Interlínea
Cuadernos de investigación en arte, diseño y educación
ISSN 2530.4968

05 | octubre de 2025

PREMIOS MESTRE A LA INVESTIGACIÓN DOCENTE 2025
Segundo Premio Josep Albert Mestre

EVALUACIÓN AUTOMATIZADA DE IMÁGENES MEDIANTE INTELIGENCIA ARTIFICIAL: UNA HERRAMIENTA DE APOYO DOCENTE EN LA ENSEÑANZA DE LAS ARTES VISUALES Y EL DISEÑO

Soraya Paniagua Moro

Interlínea | 05 2025 | ISSN 2530.4968

Resumen. Este proyecto desarrolla una herramienta educativa basada en inteligencia artificial (IA) orientada a la evaluación automatizada de imágenes producidas por estudiantes en contextos de artes visuales y diseño. A partir de criterios definidos previamente por el docente, el sistema es capaz de analizar composiciones visuales generadas por el alumnado y ofrecer una valoración textual de carácter formativo, útil tanto para el docente como para el propio estudiante. El objetivo principal es proporcionar una solución de apoyo docente que permita agilizar, objetivar y enriquecer el proceso de corrección, sin sustituir el juicio profesional del profesorado. La herramienta no impone una evaluación cerrada, sino que actúa como asistente pedagógico configurable, ofreciendo sugerencias basadas en parámetros ajustables. El desarrollo se apoya en tecnologías de código abierto en lenguaje Python e incluye modelos de visión por computadora y procesamiento de lenguaje natural. El entorno ha sido implementado en Google Colab1, integrando bibliotecas como TensorFlow2, BLIP-23 y herramientas accesibles como Gradio4 para facilitar su uso, adaptabilidad y replicabilidad en contextos educativos reales.
La presente memoria describe la primera fase del proyecto, que se espera que durante los próximos años mejore en relación a la fiabilidad de los resultados y la experiencia del usuario en relación a un uso e instalación más sencillas para alguien ajeno a las tecnologías antes descritas.
Palabras clave: evaluación automatizada, Inteligencia Artificial, educación artística.

Interlínea es una publicación independiente que tiene por objeto difundir textos originales centrados en la convergencia del arte, el diseño y la educación. Interlínea es una publicación de la Confederación de Escuelas de Artes Plásticas y Diseño.

Interlínea, 2025. Creative Commons

Premio Mestre a la Investigación para la Mejora de la Calidad Educativa. Cuarta edición

Evaluación automatizada de imágenes mediante Inteligencia Artificial: una herramienta de apoyo docente en la enseñanza de las artes visuales y el diseño

Soraya Paniagua Moro

ÍNDICE

Breve resumen..... 3
Justificación..... 3
Objetivos del mismo..... 5
Objetivo general..... 5
Objetivos específicos:..... 5
Metodología..... 6
Fase de investigación..... 6
Fase de desarrollo técnico..... 6
Fase de validación..... 7
Fase de mejora..... 7
Recursos necesarios..... 8
Temporalización o periodos de realización..... 10
Diseño y planificación inicial..... 10
Desarrollo técnico del sistema..... 10
Prueba de predicción y validación inicial..... 11
Documentación del proceso y análisis pedagógico..... 12
Fases de mejora futura (propuesta)..... 12
Desarrollo y resultados..... 14
Resultados esperados..... 14
Criterios de evaluación seleccionados..... 14
Preparación del dataset..... 14
Entrenamiento de los modelos..... 15
Predicción con nuevas imágenes..... 16
Verificación de alineación con los objetivos..... 16
Criterios de evaluación..... 18
Criterios técnicos..... 18
Criterios pedagógicos..... 18
Criterios de funcionamiento del sistema..... 19
Criterios pedagógicos..... 19
Cuestionarios de satisfacción para docentes..... 20
Encuesta breve para estudiantes..... 20
Rúbrica de validación cruzada (IA vs. docente)..... 20
Análisis de logs y tiempos de corrección..... 21
Diario reflexivo o informe cualitativo del docente implementador..... 21
Indicadores de implementación organizativa..... 21
Posibilidades de generalización del trabajo a otros niveles y contextos educativos... 22

Posibilidades de ampliación y mejoras futuras.....22
Ampliaciones funcionales..... 22
Ampliaciones en inteligencia del sistema..... 23
Valor educativo generado..... 24
Otros aspectos a destacar del trabajo presentado..... 25
Originalidad e innovación educativa.....25
Arquitectura funcional del sistema.....25
Transparencia, explicabilidad y control docente.....26
Aplicabilidad real y sostenibilidad.....26
Nota final del autor..... 27

Breve resumen

Este proyecto desarrolla una herramienta educativa basada en **inteligencia artificial (IA)** orientada a la **evaluación automatizada de imágenes producidas por estudiantes** en contextos de artes visuales y diseño. A partir de **criterios definidos previamente por el docente**, el sistema es capaz de analizar composiciones visuales generadas por el alumnado y ofrecer una **valoración textual de carácter formativo**, útil tanto para el docente como para el propio estudiante.

El objetivo principal es proporcionar una **solución de apoyo docente** que permita **agilizar, objetivar y enriquecer el proceso de corrección**, sin sustituir el juicio profesional del profesorado. La herramienta no impone una evaluación cerrada, sino que actúa como **asistente pedagógico configurable**, ofreciendo sugerencias basadas en parámetros ajustables.

El desarrollo se apoya en tecnologías de código abierto en lenguaje Python e incluye modelos de visión por computadora y procesamiento de lenguaje natural. El entorno ha sido implementado en **Google Colab**[1], integrando bibliotecas como **TensorFlow**[2], **BLIP-2**[3] y herramientas accesibles como **Gradio**[4] para facilitar su uso, adaptabilidad y replicabilidad en contextos educativos reales.

La presente memoria describe la primera fase del proyecto, que se espera que durante los próximos años mejore en relación a la fiabilidad de los resultados y la experiencia del usuario en relación a un uso e instalación más sencillas para alguien ajeno a las tecnologías antes descritas.

Palabras clave: evaluación automatizada, Inteligencia Artificial, educación artística

Justificación

La evaluación en las enseñanzas artísticas supone un desafío particular, al requerir la valoración de dimensiones tanto **técnicas como expresivas y creativas**. Este proceso,

[1] Google Colab es una plataforma gratuita de Google que permite ejecutar código Python en la nube mediante notebooks Jupyter. Ofrece acceso a GPUs y TPUs sin configuración previa, ideal para aprendizaje automático y análisis de datos. Facilita la colaboración en tiempo real y la integración con Google Drive.

[2] TensorFlow es una biblioteca de código abierto desarrollada por Google para el aprendizaje automático y el desarrollo de redes neuronales. Permite construir y entrenar modelos de machine learning de manera eficiente, tanto en CPUs como en GPUs.

[3] BLIP-2 es un modelo multimodal desarrollado por Salesforce Research, diseñado para la comprensión y generación de lenguaje a partir de imágenes. Funciona mediante el uso de un codificador de imágenes estilo CLIP, un Transformador de Consultas (Q-Former) y un modelo de lenguaje grande.

[4] Gradio es una biblioteca ligera de Python que nos permite crear rápidamente aplicaciones de aprendizaje automático con interfaces basadas en la web

de gran importancia formativa, tiende a ser **subjetivo, heterogéneo y demandante en tiempo y carga mental**, especialmente en situaciones con **altos ratios de alumnado o múltiples grupos**. En este contexto, surge la necesidad de herramientas que ayuden a **sistematizar, objetivar y agilizar la evaluación sin deshumanizarla**.

La integración de la inteligencia artificial como **asistente evaluador** constituye una propuesta innovadora con potencial para:

- **Complementar la mirada crítica del docente**, reforzando el proceso con una capa adicional de análisis técnico.
- **Detectar patrones visuales objetivos o indicadores de calidad** difíciles de sistematizar manualmente.
- **Ofrecer retroalimentación formativa inmediata**, comprensible y adaptable para el estudiante.
- **Favorecer la replicabilidad** del sistema en distintas materias, niveles educativos y asignaturas de corte visual.

Este proyecto se justifica, además, por su alineación con los **objetivos estratégicos de digitalización educativa**, la mejora de la **competencia digital docente**, y la **exploración responsable de la IA en entornos pedagógicos reales**. En particular, en el ámbito del arte y el diseño —a menudo rezagado en la aplicación de estas tecnologías—, esta propuesta representa una oportunidad para **ampliar horizontes metodológicos** sin renunciar a la especificidad del lenguaje visual.

Objetivos del mismo

Objetivo general

Diseñar, desarrollar y validar una herramienta basada en inteligencia artificial que permita la **evaluación automatizada, configurable y formativa de imágenes** generadas por el alumnado en contextos de enseñanza artística, facilitando al profesorado un sistema de apoyo ágil, transparente y adaptable.

Objetivos específicos:

1. **Explorar el potencial actual de los modelos de inteligencia artificial multimodal** (imagen + texto) aplicados al análisis visual desde una perspectiva educativa.
2. **Diseñar una arquitectura técnica** que combine modelos de visión por computadora y generación de lenguaje natural para la interpretación y retroalimentación sobre imágenes.
3. **Desarrollar una interfaz accesible y funcional** que permita al profesorado definir y personalizar criterios de evaluación sin necesidad de conocimientos de programación.
4. **Probar y validar el sistema en un entorno real** de Escuela de Arte, recopilando valoraciones de docentes y estudiantes sobre su usabilidad, precisión y utilidad pedagógica (en proceso).
5. **Analizar la viabilidad de generalización del sistema** a otros niveles, contextos o especialidades dentro de la educación artística (en proceso).

Metodología

El desarrollo del proyecto ha seguido una metodología estructurada en cuatro fases, combinando investigación teórica, desarrollo técnico, experimentación práctica y análisis de resultados.

Fase de investigación

- Revisión del estado del arte sobre evaluación en educación artística y experiencias previas con inteligencia artificial (IA) en el ámbito educativo.
- Estudio de modelos de IA aplicados a la visión por computadora y la generación de lenguaje natural (IA multimodal).
- Identificación de los criterios de evaluación más frecuentes en disciplinas como el dibujo artístico, el diseño gráfico y la ilustración, prestando especial atención a factores visuales como la legibilidad o el contraste.

Fase de desarrollo técnico

- Primera exploración con modelos generativos tipo BLIP-2 y LLMs[5] (como GPT) para generar descripciones de imágenes y realizar inferencias textuales.
- Rediseño del enfoque hacia el entrenamiento de modelos de clasificación visual (MobileNetV2[6]) directamente con datasets etiquetados por criterio.
- Implementación de un entorno de desarrollo en Google Colab, integrando código en Python con librerías como TensorFlow, Keras[7] y PIL[8].
- Automatización de procesos como la limpieza de datos, el padding de imágenes, el entrenamiento por criterios y la exportación de resultados (a CSV, TXT o Google Sheets).
- Diseño de una arquitectura modular que permite personalizar los criterios sin necesidad de modificar el código base.

[5] LLM, o Large Language Model (Modelo de Lenguaje de Gran Escala en español), es un tipo de inteligencia artificial diseñado para comprender y generar texto en lenguaje natural. Estos modelos se entrenan con grandes cantidades de datos de texto y son capaces de realizar diversas tareas como responder preguntas, traducir idiomas, crear texto, entre otros, con un alto nivel de precisión y fluidez

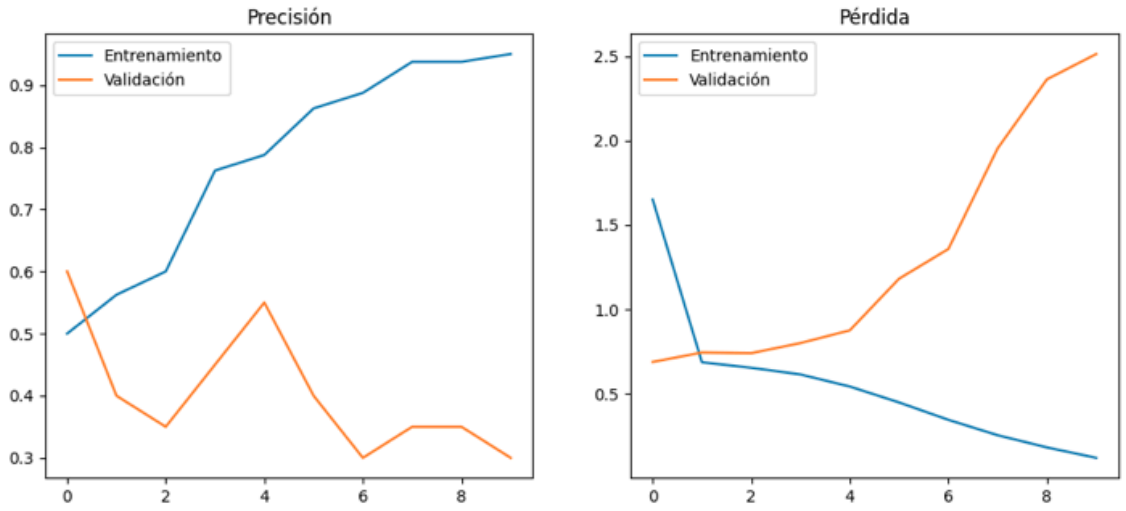
[6] MobileNetV2 es una red neuronal profunda ligera utilizada principalmente para la clasificación de imágenes.

[7] Keras es una biblioteca de código abierto para aprendizaje profundo (deep learning) escrita en Python. Actúa como una interfaz de alto nivel para redes neuronales, permitiendo a los desarrolladores crear y entrenar modelos de manera más rápida y sencilla, abstraíendolos de las complejidades de

[8] Python, una imagen en formato PIL (Python Imaging Library) se refiere a una imagen que puede ser manipulada y procesada utilizando la biblioteca Pillow, que es una bifurcación de PIL.

Fase de validación

- Pruebas preliminares con docentes para verificar la utilidad y comprensión del sistema.
- Evaluación del comportamiento del modelo frente a imágenes nuevas, analizando la coherencia de sus predicciones.
- Comparación entre las valoraciones generadas por la IA y el juicio profesional del docente.
- Revisión de la capacidad del sistema para exportar informes comprensibles y útiles pedagógicamente.

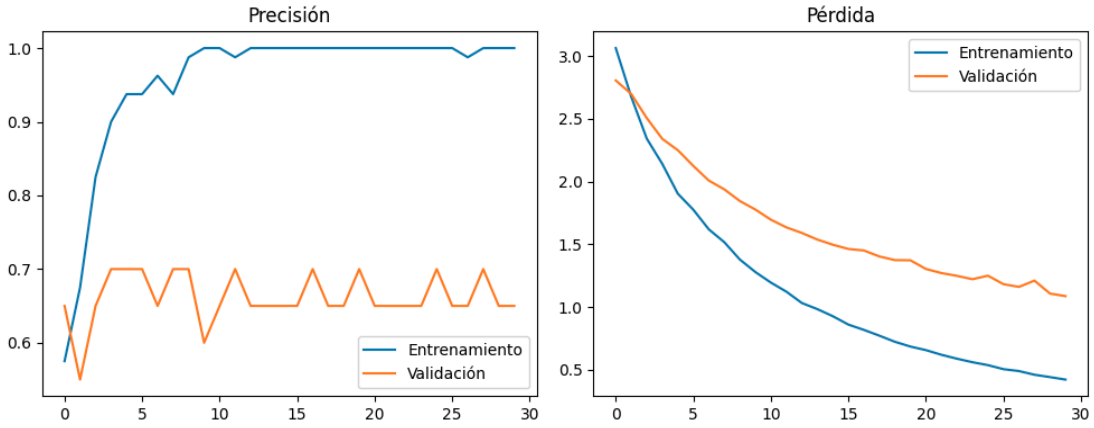


Gráficas de precisión y pérdida respecto al entrenamiento y su validación con CNN

Fase de mejora

- Ajustes en los modelos mediante técnicas de regularización (Dropout[9], L2), aumentación de datos y entrenamiento con MobileNetV2 como base preentrenada.
- Reentrenamiento con datasets ampliados (de 50 a 100 imágenes por clase) para mejorar la generalización del sistema.
- Mejoras en la estructura del código para facilitar la extensión a nuevos criterios visuales (espaciado, tamaño de fuente, alineación, etc.).
- Propuesta de futuras integraciones con plataformas educativas y mejoras en la interfaz.

[9] En el contexto de redes neuronales, dropout es una técnica de regularización que ayuda a prevenir el sobreajuste (overfitting) al desactivar aleatoriamente un conjunto de neuronas durante el entrenamiento



Gráficas de precisión y pérdida respecto al entrenamiento y su validación con MobileNetV2

Recursos necesarios

Para la implementación efectiva de este proyecto de evaluación automatizada de imágenes en educación artística mediante inteligencia artificial, se han empleado y estimado los siguientes recursos, clasificados por categoría:

Tipo de recurso	Descripción
Hardware	Entorno en la nube basado en Google Colab, sin GPU dedicada, lo que ha condicionado el tiempo de entrenamiento, pero ha permitido desarrollar el prototipo sin necesidad de infraestructuras locales potentes. Integración con Google Drive para almacenamiento de modelos, métricas y resultados[10].
Software	Lenguaje de programación Python junto con librerías especializadas: TensorFlow y Keras para el entrenamiento de modelos, PIL y OpenCV para el tratamiento de imágenes, Pandas y Matplotlib[11] para análisis y visualización de resultados. Se ha trabajado con el modelo preentrenado MobileNetV2[12]. Para la interfaz de usuario y pruebas preliminares se consideró el uso de Gradio. En fases anteriores, se exploró el uso de BLIP-2 y modelos LLM para generar descripciones automáticas.

[10] Cabe resaltar el impacto positivo de haber utilizado recursos gratuitos (Colab, Drive) como una fortaleza para la escalabilidad del sistema.

[11] Matplotlib es una biblioteca de Python para crear visualizaciones de datos de alta calidad. Permite generar una amplia variedad de gráficos, como diagramas de barras, histogramas, diagramas de dispersión y más, con gran flexibilidad y personalización.

[12] En la primera versión del proyecto se usó una arquitectura CNN sencilla (Convolutional Neural Network) creada desde cero con keras.Sequential(). Esta arquitectura tenía varias capas Conv2D, MaxPooling2D, y Dense, y se entrenó para predecir los criterios de legibilidad o contraste. Tras la prueba de concepto, se migró a un modelo preentrenado para ganar robustez y evitar entrenar desde cero.

Datos	Conjuntos de imágenes educativas organizadas por clases (legibilidad alta/baja y contraste alto/bajo). Estas imágenes fueron recopiladas, adaptadas y preprocesadas con técnicas como padding, limpieza y aumentación de datos. Se consideró necesario aumentar el número por clase a más de 100 para una mayor robustez.
Recursos humanos	Participación de docentes de áreas artísticas en la definición de criterios de evaluación y validación de resultados. Alumnado voluntario para generar imágenes de prueba y retroalimentar el funcionamiento del sistema. Asesoramiento externo en IA educativa para evaluar enfoques viables.
Conectividad y almacenamiento	Integración con Google Drive para persistencia de modelos, exportación de resultados y colaboración. Exportación automatizada a Google Sheets y archivos descargables en .csv, .txt y futuras versiones en .pdf.
Tiempo estimado	Aproximadamente *90 días* efectivos para el desarrollo técnico, pruebas iterativas y entrenamientos múltiples. Adicionalmente, se han destinado sesiones específicas para validación con docentes, análisis pedagógico y documentación final.

Temporalización o periodos de realización

Este proyecto se ha desarrollado durante el segundo y tercer trimestre del curso escolar, iniciándose en enero. A continuación se detalla la planificación seguida, distribuida por fases, con objetivos, actividades y resultados esperados o alcanzados.

Diseño y planificación inicial

Objetivos

- Definir el alcance y propósito general del proyecto.
- Identificar asignaturas piloto y profesorado implicado.
- Recoger y consensuar criterios de evaluación aplicables a producciones artísticas.

Actividades

- Revisión bibliográfica sobre evaluación en educación artística e IA educativa.
- Reuniones preliminares con docentes de diferentes disciplinas.
- Estudio de herramientas IA accesibles (Gradio, BLIP, modelos LLM), con una primera hipótesis basada en generación de lenguaje natural a partir de imagen.
- Primer esbozo del flujo de trabajo deseado (entrada de imágenes, análisis y generación de informe).

Resultados esperados

- Documento base del proyecto con objetivos y justificación.
- Plan técnico inicial con propuesta de arquitectura, modelos y tecnologías a explorar.

Desarrollo técnico del sistema

Objetivos

- Establecer los criterios de evaluación.
- Evaluar alternativas metodológicas con redes neuronales frente a modelos generativos.
- Preparar los datasets y entrenar modelos para cada criterio.
- Implementar limpieza y preprocesamiento de imágenes.
- Diseñar una arquitectura de entrenamiento robusta en Colab.

Actividades

- Creación de datasets por clases con imágenes propias y de ejemplo.
- Aplicación de técnicas de padding, normalización y data augmentation.
- Entrenamiento de modelos[13] CNN y luego MobileNetV2 con validación cruzada.
- Implementación de regularización para mitigar el sobreajuste.

Resultados

- Criterios seleccionados: legibilidad y contraste.
- Modelos funcionales entrenados y exportados (.keras).
- Planificación del nuevo sistema basado en MobileNetV2.
- Código optimizado en Google Colab.
- Guardado automático de métricas y modelos en Google Drive.

Prueba de predicción y validación inicial

Objetivos

- Evaluar la capacidad de los modelos con nuevas imágenes reales.
- Analizar los resultados y definir márgenes de mejora.

Actividades

- Creación de carpeta con nuevas imágenes para testeo[14].
- Ejecución de script de predicción automático.
- Generación de CSV, TXT y exportación a Google Sheets.
- Interpretación de resultados por criterio.
- Evaluación de problemas de escalabilidad y precisión.

Resultados

- Predicción acertada en contraste y parcialmente en legibilidad.
- Verificación de robustez básica del modelo.

[13] Tras pruebas iniciales, se descartó el enfoque con BLIP-2 debido a limitaciones para clasificación masiva, optando por modelos CNN entrenables como MobileNetV2.

[14] Las imágenes utilizadas no formaban parte del conjunto de entrenamiento, lo que permitió evaluar la capacidad de generalización del modelo.

- Identificación de posibles mejoras: dataset, regularización, feedback textual.

Documentación del proceso y análisis pedagógico

Objetivos

- Elaborar la memoria del proyecto con todos los avances y código.
- Integrar mejoras sugeridas y vinculación pedagógica.

Actividades

- Redacción de epígrafes técnicos, metodológicos y educativos.
- Vinculación de resultados con los objetivos planteados.
- Evaluación pedagógica de la utilidad del sistema.

Resultados

- Memoria completa con comentarios y mejoras aplicadas.
- Tabla de criterios de evaluación organizados (técnicos, pedagógicos, organizativos).
- Valoración de los modelos en contexto educativo realista.

Fases de mejora futura (propuesta)

Objetivos

- Ampliar el sistema hacia una versión más robusta y usable.
- Implementar interfaz para el docente.

Actividades propuestas

- Aumento del dataset a 100+ imágenes por clase.
- Implementación de una interfaz Gradio[15] para subir imágenes, ver predicciones y exportar.
- Entrenamiento con dataset real de aula (legibilidad de trabajos reales).

Resultados esperados

- Herramienta completa lista para piloto docente real.

[15] Se proyecta la implementación de una interfaz Gradio que permita a cualquier docente subir imágenes, obtener predicciones y descargar informes sin necesidad de conocimientos técnicos.

- Mayor precisión y detalle de los resultados.
- Uso autónomo por parte de docentes no programadores.

Desarrollo y resultados

Resultados esperados

- Prototipo funcional de una herramienta de evaluación automatizada de imágenes.
- Documento técnico con la arquitectura, código y modelos utilizados.
- Guía de uso para docentes.
- Informe de validación y análisis de impacto pedagógico.
- Posibilidad de replicación y ampliación del sistema en otras áreas artísticas (fotografía, animación, diseño digital, etc.).

El proyecto comenzó con un enfoque centrado en el uso de modelos de lenguaje como BLIP para generar descripciones automáticas a partir de imágenes. A partir de estas descripciones, se aplicaban prompts dirigidos para obtener evaluaciones cualitativas sobre aspectos como la legibilidad o el contraste. Sin embargo, este enfoque presentaba limitaciones a nivel técnico, especialmente en la evaluación masiva y objetiva de imágenes.

Comentario: *Se ha mantenido la idea del enfoque inicial para mostrar evolución metodológica.*

Por ello, se rediseñó la estrategia del proyecto hacia una solución basada en modelos de visión por computadora entrenados directamente con imágenes clasificadas. Esta nueva dirección permitió una evaluación automática, objetiva y escalable de los criterios seleccionados.

Criterios de evaluación seleccionados

Se seleccionaron dos criterios visuales fundamentales para la comprensión lectora en contextos educativos:

- Legibilidad: imágenes con texto fácil o difícil de leer
- Contraste: imágenes con alto o bajo contraste visual entre el texto y el fondo.

Comentario: *Se pueden añadir otros criterios si el docente decide ampliar el proyecto (espaciado, tamaño de letra, etc.)*

Preparación del dataset

Se crearon dos conjuntos de imágenes organizadas en carpetas según su clase (ej. *legibles/ilegibles, alto_contraste/bajo_contraste*).
Cada conjunto constaba inicialmente de unas 50 imágenes por clase. Posteriormente

se recomendó aumentarlo a 100 o más por clase para mejorar la robustez del entrenamiento.

Antes del entrenamiento, se aplicaron los siguientes pasos de preprocesamiento:

- Verificación de validez y formato de las imágenes.
- Eliminación de imágenes corruptas o con extensiones no válidas.
- Conversión a formato cuadrado con relleno negro (padding).
- Redimensionado a 256x256 píxeles.
- Normalización y aumentación de datos (flip horizontal, rotaciones, etc.).

Comentario: *Se han descrito los pasos reales aplicados en el código que desarrollamos.*

Entrenamiento de los modelos

Para cada criterio se entrenó un modelo independiente utilizando MobileNetV2 como base pre-entrenada sobre **ImageNet**[16], con capas densas adicionales y una salida binaria (sigmoid) para clasificación.

Esto se realizó en Google Colab con conexión a Google Drive, permitiendo guardar modelos, métricas y resúmenes.

Los modelos fueron entrenados durante 30 épocas, con validación sobre el 20% del dataset. Se evaluaron las métricas de pérdida (loss) y precisión (accuracy) tanto en entrenamiento como en validación.

Ejemplo de resultados obtenidos

Criterio	Precisión (val)	Pérdida (val)
Legibilidad	0.65	1.0879
Contraste	0.7368	0.9975

Comentario: *Se pueden actualizar los valores con más entrenamientos o mejoras el dataset.*

A pesar de los buenos resultados en entrenamiento, se detectó overfitting en algunos casos (precisión muy alta en entrenamiento, pero baja en validación). Para mitigar esto se aplicaron técnicas de:

- Aumentación de datos.
- Regularización con Dropout y L2.

[16] ImageNet es un enorme conjunto de datos de acceso público con más de 14 millones de imágenes que han sido anotadas a mano para indicar qué objetos representan.

- Congelación de pesos del modelo base.

Predicción con nuevas imágenes

Se creó un script para evaluar automáticamente nuevas imágenes cargadas en una carpeta. Este sistema aplicaba el modelo entrenado correspondiente a cada criterio, devolviendo la probabilidad de que la imagen pertenezca a cada clase.

Se probaron tres imágenes reales, obteniendo predicciones satisfactorias en contraste, y más variables en legibilidad, lo que confirma que el modelo puede generalizar, pero que la calidad y cantidad del dataset son clave para mejorar la fiabilidad.

Los resultados se exportaron automáticamente a:

- CSV descargable.
- Informe en formato .txt.
- Hoja de cálculo de Google Sheets.
- Gráficas resumen con matplotlib.

Comentario: *Se podría enriquecer esta sección incluyendo un ejemplo de interpretación pedagógica del informe generado, mostrando cómo un docente podría utilizarlo como base de retroalimentación formativa para el alumno.*

Verificación de alineación con los objetivos

Objetivos definidos

- Diseñar e implementar un sistema IA para evaluar imágenes educativas.
- Investigar IA multimodal (imagen + texto).
- Crear una interfaz configurable por el docente.
- Validar el sistema con docentes y estudiantes.

Resultados obtenidos

- Se ha implementado un sistema funcional con MobileNetV2.
- Se ha trabajado con datasets propios y preprocesamiento avanzado.
- Se han generado informes y visualizaciones de resultados.
- Se han entrenado modelos por criterio (legibilidad, contraste).
- Se ha desarrollado un script de predicción para imágenes nuevas.

Aunque el sistema es funcional, no se documenta si se ha realizado una validación formal con docentes y estudiantes (objetivo 4). Sería recomendable mencionar si ha habido validación informal o si está prevista para fases futuras.

Aunque se han conseguido resultados técnicos prometedores, la validación pedagógica (por parte de docentes y estudiantes) está todavía pendiente de implementación. Se recomienda incluir esta fase como parte del desarrollo futuro del sistema para evaluar su aceptación y aplicabilidad real en el aula.

A nivel técnico, todos los objetivos se han cumplido o se encuentran muy avanzados. A nivel pedagógico, se recomienda en fases futuras realizar pruebas controladas con usuarios reales y recoger evidencias cuantitativas de impacto educativo.

Criterios de evaluación

La evaluación de este proyecto se ha estructurado en tres bloques: **criterios técnicos**, **criterios pedagógicos, de funcionamiento del sistema** y **criterios organizativos y de gestión**. Cada criterio se acompaña de indicadores observables y medibles que permitirán valorar su grado de cumplimiento.

Criterios técnicos

Criterio	Indicador
Utilidad docente	Los docentes indican que la herramienta mejora la eficiencia en la evaluación[17].
Utilidad para el estudiante	El feedback generado por el sistema es claro, comprensible y formativo para el alumnado.
Coherencia pedagógica	Las valoraciones generadas por la IA son coherentes con los criterios educativos del aula.
Mejora del proceso	Se evidencia una reducción del tiempo de corrección y una mejora en la retroalimentación ofrecida.

Criterios pedagógicos

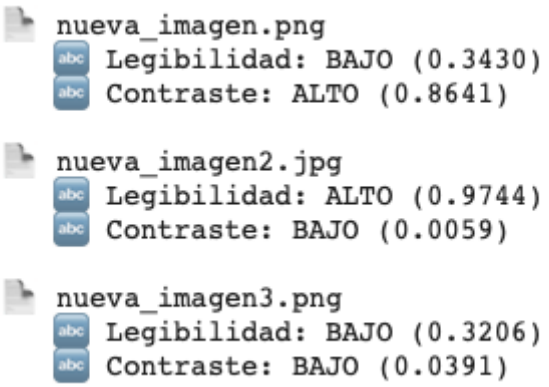
Criterio	Indicador
Participación del profesorado	Al menos dos asignaturas distintas incorporan el sistema en su evaluación.
Planificación cumplida	El desarrollo del proyecto se ajusta al cronograma establecido inicialmente.
Documentación generada	Se dispone de manuales de uso, informes técnicos y propuestas de continuidad.

[17] Esta información se recoge mediante encuestas a los profesores.

Transferencia	El sistema demuestra ser aplicable a otros contextos educativos similares (niveles, áreas, centros).

Criterios de funcionamiento del sistema

Criterio	Indicador
Funcionamiento del sistema	El sistema genera informes automáticos coherentes y configurables. Estos pueden ser visuales o descargables en varios formatos.
Precisión de análisis visual	El modelo reconoce y describe correctamente elementos relevantes en las imágenes con un porcentaje de acierto superior al 70% en pruebas de validación.
Personalización de criterios	El docente puede modificar o añadir criterios sin editar código[18].
Accesibilidad	La herramienta se ejecuta en la nube y es accesible desde distintos dispositivos (PC, móvil, tablet).



Predicción en imágenes nuevas con los modelos de legibilidad y contraste

Criterios pedagógicos

A continuación se detallan los mecanismos de evaluación para comprobar el funcionamiento del sistema

[18] Este actualmente es un objetivo de futuro porque no se ha desarrollado aún la interfaz.

Cuestionarios de satisfacción para docentes

- **Objetivo**: Valorar la utilidad percibida del sistema, facilidad de uso y fiabilidad.
- **Tipo**: Escala tipo Likert (1–5) con ítems como:
    - "La herramienta ha reducido el tiempo que dedico a la evaluación."
    - "La retroalimentación generada por la IA se alinea con mis criterios docentes."
- **Indicadores cuantitativos vinculados**
    - ≥80% de respuestas con puntuaciones ≥4 indica alta satisfacción.
    - Tiempo medio de corrección disminuye ≥30% respecto al método manual.

Encuesta breve para estudiantes

- **Objetivo**: Valorar la claridad y utilidad del feedback recibido.
- **Tipo**: Cuantitativa + cualitativa. Por ejemplo:
    - "¿Comprendiste por qué tu imagen obtuvo una puntuación baja en contraste?"
    - "¿Te ayudó el informe de IA a mejorar tu trabajo?"
- **Indicadores cuantitativos**
    - ≥70% de estudiantes afirma haber mejorado tras recibir retroalimentación de la IA.
    - ≥80% declara entender los criterios de evaluación tras usar la herramienta.

Rúbrica de validación cruzada (IA vs. docente)

- **Objetivo:** Evaluar la **precisión y coherencia** del sistema respecto al juicio humano.
- **Método:** El docente etiqueta una muestra de trabajos (por ejemplo, 20 imágenes) y se compara con la salida de la IA.
- **Indicadores cuantitativos**
    - ≥85% de coincidencia en etiquetas ALTO/BAJO entre IA y docente.
    - Discrepancia ≤15% entre las valoraciones para considerarse aceptable.

Análisis de logs y tiempos de corrección

- **Objetivo**: Medir de forma objetiva la mejora del proceso.
- **Método**: Se registran automáticamente los tiempos de ejecución del modelo y la generación de informes.
- **Indicadores cuantitativos**
    - Tiempo medio por imagen ≤2 segundos.
    - Reducción ≥40% en el tiempo total invertido en la evaluación de trabajos visuales.

Diario reflexivo o informe cualitativo del docente implementador

- **Objetivo**: Recoger observaciones no cuantificables: dificultades, adaptaciones, cambios propuestos.
- **Indicadores indirectos**
    - Se identifican al menos 2 propuestas de mejora para futuras versiones.
    - El docente valora positivamente (≥4/5) la experiencia de innovación metodológica.

Indicadores de implementación organizativa

- **Método**: Observación directa y análisis documental.
- **Criterios evaluables**
    - Participan ≥2 asignaturas o niveles educativos.
    - Se generan al menos 3 documentos: manual, informe técnico y propuesta de transferencia.
    - El sistema se reutiliza en el curso siguiente → indicador de sostenibilidad.

Posibilidades de generalización del trabajo a otros niveles y contextos educativos.

El sistema desarrollado en este proyecto —basado en visión por computadora y modelos de IA entrenados con criterios docentes— presenta una alta escalabilidad y adaptabilidad a otros niveles educativos, especialidades y contextos de evaluación. Su estructura modular, su implementación en Google Colab (accesible sin instalación local) y el uso de redes neuronales personalizables, lo convierten en una herramienta flexible para extender su aplicación más allá del contexto inicial.

Aplicación en otros niveles educativos

- Educación Secundaria Obligatoria y Bachillerato artístico: Adaptando los criterios al nivel competencial del alumnado, es posible evaluar ejercicios de dibujo técnico, composiciones gráficas o tareas visuales.
- Formación Profesional: Especialmente útil en módulos de imagen y sonido, diseño gráfico, fotografía o producción audiovisual.
- Educación Superior y Escuelas de Arte: Permite automatizar evaluaciones formativas en asignaturas como tipografía, composición visual, maquetación, diseño editorial o ilustración digital.

Adaptabilidad a otros contextos

- Evaluación formativa continua: Se puede emplear para generar informes periódicos que documenten el progreso del estudiante.
- Autoevaluación guiada: El alumnado puede utilizar la herramienta para analizar sus propias producciones antes de la entrega final.
- Evaluación por competencias: Puede integrarse en rúbricas competenciales con descriptores visuales.

Posibilidades de ampliación y mejoras futuras

Se plantean a continuación algunas líneas de mejora (ampliaciones) que permitirían evolucionar el sistema hacia un modelo más completo, inclusivo y sostenible en el tiempo:

Ampliaciones funcionales

Mejora	Descripción	Dificultad

Exportación a PDF	Posibilidad de generar un informe descargable en formato profesional.	Media
Histórico de evaluaciones	Guardar informes por alumno, fecha y asignatura, con visualización de progreso.	Alta
Integración con plataformas educativas	Enlace directo con Moodle, Classroom o similares para importar listas de clase y exportar notas.	Alta
Evaluación multilingüe	Posibilidad de configurar la interfaz y los informes en varios idiomas.	Baja
Generación de feedback con GPT (vía API)	Usar modelos GPT para generar observaciones escritas con tono personalizado.	Media
Avatar animado y síntesis de voz	Facilitar la accesibilidad mediante una IA parlante y expresiva.	Alta
Interfaz docente con Gradio	Desarrollo de una interfaz sencilla y visual que permita a los docentes subir imágenes, seleccionar criterios y obtener un informe inmediato. Facilita su uso en el aula sin necesidad de conocimientos técnicos.	Media

La dificultad en todos los casos es una estimación

Ampliaciones en inteligencia del sistema

Línea de mejora	Descripción	Valor educativo
Entrenamiento con dataset local	Entrenar modelos con producciones reales del centro para mayor adecuación.	Elevado
Rúbricas adaptativas por nivel o ciclo	Permitir definir escalas distintas de valoración según la asignatura o curso.	Alto
Reconocimiento multimodal (imagen + texto)	Añadir OCR para evaluar presentaciones que mezclen imágenes con títulos, rótulos o descripciones.	Medio

Valor educativo generado

Este proyecto abre la puerta a una nueva forma de evaluación visual asistida por IA, permitiendo al docente ahorrar tiempo, ganar objetividad y ofrecer retroalimentación inmediata y coherente. Su capacidad de escalarse a nuevos niveles, lenguas y contextos, lo convierte en un modelo transferible no solo dentro de una escuela de arte, sino en cualquier entorno educativo donde se trabajen producciones visuales: desde la educación artística básica hasta el diseño profesional.

Otros aspectos a destacar del trabajo presentado

Originalidad e innovación educativa

Este proyecto constituye una **propuesta innovadora y singular en el ámbito de la enseñanza artística**, al situar la inteligencia artificial como un **agente de apoyo pedagógico en procesos de evaluación visual**, un campo en el que su aplicación es aún incipiente.

A diferencia de los usos más habituales de la IA como generadora de contenido (imágenes, texto, música, etc.), aquí se le otorga un **rol evaluador** mediante el análisis de producciones estudiantiles con base en **criterios definidos por docentes reales**. Esta función no sustituye la mirada del profesorado, sino que la **refuerza y sistematiza**, brindando una nueva forma de retroalimentación que es más ágil, objetiva y personalizable.

El enfoque del proyecto **integra lo humano y lo artificial en una sinergia inteligente**, donde la tecnología se adapta al contexto educativo sin imponer lógicas externas. La herramienta se concibe como un **asistente docente configurable**, no como un evaluador cerrado.

Arquitectura funcional del sistema

El sistema ha sido diseñado con una lógica modular, intuitiva y abierta, que permite su comprensión y reutilización por parte de equipos docentes, incluso sin formación técnica avanzada.

A continuación, se describe la arquitectura básica del flujo de funcionamiento:

El estudiante o docente sube una imagen artística

↓

La IA analiza la imagen usando redes neuronales entrenadas

↓

Se evalúa según los criterios definidos: contraste, legibilidad, etc.

↓

Se genera un informe textual o numérico con valoración o sugerencias

↓

El docente puede revisar, adaptar o completar el informe generado

Este proceso es **totalmente configurable y extensible**: permite añadir nuevos criterios, nuevos modelos, más ejemplos de entrenamiento e incluso cambiar el idioma, el formato del informe o su integración con otras plataformas educativas.

Transparencia, explicabilidad y control docente

Otro elemento destacable es que el sistema ha sido desarrollado con enfoque de transparencia:

- El docente define los criterios y puede entrenar sus propios modelos.
- El código está comentado, documentado y accesible desde Google Colab.
- Se generan informes con puntuaciones comprensibles para alumnado y profesorado.
- Se evita la "caja negra" típica de muchos sistemas automáticos, facilitando la comprensión del proceso de evaluación.

Aplicabilidad real y sostenibilidad

Durante el desarrollo del proyecto se ha demostrado que el sistema es funcional y útil con recursos mínimos, sin necesidad de infraestructuras complejas. Esta característica lo hace especialmente valioso para centros con recursos limitados, permitiendo:

- Uso en Chromebooks o equipos antiguos vía Colab.
- Acceso multiplataforma (PC, tablet, móvil).
- Entrenamiento y predicción con datasets reducidos y locales.

Además, su diseño está orientado a una implementación sostenible y evolutiva, donde cada centro puede adaptar los modelos a su contexto real de aula.

Nota final del autor

INQUIETUD QUE VIENE DE ATRÁS

Este proyecto es resultado de una inquietud que tuve como estudiante cuando me preguntaba cómo los profesores evaluaban realmente mi trabajo. Luego como docente, pese al uso de rúbricas y criterios cuantificables, y bajo los mismos criterios de evaluación, a veces se observan discrepancias en la evaluación entre docentes. Por ejemplo en los proyectos integrados.

DESEO DE TENER UNA HERRAMIENTA

El proyecto también es mi deseo de tener una herramienta de ayuda en la evaluación, sobre todo para criterios que a menudo se encasillan como subjetivos. Y este paso creado, que ya es más que un concepto, puede hacerse realidad. Y ayudar tanto alumnos de forma autónoma como a docentes.

QUÉ SE HA REALIZADO REALMENTE

Con el fin de mejorar la credibilidad, trazabilidad y valor pedagógico del trabajo se han incluido las decisiones descartadas, los límites encontrados y el tipo de datos usados para testeo.

Seguramente esta tecnología ya existe, pero no está disponible de forma gratuita. Hay herramientas similares en empresas como Duolingo para la evaluación del lenguaje. Pero no a nivel artístico. Tenemos las herramientas para construir algo potencialmente muy valioso: un modelo objetivo, implementarlo y ofrecérselo a la comunidad educativa de manera.

MIEDO

Un rechazo podría ser el consenso de los criterios de evaluación y el cómo determinar aquellos criterios subjetivos. Lejos de normalizar el arte, pese a que hay criterios establecidos universales, se hace notar la necesidad de hacer que esta herramienta sea configurable, y así en cada materia, etapa o Escuela de Arte puedan tener facilidad para personalizarlo.