiere a "acciones privadas" en su artículo 19, el cual ha sido interpretado por la Corte Suprema de Argentina como consagrando el derecho a la privacidad. En materia de datos personales, dicha protección es regulada en Argentina por la Ley N° 25.326.

generados artificialmente. A medida que ambos modelos compiten entre sí, el generador mejora su capacidad para producir imágenes cada vez más realistas, intentando engañar al modelo discriminador en un proceso iterativo.

En materia de aplicaciones de este modelo, las GANs han demostrado ser eficaces en numerosos campos. Entre esos campos, uno de nuestro interés para este trabajo es el de generación de imágenes sintéticas. Un ejemplo de un proyecto que aplica esta tecnología es *This Person Does Not Exist* (Karras et al., 2019), que básicamente genera rostros humanos que a simple vista, no es posible determinar si son sintéticos o reales.

Imagen N° 1: Muestra de ejemplo de https://thispersondoesnotexist.com/

Por otro lado, esta tecnología también es conocida por su importancia en la generación de Deepfakes y manipulación de videos, donde las GANs han sido utilizadas en la síntesis de videos en los que se superpone la cara de una persona sobre otra, lo que plantea, en palabras de Yisroel Mirsky y Wenke Lee (2021), nuevos desafíos en el ámbito de la seguridad digital y la detección de contenido falso.

3.2. Modelos de Difusión

Los modelos de difusión han surgido como una alternativa a las Redes Generativas Antagónicas (GANs) en la generación de imágenes y videos sintéticos de alta calidad. Su funcionamiento se basa en un proceso probabilístico de difusión y denoising (eliminación de ruido), lo que les permite generar contenido visual de manera estable y detallada.

La primera formulación de estos modelos fue presentada por Sohl-Dickstein et al. (2015), quienes propusieron un mecanismo en el que los datos se transforman gradualmente en ruido puro a través de un proceso de difusión. Posteriormente, un modelo aprende a invertir este proceso, reconstruyendo los datos originales a partir del ruido, lo que da lugar a la generación

de nuevas muestras sintéticas. Este enfoque ha demostrado ser particularmente útil en la creación de imágenes y videos ultra realistas (Dhariwal & Nichol, 2021).

A diferencia de las GANs, donde el generador y el discriminador compiten entre sí, los modelos de difusión emplean un proceso de ruido gradual que se desarrolla en dos fases:

1. **Difusión directa (Forward process)**: Se añade progresivamente ruido gaussiano a una imagen real hasta que se convierte en ruido puro.
2. **Difusión inversa (Reverse process)**: Se entrena un modelo para aprender a revertir este proceso, eliminando el ruido paso a paso hasta reconstruir una imagen sintética de alta calidad.

El modelo más representativo en esta línea es Stable Diffusion, desarrollado por Rombach (2022), que introduce un enfoque de difusión latente para reducir la carga computacional y mejorar la eficiencia. Gracias a esta optimización, Stable Diffusion se ha convertido en una de las herramientas más utilizadas[4] en la conversión de texto a imagen, permitiendo la generación de contenido visual a partir de descripciones textuales.

3.3. Modelos de transformadores multimodales

Los modelos de transformadores multimodales han emergido como una solución innovadora en inteligencia artificial, permitiendo la integración y generación de contenido a partir de múltiples tipos de datos (texto, imágenes y video). Estos modelos están basados en la arquitectura de transformadores, introducida originalmente en el contexto del procesamiento de lenguaje natural (Vaswani et al., 2017), pero que ha sido adaptada para comprender y generar contenido visual de manera contextualizada.

Uno de los modelos más influyentes en este campo es CLIP[5] (Contrastive Language–Image Pretraining), desarrollado por OpenAI (Radford et al., 2021). CLIP es capaz de comprender relaciones entre texto e imágenes, permitiendo generar contenido visual basado en descripciones textuales. Este avance ha sido fundamental para el desarrollo de herramientas como DALL·E[6], que utiliza transformadores multimodales para la generación de imágenes realistas a partir de instrucciones en lenguaje natural (Ramesh et al., 2022).

Los transformadores multimodales aprenden representaciones conjuntas entre diferentes tipos de datos, utilizando mecanismos de atención cruzada (cross-attention) que permiten a la IA captar relaciones complejas entre texto, imagen y video. En forma resumida, el proceso se desarrolla en los siguientes pasos:

1. **Codificación del texto y la imagen:** Se generan representaciones numéricas (embeddings) para cada entrada (texto, imagen o video).
2. **Mecanismo de atención cruzada:** Se alinean los embeddings, permitiendo que la IA aprenda correlaciones entre distintos tipos de datos.

[4] Estadísticas según https://journal.everypixel.com/ai-image-statistics
[5] Información oficial en OpenIA: https://openai.com/es-ES/index/clip/
[6] Información oficial en OpenIA: https://openai.com/es-419/index/dall-e-3/

3. **Generación de contenido:** A partir de un instrucción concretar por texto (*prompt*), el modelo produce una imagen o video que mantiene coherencia semántica con la instrucción original.

Un ejemplo reciente es Imagen[7], un modelo de Google basado en transformadores multimodales que ha demostrado una capacidad avanzada en la síntesis de imágenes realistas a partir de texto, superando en calidad a modelos previos como DALL·E.

3.4. Redes Neuronales Convolucionales (CNN)

Las Redes Neuronales Convolucionales (CNNs, por sus siglas en inglés) son un tipo de arquitectura de inteligencia artificial diseñada específicamente para el procesamiento y análisis de imágenes. Introducidas por LeCun et al. (1998), las CNNs han sido la base del reconocimiento visual moderno y han evolucionado hacia arquitecturas avanzadas capaces de generar, modificar y mejorar imágenes y videos realistas.

A diferencia de las Redes Neuronales Artificiales (ANNs) tradicionales, las CNNs utilizan capas convolucionales que detectan patrones espaciales en los datos visuales, permitiendo la extracción jerárquica de características, como bordes, texturas y formas. Este enfoque ha hecho que las CNNs sean la base de muchos modelos de generación de imágenes, incluidas las GANs y los modelos de difusión (Simonyan & Zisserman, 2015).

Las CNNs funcionan a través de una serie de capas especializadas, cada una con un papel clave en el procesamiento de imágenes. De forma resumida, estas capas son:

1. **Capa Convolucional:** Extrae características esenciales de la imagen a través de filtros que detectan bordes, texturas y formas.
2. **Capa de Pooling (Submuestreo):** Reduce la dimensionalidad de los datos para mejorar la eficiencia y reducir la redundancia en la información.
3. **Capas de Normalización y Activación:** Permiten mejorar la estabilidad y la velocidad del entrenamiento de la red.
4. **Capas Densas o Fully Connected:** Procesan las características extraídas para la generación o clasificación de imágenes.

El entrenamiento de CNNs avanzadas implica el uso de grandes volúmenes de datos, donde la red aprende a identificar patrones y estructuras de imágenes reales, lo que posteriormente permite generar contenido visual coherente.

3.5. Redes Generativas de Video

Las Redes Generativas de Video son un área emergente dentro de la inteligencia artificial, enfocada en la generación y síntesis de secuencias de video realistas. A diferencia de la generación de imágenes estáticas, la creación de videos requiere mantener coherencia temporal entre múltiples fotogramas, lo que supone un desafío técnico significativo.

En los últimos años, han surgido modelos avanzados que combinan Redes Generativas Antagónicas (GANs), Modelos de Difusión, Transformadores Temporales y Redes Neuronales

[7]Disponible en línea: https://cloud.google.com/vertex-ai/generative-ai/docs/image/overview?hl=es-419

Recurrentes (RNNs) para mejorar la calidad y continuidad de los videos generados. Ejemplos destacados incluyen MoCoGAN[8], VideoGPT[9] y el reciente Sora[10] de OpenAI (Clark et al., 2024).

Los modelos de redes generativas de video deben resolver múltiples problemas, como la generación de movimientos fluidos, la preservación de detalles visuales y la consistencia entre diferentes fotogramas. Para lograrlo, se utilizan diversas arquitecturas:

1. **Modelos basados en GANs**
    - Redes como MoCoGAN (Tulyakov et al., 2018) utilizan una separación entre contenido y movimiento, lo que permite generar secuencias de video con coherencia temporal.
    - StyleGAN-V (Skorokhodov et al., 2022) aplica la arquitectura de StyleGAN en la generación de videos cortos de alta fidelidad.
2. **Modelos de Difusión para Videos**
    - Imagen Video (Ho et al., 2022) y Sora de OpenAI utilizan modelos de difusión condicionados a la información de cuadros anteriores, asegurando una progresión fluida de los videos.
3. **Modelos de Transformadores Temporales**
    - VideoGPT (Yan et al., 2021) aplica transformadores auto-regresivos para predecir secuencias de video, mejorando la consistencia de movimiento.
    - TimeSformer (Bertasius et al., 2021) introduce mecanismos de atención espacial y temporal para mejorar la generación de videos de larga duración.

4. Construcción de datasets de prueba

A fin de llevar adelante las pruebas forenses establecidas entre los objetivos, avanzaremos con la generación de distintos conjuntos (dataset) de imágenes.

Para ello, hemos generado 6 (seis) dataset de 10 (diez) imágenes cada uno, que a posteriori serán analizadas con las distintas técnicas y herramientas -siempre que sea posible-, a fin de poder comprobar la eficacia en la detección. La cantidad de datasets, así como de imágenes que serán incluídas en cada uno, responder a una limitación técnica y temporal para la realización del mismo, estimando la cantidad de pruebas que aplicamos.

4.1. Creación de datasets de imágenes reales

Para la construcción del primer dataset de referencia, se procedió a seleccionar 10 (diez) imágenes reales, extraídas de la galería de fotos de un dispositivo celular (Iphone 13 PRO) de propiedad del autor del estudio. Se ha considerado apropiado utilizar imágenes propias a fin de garantizar que las imágenes efectivamente son reales, y a la vez, garantizando que las imágenes son extraídas del dispositivo original que ha capturado la imagen, favoreciendo la existencia de

[8] MoCoGAN: Decomposing Motion and Content for Video Generation; Disponible en: https://arxiv.org/abs/1707.04993
[9] https://videogpt.io/
[10] Información oficial de OpenIA en https://openai.com/es-ES/sora/

metadatos reales (simulando una potencial extracción forense que podría tenerse en el marco de una pericia informática real). Las imágenes reales se acompañan en el ANEXO I del presente estudio.

4.2. Creación de datasets de imágenes generadas con inteligencia artificial

Metodológicamente, su generación fue realizada de la siguiente manera:

1) Se procedió a utilizar una herramienta denominada "ImagePrompt"[11], cuya función es detectar lo que existe sobre una imagen y generar un prompt (instrucción por texto) de forma detallada y precisa, para que luego podamos utilizarlos para generar nuevas imágenes sintéticas que logren una imagen similar.

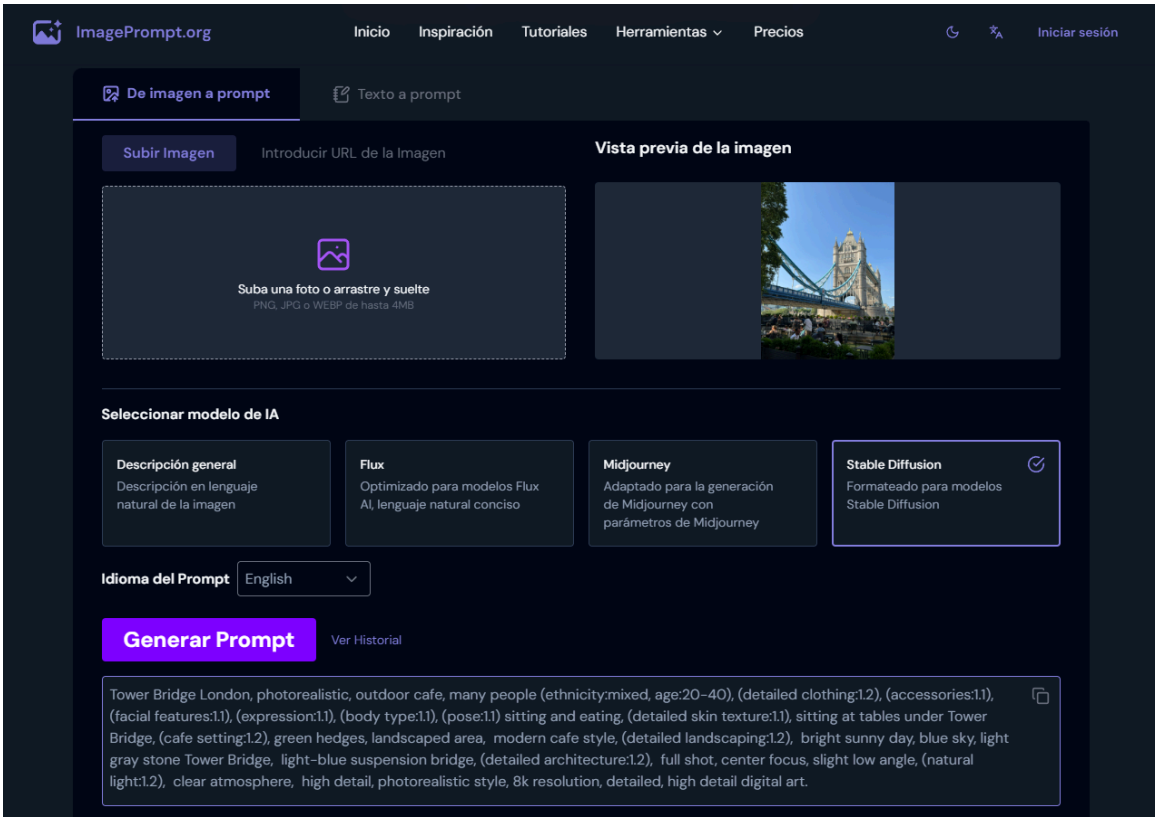


Imagen N° 2. Herramienta utilizada para la generación de prompts profesionales a partir de imágenes.

2) A través de dicha herramienta, se ingresaron las 10 (diez) imágenes reales, generándose por cada una el prompt correspondiente. Los *prompts* generados se pueden observar en el ANEXO I del presente estudio.

[11] ImagePrompt es una herramienta disponible en: https://imageprompt.org/es/about-us

3) Los *prompts* generados fueron introducidos en cuatro herramientas de generación de imágenes con inteligencia artificial: Stable Difussion[12], Dall-e[13], Grok[14] y Fooocus[15], generando así 10 (diez) imágenes sintéticas por cada una de ellas.

Se deja constancia que si bien al momento de la generación del trabajo, Midjourney[16] también es una de las herramientas más populares de generación de imágenes sintéticas, ha sido excluída del estudio toda vez que no es posible su utilización de forma gratuita, por lo que se han priorizado otras tecnologías de generación de imágenes sintéticas que sean opensource, o al menos, permitan un uso gratuito (en algunos casos limitado a determinada cantidad de tokens diarios, como en el caso de Dall-e).

A continuación, generamos una tabla comparativa, resumiendo los datasets generados, incluyendo el tipo de licencia utilizada, versión, forma de ejecución y formato de salida de las imágenes generadas.

Dataset N°	Generado a partir de	Versión y características	Licencia	Ejecución	Formato
1	Imágenes reales	No aplica	No aplica	No aplica	jpg
2	Stable Difussion	Version: v1.10.1 Checkpoint: 463d6a9fe8	Creative ML OpenRAIL-M	Local	png
3	Dall-e	Versión: 3 Propietario: Microsoft	Propietaria Microsoft	En línea a través de https://www.bing.com/images/create	jpeg
4	Grok	Versión: 3 Propietario: xAI	Propietaria xAI	En línea a través de www.grok.com	jpg
5	Fooocus	Versión: 2.5.5 Gradio: 3.41.2	GNU GPL	Local	png
6	Imágenes híbridas	Versión: 2.5.5 Gradio: 3.41.2 Herramienta: INPAINT	GNU GPL	Local	png

Tabla Nº 1: Comparativa de tecnologías de generación de contenido sintético

4) Generadas las imágenes, se guardaron en carpetas diferentes, indicándose a qué IA generativa corresponden, y con nombres secuenciales (imagen1, imagen2, etc.). Los formatos de guardado, han sido aquellos formatos originales de exportación de cada herramienta.

En al ANEXO I se pueden observar los datasets de imágenes generados, organizadas de acuerdo al tipo de tecnología de inteligencia artificial generativa utilizada.

[12] Modelo generativo de imágenes de Stability difusión, disponible en: https://stability.ai/stable-image
[13] Modelo generativo de imágenes de OpenIA, disponible en: https://openai.com/es-419/index/dall-e-3/
[14] Modelo generativo de imágenes de xAI Corp, disponible en: github.com/xai-org/grok-1
[15] Modelo generativo de imágenes, basado en Stable Difusión XL: https://github.com/lllyasviel/Fooocus
[16] Modelo generativo de imágenes de Midjourney, Inc., disponible en: https://www.midjourney.com/

4.3. Generación de dataset extra con imágenes reales modificadas por IA

Adicionalmente incorporamos un sexto dataset que, a diferencia de los dataset anteriores (que eran 100% generados sintéticamente), tienen una combinación basada en mayor parte por una imagen real, pero que fueron modificados por algún modelo de inteligencia artificial. Entendemos que su incorporación es apropiada para ampliar el marco del estudio de las herramientas y técnicas alcanzadas, permitiendo analizar su comportamiento ante este tipo de imágenes.

Para la generación de este dataset denominado como "híbrido", se ha partido de las imágenes reales (Dataset N°1), sobre las que se han realizado algún tipo de modificación utilizando inteligencia artificial. Para realizar la alteración sobre la imagen, se utilizó "Fooocus" (utilizada para la generación del Dataset N° 5), a través de la técnica de INPAINT.

El "inpainting" es una técnica de restauración de imágenes que utiliza la inteligencia artificial para rellenar áreas dañadas o faltantes. De esta forma, el usuario pinta la parte de la imágen que desea retocar, y a través de un prompt, la inteligencia artificial intenta modificar esa sección, teniendo en consideración el contexto de la imagen. Sin embargo, el inpainting no sólo puede corregir imperfecciones, sino que puede eliminar objetos no deseados e incluso crear elementos u objetos que no existen en una fotografía, por lo que consideramos de gran interés para el objeto del presente estudio.

Todas las imágenes generadas podrán ser visualizadas en el ANEXO I del presente estudio.

5. Técnicas y Herramientas disponibles para la detección de contenido sintético

En ciencias forenses, una herramienta debe cumplir con principios de validez científica y reproducibilidad. Por eso, en cada prueba de herramienta o técnica, realizamos un cálculo básico de tasa de error, considerando en el resultado obtenido, si la herramienta acertó o no en definición acerca de si el contenido es o no sintético, de acuerdo a las siguientes consideraciones:

- **Interpretación estadística:** En muchos casos las herramientas manejan variables estadísticas, por lo que a los fines del presente trabajo, si al analizar una imagen la herramienta indica (a modo de ejemplo) que dicho contenido en un 60% fue realizado con IA, e indica un restante de 40% de autenticidad, interpretaremos que la herramienta (por mayoría) indica que la imagen ha sido generada sintéticamente.
- **Categorización del resultado:** El resultado podrá ser categorizado como verdadero positivo (VP), falso positivo (FP), falso negativo (FN) o verdadero negativo (VN), de acuerdo al siguiente criterio:

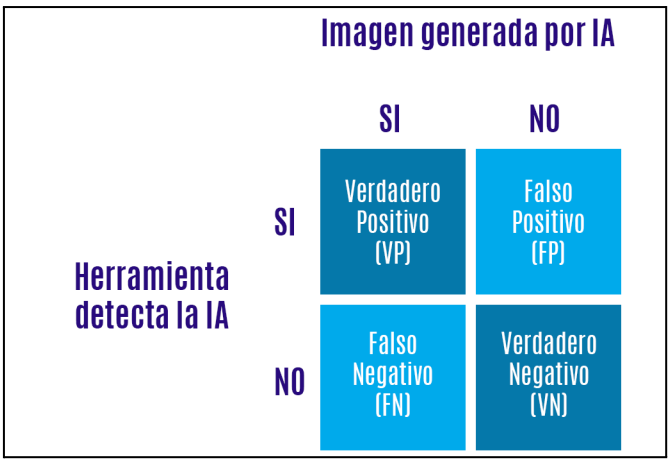


Imagen N° 3. Cuadro de categorización de resultados

- **Cálculo de la tasa de error:** Basado en estos resultados de las pruebas, calcularemos la exactitud de cada herramienta, con la fórmula siguiente:
    - Exactitud (%)= E = ((VP + VN) / (VP + VN + FN + FP)) x 100
    - Tasa de error (%) = (1 - E)
- **Tasa de error sugerida por normas internacionales:** Según la National Institute of Standards and Technology (NIST) y la Scientific Working Group on Digital Evidence (SWGDE), una herramienta de análisis forense debe demostrar un nivel de precisión y confiabilidad que minimice el riesgo de falsos positivos y falsos negativos (NIST, 2021). De acuerdo a estos organismos, una tasa de error aceptable para herramientas utilizadas en investigaciones forenses, deberían tener un error inferior al 5% (equivalente a una precisión del 95%).
- **Precisión sobre IA generativa (Índice PIAG):** Con algunas de las técnicas o herramientas utilizadas, se podrá determinar con qué modelo de IA generativa de imágenes se realizó el contenido. Este índice se calcula basado en primer lugar, en los aciertos de la herramienta (la herramienta debe primero haber acertado si se trata de un contenido sintético), y dentro de ese acierto, se evaluará en cuantos casos la herramienta acertó en la predicción sobre cuál es el modelo de IA generativa utilizada para el contenido analizado.
- **Criterio de interpretación para la determinación de eficacia en el Dataset N° 6:** En estos casos, considerando que muchas de las herramientas utilizadas determinan sus resultados a nivel estadísticos (en porcentajes %), aplicaremos un criterio interpretativo amplio, considerando como “acierto” los casos donde la herramienta utilizada detecte que el componente de IA detectado, oscila entre un 5 y un 40%. La justificación para dicho criterio, parte de considerar que las modificaciones realizadas (de acuerdo a ANEXO I), en todos los casos modifica menos del 30% de la imagen completa. Es decir, de forma aproximada, un 70% de la imagen se conserva real, mientras sólo un 30% sería contenido sintético.

5.1. Herramienta N° 1: EXIFTool

Los metadatos son información incrustada en los archivos de imagen y video, que incluye detalles sobre la cámara utilizada, la ubicación GPS, la fecha y hora de captura, y otros datos clave (Gioia, C. V., 2017).

Dentro de un primer bloque de análisis forense, encontramos que la revisión de metadatos y huellas digitales es una técnica fundamental en informática forense para la verificación de la autenticidad de imágenes, y por lo tanto, podría ser de utilidad para el caso objeto del estudio.

A fin de analizar los metadatos de las imágenes, utilizamos EXIFTool (ExifTool.org by Phil Harvey). En relación a la licencia, es software libre; que puede distribuirse y/o modificarse bajo los mismos términos legales establecidos en la Licencia de PERL (https://dev.perl.org/licenses/).

ExifTool proporciona una biblioteca que permite (conjunto de módulos en Perl) leer y escribir metainformación en una amplia variedad de archivos de imágenes, audio, vídeo y documentos. Es una herramienta con muchos años de desarrollo, muy utilizada para realizar análisis y extracción de metadatos sobre imágenes.

5.1.1. Técnica o Modelo utilizado

El examen de metadatos EXIF es una técnica fundamental en informática forense para la autenticación de imágenes digitales. Los archivos de imagen contienen metadatos EXIF (Exchangeable Image File Format), los cuales registran información detallada sobre la captura, incluyendo la marca y modelo de la cámara, la resolución, la configuración de exposición, la fecha y hora de toma, e incluso la ubicación GPS si está habilitada. Estos datos permiten verificar la procedencia de una imagen y detectar posibles manipulaciones, ya que cada dispositivo de captura deja una huella digital única en los metadatos del archivo.

De acuerdo a algunos estudios (Fan et al., 2020), los deepfakes y otros contenidos sintéticos generados por inteligencia artificial suelen carecer de metadatos consistentes, lo que los hace detectables mediante herramientas de análisis forense. En el caso de imágenes generadas por inteligencia artificial, los metadatos EXIF suelen estar incompletos, modificados o completamente ausentes, ya que los modelos generativos no capturan imágenes del mundo real, sino que las crean a partir de patrones estadísticos. Por supuesto, debemos considerar que si accedemos a la imagen de interés a través de fuentes como redes sociales, este tipo de técnica no pueda ser de utilidad, ya que su procesamiento previo elimina muchos de los metadatos de interés.

No obstante, en el caso que sea posible acceder a la imagen en su formato original, es posible analizar sus metadatos intentando detectar inconsistencias, como fechas de creación incoherentes, ausencia de datos de la cámara o ediciones sospechosas, o incluso, rastros del prompt utilizado, puede ser un indicador clave de que una imagen ha sido generada sintéticamente o manipulada digitalmente.

5.1.2. Pruebas con la herramienta

En relación a la herramienta, a fin de llevar adelante las pruebas, la misma fue descargada y utilizada de forma local, con el comando simple de “exiftool.exe -lang es imagen”

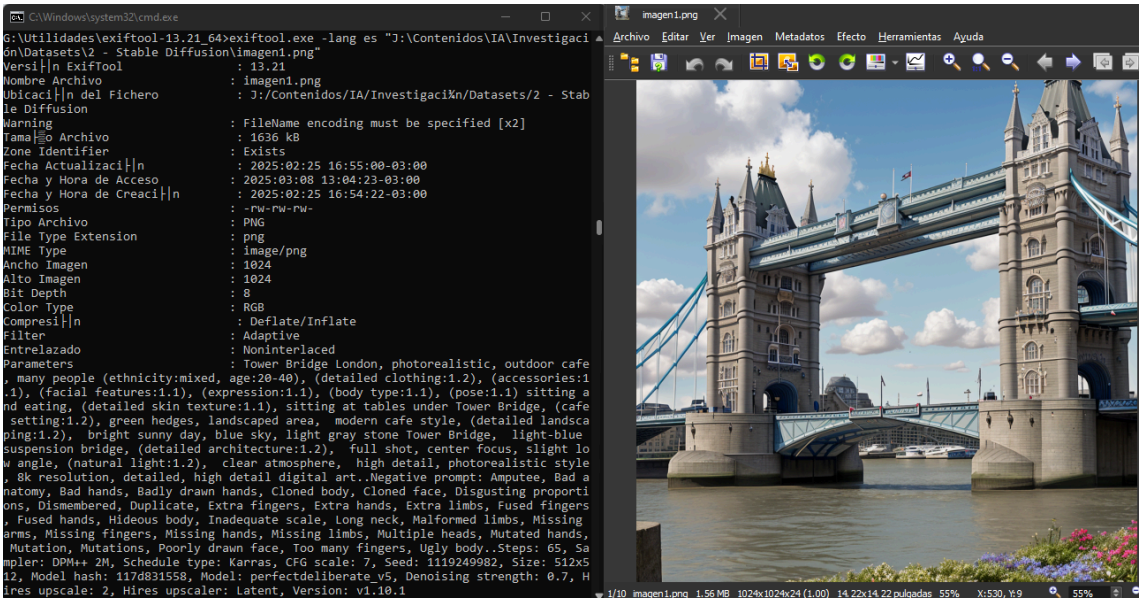


Imagen N° 4: Prueba positiva realizada con una imagen generada desde Stable Difussion

Imagen N° 5: Prueba positiva realizada con una imagen generada desde GROK

A continuación exponemos la tabla resumen de los resultados obtenidos, aplicando dicha herramienta por cada uno de los datasets de prueba:

Dataset	Conclusión	Resultado
N°1 - Reales	Se observan datos completos, incluyendo dispositivo de captura, coordenadas gps, entre otros datos.	VP

N°2 - Stable Diffusion	Se encontró información relevante en el campo “parámetros”, con información del prompt positivo y el negativo (en caso de existir).	VP
N°3 - Dall-e	No se encontró información relevante	-
N°4 - Grok	Se encontró información relevante en el campo “parámetros”, con información del prompt positivo y el negativo (en caso de existir).	VP
N°5 - Fooocus	No se encontró información relevante	-
N°6 - Híbridas	No se encontró información relevante	-

Tabla Nº 2: Comparativa de resultados obtenidos por análisis de metadatos

Por el tipo de resultado técnica utilizada, consideramos que no es posible realizar un cálculo sobre la tasa de error de la técnica, ya que sólo en dos casos ha sido posible considerar la existencia de un VP (dataset real, stable diffusion y grok). En el resto de los modelos, al no encontrarse información relevante, no permitiría concluir si la imagen es o no sintética, por lo que no consideramos apropiado categorizarlo como un FP o VN. Para el caso de estudio, podemos afirmar que la herramienta ha sido eficaz en un 50%, considerando que ha logrado brindar resultados positivos en 3 de 6 casos.

5.2. Herramienta N° 2: Susy

Herramienta desarrollada en Python por parte del equipo del Centro de Supercomputación de Barcelona (Bernabeu-Pérez et al., 2024), producto de un trabajo de investigación[17] en el cuál se ha realizado un análisis sistemático sobre modelos de detección de contenido sintéticos existentes, para desarrollar pautas prácticas para entrenar detectores de imágenes sintéticos robustos.

5.2.1. Técnica o Modelo utilizado

SuSy está basado en un modelo basado en una red neuronal convolucional de reconocimiento y detección de imágenes sintéticas basadas en el espacio, que ha sido diseñado y entrenado para detectar imágenes sintéticas y atribuirlas a un modelo generativo (es decir, dos modelos StableDiffusion, dos versiones Midjourney y DALL·E 3). El modelo toma parches de imagen de tamaño 224x224 como entrada y genera la probabilidad de que la imagen sea auténtica o haya sido creada por cada uno de los modelos generativos antes mencionados.

Las capacidades de generalización del modelo se evalúan en diferentes configuraciones (por ejemplo: escala, fuentes, transformaciones, etc), incluidas las condiciones de implementación en el mundo real. El estudio muestra cómo, los niveles de detección de diferentes modelos existentes, varían de acuerdo al escalamiento, tipos de modelos, incluso, si se trata de imágenes de paisajes o de personas. Alentamos a los lectores a realizar una revisión del estudio citado. De acuerdo al estudio previamente citado, el modelo publicado se basa en una arquitectura CNN y se entrena utilizando un enfoque de aprendizaje supervisado. Su diseño se basa en trabajos anteriores, originalmente pensados para la detección de superresolución de

[17] Pablo Bernabeu-Perez, Enrique Lopez-Cuena, Dario Garcia-Gasulla; “Present and Future Generalization of Synthetic Image Detectors”: Disponible en: https://arxiv.org/abs/2409.14128

vídeo, adaptados aquí para las tareas de detección y reconocimiento de imágenes sintéticas. La arquitectura consta de dos módulos: un extractor de características y un perceptrón multicapa (MLP), según se observa en el siguiente diagrama.

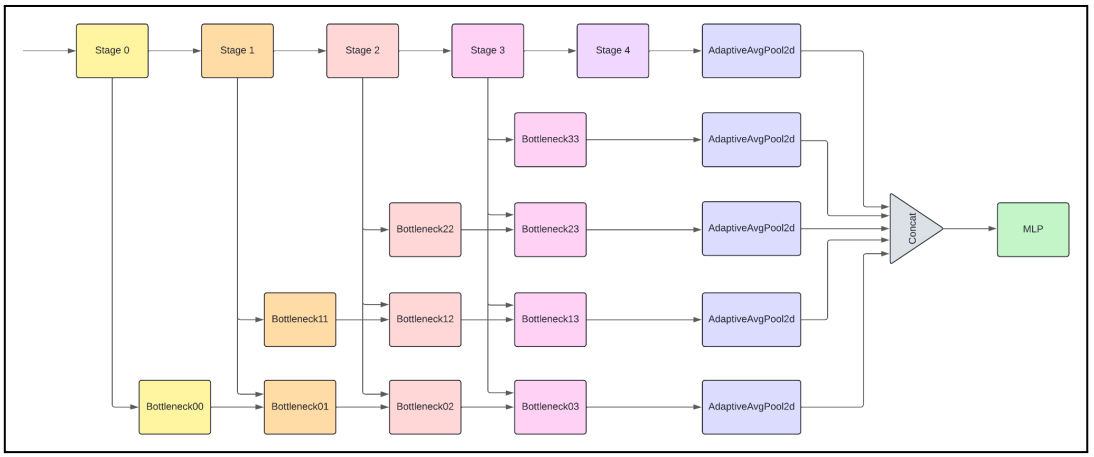


Imagen 6: Diagrama de flujo del modelo utilizado por la herramienta Susy (Fuente: https://huggingface.co/HPAI-BSC/SuSy)

El extractor de características CNN consta de cinco etapas que siguen un esquema de red neuoranl residual de 18 capas (ResNet-18). La salida de cada uno de los bloques se utiliza como entrada para varios módulos de cuello de botella que están dispuestos en un patrón de escalera. Los módulos de cuello de botella constan de tres capas convolucionales 2D. Cada nivel de cuello de botella toma la entrada en una etapa posterior al nivel anterior, y cada módulo de cuello de botella toma la entrada de la etapa actual y, excepto el primer cuello de botella de cada nivel, del módulo de cuello de botella anterior.

5.2.2. Pruebas con la herramienta

La herramienta “Susy” analizada es opensource, basada en los términos y condiciones legales de la licencia de Apache 2.0. y su código se encuentra disponible en github, junto con los modelos y los datasets utilizados, a fin de poder testearlo localmente.

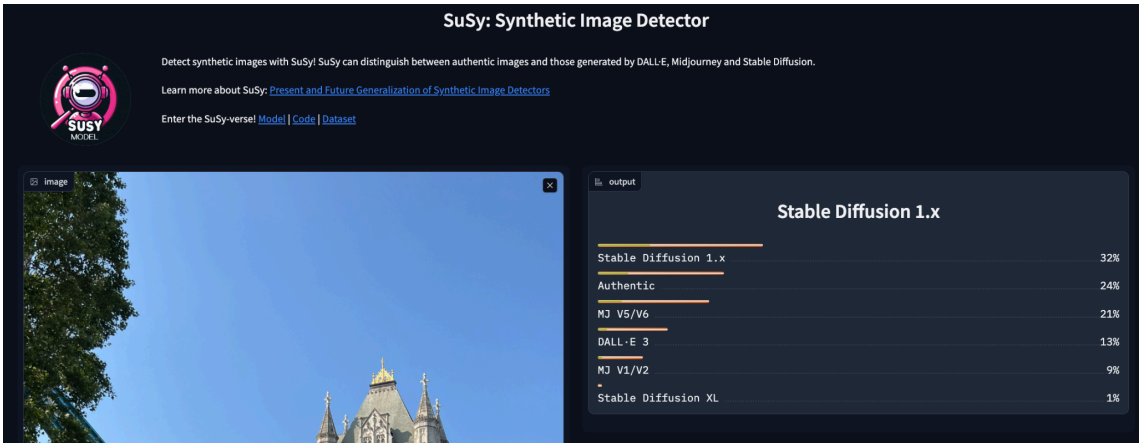


Imagen 7: Captura de pantalla de Herramienta Susy, analizando la imagen01 del Dataset N°1

Sus autores han puesto a disposición la herramienta de forma online[18], desde donde se ha realizado las pruebas de nuestro trabajo. A continuación, exhibimos una captura de pantalla con el primero de los resultados obtenidos a forma de ejemplo y posteriormente, las tablas que ilustran los resultados de todas las pruebas:

5.2.2.1. Pruebas Dataset Nº 1: Imágenes reales

Dataset	Nº de Imagen	Resultado de Susy obtenidos (en %)						Conclusión	Resultado
		SD 1.x	Authentic	MJ V5/V6	DALL-E 3	MJ V1/V2	SD XL		
Dataset Nº 1 - Imágenes Reales	1	32	24	21	13	9	1	Stable Diffusion 1.x	FP
	2	43	1	44	11	1	0	MJ V5/V6	FP
	3	4	0	48	2	45	0	MJ V5/V6	FP
	4	2	1	75	3	20	0	MJ V5/V6	FP
	5	23	3	66	2	6	0	MJ V5/V6	FP
	6	18	57	0	0	24	0	**Authentic**	**VP**
	7	18	27	15	40	1	0	DALL-E 3	FP
	8	2	61	28	9	0	0	**Authentic**	**VP**
	9	21	0	73	4	0	1	MJ V5/V6	FP
	10	14	1	27	16	42	1	MJ V1/V2	FP

Tabla Nº 3: Resultados obtenidos con herramienta Susy sobre Dataset Nº 1 - Imágenes Reales

5.2.2.2. Pruebas Dataset Nº 2: Stable Diffusion

Dataset	Nº de Imagen	Resultado de Susy obtenidos (en %)						Conclusión	Resultado
		SD 1.x	Authentic	MJ V5/V6	DALL-E 3	MJ V1/V2	SD XL		
Dataset Nº 2 - Stable Diffusion	1	3	77	0	0	0	20	Authentic	FN
	2	22	78	0	0	0	0	Authentic	FN
	3	78	0	0	0	8	13	**Stable Diffusion 1.x**	**VP**
	4	53	41	1	3	0	1	**Stable Diffusion 1.x**	**VP**
	5	14	46	0	0	0	39	Authentic	FN
	6	4	79	3	14	0	0	Authentic	FN

[18] Servicio montado sobre la infraestructura de HuggingFace, disponible en: https://huggingface.co/spaces/HPAI-BSC/SuSy

	7	8	54	2	35	0	1	Authentic	FN
	8	3	77	19	1	1	0	Authentic	FN
	9	41	57	0	1	0	0	Authentic	FN
	10	36	63	0	0	0	0	Authentic	FN

Tabla Nº 4: Resultados obtenidos con herramienta Susy sobre Dataset Nº 2 - Stable Diffusion

5.2.2.3. Pruebas Dataset Nº 3: Dall-e

Dataset	Nº de Imagen	Resultado de Susy obtenidos (en %)						Conclusión	Resultado
		SD 1.x	Authentic	MJ V5/V6	DALL-E 3	MJ V1/V2	SD XL		
Dataset Nº 3 - Dall-e	1	1	81	0	17	0	0	Authentic	FN
	2	24	3	0	73	0	0	**DALL·E 3**	**VP**
	3	0	1	1	95	0	3	**DALL·E 3**	**VP**
	4	0	12	50	38	0	0	MJ V5/V6	**VP**
	5	44	1	4	51	0	0	**DALL·E 3**	**VP**
	6	36	46	1	17	0	0	Authentic	FN
	7	23	3	21	53	0	0	**DALL·E 3**	**VP**
	8	6	17	1	75	0	0	**DALL·E 3**	**VP**
	9	29	19	1	43	0	8	**DALL·E 3**	**VP**
	10	19	45	2	33	0	1	Authentic	FN

Tabla Nº 5: Resultados obtenidos con herramienta Susy sobre Dataset Nº 3 - Dall-e

5.2.2.4. Pruebas Dataset Nº 4: Grok

Dataset	Nº de Imagen	Resultado de Susy obtenidos (en %)						Conclusión	Resultado
		SD 1.x	Authentic	MJ V5/V6	DALL-E 3	MJ V1/V2	SD XL		
Dataset Nº 4- Grok	1	17	43	0	30	0	9	Authentic	FN
	2	28	68	0	5	0	0	Authentic	FN
	3	12	0	76	12	0	0	MJ V5/V6	**VP**
	4	1	57	22	20	0	0	Authentic	FN
	5	31	37	17	15	0	0	Authentic	FN

	6	0	97	0	2	0	0	Authentic	FN
	7	3	25	1	70	0	0	DALL·E 3	**VP**
	8	72	27	0	1	0	0	Stable Diffusion 1.x	**VP**
	9	40	0	5	55	0	0	DALL·E 3	**VP**
	10	0	92	0	8	0	0	Authentic	FN

Tabla Nº 6: Resultados obtenidos con herramienta Susy sobre Dataset Nº 4 - Grok

5.2.2.5. Pruebas Dataset Nº 5: Fooocus

Dataset	Nº de Imagen	Resultado de Susy obtenidos (en %)						Conclusión	Resultado
		SD 1.x	Authentic	MJ V5/V6	DALL-E 3	MJ V1/V2	SD XL		
Dataset Nº 5-Fooocus	1	6	63	3	27	0	0	Authentic	FN
	2	14	40	7	36	2	0	Authentic	FN
	3	4	6	38	26	3	23	MJ V5/V6	**VP**
	4	40	15	35	9	1	1	Stable Diffusion 1.x	**VP**
	5	2	6	37	44	1	10	DALL·E 3	**VP**
	6	4	27	6	62	0	0	DALL·E 3	**VP**
	7	15	12	2	70	0	0	DALL·E 3	**VP**
	8	0	76	16	7	0	0	Authentic	FN
	9	25	25	4	43	0	2	DALL·E 3	**VP**
	10	6	84	2	5	0	2	Authentic	FN

Tabla Nº 7: Resultados obtenidos con herramienta Susy sobre Dataset Nº 5 - Fooocus

5.2.2.6. Pruebas Dataset Nº 6: Híbridas

Dataset	Nº de Imagen	Resultado de Susy obtenidos (en %)						Conclusión	Resultado
		SD 1.x	Authentic	MJ V5/V6	DALL-E 3	MJ V1/V2	SD XL		
Dataset Nº 6-Híbridas	1	10	17	26	2	44	1	MJ V1/V2	**VP**
	2	43	1	43	11	1	0	MJ V5/V6	**VP**
	3	4	0	48	2	45	0	MJ V5/V6	**VP**
	4	2	1	75	3	20	0	MJ V5/V6	**VP**
	5	23	3	66	2	6	0	MJ V5/V6	**VP**

	6	18	57	0	0	24	0	Authentic	FN
	7	18	27	15	40	1	0	DALL-E 3	**VP**
	8	2	61	28	9	0	0	Authentic	FN
	9	21	0	73	4	0	0	MJ V5/V6	**VP**
	10	14	1	27	16	42	1	MJ V1/V2	**VP**

Tabla Nº 8: Resultados obtenidos con herramienta Susy sobre Dataset Nº 6 - Híbridas

5.2.3. Análisis de resultados obtenidos

Las pruebas sobre el primer dataset, donde las imágenes eran reales (Tabla Nº 3), los resultados han sido muy poco confiables, ya que de 10 imágenes reales, sólo en 2 casos (imagen 6 y 8) la herramienta acertó, indicando que era auténtica con un 57% de probabilidades. Lo llamativo es que en la mitad de las pruebas, el índice de autenticidad, estuvo por debajo del 3% (imagen 3, 4, 5, 9 y 10), evidenciando serios problemas en la detección de imágenes reales.

Las pruebas sobre el segundo dataset, donde las imágenes fueron generadas con Stable Difussion (Tabla Nº 4), los resultados han sido muy poco confiables, toda vez que de 10 imágenes sintéticas, sólo en 2 casos (imagen 3 y 4) la herramienta acertó, indicando que eran sintéticas y realizadas con Stable Diffusion. Sin embargo, las restantes 8 imágenes, según la herramienta fueron identificadas como imágenes auténticas, incluso con niveles bastante más altos que los que dieron con las verdaderas imágenes reales (a modo de ejemplo, la imagen 8 dió un 77% de autenticidad). Incluso en imágenes hasta cuya generación sintética son muy obvios (como puede verse en la imagen 10, donde pueden observarse varios detalles), ha fallado. Los resultados se consideran muy poco confiables como para ser utilizados en el marco de una prueba forense.

Las pruebas sobre el tercer dataset, donde las imágenes fueron generadas con Dall-e (Tabla Nº 5), los resultados han mejorado bastante en relación a los anteriores. La herramienta acertó en 6 de 10 imágenes, indicando no sólo que eran sintéticas sino que además, la herramienta generativa fue Dall-e. De las restantes 4, la herramienta acertó en que era sintética en la imagen 4, pero indicando que la IA generativa había sido MJ V5/V6, aunque Dall-e si estuvo en segundo lugar a nivel porcentual, con un 38%. En la imagen 10, si bien la herramienta dictaminó que era auténtica, los resultados obtenidos son más equilibrados, quedando también Dall-e en ese caso, como la segunda opción con un 33%. Los errores más grandes se obtuvieron en la imagen 1, donde se dictaminó que en un 81% la imagen era auténtica. En cambio, en la imagen 6, también fue equilibrado, dando autenticidad por un 46%, pero teniendo en segundo lugar a Stable Diffusion con un 36% y a Dall-e con un 17%.

Las pruebas sobre el cuarto dataset, de imágenes generadas con Grok (Tabla Nº 6), los resultados volvieron a la media normal de este trabajo, con muchos errores. De las 10 imágenes generadas por GROK, la herramienta dió como auténticas (con niveles de incluso hasta un 97%) a 6 de ellas. De las 4 restantes, hubo confirmación de que eran imágenes sintéticas, pero se dividió entre si fueron generadas por Dall-e o por MJ o Stable Diffusion.

Las pruebas sobre el quinto dataset, de imágenes generadas con Fooocus (Tabla Nº 7), los resultados volvieron a la media normal de este trabajo, con muchos errores. De las 10 imágenes generadas por Fooocus, la herramienta dió como auténticas (con niveles de incluso hasta un 84%) a 4 de ellas. De las 6 restantes, hubo confirmación de que eran imágenes sintéticas, detectando en su mayoría que habían sido generadas por DALL-E 3. Recordemos que Fooocus, está construído integramente sobre la arquitectura de una de las ramas de Stable Diffusion, por lo que teniendo un criterio interpretativo amplio, hemos considerado que la predicción de la imagen 4, es la única correcta.

Las pruebas sobre el último dataset, de imágenes híbridas (Tabla Nº 8), ha sido quizás la más complicada para medir, toda vez que las imágenes son en su origen reales, pero que tienen, en alguna parte, alguna modificación realizada sintéticamente por IA. De las 10 imágenes, se ha detectado que 2 de ellas son auténticas, y las restantes 8 han sido generadas sintéticamente. Es interesante observar una comparativa entre la Tabla Nº 3, donde la única diferencia que podemos notar, se encuentra en la imagen Nº 1, donde en este dataset ha dado como sintético de MJ V1/V2 (con un 44%), y sin embargo, en la Tabla Nº 3 de imágenes reales, había dado como sintético, con una mayoría de stable diffusion (32%).

5.2.4. Tasa de Error de la herramienta

A continuación, realizamos un repaso por las tasas de error calculadas a nivel individual por las distintas pruebas, calculando un promedio de todas ellas. Dicho cálculo fue calculado por separado considerando primero los datasets de 1 a 5, considerando que en ellas se ha evaluado determinar si la imagen ha sido o no generada sintéticamente y en su caso, si la herramienta lo permite, indicar con qué modelo o herramienta de IA ha sido generada.

Dataset N°	Generado a partir de	Exactitud	Tasa de Error	Índice PIAG
1	Imágenes reales	20%	80%	No aplica
2	Stable Difussion	20%	80%	100%
3	Dall-e	60%	40%	85%
4	Grok	60%	40%	0%
5	Fooocus	60%	40%	16%
Promedio calculado		44%	**56%**	**50,25%**

Tabla Nº 9: Resultados obtenidos con Herramienta Susy para la identificación de imágenes reales o generadas por inteligencia artificial.

Por separado presentamos los resultados obtenidos para intentar determinar si una imagen real, ha sido o no modificada por IA (dataset 6 - imágenes híbridas).

Dataset N°	Generado a partir de	Exactitud	Tasa de Error	Índice PIAG
6	Imágenes híbridas	80%	20%	0%

Tabla Nº 10: Resultados obtenidos con Herramienta Susy para la identificación de imágenes reales modificadas con inteligencia artificial.

5.3. Herramienta N° 3: Hive IA Detector

La tercer herramienta que analizamos se llama “Hive Moderation”, y es una herramienta comercial, que ofrece una variedad de servicios, muchos de ellos relacionados con el tema de moderación automática de contenidos (una necesidad en materia de redes sociales). Entre los distintos servicios de moderación, han incorporado uno dedicado exclusivamente a la detección de contenidos generados por inteligencia artificial.

5.3.1. Técnica o Modelo utilizado

De acuerdo a la documentación oficial, la API de detección de imágenes y videos generados por IA de Hive es un único punto final que ejecuta dos modelos independientes: uno para detectar imágenes generadas por un motor de IA como Midjourney, DALL-E o Firefly y otro para detectar deepfakes o imágenes en las que se ha utilizado IA para mapear el rostro de una persona con el de otra. Su resultado combinado se devuelve como una lista de clases.

Dada una imagen o un vídeo de entrada, el modelo de detección de imágenes y vídeos generados por IA de Hive determina si la entrada es completamente generada por IA. Según la documentación de la empresa, el modelo se entrenó con un amplio conjunto de datos compuesto por millones de imágenes generadas artificialmente y por personas, como fotografías, arte digital y tradicional, y memes provenientes de la web.

El modelo de detección de imágenes y vídeos generados por IA tiene dos aspectos:

- **Clasificación de generación:** ai_generated, not_ai_generated
- **Clasificación de fuente:** sora, pika, haiper, kling, luma, hedra, runway, hailuo, mochi, flux, hallo, hunyuan, recraft, leonardo, luminagpt, var, liveportrait, mcnet, pyramidflows, sadtalker, aniportrait, cogvideos, makeittalk, sdxlinpaint, stablediffusioninpaint, bingimagecreator, adobefirefly, lcm, dalle, pixart, glide, stablediffusion, imagen, amused, stablecascade, midjourney, deepfloyd, gan, stablediffusionxl, vqdiffusion, kandinsky, wuerstchen, titan, ideogram, sana, emu3, omnigen, other_image_generators (generador de imágenes distinto a los indicados), inconcluso, inconcluso_video (no se ha identificado ninguna fuente de vídeo) o ninguno (los medios no están generados por IA). Los puntajes de confianza para cada modelo suman 1.

Si bien la herramienta es comercial, brinda la posibilidad de utilizar una versión gratuita en línea, que se puede usar tanto desde el sitio web, así como a través de una extensión que se instala en el navegador Google Chrome. Según la documentación de la herramienta, Hive promete una precisión del 98%, con una tasa de error en falsos negativos del 3%.

5.3.2. Pruebas con la herramienta

A continuación, exhibimos una captura de pantalla con el primero de los resultados obtenidos a forma de ejemplo y posteriormente, las tablas que ilustran los resultados de todas las pruebas:

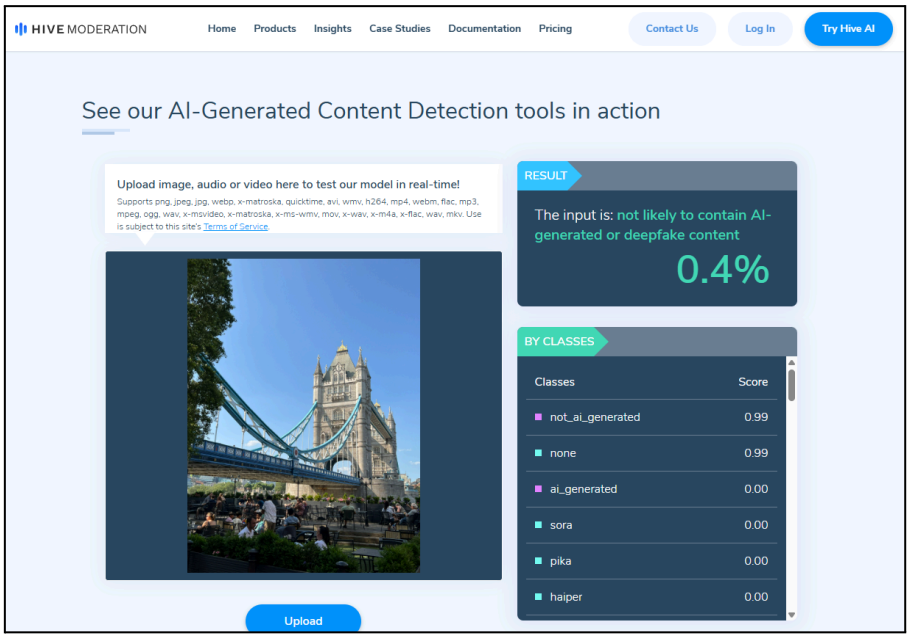


Imagen 8: Captura de pantalla de Hive

5.3.2.1. Pruebas Dataset Nº 1: Imágenes reales

Dataset evaluado	Nº de Imagen	Resultado obtenido		Conclusión	Resultado
		not_ai_generated	ai_generated		
Dataset Nº 1 - Imágenes Reales	1	99%	0,4%	99% Auténtica	VN
	2	99%	0,3%	99% Auténtica	VN
	3	99%	0,5%	99% Auténtica	VN
	4	99%	0,1%	99% Auténtica	VN
	5	99%	0%	99% Auténtica	VN
	6	99%	0,1%	99% Auténtica	VN
	7	99%	0,4%	99% Auténtica	VN
	8	99%	0,4%	99% Auténtica	VN
	9	98%	1,9%	98% Auténtica	VN
	10	99%	0%	99% Auténtica	VN

Tabla Nº 11: Resultados obtenidos con herramienta Hive sobre Dataset Nº 1 - Imágenes Reales

5.3.2.2. Pruebas Dataset Nº 2: Stable Diffusion

Dataset evaluado	Nº de Imagen	Resultado obtenido		Conclusión	Resultado
		not_ai_generated	ai_generated		
Dataset Nº 2 - Stable Diffusion	1	0%	99%	94% SD	VP
	2	0%	99%	58% MJ 41% SD	VP
	3	0%	99%	99% MJ	VP
	4	0%	99%	99% SD	VP
	5	0%	99%	93% SD	VP
	6	0%	99%	77% SD	VP
	7	0%	99%	82% MJ	VP
	8	0%	99%	64% MJ	VP
	9	0%	99%	96% SD	VP
	10	0%	99%	72% SDXL	VP

Tabla Nº 12: Resultados obtenidos con herramienta Hive sobre Dataset Nº 2 - Stable Diffusion -

Nota del autor : Los resultados marcados en color rojo indican que la herramienta falló en la detección. En verde, implican que la herramienta acertó.

5.3.2.3. Pruebas Dataset Nº 3: Dall-e

Dataset evaluado	Nº de Imagen	Resultado obtenido		Conclusión	Resultado
		not_ai_generated	ai_generated		
Dataset Nº 3 - DAll-e	1	0%	99,9%	98% Dall-e	VP
	2	0%	99,8%	99% Dall-e	VP
	3	0%	99,6%	99% Dall-e	VP
	4	0%	99,9%	99% Dall-e	VP
	5	0%	99,6%	99% Dall-e	VP
	6	0%	99,9%	99% Dall-e	VP
	7	0%	99,9%	99% Dall-e	VP

	8	0%	99,4%	88% Dall-e	**VP**
	9	0%	99,8%	95% Dall-e	**VP**
	10	0%	99,9%	99% Dall-e	**VP**

Tabla Nº 13: Resultados obtenidos con herramienta Hive sobre Dataset Nº 3 - Dall-e

5.3.2.4. Pruebas Dataset Nº 4: Grok

Dataset evaluado	Nº de Imagen	Resultado obtenido		Conclusión	Resultado
		not_ai_generated	ai_generated		
Dataset Nº 4 - Grok	1	0,01%	98,6%	93% Stable Diffusion	**VP**
	2	0%	99,9%	94% Stable Diffusion	**VP**
	3	0%	99,1%	97% Stable Diffusion	**VP**
	4	0%	99,9%	65% Stable Diffusion	**VP**
	5	0%	99,4%	97% Stable Diffusion	**VP**
	6	0%	99,9%	65% SDXL Inpaint	**VP**
	7	0%	99,6%	65% Stable Diffusion	**VP**
	8	0%	99,9%	84% Stable Diffusion	**VP**
	9	0%	99,8%	95% Stable Diffusion	**VP**
	10	0%	99%	82% Stable Diffusion	**VP**

Tabla Nº 14: Resultados obtenidos con herramienta Hive sobre Dataset Nº 4 - Grok

5.3.2.5. Pruebas Dataset Nº 5: Fooocus

Dataset evaluado	Nº de Imagen	Resultado obtenido		Conclusión	Resultado
		not_ai_generated	ai_generated		
Dataset Nº 5 - Fooocus	1	0%	99,9%	85% Stable Diffusion	**VP**
	2	0%	99,9%	95% Stable Diffusion	**VP**
	3	0%	99,8%	91% Stable Diffusion	**VP**
	4	0%	99,9%	99% Stable Diffusion	**VP**
	5	0%	99,9%	95% Stable Diffusion	**VP**
	6	0%	99,9%	71% Stable Diffusion	**VP**
	7	0%	99,9%	88% Stable Diffusion	**VP**

	8	0%	99,8%	54% Stable Diffusion	**VP**
	9	0%	99,9%	67% Stable Diffusion	**VP**
	10	0%	99,7%	92% Stable Diffusion	**VP**

Tabla Nº 15: Resultados obtenidos con herramienta Hive sobre Dataset Nº 5 - Fooocus

5.3.2.6. Pruebas Dataset Nº 6: Híbridas

Dataset evaluado	Nº de Imagen	Resultado obtenido		Conclusión	Resultado
		not_ai_generated	ai_generated		
Dataset Nº 6 - Híbrido	1	99%	0,1%	99% Auténtica	**FN**
	2	99%	0%	99% Auténtica	**FN**
	3	99%	1%	99% Auténtica	**FN**
	4	99%	0,3%	99% Auténtica	**FN**
	5	100%	0%	100% Auténtica	**FN**
	6	100%	0,3%	100% Auténtica	**FN**
	7	100%	0,5%	100% Auténtica	**FN**
	8	100%	0,2%	100% Auténtica	**FN**
	9	95%	4,8%	95% Auténtica (4% Stable Diffusion)	**FN**
	10	100%	0%	100% Auténtica	**FN**

Tabla Nº 16: Resultados obtenidos con herramienta Hive sobre Dataset Nº 6 - Híbrido

5.3.3. Análisis de resultados obtenidos

Las pruebas sobre el primer dataset, donde las imágenes fueron reales (Tabla Nº 11), los resultados son correctos, acertando en las 10 imágenes con un 99% de precisión en todos los casos.

Las pruebas sobre el segundo dataset (Tabla Nº 12), utilizando las imágenes sintéticas de stable diffusion (SD), los resultados continúan siendo correctos, acertando en las 10 imágenes con un 99% de precisión en todos los casos, en la detección si la imágen es o no sintética. También ha tenido buenos resultados detectando la fuente de generación de la imagen sintética, observándose que sólo en 4 de los 10 casos, ha detectado MidJourney por encima de Stable Diffusion.

Las pruebas sobre el tercer dataset (Tabla Nº 13), utilizando las imágenes sintéticas de Dall-e, los resultados continúan siendo correctos, acertando en las 10 imágenes con un 99% de precisión en todos los casos, en la detección si la imágen es o no sintética. También se observan excelentes resultados detectando la fuente de generación de la imagen sintética,

observándose que en los 10 casos detectó que fueron generados con Dall-e, teniendo en 9 de los 10 casos, una precisión mayor al 98%. El punto de precisión más bajo fue en la imagen 8, donde detectó Dall-e con un 88% de precisión.

Las pruebas sobre el cuarto dataset (Tabla Nº 14), utilizando las imágenes sintéticas de Grok, los resultados continúan siendo correctos, acertando en las 10 imágenes con un promedio de más del 99% de precisión en la detección si la imágen es o no sintética, excepto en un sólo caso (imagen 1) que descendio al 98,6%,. En relación a los resultados sobre el tipo de IA generativa utilizado, debemos advertir que el software utilizado no tiene a GROK entre los modelos listados (probablemente por ser uno de los últimos en llegar). En 9 de los 10 casos, detectó que las imágenes fueron originadas por SD, excepto en la imagen 6, donde detecto que fue con SDXL con Inpaint.

Las pruebas sobre el quinto dataset (Tabla Nº 15), utilizando las imágenes sintéticas generadas por Fooocus, los resultados continúan siendo muy alentadores, acertando en las 10 imágenes con un promedio de más del 99% de precisión en la detección si la imágen es o no sintética. En relación a los resultados sobre el tipo de IA generativa utilizado, en todos los casos detecto que fueron realizadas por Stable Diffusion, lo que técnicamente sería correcto ya que como hemos visto anteriormente, Fooocus es una rama de Stable Diffusion. No obstante, los índices, a diferencia por ejemplo de las imágenes que realmente fueron realizadas en SD, fueron índices mucho más bajos, llegando en el peor caso al 54% de precisión (imagen 8).

Las pruebas sobre el sexto dataset (Tabla Nº 16), utilizando las imágenes híbridas (imágenes reales con modificaciones realizadas por IA), los resultados obtenidos han indicando que las imágenes son entre 99 y 100% auténticas. Comparativamente con la Tabla Nº 9 (imágenes 100% reales), donde allí todas las imágenes dieron resultados entre un 98 y 99%. Aquí en cambio, donde todas las imágenes fueron modificadas por IA, los resultados en algunas imágenes como la 5,6, 7 y 8, dieron un resultado de 100% de autenticidad, cuando en los casos de las imágenes reales, dieron un 99%.

5.3.4. Tasa de Error de la herramienta

A continuación, realizaremos un repaso por las tasas de error calculadas a nivel individual por las distintas pruebas, calculando un promedio de todas ellas:

Dataset N°	Generado a partir de	Exactitud	Tasa de Error	Índice PIAG
1	Imágenes reales	100%	0%	No aplica
2	Stable Difussion	100%	0%	60%
3	Dall-e	100%	0%	100%
4	Grok	100%	0%	0%
5	Fooocus	100%	0%	100%
Promedio calculado		100%	**0%**	**65%**

Tabla Nº 17: Resultados obtenidos con Herramienta HIVE para la identificación de imágenes reales o generadas por inteligencia artificial.

De acuerdo a los cálculos realizados en los datasets donde se ha buscado determinar si la imagen fue o no generada sintéticamente (Datasets de 1 a 5), HIVE ha acertado en todos los casos, logrando obtener el puntaje ideal de una tasa de error promedio del 0%. No obstante, en el segundo nivel de análisis ha mostrado un índice del 65% de acierto al intentar determinar el índice PIAG.

Aún dentro de ese 65% de índice PIAG, los niveles de acierto con respecto a la IA generativa han sido excelentes (100%) en los casos de Dall-e y Fooocus, pero su promedio se ha visto afectado de forma importante por los malos resultados en GROK (una de las IA más moderna) y teniendo un resultado mediocre en los casos de Stable Difussion.

Dataset N°	Generado a partir de	Exactitud	Tasa de Error	Índice PIAG
6	Imágenes híbridas	0%	100%	100%

Tabla Nº 18: Resultados obtenidos con Herramienta HIVE para la identificación de imágenes reales modificadas con inteligencia artificial.

5.4. SightEngine

SightEngine es otra herramienta comercial, que al igual que HIVE, en sus orígenes ofrece servicios dedicados a la moderación de contenido (detectado desnudez, odio, violencia, drogas, armas, autolesiones, etc.). En ese abanico de servicios, han incorporado una sección de moderación que permite la detección de contenido generado sintéticamente y un servicio dedicado a la detección de deepfakes.

Según un estudio realizado entre la Universidad de Rochester y la Universidad de Kansas (Li et al, 2024), donde se emplearon casi 80 mil imágenes para probar la precisión de los modelos utilizados, SightEngine alcanzó 98.3% de precisión en la detección de contenido sintético.

5.4.1. Técnica o Modelo utilizado

De acuerdo a la información oficial, este modelo se entrenó con millones de imágenes, tanto artificiales como humanas, que abarcan todo tipo de contenido, como fotografía, arte, dibujos, memes y más. Según la empresa detrás de la herramienta, el modelo funciona analizando el contenido visual (píxeles) de la imagen. Afirman textualmente que “No se utilizan metadatos en el análisis. Por lo tanto, la manipulación de metadatos, como los datos EXIF, no afecta la puntuación”, aspecto que hemos señalado a tener en cuenta en la sección 5.2.3 del presente estudio.

Al momento de la realización de las pruebas, el modelo fue entrenado para detectar imágenes generadas por los principales modelos actualmente en uso: Stable Diffusion, Stable Diffusion XL, MidJourney, Dall-E, Adobe Firefly, Flux, Recraft, GANs.

La herramienta tiene la posibilidad de utilizar una demostración de forma manual (subiendo las imágenes de forma individual) o bien, se da la posibilidad de utilizarlo vía API, obteniendo un JSON con la respuesta sobre el análisis realizado.

5.4.2. Pruebas con la herramienta

A continuación, exhibimos una captura de pantalla con el primero de los resultados obtenidos a forma de ejemplo y posteriormente, las tablas que ilustran los resultados de todas las pruebas:

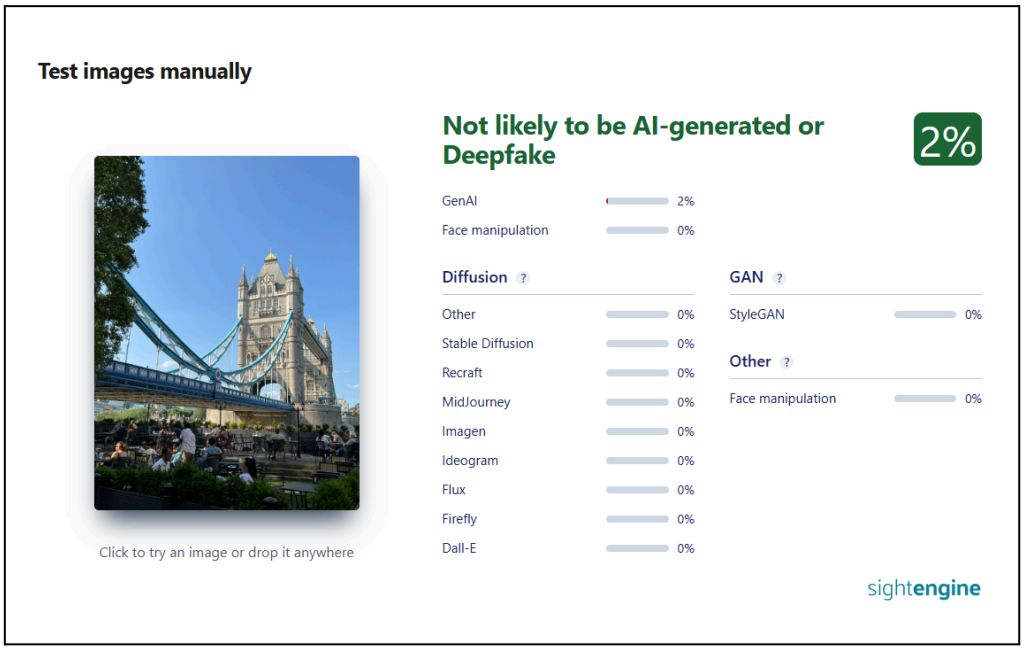


Imagen 9: Captura de pantalla de la herramienta

5.4.2.1. Pruebas Dataset Nº 1: Imágenes reales

Dataset evaluado	Nº de Imagen	Resultado obtenido (GenAI)	Conclusión	Resultado
Dataset Nº 1 - Imágenes Reales	1	2%	0%	VN
	2	7%	6% Imagen	VN
	3	3%	2% MidJourney	VN
	4	1%	1% Flux	VN
	5	1%	0%	VN
	6	1%	1% Flux	VN
	7	1%	0%	VN
	8	1%	0%	VN
	9	1%	0%	VN
	10	1%	0%	VN

Tabla Nº 19: Resultados obtenidos con SightEngine sobre Dataset Nº 1 - Imágenes Reales

5.4.2.2. Pruebas Dataset Nº 2: Stable Diffusion

Dataset evaluado	Nº de Imagen	Resultado obtenido (GenAI)	Conclusión	Resultado
Dataset Nº 2 - Stable Diffusion	1	99%	97% Imagen	**VP**
	2	99%	91% Stable Diffusion	**VP**
	3	99%	92% Stable Diffusion	**VP**
	4	99%	98% Stable Diffusion	**VP**
	5	99%	97% Stable Diffusion	**VP**
	6	99%	99% Stable Diffusion	**VP**
	7	99%	97% Stable Diffusion	**VP**
	8	99%	87% Stable Diffusion	**VP**
	9	99%	99% Stable Diffusion	**VP**
	10	99%	98% Stable Diffusion	**VP**

Tabla Nº 20: Resultados obtenidos con SightEngine sobre Dataset Nº 2 - Stable Diffusion

5.4.2.3. Pruebas Dataset Nº 3: Dall-e

Dataset evaluado	Nº de Imagen	Resultado obtenido (GenAI)	Conclusión	Resultado
Dataset Nº 3 - DAll-e	1	99%	98% Dall-E	**VP**
	2	99%	98% Dall-E	**VP**
	3	99%	96% Dall-E	**VP**
	4	99%	97% Dall-E	**VP**
	5	99%	98% Dall-E	**VP**
	6	99%	98% Dall-E	**VP**
	7	99%	96% Dall-E	**VP**
	8	99%	91% Dall-E	**VP**
	9	99%	90% Dall-E	**VP**
	10	99%	96% Dall-E	**VP**

Tabla Nº 21: Resultados obtenidos con SightEngine sobre Dataset Nº 3 - Dall-e

5.4.2.4. Pruebas Dataset Nº 4: Grok

Dataset evaluado	Nº de Imagen	Resultado obtenido (GenAI)	Conclusión	Resultado
Dataset Nº 4 - Grok	1	99%	97% Imagen	VP
	2	99%	60% Imagen	VP
	3	99%	94% Stable Diffusion	VP
	4	99%	95% Imagen	VP
	5	99%	96% Imagen	VP
	6	99%	94% Imagen	VP
	7	99%	86% Imagen	VP
	8	99%	99% Imagen	VP
	9	99%	99% Imagen	VP
	10	99%	48% Stable Diffusion	VP

Tabla Nº 22: Resultados obtenidos con SightEngine sobre Dataset Nº 4 - Grok

5.4.2.5. Pruebas Dataset Nº 5: Fooocus

Dataset evaluado	Nº de Imagen	Resultado obtenido (GenAI)	Conclusión	Resultado
Dataset Nº 5 - Fooocus	1	99%	99% Stable Diffusion	VP
	2	99%	98% Stable Diffusion	VP
	3	99%	98% Stable Diffusion	VP
	4	99%	99% Stable Diffusion	VP
	5	99%	99% Stable Diffusion	VP
	6	99%	98% Stable Diffusion	VP
	7	99%	99% Stable Diffusion	VP
	8	99%	98% Stable Diffusion	VP
	9	99%	99% Stable Diffusion	VP
	10	99%	98% Stable Diffusion	VP

Tabla Nº 23: Resultados obtenidos con SightEngine sobre Dataset Nº 5 - Fooocus

5.4.2.6. Pruebas Dataset Nº 6: Híbridas

Dataset evaluado	Nº de Imagen	Resultado obtenido (GenAI)	Conclusión	Resultado
Dataset Nº 6- Híbridas	1	2%	0%	**FN**
	2	1%	0%	**FN**
	3	4%	2% MJ - 1% SD	**VP**
	4	2%	0%	**FN**
	5	1%	0%	**FN**
	6	1%	0%	**FN**
	7	2%	0%	**FN**
	8	1%	0%	**FN**
	9	1%	0%	**FN**
	10	1%	0%	**FN**

Tabla Nº 24: Resultados obtenidos con SightEngine sobre Dataset Nº 6- Híbridas

5.4.3. Análisis de resultados obtenidos

Las pruebas sobre el primer dataset (Tabla Nº 19), utilizando las imágenes reales, los resultados son correctos, acertando en las 10 imágenes, con verdadero negativo (VN).

Las pruebas sobre el segundo dataset (Tabla Nº 20), utilizando las imágenes sintéticas de stable diffusion (SD), los resultados continúan siendo correctos, acertando en las 10 imágenes con un 99% de precisión en todos los casos, en la detección si la imágen es o no sintética. También ha tenido buenos resultados detectando la fuente de generación de la imagen sintética, observándose que sólo en 1 de los 10 casos, ha detectado una fuente de IA generativa distinta a la verdadera.

Las pruebas sobre el tercer dataset (Tabla Nº 21), los resultados continúan siendo correctos, acertando en las 10 imágenes con un 99% de precisión en todos los casos, en la detección si la imágen es o no sintética. En el segundo nivel de análisis, también los resultados son excelentes, ya que en los 10 casos, acertó con respecto a la IA generativa, con un mínimo de 90% de precisión.

Las pruebas sobre el cuarto dataset (Tabla Nº 22), los resultados continúan siendo correctos en el primer nivel de análisis, acertando en las 10 imágenes con un 99% de precisión en todos los casos, en la detección si la imágen es o no sintética. En el segundo nivel de análisis, los resultados han sido muy negativos. A modo de comentario, y como hemos mencionado antes, al momento de la realización del estudio, GROK es una de las IA generativas más novedosas, por lo que muchas de las herramientas aún no han incorporado a sus modelos las características propias.

Las pruebas sobre el quinto dataset (Tabla Nº 23), los resultados continúan siendo correctos en el primer nivel de análisis, acertando en las 10 imágenes con un 99% de precisión en todos los casos, en la detección si la imágen es o no sintética. En el segundo nivel de análisis, los resultados han sido muy positivos ya que se ha detectado que en todos los casos la fuente es Stable Diffusion. Recordemos que Fooocus, está construído integramente sobre la arquitectura de una de las ramas de Stable Diffusion, por lo que teniendo un criterio interpretativo amplio, hemos considerado que la predicción realizada por la herramienta son correctas.

Las pruebas sobre el sexto dataset con las imágenes híbridas (reales modificadas por IA - Tabla Nº 24), los resultados se han afectado significativamente, considerando solamente un caso positivo en la imagen 3, donde la herramienta (a diferencia de la imagen 3 de la Tabla Nº 19), se ha incrementado en un 1% de Stable Diffusion, lo que podríamos interpretar como un acierto por parte de la herramienta. En el resto de las imágenes evaluadas, se ha considerado que no ha detectado la modificación de la IA.

Marcando otra diferencia teniendo en cuenta el análisis sobre el dataset de imágenes reales notamos dos observaciones. Por un lado, la imagen Nº 2 que en el caso real, se detectó un 7% de IA y sin embargo, en el caso híbrido, se ha bajado en un 1%. Por otro lado, en la imagen Nº 7, en el caso real se detectó un 1% de IA, y en el caso de la híbrida se detecto un 2%, pero sin identificar ningún tipo de IA generativa en particular.

5.4.4. Tasa de Error de la herramienta

A continuación, realizaremos un repaso por las tasas de error calculadas a nivel individual por las distintas pruebas, calculando un promedio de todas ellas:

Dataset N°	Generado a partir de	Reales vs. Sintéticas		
		Exactitud	Tasa de Error	Índice PIAG
1	Imágenes reales	100%	0%	No aplica
2	Stable Difussion	100%	0%	90%
3	Dall-e	100%	0%	100%
4	Grok	100%	0%	0%
5	Fooocus	100%	0%	100%
Promedio calculado		100%	**0%**	**72,5%**

Tabla Nº 25: Tabla de Error de Herramienta SightEngine

De acuerdo a los cálculos realizados en nuestro primer nivel de análisis (determinar si las imágenes han sido generadas o no sintéticamente), su nivel de exactitud ha sido excelente, del 100% de exactitud y con 0% de tasa de error.

Adicionalmente, debemos agregar que dentro del marco de los resultados acertados, los niveles de acierto con respecto a la IA generativa (índice PIAG), han sido excelentes en los casos

de Stable Diffusion (90%), Dall-e (100%), Focus (100%) y nuevamente han fracasado estrepitosamente con GROK (al igual que sucedió con la anterior herramienta).

Dataset N°	Generado a partir de	Exactitud	Tasa de Error	Índice PIAG
6	Imágenes híbridas	10%	90%	78%

Tabla Nº 26: Resultados obtenidos con Herramienta SightEngine para la identificación de imágenes reales modificadas con inteligencia artificial.

5.5. AMPED AUTHENTICATE

La última herramienta que analizamos ha sido AMPED AUTHENTICATE, de la empresa AMPED FIVE, empresa reconocida por su software comercial que permite el mejoramiento de imágenes. Considerando que la herramienta no se encuentra abiertamente disponible para la raelización de pruebas, nos hemos puesto en contacto con la empresa, explicando la finalidad de la presente investigacióny desde AMPED han tenido la generosidad de apoyar la investigación otorgando un trial de 15 días de su herramienta AUTHENTICATE, a fin de que pueda ser evaluada.

Al instalar la herramienta, se brinda la posibilidad de ejecutar Authenticate Video o Authenticate Image. Considerando el objeto del presente trabajo, utilizaremos Authenticate Image para realizar las pruebas con el mismo dataset de imágenes generados anteriormente.

5.5.1. Técnica o Modelo utilizado

AUTHENTICATE posee internamente una gran cantidad de herramientas que podrán ser utilizadas por un especialista informático forense a fin de determinar la autenticidad de los contenidos analizados. En particular señalamos sólo algunas de las técnicas que hemos considerado como más aplicables a nuestro estudio:

1) Error Level Analysis: ELA es una técnica forense utilizada para detectar áreas alteradas en una imagen comprimida, especialmente en formatos con compresión con pérdida como JPEG. Su fundamento se basa en el principio de que una imagen que no ha sido manipulada debería presentar una distribución uniforme de errores de compresión en todos sus sectores. Para ello, ELA toma una imagen, la vuelve a guardar con un nivel fijo de compresión y luego calcula la diferencia (el “error”) entre la imagen original y la recomprimida. Las zonas con diferencias anómalas pueden indicar modificaciones. Sin embargo, ELA no prueba una manipulación por sí sola. Puede dar falsos positivos si diferentes regiones de la imagen han sido comprimidas de manera desigual, como ocurre con muchas cámaras digitales que aplican algoritmos de compresión adaptativa. Por eso, ELA debe usarse como indicador preliminar en combinación con otras técnicas como DCT Map o análisis de metadatos, y siempre acompañado de criterio técnico y experiencia. (Farid, 2009)

2) DCT Map: La técnica DCT Map se basa en el análisis del espacio de la transformada discreta del coseno (DCT), una operación matemática utilizada en la compresión JPEG para dividir la imagen en bloques de 8x8 píxeles y representar cada bloque como una suma de frecuencias. Durante esta compresión, los valores DCT se cuantifican, y este proceso deja patrones

característicos que pueden ser analizados para identificar inconsistencias. Un DCT Map permite visualizar la distribución de estos coeficientes de frecuencia a lo largo de toda la imagen. Cuando una imagen ha sido manipulada (por ejemplo, copiando y pegando elementos de otro origen), los patrones de cuantización de los bloques afectados pueden diferir notablemente del resto. Estas discrepancias se traducen en regiones anómalas en el mapa DCT, ayudando a detectar modificaciones encubiertas que no son fácilmente visibles a simple vista. (Popescu et al., 2005)

3) JPEG Ghost Map: Técnica forense avanzada que busca detectar manipulaciones en imágenes JPEG mediante la creación de "fantasmas" de compresión. Su principio se basa en que si una parte de la imagen fue editada y recomprimida con diferentes parámetros, al volver a guardar la imagen entera bajo una compresión controlada, las áreas editadas responderán de manera diferente que las originales. Al comparar visualmente las diferencias entre ambas versiones se genera un "ghost map", donde las alteraciones aparecen como zonas contrastadas. El proceso implica volver a comprimir la imagen a distintos niveles de calidad y luego restar estas nuevas versiones de la original. Las áreas que no fueron manipuladas tienden a tener un comportamiento uniforme, mientras que los sectores alterados responden de forma distinta a la nueva compresión, revelando su falta de cohesión. A menudo, estas diferencias se visualizan en escala de grises o mediante resaltado de bordes, facilitando la identificación de montajes. JPEG Ghost Map es especialmente útil para descubrir montajes sofisticados que han sido cuidadosamente integrados y no presentan bordes duros o diferencias visibles. Es complementaria a ELA y DCT Map, y destaca en casos donde se sospecha que una región fue incrustada desde otra imagen JPEG con diferente configuración de compresión. Como toda técnica, su análisis requiere experiencia, ya que imágenes recomprimidas varias veces también pueden generar artefactos que deben distinguirse de una manipulación real. (Bianchi et al, 2012).

4) Diffusion Model Deepfake: Busca artefactos introducidos por herramientas de generación de imágenes basadas en IA, como rostros generados por una GAN o imágenes sintetizadas completamente a partir de un mensaje de texto mediante un modelo de difusión (por ejemplo, Midjourney, Dall-E, Stable Diffusion). Adicionalmente tiene un módulo de Face GAN Deepfake, especializado en el análisis de rostros, a fin de evaluar si los mismos son reales o producto de una generación sintética.

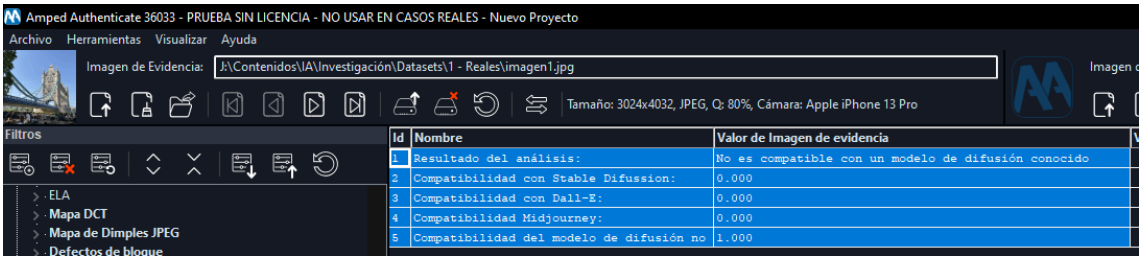


Imagen 10: Captura de resultados obtenidos de la herramienta Amped Authenticate

5) Análisis geométrico (sombras y reflejos): Los modelos generativos de IA pueden producir errores en la iluminación, creando sombras inconsistentes o reflejos poco naturales. Un análisis forense detallado puede identificar desviaciones en la dirección de la luz dentro de una imagen o video. En esta línea, hay herramientas basadas en modelos físicos de iluminación que han

sido desarrolladas para evaluar la autenticidad de retratos y fotografías con múltiples fuentes de luz (Zhou et al., 2018). AMPED AUTHENTICATE posee una herramienta exclusiva dedicada al trazado de los puntos de fuga, que permiten detectar este tipo de inconsistencias en la iluminación. Destacamos que la herramienta no detecta las sombras y realiza los trazados automáticamente, sino que precisa que la persona que está llevando adelante la pericia sobre la imagen, realice los trazados sobre la imagen, detectando así las potenciales inconsistencias que puedan revelar si la imagen es o no auténtica.

5.5.2. Pruebas con la herramienta

A continuación, utilizaremos AMPED Authenticate como herramienta para realizar las pruebas de los datasets previamente generados. La abreviatura MC significa “Modelo conocido”.

5.5.2.1. Pruebas Dataset Nº 1: Imágenes reales

Dataset evaluado	Nº de Imagen	Resultado obtenido				Conclusión (Resultado del Análisis) (MC = Modelo Conocido)	Resultado
		Compatibilidad con SD	Compatibilidad con Dall-e	Compatibilidad Midjourney	Compatibilidad del modelo de difusión no conocida		
Dataset Nº 1 - Imágenes Reales	1	0.000	0.000	0.000	1.000	No compatible con MC	VN
	2	0.002	0.000	0.000	0.998	No compatible con MC	VN
	3	0.002	0.019	0.06	0.972	No compatible con MC	VN
	4	0.000	0.000	0.000	1.000	No compatible con MC	VN
	5	0.000	0.000	0.000	1.000	No compatible con MC	VN
	6	0.000	0.000	0.000	1.000	No compatible con MC	VN
	7	0.008	0.076	0.621	0.295	Compatible con MC	FP
	8	0.001	0.000	0.000	0.999	No compatible con MC	VN
	9	0.000	0.000	0.000	1.000	No compatible con MC	VN
	10	0.000	0.002	0.005	0.993	No compatible con MC	VN

Tabla Nº 27: Resultados obtenidos con Amped Authenticate en Dataset Nº 1 - Imágenes Reales

5.5.2.2. Pruebas Dataset Nº 2: Stable Diffusion

Dataset evaluado	Nº de Imagen	Resultado obtenido				Conclusión (Resultado del Análisis) (MC = Modelo Conocido)	Resultado
		Compatibilidad con SD	Compatibilidad con Dall-e	Compatibilidad Midjourney	Compatibilidad del modelo de difusión no conocida		
Dataset Nº 2 -	1	0.001	0.003	0.030	0.965	No compatible con MC	FN

Stable Difussion	2	0.007	0.138	0.648	0.208	Compatible con MC	VP
	3	0.007	0.088	0.110	0.796	No compatible con MC	FN
	4	0.006	0.086	0.684	0.224	Compatible con MC	VP
	5	0.001	0.007	0.024	0.967	No compatible con MC	FN
	6	0.008	0.117	0.100	0.776	No compatible con MC	FN
	7	0.001	0.132	0.863	0.004	Compatible con MC	VP
	8	0.001	0.001	0.020	0.978	No compatible con MC	FN
	9	0.002	0.014	0.031	0.953	No compatible con MC	FN
	10	0.035	0.023	0.195	0.747	No compatible con MC	FN

Tabla Nº 28: Resultados obtenidos con Amped Authenticate en Dataset Nº 2 - Stable Diffusion

5.5.2.3. Pruebas Dataset Nº 3: Dall-e

Dataset evaluado	Nº de Imagen	Resultado obtenido				Conclusión (Resultado del Análisis) (MC = Modelo Conocido)	Resultado
		Compatibilidad con SD	Compatibilidad con Dall-e	Compatibilidad Midjourney	Compatibilidad del modelo de difusión no conocida		
Dataset Nº 3 - Dall-e	1	0.005	0.927	0.011	0.057	Compatible con MC	VP
	2	0.000	0.962	0.037	0.001	Compatible con MC	VP
	3	0.000	0.969	0.030	0.000	Compatible con MC	VP
	4	0.001	0.951	0.047	0.002	Compatible con MC	VP
	5	0.001	0.991	0.001	0.008	Compatible con MC	VP
	6	0.008	0.907	0.059	0.026	Compatible con MC	VP
	7	0.004	0.961	0.034	0.000	Compatible con MC	VP
	8	0.001	0.992	0.004	0.004	Compatible con MC	VP
	9	0.001	0.949	0.011	0.039	Compatible con MC	VP
	10	0.002	0.974	0.021	0.003	Compatible con MC	VP

Tabla Nº 29: Resultados obtenidos con Amped Authenticate enDataset Nº 3 - Dall-e

5.5.2.4. Pruebas Dataset Nº 4: Grok

Dataset evaluado	Nº de Imagen	Resultado obtenido				Conclusión (Resultado del Análisis) (MC = Modelo Conocido)	Resultado
		Compatibilidad	Compatibilidad con	Compatibilidad	Compatibilidad del modelo de difusión		

		con SD	Dall-e	Midjourney	no conocida		
Dataset Nº 4 - GROK	1	0.002	0.008	0.019	0.971	No Compatible con MC	VN
	2	0.029	0.105	0.035	0.830	No Compatible con MC	VN
	3	0.019	0.788	0.028	0.164	Compatible con MC	FP
	4	0.003	0.009	0.003	0.985	No Compatible con MC	VN
	5	0.008	0.036	0.024	0.932	No Compatible con MC	VN
	6	0.043	0.027	0.028	0.902	No Compatible con MC	VN
	7	0.024	0.163	0.225	0.588	No Compatible con MC	VN
	8	0.001	0.001	0.003	0.995	No Compatible con MC	VN
	9	0.02	0.013	0.019	0.965	No Compatible con MC	VN
	10	0.067	0.323	0.163	0.447	No Compatible con MC	VN

Tabla Nº 30: Resultados obtenidos con Amped Authenticate en Dataset Nº 4 - Grok

5.5.2.5. Pruebas Dataset Nº 5: Fooocus

Dataset evaluado	Nº de Imagen	Resultado obtenido				Conclusión (Resultado del Análisis) (MC = Modelo Conocido)	Resultado
		Compatibilidad con SD	Compatibilidad con Dall-e	Compatibilidad Midjourney	Compatibilidad del modelo de difusión no conocida		
Dataset Nº 5 - Fooocus	1	0.006	0.003	0.006	0.986	No Compatible con MC	FN
	2	0.016	0.106	0.027	0.851	No Compatible con MC	FN
	3	0.003	0.070	0.030	0.897	No Compatible con MC	FN
	4	0.023	0.247	0.118	0.612	No Compatible con MC	FN
	5	0.05	0.014	0.011	0.970	No Compatible con MC	FN
	6	0.142	0.068	0.083	0.707	No Compatible con MC	FN
	7	0.013	0.064	0.648	0.275	No Compatible con MC	FN
	8	0.019	0.019	0.006	0.956	No Compatible con MC	FN
	9	0.035	0.180	0.561	0.225	Compatible con MC	VP
	10	0.023	0.245	0.712	0.020	Compatible con MC	VP

Tabla Nº 31: Resultados obtenidos con Amped Authenticate en Dataset Nº 5 - Fooocus

5.5.2.6. Pruebas Dataset Nº 6: Híbridas

Dataset evaluado	Nº de Imagen	Resultado obtenido				Conclusión (Resultado del Análisis)	Resultado
		Compatibilidad con SD	Compatibilidad con Dall-e	Compatibilidad Midjourney	Compatibilidad del modelo de difusión no conocida		
Dataset Nº 6 - Híbridas	1	0.001	0.048	0.009	0.942	No Compatible con MC	VP
	2	0.001	0.000	0.000	0.999	No Compatible con MC	FN
	3	0.003	0.055	0.005	0.937	No Compatible con MC	VP
	4	0.010	0.316	0.146	0.529	No Compatible con MC	VP
	5	0.001	0.007	0.014	0.978	No Compatible con MC	FN
	6	0.001	0.005	0.027	0.967	No Compatible con MC	FN
	7	0.005	0.054	0.599	0.341	Compatible con MC	FP
	8	0.001	0.000	0.000	0.999	No Compatible con MC	FN
	9	0.000	0.000	0.000	1.000	No Compatible con MC	FN
	10	0.001	0.015	0.021	0.963	No Compatible con MC	FN

Tabla Nº 32: Resultados obtenidos con Amped Authenticate en Dataset Nº 6- Híbridas

5.5.3. Prueba extra: Análisis de imagen con filtro JPEG Ghost Map

A diferencia del resto de las herramientas analizadas, AMPED AUTHENTICATE posee una serie de filtros que pueden ser de utilidad para distintos casos. Después de probar varias herramientas, confirmamos la existencia de una herramienta que nos permite, con un nivel de precisión importante, determinar qué sector de la imagen ha sido modificada o alterada (en nuestro caso, por IA).

Hemos identificado que utilizando el filtro “JPEG Ghost Map”, es posible visualizar la parte (con mayor o menor precisión) de la imagen modificada o alterada por IA. Dicho filtro, se ha utilizado configurando un tamaño de bloque de 32, modelo simplificado y en tipo de mapa, hemos seleccionado “Superponer Color Invertido” (para visualizar más rápido la parte alterada). A continuación, exhibimos los resultados obtenidos:

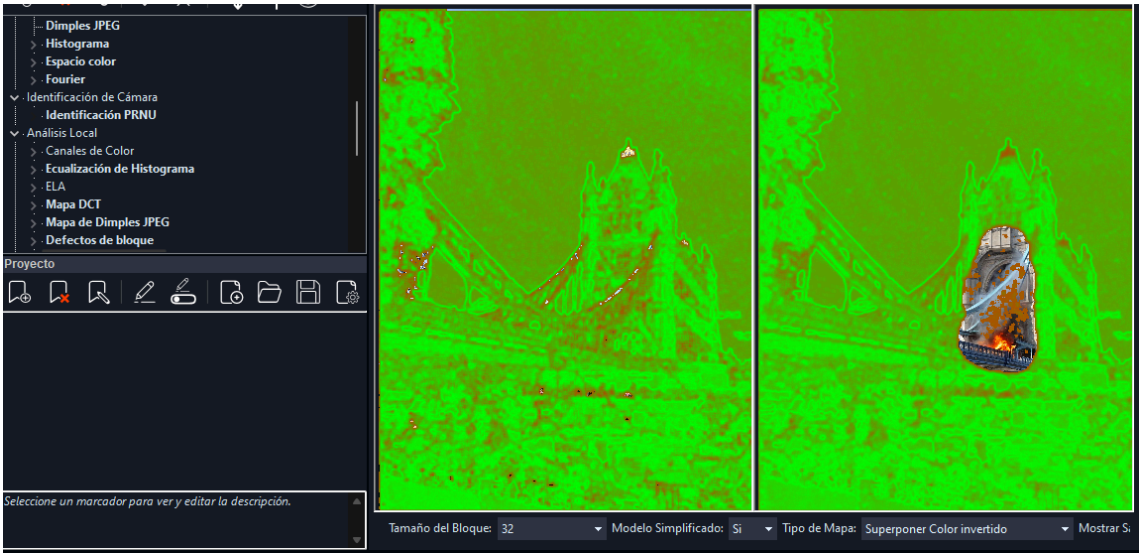


Imagen Nº 11: Aplicación de filtro JPEG Ghost Map sobre Imagen Nº 1 (Dataset 1 vs Dataset 6)

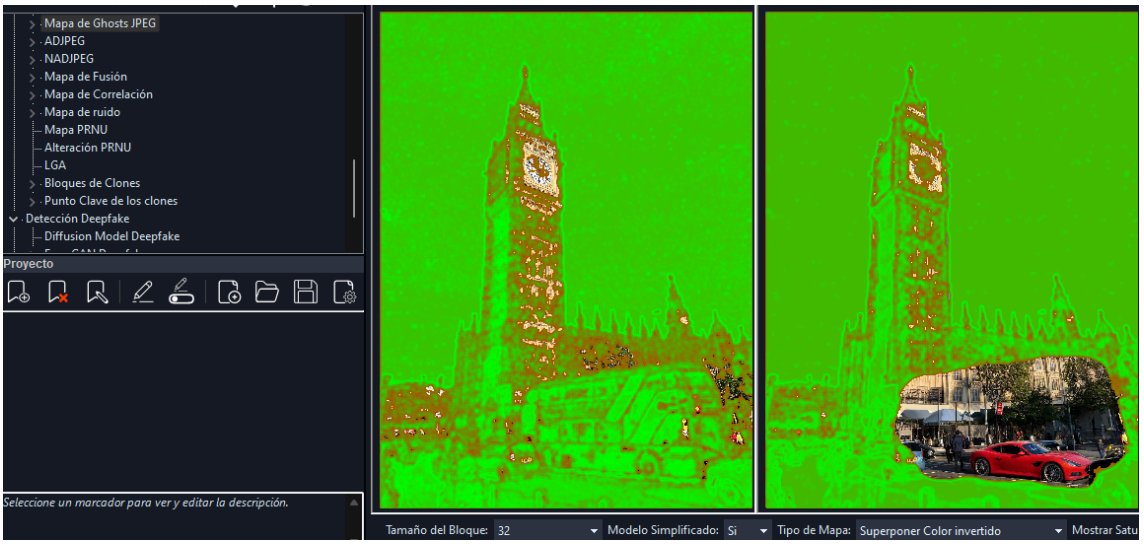


Imagen Nº 12: Aplicación de filtro JPEG Ghost Map sobre Imagen Nº 2 (Dataset 1 vs Dataset 6)

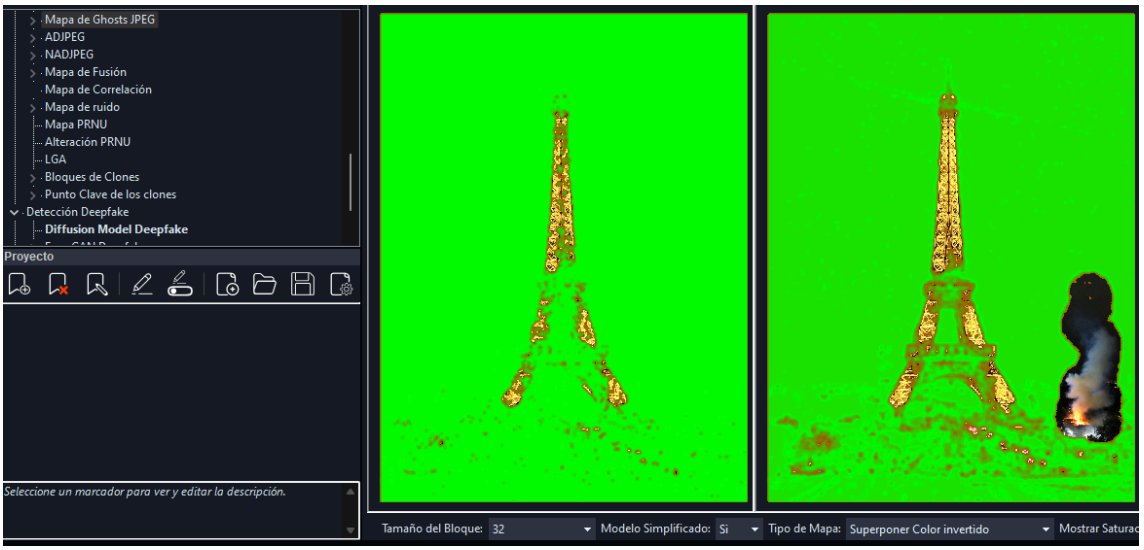


Imagen Nº 13: Aplicación de filtro JPEG Ghost Map sobre Imagen Nº 3 (Dataset 1 vs Dataset 6)

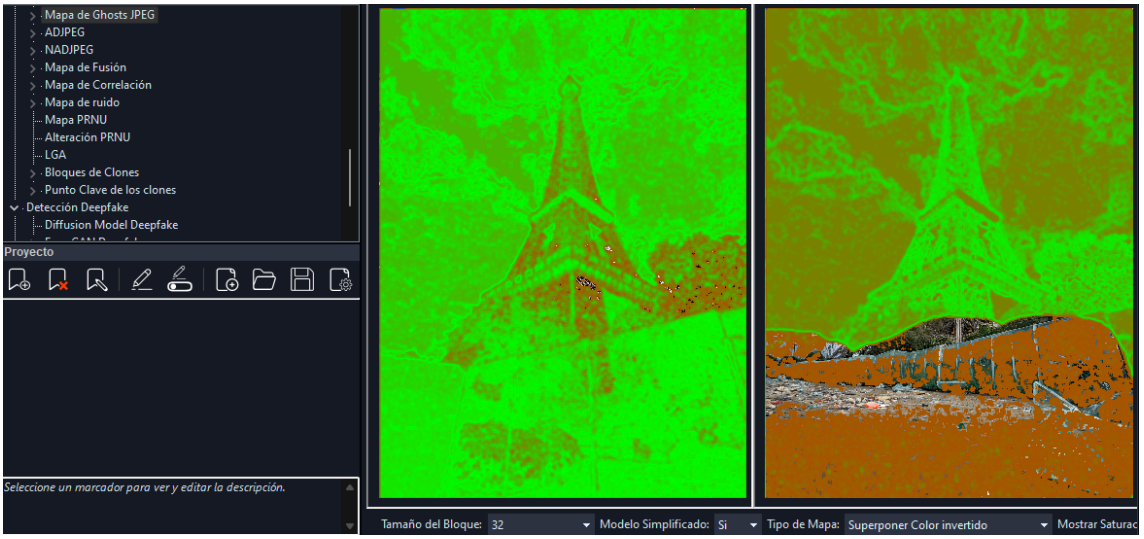


Imagen Nº 14: Aplicación de filtro JPEG Ghost Map sobre Imagen Nº 4 (Dataset 1 vs Dataset 6)

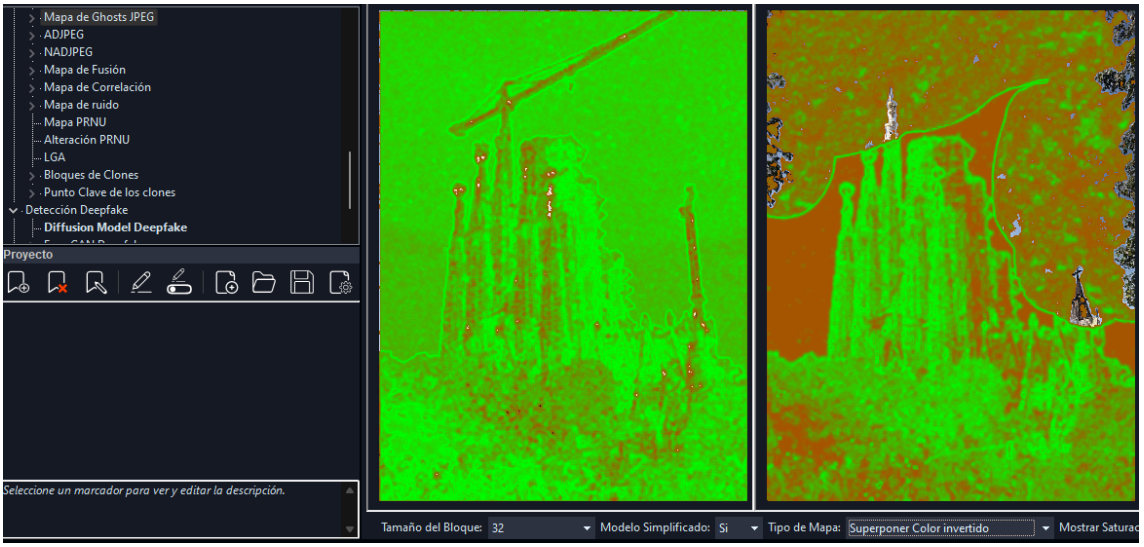


Imagen Nº 15: Aplicación de filtro JPEG Ghost Map sobre Imagen Nº 5 (Dataset 1 vs Dataset 6)

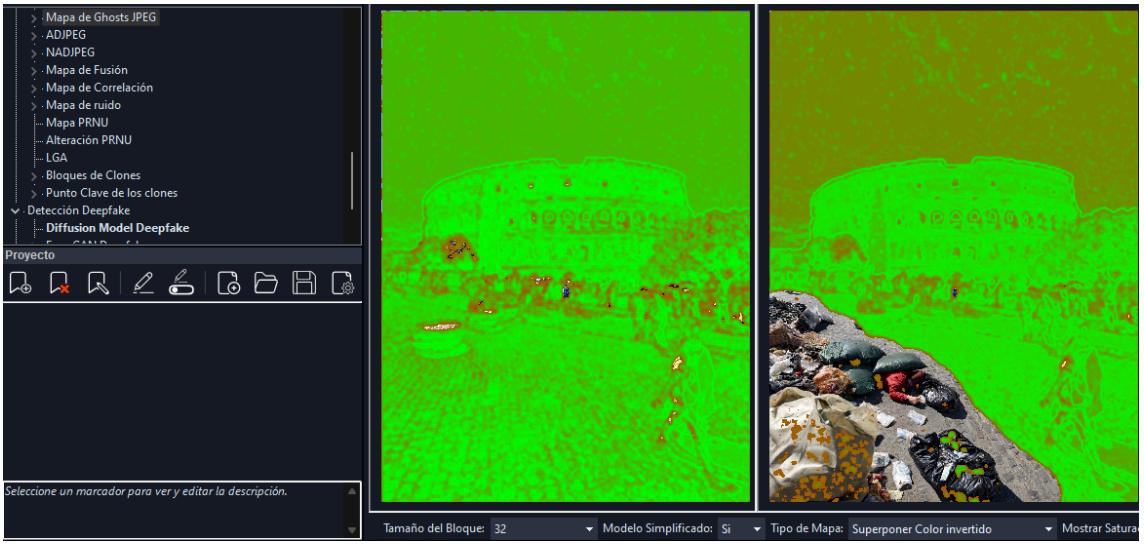


Imagen Nº 16: Aplicación de filtro JPEG Ghost Map sobre Imagen Nº 6 (Dataset 1 vs Dataset 6)

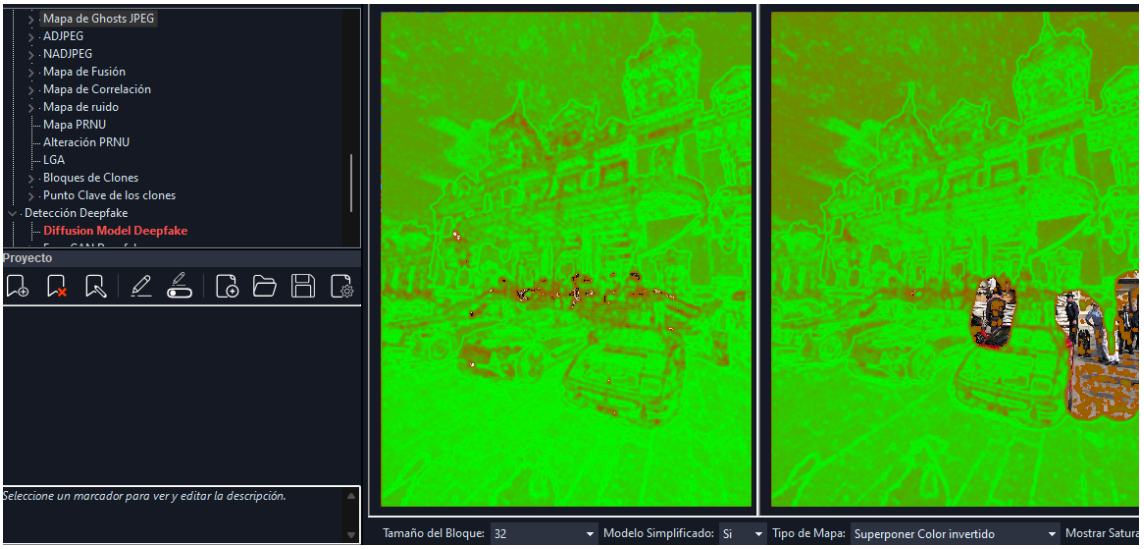


Imagen Nº 17: Aplicación de filtro JPEG Ghost Map sobre Imagen Nº 7 (Dataset 1 vs Dataset 6)

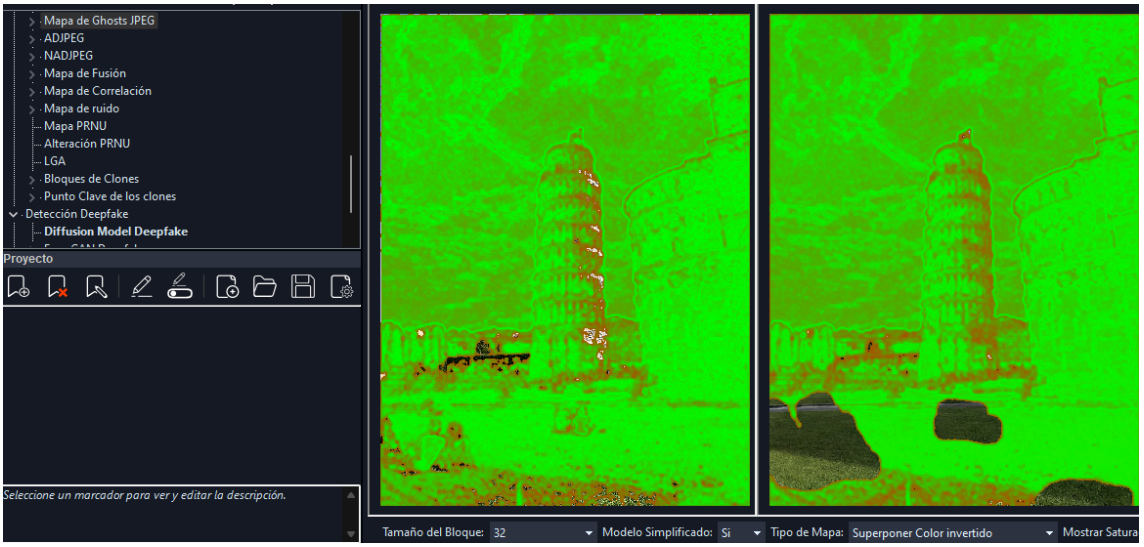


Imagen Nº 18: Aplicación de filtro JPEG Ghost Map sobre Imagen Nº 8 (Dataset 1 vs Dataset 6)

Imagen Nº 19: Aplicación de filtro JPEG Ghost Map sobre Imagen Nº 9 (Dataset 1 vs Dataset 6)

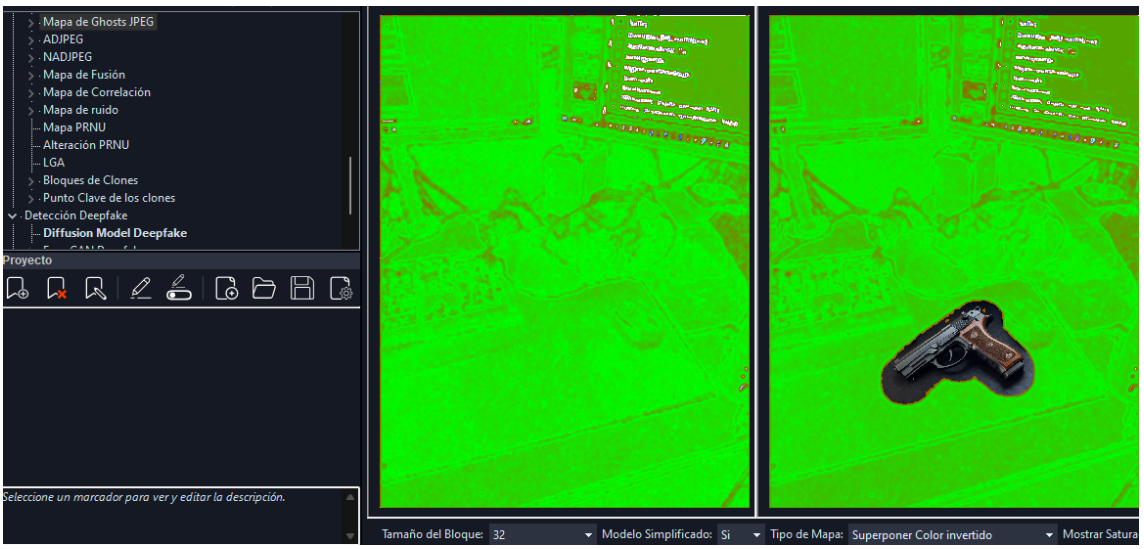


Imagen Nº 20: Aplicación de filtro JPEG Ghost Map en Imagen Nº 10 (Dataset 1 vs Dataset 6)

5.5.4. Análisis de resultados obtenidos

Las pruebas sobre el primer dataset (Tabla Nº 27), se puede observar buenos resultados, ya que en 9 de las 10 imágenes, la herramienta determinó que las imágenes cargadas (reales) no eran compatibles con ninguno de los modelos de entrenamiento, excepto por el caso de la Imagen Nº 7, donde en un 62% dió compatibilidad con Midjourney, generando el único FP. Del resto de los resultados, también destacamos que el índice de "modelo no conocido", estuvo en 5 de los 9 resultados correctos, en el 100%. En el que dio menos índice, fue en la imagen 3, donde el índice fue de 97,2%.

Las pruebas sobre el segundo dataset (Tabla Nº 28), se puede observar resultados bastante negativos, considerando que las imágenes sintéticas fueron generadas con uno de los modelos que supuestamente están incluídos en el análisis de la herramienta, Stable Difussion. En este caso, sólo en 3 de los 10 casos, la herramienta detectó que eran imágenes sintéticas y en ninguno de los 3 determinó que eran de SD, sino que eran de MJ.

Las pruebas sobre el tercer dataset (Tabla Nº 29), se puede observar el mejor resultado alcanzado por esta herramienta, donde ha acertado en los 10 casos, no solamente indicando que respondían a un modelo conocido, sino además acertando en el segundo nivel de análisis, al determinar con éxito que todas las imágenes fueron realizadas con Dall-e, con porcentajes de compatibilidad que en superaron el 90% de compatibilidad.

Las pruebas sobre el cuarto dataset (Tabla Nº 30), la herramienta determinó en 9 de los 10 casos, que las imágenes no responden a un modelo conocido (algo que, técnicamente es cierto, ya que la herramienta no ha sido entrenada para detectar contenidos de GROK). Por este motivo, se ha considerado estos resultados como Verdaderos Negativos. No obstante, se detectó en un caso, que la imagen había sido generada por Dall-e, considerándolo un Falso Positivo.

Las pruebas sobre el quinto dataset (Tabla Nº 31), la herramienta sólo acertó en dos casos (considerando que Fooocus es un modelo derivado de SD, consideramos que debería haberlo

reconocido bajo esa rama), en las imágenes 9 y 10, pero en ninguna acertó en cuanto al modelo (en ambos, sugirió que fueron realizados por MJ).

Las pruebas sobre el sexto dataset (Tabla Nº 32), consideramos que la herramienta sólo acertó en 3 de los 10 casos, aplicando el criterio interpretativo sostenido anteriormente. Es decir, en aquellos casos donde la herramienta detectó modelos conocidos en la imagen, que oscilan entre 4 y un 40% de IA. Se consideró como error aquellos casos, como el de la imagen 7, donde la herramienta detectó que fue realizada por MJ (arrastrando el mismo error ya indicado en las pruebas del Dataset Nº 1).

Las pruebas extras realizadas, sobre el sexto dataset pero utilizando la técnica de JPEG Ghost Map, se pudo confirmar que en todos los casos la técnica ha podido indicar la sección o área de la imagen adulterada sintéticamente. Si bien no es posible aplicar la fórmula de tasa de error que hemos utilizado a lo largo del trabajo, es posible concluir que, este tipo de técnica, ha tenido un 100% de efectividad sobre los análisis realizados.

5.5.5. Tasa de Error de la herramienta

A continuación, realizaremos un repaso por las tasas de error calculadas a nivel individual por las distintas pruebas, calculando un promedio de todas ellas:

Dataset N°	Generado a partir de	Reales vs. Sintéticas		
		Exactitud	Tasa de Error	Índice PIAG
1	Imágenes reales	90%	10%	No aplica
2	Stable Difussion	30%	70%	0%
3	Dall-e	100%	0%	100%
4	Grok	90%	10%	0%
5	Fooocus	20%	80%	0%
Promedio calculado		66%	**34%**	**25%**

Tabla Nº 33: Tabla de Error de Herramienta Amped Authenticate para detección de imágenes generadas sintéticamente

De acuerdo a los cálculos realizados en nuestro primer nivel de análisis (determinar si las imágenes han sido generadas o no sintéticamente), su nivel de exactitud no ha sido tan bueno como otras herramientas, alcanzando un 66% de exactitud y con 34% de tasa de error.

Adicionalmente, debemos agregar que dentro del marco de los resultados acertados, los niveles de acierto con respecto a la IA generativa (índice PIAG), ha sido excelente sólo en el caso de Dall-e, que pareciera ser el modelo con mayor nivel de entrenamiento / aprendizaje de la herramienta. En los otros casos, sorprende el escaso nivel de detección en SD y Fooocus, ambos basados en el mismo modelo.

Dataset	Generado a partir	Técnica	Exactitud	Tasa de	Índice PIAG

N°	de			Error	
6	Imágenes híbridas	Identificación IA	30%	70%	0%
6	Imágenes híbridas	JPEG Ghost Map	100%	-	-

Tabla Nº 34: Resultados obtenidos con Amped Authenticate para la identificación de imágenes reales modificadas con inteligencia artificial.

En el caso del dataset híbrido, la utilización de la técnica de detección de IA de la herramienta, presenta niveles de tasa de error demasiado altos para ser considerado aplicable en el ámbito forense. No obstante, como afirmamos anteriormente, utilizando la técnica de JPEG Ghost Map, se pudo confirmar que en todos los casos la técnica ha podido indicar la sección o área de la imagen adulterada sintéticamente.

5.6. Resultados Preliminares

Como nota preliminar, destacamos que nuestras apreciaciones deben ser tomadas como un indicio de uso para herramientas, ya que consideramos pertinente realizar pruebas y estudios más complejos y cuantitativos sobre las herramientas seleccionadas, a fin de confirmar nuevamente su capacidad para uso forense.

5.6.1. Herramienta EXIFTool

Podemos observar que el análisis de metadatos como técnicas de análisis de imágenes sintéticas a nivel forense, podría ser de utilidad sólo ante casos concretos (50% de los casos analizados), como el caso de haber sido generadas con Stable Difussion y GROK, podríamos encontrar evidencia del prompt de generación utilizado, o bien, para confirmar que una imagen ha sido capturada con un determinado dispositivo, confirmando su origen. No obstante, siempre debemos tener presente que sería posible, a través de la modificación o alteración dolosa de los metadatos, generar artificialmente dicha información.

Cabe aclarar que el éxito de aplicación de la técnica -al igual que sucede con el análisis de metadatos de cualquier imagen- quedará condicionado a variables como la de poder analizar el contenedor original, ya que muchos de los procesamientos posteriores (por ejemplo, su transmisión a través de Whatsapp), afectaría los resultados.

No obstante lo dicho, entendemos que el análisis de los metadatos, es una técnica que por su nivel de divulgación y facilidad de implementación, debe ser tenida en cuenta en la etapa inicial de análisis del perito informático forense al momento de intentar un análisis que permita detectar si el contenido ha sido generado por inteligencia artificial, junto a otras herramientas que puedan abordar técnicas diferentes. El aspecto positivo es que, de encontrarse resultados positivos (como en los modelos de Stable Difussion o GROK), entendemos que su hallazgo permitiría tener elementos técnicos para confirmar que la misma ha sido generada sintéticamente.

5.6.2. Herramienta Susy

A nivel general, la herramienta Susy ha mostrado resultados más precisos (alrededor del 80%) con sólo algunos de los modelos de generación de imágenes sintéticas más tradicionales, pero ha mostrado resultados mucho peores, llegando al 40% con otros modelos más modernos. En promedio, la tasa de error del 56% es un valor que es demasiado alto para ser considerada una herramienta apta para su aplicación forense. A su vez, dentro de los aciertos realizados, el índice de Precisión sobre la IA Generativa detectada, sólo ha tenido, en promedio, un 40% de acierto, muy por debajo de los estándares internacionales para considerar la potencial aplicación de la herramienta a nivel forense.

En relación a su aplicación para la detección de imágenes reales que han sido modificadas por IA, su nivel de precisión aumentó a un 80%, pero con un índice PIAG del 0%, es decir que en ningún caso acertó con la IA realmente utilizada (Stable Diffusion o Fooocus). Si bien las tasas mejoraron en este aspecto, también consideramos que están por debajo de los estándares internacionales para considerar la potencial aplicación de la herramienta a nivel forense.

5.6.3. Herramienta HIVE

Consideramos que en base a los resultados obtenidos y su tasa de error obtenida, HIVE muestra valores realmente interesantes para tener en cuenta y que pueden, para casos concretos (donde se desee analizar si la imagen es o no sintética), ser una buena opción a tener en cuenta en el marco de un análisis informático forense.

No obstante, se debe considerar que la herramienta ha tenido serias dificultades al analizar el dataset de imágenes híbridas N° 6, donde sólo ha acertado en uno de los 10 casos procesados, dando para nuestra evaluación, una tasa de error insostenible del 90%, insostenible para cualquier tipo de aplicación práctica.

5.6.4. Herramienta SightEngine

Consideramos que es una herramienta interesante para tener en cuenta en los casos donde sea necesario evaluar si una imagen es o no sintética, ya que a nivel comparativo, SightEngine muestra valores realmente interesantes para tener en cuenta y que pueden, para casos concretos, ser una buena opción a tener en cuenta en el marco de un análisis informático forense.

En el caso del dataset híbrido, se ha considerado un sólo caso de acierto, descartando por completo la utilización de esta herramienta para este tipo de casos. Es decir, al igual que sucedió con HIVE, el problema detectado para la utilización de la herramienta, es para los casos donde estemos hablando de imágenes reales en su esencia, con modificaciones o alteraciones realizadas a través de IA.

5.6.5. Herramienta AMPED AUTHENTICATE

Del análisis de los resultados obtenidos, si bien en el caso de las imágenes realizadas por Dall-e ha funcionado perfectamente, en el resto de las pruebas los resultados no han alcanzado los mínimos recomendados. A modo general, de la primer tanda de pruebas concluimos en que la

herramienta analizada aún no posee los niveles de exactitud / tasa de error necesarios para su aplicación en la detección sobre si una imagen ha sido o no generada sintéticamente.

Sin embargo, de los resultados obtenidos de la prueba extra (Sección 5.6.3) podemos observar que la aplicación de la técnica JPEG Ghost Map, ha permitido en el 100% de los casos analizados sobre detección de imágenes reales alteradas sintéticamente, obtener resultados positivos en un aspecto donde todas las otras herramientas habían fracasado. Los resultados obtenidos permitirían afirmar que esta técnica (junto a otras que podrían evaluarse a futuro), podría ser aplicada para evaluar casos donde se cuestione si imágenes reales han sido manipuladas, detectando con un importante nivel de precisión, las zonas que han sido alteradas. Destacamos también, que estas "zonas modificadas", también presentan falsos positivos, observándose que en algunos resultados se indican como zonas alteradas, algunas que no han tenido ningún tipo de alteración.

5.6.6. Comparativo de Herramientas analizadas

Id	Herramienta	Dataset	Técnica utilizada	Tasa de Error	Índice PIAG
1	ExifTool	1 a 6	Análisis de Metadatos	No es posible calcular (50% efectividad)	
2	Susy	1 a 5	IA entrenada con Modelos de Difusión Conocidos	56%	50,25%
		6		20%	0%
3	Hive IA Detector	1 a 5	IA entrenada con Modelos de Difusión Conocidos	0%	65%
		6		100%	100%
4	SightEngine	1 a 5	IA entrenada con Modelos de Difusión Conocidos	0%	72,5%
		6		90%	78%
5	Amped Five Authenticate	1 a 5	IA entrenada con Modelos de Difusión Conocidos	34%	25%
		6	IA entrenada con Modelos de Difusión Conocidos	70%	0%
		6	ELA (Error Level Analysis) y JPEG Ghost Map	No es posible calcular (100% efectividad)	

6. Guía Forense de Buenas Prácticas para la detección de imágenes generadas por IA

6.1. Aspectos preliminares

Esta guía tiene como objetivo ser un primer acercamiento en el desafío del análisis forense de imágenes digitales sospechadas de haber sido generadas o modificadas mediante IA. Como punto de partida, sugerimos tener en consideración el Modelo PURI (Proceso Unificado de Recuperación de Información), planteado en la "Guía Integral de Empleo de la Informática Forense en el Proceso Penal – Modelo PURI"[19] en relación a todas las etapas aplicables al proceso tradicional de la informática forense. Dentro de dicho esquema, la presente guía pretende sugerir y brindar una serie de procedimientos aplicables en las etapas de extracción y análisis del objeto pericial, así como algunas sugerencias a tener en cuenta en la etapa de presentación.

6.2. Guía de Buenas Prácticas: Fase de extracción y análisis

Considerando que la fiabilidad de los resultados dependerán del nivel de calidad del objeto de estudio, resulta fundamental que la adquisición / extracción de los contenidos objeto del análisis sean realizadas en su contenedor original, por lo que deberá analizarse la forma de extracción más adecuada de acuerdo a cada caso.

Seguidamente, se propone la realización de las siguientes fases de análisis:

Fase 1: Análisis de Metadatos	Herramientas sugeridas	Exiftool (ExifTool.org)
	Procedimiento	Verificar a través del análisis de los metadatos existen rastros del prompt utilizado para la generación de la imagen
	Precauciones	- Posibilidad de modificación de metadatos - Posibilidad de no contar con contenedor original
Fase 2: Análisis de Inspección Visual Asistida	Herramientas sugeridas	AMPED Authenticate (https://ampedsoftware.com/)
	Procedimiento	Verificar a través las herramientas, la existencia de anomalías en la imagen. Se sugiere revisar: (a) proporciones, (b) simetrías, (c) iluminación, (d) detalles del cuerpo humano (manos, ojos, dientes, etc) donde suelen

[19] Di Iorio, A. H., [et al.]. (2016). Guía integral de empleo de la informática forense en el proceso penal (2ª ed.). Universidad FASTA. Recuperado de: http://redi.ufasta.edu.ar:8082/jspui/bitstream/123456789/1592/2/PAIF.pdf

		existir errores de IA.
	Precauciones	- Posibilidad de no contar con buena calidad de imagen - Posibilidad de que la imagen haya sido realizada con modelos con mucho entrenamiento (menos errores)
Fase 3: Análisis de detección de Modelos de IA	Herramientas sugeridas	- HIVE (https://hivemoderation.com/) - SightEngine (https://sightengine.com/)
	Procedimiento	Verificar a través de las herramientas, la detección de patrones de modelos entrenamiento conocidos.
	Precauciones	- La eficacia dependerá de la cantidad de modelos con los que la herramienta ha sido entrenada, así como su nivel de actualización - Si el modelo con el cuál ha sido generada la imagen, no ha sido parte de los modelos de entrenamiento, puede dar falsos negativos. - En algunos casos, las herramientas sugeridas han dado un verdadero positivo al detectar que la imagen ha sido generada con IA, pero han tenido errores al indicar el tipo de modelo utilizado.
Fase 4: Análisis de detección de anomalías en compresión	Herramientas sugeridas	- AMPED Authenticate (https://ampedsoftware.com/) - Forensically (https://29a.ch/photo-forensics)
	Procedimiento	Verificar la compresión de la imagen a fin de detectar artefactos inusuales que puedan ser rastros de modificaciones. Se sugiere la revisión mínima con ELA (Error Level Analysis) y JPEG Ghost.
	Precauciones	- Posibilidad de no contar con buena calidad de imagen - Posibilidad de que la imagen haya sido realizada con modelos con mucho entrenamiento

6.3. Consideraciones finales

Las herramientas sugeridas se basan en un estudio de investigación realizado en el año 2025, bajo el título “Técnicas forenses para la detección de contenido generado por inteligencia artificial”, realizado por el Dr. Marcelo Temperini en el marco del Trabajo Final de la Especialización en Informática Forense de la Universidad FASTA. Por lo tanto, sugerimos siempre consultar por las últimas versiones de las herramientas, así como analizar la posibilidad de aplicación de otras herramientas o técnicas similares, de acuerdo a las necesidades del caso concreto.

Del estudio citado, se puede concluir que al momento de su realización y evaluación de herramientas especializadas, se ha evidenciado que, si bien algunas tecnologías muestran resultados alentadores en determinadas circunstancias, ninguna herramienta ofrece todavía una confiabilidad absoluta para su uso forense sin validación complementaria.

La presente Guía de Buenas Prácticas pretende ser un primer acercamiento, destacando las conclusiones alcanzadas son preliminares y se destaca la necesidad de continuar con la

realización de estudios e investigaciones más profundas y sobre todo, más cuantitativas a nivel de muestras, a fin de confirmar su capacidad para uso forense. También destacamos la importancia de conformar grupos de trabajo donde puedan incorporarse el expertise y conocimientos de peritos especializados en el tratamiento de imágenes y videos.

Subrayamos la obligación de confidencialidad que todo profesional forense debe resguardar sobre el objeto pericial, por lo que sugerimos siempre optar por herramientas que puedan ser utilizadas y ejecutadas localmente en el equipo de análisis forense, evitando la transferencia de la información del caso hacia servidores de terceros.

En aquellos casos (depende de la herramienta y el tipo de licencia) donde no es posible el uso local (es exclusivo en línea), sugerimos revisar previamente los contratos (términos y condiciones legales, así como las políticas de privacidad) a fin de poder identificar los potenciales riesgos de la utilización de la herramienta.

La diversidad en la performance de las herramientas examinadas refuerza la necesidad de establecer una guía de actuación estructurado que guíe a los peritos informáticos en su labor, considerando las limitaciones técnicas y los aspectos legales involucrados, por lo que se concluye que el abordaje forense de imágenes generadas o manipuladas por IA debe combinar métodos automáticos de herramientas, administradas por parte de un especialista en informática forense.

7. Conclusiones finales

A modo general, entendemos que el trabajo permitió realizar un primer acercamiento, desde una perspectiva forense, al desafío que implica el emergente crecimiento en la generación y manipulación de imágenes mediante inteligencia artificial.

Consideramos también apropiado advertir que este trabajo no ha tenido pretensiones de realizar un desarrollo acabado y minucioso de todas las técnicas forenses que podrían ser aplicable, ya que llevar adelante implicaría un desarrollo mucho más extenso de lo que podríamos abordar, no solamente de forma individual, sino además, en los tiempos y la extensión asignada para el presente trabajo. Por tal motivo, en diferentes pasajes del estudio, dejamos referencias a temas de interés, que podrían ser puntos de partida para potenciales futuras líneas investigaciones, tanto de este autor, como de otros y otras que tengan interés en continuar desarrollando conocimiento en esta área.

A lo largo del estudio se han podido realizar pruebas sobre una serie de herramientas que podrían ser tenidas en cuenta por parte del especialista en informática forense al momento de tener el desafío técnico de dictaminar sobre si una imagen ha sido o no generada o modificada utilizando inteligencia artificial. A partir de la construcción de un conjunto de datasets y la evaluación de herramientas especializadas, se ha evidenciado que, si bien algunas tecnologías muestran resultados alentadores en determinadas circunstancias, ninguna herramienta ofrece todavía una confiabilidad absoluta para su uso forense. No obstante, entre los hallazgos principales que más podemos destacar, encontramos:

- La utilidad de los metadatos como técnica inicial, aunque limitada por la posibilidad de alteración dolosa.
- El valor del análisis de errores de compresión (ELA), en particular el uso del filtro JPEG Ghost Map, que demostró una eficacia superior en imágenes híbridas.
- La alta precisión de herramientas como Hive o SightEngine para detectar imágenes completamente generadas por IA, siempre que sean realizadas con determinados modelos con los que fueron entrenados. Destacamos que estas mismas herramientas presentaron limitaciones serias para identificar modificaciones parciales (Dataset 6).
- La necesidad de validar las herramientas con datasets propios y controlados, dada la variabilidad de resultados según el modelo de IA generativa y el tipo de imagen.

Que, en el marco de la propuesta innovadora del presente estudio, proponermos una “Guía de Buenas Prácticas”, con el fin que sea un recurso de utilidad para aquellas personas que se encuentre con el desafío de tener que participar en el análisis forense de una imagen que potencialmente pueda estar modificada por inteligencia artificial.

Que, es importante que se comprenda que las conclusiones alcanzadas son preliminares y limitadas a los alcances del presente estudio de final de carrera de Especialización en Informática Forense, destacando este autor la necesidad de continuar con la realización de estudios e investigaciones más profundas y sobre todo, más cuantitativas a nivel de muestras, a fin de confirmar su capacidad para uso forense. También destacamos la importancia de conformar grupos de trabajo donde puedan incorporarse la experiencia y conocimientos de peritos especializados en el tratamiento de imágenes y videos.

Destacamos además la obligación de confidencialidad que todo profesional forense debe resguardar sobre el objeto pericial, por lo que sugerimos siempre optar por herramientas que puedan ser utilizadas y ejecutadas localmente en el equipo de análisis forense, evitando la transferencia de la información del caso hacia servidores de terceros. En aquellos casos (depende de la herramienta y el tipo de licencia) donde no es posible el uso local (es exclusivo en línea), sugerimos revisar previamente los contratos (términos y condiciones legales, así como las políticas de privacidad) a fin de poder identificar los potenciales riesgos de la utilización de la herramienta.

Se concluye que el abordaje forense de imágenes generadas o manipuladas por IA debe combinar métodos automáticos, siempre administradas por un especialista en informática forense. La diversidad en la performance de las herramientas examinadas refuerza la necesidad de establecer una guía de actuación estructurado que permita orientar a los especialistas en informática forense en su labor, considerando las limitaciones técnicas y los aspectos legales involucrados.

8. Bibliografía

- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. Advances in neural information processing systems, 27.
- Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 4401-4410.
- Mirsky, Y., & Lee, W. (2021). The creation and detection of deepfakes: A survey. ACM Computing Surveys (CSUR), 54(1), 1-41.
- Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., & Ganguli, S. (2015). Deep unsupervised learning using nonequilibrium thermodynamics. Proceedings of the International Conference on Machine Learning, 2256-2265.
- Dhariwal, P., & Nichol, A. (2021). Diffusion models beat GANs on image synthesis. Advances in Neural Information Processing Systems, 34, 8780-8794.
- Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 10684-10695.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). *Attention is all you need*. Advances in Neural Information Processing Systems, 30.
- Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021). *Learning transferable visual models from natural language supervision*. arXiv
- Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., ... & Sutskever, I. (2022). *Hierarchical text-conditional image generation with CLIP latents*.
- LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.
- Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition.
- Clark, J., Ramesh, A., Nichol, A., Dhariwal, P., & Radford, A. (2024). *Sora: High-resolution text-to-video generation with latent diffusion models*. OpenAI Research.
- Ho, J., Saharia, C., Chan, W., Fleet, D. J., Norouzi, M., & Salimans, T. (2022). *Imagen Video: High-definition video generation with diffusion models*.
- Skorokhodov, I., Schwarz, K., & Lempitsky, V. (2022). *StyleGAN-V: A continuous video generator with the style-based architecture*.
- Tulyakov, S., Liu, M. Y., Yang, X., & Kautz, J. (2018). *MoCoGAN: Decomposing motion and content for video generation*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1526-1535.
- Yan, Z., Rombach, R., Esser, P., & Ommer, B. (2021). *VideoGPT: A generative pre-trained transformer for video generation*.
- Wang, S. Y., Zhang, O., & Chang, S. F. (2019). CNN-generated images are surprisingly easy to spot… for now. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 8695-8704.

- Zhou, Z., Shi, Y., Zhang, X., & Fan, H. (2018). Illumination inconsistency detection for image forensics. Journal of Visual Communication and Image Representation, 55, 523-531.
- Cozzolino, D., Thies, J., Rössler, A., Nießner, M., & Verdoliva, L. (2019). SpoC: Spoofing camera fingerprints. IEEE/CVF International Conference on Computer Vision Workshops, 2312-2320.
- Fan, Y., Li, S., & Lyu, S. (2020). Finding messages in a haystack: Deepfake detection with metadata forensics. IEEE Conference on Computer Vision and Pattern Recognition Workshops, 1442-1451.
- Marra, F., Gragnaniello, D., Cozzolino, D., & Verdoliva, L. (2019). Do GANs leave artificial fingerprints?. IEEE Transactions on Information Forensics and Security, 14(11), 2726-2739.
- Fan, Y., Li, S., & Lyu, S. (2020). Finding messages in a haystack: Deepfake detection with metadata forensics. IEEE Conference on Computer Vision and Pattern Recognition Workshops, 1442-1451.
- Gioia, C. V. (2017). Metodología de análisis forense informático para la obtención de evidencia digital en base de datos. Universidad Nacional de La Matanza.
- Schinas, M., & Papadopoulos, S. (2024). SIDBench: A Python Framework for Reliably Assessing Synthetic Image Detection Methods.
- Bernabeu-Perez, P., & Lopez-Cuena E., & Garcia-Gasulla D. (2024). Present and Future Generalization of Synthetic Image Detectors.
- Li, Y., Liu, Z., Zhao, J., Ren, L., Li, F., Luo, J., & Luo, B. (2024). The adversarial AI-art: Understanding, generation, detection, and benchmarking.
- Farid, H. (2009). Exposing digital forgeries from JPEG ghosts. IEEE Transactions on Information Forensics and Security, 4(1), 154–160.
- Popescu, A. C., & Farid, H. (2005). Exposing digital forgeries by detecting traces of resampling. IEEE Transactions on Signal Processing, 53(2), 758–767.
- Bianchi, T., & Piva, A. (2012). Image forgery localization via block-grained analysis of JPEG artifacts. IEEE Transactions on Information Forensics and Security.

ANEXO I: Generación de Datasets para el estudio comparativo

Imagen Nº 1

Prompt generado basado en Imagen real (imagen1.jpg): Tower Bridge London, photorealistic, outdoor cafe, many people (ethnicity:mixed, age:20-40), (detailed clothing:1.2), (accessories:1.1), (facial features:1.1), (expression:1.1), (body type:1.1), (pose:1.1) sitting and eating, (detailed skin texture:1.1), sitting at tables under Tower Bridge, (cafe setting:1.2), green hedges, landscaped area, modern cafe style, (detailed landscaping:1.2), bright sunny day, blue sky, light gray stone Tower Bridge, light-blue suspension bridge, (detailed architecture:1.2), full shot, center focus, slight low angle, (natural light:1.2), clear atmosphere, high detail, photorealistic style, 8k resolution, detailed, high detail digital art

Reales

Híbridas

Stable Difussion

	Dall-e
 Grok 3	 Fooocus

Imagen Nº 2

Prompt generado basado en Imagen real (imagen2.jpg): Big Ben, Houses of Parliament, London, UK, photorealistic, historic landmark, golden sandstone tower, detailed clock face, red double-decker bus, street scene, cars, pedestrians, clear blue sky, bright sunlight, mid-day light, city life, urban landscape, (architecture:1.3), (building details:1.2), (tower height:1.2), (classic bus:1.1), (vehicles:1.1), street level perspective, full shot, center composition, (bright colors:1.1), (sharp focus:1.2), detailed urban environment, photographic quality, daylight, detailed stonework, city streets, bus in foreground, tower in background, photorealistic style, (street scene detail:1.2)

 Reales	 Híbridas

Stable Difussion

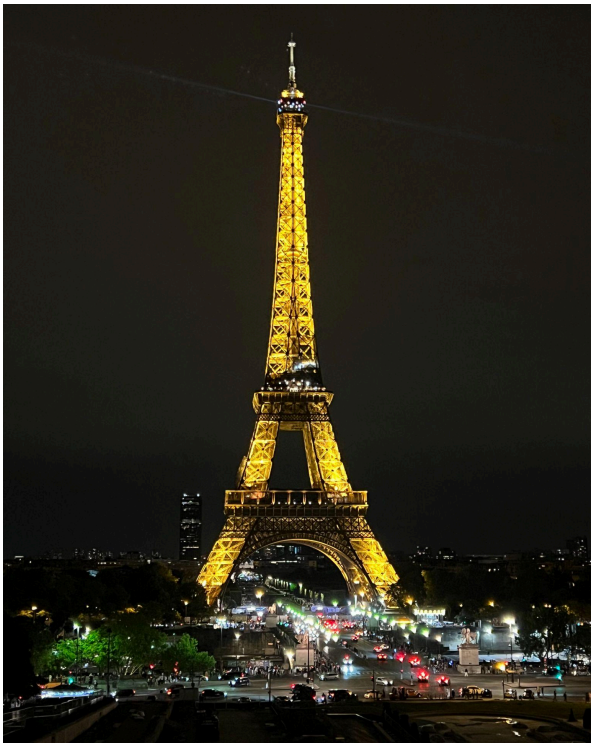
Dall-e

Grok 3

Fooocus

Imagen Nº 3

Prompt generado basado en Imagen real (imagen3.jpg): Eiffel Tower at night, golden light illuminating the structure, (detailed metalwork:1.3), (tower lights:1.2), city lights surrounding the base, low level shot, (cityscapes:1.2), street lights, cars, pedestrians, dark night sky, (night atmosphere:1.2), photorealistic, detailed city scene, (urban environment:1.3), Paris, France, (architecture:1.5), (perspective:1.2), wide angle shot, (composition:1.3), golden-yellow hues, warm lighting, high-detail, 8k resolution.



Reales

Híbridas

Stable Difussion

Dall-e

Grok 3

Fooocus

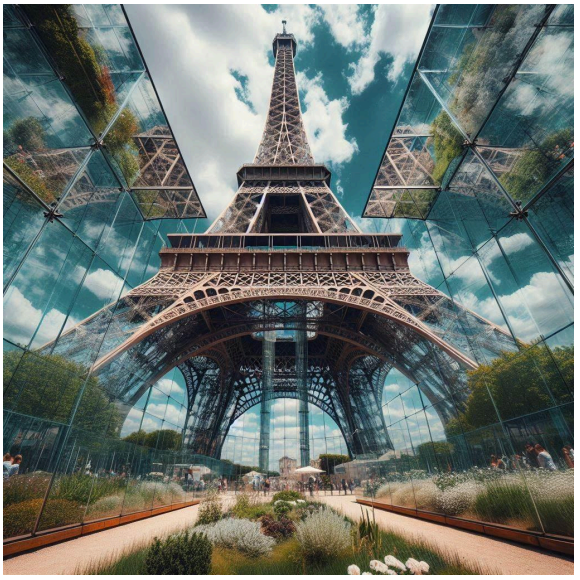
Imagen Nº 4

Prompt generado basado en Imagen real (imagen4.jpg): A low-angle view of the Eiffel Tower in Paris. The tower, a large, intricate structure of dark metal, is the central focus, taking up most of the upper portion of the image. A reflective glass barrier surrounds the base of the tower, with greenery and flowers visible through the glass. White flowers, bushes, and green grass fill the foreground. Some people are visible as small figures in the distance through the glass barrier. The sky is a vibrant blue with scattered white clouds. Sunlight illuminates the scene, casting highlights on the tower's structure. The overall atmosphere is one of a sunny, pleasant day in a major city. The composition is dominated by the tall, imposing tower, creating a sense of scale and awe. The perspective is looking upward from the base of the tower. Colors are predominantly blues, grays, and greens, with pops of white from the flowers.

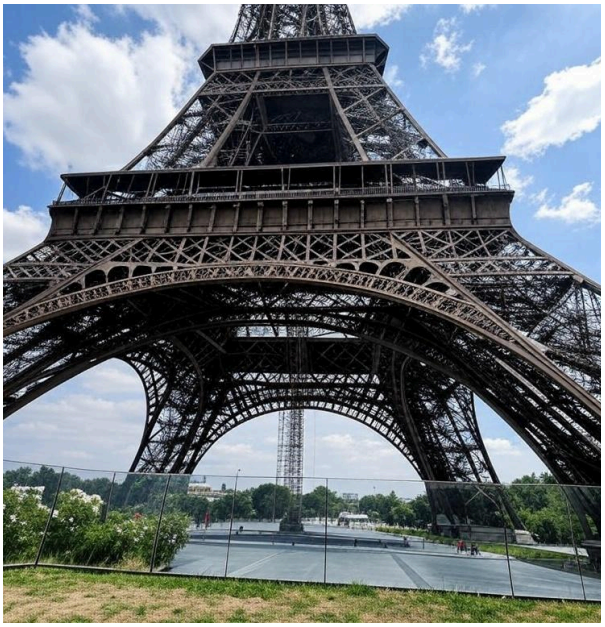
Reales

Híbridas

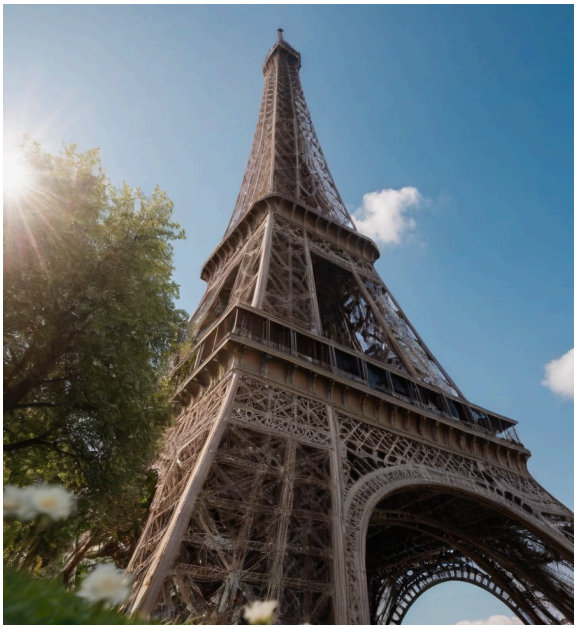
Stable Difussion



Dall-e



Grok 3



Fooocus

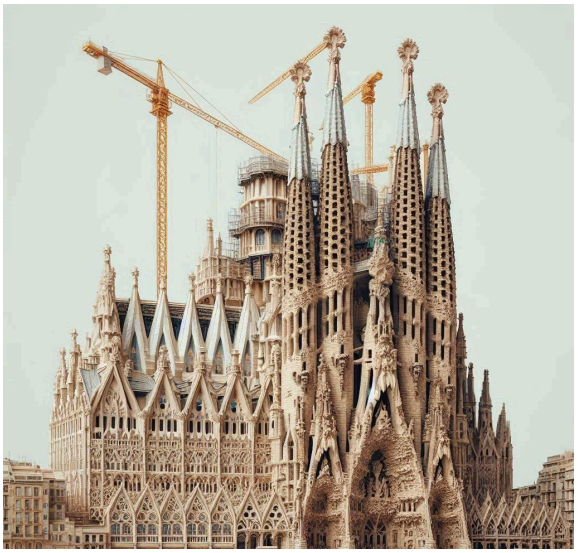
Imagen Nº 5

Prompt generado basado en Imagen real (imagen5.jpg): A large, ornate cathedral, the Sagrada Família in Barcelona, is the central subject. The cathedral is under construction, with scaffolding visible on various parts of the structure. Numerous tall, pointed towers and spires, adorned with intricate details and decorative elements, are prominent. The overall color palette is light beige, tan, and light gray, typical of stonework. A large construction crane is positioned above a section of the building, angled slightly to the upper right. The cathedral is situated within a landscaped urban area, with lush green trees and foliage forming a backdrop around the lower portion of the image. The perspective is from ground level, looking upward toward the cathedral. The sky is a clear, light blue. The composition is a full shot, showcasing the grandeur and scale of the structure. The scene conveys a sense of ongoing construction and architectural artistry.

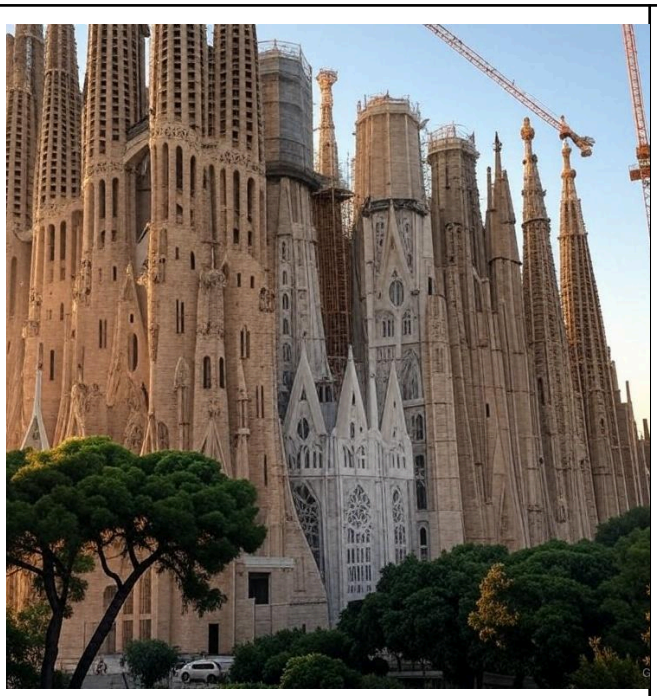
Reales

Híbridas

Stable Difussion



Dall-e



Grok 3

Fooocus

Imagen Nº 6

Prompt generado basado en Imagen real (imagen6.jpg): A wide shot of the Colosseum in Rome, Italy. The Colosseum, a large ancient amphitheater, is situated in the background, taking up a significant portion of the image's upper and middle area. The foreground shows a cobblestone street, filled with numerous tourists and sightseers walking around in various directions. Many people are wearing casual clothing in various colors and shades. Several children, including a girl wearing a light yellow shirt and light blue pants, are visible in the lower right quadrant, walking. The light is bright and sunny, casting shadows from the people and the Colosseum itself. The overall atmosphere is one of sightseeing and touristic activity. The composition is panoramic, showcasing a broad perspective of the area. The colors are primarily earth tones, grays, tans, and light blues of the sky. The perspective is from slightly above the cobblestone street.

Reales

Híbridas

Stable Difussion

Dall-e

Grok 3

Fooocus

Imagen Nº 7

Prompt generado basado en Imagen real (imagen7.jpg): A wide shot of the Monte Carlo Casino in Monaco. A large, ornate, light beige building with multiple levels and a clock tower is the central focus. The building's facade is detailed with statues and ornate architectural elements. Luxury cars, including a black Bentley and a red Ferrari, are parked in front of the casino's entrance. A group of people, dressed in suits and white shirts, are standing near the cars and the entrance steps. The scene is sunny, with a clear blue sky and bright sunlight. The cars are positioned in the foreground, centered. The casino building takes up most of the background, with the cars being in the lower section of the frame. The perspective is from the street level looking up towards the building entrance. The overall atmosphere is luxurious and upscale. The cars and building's architecture are the prominent features.

Reales

Híbridas

Stable Difussion

Dall-e

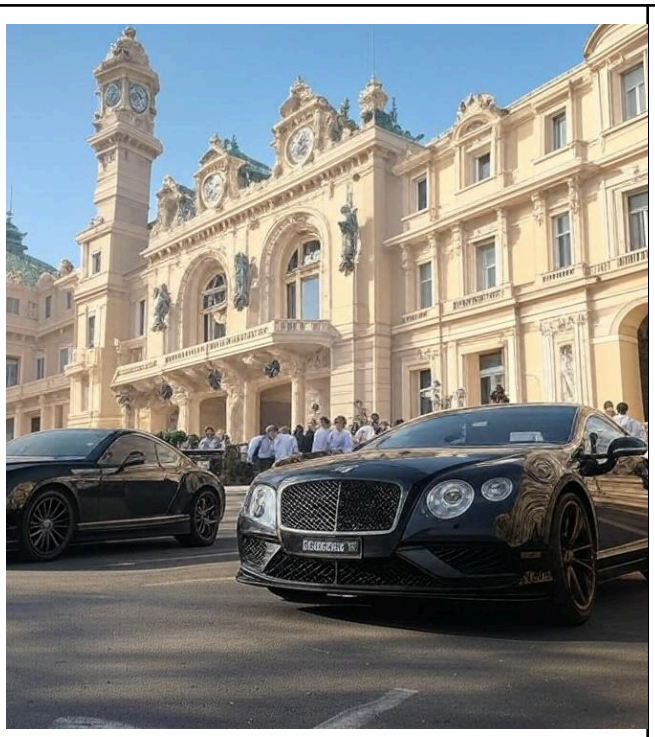
	
Grok 3	Fooocus

Imagen Nº 8

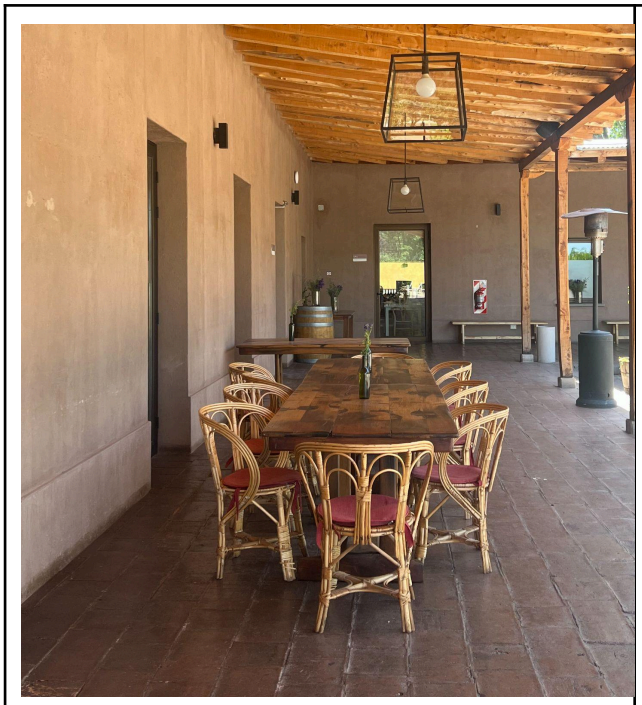
Prompt generado basado en Imagen real (imagen8.jpg): A wide-shot view of the Leaning Tower of Pisa and Piazza dei Miracoli. A light beige-white leaning tower, with numerous levels and decorative moldings, is the main subject, positioned slightly off-center to the left. A large, light gray cathedral-like building is visible behind and to the right of the tower. The foreground shows a grassy area where many people of various ages and ethnicities are sitting and socializing; some are relaxing and chatting, couples and families. Clothing styles are casual, ranging from t-shirts and shorts to sundresses and jeans. The lighting is bright and sunny, with the light source originating from above, casting soft shadows around the people on the lawn. The sky is a clear blue with scattered clouds. The perspective is from ground level, looking up at the tower and cathedral. The composition is balanced and inviting, with a focus on the historic buildings, people enjoying the day, and the vast open space. The overall atmosphere is one of relaxed tourism and enjoyment of a beautiful Italian landmark.

Reales	Híbridas
 Stable Difussion	 Dall-e
 Grok 3	 Fooocus

Imagen Nº 9

Prompt generado basado en Imagen real (imagen9.jpg): A long outdoor dining area under a wooden-beamed pergola. A large, rustic wooden table with woven rattan chairs, upholstered in reddish-brown fabric, sits in the center of the image, stretching almost the full length. The table is set for a meal, with a few bottles and vases on the surface. The walls are a light beige stucco. The floor is terracotta-colored tile. The pergola's wooden beams and supports are light brown. Several hanging light fixtures are simple, with black metal frames and clear glass. The background shows the entrance to a building or an outdoor space with similar light beige walls. The lighting is natural, bright, and evenly distributed, casting no significant shadows. The atmosphere is relaxed and inviting, suggesting a patio cafe or restaurant. The perspective is looking down the length of the patio towards the back of the structure. The composition is straightforward and balanced, with the dining set as the focal point. No people are visible, but the scene evokes a sense of leisurely dining outdoors.

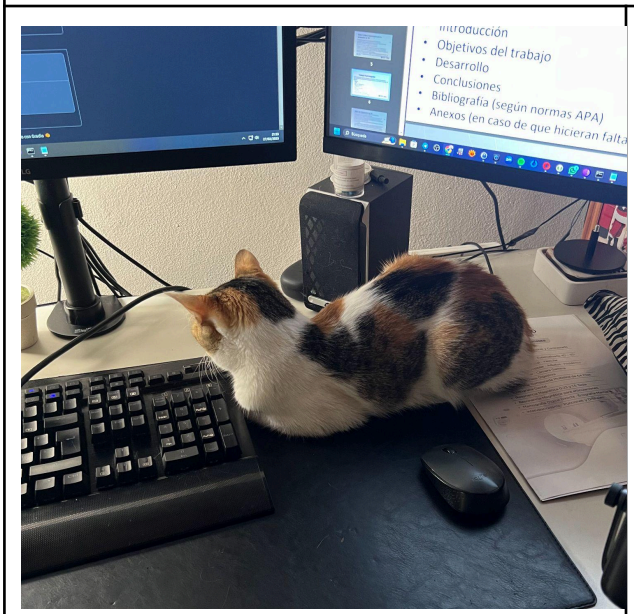
 Reales	 Híbridas
 Stable Difussion	 Dall-e

Grok 3

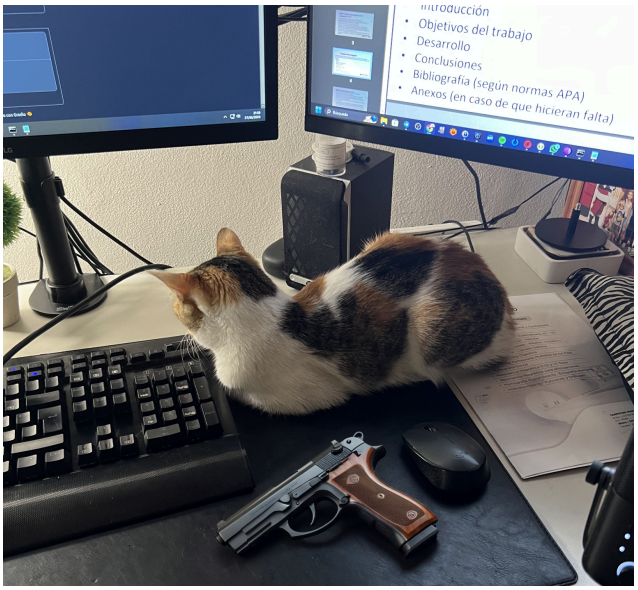
Fooocus

Imagen Nº 10

Prompt generado basado en Imagen real (imagen10.jpg): A calico cat is resting on a computer desk. The cat is a medium size with a mix of orange, white, and brown fur. It is positioned on the lower-middle portion of the image, laying on a black computer mousepad. The cat is facing slightly to its right, looking down, and its body is oriented towards the middle of the image. The cat appears relaxed and comfortable. The desk is white with two computer monitors, showing open document windows, and other office supplies. A black computer keyboard, a wireless mouse, and a microphone stand are also visible on the desk. The image's lighting is bright and even, casting no significant shadows on the desk. The overall atmosphere is calm and neutral, indicative of a home office setting. The perspective is directly above the desk's surface. The composition is straightforward and centered on the cat. The colors are primarily neutral tones of white, black, gray, and muted orange and brown. There are several documents on the desk and a small plant is partially visible in the background.

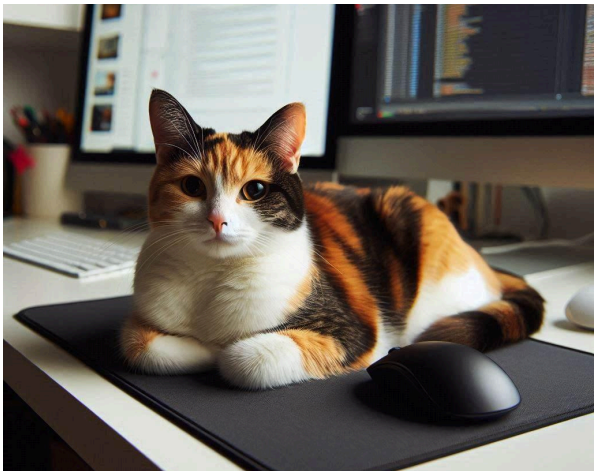


Reales

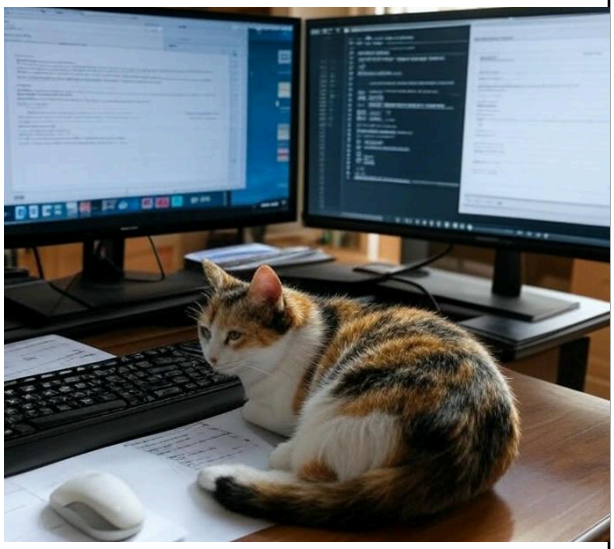


Híbridas

Stable Difussion



Dall-e



Grok 3

Fooocus