



Universidad Internacional de La Rioja  
Escuela Superior de Ingeniería y  
Tecnología

Máster Universitario en Inteligencia Artificial

**Probador virtual basado en redes neuronales  
convolucionales (CNN) y redes generativas antagónicas  
(GAN)**

Trabajo fin de estudio presentado por:	Marina Palma Lorita
Tipo de trabajo:	Desarrollo de Software
Director/a:	David Sánchez Alonso
Fecha:	20/07/2023

## Resumen

El auge del comercio electrónico ha amplificado la necesidad de soluciones tecnológicas que aborden los desafíos inherentes a la compra de ropa en línea. El presente trabajo se enfoca en la implementación de un probador virtual basado en Inteligencia Artificial que simula el ajuste de la ropa en un cuerpo tridimensional generado. La propuesta combina técnicas de visión por computadora y aprendizaje profundo para recrear cuerpos humanos y superponer ropa de manera realista, proporcionando así una vista previa del producto en un modelo personalizado. El sistema demostró su eficacia alcanzando una precisión del 96.6% en la segmentación de imágenes, un 57.9% en la estimación de poses, y un equilibrio sólido en la generación y discriminación de imágenes en el modelo GAN. Adicionalmente, obtuvo una calificación media del 95% en la evaluación con usuarios, reflejando su potencial de transformar la industria de la moda *online*, favoreciendo la personalización y reduciendo las devoluciones de productos.

**Palabras Clave:** Deep Learning, CNN, GAN, Transfer Learning, Virtual Fitting Room

## Abstract

The rise of e-commerce has amplified the need for technological solutions that address the challenges inherent in online clothing shopping. This work focuses on the implementation of an AI-based virtual fitting room that simulates the fit of clothing on a generated 3D body. The proposal combines computer vision and deep learning techniques to recreate human bodies and realistically overlay clothes, thus providing a preview of the product on a personalised model. The system proved its effectiveness by achieving 96.6% accuracy in image segmentation, 57.9% accuracy in pose estimation, and a solid balance in image generation and discrimination in the GAN model. Additionally, it scored an average rating of 95% in user evaluation, reflecting its potential to transform the online fashion industry, favouring personalisation and reducing product returns.

**Keywords:** Deep Learning, CNN, GAN, Transfer Learning, Virtual Fitting Room

# Índice de contenido

1.	Introducción	1
1.1.	Motivación	1
1.2.	Planteamiento del problema	2
1.3.	Estructura de la memoria	3
2.	Análisis del contexto	5
2.1.	Análisis del mercado de venta de ropa online	5
2.2.	Plataformas con soluciones similares	7
2.3.	Tendencias en la experiencia del usuario en el comercio electrónico de moda	8
2.4.	Beneficios y desafíos de implementar probadores virtuales en el comercio electrónico	10
2.4.1.	Beneficios de los probadores virtuales	10
2.4.2.	Desafíos de los probadores virtuales	11
3.	Estado del arte	12
3.1.	Hitos destacados de la inteligencia artificial para probadores virtuales	12
3.2.	Conjuntos de datos para el entrenamiento de modelos	13
3.3.	Soluciones actuales para la segmentación de imágenes y la estimación de poses con IA	15
3.3.1.	Segmentación de imágenes con CNN	15
3.3.2.	Estimación de poses con CNN	17
3.4.	Soluciones actuales a la generación de imágenes con IA	18
3.4.1.	Generación de imágenes con GAN	18
3.4.2.	Generación de imágenes con VAE	20
3.5.	Limitaciones en los modelos de segmentación de imágenes, estimación de poses y generación de imágenes	22
3.6.	Soluciones a algunas de las limitaciones	23
3.6.1.	Transfer learning para mejorar el rendimiento del modelo	23
3.6.2.	Enfoques de aprendizaje multi-tarea	24
3.7.	Tecnologías para la implementación	25
3.7.1.	Frameworks y bibliotecas de aprendizaje profundo	25

3.7.2.	Servicios en la nube para el despliegue	27
3.7.3.	Tecnologías web y aplicaciones móviles	28
3.8.	Limitaciones de las tecnologías para la implementación	29
3.8.1.	Dependencia del hardware	29
3.8.2.	Tiempos de entrenamiento prolongados	30
3.8.3.	Privacidad y seguridad de los datos	30
3.8.4.	Adaptabilidad a diferentes dominios y estilos	31
3.9.	Resumen del estado del arte	31
4.	Objetivos y metodología del trabajo	33
4.1.	Objetivo general	33
4.2.	Objetivos específicos	33
4.3.	Metodología de trabajo	34
4.3.1.	Estructura de desglose de trabajo	35
4.3.2.	Control de calidad, pruebas y validación del modelo de IA	37
4.4.	Herramientas, bibliotecas y librerías utilizadas	37
4.4.1.	Fuentes de datos utilizadas	37
4.4.2.	Herramientas utilizadas	38
4.4.3.	Bibliotecas y librerías utilizadas	38
5.	Identificación de requisitos	39
5.1.	Requisitos funcionales	39
5.2.	Requisitos no-funcionales	39
5.3.	Restricciones del sistema	40
6.	Descripción de la herramienta software desarrollada	41
6.1.	Desarrollo del Sistema de Inteligencia Artificial	41
6.1.1.	Generación de un dataset adecuado	41
6.1.2.	Modelo de segmentación de imágenes con CNN	45
6.1.3.	Modelo de estimación de poses con CNN	50
6.1.4.	Modelo de generación de imágenes con GAN	53
6.1.5.	Integración de los modelos	58
6.1.6.	Limitaciones del Sistema de Inteligencia Artificial	59

6.2.	Desarrollo de la página web	59
6.2.1.	Introducción y Tecnologías Usadas	60
6.2.2.	Estructura de la Página Web	60
6.2.3.	Términos y condiciones	63
6.2.4.	Almacenamiento, seguridad y protección de los datos	64
7.	Evaluación de la herramienta desarrollada	66
7.1.	Evaluación de usabilidad	66
7.2.	Evaluación de impacto y utilidad	67
8.	Conclusiones y trabajo futuro	69
8.1.	Conclusiones	69
8.2.	Trabajo futuro	71
	Referencias bibliográficas	72
	Anexo I: Capturas de pantalla de la página web	II
	Anexo II: Resultados de las pruebas de usuario	V
	Anexo III: Preguntas y respuestas de la encuesta realizada	VII
	Anexo IV: Artículo de investigación	IX

## Índice de figuras

Figura 1: Pronóstico de ingresos del e-commerce global	5
Figura 2: Crecimiento (2020 - 2023) en las ventas minoristas de moda de lujo versus resto	6
Figura 3: Probador online de una selección de prendas a modelos de distintas características	8
Figura 4: Principales plataformas de e-commerce respecto al nivel de personalización y ganancias	9
Figura 5: Probador online con AR	9
Figura 6: Espejo probador con AR	11
Figura 7: Prueba de funcionamiento del probador virtual VITON	13
Figura 8: Ejemplos y detalles de la base de datos DeepFashion2	14
Figura 9: Ejemplos de la estimación DensePose en la base de datos DensePose-COCO	15
Figura 10: Arquitectura de la red AFGN	16
Figura 11: Resultados del enfoque HRNet en imágenes de diferentes características	18
Figura 12: Diagrama del funcionamiento de la arquitectura VITON	20
Figura 13: Ejemplo de la arquitectura de VAE	21
Figura 14: Casos de error en la generación de imágenes con el método de VITON	22
Figura 15: Aprendizaje por transferencia del dataset de ImageNet aplicado a imágenes médicas	24
Figura 16: Comparativa del número de referencias académicas a TensorFlow y PyTorch	27
Figura 17: Arquitectura del ciclo de vida de un modelo de Machine Learning en AWS SageMaker	28
Figura 18: EDT del proyecto	35

Figura 19: Diagrama de Gantt del proyecto	36
Figura 20: Muestra de los diferentes tipos de datos de DeepFashion-MultiModal	42
Figura 21: Subconjunto del dataset de imágenes originales	42
Figura 22: Subconjunto de las imágenes segmentas	43
Figura 23: Visualización de la posición a partir de las coordenadas	44
Figura 24: Subconjunto de la visualización de posiciones	44
Figura 25: Arquitectura del modelo pre-entrenado VGG16	45
Figura 26: Arquitectura del modelo de segmentación de imágenes	47
Figura 27: Resultados de la última época del entrenamiento de segmentación	48
Figura 28: Evolución de la pérdida y precisión del modelo de segmentación	48
Figura 29: Evaluación de los resultados del modelo de segmentación	49
Figura 30: Arquitectura del modelo pre-entrenado RedNet50	50
Figura 31: Arquitectura del modelo de estimación de poses	51
Figura 32: Resultados de la última época del entrenamiento de estimación de poses	52
Figura 33: Evolución de la pérdida y precisión del modelo de estimación de poses	52
Figura 34: Evaluación de los resultados del modelo de estimación de poses	53
Figura 35: Arquitectura del modelo pix2pix	53
Figura 36: Arquitectura de la fase de adaptación de la prenda (generador) del modelo de generación de imágenes	54
Figura 37: Arquitectura de la fase de generación de partes no cubiertas (generador) del modelo de generación de imágenes	55
Figura 38: Arquitectura del discriminador en el modelo de generación de imágenes	55
Figura 39: Evolución de la pérdida en el modelo de generación de imágenes	56
Figura 40: Ejemplo de la aplicación del modelo de generación de imágenes	57

Figura 41: Resultados de la evaluación del modelo de generación de imágenes	57
Figura 42: Diagrama de la integración de la herramienta	58
Figura 44: Página de inicio de la web	61
Figura 45: Página del probador mientras se añade una imagen	61
Figura 46: Página del probador mostrando la nueva imagen generado	62
Figura 47: Página de información de la web	63
Figura 48: Ejemplos de las pruebas de usuario realizadas	66
Figura 49: Resultados a la puntuación general de la herramienta	67
Figura 50: Respuestas a un subconjunto de las cuestiones planteadas a los usuarios	68

## Índice de tablas

Tabla 1: Descripción de tareas con fecha de inicio y fin _____	35
Tabla 2: Etiquetas de las imágenes segmentadas por DeepFashion-MultiModal _____	43

# 1. Introducción

En el mundo de la moda y la venta de ropa en línea, la personalización y la experiencia del cliente son aspectos cruciales para mejorar la satisfacción y la fidelidad del consumidor. Actualmente, muchas tiendas en línea ofrecen fotografías de modelos con distintas prendas, pero la experiencia de compra podría mejorarse significativamente si los clientes pudieran visualizar cómo les quedaría la ropa a ellos mismos. Por otro lado, este tipo de personalización también podría ser útil para los fabricantes de ropa que quisieran tener una idea de cómo se verían sus productos en diferentes personas antes de iniciar la producción en masa.

La IA y el aprendizaje profundo poseen un gran potencial en la generación y manipulación de imágenes. En particular, se ha observado que las redes generativas antagónicas (GAN, por sus siglas en inglés) y las redes neuronales convolucionales (CNN, por sus siglas en inglés) presentan una notable capacidad para generar imágenes realistas, realizar segmentaciones y estimar poses ([Goodfellow et al., 2018](#)). Estas tecnologías pueden aplicarse para crear una solución que permita a los usuarios visualizar cómo les quedaría una prenda de vestir específica, basándose en imágenes personales y de la prenda deseada.

## 1.1. Motivación

La principal motivación de este proyecto reside en revolucionar la experiencia de compra *online* y la toma de decisiones en la industria de la moda. Al ofrecer una representación visual cautivadora de cómo se ajusta la ropa a los clientes, se les brinda la oportunidad de tomar decisiones de compra mejor informadas y seguras, lo que, en última instancia, podría disminuir las devoluciones y elevar la satisfacción del cliente.

Además, al adoptar un enfoque visual, se beneficia a los usuarios al proporcionar imágenes personalizadas y realistas que abarcan la diversidad de cuerpos, edades, géneros y razas en la sociedad actual. Al alejarse de las representaciones típicas de modelos con cuerpos normativos y ofrecer imágenes más inclusivas, se fomenta una percepción más saludable y realista de cómo las personas pueden lucir las prendas de vestir.

Asimismo, este proyecto tiene el potencial de brindar ventajas a los fabricantes de ropa, ya que podrían visualizar cómo se verían sus productos en diferentes personas antes de la producción en masa. De esta manera, podrían tomar decisiones mejor fundamentadas sobre el diseño y la selección de tallas, lo que permitiría optimizar la producción y la logística.

En resumen, este proyecto aspira a transformar la industria de la moda, brindando beneficios tanto a los consumidores como a los fabricantes y promoviendo una mayor inclusión en el proceso.

## **1.2. Planteamiento del problema**

La estructura del proyecto se desarrollará en varias etapas para abordar de manera eficiente y efectiva el desarrollo del probador virtual.

En primer lugar, se llevará a cabo un análisis exhaustivo de las principales soluciones basadas en IA en el campo de la generación de imágenes, la segmentación semántica y la estimación de poses. El análisis permitirá identificar las técnicas y arquitecturas más adecuadas para abordar el problema específico de generar imágenes de personas con ropa personalizada.

A continuación, se procederá a la preparación del conjunto de datos de imágenes, que incluirá personas con diferentes poses, pesos, alturas y prendas de vestir, así como información sobre las tallas y otros detalles relevantes de las prendas. El conjunto de datos será crucial para entrenar y evaluar los modelos de segmentación semántica, estimación de poses y generación de imágenes.

Posteriormente, se entrenará un modelo de segmentación semántica y estimación de poses, utilizando CNN para segmentación de imágenes. Este modelo permitirá detectar personas y sus poses en las imágenes de entrada, lo cual es fundamental para generar imágenes realistas de personas con ropa personalizada.

En paralelo, se entrenará una GAN condicional, que será responsable de generar imágenes de personas con ropa basándose en las imágenes de entrada y las tallas de ropa proporcionadas. Este modelo se beneficiará del *transfer learning*, aprovechando modelos preentrenados y conocimientos adquiridos en tareas anteriores de generación de imágenes.

Una vez que ambos modelos estén entrenados y optimizados, se combinarán en un sistema integrado que tome como entrada imágenes de personas, pesos, alturas y prendas de vestir, con el fin de generar imágenes de la persona con la ropa puesta. Este sistema permitirá la interacción con el usuario final y proporcionará resultados visuales de cómo le quedaría la ropa.

Finalmente, se desarrollará el software en forma de aplicación o página web, accesible para el usuario final. Esta plataforma permitirá a los usuarios cargar sus imágenes, seleccionar prendas de vestir y tallas, y obtener una representación visual de cómo les quedaría la ropa, mejorando así la experiencia de compra en línea y la toma de decisiones en la industria de la moda.

### **1.3. Estructura de la memoria**

En el **primer capítulo**, se establece el propósito y la motivación del trabajo, así como el planteamiento del problema a resolver con la propuesta de un probador virtual en línea. Este capítulo también describe cómo está organizada la memoria.

El **segundo capítulo** se sumerge en un análisis profundo del contexto, donde se explora el mercado de venta de ropa en línea, se analizan las plataformas con soluciones similares y se discuten las tendencias actuales en la experiencia del usuario en el comercio electrónico de moda. Este capítulo concluye con un examen de los beneficios y desafíos de implementar probadores virtuales en el comercio electrónico.

El **tercer capítulo** es un recorrido por el estado del arte. Se describen las tecnologías actuales en inteligencia artificial aplicadas a probadores virtuales, incluyendo el uso de conjuntos de datos, técnicas de segmentación de imágenes y estimación de poses, generación de imágenes y sus limitaciones. Se presenta también un resumen de las soluciones y tecnologías de implementación disponibles.

En el **cuarto capítulo**, se definen los objetivos generales y específicos del trabajo y se describe la metodología seguida. Además, se detallan las herramientas, bibliotecas y librerías utilizadas en el proyecto.

El **quinto capítulo** presenta los requisitos funcionales y no funcionales identificados para el sistema, así como las restricciones del sistema propuesto.

En el **sexto capítulo**, se detalla el desarrollo del sistema de inteligencia artificial y de la página web. Se discute desde la generación de un conjunto de datos adecuado hasta la integración de los modelos. Este capítulo continúa con una descripción de las limitaciones del sistema de inteligencia artificial y del desarrollo de la página web.

El **séptimo capítulo** evalúa la herramienta desarrollada en términos de usabilidad, impacto y utilidad. Se proporcionan detalles de cómo se realizaron estas evaluaciones.

En el **octavo capítulo** se extraen conclusiones basadas en el trabajo realizado y se esbozan posibles direcciones para futuras investigaciones y mejoras del sistema.

En el **anexo I**, se adjuntas capturas de pantalla de la página web desarrollada.

En el **anexo II**, se adjuntas los resultados de las 4 pruebas de usuario.

En el **anexo III**, se adjunta la encuesta completa realizada a los usuarios de la página web.

En el **anexo IV**, se adjunta el artículo de investigación correspondiente al trabajo realizado.

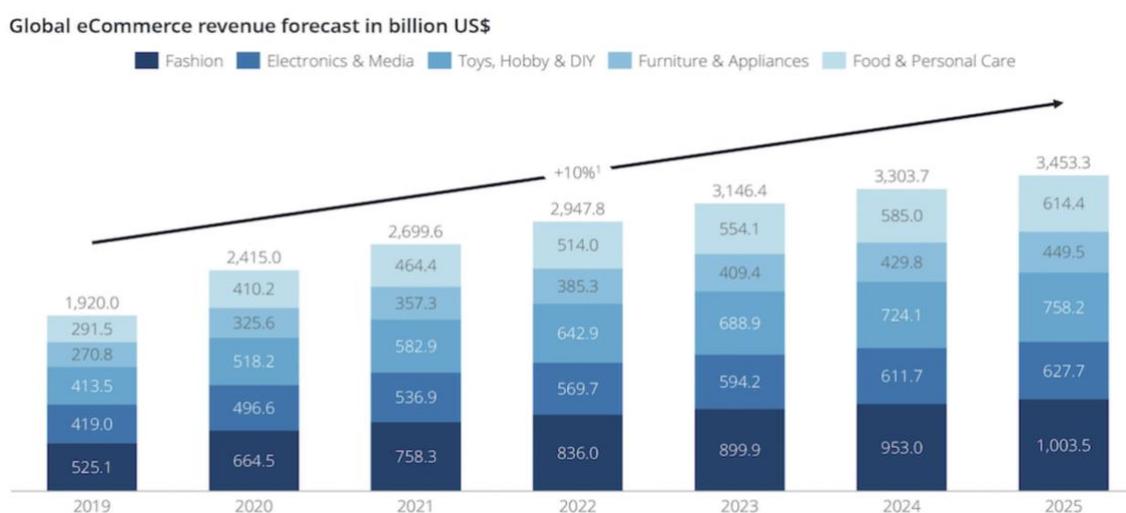
## 2. Análisis del contexto

La era digital ha transformado la forma en que los consumidores compran productos y servicios, y la industria de la moda no es una excepción. El comercio electrónico ha experimentado un crecimiento significativo en los últimos años, y se espera que continúe expandiéndose ([Statista, 2021](#)). Las marcas y minoristas de moda se enfrentan a la necesidad de adaptarse a este nuevo paradigma y ofrecer experiencias de compra en línea que satisfagan las expectativas de los consumidores.

### 2.1. Análisis del mercado de venta de ropa *online*

El mercado de venta de ropa en línea ha experimentado un rápido crecimiento en los últimos años, gracias a factores como el incremento en la penetración de Internet, la mayor adopción de dispositivos móviles y la expansión de plataformas de comercio electrónico ([Statista, 2021](#)). En la Figura 1 se observa cómo el sector de la moda tiene, y se pronostica que tendrá, los mayores ingresos en *e-commerce* a nivel mundial. A medida que más consumidores adoptan la compra de ropa *online*, las expectativas sobre la experiencia de compra también aumentan. Los clientes buscan comodidad, variedad, precios competitivos y una experiencia de compra atractiva que les permita explorar y probar productos virtualmente antes de tomar una decisión de compra ([Deloitte, 2020](#)).

Figura 1: Pronóstico de ingresos del e-commerce global

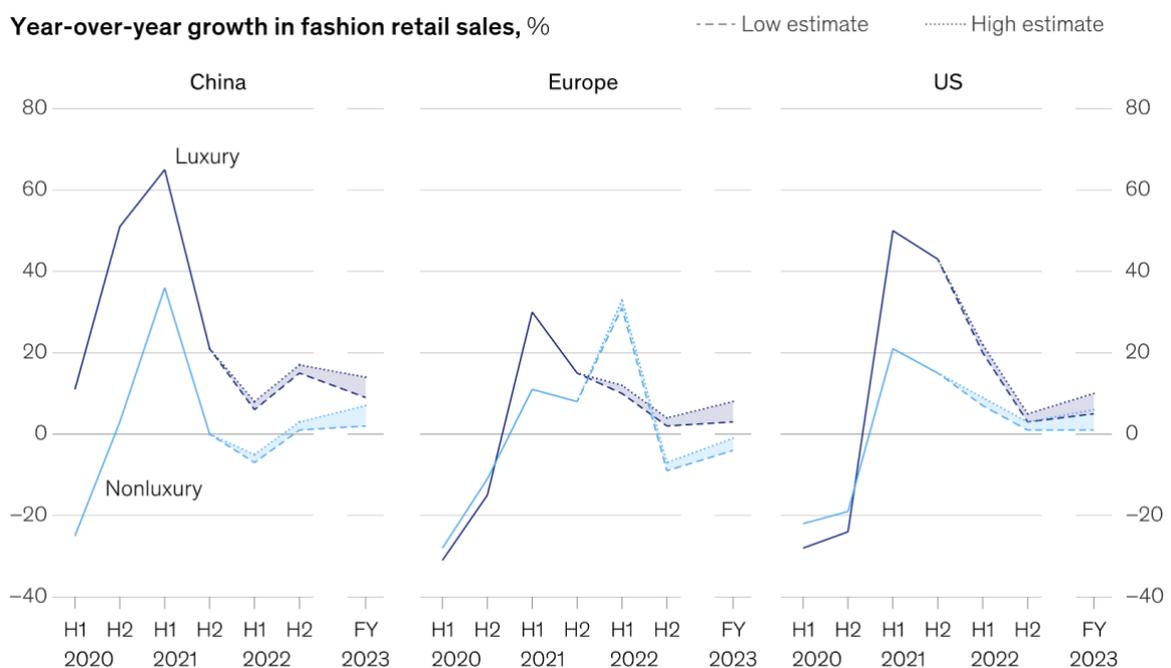


Fuente: ([Statista, 2021](#))

Según el informe de Statista (2021), se espera que el crecimiento del mercado de venta de ropa *online* continúe en los próximos años, lo que representa oportunidades significativas para las marcas y minoristas que buscan innovar y mejorar la experiencia del cliente en línea.

La pandemia de COVID-19 aceleró la adopción del comercio electrónico de moda, ya que las restricciones de movilidad y el cierre de tiendas físicas impulsaron a los consumidores a buscar soluciones de compra en línea (McKinsey & Company, 2023). En la Figura 2 se observa como el análisis de McKinsey de las previsiones de moda para 2023 proyecta un crecimiento de las ventas relativamente lento, lastrado por una contracción en el mercado europeo y el efecto de la inflación.

Figura 2: Crecimiento (2020 - 2023) en las ventas minoristas de moda de lujo versus resto



Fuente: (McKinsey & Company, 2023)

En este contexto de incertidumbre sobre las ventas, las empresas de moda enfrentan el desafío de ofrecer experiencias de compra en línea que satisfagan las crecientes expectativas de los consumidores. Incluyendo el desarrollo e implementación de soluciones innovadoras, como los probadores virtuales, se permite a los clientes explorar y probar productos de manera virtual antes de realizar una compra. La inversión en tecnologías y soluciones avanzadas es esencial para mantenerse competitivo en el mercado de venta de ropa en línea, que se caracteriza por su dinamismo y rapidez en el cambio de tendencias y demandas de los consumidores.

Los probadores virtuales representan una oportunidad para mejorar la experiencia del cliente *online*, al mismo tiempo que abordan algunos de los desafíos clave que enfrentan las empresas de moda en el comercio electrónico, como las devoluciones y la falta de interacción personalizada. Al adoptar tecnologías de inteligencia artificial (IA), aprendizaje automático y procesamiento de imágenes, las empresas pueden ofrecer soluciones de probador virtual que proporcionen una experiencia de compra más atractiva y personalizada, lo que a su vez puede impulsar la satisfacción del cliente y la lealtad a la marca.

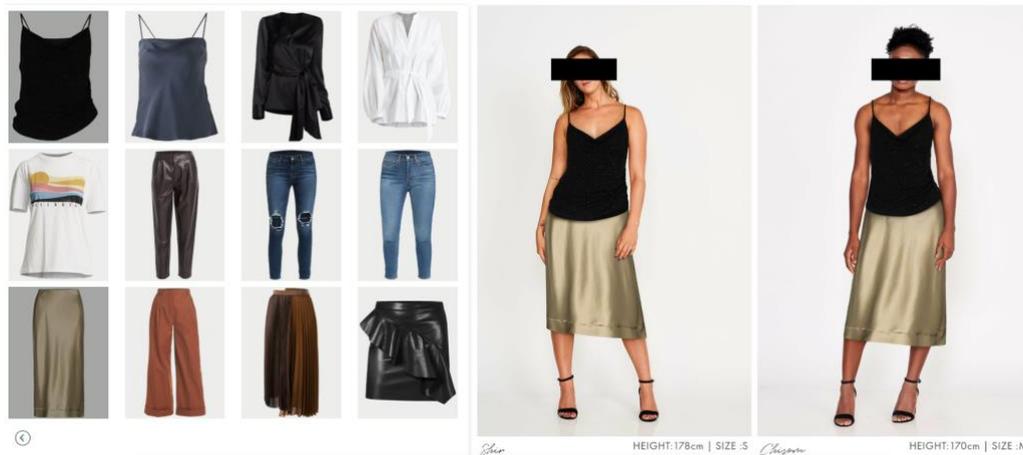
## 2.2. Plataformas con soluciones similares

En respuesta a la creciente demanda de experiencias de compra en línea más personalizadas e interactivas, varias plataformas han comenzado a implementar soluciones similares a los probadores virtuales. Estas soluciones, impulsadas por la IA, permiten a los clientes probar ropa virtualmente, lo que ayuda a superar las barreras de la compra *online* y reduce las devoluciones de productos ([Deloitte, 2020](#)).

Una de las plataformas más conocidas en este ámbito es Zeekit, un startup israelí adquirida por Walmart en 2021 ([Zeekit, 2021](#)). Zeekit ha desarrollado una tecnología de probador virtual que permite a los clientes probar distintas prendas de ropa a una amplia gama de modelos en tiempo real a través de la web. Además, los usuarios pueden cargar imágenes de sí mismos y "probarse" virtualmente prendas de ropa de diferentes tallas y estilos. En la Figura 3 se puede ver la interfaz del probador online de la plataforma Zeekit. Otras marcas y minoristas también han adoptado soluciones similares, como ASOS con su función "See My Fit" ([ASOS, 2020](#)), que utiliza la realidad aumentada (AR, por sus siglas en inglés) para mostrar cómo quedan las prendas en diferentes tipos de cuerpos.

Además, algunas empresas se han decantado por ofrecer servicios de personalización de prendas y asesoramiento de estilo en línea, como Stitch Fix, que utiliza algoritmos de aprendizaje automático para seleccionar prendas y accesorios adecuados para sus clientes según sus preferencias y medidas ([Stitch Fix, 2021](#)).

Figura 3: Probador online de una selección de prendas a modelos de distintas características



Fuente: [\(Zeekit, 2021\)](#)

A pesar de los avances en las soluciones de probadores virtuales, aún existen desafíos y oportunidades para mejorar la precisión y la calidad sobre la experiencia de las mismas. La implementación de tecnologías más avanzadas, puede ayudar a superar algunas de las limitaciones actuales y ofrecer experiencias de probador virtual más realistas y personalizadas.

### 2.3. Tendencias en la experiencia del usuario en el comercio electrónico de moda

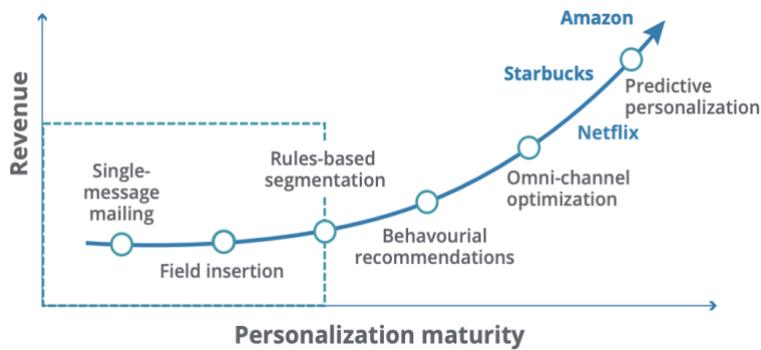
La experiencia del usuario en el comercio electrónico de moda ha experimentado cambios significativos en los últimos años, impulsados por las expectativas cambiantes de los consumidores y los avances en la tecnología. Entre las tendencias más destacadas se encuentran la sostenibilidad, la personalización, la AR y la integración de la IA.

La sostenibilidad se ha vuelto cada vez más relevante en la experiencia del usuario en el comercio electrónico de moda. Los consumidores están cada vez más interesados en la procedencia y el impacto ambiental de las prendas que compran [\(McKinsey, 2020\)](#). Las marcas están respondiendo a esta tendencia al ofrecer información detallada sobre sus prácticas de producción sostenible introduciendo líneas de ropa ecológicas.

La personalización también se ha convertido en una característica clave para mejorar la experiencia del usuario en el comercio electrónico de moda, ya que los consumidores buscan experiencias de

compra más individualizadas y relevantes ([Deloitte, 2020](#)). Se puede observar como se relaciona la facturación y el nivel de personalización en las principales plataformas de comercio electrónico en la Figura 4. La IA y el aprendizaje automático están permitiendo a las marcas ofrecer recomendaciones de productos y promociones más específicas, adaptadas a las preferencias y el historial de compras de cada cliente.

Figura 4: Principales plataformas de e-commerce respecto al nivel de personalización y ganancias



Fuente: ([Deloitte, 2020](#))

Una tecnología inmersiva que gana cada vez más atención en este contexto es la AR. A través de dispositivos como teléfonos o tabletas, cambia la percepción del mundo físico al superponer su entorno real con elementos virtuales en tiempo real, lo que permite a los clientes probar virtualmente las prendas antes de comprarlas ([Caboni & Hagberg, 2019](#); [Cosco, 2020](#)). Además, la AR está siendo utilizada para enriquecer la experiencia de compra en línea con visualizaciones 3D de productos y experiencias interactivas. En la Figura 5 se muestra un ejemplo de un probador virtual con AR.

Figura 5: Interfaz del probador online de Snapchat con AR



Fuente: ([L. London, 2021](#))

La integración de la IA en la experiencia del usuario en el comercio electrónico de moda está permitiendo a las empresas mejorar la eficiencia de sus operaciones y proporcionar un mejor servicio al cliente. Los *chatbots* y los sistemas de soporte al cliente impulsados por IA están mejorando la comunicación con los clientes, mientras que la IA también se utiliza para optimizar la cadena de suministro y la gestión de inventario ([Buttler, 2021](#)).

## **2.4. Beneficios y desafíos de implementar probadores virtuales en el comercio electrónico**

Los probadores virtuales han surgido como una solución innovadora en el comercio electrónico de moda para mejorar la experiencia del cliente y aumentar las ventas. Sin embargo, la implementación de esta tecnología también presenta desafíos que deben abordarse para maximizar sus beneficios.

### **2.4.1. Beneficios de los probadores virtuales**

Uno de los principales beneficios de los probadores virtuales es mejorar la experiencia del cliente al permitir probar prendas virtualmente antes de realizar una compra ([Lee & Xu, 2020](#)). Esta capacidad ayuda a los consumidores a tomar decisiones más informadas y a encontrar la talla y el estilo adecuado, lo que puede conducir a una mayor satisfacción y lealtad del cliente ([Liaw & Chen, 2013](#)).

Además, los probadores virtuales pueden aumentar las ventas al reducir las tasas de devolución de productos y mejorar la tasa de conversión de visitantes a compradores ([Yang et al., 2019](#)). La capacidad de probar virtualmente la ropa antes de comprarla puede disminuir la incertidumbre sobre cómo se verá y se ajustará el producto, lo que reduce la probabilidad de devoluciones y, en última instancia, puede aumentar los ingresos para el vendedor.

Desde la perspectiva de la experiencia del cliente, Lee y Xu (2020) señalan que las tecnologías de probadores virtuales se pueden clasificar en función de su capacidad para proporcionar interacción, inmersión y realismo. Estos elementos pueden mejorar significativamente la satisfacción del cliente y la intención de compra al proporcionar una experiencia de compra en línea más personalizada y atractiva. La Figura 6 muestra el prototipo de un espejo probador implementado con AR en una tienda de indonesia ([Werdayani & Widiaty, 2020](#)).

Figura 6: Prototipo de espejo probador con tecnología AR



Fuente: [\(Werdayani & Widiaty, 2020\)](#)

#### 2.4.2. Desafíos de los probadores virtuales

Los probadores virtuales también enfrentan desafíos en su implementación, a pesar de sus beneficios. Uno de los principales obstáculos es la calidad y precisión de la representación de la ropa en el cuerpo virtual del usuario. Si las imágenes generadas no son lo suficientemente realistas o precisas, los consumidores pueden verse decepcionados y perder la confianza en la plataforma [\(Lee et al., 2022\)](#).

Otro desafío es garantizar la facilidad de uso y accesibilidad de la tecnología para una amplia gama de consumidores. La complejidad técnica y las dificultades para utilizar la plataforma pueden disuadir a los clientes de utilizar los probadores virtuales y, en última instancia, limitar su adopción [\(Liaw & Chen, 2013\)](#).

Además, los probadores virtuales pueden presentar preocupaciones de privacidad para los consumidores. La recopilación y el uso de datos personales, como imágenes corporales o medidas, pueden generar preocupaciones sobre la seguridad y el uso indebido de estos datos [\(Yang et al., 2019\)](#). Los vendedores deben abordar estos problemas de privacidad y garantizar que los datos del cliente estén protegidos y se utilicen de manera responsable.

### 3. Estado del arte

El estado del arte en técnicas de generación de imágenes, segmentación de imágenes y estimación de poses ha evolucionado rápidamente en la última década, gracias al avance en el aprendizaje profundo y las redes neuronales. A continuación, se presenta una descripción detallada del estado del arte técnico relevante para este proyecto.

#### 3.1. Hitos destacados de la inteligencia artificial para probadores virtuales

La IA ha experimentado un rápido desarrollo en la última década, impulsando avances significativos en áreas como el reconocimiento y la generación de imágenes, la segmentación de imágenes y la estimación de poses. Estos desarrollos han tenido un impacto considerable en la industria de la moda, dando lugar a numerosos hitos en el campo de los probadores virtuales, como herramientas que hacen uso de la IA para ofrecer a los clientes una experiencia más enriquecedora ([Gao et al., 2019](#)).

Uno de los pilares fundamentales en el desarrollo de la IA para probadores virtuales ha sido la creación de conjuntos de datos sólidos. DeepFashion ([Liu et al., 2016](#)), es un conjunto de datos ampliamente utilizado que contiene más de 800.000 imágenes etiquetadas de prendas de vestir. Al ser una de las primeras bases de datos robustas en el ámbito de la moda, ha impulsado el avance de algoritmos de reconocimiento y generación de imágenes en este sector.

Con la creciente adopción de la IA en la industria de la moda, los investigadores han desarrollado enfoques innovadores para generar imágenes de personas usando ropa virtual. Un ejemplo notable es VITON ([Han et al., 2018](#)), un sistema de probador virtual que utiliza GAN y CNN para crear imágenes realistas de personas vistiendo ropa seleccionada a partir de imágenes de catálogos.

Además de las investigaciones académicas, tales como DeepFashion ([Liu et al., 2016](#)), DeepFashion2 ([Ge et al., 2019](#)) y Viton ([Han et al., 2018](#)), varias empresas y aplicaciones han surgido para ofrecer soluciones de probadores virtuales basadas en la IA. La Figura 7 muestra algunos de los resultados del proyecto VITON. Zeekit, mencionada anteriormente, utiliza algoritmos de aprendizaje profundo y segmentación de imágenes para ajustar automáticamente el tamaño y el ajuste de las prendas en función del cuerpo del usuario ([Zeekit, 2021](#)).

Figura 7: Prueba de funcionamiento del probador virtual VITON



Fuente: [\(Han et al., 2018\)](#)

Otra aplicación innovadora en este campo es Metail, que ha desarrollado un probador virtual basado en avatares 3D [\(Metail, 2019\)](#). Los usuarios pueden crear avatares personalizados con sus medidas y probar diferentes prendas en el mismo. Metail utiliza tecnologías como la visión por computadora y el aprendizaje automático para generar avatares realistas y simular la forma en que las prendas se ajustan al cuerpo del usuario.

Estos hitos en la IA para probadores virtuales han sido posibles gracias a la confluencia de avances en tecnologías y algoritmos, como GAN, CNN y *transfer learning*, como en el proyecto Viton [\(Han et al., 2018\)](#). Juntos, han sentado las bases para una nueva generación de soluciones de moda en línea, permitiendo una experiencia de compra más personalizada y realista para los usuarios.

### 3.2. Conjuntos de datos para el entrenamiento de modelos

La disponibilidad de conjuntos de datos adecuados es esencial para el entrenamiento y evaluación de modelos de IA en el ámbito de los probadores virtuales. Estos conjuntos de datos deben contener imágenes de personas con diferentes características, como poses, pesos y alturas, así como imágenes de prendas de vestir con información pertinente, como tallas y detalles de diseño.

Un conjunto de datos clave en este campo es DeepFashion, mencionado anteriormente, que fue presentado en 2016 [\(Liu et al., 2016\)](#). Este conjunto de datos incluye más de 800.000 imágenes

etiquetadas de prendas de vestir, lo que lo convierte en una fuente de información valiosa para el desarrollo de algoritmos de reconocimiento y generación de imágenes en el ámbito de la moda.

Más tarde aparecería DeepFashion2, lanzado en 2019, con más de 490.000 imágenes con anotaciones detalladas de instancias de ropa, categorías, atributos y puntos clave, siendo útil para la segmentación de imágenes y la estimación de poses ([Ge et al., 2019](#)). Un ejemplo del detalle y especificaciones de la base de datos DeepFashion2 se muestra en la Figura 8.

Figura 8: Ejemplos y detalles de la base de datos DeepFashion2



Fuente: ([Ge et al., 2019](#))

Otro conjunto de datos relevante es iMaterialist Challenge (Fashion), creado para la Conferencia sobre la Visión por Computadora y el Reconocimiento de Patrones (CVPR) 2019 ([Bai et al., 2020](#)). Contiene más de 50.000 imágenes con anotaciones detalladas de atributos y etiquetas semánticas de ropa.

Un enfoque alternativo para la adquisición de datos es utilizar conjuntos de datos generales de personas y segmentarlos para obtener información relevante. El conjunto de datos DensePose-COCO ([Güler et al., 2018](#)) es una extensión del conjunto de datos MS COCO ([Lin et al., 2014](#)) que proporciona anotaciones de correspondencias de superficie densa para la estimación de poses humanas. Fue lanzado en 2018 y contiene anotaciones para más de 200.000 personas en más de 50.000 imágenes. La Figura 9 muestra la calidad de las estimaciones de DensePose en distintas situaciones.

Figura 9: Ejemplos de la estimación DensePose en la base de datos DensePose-COCO



Fuente: [\(Güler et al., 2018\)](#)

Es importante destacar que, en algunos casos, puede ser necesario crear conjuntos de datos personalizados para abordar necesidades específicas de un proyecto. La recopilación de imágenes de personas y prendas de vestir, así como la anotación de información relevante, pueden conllevar tiempo y recursos, pero es crucial para garantizar la precisión y efectividad del modelo entrenado.

### **3.3. Soluciones actuales para la segmentación de imágenes y la estimación de poses con IA**

En el ámbito de los probadores virtuales, la segmentación de imágenes y la estimación de poses son dos tareas fundamentales que permiten analizar y comprender las imágenes de personas y prendas de vestir. Estas tareas han experimentado avances significativos en los últimos años gracias a las investigaciones en IA y, en particular, al uso de CNN [\(Wang et al., 2018; Güler et al., 2018\)](#).

#### **3.3.1. Segmentación de imágenes con CNN**

La segmentación de imágenes es el proceso de dividir una imagen en varias regiones o segmentos, asignando una etiqueta a cada píxel para identificar los objetos y regiones de interés. En el contexto de los probadores virtuales, la segmentación de imágenes es crucial para separar a las personas y las prendas de vestir del fondo, lo que permite la manipulación y la generación de imágenes realistas de personas usando ropa personalizada.

Las CNN han demostrado ser particularmente efectivas para abordar problemas de segmentación de imágenes, ya que han sido diseñadas específicamente para manejar datos visuales [\(Long et al., 2015\)](#). Estas redes utilizan capas convolucionales para analizar localmente las imágenes y extraer

características jerárquicas, lo que les permite aprender representaciones efectivas de los objetos y sus contextos espaciales.

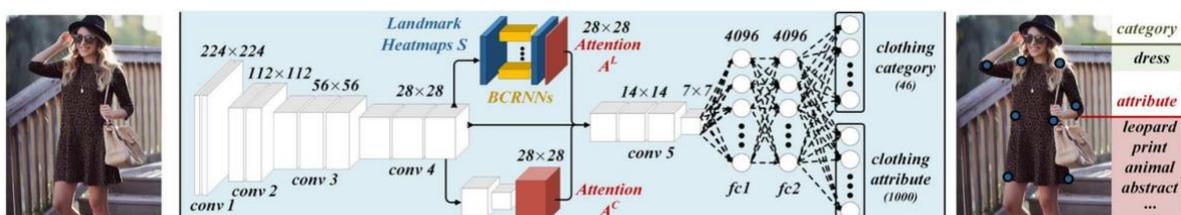
Un enfoque popular en la segmentación de imágenes con CNN es el uso de arquitecturas de codificador-decodificador, como la U-Net ([Giang et al., 2020](#)), la SegNet ([Badrinarayanan et al., 2017](#)) o el modelo VGG16 ([Simonyan et al., 2014](#)). Este último, consiste en 13 capas convolucionales, organizadas en bloques con un número creciente de filtros, y 3 capas completamente conectadas. Los filtros extraídos por este modelo son altamente generalizables y, por lo tanto, proporcionan una base sólida para muchas tareas de *deep learning*.

Estas arquitecturas consisten en un codificador que extrae características de la imagen de entrada, reduciendo progresivamente su dimensionalidad mientras aumenta la profundidad de las características. Seguidamente, un decodificador reconstruye la segmentación a partir de estas características, restaurando gradualmente la dimensionalidad de la imagen original y ubicando precisamente cada píxel en una categoría específica.

El proceso que sigue una CNN se asemeja a comprimir y luego descomprimir la información, lo que permite que la red aprenda y mantenga las relaciones espaciales entre los píxeles, un factor crítico en la segmentación de imágenes ([Zhou et al., 2018](#)). Esta capacidad de las arquitecturas de codificador-decodificador ha resultado en su uso extendido en aplicaciones de segmentación de imágenes, proporcionando rendimientos notables en diversas tareas.

Un estudio que aborda la segmentación de imágenes de personas y ropa es el trabajo de Wang et al., (2018). En su investigación, los autores propusieron una red de gramática de moda atenta, la Attentive Fashion Grammar Network (AFGN), para la detección de puntos de referencia de moda y la clasificación de categorías de ropa. La AFGN utiliza una estructura jerárquica y módulos de atención para capturar información semántica y detalles de prendas de vestir de manera eficiente, como se puede ver en la Figura 10.

Figura 10: Arquitectura de la red AFGN



Fuente: ([Wang et al., 2018](#))

Avances recientes en la segmentación de imágenes con CNN han permitido el desarrollo de sistemas de probadores virtuales más efectivos y precisos ([Han et al., 2018](#)). Sin embargo, aún existen desafíos en la segmentación de imágenes de personas y ropa, como la variabilidad en la apariencia de las prendas, las poses y las condiciones de iluminación. Investigaciones futuras podrían explorar nuevas arquitecturas y técnicas de aprendizaje profundo para abordar estos desafíos y mejorar aún más el rendimiento de los sistemas de segmentación de imágenes.

### 3.3.2. Estimación de poses con CNN

La estimación de poses es otro componente clave en el desarrollo de probadores virtuales, ya que permite comprender la postura y la orientación del cuerpo humano en las imágenes. La tarea de estimación de poses consiste en localizar y asociar las articulaciones del cuerpo, como los hombros, los codos y las rodillas, para obtener una representación estructural de la postura humana. Esta información es muy útil para posteriormente generar imágenes realistas de personas usando ropa personalizada, ya que permite ajustar las prendas de vestir a la postura y la forma del cuerpo de manera adecuada. Las CNN han demostrado ser efectivas en la estimación de poses humanas, debido a su capacidad para capturar características visuales y patrones espaciales ([Yang et al., 2019](#)).

Un enfoque reciente en la estimación de poses es el uso de arquitecturas de CNN basadas en la representación de mapas de calor para las articulaciones del cuerpo humano. En este enfoque, la CNN predice un mapa de calor para cada articulación, donde los valores de los píxeles representan la probabilidad de la presencia de una articulación en una ubicación específica. Los mapas de calor se combinan luego para obtener la postura humana completa ([Chen et al., 2018](#)). También es ampliamente utilizado el modelo ResNet50 ([Theckedath et al., 2020](#)), una red convolucional profunda que se destaca por sus conexiones residuales, permitiendo la propagación directa de la señal de entrada a capas más profundas y aportando una alta precisión en la estimación de poses.

Por ejemplo, Sun et al. (2019) propusieron un enfoque llamado *Deep High-Resolution Representation Learning* (HRNet) para la estimación de poses humanas. HRNet utiliza una arquitectura de CNN que mantiene una resolución espacial alta en todas las etapas del proceso, lo que permite capturar detalles finos y contextos globales de manera efectiva. Su enfoque logró un rendimiento superior en la estimación de poses humanas en comparación con los enfoques anteriores ([Sun et al., 2019](#)). Un ejemplo de sus resultados puede verse en la Figura 11.

Figura 11: Resultados del enfoque HRNet en imágenes de diferentes características



Fuente: [\(Sun et al., 2019\)](#)

Otro enfoque es el de DensePose, mencionado anteriormente por proporcionar, además de su método de estimación de poses, la base de datos DensePose-COCO [\(Güler et al., 2018\)](#). DensePose proporciona una estimación de poses humanas en imágenes cotidianas. Utiliza una CNN para predecir correspondencias de densidad entre la imagen 2D y un modelo tridimensional del cuerpo humano. Este enfoque permite obtener una estimación más precisa y detallada de la postura y la forma del cuerpo en escenas del mundo real, lo que es útil para aplicaciones de probadores virtuales.

Los avances en la estimación de poses con CNN han permitido el desarrollo de sistemas de probadores virtuales más precisos y realistas. A medida que se siguen investigando nuevas arquitecturas y técnicas de aprendizaje profundo, es probable que la estimación de poses continúe mejorando, lo que a su vez mejorará la calidad y la efectividad de los probadores virtuales.

### 3.4. Soluciones actuales a la generación de imágenes con IA

La generación de imágenes es una tarea esencial en el contexto de los probadores virtuales, ya que permite crear imágenes realistas y convincentes de personas con prendas de vestir virtuales. En los últimos años, se han desarrollado varios enfoques y arquitecturas basadas en IA para abordar este desafío [\(Goodfellow et al., 2018\)](#), como las que se especifican a continuación.

#### 3.4.1. Generación de imágenes con GAN

Las GAN son una herramienta poderosa y efectiva para generar imágenes realistas a partir de datos de entrada. Consisten en dos redes neuronales, un generador y un discriminador, que compiten entre sí

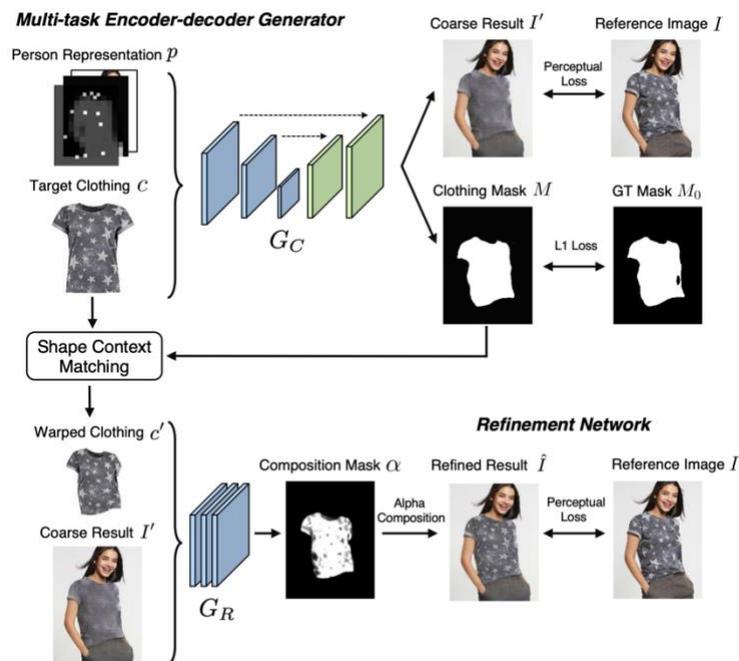
en un juego de suma cero. El generador crea imágenes sintéticas, mientras que el discriminador evalúa si las imágenes son reales o falsas. A medida que el generador mejora en la creación de imágenes realistas, el discriminador debe mejorar en la identificación de imágenes falsas. Este proceso iterativo continúa hasta que el generador crea imágenes que el discriminador no puede distinguir de entre las imágenes reales.

En el contexto de los probadores virtuales, las GAN pueden utilizarse para generar imágenes de personas con diferentes prendas de vestir y posturas. En los últimos años, se han propuesto varias arquitecturas de GAN y técnicas de entrenamiento para mejorar la calidad y diversidad de las imágenes generadas ([Goodfellow et al., 2018](#)). Una arquitectura GAN relevante en este contexto es el modelo pix2pix ([Li et al., 2022](#)), que se centra en tareas de traducción de imagen a imagen. Su capacidad para retener detalles finos en las imágenes de salida es particularmente útil para la generación de imágenes de personas y animales.

Por ejemplo, Zhu et al. (2017) presentaron un enfoque llamado SwapNet para generar imágenes de personas con ropa nueva y posturas personalizadas ([Zhu et al., 2017](#)). SwapNet utiliza una arquitectura GAN condicional, donde el generador crea imágenes de personas con ropa nueva basándose en las características extraídas de una imagen de entrada. El discriminador, a su vez, evalúa si las imágenes generadas son realistas y corresponden a la ropa y postura objetivo. El enfoque de SwapNet demostró ser efectivo para generar imágenes realistas de personas con diferentes estilos de ropa y posturas.

Otro ejemplo es el trabajo de Han et al. (2018), mencionado anteriormente, quienes propusieron una arquitectura llamada VITON, que utiliza GAN condicionales junto con un módulo de atención para generar imágenes realistas de personas con ropa nueva ([Han et al., 2018](#)). VITON combina información de las imágenes de entrada, como la silueta de la persona y la ropa objetivo, y la utiliza para sintetizar imágenes realistas de personas usando la ropa seleccionada. Consta de una etapa generadora de codificador-decodificador y una etapa de refinamiento, como se puede observar en la Figura 12. Este enfoque también ha demostrado ser efectivo en la creación de imágenes realistas de personas con diferentes prendas de vestir y posturas.

Figura 12: Diagrama del funcionamiento de la arquitectura VITON

Fuente: [\(Han et al., 2018\)](#)

Los avances mencionados en la generación de imágenes con GAN han permitido el desarrollo de probadores virtuales más realistas y personalizados. A medida que continúa la investigación en arquitecturas GAN y técnicas de aprendizaje profundo, se espera que la calidad y diversidad de las imágenes generadas siga mejorando, enriqueciendo aún más la experiencia del usuario en los probadores virtuales.

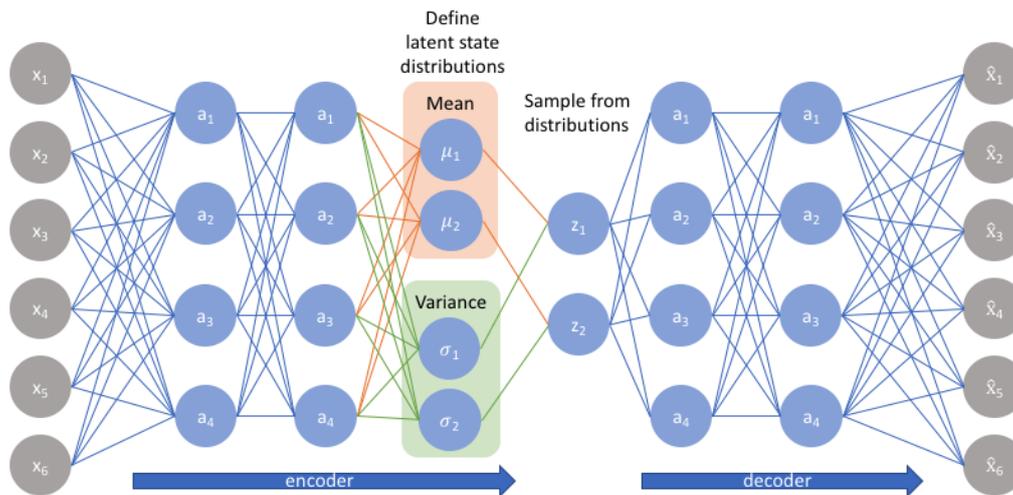
### 3.4.2. Generación de imágenes con VAE

Otra técnica popular para generar imágenes es el uso de Autoencoders Variacionales (VAE, por sus siglas en inglés). Los VAE son una clase de modelos generativos que aprenden a comprimir y reconstruir datos de alta dimensión, como imágenes, en un espacio latente de baja dimensión. La característica distintiva de los VAE es que imponen una estructura probabilística en el espacio latente, lo que permite la generación de imágenes nuevas y realistas al muestrear puntos en dicho espacio.

Como muestra la Figura 13, los VAE constan de dos componentes principales: un codificador y un decodificador. El codificador toma una imagen de entrada y aprende a comprimir sus características en un espacio latente. Este espacio latente es modelado como una distribución de probabilidad, generalmente una distribución gaussiana multivariante. Luego, el decodificador toma puntos

muestreados en el espacio latente y reconstruye imágenes a partir de ellos. La Figura 13 muestra un esquema de ejemplo de la arquitectura VAE.

Figura 13: Ejemplo de la arquitectura de VAE



Fuente: [\(Jordan, 2023\)](#)

En el contexto de los probadores virtuales, los VAE pueden utilizarse para generar imágenes realistas de personas usando diferentes prendas de vestir y adoptando diferentes posturas. Aunque los VAE no han sido tan ampliamente utilizados en esta área como las GAN, son capaces de generar imágenes realistas en otros dominios y podrían adaptarse para su uso en probadores virtuales.

El-Kaddoury et al. (2019) realizaron una comparación práctica entre VAE y GAN para la generación de imágenes [\(El-Kaddoury et al., 2019\)](#). En su estudio, analizaron las ventajas y desventajas de ambas técnicas y proporcionaron una visión detallada de su desempeño en varias tareas de generación de imágenes. En cuanto a calidad de las imágenes generadas, las GAN generalmente produjeron imágenes de mayor calidad en términos de nitidez y diversidad, mientras que los VAE, por otro lado, tendían a generar imágenes más borrosas y suaves. Por otro lado, los VAE tienen un proceso de entrenamiento más estable, lo que facilita la optimización de sus parámetros.

En resumen, los VAE son una técnica prometedora para la generación de imágenes en probadores virtuales, aunque su aplicación en esta área ha sido menos explorada en comparación con las GAN. A medida que continúe la investigación en modelos generativos y aprendizaje profundo, es probable que los VAE sigan mejorando y puedan ser adaptados para su uso en probadores virtuales.

### 3.5. Limitaciones en los modelos de segmentación de imágenes, estimación de poses y generación de imágenes

A pesar de los avances en los modelos de segmentación de imágenes, estimación de poses y generación de imágenes, se analizarán las limitaciones que pueden afectar el rendimiento y la calidad de los probadores virtuales.

En el caso de la segmentación de imágenes, una limitación es que los modelos como U-Net pueden verse afectados por la variabilidad en las posturas, las apariencias y las condiciones de iluminación presentes en las imágenes [\(He et al., 2017\)](#). Esto puede resultar en una segmentación imprecisa, lo que afecta a la calidad de las imágenes generadas en los probadores virtuales.

En cuanto a la estimación de poses, modelos como OpenPose y HRNet pueden enfrentarse a desafíos cuando las personas están parcialmente ocluidas, en poses extremas o en entornos con múltiples personas [\(Cao et al., 2017\)](#). Estas situaciones pueden llevar a una estimación de poses inexacta, lo que dificulta la correcta adaptación de las prendas de vestir virtuales a la postura del cuerpo.

En el ámbito de la generación de imágenes, las GAN y VAE pueden presentar limitaciones en términos de calidad y diversidad de las imágenes generadas. Las GAN, por ejemplo, pueden sufrir de colapso de modos, lo que significa que el generador produce imágenes similares o poco variadas [\(Arjovsky et al., 2017\)](#). Además, las GAN pueden generar artefactos visuales o inconsistencias en las imágenes, lo que disminuye la calidad y el realismo de las representaciones de personas con ropa virtual [\(Karras et al., 2019\)](#). La Figura 14 muestra algunos ejemplos donde las GAN generan imágenes con malformaciones o poco realistas. Las VAE tienden a generar imágenes más distintas a la original, pero con más probabilidad de ser inverosímiles y menos detalladas.

*Figura 14: Casos de error en la generación de imágenes con el método de VITON*



Fuente: [\(Han et al., 2018\)](#)

A pesar de las limitaciones, los avances en las técnicas de segmentación de imágenes, estimación de poses y generación de imágenes siguen mejorando la calidad y el rendimiento de los probadores virtuales. Investigaciones futuras en estos campos pueden abordar las limitaciones y mejorar aún más la experiencia del usuario en aplicaciones de moda virtual.

### **3.6. Soluciones a algunas de las limitaciones**

A pesar de las limitaciones mencionadas en los modelos de segmentación de imágenes, estimación de poses y generación de imágenes, la comunidad de investigación ha propuesto varias soluciones para abordar estos desafíos y mejorar la calidad y el rendimiento de los probadores virtuales.

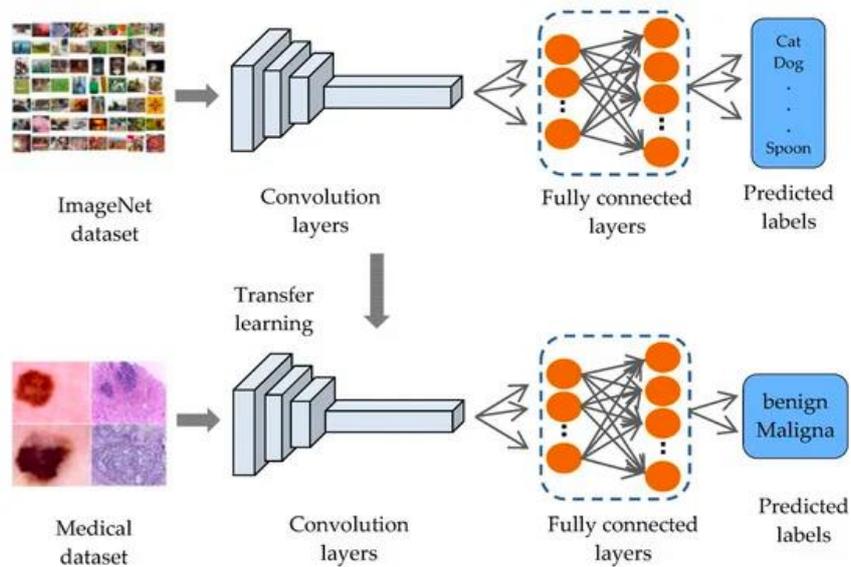
#### **3.6.1. *Transfer learning* para mejorar el rendimiento del modelo**

El aprendizaje por transferencia, denominado *transfer learning*, es una técnica que permite a los modelos de IA aprender de un conjunto de datos previamente trabajado y aplicar ese conocimiento a otro conjunto de datos relacionado. Esta técnica es especialmente útil cuando se tiene una cantidad limitada de datos etiquetados. Permite aprovechar el conocimiento adquirido previamente en un conjunto de datos ya preentrenado, mejorando así el rendimiento y reduciendo el tiempo de entrenamiento.

El aprendizaje por transferencia funciona mediante el uso de las capas iniciales de un modelo previamente entrenado para extraer características generales y relevantes de las imágenes. Luego, se pueden agregar y entrenar capas adicionales para adaptar el modelo a otras tareas específicas, como la segmentación de imágenes, estimación de poses y generación de imágenes ([Kora et al., 2022](#)).

Un estudio reciente de Mukhlif et al. (2023) propone una metodología de *transfer learning* dual en la que se aprovechan dos modelos preentrenados para mejorar el rendimiento en la clasificación de imágenes médicas ([Mukhlif et al., 2023](#)), como se muestra en la Figura 15. Este enfoque innovador demuestra cómo el *transfer learning* puede ser aplicado de manera efectiva en diversos contextos y tareas, sugiriendo su posible potencial para mejorar la generación de imágenes de personas con ropa personalizada.

Figura 15: Aprendizaje por transferencia del dataset de ImageNet aplicado a imágenes médicas



Fuente: [\(Mukhlif et al., 2023\)](#)

La aplicación del aprendizaje por transferencia puede aportar varios beneficios. Entre ellos se encuentra la mejora del rendimiento de los modelos, al aprovechar el conocimiento adquirido previamente en conjuntos de datos similares. Además, se reduce el tiempo de entrenamiento al no tener que entrenar desde cero. Finalmente, también disminuye la necesidad de grandes cantidades de datos etiquetados, lo cual es especialmente útil para casos de uso concretos, como obtener datos de personas con diferentes tipos de ropa y poses.

### 3.6.2. Enfoques de aprendizaje multi-tarea

El aprendizaje multi-tarea es otra de las técnicas utilizadas para abordar las limitaciones de la segmentación, estimación y generación de imágenes. Este enfoque consiste en entrenar un único modelo para realizar varias tareas relacionadas simultáneamente, en lugar de entrenar modelos individuales para cada tarea. Al hacerlo, el modelo puede compartir información y aprender características comunes entre las tareas, mejorando así el rendimiento general y reduciendo el tiempo de entrenamiento [\(Ruder, 2017\)](#).

Una de las ventajas del aprendizaje multi-tarea es que puede mejorar la generalización del modelo al aprender características comunes entre las tareas. También puede reducir el riesgo de sobreajuste, ya que el modelo se entrena en varias tareas y no se centra únicamente en una [\(Kendall et al., 2018\)](#).

Un ejemplo reciente de la aplicación del aprendizaje multi-tarea en la segmentación de imágenes es el estudio de Bischke et al. (2019), en el que se utiliza un enfoque de aprendizaje multi-tarea para la segmentación de huellas de edificios en imágenes satelitales. El modelo propuesto combina múltiples tareas, como la segmentación de edificios y la clasificación de uso del suelo, lo que permite mejorar el rendimiento en la segmentación de las huellas de los edificios.

La adopción de enfoques de aprendizaje multi-tarea podría mejorar la eficiencia y la precisión del modelo, al permitir aprovechar las características compartidas entre las diferentes tareas. Además, el aprendizaje multi-tarea puede mejorar la capacidad de generalización del modelo, lo que podría resultar en un rendimiento más sólido en contextos no vistos previamente.

### 3.7. Tecnologías para la implementación

Para desarrollar un probador virtual que aborde las limitaciones discutidas en los puntos anteriores, es fundamental conocer las tecnologías adecuadas para la implementación. Algunas de las tecnologías clave que pueden ser útiles en este contexto incluyen *frameworks* y bibliotecas de aprendizaje profundo, servicios en la nube y tecnologías web.

#### 3.7.1. *Frameworks* y bibliotecas de aprendizaje profundo

El aprendizaje profundo ha experimentado un rápido crecimiento en los últimos años, y como resultado, han surgido varios *frameworks* y bibliotecas de aprendizaje profundo para facilitar el desarrollo e implementación de modelos de IA.

Algunos de los más populares y ampliamente utilizados incluyen TensorFlow, PyTorch y Keras ([Géron, 2019](#); [Paszke et al., 2019](#)). A continuación, se presenta una descripción detallada de estos tres *frameworks* y cómo podrían adaptarse a este proyecto:

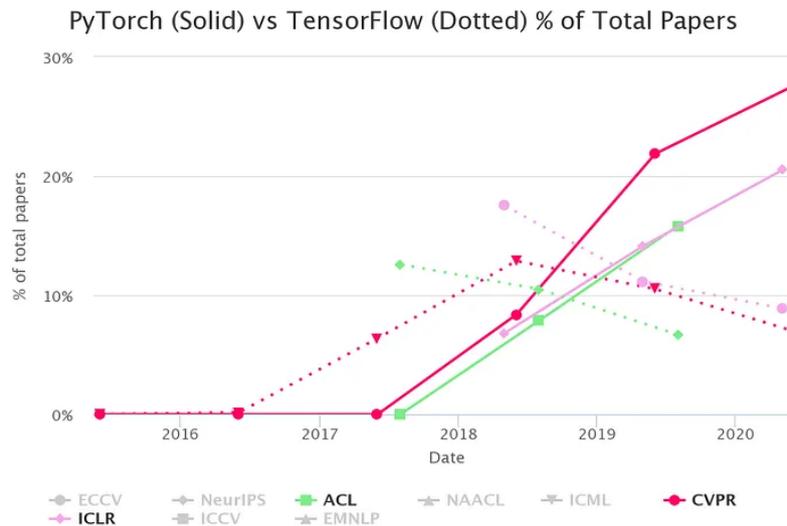
- **TensorFlow:** Desarrollado por Google Brain, TensorFlow es un framework de código abierto para el aprendizaje automático y el aprendizaje profundo ([TensorFlow, 2023](#)). TensorFlow ofrece una amplia gama de herramientas, bibliotecas y recursos para facilitar el diseño, entrenamiento e implementación de modelos de aprendizaje profundo en una variedad de aplicaciones ([Ramsundar & Zadeh, 2020](#)). TensorFlow es especialmente útil debido a su capacidad para manejar fácilmente operaciones con tensores y su flexibilidad para trabajar

con diferentes arquitecturas de redes neuronales, como CNN y GAN. Además, es compatible con la aceleración de hardware, como GPU y TPU, lo que permite un entrenamiento y una inferencia más rápidos ([Géron, 2019](#)).

- **PyTorch:** PyTorch es un framework de aprendizaje profundo desarrollado por Facebook AI Research (FAIR) y es conocido por su facilidad de uso y flexibilidad ([PyTorch, 2023](#)). PyTorch proporciona un enfoque más dinámico para la ejecución de gráficos computacionales en comparación con TensorFlow, lo que facilita la depuración y el prototipado rápido de modelos ([Géron, 2019](#)). PyTorch también permite la implementación de diversas arquitecturas de redes neuronales, como CNN, GAN y VAE. Además, es compatible con la aceleración de hardware y cuenta con una amplia comunidad de desarrolladores que contribuyen con herramientas y recursos adicionales ([Paszke et al., 2019](#)).
- **Keras:** Keras es una biblioteca de alto nivel para el aprendizaje profundo que proporciona una interfaz fácil de usar y simplificada para construir y entrenar modelos de aprendizaje profundo ([Keras, 2023](#)). Keras está diseñado para ser fácil de entender y se ejecuta sobre TensorFlow o Theano como backend ([Géron, 2019](#)). Aunque Keras no ofrece la misma flexibilidad que TensorFlow y PyTorch en términos de personalización de modelos, su simplicidad y facilidad de uso lo convierten en una excelente opción para el prototipado rápido y la implementación de modelos. Keras proporciona soporte para la implementación de arquitecturas de redes neuronales populares como CNN y GAN, y es compatible con la aceleración de hardware ([Ramsundar & Zadeh, 2020](#)).

Al estudiar los distintos *frameworks* o bibliotecas de aprendizaje profundo, hay que tener en cuenta factores como la flexibilidad, la facilidad de uso, la capacidad de personalización y el soporte para la aceleración de hardware. TensorFlow, PyTorch y Keras son opciones sólidas que ofrecen una amplia gama de herramientas y recursos para facilitar la implementación de las técnicas de aprendizaje profundo. En la Figura 16 se comparan el número de veces que se han referenciado estos frameworks en artículos académicos entre 2016 y 2020. La evolución muestra como el número de referencias, y por consiguiente su uso en investigación, han aumentado considerablemente en los últimos años.

Figura 16: Comparativa del número de referencias académicas a TensorFlow y PyTorch



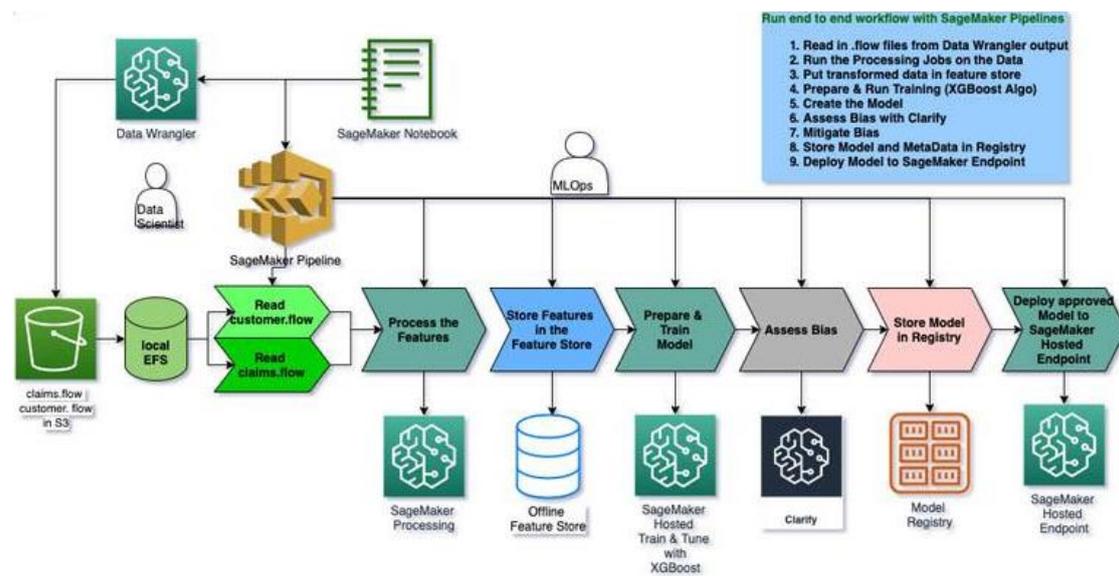
Fuente: [\(He et al., 2020\)](#)

### 3.7.2. Servicios en la nube para el despliegue

La implementación y el despliegue de soluciones basadas en IA, la generación de imágenes de personas usando prendas de vestir personalizadas y estimación de poses, puede beneficiarse significativamente de los servicios en la nube. Estos servicios ofrecen una amplia gama de recursos y herramientas que facilitan el desarrollo, entrenamiento, y despliegue de modelos de aprendizaje profundo, así como el escalado y la gestión de aplicaciones. Algunos de los principales proveedores de servicios en la nube incluyen Amazon Web Services (AWS), Google Cloud Platform (GCP) y Microsoft Azure.

- **Amazon Web Services (AWS)** ofrece una variedad de servicios específicos para el aprendizaje profundo y el machine learning, como Amazon SageMaker, un servicio completamente administrado que permite a los desarrolladores y científicos de datos crear, entrenar y desplegar modelos de machine learning rápidamente [\(AWS, 2023\)](#). AWS también proporciona AWS Deep Learning AMIs (Amazon Machine Images), que son imágenes de máquinas preconfiguradas con entornos de aprendizaje profundo populares como TensorFlow, PyTorch y Keras. En la Figura 17 podemos ver un ejemplo de arquitectura de un modelo de Machine Learning con AWS SageMaker.

Figura 17: Arquitectura del ciclo de vida de un modelo de Machine Learning en AWS SageMaker



Fuente: [AWS Blog, 2021](#)

- **Google Cloud Platform (GCP)** también ofrece una amplia gama de servicios en la nube para el aprendizaje profundo y el machine learning ([GCP, 2023](#)). Google Cloud AI Platform es un conjunto de herramientas y servicios para el desarrollo, el entrenamiento y la implementación de modelos de machine learning en la nube. Además, GCP proporciona TensorFlow Enterprise, una plataforma optimizada y de alto rendimiento para TensorFlow en la nube de Google, facilitando el despliegue y la escalabilidad de los modelos.
- **Microsoft Azure**, por otro lado, ofrece Azure Machine Learning, un servicio basado en la nube que permite a los científicos de datos y desarrolladores crear, entrenar y desplegar modelos de machine learning utilizando diversas bibliotecas y *frameworks*, incluyendo TensorFlow, PyTorch y Keras ([Azure, 2023](#)). Azure también ofrece una amplia gama de servicios de IA, como Computer Vision, que facilitan el desarrollo de soluciones en este campo.

### 3.7.3. Tecnologías web y aplicaciones móviles

Las tecnologías web y las aplicaciones móviles juegan un papel crucial en la implementación y el acceso a soluciones basadas en IA. Permiten a los usuarios interactuar con los modelos a través de interfaces de usuario intuitivas y accesibles en diferentes dispositivos y plataformas.

Para el desarrollo de aplicaciones web, existen múltiples *frameworks* y bibliotecas disponibles que facilitan la creación de interfaces de usuario y la comunicación con los modelos de aprendizaje profundo. Algunos de los *frameworks* más populares son React ([React, 2023](#)) o Angular ([Angular, 2023](#)). Estos *frameworks* facilitan la creación de interfaces de usuario completas y dinámicas, lo que podría permitir a los usuarios cargar imágenes y recibir resultados del modelo en tiempo real.

En cuanto a las aplicaciones móviles, existen varias opciones para el desarrollo de aplicaciones nativas y multiplataforma. Para el desarrollo nativo, los desarrolladores pueden utilizar Swift para iOS ([Swift, 2023](#)) y Kotlin para Android ([Kotlin, 2023](#)). Ambos lenguajes de programación ofrecen un alto rendimiento y acceso a todas las características específicas de la plataforma. Para el desarrollo multiplataforma, React Native ([React Native, 2023](#)) y Flutter ([Flutter, 2023](#)) son opciones populares que permiten a los desarrolladores crear aplicaciones móviles para iOS y Android utilizando un único código base. Estos frameworks también proporcionan acceso a bibliotecas y herramientas para la integración de modelos de aprendizaje profundo y servicios en la nube en aplicaciones móviles.

### **3.8. Limitaciones de las tecnologías para la implementación**

Aunque las tecnologías mencionadas anteriormente pueden ser útiles para desarrollar un probador virtual eficiente y eficaz, también es importante tener en cuenta sus limitaciones. Estas limitaciones pueden afectar la calidad, el rendimiento y la accesibilidad del probador virtual en diferentes situaciones.

#### **3.8.1. Dependencia del hardware**

Los modelos de IA, como CNN y GAN, dependen en gran medida del hardware, como las unidades de procesamiento gráfico (GPU) y las unidades de procesamiento tensorial (TPU), para llevar a cabo cálculos intensivos en paralelo de manera eficiente ([Chen et al., 2015](#)). Las GPU y las TPU son dispositivos de hardware especializados diseñados para acelerar las operaciones matriciales y tensoriales comunes en el aprendizaje profundo. Sin embargo, su adquisición y mantenimiento pueden ser costosos, especialmente para pequeñas empresas y proyectos de investigación.

Una posible solución para abordar la dependencia del hardware es utilizar servicios en la nube, como AWS, GCP y Azure, vistos anteriormente, que ofrecen infraestructura de hardware escalable y optimizada para el aprendizaje profundo. Otra opción es utilizar técnicas de optimización de modelos,

como la cuantización y la poda, para reducir la dependencia del hardware y mejorar la eficiencia computacional ([Zhang et al., 2018](#)). Estas técnicas pueden ser especialmente útiles para implementaciones en dispositivos móviles y dispositivos con recursos limitados.

### 3.8.2. Tiempos de entrenamiento prolongados

El entrenamiento de modelos de aprendizaje profundo puede requerir tiempos de entrenamiento prolongados debido a la gran cantidad de parámetros y capas que componen estos modelos. Los tiempos de entrenamiento largos pueden presentar desafíos en términos de costos y disponibilidad de recursos para las aplicaciones que implementan estos modelos.

Una solución para abordar los tiempos de entrenamiento prolongados es utilizar técnicas de aprendizaje acelerado, como el aprendizaje distribuido y el aprendizaje federado. El aprendizaje distribuido divide el proceso de entrenamiento entre múltiples dispositivos de hardware, como GPU y TPU, para acelerar el entrenamiento y mejorar la eficiencia. El aprendizaje federado, por otro lado, permite entrenar modelos de manera colaborativa entre varios dispositivos, sin compartir datos brutos, lo que puede ser útil para proteger la privacidad de los datos y reducir la carga de comunicación ([Yang et al., 2019](#)).

Otra solución es utilizar *transfer learning*, mencionado anteriormente, que permite aprovechar modelos previamente entrenados en tareas relacionadas para acelerar el entrenamiento en una tarea específica ([Kora et al., 2022](#)). El *transfer learning* puede reducir significativamente el tiempo de entrenamiento y mejorar el rendimiento del modelo al aprovechar el conocimiento adquirido en tareas previas.

### 3.8.3. Privacidad y seguridad de los datos

La privacidad y seguridad de los datos es una preocupación fundamental en la implementación de tecnologías de aprendizaje profundo. La recopilación, el almacenamiento y el procesamiento de datos de usuarios, especialmente de imágenes y de datos personales, deben cumplir con las leyes y regulaciones de protección de datos, como el Reglamento General de Protección de Datos (GDPR) en la Unión Europea ([European Parliament & Council of the European Union, 2018](#)).

La conformidad con el GDPR implica asegurar la transparencia en la recopilación y uso de datos personales, obtener el consentimiento informado de los usuarios y garantizar la privacidad y seguridad de los datos en todas las etapas del procesamiento. En el contexto de este proyecto, es imprescindible garantizar que los usuarios estén informados sobre cómo se utilizarán sus imágenes y datos personales, y que se obtenga su consentimiento antes de la recopilación y procesamiento de dichos datos. Además, la utilización de técnicas de anonimización puede contribuir a la protección de la identidad y la privacidad de los usuarios.

#### **3.8.4. Adaptabilidad a diferentes dominios y estilos**

La adaptabilidad a diferentes dominios y estilos de ropa es un desafío clave en la implementación de probadores virtuales basados en aprendizaje profundo. Los modelos de aprendizaje profundo, como las CNN y GAN, requieren una gran cantidad de datos etiquetados para entrenamiento y pueden ser sensibles a cambios en los dominios y estilos de ropa. Los cambios en la moda, las tendencias y los estilos de ropa pueden afectar la capacidad de los modelos entrenados para generalizar y proporcionar resultados precisos en la generación y adaptación de imágenes de ropa en diferentes dominios ([Zhu et al., 2017](#)).

Una solución para abordar este desafío es el uso de *transfer learning*, mencionado en varias ocasiones, que permite a los modelos aprovechar los conocimientos adquiridos en tareas relacionadas para mejorar la adaptabilidad y el rendimiento en nuevos dominios y estilos de ropa ([Kora et al., 2022](#)). Además, el uso de datos sintéticos y aumentados, junto con técnicas de regularización y arquitecturas de redes más eficientes, puede mejorar la capacidad de generalización de los modelos y su adaptabilidad a diferentes dominios y estilos de ropa.

### **3.9. Resumen del estado del arte**

En los últimos años, se ha observado un avance significativo en las herramientas y plataformas de generación de imágenes con IA. La investigación en este campo se ha centrado en diversas áreas clave, que incluyen la segmentación de imágenes, la estimación de poses, la generación de imágenes, y la aplicación de enfoques como el *transfer learning* y el aprendizaje multitarea.

En la segmentación de imágenes, las CNN han demostrado ser efectivas para identificar y separar personas y ropa en imágenes ([Wang et al., 2018](#)). La estimación de poses también ha experimentado

avances significativos, con enfoques como DensePose ([Güler et al., 2018](#)) que proporcionan una representación detallada del cuerpo humano en imágenes.

La generación de imágenes ha sido impulsada por el desarrollo de GAN y VAE. Los GAN son especialmente útiles para generar imágenes realistas y de alta calidad, mientras que los VAE ofrecen un enfoque probabilístico que puede ayudar a aprender representaciones latentes de los datos ([El-Kaddoury et al., 2019](#)).

El *transfer learning* ha surgido como una solución eficaz para superar la escasez de datos etiquetados en dominios específicos, permitiendo el uso de modelos previamente entrenados en tareas similares para mejorar el rendimiento ([Mukhlif et al., 2023](#)). Además, el aprendizaje multitarea permite entrenar modelos para realizar varias tareas simultáneamente, mejorando la eficiencia y el rendimiento del modelo ([Bischke et al., 2019](#)).

En cuanto a las tecnologías para la implementación, los *frameworks* de aprendizaje profundo como TensorFlow, PyTorch y Keras son ampliamente utilizados en la comunidad de investigadores y desarrolladores. Los servicios en la nube, como AWS, Google Cloud y Microsoft Azure, proporcionan recursos de cómputo y almacenamiento escalables y asequibles para el entrenamiento y la implementación de modelos de aprendizaje profundo. Las tecnologías web y aplicaciones móviles permiten el desarrollo de interfaces de usuario atractivas e interactivas para probadores virtuales.

Sin embargo, también existen limitaciones en la implementación de estas tecnologías, como la dependencia del hardware, los tiempos de entrenamiento prolongados y la adaptabilidad a diferentes dominios y estilos. Además, la privacidad y la seguridad de los datos son preocupaciones importantes, ya que las leyes de protección de datos, como el GDPR, deben ser tenidas en cuenta al desarrollar e implementar probadores virtuales.

## 4. Objetivos y metodología del trabajo

El desarrollo y la implementación de este proyecto requieren una metodología y objetivos definidos de manera precisa. El rápido avance de la inteligencia artificial y del aprendizaje profundo ha proporcionado herramientas valiosas y enfoques innovadores para abordar este desafío. A continuación, se presenta un análisis detallado de los objetivos generales y específicos, la metodología de trabajo y las herramientas utilizadas, fundamentales para el éxito de este proyecto.

### 4.1. Objetivo general

El objetivo general de este proyecto es **desarrollar e implementar un sistema de probadores virtuales basado en inteligencia artificial que permita a los usuarios visualizar de forma precisa y realista cómo se verían con diferentes prendas de vestir**. El usuario facilitará las imágenes, tanto de la prenda como del usuario, y la aplicación generará una imagen sintética realista de cómo se ajustaría la prenda al cuerpo del usuario. Este sistema debe estar diseñado de tal manera que garantice un tiempo de respuesta rápido, no superior a cinco segundos, desde la carga de las imágenes hasta la generación de la imagen sintética. Además, el sistema debe cumplir con las normas de protección de datos (como el GDPR) para garantizar la privacidad y seguridad de los datos personales de los usuarios.

Como medida de éxito, se establece que el sistema obtenga una calificación promedio de satisfacción del usuario de al menos 4 sobre 5 en la primera versión de prueba.

### 4.2. Objetivos específicos

Los objetivos específicos del proyecto son los siguientes:

1. **Estudiar las técnicas IA:** Profundizar en las técnicas actuales en inteligencia artificial, con énfasis en la generación de imágenes, segmentación de imágenes y estimación de poses.
2. **Seleccionar las herramientas IA:** Identificar y escoger las bibliotecas y frameworks de inteligencia artificial óptimos para el desarrollo del sistema de probadores virtuales.

3. **Crear un *dataset*:** Generar un conjunto de datos adecuado que facilite el entrenamiento y evaluación de los algoritmos y modelos utilizados en el sistema de probadores virtuales.
4. **Desarrollar algoritmos de segmentación, estimación y generación de imágenes:** Crear e implementar algoritmos de segmentación de imágenes y estimación de poses para un ajuste preciso de la prenda a la imagen del usuario. Construir y entrenar un modelo de generación de imágenes capaz de elaborar imágenes sintéticas realistas que muestren al usuario con la prenda seleccionada.
5. **Adaptarse a la protección de datos:** Establecer medidas de protección de datos y privacidad para asegurar el cumplimiento con las regulaciones de protección de datos, como el GDPR.
6. **Desarrollar una página web para interactuar con la solución:** Desarrollar una página web que contenga el modelo entrenado, una base de datos de almacenamiento de imágenes y la interfaz virtual para que el usuario pruebe la solución propuesta.
7. **Evaluar la solución propuesta:** Llevar a cabo pruebas de calidad y conducir pruebas de usabilidad para evaluar la facilidad de uso, la intuitividad y la satisfacción general de la solución desarrollada.

### 4.3. Metodología de trabajo

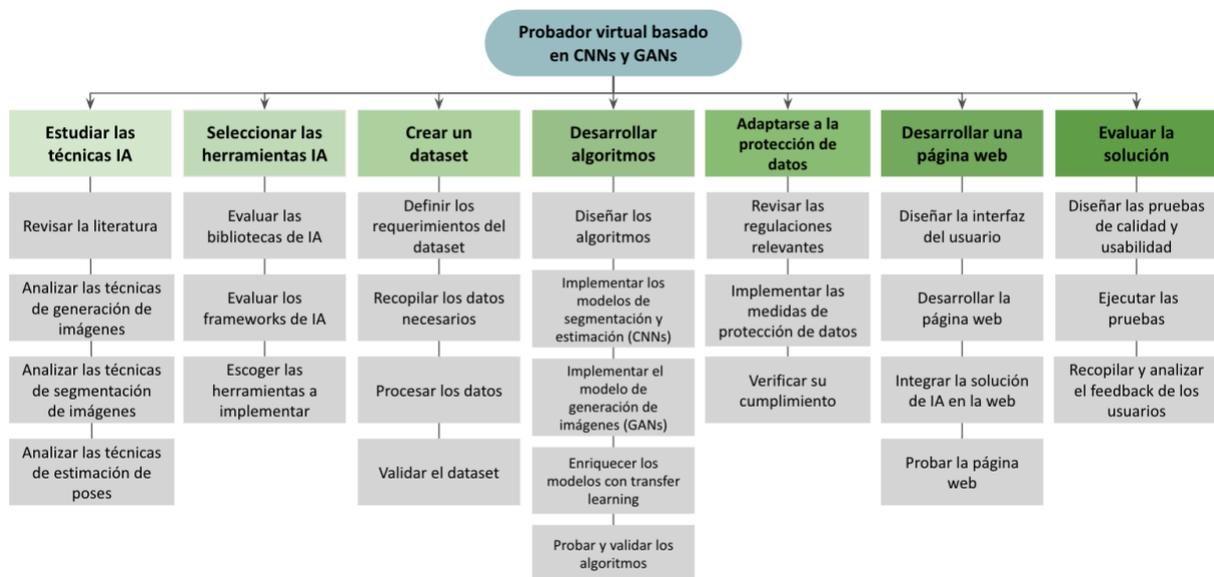
La metodología de trabajo adoptada para este proyecto se basa en un enfoque iterativo y centrado en el usuario, combinando las mejores prácticas de las metodologías ágiles y la ingeniería de software.

Aun siendo un proceso iterativo, se puede definir la hoja de ruta que se seguirá para alcanzar los objetivos específicos, culminando en el logro del objetivo general. Esta metodología aborda el proceso de investigación y análisis, el diseño y desarrollo, la implementación, las pruebas y la evaluación.

### 4.3.1. Estructura de desglose de trabajo

Una Estructura de Desglose de Trabajo (EDT) es una herramienta de gestión de proyectos que divide un proyecto en tareas más pequeñas y manejables, facilitando así su planificación, gestión y control. En este proyecto, la EDT ayudará a desglosar los objetivos específicos en subtareas, definiendo claramente qué se requiere para alcanzar cada objetivo. En la Figura 18 se muestra el EDT correspondiente a este proyecto:

Figura 18: EDT del proyecto



#### 4.3.1.1. Descripción de tareas

En la Tabla 1 se muestra la descripción de las tareas con la fecha de inicio y fin de la realización:

Tabla 1: Descripción de tareas con fecha de inicio y fin

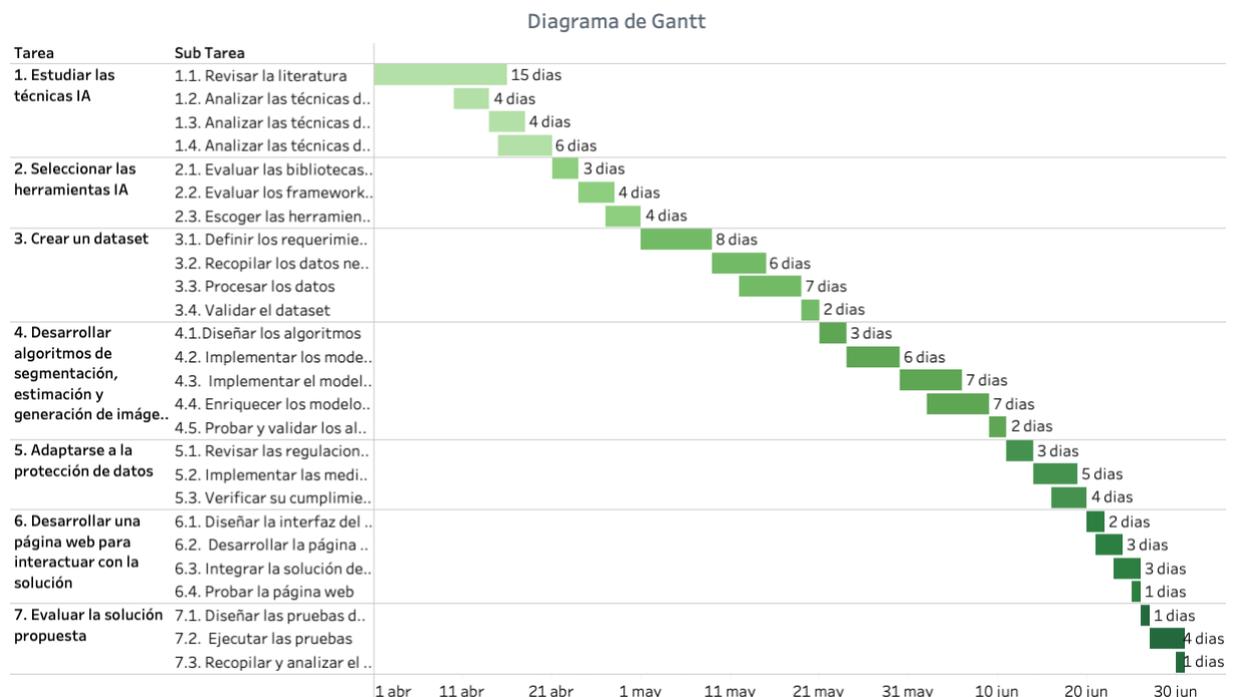
Tarea	Fecha Inicio	Fecha Fin
<b>1. Estudiar las técnicas IA</b>	<b>1/4/23</b>	<b>20/4/23</b>
1.1. Revisar la literatura	1/4/23	15/4/23
1.2. Analizar las técnicas de generación de imágenes	10/4/23	13/4/23
1.3. Analizar las técnicas de segmentación de imágenes	14/4/23	17/4/23
1.4. Analizar las técnicas de estimación de poses	15/4/23	20/4/23
<b>2. Seleccionar las herramientas IA</b>	<b>21/4/23</b>	<b>30/4/23</b>
2.1. Evaluar las bibliotecas de IA	21/4/23	23/4/23
2.2. Evaluar los frameworks de IA	24/4/23	27/4/23
2.3. Escoger las herramientas a implementar	27/4/23	30/4/23
<b>3. Crear un dataset</b>	<b>1/5/23</b>	<b>20/5/23</b>
3.1. Definir los requerimientos del dataset	1/5/23	8/5/23
3.2. Recopilar los datos necesarios	9/5/23	14/5/23
3.3. Procesar los datos	12/5/23	18/5/23
3.4. Validar el dataset	19/5/23	20/5/23

<b>4. Desarrollar algoritmos de segmentación, estimación y generación de imágenes</b>	<b>21/5/23</b>	<b>10/6/23</b>
4.1. Diseñar los algoritmos	21/5/23	23/5/23
4.2. Implementar los modelos de segmentación y estimación (CNNs)	24/5/23	29/5/23
4.3. Implementar el modelo de generación de imágenes (GANs)	30/5/23	5/6/23
4.4. Enriquecer los modelos con transfer learning	2/6/23	8/6/23
4.5. Probar y validar los algoritmos	9/6/23	10/6/23
<b>5. Adaptarse a la protección de datos</b>	<b>11/6/23</b>	<b>19/6/23</b>
5.1. Revisar las regulaciones relevantes	11/6/23	13/6/23
5.2. Implementar las medidas de protección de datos	14/6/23	18/6/23
5.3. Verificar su cumplimiento	16/6/23	19/6/23
<b>6. Desarrollar una página web para interactuar con la solución</b>	<b>20/6/23</b>	<b>25/6/23</b>
6.1. Diseñar la interfaz del usuario	20/6/23	21/6/23
6.2. Desarrollar la página web	21/6/23	23/6/23
6.3. Integrar la solución de IA en la web	23/6/23	25/6/23
6.4. Probar la página web	25/6/23	25/6/23
<b>7. Evaluar la solución propuesta</b>	<b>26/6/23</b>	<b>30/6/23</b>
7.1. Diseñar las pruebas de calidad y usabilidad	26/6/23	26/6/23
7.2. Ejecutar las pruebas	27/6/23	30/6/23
7.3. Recopilar y analizar el feedback de los usuarios	30/6/23	30/6/23

### 4.3.1.2. Diagrama de Gantt

El Diagrama de Gantt se utiliza para visualizar la dedicación prevista para cada tarea y subtarea dentro de un proyecto. En la Figura 19 se puede ver el Diagrama de Gantt para este proyecto, previsto en un margen de 3 meses.

Figura 19: Diagrama de Gantt del proyecto



### **4.3.2. Control de calidad, pruebas y validación del modelo de IA**

El control de calidad y las pruebas desempeñan un papel crucial en la metodología de trabajo. Específicamente, se realizarán pruebas y validaciones tanto a nivel de los componentes individuales del sistema como del sistema en su totalidad.

El control de calidad del modelo de IA implica pruebas de rendimiento y precisión del algoritmo de generación de imágenes. Se establecerán métricas de rendimiento y precisión basadas en las necesidades del proyecto y se realizarán pruebas continuas durante el desarrollo para asegurar que estas métricas se mantengan.

La validación de la solución final implicará pruebas de usabilidad con usuarios reales, así como la recopilación de *feedback* para entender cómo el sistema se desempeña en un escenario real y cómo los usuarios interactúan con él.

## **4.4. Herramientas, bibliotecas y librerías utilizadas**

Para lograr los objetivos de este proyecto, se emplearán una variedad de herramientas, bibliotecas y librerías, así como diversas fuentes de datos.

### **4.4.1. Fuentes de datos utilizadas**

Las fuentes de datos para este proyecto consistirán principalmente en conjuntos de datos de imágenes de personas con distintas prendas de vestir. Las imágenes representarán a individuos de diferentes tamaños, formas y en distintas posiciones. Los conjuntos de datos serán extraídos de bases de dominio público, como el *dataset* de moda DeepFashion ([Ge et al., 2019](#)), ya utilizado en proyectos de probadores virtuales como VITON ([Han et al., 2018](#)).

El uso de esta base de datos requiere de tiempo y recursos para categorizar y procesar cada imagen. Todas las imágenes utilizadas estarán sujetas a las correspondientes regulaciones de privacidad y consentimiento para garantizar el respeto a los derechos de los individuos representados en las mismas.

#### 4.4.2. Herramientas utilizadas

En cuanto a las herramientas, se utilizará Python como lenguaje de programación principal debido a su versatilidad y amplio soporte en el campo de la inteligencia artificial y el aprendizaje automático. El entrenamiento de los modelos se realiza en plataformas del estilo Google Colab ([Bisong, 2019](#)), que proporcionan recursos computacionales suficientes para el procesamiento de grandes volúmenes de datos.

Para el desarrollo de la página web, se combinará el lenguaje Python y frameworks especializados para la creación de páginas web como Django y Flask ([Flask, 2023](#)).

#### 4.4.3. Bibliotecas y librerías utilizadas

En el desarrollo de los algoritmos de segmentación de imágenes, estimación de poses y generación de imágenes, se emplearán diversas bibliotecas y librerías. [TensorFlow](#) ([TensorFlow, 2023](#)) y [PyTorch](#) ([PyTorch, 2023](#)) son dos de las bibliotecas de aprendizaje profundo más utilizadas que se utilizan para este proyecto. También se utilizar librerías adicionales para tareas específicas, como [OpenCV](#) ([OpenCV, 2023](#)) para el procesamiento de imágenes, [Pandas](#) ([Pandas, 2023](#)) para la manipulación de datos y [Matplotlib](#) ([Matplotlib, 2023](#)) para la visualización de resultados. Además, para garantizar la conformidad con las normas de protección de datos, utilizamos la biblioteca [PyDP](#) ([PyDP, 2023](#)) que proporciona funcionalidades para el cifrado de datos y el mantenimiento de la privacidad.

## 5. Identificación de requisitos

En este capítulo, se identifican los requisitos funcionales y no funcionales de nuestra solución, así como las posibles restricciones del propio sistema. Esta identificación de requisitos se ha basado en una comprensión detallada del problema que nuestra aplicación está diseñada para resolver, así como en la consideración del contexto habitual en el que se utilizará la aplicación.

### 5.1. Requisitos funcionales

Los requisitos funcionales son las funcionalidades que debe proporcionar el sistema. En nuestro caso, se incluirían:

- FR1.** La aplicación debe ser capaz de generar imágenes realistas de los usuarios vistiendo diferentes prendas de vestir.
- FR2.** La aplicación debe permitir a los usuarios subir sus propias imágenes para la generación de imágenes.
- FR3.** La aplicación debe proporcionar una interfaz de usuario intuitiva y fácil de usar para interactuar con la solución de IA.
- FR4.** La aplicación debe proporcionar una visualización clara y comprensible de las imágenes generadas.

### 5.2. Requisitos no-funcionales

Los requisitos no funcionales son características que el sistema debe tener, como la eficiencia, la usabilidad, la privacidad, etc. En este caso, son:

- NFR1.** El sistema debe garantizar la privacidad de los datos del usuario y cumplir con las regulaciones de protección de datos, como el GDPR.
- NFR2.** El sistema debe ser eficiente y proporcionar tiempos de respuesta razonables.
- NFR3.** El sistema debe ser robusto y capaz de manejar errores y problemas de manera adecuada.

### **5.3. Restricciones del sistema**

Las restricciones del sistema son las limitaciones bajo las cuales debe operar el sistema. Para este proyecto se identifican las siguientes:

- SR1.** La solución debe ser implementada como una aplicación web, para facilitar el acceso y uso de los usuarios.
- SR2.** El sistema debe funcionar dentro de los límites de recursos disponibles, como el almacenamiento y la capacidad de computación.

## 6. Descripción de la herramienta software desarrollada

En esta sección, se describe en profundidad la herramienta de software desarrollada para este proyecto. Esta solución, basada en técnicas avanzadas de inteligencia artificial, permite transformar imágenes de personas en una segmentación detallada y una estimación de pose para, finalmente, generar una nueva imagen con una prenda de vestir distinta. Además, proporciona una interfaz web intuitiva y fácil de usar para interactuar con estos modelos de IA. En las siguientes secciones, se explorarán los componentes clave de esta herramienta y se describirá el proceso seguido para su desarrollo.

### 6.1. Desarrollo del Sistema de Inteligencia Artificial

El *core* de este proyecto reside en un sistema de inteligencia artificial robusto y eficiente que se encarga de procesar y transformar las imágenes de los usuarios. A continuación, se detalla cada paso en el proceso de desarrollo de este sistema de IA.

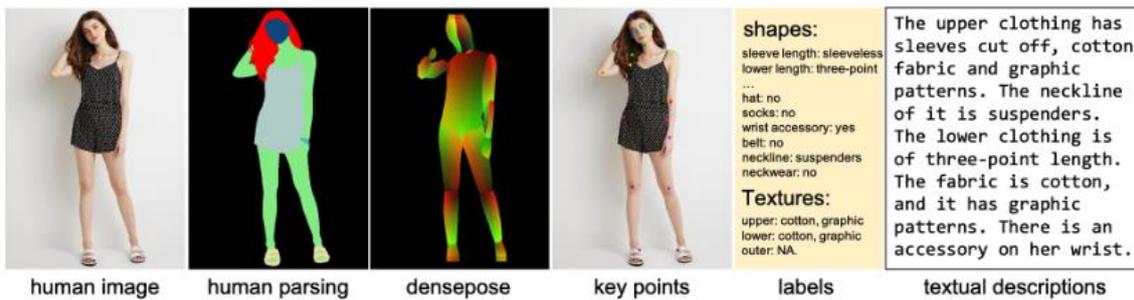
#### 6.1.1. Generación de un *dataset* adecuado

A continuación, se detallan los pasos para la generación del *dataset*.

##### 6.1.1.1. Adquisición de Datos

En la etapa de adquisición de datos, se utilizó el *dataset* DeepFashion-MultiModal ([Jiang et al., 2022](#)), un recurso amplio y de alta calidad por sus anotaciones. Este conjunto de datos consiste en 44.096 imágenes humanas de alta resolución. También contiene etiquetas manualmente anotadas que segmentan 12.701 de las imágenes originales en 24 categorías distintas. Además, cada imagen viene con anotaciones de puntos clave y descripciones textuales. El *dataset* también incluye una representación del modelo DensePose ([Güler et al., 2018](#)) que extrae para cada imagen un mapa detallado de la pose y la forma del cuerpo humano. En la Figura 20 se puede ver una muestra de los diferentes tipos de datos que contiene el *dataset*. En total, esta base de datos ocupa aproximadamente 12GB de memoria.

Figura 20: Muestra de los diferentes tipos de datos de DeepFashion-MultiModal



Fuente: [\(Jiang et al., 2022\)](#)

### 6.1.1.2. Procesamiento de los datos

El procesamiento de los datos es esencial para la generación de un conjunto de datos adecuado. Para este proyecto se utilizan tres componentes del conjunto de datos DeepFashion-MultiModal: (1) Las imágenes originales, (2) las imágenes segmentadas, (3) las coordenadas que ayudan a identificar la postura.

Las imágenes originales de este *dataset* son 44.096 archivos con la extensión .jpg que ocupan en total 5,6 GB. Tienen unas dimensiones de 750 x 1101 píxeles x 3 (los 3 canales RGB). Aunque en los proyectos de Deep Learning es común trabajar en escala de grises, en este caso de uso es importante mantener los canales de color para reproducir imágenes con los mismos tonos. Para reducir el coste computacional de procesar las imágenes, se reduce la dimensionalidad de estas a 224 x 350 píxeles x 3 (aprox. un tercio de las medidas originales). En la Figura 21 se puede ver un ejemplo del tipo de imágenes que se puede encontrar en este *dataset*.

Figura 21: Subconjunto del dataset de imágenes originales





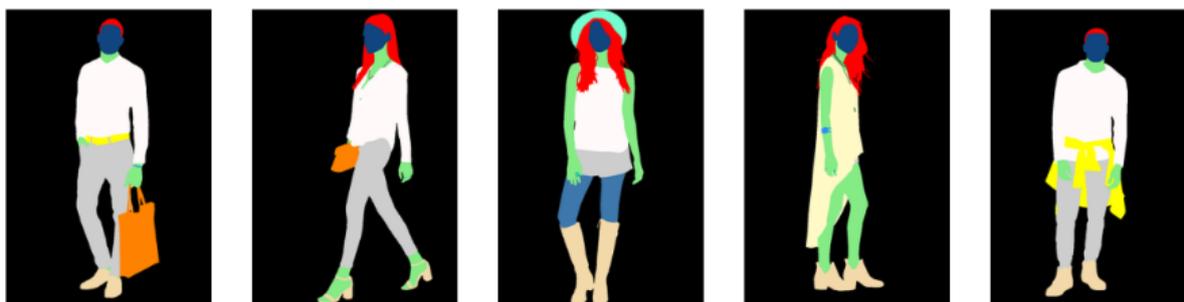
Las imágenes segmentadas, por su parte, son clasificaciones generadas manualmente y corresponden a la segmentación de las imágenes originales. Cada parte de la vestimenta en la imagen esta etiquetada con una de las 24 categorías disponibles, las cuales se pueden ver en la Tabla 2:

Tabla 2: Etiquetas de las imágenes segmentadas por DeepFashion-MultiModal

0	fondo	1	parte superior	2	chaquetas	3	faldilla	4	vestido
5	pantalón	6	leggings	7	gorro/sombrero	8	gafas de sol	9	collares
10	cinturon	11	calzado	12	mochila	13	cabello	14	cara
15	piel	16	anillo	17	pulsera/reloj	18	calcetines	19	guante
20	collar	21	mono	22	pendiente	23	corbata		

El *dataset* de imágenes segmentadas está compuesto de 12.701 ejemplos, los cuales tenían una dimensión original de 750 x 1101 pixeles x 1 (categoría de la imagen) y han sido redimensionados a 224 x 350 pixeles x 3. En la Figura 22 se pueden ver algunos ejemplos de las imágenes segmentadas en las categorías anteriores:

Figura 22: Subconjunto de las imágenes segmentas



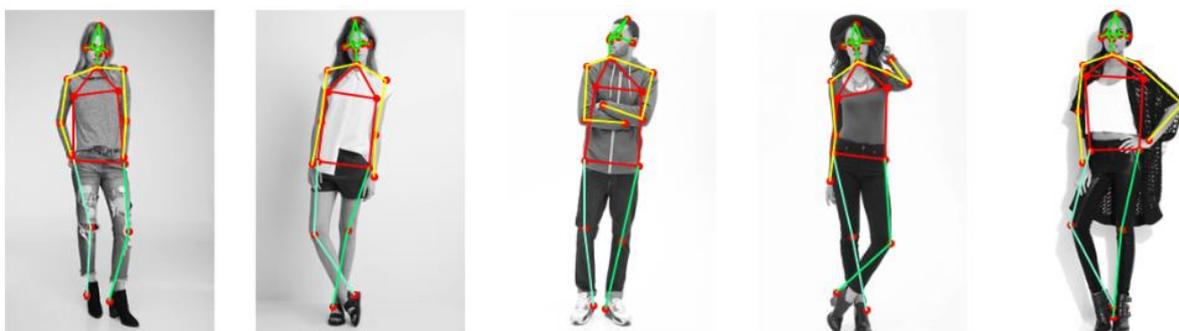
Por último, las coordenadas para estimar la posición de la persona también son anotaciones manuales, y forman un conjunto de 21 puntos para cada imagen (42 coordenadas, teniendo en cuenta que son en 2D). Al igual que las imágenes, es necesario redimensionar las coordenadas para mantener la

proporción. Para gestionar este tipo de datos y visualizarlos en las imágenes, se transformaron las anotaciones a una visualización 2D más intuitiva superpuesta a la imagen original, como se puede ver en la Figura 23 y Figura 24.

Figura 23: Visualización de la posición a partir de las coordenadas



Figura 24: Subconjunto de la visualización de posiciones



En un principio, se planteó el aplicar *Data Augmentation* (aumento de los datos) para aumentar el volumen de los datos de entrenamiento y para introducir diversidad adicional, como diferentes ángulos o condiciones de iluminación. No obstante, estas características ya están presentes en el conjunto de datos actual. Además, la transformación de las anotaciones detalladas que acompañan a cada imagen, incluyendo las etiquetas de segmentación y los puntos clave de las poses, añadiría una

capa de complejidad al intentar aumentar los datos. Este factor, unido al coste computacional adicional que implicaría, ha conducido a la decisión de no aplicarla en este proyecto. En su lugar, los recursos se han dirigido hacia el entrenamiento y optimización de los modelos.

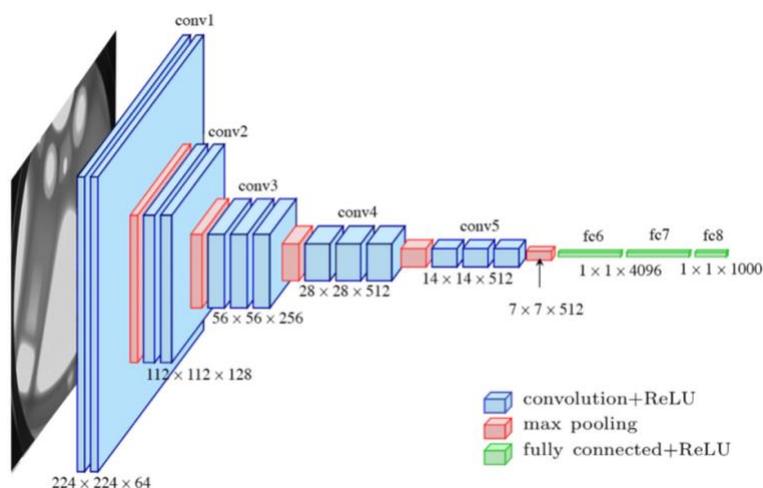
### 6.1.2. Modelo de segmentación de imágenes con CNN

Para este proyecto, se ha desarrollado un modelo de segmentación de imágenes basado en CNN. Dicho modelo tiene el objetivo de identificar y clasificar diferentes regiones de una imagen según las características de las prendas puestas.

#### 6.1.2.1. Selección del modelo candidato para *Transfer Learning*

El modelo utilizado para la extracción de características en la fase de codificación es VGG16 (Pre-trained Visual Geometry Group 16) (Simonyan et al., 2014). La selección de este modelo se basó en su capacidad para extraer características de imágenes a varios niveles debido a su profundidad y arquitectura, lo que facilita el aprendizaje de representaciones más complejas de los datos. Las capas del modelo base se han ajustado para ser no entrenables, por lo tanto, no se actualizarán durante el entrenamiento. En la Figura 25 se puede observar el detalle por capas de la arquitectura del modelo VGG16.

Figura 25: Arquitectura del modelo pre-entrenado VGG16



Fuente: (Le, K., 2022)

### 6.1.2.2. Arquitectura del modelo desarrollado

La arquitectura final del modelo de segmentación de imágenes se ha inspirado en redes de tipo U-Net ([Giang et al., 2020](#)). Consiste en un codificador (parte de contracción) y un decodificador (parte de expansión).

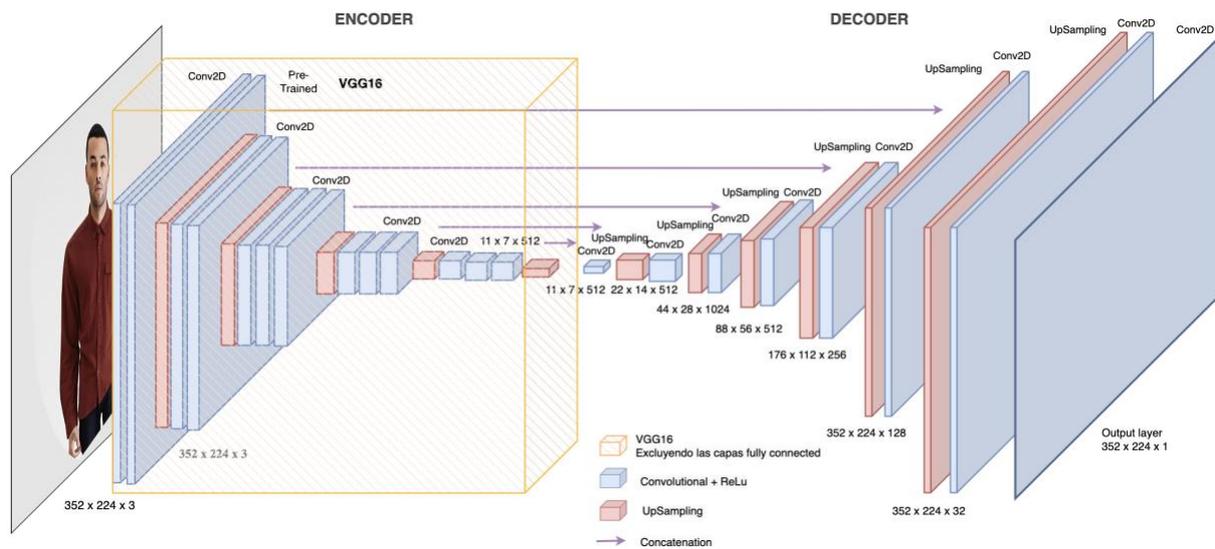
En la parte del codificador se utiliza el modelo por *transfer learning*, VGG16, excluyendo las capas superiores (*fully-connected layers*). Para extraer características de las imágenes de entrada, se mantienen sus cinco bloques de convolución y el último bloque de *pooling*.

En cuanto al decodificador, su función es la de reconstruir la imagen a partir de las características extraídas por el codificador, produciendo finalmente una imagen segmentada. El decodificador se compone de una serie de capas de convolución (Conv2D) y de aumento de la resolución (UpSampling2D). Estas capas permiten incrementar gradualmente la dimensión espacial de las salidas del codificador hasta alcanzar la dimensión de la imagen original.

Además, se utiliza una operación de concatenación en cada etapa del decodificador para fusionar las características del decodificador con las características de las correspondientes capas del codificador. Esta operación se conoce como una conexión de salto y es crucial para permitir que la red aprenda de las características a diferentes niveles de abstracción.

Finalmente, el modelo tiene una capa de convolución final que utiliza una función de activación softmax para producir la segmentación final de la imagen. Esta capa se encarga de clasificar cada píxel de la imagen en las diferentes clases definidas para la segmentación, incluyendo el fondo. Así, el resultado final es una representación de la imagen original en la que cada píxel ha sido clasificado en una de las categorías definidas. En la Figura 26 se puede observar el detalle por capas del modelo final para la segmentación de imágenes.

Figura 26: Arquitectura del modelo de segmentación de imágenes



### 6.1.2.3. Entrenamiento del modelo

El entrenamiento del modelo se realiza con el objetivo de ajustar los parámetros del decodificador de la red neuronal, con base en el error de predicción que genera el modelo en comparación con las segmentaciones verdaderas. Se utilizó una división de los datos en un conjunto de entrenamiento y uno de validación. El 80% de las imágenes se utilizó para el entrenamiento y el restante 20% para la validación. De esta manera, se puede monitorizar el rendimiento del modelo en datos no vistos durante el entrenamiento, consiguiendo una medida más realista de la capacidad de generalización del modelo.

La búsqueda de un modelo óptimo fue un proceso iterativo que implicó la exploración de diferentes configuraciones e hiperparámetros. Se llevaron a cabo varios experimentando con diversas tasas de aprendizaje, tamaños de lote y funciones de activación, entre otros. Durante este proceso de optimización, se utilizó el algoritmo de optimización Adam. Después de un riguroso proceso de prueba y error, se llegó a la configuración final del modelo, que es la que se presenta y discute en este documento.

El criterio que guía la optimización es la minimización de la función de pérdida de entropía cruzada categórica, que es comúnmente utilizada para problemas de clasificación de múltiples clases. Esta función cuantifica la diferencia entre la distribución de probabilidad generada por el modelo y la distribución de probabilidad verdadera (es decir, las etiquetas verdaderas de los píxeles).

En el entrenamiento con mejores resultados, se utilizó un tamaño de lote de 16 imágenes, lo que significa que los pesos del modelo se actualizan después de haber calculado la pérdida sobre 16 imágenes. Además, el entrenamiento se llevó a cabo durante casi 38 horas, correspondientes a 10 épocas. En la Figura 27 se puede observar el resumen de la última época y las métricas finales del modelo.

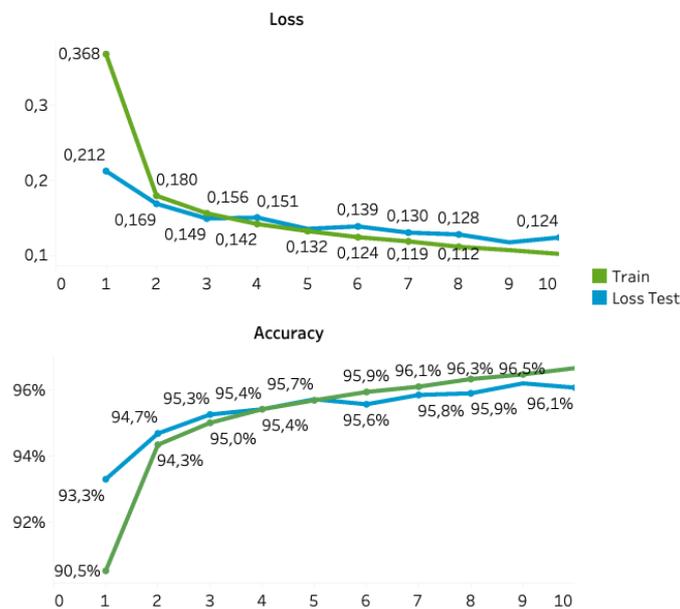
Figura 27: Resultados de la última época del entrenamiento de segmentación

```
Epoch 10/10
635/635 [=====] - 25392s 40s/step - loss: 0.1020 - accuracy: 0.9665 - val_loss: 0.1239 - val_accuracy: 0.9606
```

Epochs	Batch Size	Training Time (min)	Min (Loss)	Max(Accuracy)
10	16	2.247	0,1020	96,65%

A lo largo del entrenamiento, se observó una disminución en la pérdida tanto en el conjunto de entrenamiento como en el de validación, lo que indica que el modelo estaba aprendiendo efectivamente a segmentar las imágenes. Además, la precisión, que es una medida de cuántos píxeles están siendo clasificados correctamente, aumentó a lo largo del tiempo, llegando a valores cercanos al 97% y 96% en el conjunto de entrenamiento y validación, respectivamente, lo que evidencia un buen rendimiento del modelo. En la Figura 28 se puede ver la evolución por época del valor de pérdida y de la precisión del modelo entrenado.

Figura 28: Evolución de la pérdida y precisión del modelo de segmentación



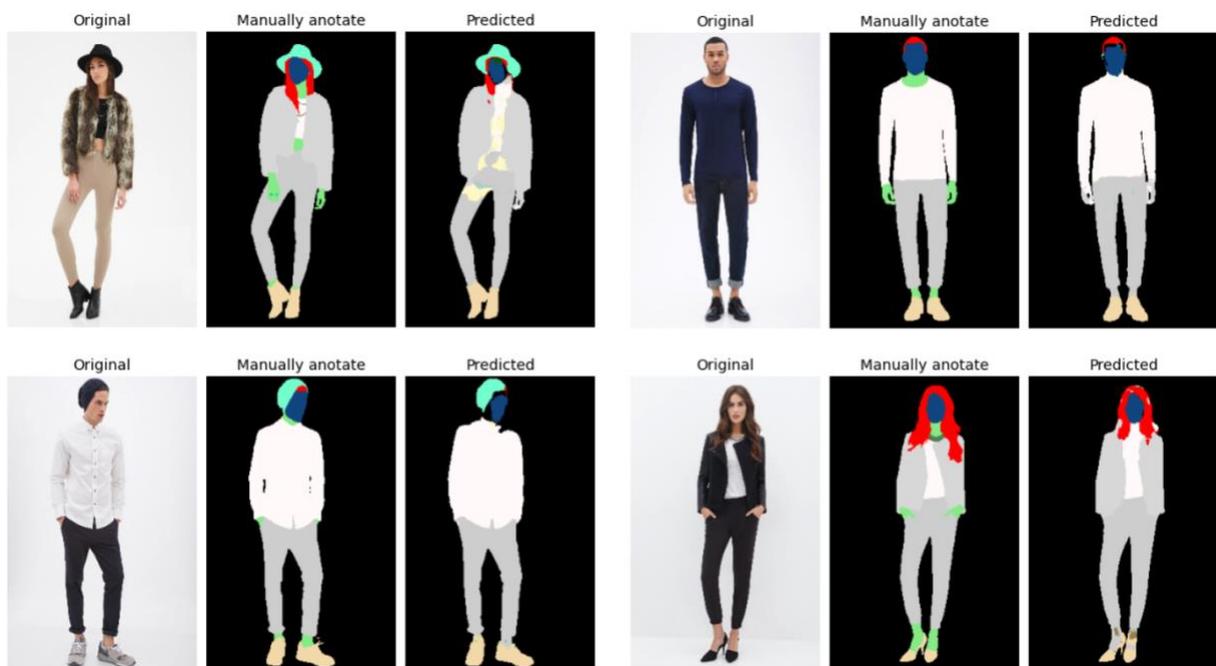
#### 6.1.2.4. Evaluación de resultados

Para la evaluación de resultados del modelo de segmentación, se llevó a cabo una evaluación cualitativa, es decir, se examinó visualmente cómo se desempeña el modelo en la segmentación de imágenes nuevas y se comparó con las segmentaciones manuales.

En el proceso de evaluación, se observó que el modelo era muy capaz de distinguir algunos segmentos específicos. Por ejemplo, demostró una alta precisión al separar partes inferiores como pantalones de las partes superiores, así como en la identificación de zapatos y sombreros. Esto sugiere la capacidad del modelo para capturar las características relevantes de estas prendas.

Aunque la mayoría de los resultados son favorables, se detectaron algunas dificultades en la capacidad de segmentación del modelo. En particular, se encontró que el modelo tenía problemas para distinguir entre la parte superior y las prendas de tipo chaqueta. Este tipo de confusión puede deberse a la similitud en la apariencia visual y la proximidad espacial de estas categorías en las imágenes. Del mismo modo, la distinción entre la piel de la persona y las prendas ha resultado poco precisa. Esto puede ser debido a las variaciones en el tono de la piel y la textura de la ropa que pueden confundir al modelo, lo que señala un área de mejora potencial. En la Figura 29 se pueden ver algunos ejemplos del análisis cualitativo.

Figura 29: Evaluación de los resultados del modelo de segmentación



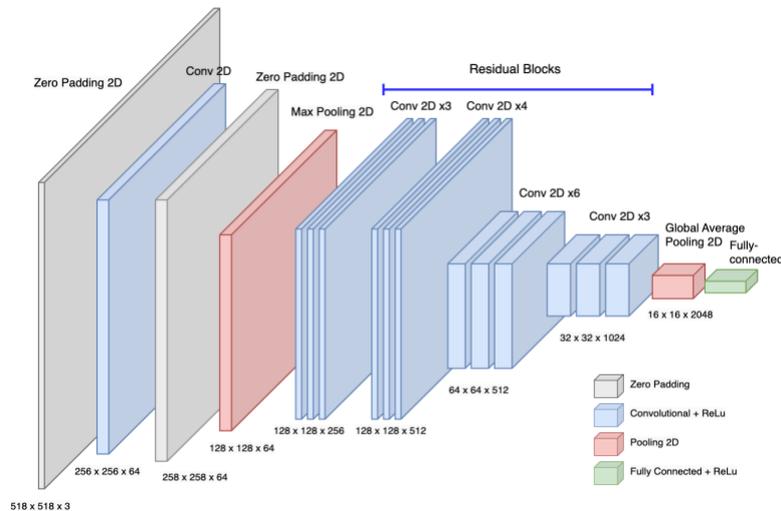
### 6.1.3. Modelo de estimación de poses con CNN

El segundo modelo de este proyecto tiene el objetivo de detallar la posición del cuerpo dada una imagen para ayudar al modelo generador a adaptar las prendas al cuerpo. El desafío de la estimación de poses es detectar y localizar una serie de puntos clave en una imagen que representan la pose del cuerpo humano. Se ha optado por un enfoque de aprendizaje por transferencia combinado con capas CNN para construir dicho modelo.

#### 6.1.3.1. Selección del modelo candidato para *Transfer Learning*

El modelo utilizado para la extracción de características es ResNet50 (Theckedath et al., 2020), una red neuronal convolucional pre-entrenada en el conjunto de datos ImageNet (Mukhlif et al., 2023). La elección de este modelo se basó en su rendimiento en la extracción de características de imágenes debido a su arquitectura residual profunda. Al igual que con el modelo de segmentación, todas las capas del modelo base se han ajustado para ser no entrenables, lo que significa que los pesos de estas capas no se actualizarán durante el entrenamiento. En la Figura 30 se puede observar arquitectura del modelo con el detalle por capas.

Figura 30: Arquitectura del modelo pre-entrenado RedNet50

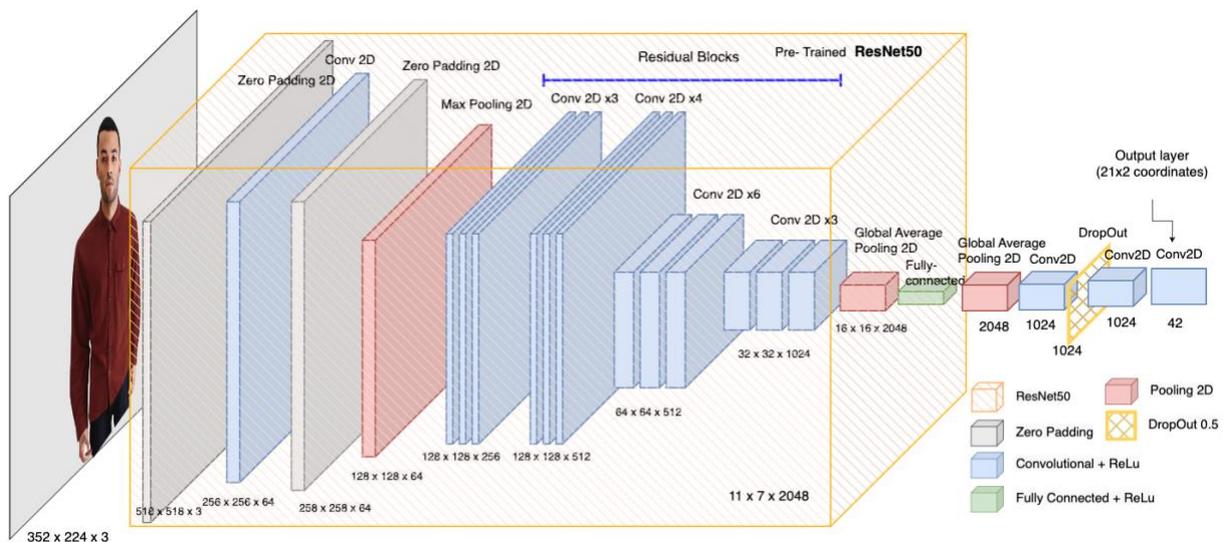


#### 6.1.3.1. Arquitectura del modelo desarrollado

La arquitectura del modelo de estimación de poses es la mezcla del modelo pre-entrenado ResNet50 y un clasificador personalizado. Como ResNet50 ya está pre-entrenada con imágenes de personas,

tiene una gran capacidad para extraer características visuales útiles para este caso de uso. Sobre el modelo escogido, se agrega una capa de GlobalAveragePooling2D para reducir las dimensiones de salida, lo que ayuda a prevenir el sobreajuste. Luego, se incluyen dos capas Densas con 1024 nodos y activación ReLU para permitir que el modelo aprenda relaciones complejas. También se introduce una capa de Dropout entre las capas densas para mejorar la generalización del modelo. La última capa es una capa Densa con 42 nodos, necesaria para predecir las 21x2 coordenadas. Se utiliza una activación lineal aquí ya que es un problema de regresión. Solo las últimas 10 son entrenables, lo que ayuda a prevenir el sobreajuste en conjuntos de datos pequeños. En la Figura 31 se puede ver el detalle de cómo ha sido diseñada la arquitectura del modelo.

Figura 31: Arquitectura del modelo de estimación de poses



### 6.1.3.2. Entrenamiento del modelo

El entrenamiento del modelo, al igual que el anterior, ha sido el resultado de una serie de pruebas con diferentes configuraciones con el fin de perfeccionar los parámetros de la red. Durante el entrenamiento de la selección final, se observa una tendencia consistente y favorable tanto en la pérdida como en la precisión. Se compila con el optimizador Adam y la pérdida de error cuadrático medio, ideal para modelos de regresión. Como en el modelo anterior, se utilizó un tamaño de lote de 16 imágenes y se llevó a cabo durante 10 épocas. En la Figura 32 se puede observar el resumen de la última época y las principales métricas del modelo.

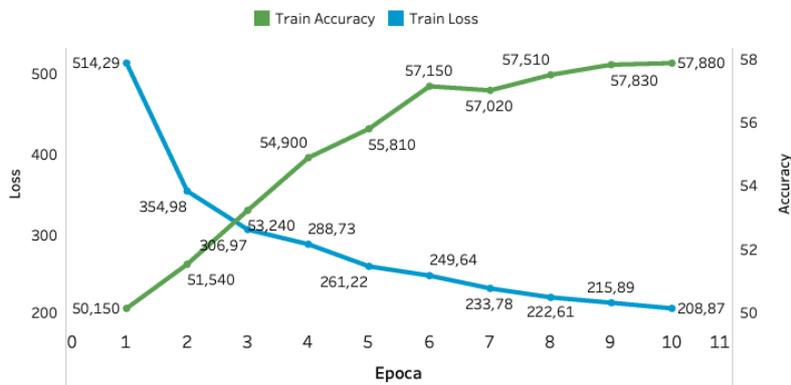
Figura 32: Resultados de la última época del entrenamiento de estimación de poses

Epoch 10/10  
 794/794 [=====] - 5741s 7s/step - loss: 208.8740 - accuracy: 0.5788

Epochs	Batch Size	Training Time (min)	Min (Loss)	Max(Accuracy)
<b>10</b>	<b>16</b>	<b>1.300</b>	<b>208,87</b>	<b>57,88%</b>

A medida que avanzaban las épocas, era más notable la disminución en la pérdida y una leve mejora en la precisión. Al finalizar la décima época, el modelo tenía una pérdida de 208,87 y una precisión del 57,88%. En la Figura 33 se puede ver la evolución en el rendimiento del modelo basado en la pérdida y la precisión.

Figura 33: Evolución de la pérdida y precisión del modelo de estimación de poses



### 6.1.3.3. Evaluación de resultados

La evaluación de los resultados para el modelo de estimación de poses consiste en una evaluación cualitativa, comparando visualmente las estimaciones de las poses producidas por el modelo con las anotaciones manuales. Para ello, se muestran imágenes no trabajadas por el modelo anteriormente, con la pose anotada manualmente en verde y la estimación del modelo en rojo.

En general, el modelo ha demostrado un buen rendimiento en la estimación de las proporciones del cuerpo, aunque se han observado algunas dificultades con los puntos que representan la posición de los hombros y los tobillos. Estas dificultades podrían deberse a la variabilidad en la posición y apariencia de estos puntos en diferentes poses e imágenes. En la Figura 34 se pueden ver algunos de los resultados obtenidos.

Figura 34: Evaluación de los resultados del modelo de estimación de poses



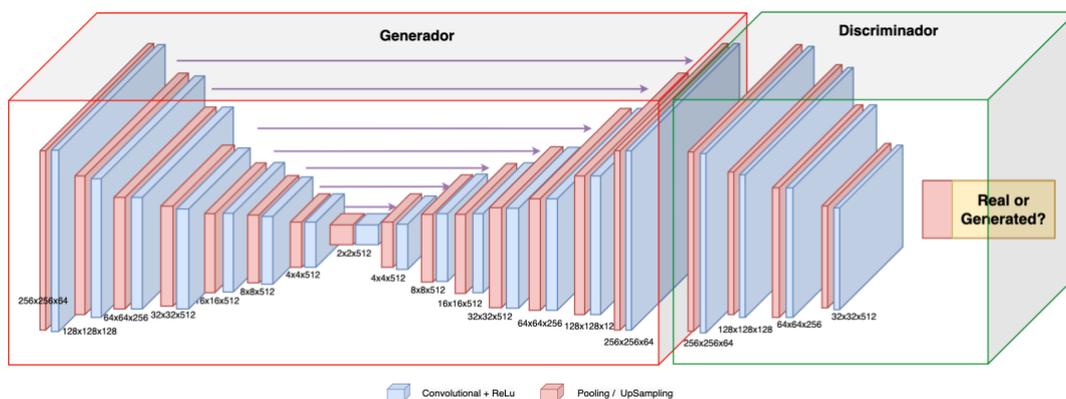
#### 6.1.4. Modelo de generación de imágenes con GAN

La última fase del desarrollo de IA de este proyecto es el modelo de generación de imágenes con GAN. En este apartado se describe el modelo desarrollado, el cual consta de varias etapas que abordan diferentes aspectos del problema: la adaptación de la prenda a la forma del cuerpo del usuario, la generación de las partes del cuerpo no cubiertas por la prenda y la superposición final de la prenda adaptada sobre el cuerpo del usuario. En las primeras capas de la red GAN, se utilizarán modelos preentrenados a través del *transfer learning*, lo cual permite acelerar el entrenamiento.

##### 6.1.4.1. Selección de los modelos candidatos para *Transfer Learning*

Para la selección de los modelos candidatos para la aplicación de *transfer learning* en la GAN condicional, se exploraron varios modelos preentrenados. Entre estos, destacaron dos modelos por su desempeño y aplicabilidad al problema en cuestión: VGG16 y pix2pix (Li et al., 2022). En la Figura 35 se puede ver una recreación de la arquitectura de pix2pix.

Figura 35: Arquitectura del modelo pix2pix



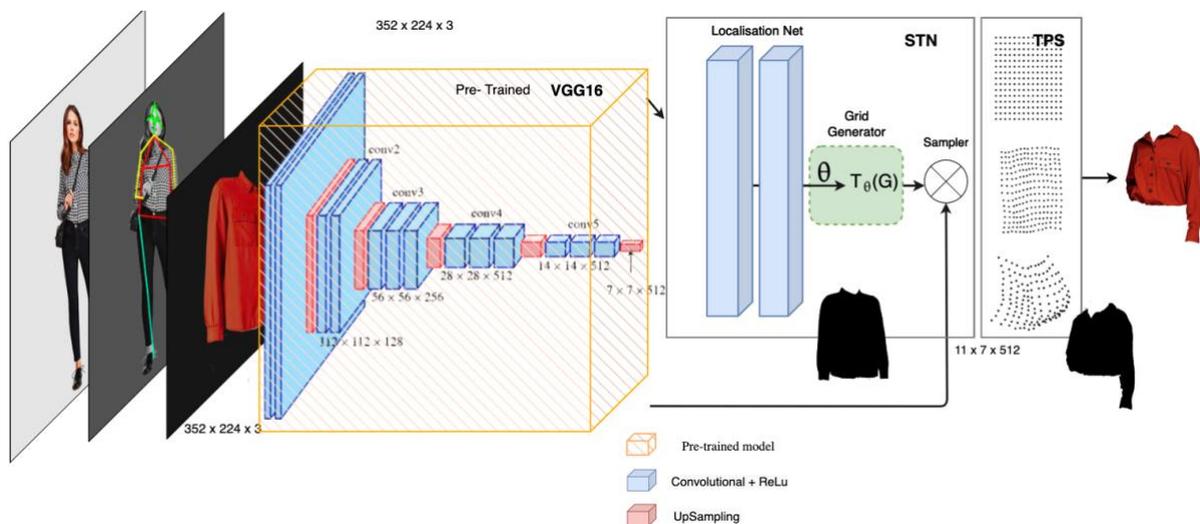
### 6.1.4.2. Arquitectura del modelo desarrollado

#### I. Generador

El generador es la primera parte de la GAN, y se encarga de generar nuevas imágenes sintéticas. Para este proyecto, el generador puede ser dividido en dos procesos: el primero para adaptar la prenda y el segundo para generar las partes no cubiertas del cuerpo.

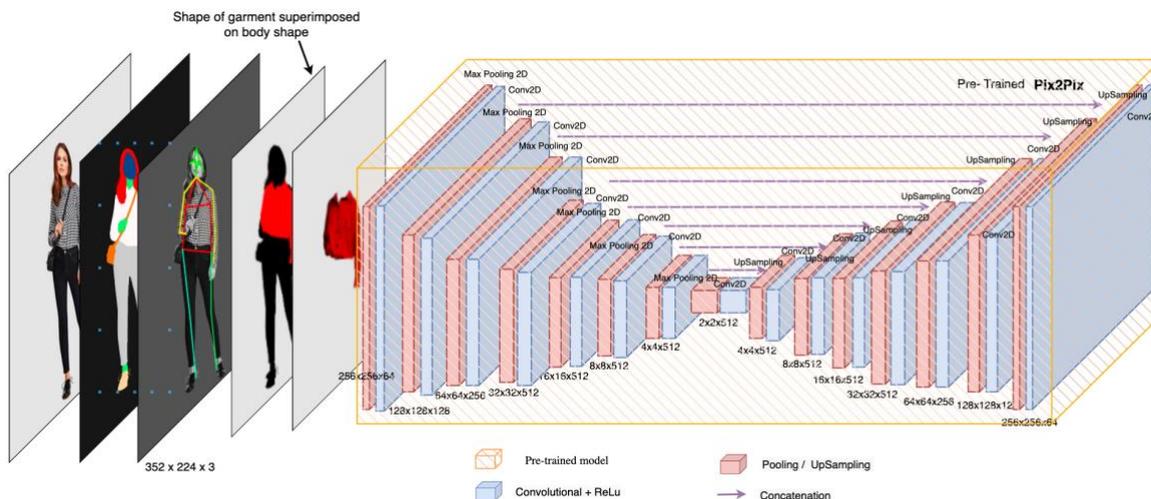
- **Adaptación de la prenda:** Se utiliza el modelo preentrenado VGG16 para la extracción de características, seguido de una Red de transformación espacial (Spatial Transformer Network, STN) con Spline de placa delgada (Thin Plate Spline, TPS) para adaptar la forma de la prenda a la pose del cuerpo. En la Figura 36 podemos ver el resultado de la primera parte de la arquitectura del generador.

Figura 36: Arquitectura de la fase de adaptación de la prenda (generador) del modelo de generación de imágenes



- **Generación de partes no cubiertas:** Se aplicarán las capas generadoras del modelo pix2pix, red de tipo cGAN, para generar las partes no cubiertas del cuerpo. En este caso el modelo está centrado en aquellos pixeles que no quedan cubiertos en la superposición de la prenda con la forma del cuerpo. En la Figura 37 se puede ver el detalle de la segunda parte del generador para el modelo de generación de imágenes.

Figura 37: Arquitectura de la fase de generación de partes no cubiertas (generador) del modelo de generación de imágenes

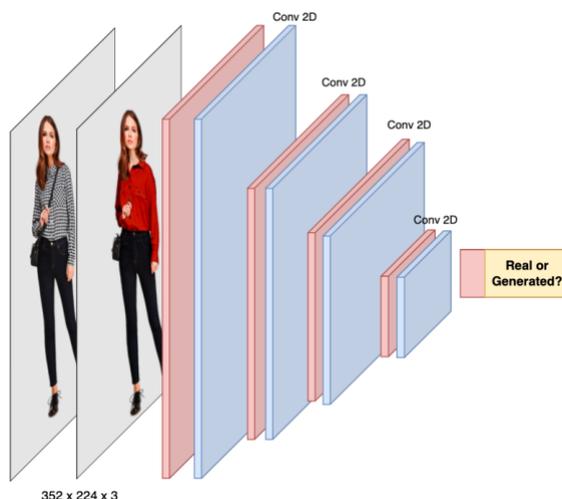


La combinación de los resultados de las dos fases se realiza utilizando técnicas de superposición y transformación. En este paso, la imagen generada de la prenda se superpone en la imagen del cuerpo, mientras que las partes del cuerpo no cubiertas se transforman en piel.

## II. Discriminador

El discriminador es una red convolucional que clasifica si una imagen es real (proveniente del conjunto de datos de entrenamiento) o falsa (generada por el generador). Este componente del modelo utiliza las imágenes de entrada y las imágenes de salida correspondientes como pares de entrenamiento. En la Figura 38 se puede ver el detalle de la arquitectura para el discriminador del modelo.

Figura 38: Arquitectura del discriminador en el modelo de generación de imágenes



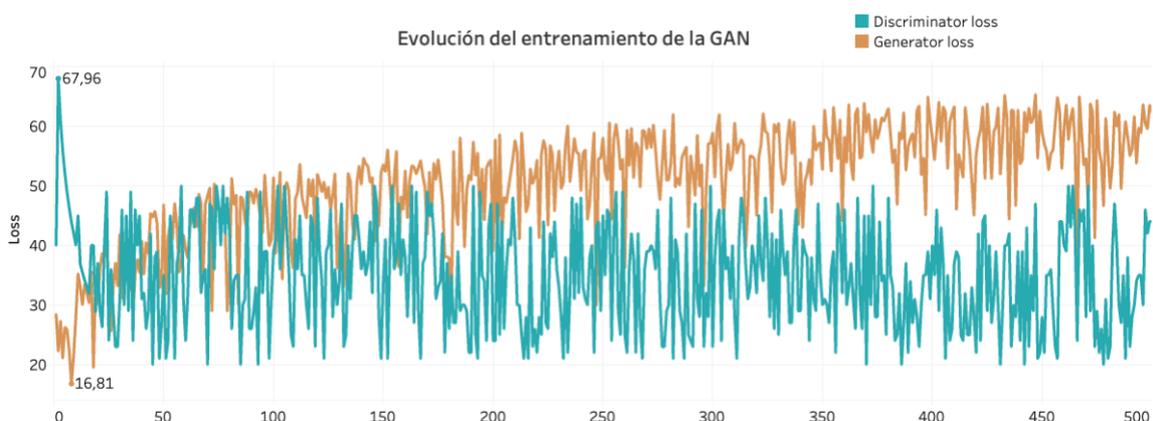
#### 6.1.4.1. Entrenamiento del modelo

El proceso de entrenamiento del modelo de generación de imágenes ha representado uno de los desafíos más significativos de este proyecto, tanto en términos de esfuerzo como de requerimientos computacionales. Debido a la complejidad del modelo y a la gran cantidad de datos manejados, ha sido necesario incrementar considerablemente los recursos de cómputo disponibles. Para ello, se ha optado por utilizar Google Colab Pro [\(Bisong, 2019\)](#), lo que ha permitido ampliar la memoria RAM del sistema, la RAM de la GPU y la memoria de almacenamiento general.

En relación a los mejores resultados de entrenamiento, tanto el generador como el discriminador han sido compilados utilizando el optimizador Adam y la función de pérdida entropía cruzada. Se han empleado lotes de 32 imágenes y se ha configurado el entrenamiento para llevar a cabo un total de 500 épocas.

Durante las primeras épocas del entrenamiento, se ha observado una disminución notable en la pérdida tanto del generador como del discriminador, lo que indica que ambos componentes estaban aprendiendo de manera efectiva a realizar sus respectivas tareas. Sin embargo, conforme avanzaba el entrenamiento, la pérdida en el discriminador se estabiliza, mientras que la del generador aumenta ligeramente. Esta situación es común en el entrenamiento de GAN y refleja la naturaleza adversarial de su diseño, donde el generador y el discriminador compiten entre sí, llevando a que ambos componentes mejoren conjuntamente hasta alcanzar un punto de equilibrio. En la Figura 39 se puede ver la evolución de la pérdida tanto para el generador como para el discriminador.

Figura 39: Evolución de la pérdida en el modelo de generación de imágenes



#### 6.1.4.2. Evaluación de resultados

En cuanto a la evaluación de resultados del modelo de generación de imágenes, se ha llevado a cabo un análisis cualitativo, realizando pruebas con diversas prendas de vestir en diferentes individuos. En términos generales, los resultados obtenidos han sido favorables. El modelo ha demostrado ser capaz de reproducir la prenda deseada sobre la persona, respetando adecuadamente las formas y los colores. En la Figura 40 se observa un ejemplo de cómo el generador aplica una camisa de color y textura distinta a la imagen original.

*Figura 40: Ejemplo de la aplicación del modelo de generación de imágenes*



No obstante, durante el proceso de evaluación, se han identificado ciertos aspectos susceptibles de mejora. Entre ellos, se han detectado ciertas zonas borrosas en las imágenes generadas, especialmente en las áreas donde el modelo debía generar piel. Esto podría deberse a la complejidad inherente de generar texturas de piel realistas. Además, el modelo parece tener dificultades con la generación de manos, resultando en ocasiones en la creación de dedos de aspecto inusual. En la Figura 41 se muestran dos ejemplos de la evaluación en los que se resaltan las zonas con incongruencias.

*Figura 41: Resultados de la evaluación del modelo de generación de imágenes*



Otro aspecto para mejorar es que, en algunas instancias, el modelo ha recortado de manera incorrecta el cabello de la persona. Esto podría ser debido a la dificultad de diferenciar entre el cabello y el fondo en algunas imágenes.

A pesar de estas áreas de mejora, los resultados obtenidos han sido en su mayoría positivos, demostrando la capacidad del modelo para generar imágenes de prendas de vestir sobre personas de una manera convincente.

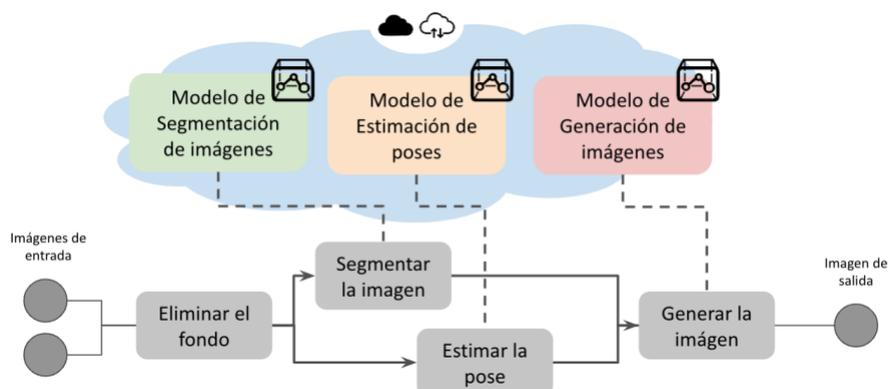
### 6.1.5. Integración de los modelos

Para poder integrar todo el proceso, los modelos de segmentación de imágenes, estimación de poses y generación de imágenes han sido almacenados como archivos pickle en Google Drive. Esta metodología permite la flexibilidad de poder descargar o conectarse desde la página web a la nube cuando sea necesario, facilitando la implementación.

Además, para adaptar el sistema a las necesidades y condiciones específicas de los usuarios finales, se ha incorporado una fase previa de eliminación de fondo para imágenes nuevas. Esta es una transformación relativamente sencilla que, no obstante, juega un papel crucial en la mejora de los resultados. La eliminación de fondo contribuye a que las imágenes 'no profesionales', como las que probablemente tengan los usuarios, se asemejen más a las imágenes con las que los modelos han sido entrenados.

En la Figura 42 se muestra el proceso que seguiría una nueva imagen subida por el usuario para ser procesada y finalmente reconstruida con la nueva prenda de vestir.

Figura 42: Diagrama de la integración de la herramienta



En resumen, el proceso de integración de los modelos ha requerido una cuidadosa consideración tanto de los aspectos técnicos, como el almacenamiento y recuperación de los modelos, como de las necesidades específicas de los usuarios finales, con el objetivo de proporcionar una herramienta efectiva y usable en diversas condiciones.

### **6.1.6. Limitaciones del Sistema de Inteligencia Artificial**

El sistema presenta varias limitaciones que deben ser tomadas en consideración. En algunos casos, se han observado resultados que no se ajustan completamente al cuerpo de la persona original, especialmente cuando la prenda a probar es compleja en términos de textura y diseño. Esta limitación puede ser atribuida, en parte, al proceso secuencial del sistema, donde un error en la estimación de la posición o la segmentación en las etapas iniciales puede tener un efecto en cascada en el modelo de generación de imágenes.

El modelo de generación de imágenes también presenta algunas incongruencias, como zonas borrosas y deformaciones en las manos y el cabello. Estas imperfecciones podrían ser el resultado de la dificultad inherente en la generación de estas características finas y detalladas.

Una de las limitaciones más relevantes del sistema, e independiente del desarrollo técnico, se encuentra en el propio conjunto de datos de entrenamiento. Aunque la intención inicial era proporcionar una solución que pudiera ser representativa de cualquier usuario, la diversidad de tipos de cuerpo en la realidad supera las representaciones disponibles en los conjuntos de datos actuales. Como resultado, la precisión de la herramienta podría estar condicionada por la similitud del aspecto del usuario con los ejemplos vistos por el modelo durante su entrenamiento.

A pesar de estas limitaciones, el sistema logra cumplir con su objetivo general y ofrece resultados satisfactorios en muchos aspectos. Sin embargo, es importante destacar que estos desafíos representan oportunidades para futuros trabajos de mejora y optimización del sistema.

## **6.2. Desarrollo de la página web**

Este apartado explica el proceso y las tecnologías utilizadas para desarrollar la página web que permite la interacción con el sistema. Esta interfaz desempeña un papel crucial, ya que proporciona una

plataforma fácil de usar y accesible para que los usuarios interactúen con los modelos de aprendizaje profundo.

### **6.2.1. Introducción y Tecnologías Usadas**

El desarrollo de la página web se realizó en el lenguaje de programación Python, como ya se hizo para el entrenamiento de los modelos, por su fácil integración entre las herramientas de aprendizaje profundo y las de desarrollo web.

Para la creación de la página web, se recurrió a Flask ([Flask, 2023](#)), un microframework de Python que provee las funcionalidades necesarias para el desarrollo de aplicaciones web. Su simplicidad y flexibilidad, ha permitido crear una interfaz web personalizada, evitando la complejidad de otros frameworks más robustos. El diseño inicial de la interfaz se ha llevado a cabo utilizando la aplicación Uizard ([Uizard, 2023](#)). Esta aplicación de prototipado y diseño de interfaz permite visualizar y planificar la estructura de la página web antes de iniciar el desarrollo.

La aplicación está configurada para poder acceder a los modelos, almacenados en Google Drive, evitando la necesidad de almacenar los modelos en el servidor de la aplicación, ayudando a una gestión eficiente de los recursos.

En las secciones siguientes, se detalla la estructura y funcionalidad de la página web, así como las medidas tomadas para garantizar la seguridad y protección de los datos del usuario.

### **6.2.2. Estructura de la Página Web**

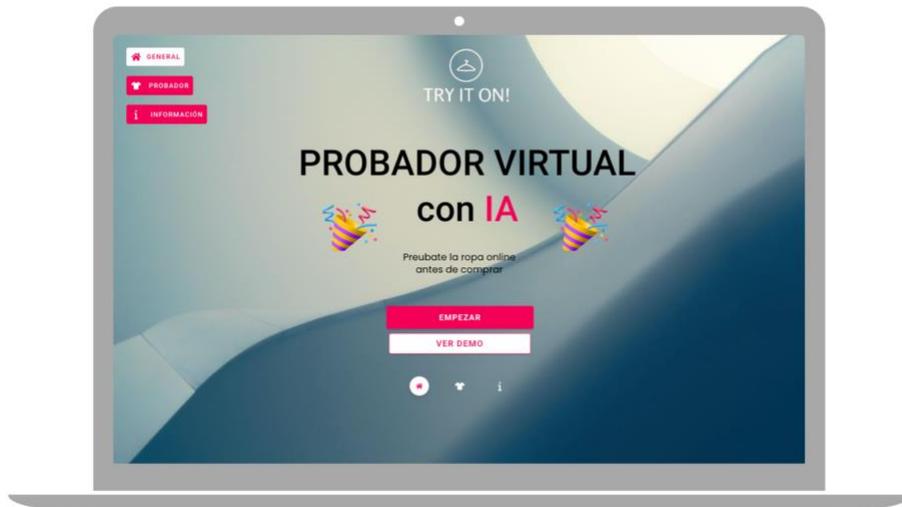
Para lograr una experiencia de usuario agradable, la página web se divide en tres secciones principales: la página de inicio, la página del probador, y la página de información.

#### **6.2.2.1. Página de inicio**

La página de inicio actúa como la puerta de entrada a la aplicación web. En la misma, se presenta un diseño limpio y minimalista, evitando distracciones innecesarias y concentrándose en la funcionalidad central del proyecto. En la parte superior de la página se puede encontrar el logo del proyecto y un título que da la bienvenida a los usuarios y ofrece una breve descripción de lo que pueden esperar de la aplicación. Además, esta página cuenta con tres enlaces destacados que dirigen a los usuarios hacia

la página del probador, la página de información y una breve demo de cómo utilizar todas las funcionalidades. En la Figura 43 se muestra una captura de pantalla de cómo se visualiza la página de inicio.

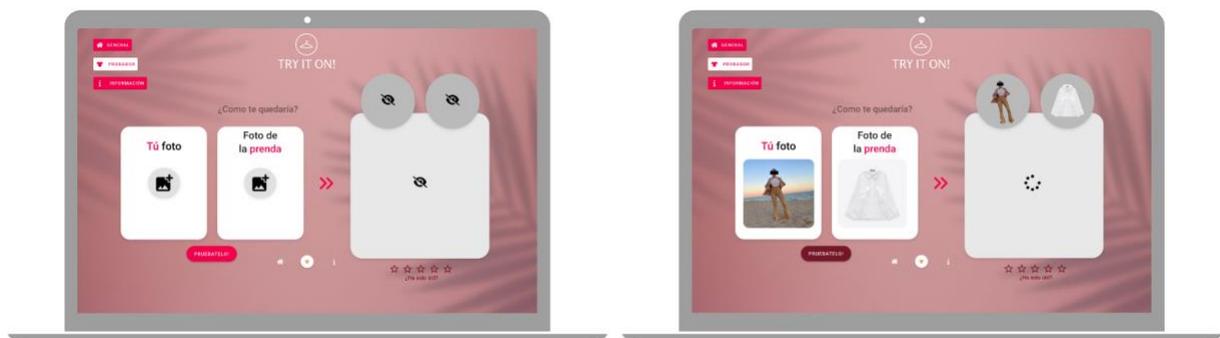
Figura 43: Página de inicio de la web



#### 6.2.2.2. Página del probador

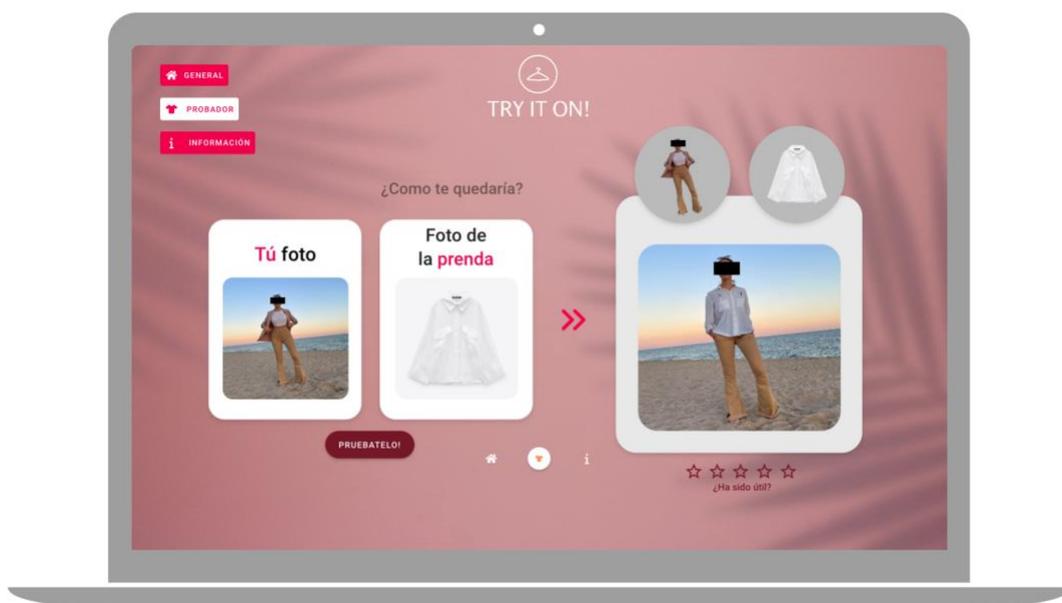
La página del probador es donde reside la funcionalidad principal de la aplicación. En ella, se solicita al usuario que suba dos imágenes: una de cuerpo entero y otra de la prenda que desea "probar". Se ofrece una interfaz de usuario sencilla que guía al usuario a través del proceso de subida de las imágenes. En la Figura 44 se puede observar la interfaz de la página del probador antes y después de añadir las imágenes para probarse.

Figura 44: Página del probador mientras se añade una imagen



Una vez cargadas las imágenes, el sistema inicia el proceso de análisis utilizando los modelos de inteligencia artificial previamente entrenados. En primer lugar, se ejecutan los modelos de estimación de poses y segmentación de imágenes con redes neuronales convolucionales (CNN). Los resultados de estos modelos se utilizan como entrada para el modelo de generación de imágenes con redes generativas antagónicas (GAN), que finalmente produce la imagen de la persona "vistiendo" la prenda seleccionada. En la figura se muestra una captura de pantalla del resultado de la generación de imagen en la página del probador.

Figura 45: Página del probador mostrando la nueva imagen generada



Después de que se genera la imagen, esta se muestra al usuario en la misma página del probador. Debajo de la imagen generada, se proporciona una funcionalidad que permite al usuario puntuar la precisión de la predicción en una escala de una a cinco estrellas. Este sistema de calificación sirve tanto para recoger *feedback* de los usuarios sobre la calidad de las predicciones como para que los usuarios se sientan más involucrados en el proceso.

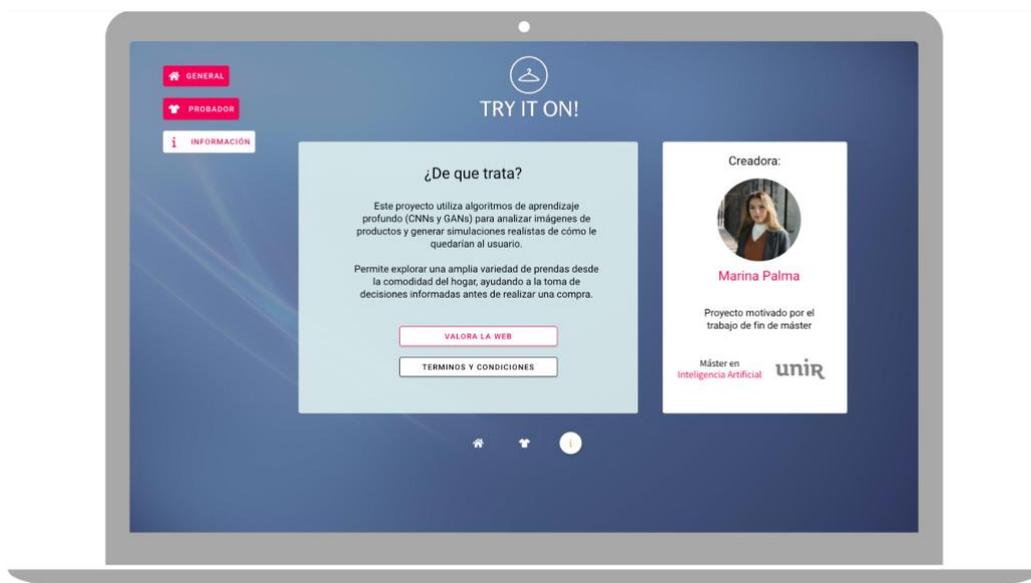
### 6.2.2.3. Página de información

La página de información proporciona contexto y transparencia acerca de la aplicación. Aquí se ofrece un resumen que explica la funcionalidad de la página web y los detalles sobre su desarrollo. Se describe la motivación detrás del proyecto, los problemas que se propone resolver y los beneficios que ofrece

a los usuarios. Además, se incluye una descripción de la creadora, añadiendo un toque personal a la página web.

En esta página también se abordan importantes cuestiones de privacidad y seguridad. Un botón de tipo *pop up* con los Términos y Condiciones aclara que las imágenes subidas por los usuarios no se almacenan de forma permanente y que se eliminan de la memoria al cerrar la página. Además, se enfatiza que estos datos no se utilizan para reentrenar los modelos en futuras versiones. En la Figura 46 se muestra una captura de la pestaña de información de la página web.

Figura 46: Página de información de la web



### 6.2.3. Términos y condiciones

Una adición importante a la página de información es una sección dedicada a los "Términos y Condiciones". Esta sección es crucial para establecer las reglas y pautas que los usuarios deben seguir al usar la aplicación, y también sirve para proteger legalmente al creador y al proyecto en general.

Los términos y condiciones explican en detalle cómo los usuarios pueden y no deben usar la aplicación. Este conjunto de reglas cubre temas como el uso aceptable de los servicios, la responsabilidad de los usuarios en relación con su conducta y sus contenidos, y las limitaciones del uso de los servicios proporcionados por la aplicación. Entre las reglas establecidas, se encuentra que los usuarios solo

pueden subir imágenes de personas físicas si ellos mismos son la persona que aparece en la imagen. Esta norma es crucial para respetar la privacidad y los derechos de las personas.

Además, esta sección también aborda temas relacionados con los derechos de propiedad intelectual. Se especifica que, aunque los usuarios tienen la libertad de subir imágenes para probar los servicios de la aplicación, siguen siendo los únicos propietarios de los derechos de propiedad intelectual de esas imágenes.

Otro aspecto importante que se cubre en los términos y condiciones es la política de privacidad. Esta política detalla cómo se recopilan, utilizan, protegen y, en algunos casos, comparten los datos personales de los usuarios. Se subraya que la página web se adhiere a las leyes y regulaciones de protección de datos vigente, como el Reglamento General de Protección de Datos (GDPR) de la Unión Europea ([European Parliament & Council of the European Union, 2018](#)), garantizando que los usuarios tengan control total sobre su información personal.

Por último, en los términos y condiciones se aclara que cualquier posible actualización o cambio en los servicios ofrecidos por la página web se comunicará a los usuarios a través de la web.

#### **6.2.4. Almacenamiento, seguridad y protección de los datos**

La seguridad de los datos del usuario y la protección de su privacidad son consideraciones clave en la construcción y operación de la página web. Se han tomado medidas para garantizar que se manejen de manera segura y se respeten plenamente los derechos del usuario.

- **Almacenamiento de datos:** Las imágenes que los usuarios cargan en la página web para usar el servicio no se almacenan de manera permanente. En cambio, las imágenes se mantienen en caché temporalmente durante la sesión activa del usuario para el procesamiento y generación de las nuevas imágenes. Una vez que el usuario cierra la página o termina la sesión, estas imágenes se eliminan de la memoria. Esta medida garantiza que no se retenga ninguna información personal sin el consentimiento del usuario.
- **Seguridad de datos:** La página web utiliza tecnologías de encriptación y seguridad para proteger cualquier dato transmitido durante el uso del servicio. Esto garantiza que la

información del usuario se mantiene segura y protegida contra cualquier intento de acceso no autorizado o uso indebido.

- **Protección de datos:** En términos de protección de datos, la página web se adhiere a el Reglamento General de Protección de Datos (GDPR) ([European Parliament & Council of the European Union, 2018](#)). Los usuarios tienen el derecho de saber qué datos se recopilan, cómo se utilizan y cómo pueden controlar o eliminar estos datos. En los términos y condiciones, se proporciona una descripción detallada de estas políticas de privacidad. Además, es importante señalar que los datos del usuario no se utilizan para reentrenar el modelo en futuras versiones.

En resumen, la página web se ha desarrollado con un enfoque en la seguridad y protección de datos, garantizando que los usuarios puedan utilizar el servicio con plena confianza en el manejo seguro y respetuoso de su información.

## 7. Evaluación de la herramienta desarrollada

La finalidad de esta sección es proporcionar una evaluación detallada de la herramienta desarrollada en términos de su usabilidad y su potencial impacto y utilidad. Se consideran aspectos tanto cuantitativos como cualitativos, con un enfoque en la experiencia real del usuario y en la eficacia de la herramienta para cumplir su propósito previsto. La metodología ha consistido en permitir a un grupo de usuarios interactuar con la herramienta para después calificarla a través de un formulario de *Google Docs*. En la Figura 47 se pueden observar dos ejemplos de los resultados obtenidos por los usuarios.

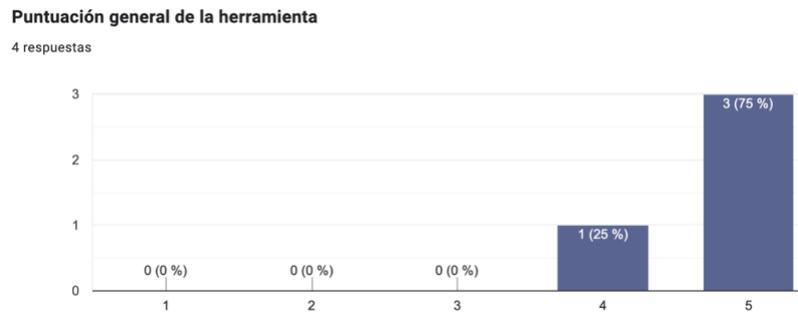
*Figura 47: Ejemplos de las pruebas de usuario realizadas*



### 7.1. Evaluación de usabilidad

La usabilidad de la herramienta se mide en términos de su facilidad de uso y de cuán intuitiva es para los usuarios. Durante la evaluación, se ha observado que los usuarios no tuvieron problemas para interactuar con la herramienta de generación de imágenes. Se proporcionó una interfaz de usuario sencilla y bien diseñada, permitiendo a los usuarios cargar fácilmente las imágenes de entrada necesarias, y recibir las imágenes generadas sin complicaciones. La valoración general de la herramienta ha obtenido una puntuación de 4,75 sobre 5. La Figura 48 muestra la distribución de los resultados.

Figura 48: Resultados a la puntuación general de la herramienta



A pesar de la alta puntuación media, un análisis más detallado de las respuestas de los usuarios y de sus interacciones con la herramienta reveló varias áreas potenciales de mejora. En particular, se observó que, aunque la mayoría de las predicciones generadas por la herramienta eran realistas, hubo ciertos casos en los que la correspondencia entre la prenda o el cuerpo del usuario y la imagen generada no era la esperada. Un problema recurrente en los resultados fue la representación borrosa de las manos, o la aparición de extremidades adicionales.

En resumen, la evaluación de la usabilidad de la herramienta fue positiva, con una respuesta general de que la herramienta era fácil de usar. No obstante, aunque la herramienta es funcional y puede generar resultados realistas en la mayoría de los casos, hay aspectos de la generación de imágenes que convendría investigar para mejorar la calidad y la precisión de las predicciones.

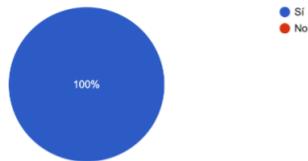
## 7.2. Evaluación de impacto y utilidad

La herramienta fue valorada positivamente en términos de su potencial impacto y utilidad. Los usuarios reconocieron que la herramienta podría transformar su experiencia de compra *online*, permitiéndoles visualizar cómo las prendas se ajustarían a su cuerpo antes de hacer la compra. Esto tiene el potencial de reducir las devoluciones de productos y aumentar la satisfacción del cliente, lo que es beneficioso tanto para los consumidores como para los vendedores.

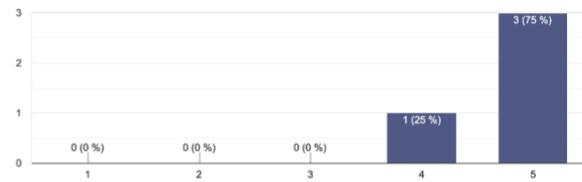
Los usuarios también indicaron que estarían dispuestos a utilizar la herramienta de forma regular durante sus compras en línea. Lo que indica que ven un valor real en el servicio proporcionado y que podría tener un impacto significativo en su comportamiento de compra. El resultado de algunas de las preguntas del formulario se puede ver en la Figura 49.

Figura 49: Respuestas a un subconjunto de las cuestiones planteadas a los usuarios

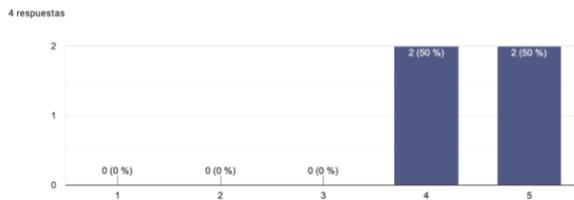
¿Crees que esta herramienta podría cambiar tu experiencia de compra online?  
4 respuestas



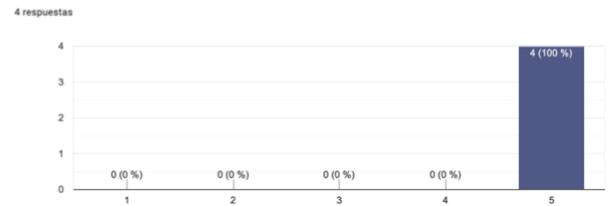
¿Cómo de útil encuentras esta herramienta?  
4 respuestas



¿Las imágenes generadas se asemejan a cómo esperabas que se viera la prenda en el cuerpo?  
4 respuestas



¿Cómo de realistas eran las imágenes? Independientemente de si han representado fielmente el cuerpo y prenda original. [Copiar](#)



En cuanto a las recomendaciones de mejora, la mayoría de las sugerencias se centraron en la calidad y el realismo de las imágenes generadas. Aunque en general se encontraron satisfactorias, se señaló que había espacio para mejorar en términos de la precisión con la que la herramienta recreaba la forma del cuerpo y la ropa. Trabajar en estos aspectos podría aumentar aún más el impacto y la utilidad de la herramienta.

En general, la evaluación de impacto y utilidad fue muy positiva. Los usuarios valoraron la herramienta como una adición útil a su experiencia de compra en línea, y expresaron su disposición a usarla regularmente. La mejora del realismo y la precisión de las imágenes generadas podría incrementar aún más el impacto y la utilidad de la herramienta.

## 8. Conclusiones y trabajo futuro

Este trabajo ha descrito el desarrollo y evaluación de una herramienta de inteligencia artificial para la generación de imágenes de prendas de vestir sobre personas. Las conclusiones se extraen de cada uno de los componentes y se presentan a continuación.

### 8.1. Conclusiones

El avance de este proyecto ha permitido desarrollar una herramienta robusta para la generación de imágenes de personas vistiendo diferentes prendas de ropa. A través del empleo de tecnologías como las redes neuronales convolucionales, redes neuronales generativas adversarias y el aprendizaje por transferencia, se han podido abordar desafíos complejos como la segmentación, la estimación de poses y la generación de imágenes. De acuerdo con los resultados obtenidos, se evidencia el progreso significativo logrado en cada una de estas áreas.

A pesar de las limitaciones observadas, como la dificultad en la segmentación de prendas similares y la distinción entre la piel de la persona y las prendas, el modelo de segmentación ha demostrado ser efectivo en la identificación y separación de diferentes categorías de prendas. La estimación de poses, a pesar de presentar dificultades en la identificación de puntos como hombros y tobillos, ha proporcionado una representación acertada de las proporciones del cuerpo. El modelo de generación de imágenes ha mostrado la capacidad de reproducir prendas sobre individuos con bastante precisión, aunque con áreas identificadas para mejorar, como la distorsión de las manos y la distinción entre cabello y fondo.

Con respecto al **primer objetivo específico**, estudiar las técnicas de IA del estado del arte, se ha logrado una comprensión profunda de tecnologías como las redes neuronales convolucionales y el aprendizaje por transferencia, particularmente en la generación, segmentación de imágenes y estimación de poses. Además, se identificaron nuevos desafíos, como la segmentación de prendas similares y la diferenciación entre piel y ropa.

En cuanto al **segundo objetivo específico**, seleccionar las herramientas de IA, se escogieron y utilizaron las bibliotecas y frameworks más destacados en inteligencia artificial para el desarrollo del sistema de

probador virtual. Cabe destacar que se incrementaron los recursos planteados inicialmente para poder soportar el coste de desarrollo del proyecto.

El **tercer objetivo específico**, crear un conjunto de datos, se cumplió combinando un dataset multi modal *open source* previamente etiquetado, y adaptando dicha base de datos para el correcto desarrollo de los modelos. El gran volumen de datos incremento el tiempo estimado de entrenamiento, pero a su vez contribuyo a la precisión de los resultados obtenidos.

En referencia al **cuarto objetivo específico**, desarrollar los algoritmos de segmentación, estimación y generación de imágenes, se lograron construir modelos efectivos. Se encontraron dificultades en la identificación de ciertos puntos, tales como poca precisión en la distinción entre la piel y las prendas de vestir y la generación de algunas imágenes con zonas borrosas. A pesar de las dificultades, se lograron resultados significativos, alcanzando una precisión del 96.6% en la segmentación de imágenes, un 57.9% en la estimación de poses, y un equilibrio sólido en la generación y discriminación de imágenes en el modelo GAN.

Respecto al **quinto objetivo**, adaptarse a la protección de datos, se establecieron medidas adecuadas, como anonimizar los identificadores de cada imagen trabajada, para garantizar la privacidad de los usuarios, cumpliendo con las regulaciones vigentes como el GDPR.

En relación con el **sexto objetivo**, desarrollar una página web para interactuar con la solución, se diseñó una interfaz intuitiva y fácil de usar. Para agilizar el desarrollo se implementó el diseño en Uizard, y posteriormente se trasladó el código a la librería Flask de Python para mantener la unicidad del lenguaje de programación entre los modelos y la web.

Finalmente, el **séptimo objetivo específico**, evaluar la solución propuesta, se cumplió realizando pruebas de calidad y usabilidad. Estas pruebas revelaron un alto nivel de satisfacción entre los usuarios, aunque también se identificaron áreas de mejora en cuanto a la calidad de las imágenes generadas. La herramienta ha sido bien recibida, logrando una alta calificación de 4.75 sobre 5 en la evaluación con usuarios.

Para concluir, a pesar de las limitaciones y desafíos, el desarrollo de la herramienta ha demostrado ser un logro significativo en la aplicación de la inteligencia artificial en la industria de la moda. La

investigación y los desarrollos realizados apuntan a un futuro prometedor en el que la inteligencia artificial puede mejorar aún más la experiencia de compra en línea y superar las limitaciones actuales.

## **8.2. Trabajo futuro**

Los resultados obtenidos en el presente trabajo sentaron una base para la continuación de las investigaciones en el campo de los probadores virtuales con tecnología de Inteligencia Artificial. Aunque la herramienta desarrollada logró satisfacer la mayoría de los requisitos identificados, también se presentaron diversas áreas de mejora y oportunidades de investigación para futuros trabajos.

Una línea de trabajo futuro relevante se refiere a mejorar la calidad y el realismo de las imágenes generadas. Se podrían explorar nuevos modelos GAN, que se vayan descubriendo en el futuro, para mejorar la precisión y calidad de las imágenes. Específicamente, la representación de texturas de piel más realistas, es un área en la que se podrían centrar futuras investigaciones.

Otra área para explorar es la inclusión de tallas de prendas en la herramienta, lo que podría mejorar la precisión en la decisión del usuario. La integración de un sistema de recomendación de tallas basado en las medidas corporales de los usuarios podría proporcionar un valor añadido a la herramienta y aumentar su utilidad y satisfacción para los usuarios.

En términos de diversidad de tipos de cuerpo, es vital extender el conjunto de datos de entrenamiento para incluir una mayor variedad de tipos de cuerpo, géneros y grupos étnicos. Esto podría mejorar la capacidad del modelo para generar imágenes precisas para una gama más amplia de usuarios aumentando su accesibilidad y aplicabilidad.

En conjunto, estos posibles trabajos futuros podrían permitir mejorar aún más la eficacia de la herramienta de probadores virtuales y potenciar su utilidad tanto para los comerciantes como para los usuarios finales.

## Referencias bibliográficas

Angular. (2023). *Angular* <https://angular.io/>

Arjovsky, M., Chintala, S., & Bottou, L. (2017, July). Wasserstein generative adversarial networks. In International conference on machine learning (pp. 214-223). PMLR.

ASOS, See My Fit (2020). *Probador online con realidad aumentada*. <https://www.asosplc.com/news/asos-trials-see-my-fit-augmented-reality-tool/>

AWS. (2023). *Amazon Web Service*. <https://aws.amazon.com>

AWS Blog. (2021). *Blog de Amazon Web Services (AWS)* <https://aws.amazon.com/es/blogs>

Azure. (2023). *Microsoft Azure*. <https://azure.microsoft.com>

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12), 2481-2495.

Bai, Y., Chen, Y., Yu, W., Wang, L., & Zhang, W. (2020). Products-10k: A large-scale product recognition dataset. *arXiv preprint arXiv:2008.10545*.

Bischke, B., Helber, P., Folz, J., Borth, D., & Dengel, A. (2019, September). Multi-task learning for segmentation of building footprints with deep neural networks. In 2019 IEEE International Conference on Image Processing (ICIP) (pp. 1480-1484). IEEE.

Bisong, E. (2019). Google Colaboratory. En *Apress eBooks* (pp. 59-64). [https://doi.org/10.1007/978-1-4842-4470-8\\_7](https://doi.org/10.1007/978-1-4842-4470-8_7)

Buttler, G. *Artificial intelligence for fashion*, Leanne Luce, Apress 2019, ISBN 978-1-4842-3930-8 how AI is revolutionizing the fashion industry. *Genet Program Evolvable Mach* 23, 159–160 (2022). <https://doi.org/10.1007/s10710-021-09422-8>

Caboni, F., & Hagberg, J. (2019). Augmented reality in retailing: a review of features, applications and value. *International Journal of Retail & Distribution Management*. <https://doi.org/10.1108/IJRDM-12-2018-0263>

Cao, Z., Simon, T., Wei, S. E., & Sheikh, Y. (2017). Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7291-7299).

- Chen, T., Li, M., Li, Y., Lin, M., Wang, N., Wang, M., ... & Zhang, Z. (2015). Mxnet: A flexible and efficient machine learning library for heterogeneous distributed systems. arXiv preprint arXiv:1512.01274.
- Chen, Y., Wang, Z., Peng, Y., Zhang, Z., Yu, G., & Sun, J. (2018). Cascaded pyramid network for multi-person pose estimation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7103-7112).
- Cosco, A. (2019, May 19). How Luxury Retailers are using Augmented Reality. Electric Runway. <https://electricrunway.com/how-luxury-retailers-are-using-augmented-reality/>
- Deloitte (2020). Omnia AI, The Future of Personalization in Retail. <https://www2.deloitte.com/content/dam/Deloitte/ca/Documents/deloitte-analytics/ca-en-omnia-ai-marketing-pov-fin-jun24-aoda.pdf>
- El-Kaddoury, M., Mahmoudi, A., & Himmi, M. M. (2019). Deep generative models for image generation: A practical comparison between variational autoencoders and generative adversarial networks. In Mobile, Secure, and Programmable Networking: 5th International Conference, MSPN 2019, Mohammedia, Morocco, April 23–24, 2019, Revised Selected Papers 5 (pp. 1-8). Springer International Publishing.
- European Parliament and Council of the European Union. (2018). Regulation (EU) 2018/1725 of the European Parliament and of the Council of 23 October 2018 on the protection of natural persons with regard to the processing of personal data by the Union institutions, bodies, offices and agencies and on the free movement of such data, and repealing Regulation (EC) No 45/2001 and Decision No 1247/2002/EC (Text with EEA relevance). Official Journal of the European Union, L295, 39-98. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32018R1725>
- Flask. (2023) *Flask* <https://flask.palletsprojects.com/en/2.3.x/>
- Flutter. (2023). *Flutter*. <https://flutter.dev/>
- GCP. (2023). *Google Cloud Platform* <https://cloud.google.com>
- Ge, Y., Zhang, R., Wang, X., Tang, X., & Luo, P. (2019). Deepfashion2: A versatile benchmark for detection, pose estimation, segmentation and re-identification of clothing images. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 5337-5345).
- Géron, A. (2019). Hands-on machine learning with SciKit-Learn, KERAS, and Tensorflow: Concepts, Tools, and Techniques to Build Intelligent Systems. «O'Reilly Media, Inc.» ISBN: 9781492032649

- T. L. Giang, K. B. Dang, Q. Toan Le, V. G. Nguyen, S. S. Tong and V. -M. Pham, "U-Net Convolutional Networks for Mining Land Cover Classification Based on High-Resolution UAV Imagery," in IEEE Access, vol. 8, pp. 186257-186273, 2020, doi: 10.1109/ACCESS.2020.3030112.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2020). Generative adversarial networks. Communications of the ACM, 63(11), 139-144.
- R. A. Güler, N. Neverova and I. Kokkinos, "DensePose: Dense Human Pose Estimation in the Wild," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 7297-7306, doi: 10.1109/CVPR.2018.00762.
- Han, X., Wu, Z., Wu, Z., Yu, R., & Davis, L. S. (2018). VITON: An Image-based Virtual Try-on Network. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). [https://openaccess.thecvf.com/content\\_cvpr\\_2018/html/Han\\_VITON\\_An\\_Image-Based\\_CVPR\\_2018\\_paper.html](https://openaccess.thecvf.com/content_cvpr_2018/html/Han_VITON_An_Image-Based_CVPR_2018_paper.html)
- He, H. (2020). The state of machine learning frameworks. The Gradient. <https://thegradient.pub/the-state-of-machine-learning-frameworks/>
- He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 2961-2969).
- iMaterialist (Fashion) 2019 at FGVC6. <https://www.kaggle.com/competitions/imaterialist-fashion-2019-FGVC6/overview>
- Jiang, Y., Yang, S., Qiu, H., Wu, W., Loy, C. C., & Liu, Z. (2022). Text2human: Text-driven controllable human image generation. ACM Transactions on Graphics (TOG), 41(4), 1-11.
- Jordan, J. (2023). Variational autoencoders. Jeremy Jordan. <https://www.jeremyjordan.me/variational-autoencoders/>
- Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 4401-4410).
- Kendall, A., Gal, Y., & Cipolla, R. (2018). Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7482-7491).
- Keras. (2023) Keras <https://keras.io/>
- Kora, P., Ooi, C. P., Faust, O., Raghavendra, U., Gudigar, A., Chan, W. Y., ... & Acharya, U. R. (2022). Transfer learning techniques for medical image analysis: A review. Biocybernetics and Biomedical Engineering, 42(1), 79-107.

- Kotlin. (2023). *Kotlin Android*. <https://kotlinlang.org/>
- L. London, Forbes (2021). Virtual Try-On is more than a pandemic trend and these brands are reaping the rewards. <https://www.forbes.com/sites/lelalondon/2021/05/20/virtual-try-on-is-more-than-a-pandemic-trendand-these-brands-are-reaping-the-rewards/?sh=11fcc2d26c82>
- Le, K. (2022). An overview of VGG16 and NiN models - MLearning.ai - Medium. Medium. <https://medium.com/mllearning-ai/an-overview-of-vgg16-and-nin-models-96e4bf398484>
- Lee, H., & Xu, Y. (2020). Classification of virtual fitting room technologies in the fashion industry: from the perspective of consumer experience. *International Journal of Fashion Design, Technology and Education*, 13(1), 1-10.
- Lee, H., Xu, Y., & Porterfield, A. (2022). Virtual fitting rooms for online apparel shopping: An exploration of consumer perceptions. *Family and Consumer Sciences Research Journal*, 50(3), 189-204.
- Li, H., Zhang, M., Yu, Z., Li, Z., & Li, N. (2022). An improved pix2pix model based on Gabor filter for robust color image rendering. *Mathematical Biosciences and Engineering*, 19(1), 86-101. <https://doi.org/10.3934/mbe.2022004>
- Liaw, C., & Chen, C. (2013). The impact of virtual fitting room technology on consumers' online purchase intention. *Management and Administrative Sciences Review*, 2(1), 23-35.
- Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., ... & Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In *Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13* (pp. 740-755). Springer International Publishing.
- Liu, Z., Luo, P., Qiu, S., Wang, X., & Tang, X. (2016). DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich Annotations. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [https://openaccess.thecvf.com/content\\_cvpr\\_2016/html/Liu\\_DeepFashion\\_Powering\\_Robust\\_CVPR\\_2016\\_paper.html](https://openaccess.thecvf.com/content_cvpr_2016/html/Liu_DeepFashion_Powering_Robust_CVPR_2016_paper.html)
- Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3431-3440).
- Matplotlib. (2023) Matplotlib <https://matplotlib.org/>
- McKinsey (2020). Fashion on Climate. <https://www.mckinsey.com/~media/mckinsey/industries/retail/our%20insights/fashion%20on%20climate/fashion-on-climate-full-report.pdf>

- McKinsey & Company (2023). The State of Fashion 2023. <https://www.mckinsey.com/~media/mckinsey/industries/retail/our%20insights/state%20of%20fashion/2021/the-state-of-fashion-2021-vf.pdf>
- Metail. (2019) *Plataforma desarrolladora de modelos 3D para adaptar la moda a distintos tipos de cuerpo*. <https://metail.com/>
- Mukhlif, A. A., Al-Khateeb, B., & Mohammed, M. A. (2023). Incorporating a Novel Dual Transfer Learning Approach for Medical Images. *Sensors*, 23(2), 570.
- OpenCV. (2023) Open CV <https://opencv.org/>
- Pandas. (2023) Pandas <https://pandas.pydata.org/>
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... & Lin, Z. (2019). PyTorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems* (pp. 8026-8037).
- PyDP. (2023) PyDP <https://github.com/OpenMined/PyDP>
- PyTorch. (2023). *PyTorch* <https://pytorch.org/>
- Ramsundar, B., & Zadeh, R. (2020). *TensorFlow for deep learning: From linear regression to reinforcement learning*. O'Reilly Media.
- React. (2023). *React* <https://es.react.dev/>
- React Native. (2023). *React Native*. <https://reactnative.dev/>
- Ruder, S. (2017). An overview of multi-task learning in deep neural networks. arXiv preprint arXiv:1706.05098.
- Simonyan, K. & Zisserman, A (2014). International Conference on Learning Representation, <http://arxiv.org/abs/1409.1556>.
- Statista. (2023). Digital Market Insights: eCommerce - Worldwide. <https://www.statista.com/outlook/dmo/ecommerce/worldwide/apparel>
- Stitch Fix (2021). *Plataforma de personalización de prendas y asesoramiento de estilo online*. <https://www.stitchfix.com/>
- Sun, K., Xiao, B., Liu, D., & Wang, J. (2019). Deep high-resolution representation learning for human pose estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 5693-5703).

- Swift. (2023). *Swift Apple*. <https://developer.apple.com/swift/>
- TensorFlow. (2023). *TensorFlow* <https://www.tensorflow.org/?hl=es-419>
- Theckedath, D., & Sedamkar, R. R. (2020). Detecting affect states using VGG16, ResNet50 and SE-ResNet50 networks. *SN Computer Science*, 1, 1-7.
- Mukti, I. Z., & Biswas, D. (2019, December). Transfer learning-based plant diseases detection using ResNet50. In 2019 4th International conference on electrical information and communication technology (EICT) (pp. 1-6). IEEE.
- Uizard. (2023). Uizard <https://uizard.io/>
- Wang, W., Xu, Y., Shen, J., & Zhu, S. C. (2018). Attentive fashion grammar network for fashion landmark detection and clothing category classification. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4271-4280).
- Werdayani, D., & Widiaty, I. (2021, March). Virtual fitting room technology in fashion design. In IOP Conference Series: Materials Science and Engineering (Vol. 1098, No. 2, p. 022110). IOP Publishing.
- Yang, S., & Xiong, G. (2019). Try it on! Contingency effects of virtual fitting rooms. *Journal of Management Information Systems*, 36(3), 789-822.
- Zeekit. (2021) *Plataforma de compra online interactiva para marcas de moda*. <https://zeekit.me/>
- Zhang, X., Zhou, X., Lin, M., & Sun, J. (2018). Shufflenet: An extremely efficient convolutional neural network for mobile devices. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 6848-6856).
- Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4 (pp. 3-11). Springer International Publishing.
- Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision (pp. 2223-2232)

## **Anexos**

Se incluyen 3 anexos, entre ellos, el artículo científico asociado al trabajo realizado.

### **Anexo I: Capturas de pantalla de la página web**

Este apartado incluye algunas capturas de pantalla de la página web desarrollada.

### **Anexo II: Resultados de las pruebas de usuario**

Este apartado recoge los resultados de las pruebas de usuario realizadas a 4 voluntarios.

### **Anexo III: Preguntas y respuestas de la encuesta realizada**

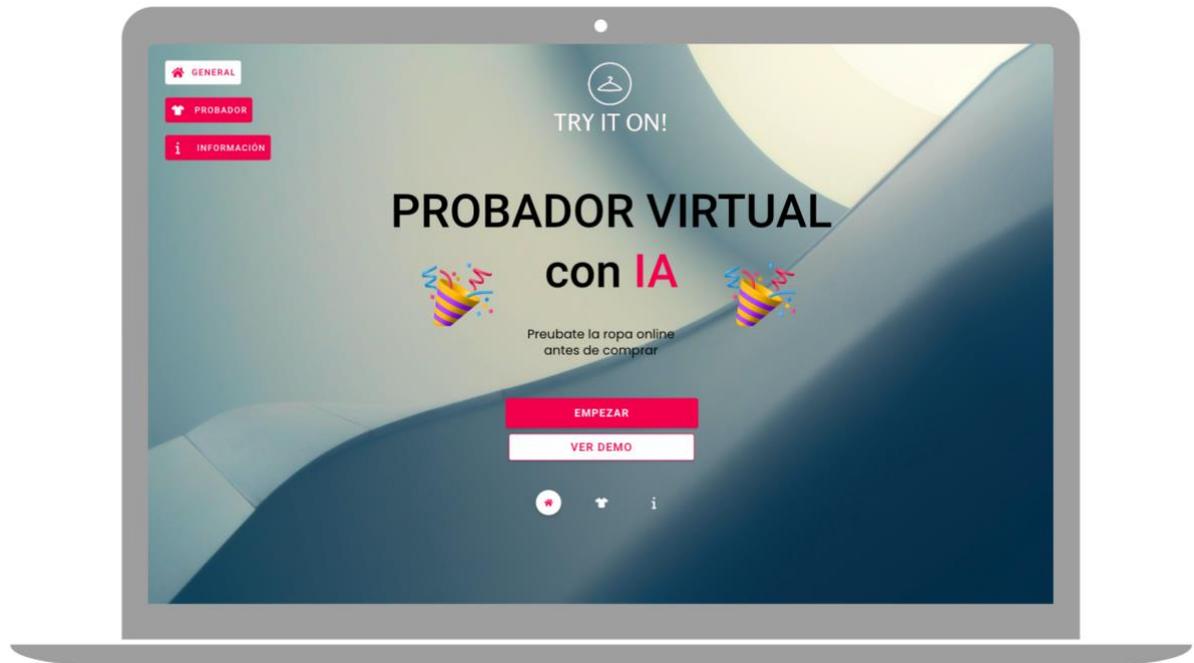
Este apartado incluye las preguntas y respuestas de la encuesta completa realizada a través de Google Docs a los usuarios de la web.

### **Anexo IV: Artículo de investigación**

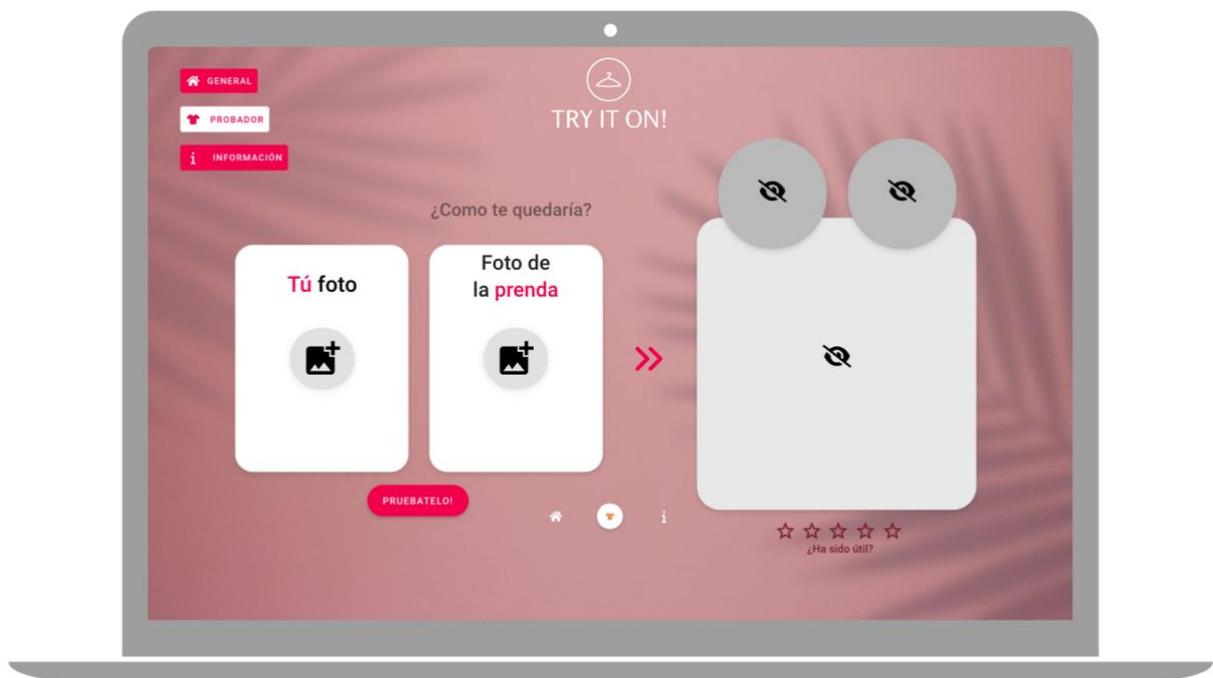
Este apartado incluye el artículo académico de investigación que resumen el trabajo realizado durante el proyecto.

## Anexo I: Capturas de pantalla de la página web

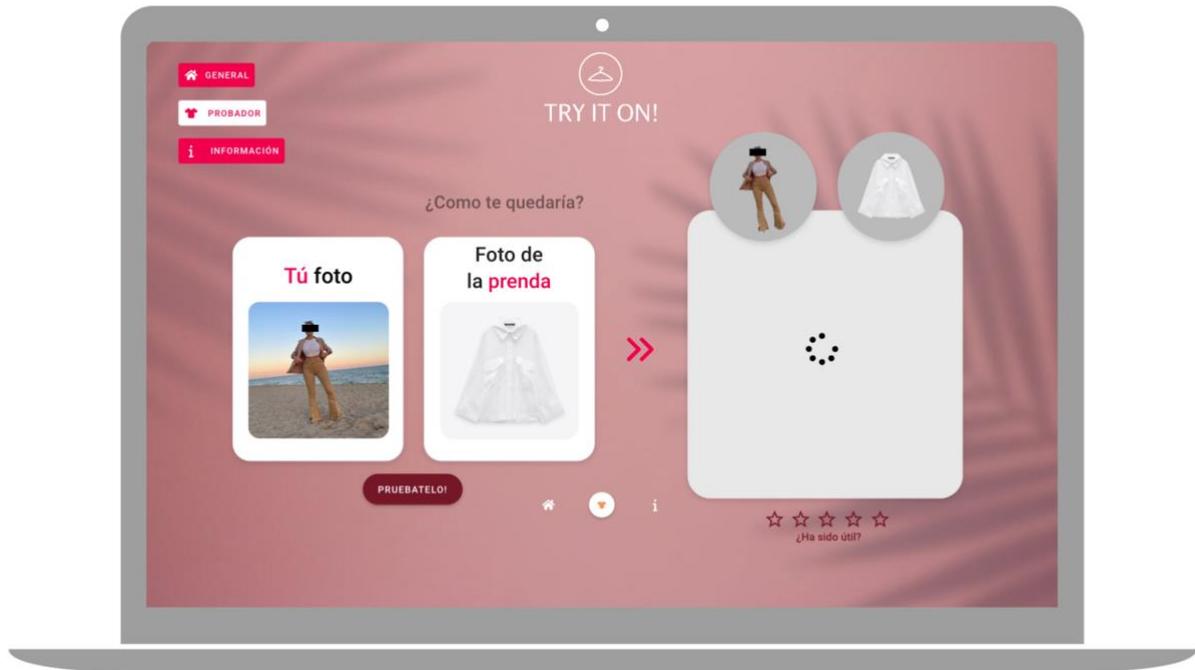
Este apartado incluye algunas capturas de pantalla de la página web desarrollada y elaborada con Flusk.



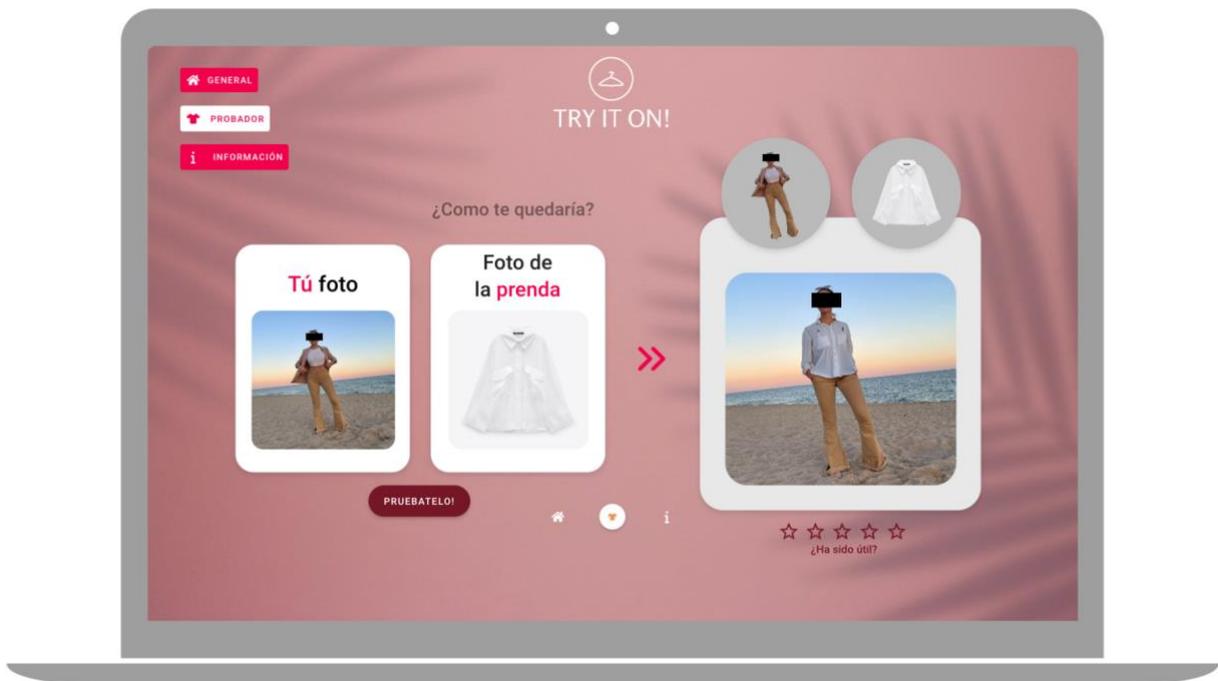
Página de inicio



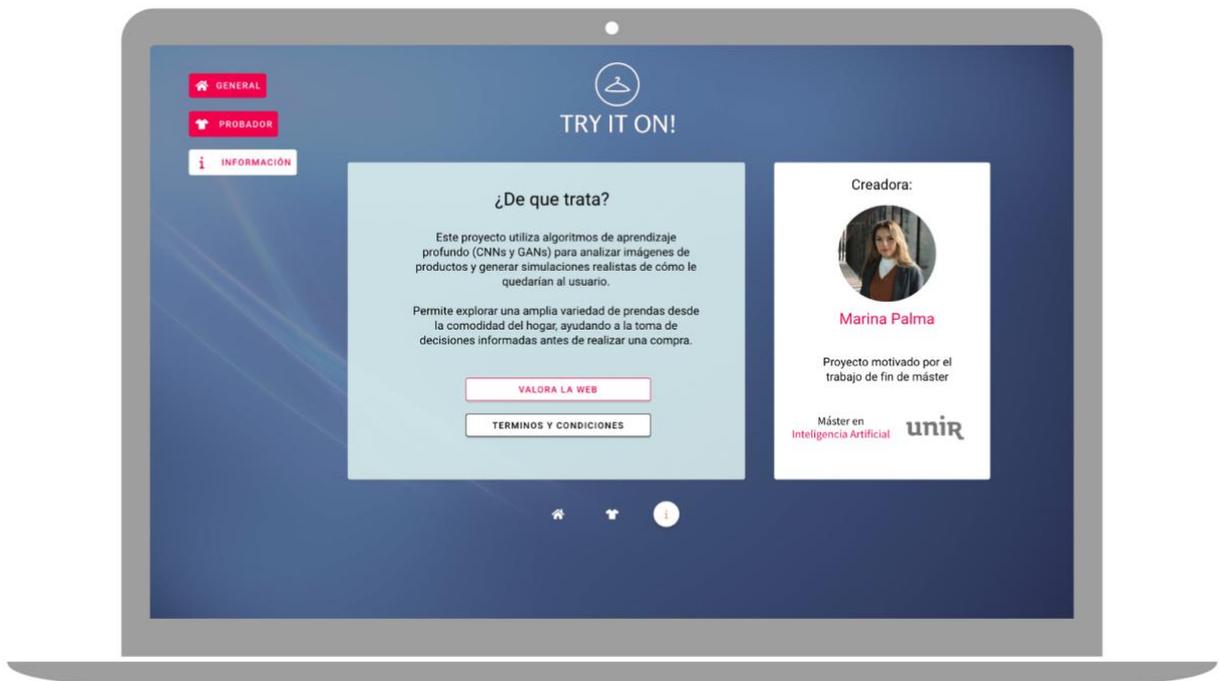
Página del probador (sin imágenes cargadas)



Página del probador (con imágenes cargadas)



Página del probador (con imágenes cargadas y la nueva imagen generada)



Página de información

## Anexo II: Resultados de las pruebas de usuario

Este apartado recoge los resultados de las pruebas de usuario realizadas a 4 voluntarios.





## Anexo III: Preguntas y respuestas de la encuesta realizada

Este apartado incluye las preguntas y respuestas de la encuesta completa realizada a través de Google Docs a los usuarios de la web.

# PROBADOR VIRTUAL

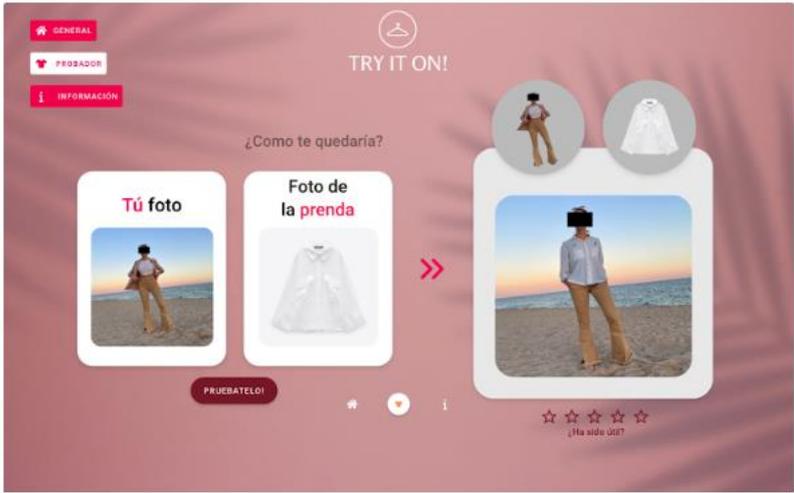
con IA

Pruebate la ropa online antes de comprar

## Encuesta sobre la herramienta

Evaluación de usabilidad, impacto y utilidad

Try it on



¿Como te quedaria?

Tú foto

Foto de la prenda

PRUEBATLO!

(Ha sido útil?)

¿Crees que esta herramienta podría cambiar tu experiencia de compra online? \*

Sí

No

¿Considerarías la posibilidad de utilizar herramientas de este tipo regularmente? \*

Sí

No

¿Cómo de útil encuentras esta herramienta? \*

Nada útil      1      2      3      4      5      Muy útil

¿Estas satisfech@ con la calidad de las imágenes generadas? \*

Nada satisfech@      1      2      3      4      5      Muy satisfech@

¿Las imágenes generadas se asemejan a cómo esperabas que se viera la prenda en el cuerpo? \*

Poco      1      2      3      4      5      Mucho

¿Cómo de realistas eran las imágenes? Independientemente de si han representado fielmente el cuerpo y prenda original.

Poco realistas      1      2      3      4      5      Muy realistas

¿Recomendarías la herramienta a otras personas? \*

Sí

No

¿Tienes alguna recomendación de mejora?

Tu respuesta \_\_\_\_\_

**Puntuación general de la herramienta**

No me gusta      1      2      3      4      5      Me encanta

# Probador virtual basado en redes neuronales convolucionales (CNN) y redes generativas antagónicas (GAN)



Marina Palma Lorita

Universidad Internacional de la Rioja, Logroño (España)

20/07/2023

## PALABRAS CLAVE

CNN,  
Deep Learning,  
GAN,  
Transfer Learning,  
Virtual Fitting Room

## RESUMEN

Este trabajo propone un sistema inteligente de probador virtual en línea para mejorar la experiencia de compra en la industria de la moda. Mediante técnicas de inteligencia artificial, específicamente Redes Neuronales Convolucionales (CNN) y Redes Generativas Adversarias (GAN), se desarrolló un sistema que permite la segmentación de imágenes, la estimación de poses y la generación de imágenes para probar diferentes prendas virtualmente. Además, se implementó una página web funcional para facilitar la interacción con el sistema. Los resultados demuestran una notable mejora en la precisión de la segmentación y estimación de poses, lo que conduce a una simulación más realista de la prueba de ropa. El sistema demostró su eficacia alcanzando una precisión del 96.6% en la segmentación de imágenes, un 57.9% en la estimación de poses, y un equilibrio sólido en la generación y discriminación de imágenes en el modelo GAN. Las conclusiones apuntan a un potencial significativo de la herramienta en la mejora de la experiencia de usuario en el comercio electrónico de moda, a la vez que sugieren oportunidades para futuras mejoras e investigaciones.

## I. INTRODUCCIÓN

La proliferación del comercio electrónico y las crecientes demandas de personalización han amplificado la necesidad de soluciones tecnológicas que aborden los desafíos inherentes a la compra de ropa en línea. Este trabajo se centra en la implementación de un probador virtual basado en inteligencia artificial (IA), específicamente, utilizando redes neuronales convolucionales (CNN) y redes generativas antagónicas (GAN), para simular el ajuste de ropa al cuerpo de cualquier usuario. Este enfoque se basa en recientes avances en visión por computadora y aprendizaje profundo que han permitido una generación y manipulación de imágenes más efectiva.

La idea principal es proporcionar a los compradores una vista previa realista de cómo les quedará la ropa, superponiéndola sobre una imagen sintética de ellos mismos. Esta solución puede transformar no solo la experiencia del consumidor, sino también la toma de decisiones en la industria de la moda al permitir a los fabricantes una vista previa de cómo se verán sus productos en diferentes tipos de cuerpos antes de iniciar la producción en masa.

El desafío de crear un probador virtual realista es considerable. Existen varios problemas técnicos a resolver, como la segmentación de la ropa y el cuerpo en las imágenes, la generación de imágenes artificiales y la superposición de la ropa de manera realista. Sin embargo, el rápido progreso en el campo de las GAN y las CNN, especialmente en el último trienio, ha hecho que estos desafíos sean menos insuperables.

Este trabajo se basa en el estudio y aplicación de los avances más recientes, utilizando conjuntos de datos existentes, como

DeepFashion [1] y el uso de transfer learning en el aprendizaje profundo, con el objetivo de superar los obstáculos técnicos y proporcionar una solución que sea tanto técnicamente sólida como comercialmente viable.

## II. ESTADO DEL ARTE

La revolución digital en la industria de la moda ha generado una ola de innovación en cuanto a la generación de imágenes, la segmentación de imágenes y la estimación de poses mediante el aprendizaje profundo y las redes neuronales. Este documento ofrece una descripción pormenorizada del estado actual de la tecnología y sus implicaciones para el desarrollo de probadores virtuales.

### 1. Conjuntos de Datos Actuales

La eficacia de los modelos de aprendizaje profundo depende, en gran medida, de la calidad y diversidad del conjunto de datos utilizado para el entrenamiento. DeepFashion [1] y DeepFashion2 [2] son conjuntos de datos prominentes en el ámbito de los probadores virtuales, que contienen más de 800.000 imágenes etiquetadas de ropa. Ambos conjuntos de datos ofrecen variedad en poses, pesos, alturas y detalles de diseño, proporcionando un recurso valioso para el entrenamiento y evaluación de modelos.

### 2. Soluciones Actuales para la Segmentación de Imágenes con CNN

La segmentación de imágenes es esencial para separar a las distintas prendas de vestir y partes de las imágenes. Las arquitecturas de codificador-decodificador, como U-Net [3] y

SegNet [4], son populares para esta tarea. Estas arquitecturas, basadas en CNN, extraen características de la imagen de entrada y reconstruyen la segmentación a partir de estas características, lo que permite una segmentación precisa.

### 3. Soluciones Actuales para la Estimación de Poses con CNN

La estimación de poses, esencial para comprender la postura y la orientación del cuerpo humano, también ha avanzado gracias al uso de las CNN. Un enfoque reciente utiliza mapas de calor para las articulaciones del cuerpo humano, que la CNN combina para obtener la postura humana completa [5]. Deep High-Resolution Representation Learning (HRNet) [6], y DensePose [7] son ejemplos destacados de estas arquitecturas.

### 4. Soluciones Actuales para la Generación de Imágenes de Personas con GAN

En la generación de imágenes, las GAN han demostrado ser una herramienta poderosa y efectiva para crear imágenes realistas. VITON [8] es un sistema notable que utiliza GAN y CNN para crear imágenes realistas de personas con ropa seleccionada a partir de imágenes de catálogos.

### 5. Limitaciones de los modelos actuales

A pesar de los avances en la generación de imágenes, la segmentación de imágenes y la estimación de poses, los modelos actuales aún enfrentan varios desafíos. Uno de ellos es la variabilidad y diversidad en los datos, como diferentes posturas y apariencias que pueden afectar el rendimiento de los modelos [1]. Además, el manejo de poses poco comunes y entornos con varias personas sigue siendo un desafío para los modelos de estimación de poses. En cuanto a la generación de imágenes, los problemas como el colapso de modos, donde el generador produce imágenes poco variadas o similares, y la generación de artefactos visuales o inconsistencias en las imágenes, limitan la calidad y el realismo de las representaciones generadas. La Fig. 1 muestra algunos ejemplos en los que la red generativa no produjo imágenes realistas en el trabajo de VITON [8].



Fig. 1. Casos de fallo en el trabajo de VITON [8]

### 6. Transfer learning para mejorar el rendimiento

Para superar estos desafíos, el aprendizaje por transferencia ha demostrado ser una herramienta eficaz para mejorar el rendimiento de los modelos de IA. Al utilizar modelos previamente entrenados en grandes conjuntos de datos, se puede mejorar el rendimiento y reducir el tiempo de entrenamiento. Se ha demostrado que el aprendizaje por transferencia puede mejorar significativamente el rendimiento de los modelos de estimación de poses, incluso en condiciones desafiantes como las oclusiones y las poses extremas [9].

#### 6.1. VGG16

El modelo preentrenado VGG16 [10] consiste en 13 capas convolucionales, organizadas en bloques con un número creciente de filtros, y 3 capas completamente conectadas. Los

filtros extraídos por este modelo son altamente generalizables y, por lo tanto, proporcionan una base sólida para muchas tareas de deep learning, cómo por ejemplo, la segmentación de imágenes.

#### 6.2. ResNet50

El modelo preentrenado ResNet50 [11] es una red convolucional profunda que se destaca por sus conexiones residuales, permitiendo la propagación directa de la señal de entrada a capas más profundas y aportando una alta precisión en la estimación de poses.

#### 6.3. Pix2pix

El modelo preentrenado Pix2pix [12] se centra en tareas de generación de imágenes. Su capacidad para retener detalles finos en las imágenes de salida es particularmente útil para la generación detallada de imágenes de personas y animales.

## 7. Tecnologías para la implementación

Para el desarrollo e implementación de los modelos de aprendizaje profundo es necesario conocer los principales frameworks y bibliotecas de aprendizaje profundo, así como servicios en la nube que amplíen los recursos locales.

### 7.1. Frameworks y bibliotecas

Para implementar estos modelos, se utilizan varios frameworks y bibliotecas, entre los que se encuentran TensorFlow, PyTorch y Keras, que ofrecen herramientas y funcionalidades para facilitar el entrenamiento y la evaluación de modelos de IA [13]. Además, las bibliotecas como OpenPose proporcionan herramientas para la estimación de poses y StyleGAN ofrece un marco para la generación de imágenes con GAN [14].

### 7.2. Tecnologías de despliegue en la nube

La implementación y el despliegue de estos modelos pueden beneficiarse de las herramientas de la nube, como AWS SageMaker, Google AI Platform y Azure Machine Learning, que proporcionan recursos para el entrenamiento, la evaluación y el despliegue de modelos de IA a gran escala.

## III. OBJETIVOS Y METODOLOGÍA

### 1. Objetivos y requisitos del sistema

Este proyecto tiene como objetivo general el desarrollo y la implementación de un sistema de probadores virtuales basado en IA que genere una imagen sintética de los usuarios vistiendo diferentes prendas de manera realista y rápida. Mas específicamente, se incluyen la investigación y aplicación de técnicas de IA, la selección de bibliotecas y frameworks adecuados, la creación de un *dataset* de entrenamiento y evaluación de los algoritmos, el desarrollo de algoritmos de segmentación, estimación de poses y generación de imágenes, y la evaluación exhaustiva de la solución propuesta.

### 2. Creación del dataset

En esta etapa, se utilizó el *dataset* DeepFashion-MultiModal [15], el cual consiste en 44.096 imágenes humanas de alta resolución con anotaciones detalladas de las prendas de vestir.

El procesamiento de los datos se aplicó a tres componentes del conjunto de datos: Las imágenes originales, las imágenes segmentadas y las coordenadas que identifican la postura. Estos datos se procesan para reducir su dimensionalidad y adaptarlos al tamaño requerido para el sistema.

Se contempló la idea de aplicar *Data Augmentation* para aumentar la diversidad de los datos, pero se decidió no hacerlo debido a la complejidad adicional que representaría transformar las anotaciones detalladas que acompañan a cada imagen y al costo computacional que esto implicaría. En su lugar, los recursos se han destinado a la optimización de los modelos y al entrenamiento.

### 3. Entrenamiento del modelo de segmentación de imágenes

Se desarrolló un modelo de segmentación de imágenes que se basa en las Redes Neuronales Convolucionales (CNN). El propósito de dicho modelo es identificar y clasificar diferentes regiones en una imagen correspondientes a las prendas de vestir representadas.

Para llevar a cabo el entrenamiento del modelo, se recurrió al enfoque de aprendizaje por transferencia, utilizando el modelo VGG16 para la extracción de características en la fase de codificación [16]. La arquitectura del modelo de segmentación está basada en U-Net [3], muy utilizada en tareas de segmentación de imágenes. La red se compone de un codificador (o fase de contracción) y un decodificador (o fase de expansión). El modelo VGG16 [10], menos las capas completamente conectadas, conforma el codificador, el cual se encarga de extraer las características de las imágenes de entrada. Por su parte, el decodificador está compuesto por una serie de capas de convolución y aumento de resolución con el propósito de reconstruir la imagen a partir de las características extraídas. En la Fig. 2 se puede ver la arquitectura del modelo de segmentación definido.

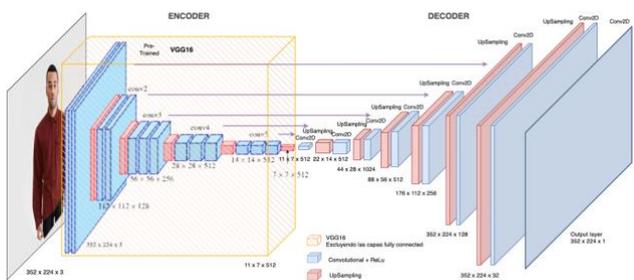


Fig. 2. Arquitectura del modelo de segmentación

### 4. Entrenamiento del modelo de estimación de poses

El segundo modelo del proyecto se desarrolla con el propósito de detallar la posición del cuerpo a partir de una imagen. Este desafío se aborda detectando una serie de puntos clave denotan la pose del cuerpo humano. Se emplea un enfoque combinado de transferencia de aprendizaje y CNN para construir este modelo.

La arquitectura del modelo de estimación de poses es una combinación del modelo preentrenado ResNet50 [11] y un clasificador personalizado. Se incorporan capas de GlobalAveragePooling2D, capas Densas y de Dropout. Finalmente, se utiliza una función de activación lineal ya que es un problema de regresión. En la Fig. 3 se puede observar el detalle

de la construcción de la arquitectura para el modelo de estimación de poses.

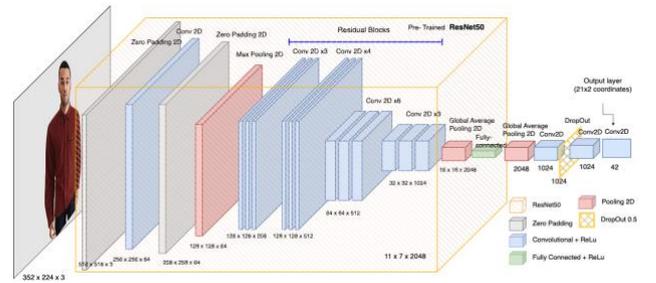


Fig. 3. Arquitectura del modelo de estimación de poses

### 5. Entrenamiento del modelo de generación de imágenes

La última etapa del desarrollo de IA en este proyecto consiste en el modelo de generación de imágenes con Redes Generativas Antagónicas (GAN). Esta fase es integral y se compone de múltiples etapas que abordan diferentes aspectos del problema: adaptación de la prenda al cuerpo del usuario, generación de partes del cuerpo no cubiertas por la prenda y la superposición final de la prenda adaptada sobre el cuerpo del usuario.

La arquitectura de la GAN se divide en dos partes: el generador y el discriminador. El generador, responsable de generar nuevas imágenes sintéticas, se divide en dos procesos: adaptación de la prenda y generación de partes no cubiertas del cuerpo. Se emplea el modelo preentrenado VGG16 [10] para la extracción de características y una Red de transformación espacial (STN) para adaptar la forma de la prenda a la pose del cuerpo. Luego, se aplican las capas generadoras del modelo pix2pix [12] para generar las partes no cubiertas del cuerpo. El discriminador, por otro lado, es una red convolucional que clasifica si una imagen es real o sintética, utilizando las imágenes de entrada y salida correspondientes como pares de entrenamiento. Las figuras Fig. 4 y Fig. 5 muestran la arquitectura de la fase de adaptación de la prenda y generación de partes no cubiertas del generador, respectivamente.

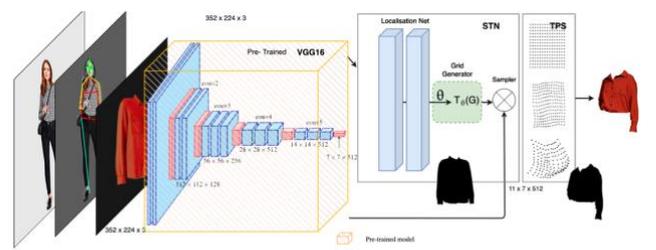


Fig. 4. Arquitectura del generador en la fase de adaptación de la prenda

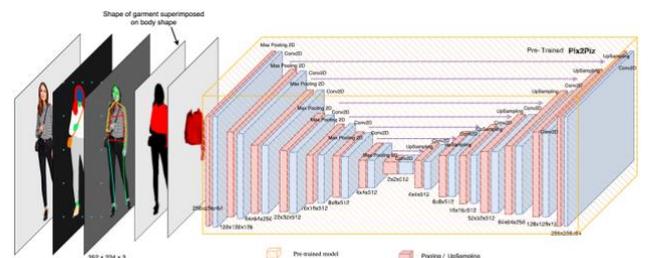


Fig. 5. Arquitectura del generador en la fase de generación de partes no cubiertas

#### IV. CONTRIBUCIÓN

Este trabajo contribuye significativamente al campo de la inteligencia artificial y el aprendizaje profundo, proporcionando un marco eficiente y sofisticado para la generación automática de imágenes de usuarios vistiendo diferentes prendas de vestir. Mediante la combinación de técnicas avanzadas como la segmentación de imágenes y la estimación de poses con redes neuronales convolucionales (CNN) y la generación de imágenes con Redes Generativas Antagónicas (GAN), este trabajo ofrece un método potente y adaptable para generar imágenes de alta calidad que se adaptan a las formas y posturas del cuerpo del usuario. Este avance podría tener aplicaciones significativas en comercio electrónico y marketing de moda, mejorando la experiencia de los clientes y acelerando la producción y promoción de nuevas líneas de ropa.

#### V. EVALUACIÓN Y RESULTADOS

A continuación, se evalúan los resultados de la metodología empleada.

##### Evaluación del modelo de segmentación de imágenes

El modelo de segmentación de imágenes se entrenó en 10 épocas con lotes de 16 imágenes durante casi 38 horas. La pérdida mínima conseguida fue de 0,1020 y la precisión máxima del 96,65%. La Tab. 1 muestra el resumen de los resultados obtenidos.

Epochs	Batch Size	Training Time (min)	Min (Loss)	Max(Accuracy)
<b>10</b>	<b>16</b>	<b>2.247</b>	<b>0,1020</b>	<b>96,65%</b>

Tab. 1. Resultados del entrenamiento del modelo de segmentación

Durante la evaluación cualitativa del modelo de segmentación, se pudo observar su desempeño al analizar visualmente la segmentación de nuevas imágenes y compararla con las segmentaciones realizadas manualmente. En general, el modelo demostró una alta capacidad para distinguir ciertos segmentos, lo que indica que el modelo puede capturar las características relevantes de las prendas. La Fig. 6 siguiente muestra algunos de los mejores resultados obtenidos.

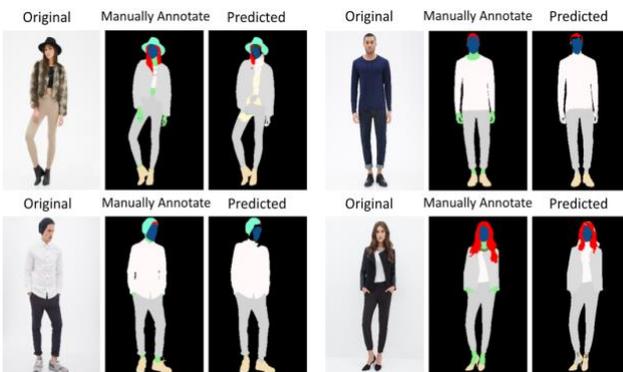


Fig. 6. Resultados del modelo de segmentación

Aun con la notable precisión del modelo, se identificaron problemas para diferenciar entre partes superiores y prendas tipo chaqueta, posiblemente debido a la similitud visual y proximidad espacial de estas categorías en las imágenes. Además, la distinción entre la piel de la persona y las prendas no fue tan precisa, probablemente debido a las variaciones en el tono de piel

y textura de la ropa. Estas áreas de mejora señalan oportunidades para continuar optimizando el modelo y su capacidad de segmentación.

##### Evaluación del modelo de estimación de poses

El modelo de segmentación de imágenes se entrenó en 10 épocas con lotes de 16 imágenes durante aproximadamente 22 horas. La pérdida mínima conseguida fue de 208,87 y la precisión máxima del 57,88%. La Tab. 2 muestra el resumen de los resultados obtenidos.

Epochs	Batch Size	Training Time (min)	Min (Loss)	Max(Accuracy)
<b>10</b>	<b>16</b>	<b>1.300</b>	<b>208,87</b>	<b>57,88%</b>

Tab. 2. Resultados del entrenamiento del modelo de estimación de poses

Durante la evaluación cualitativa del modelo de estimación de poses, se compararon visualmente las predicciones del modelo con las anotaciones manuales. En la Fig. 7 se pueden ver algunos resultados de las poses anotadas manualmente en verde y las estimaciones del modelo en rojo.



Fig. 7. Resultados del modelo de estimación de poses

En líneas generales, el modelo ha demostrado un rendimiento satisfactorio en la estimación de las proporciones del cuerpo. Sin embargo, se observaron algunas dificultades específicas en la representación de los puntos que indican la posición de los hombros y los tobillos. Estas dificultades pueden atribuirse a la variabilidad en la posición y apariencia de estos puntos en diferentes poses e imágenes. Estas áreas identificadas para mejorar sugieren oportunidades de optimización para lograr estimaciones de poses aún más precisas y consistentes.

##### Evaluación del modelo de generación de imágenes

El entrenamiento del modelo de generación de imágenes ha sido un desafío significativo en términos de esfuerzo y requerimientos computacionales. El entrenamiento ha consistido en 500 épocas con lotes de 32 imágenes. En la Fig. 8 se puede observar como la pérdida en el discriminador se estabilizaba mientras a lo largo de las épocas, mientras que en el generador aumentaba ligeramente, reflejando la naturaleza adversarial de las GAN.

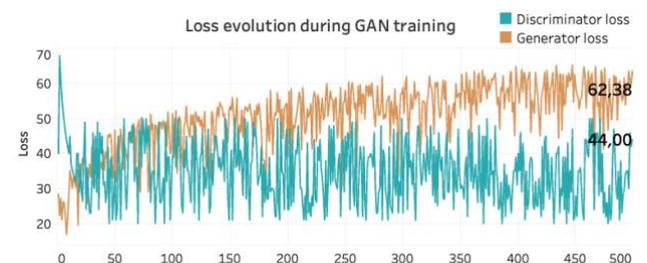


Fig. 8. Evolución de la pérdida durante el entrenamiento de la GAN

En cuanto al análisis cualitativo, los resultados fueron favorables, con el modelo reproduciendo las prendas deseadas de manera precisa en términos de formas y colores. La Fig. 9 muestra un ejemplo de cómo una nueva prenda se ajusta al cuerpo de la imagen original.



Fig. 9. Resultados del modelo de generación de imágenes

Aún así, se identificaron áreas de mejora, como zonas borrosas en la generación de piel y dificultades en la creación de manos y el recorte de cabello. La Fig. 10 muestra en aumento algunas de estas incongruencias.

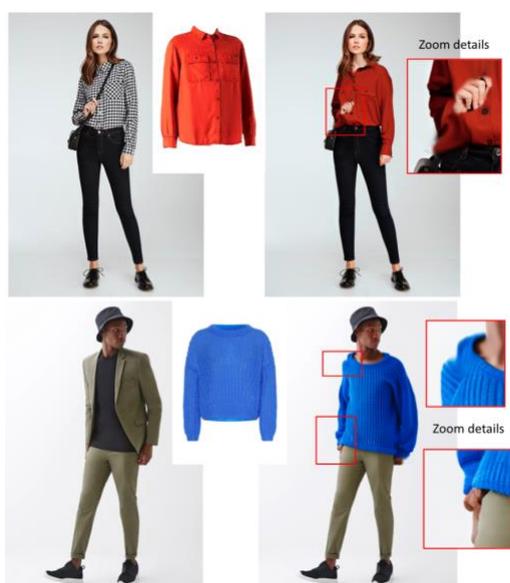


Fig. 10. Ejemplos de áreas de mejora en el modelo GAN

A pesar de estas áreas de mejora, en general los resultados fueron positivos, demostrando la capacidad del modelo para generar imágenes convincentes de prendas de vestir sobre personas.

## VI. DISCUSIÓN

La discusión de los resultados revela varias limitaciones a abordar en futuras iteraciones del sistema de Inteligencia Artificial. Uno de los retos clave radica en la precisión en la adaptación de las prendas al cuerpo al usuario, especialmente en casos de prendas con diseños y texturas complejas. Este desafío podría verse exacerbado por errores en las etapas iniciales del proceso, tales como la estimación de la pose y la segmentación de la imagen, que pueden propagarse y afectar la calidad del generador. Adicionalmente, el modelo de generación de imágenes ha demostrado dificultades en la reproducción precisa de ciertos detalles, como las manos y el cabello, generando en ocasiones áreas borrosas o deformadas. Esto podría atribuirse a la complejidad de generar estas características tan detalladas.

Una limitación adicional, reside en el conjunto de datos de entrenamiento utilizado. A pesar de la intención inicial de crear un sistema que sea representativo de cualquier usuario, la realidad es que la diversidad de los tipos de cuerpo en la población supera las representaciones disponibles en los conjuntos de datos actuales. Como resultado, la eficacia del sistema puede verse limitada por la similitud entre el usuario y los ejemplos utilizados durante el entrenamiento del modelo.

A pesar de estos desafíos, es crucial destacar que el sistema en general logra cumplir con sus objetivos y produce resultados satisfactorios en múltiples aspectos.

## VII. CONCLUSIONES

A través del desarrollo de un sistema de inteligencia artificial secuencial que integra tres modelos de aprendizaje profundo: segmentación de prendas, estimación de poses y generación de imágenes, se ha avanzado considerablemente en la adaptación de prendas virtuales al cuerpo de los usuarios. En particular, el sistema propuesto ha demostrado su capacidad para segmentar prendas de una imagen, estimar la pose del cuerpo humano y generar una nueva imagen que superpone la prenda segmentada en la pose estimada.

Respecto a los objetivos planteados inicialmente, los resultados obtenidos revelan que se ha conseguido un grado notable de éxito. El sistema demostró su eficacia alcanzando una precisión del 96.6% en la segmentación de imágenes, un 57.9% en la estimación de poses, y un equilibrio sólido en la generación y discriminación de imágenes en el modelo GAN. Aunque se han identificado desafíos y limitaciones, como la precisión en la adaptación de prendas y la representación detallada de ciertos aspectos, en términos generales, el sistema propuesto demuestra su capacidad para producir resultados satisfactorios en una amplia gama de casos.

Este trabajo abre numerosas líneas de investigación y desarrollo futuro. Las limitaciones identificadas proporcionan un claro marco para futuras mejoras del sistema. Por ejemplo, podrían desarrollarse técnicas más sofisticadas para la generación de imágenes, para manejar mejor la complejidad de las texturas y diseños de las prendas.

Por último, la contribución de este trabajo tiene aplicaciones potenciales más allá de la prueba de vestimenta virtual. Podría emplearse en el comercio electrónico, la moda, la industria del entretenimiento, entre otros, proporcionando un valor añadido significativo en estos campos. De hecho, el sistema podría ser un paso crucial hacia una experiencia de compra en línea más personalizada y atractiva, que podría transformar el panorama del comercio electrónico y de la moda en el futuro.

## REFERENCIAS

- [1] Liu, Z., Luo, P., Qiu, S., Wang, X., & Tang, X. (2016). DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich Annotations. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). [https://openaccess.thecvf.com/content\\_cvpr\\_2016/html/Liu\\_DeepFashion\\_Powering\\_Robust\\_CVPR\\_2016\\_paper.html](https://openaccess.thecvf.com/content_cvpr_2016/html/Liu_DeepFashion_Powering_Robust_CVPR_2016_paper.html)

- [2] Ge, Y., Zhang, R., Wang, X., Tang, X., & Luo, P. (2019). Deepfashion2: A versatile benchmark for detection, pose estimation, segmentation and re-identification of clothing images. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 5337-5345).
- [3] Giang, T. L., Dang, K. B., Le, Q. T., Nguyen, V. G., Tong, S. S., & Pham, V. M. (2020). U-Net convolutional networks for mining land cover classification based on high-resolution UAV imagery. *Ieee Access*, 8, 186257-186273.
- [4] Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12), 2481-2495.
- [5] Chen, Y., Wang, Z., Peng, Y., Zhang, Z., Yu, G., & Sun, J. (2018). Cascaded pyramid network for multi-person pose estimation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7103-7112).
- [6] Sun, K., Xiao, B., Liu, D., & Wang, J. (2019). Deep high-resolution representation learning for human pose estimation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 5693-5703).
- [7] Güler, R. A., Neverova, N., & Kokkinos, I. (2018). Densepose: Dense human pose estimation in the wild. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7297-7306).
- [8] Han, X., Wu, Z., Wu, Z., Yu, R., & Davis, L. S. (2018). VITON: An Image-based Virtual Try-on Network. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). [https://openaccess.thecvf.com/content\\_cvpr\\_2018/html/Han\\_VITON\\_An\\_Image-Based\\_CVPR\\_2018\\_paper.html](https://openaccess.thecvf.com/content_cvpr_2018/html/Han_VITON_An_Image-Based_CVPR_2018_paper.html)
- [9] Kora, P., Ooi, C. P., Faust, O., Raghavendra, U., Gudigar, A., Chan, W. Y., ... & Acharya, U. R. (2022). Transfer learning techniques for medical image analysis: A review. *Biocybernetics and Biomedical Engineering*, 42(1), 79-107.
- [10] Simonyan, K. & Zisserman, A (2014). International Conference on Learning Representation, <http://arxiv.org/abs/1409.1556>.
- [11] Mukti, I. Z., & Biswas, D. (2019, December). Transfer learning based plant diseases detection using ResNet50. In 2019 4th International conference on electrical information and communication technology (EICT) (pp. 1-6). IEEE.
- [12] Li, H., Zhang, M., Yu, Z., Li, Z., & Li, N. (2022). An improved pix2pix model based on Gabor filter for robust color image rendering. *Mathematical Biosciences and Engineering*, 19(1), 86-101. <https://doi.org/10.3934/mbe.2022004>
- [13] He, H. (2020). The state of machine learning frameworks. *The Gradient*. <https://thegradient.pub/the-state-of-machine-learning-frameworks/>
- [14] Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 4401-4410).
- [15] Jiang, Y., Yang, S., Qiu, H., Wu, W., Loy, C. C., & Liu, Z. (2022). Text2human: Text-driven controllable human image generation. *ACM Transactions on Graphics (TOG)*, 41(4), 1-11.