Facultad de Ciencias Económicas y Empresariales

Doble Grado en Ade y Business Analytics (E2-Analytics)

Evaluación comparativa de pictogramas generados por Inteligencia Artificial para niños con autismo: modelo chino vs modelo estadounidense

Autor: Nuria Castillo García

Director: Eduardo César Garrido Merchán

202008199@alu.comillas.edu

MADRID | Junio 2025

Resumen

Con el objetivo de contribuir a paliar las dificultades de comunicación en niños con Trastorno del Espectro Autista (TEA) severo, este trabajo analiza el potencial de la inteligencia artificial generativa para crear pictogramas basados en el estilo ARASAAC. Se comparan dos modelos, DALL·E 3 (OpenAI) y Qwen2.5-Max (Alibaba DAMO Academy), a partir de 25 acciones cotidianas transformadas en pictogramas mediante prompts visuales y textuales. La evaluación se realiza mediante una rúbrica específica que puntúa cada imagen en cinco dimensiones clave. Posteriormente, se aplica un contraste de hipótesis para valorar el rendimiento de ambos modelos y determinar cuál ofrece mejores resultados en este contexto educativo y terapéutico.

Palabras clave: inteligencia artificial generativa, pictogramas, autismo, comunicación aumentativa, DALL·E 3, Qwen, ChatGPT

Abstract

In order to help mitigate communication difficulties in children with severe Autism Spectrum Disorder (ASD), this study analyzes the potential of generative artificial intelligence to create pictograms based on the ARASAAC style. Two models, DALL·E 3 (OpenAI) and Qwen2.5-Max (Alibaba DAMO Academy), are compared using 25 everyday actions transformed into pictograms through visual and textual prompts. The evaluation is conducted using a specific rubric that scores each image across five key dimensions. A hypothesis test is then applied to assess the performance of both models and determine which one delivers better results in this educational and therapeutic context.

Keywords: generative artificial intelligence, pictograms, autism, augmentative communication, DALL·E 3, Qwen, ChatGPT

ÍNDICE

Lista de tablas

Lista de figuras

Lista de ecuaciones

1. Introducción

En el último año, la inteligencia artificial (IA) ha experimentado un crecimiento sin precedentes, impulsado por avances en modelos generativos capaces de transformar texto e imágenes en contenidos originales de alta calidad. Este desarrollo ha dado lugar a una amplia proliferación de herramientas en el mercado, cada vez más accesibles y potentes. Hasta el momento, el modelo más conocido y utilizado ha sido ChatGPT, desarrollado por OpenAI, cuya popularidad lo ha convertido en una referencia mundial en el ámbito de la IA generativa (Peñalvo et al., 2024). Sin embargo, la supremacía de este modelo se está viendo cuestionada por la aparición de nuevos competidores, especialmente provenientes del ecosistema tecnológico chino.

La inteligencia artificial china ha cobrado una creciente relevancia en el panorama internacional, generando titulares en medios globales por su potencial para rivalizar e incluso superar al modelo estadounidense en ciertos aspectos clave. Tal y como apunta un artículo de la BBC, esta rivalidad tecnológica ha adquirido una dimensión estratégica y simbólica, consolidando la IA como un nuevo frente de competencia entre potencias (BBC, 2024).

Más allá de la dimensión geopolítica, esta competición también plantea oportunidades concretas en ámbitos de alto impacto social. Entre las múltiples aplicaciones que permite la IA generativa, destaca su potencial para facilitar la vida de colectivos con necesidades especiales. En este trabajo se explora una de estas aplicaciones: el uso de modelos de inteligencia artificial generativa de imagen para apoyar la comunicación de niños con Trastorno del Espectro Autista (TEA).

El TEA es un trastorno del neurodesarrollo que afecta a millones de personas en todo el mundo y se caracteriza principalmente por alteraciones en la interacción social, la comunicación verbal y no verbal, y la presencia de patrones de conducta repetitivos o restringidos (American Psychiatric Association, 2013). Estas manifestaciones pueden variar considerablemente entre individuos, tanto en intensidad como en forma, lo que ha llevado a su conceptualización como un espectro amplio y diverso. Esta variabilidad implica que no existen dos casos iguales de TEA, lo que representa un desafío importante tanto para el diagnóstico como para la intervención clínica y educativa. El manual diagnóstico DSM-5 contempla esta heterogeneidad clasificando el TEA en tres niveles de

severidad, definidos en función del grado de apoyo necesario para desenvolverse en la vida cotidiana (Lord et al., 2020). En este marco, el nivel 3 representa la categoría más

severa y es en la que se focaliza este estudio, al estar caracterizada por limitaciones comunicativas extremadamente significativas, con escasa o nula capacidad de lenguaje funcional, junto con una alta dependencia de apoyos externos para poder establecer cualquier tipo de interacción significativa con el entorno. Ante esta realidad, la intervención temprana, adaptada a las particularidades de cada caso, se ha consolidado como una de las estrategias más eficaces para favorecer el desarrollo de habilidades comunicativas, cognitivas y sociales. La personalización de estas estrategias no solo mejora la calidad de vida de los niños afectados, sino que también contribuye de manera decisiva a fomentar su autonomía progresiva, su bienestar emocional y su integración en contextos educativos, familiares y sociales (Rojas-Torres, Alonso-Esteban y Alcantud-Marín, 2020).

En este contexto, el presente trabajo se propone evaluar si los modelos de IA generativa actuales, representados por el modelo estadounidense DALL·E 3 y el modelo chino Qwen-VL, pueden ser utilizados para generar pictogramas comprensibles para niños con autismo severo. En concreto, se pretende analizar si la supuesta amenaza tecnológica de la IA china frente a la estadounidense se traduce también en una diferencia significativa de calidad y funcionalidad en tareas de alto valor social como la generación de herramientas de comunicación aumentativa

2. Estado del Arte

El interés científico por el autismo comenzó a tomar forma en la década de 1940, cuando Leo Kanner (1943) introdujo el término "autismo infantil precoz" en la literatura psiquiátrica, a partir de la observación de once niños con comportamientos inusuales, dificultades para establecer relaciones sociales y patrones de conducta altamente repetitivos. Poco después, Hans Asperger (1944) describió un cuadro clínico similar en varios niños que, pese a mostrar una inteligencia media o superior, presentaban un estilo de comunicación reiterativo, escasa expresión emocional y una fuerte fijación por temas concretos. Los denominó "pequeños profesores", debido a su tendencia a exponer con gran detalle sus intereses especializados.

Durante décadas, estos perfiles se consideraron diagnósticos independientes. No fue hasta la publicación del DSM-5 por la American Psychiatric Association (2013) cuando se consolidó el concepto unificado de Trastorno del Espectro Autista (TEA), integrando los subtipos previos en un único espectro continuo. Esta clasificación introdujo además tres niveles de severidad en función del grado de apoyo requerido, lo que ha permitido adaptar mejor las intervenciones a las características individuales de cada caso. Esta perspectiva reconoce la gran heterogeneidad del trastorno, tanto en lo cognitivo como en lo comunicativo, y ha favorecido un enfoque más flexible e inclusivo, no solo en contextos clínicos sino también en entornos familiares y sociales.

En este marco, la comunicación se ha consolidado como uno de los grandes desafíos asociados al TEA. Numerosos estudios han documentado las dificultades que presentan los niños con autismo para desarrollar un lenguaje funcional, interpretar gestos, comprender intenciones o expresar emociones de forma adecuada (Rojas-Torres et al., 2020). Estas barreras no solo afectan su capacidad para interactuar con los demás, sino que también condicionan su autonomía, su integración social y su bienestar emocional. La necesidad de encontrar formas alternativas y accesibles de comunicación ha sido uno de los grandes motores de innovación en este ámbito.

Con este objetivo, las tecnologías digitales han comenzado a desempeñar un papel creciente en el desarrollo de estrategias de comunicación aumentativa, especialmente entre niños con TEA que presentan escaso o nulo lenguaje verbal. Tal como señalan Moraiti et al. (2023), dispositivos como tablets, aplicaciones móviles o software

interactivo se han convertido en herramientas clave para facilitar la expresión de necesidades, emociones o ideas en personas con trastornos del neurodesarrollo. Entre las soluciones más utilizadas destacan aplicaciones como Proloquo2Go o Avaz, que permiten construir mensajes mediante pictogramas, aunque su personalización sigue siendo limitada en muchos casos.

No obstante, la adopción de estas tecnologías no ha estado exenta de dificultades. Por un lado, persiste la falta de contenidos realmente adaptados a las distintas necesidades comunicativas; por otro, muchos cuidadores y profesionales carecen de formación específica para seleccionar y utilizar estas herramientas de forma efectiva (Pino y Guerrero, 2021). A pesar de ello, el avance de la tecnología ha abierto nuevas líneas de investigación que buscan superar estas barreras, incorporando soluciones más dinámicas, interactivas y ajustables en tiempo real, como es el caso de la inteligencia artificial generativa, que se abordará en los siguientes apartados.

La progresiva incorporación de tecnologías digitales ha abierto nuevas vías para apoyar la comunicación de personas con TEA, especialmente en aquellos casos donde el lenguaje verbal está limitado o ausente. Sin embargo, las herramientas tradicionales, como los softwares de pictogramas o los tableros visuales, presentan limitaciones en cuanto a adaptabilidad, personalización y capacidad de respuesta. En este contexto, emerge una alternativa cada vez más prometedora: la inteligencia artificial (IA).

La IA ha avanzado notablemente en la última década, dejando de ser una disciplina experimental para convertirse en una tecnología transversal con aplicaciones en salud, asistencia personal, comunicación y diversidad funcional. Su capacidad para reconocer patrones complejos, aprender de datos y generar respuestas en tiempo real la posiciona como un recurso altamente valioso en contextos donde la comunicación humana se ve comprometida, como ocurre en muchos casos del Trastorno del Espectro Autista.

Tal como señalan Miao et al. (2021), la IA ofrece soluciones a algunos de los grandes retos de la neurodiversidad, como la personalización del acompañamiento comunicativo, la estimulación de habilidades sociales o el fomento del aprendizaje autónomo. En entornos con discapacidad cognitiva, estas tecnologías ya han demostrado su potencial a través de la creación de espacios interactivos personalizados que ajustan la forma en que una persona recibe o emite información.

Una de las áreas con mayor desarrollo es la Comunicación Aumentativa y Alternativa (CAA), un campo centrado en facilitar la expresión de personas con dificultades de lenguaje oral mediante sistemas visuales, simbólicos o tecnológicos. En este ámbito, la revisión de Barua et al. (2022) destaca cómo la IA ha permitido diseñar herramientas asistivas más eficaces, desde interfaces con síntesis de voz hasta plataformas capaces de reconocer gestos y emociones.

La adaptabilidad es uno de los factores diferenciales más relevantes. Como explican Iannone y Giansanti (2024), la IA es capaz de analizar en tiempo real las respuestas y patrones de comportamiento de un niño con autismo, lo que permite sugerir recursos comunicativos personalizados, como pictogramas, secuencias visuales o asistentes virtuales, que se ajustan a sus capacidades y necesidades específicas. Así, la intervención deja de ser uniforme y pasa a ser genuinamente individualizada.

Además de facilitar la producción de mensajes, la IA está contribuyendo de forma significativa a mejorar la capacidad de interpretar y expresar emociones, un aspecto esencial para lograr una comunicación funcional. Esta dimensión emocional, a menudo descuidada en las herramientas convencionales, resulta especialmente relevante para los niños con TEA, ya que aprender a expresar cómo se sienten o comprender el estado emocional de los demás mejora tanto su autonomía como su bienestar psicológico.

Tecnologías emergentes como el reconocimiento facial, el análisis del tono de voz y los modelos de lenguaje generativo permiten desarrollar sistemas que acompañan al niño en su aprendizaje emocional y social, actuando como mediadores en contextos familiares, terapéuticos o comunitarios (Tang et al., 2024; Iannone & Giansanti, 2024). Gracias a estos avances, la IA puede generar respuestas empáticas, adaptar el tipo de interacción en función del estado emocional del usuario y facilitar entornos más comprensivos y respetuosos con las necesidades neurodivergentes (Cavallaro, Sica y Bloisi, 2023).

Este enfoque ha despertado un interés creciente entre investigadores, profesionales de la salud y familias, al ofrecer apoyos más sostenibles, dinámicos y adaptativos que las herramientas tradicionales. En consecuencia, la IA comienza a consolidarse no como un complemento puntual, sino como una tecnología estructuralmente integrada en los procesos de comunicación aumentativa para personas con autismo (Barua et al., 2022; Jesse, 2024).

Gracias a estos desarrollos, la inteligencia artificial ha dejado de ser una herramienta complementaria para convertirse en un recurso central en la comunicación aumentativa y alternativa (CAA). Una de las ramas más prometedoras en este campo es la inteligencia artificial generativa (IAG), que ha irrumpido con fuerza en los últimos años y permite no solo analizar datos, sino también generar contenido nuevo, como textos, imágenes, voces sintéticas o diálogos personalizados, a partir de la comprensión del contexto. Esta capacidad resulta especialmente útil para crear experiencias comunicativas más humanas, adaptadas a las necesidades de personas con trastornos del espectro autista.

Uno de los estudios más destacados es el de Tang et al. (2024), quienes desarrollaron EmoEden, una herramienta basada en IAG pensada para mejorar el aprendizaje emocional en niños con autismo de alto funcionamiento. El sistema utiliza modelos de lenguaje generativo para enseñar a identificar, interpretar y expresar emociones a través de escenarios sociales simulados, adaptándose en tiempo real a las respuestas del usuario y proporcionando un acompañamiento ajustado a su evolución.

En la misma línea, Jesse (2024) analiza la integración de IA generativa en entornos educativos neuro-inclusivos, planteando un enfoque adaptable en el que la tecnología selecciona dinámicamente los recursos comunicativos más eficaces según los resultados observados. Aunque su propuesta se enmarca en el ámbito escolar, sus conclusiones son extrapolables a otros espacios, como los entornos clínicos o familiares, donde también se requiere flexibilidad y accesibilidad para promover la comunicación.

Desde una vertiente más técnica, Hackbarth (2024) destaca el salto cualitativo que representa la IA generativa frente a los antiguos sistemas de CAA. En lugar de emplear bancos fijos de pictogramas o frases pregrabadas, estos sistemas pueden generar mensajes personalizados según la situación, interpretar emociones en una imagen, o construir un discurso visual adaptado a la intención comunicativa del usuario. Esta personalización contribuye a que el niño se exprese de manera más espontánea y significativa.

Paralelamente, autores como Johnson, Smart y Mahar (s.f.) plantean una mirada crítica, señalando que aunque la IA generativa tiene un gran potencial para mejorar la comunicación en contextos de diversidad funcional, también presenta retos éticos y técnicos. Entre ellos destacan el riesgo de dependencia tecnológica, la escasa regulación

del sector y la importancia de implicar activamente a cuidadores y familias en su implementación.

En cuanto a revisiones amplias, Barua et al. (2022) ofrecen una panorámica del uso de tecnologías generativas en herramientas diseñadas para mejorar la comunicación de niños con TEA. Subrayan cómo estos sistemas permiten desarrollar soluciones más accesibles y escalables, especialmente valiosas para quienes presentan un lenguaje verbal limitado. A su vez, Cavallaro, Sica y Bloisi (2023) insisten en la versatilidad de estos modelos, capaces de ajustarse en tiempo real al perfil comunicativo del niño y al contexto emocional, incorporando incluso rasgos de empatía artificial.

Finalmente, cabe destacar el trabajo de Maldonado Gilarranz (2024), cuyo Trabajo de Fin de Grado propone una aplicación práctica del modelo DALL·E 3, a través de ChatGPT, para generar pictogramas personalizados mediante descripciones textuales. Su objetivo es crear imágenes que sean instantáneas, accesibles y adaptadas al usuario, eliminando la necesidad de búsquedas visuales tradicionales. Este enfoque abre la puerta a nuevos métodos de intervención comunicativa más eficientes y centrados en la experiencia del niño, anticipando posibles líneas de investigación como la que se propone en el presente trabajo.

Tabla 2.1. Principales aportaciones científicas sobre autismo, comunicación aumentativa e inteligencia artificial generativa

Artículo	Conclusiones clave
Kanner (1943)	Definió por primera vez el concepto de "autismo infantil precoz" observando patrones de comportamiento repetitivo y dificultades sociales.
Asperger (1944)	Identificó perfiles similares con habilidades cognitivas altas pero escasa comunicación emocional; acuñó el término "pequeños profesores".
Rojas-Torres et al. (2020)	Demostraron que intervenciones personalizadas mejoran la calidad de vida de niños con TEA, especialmente en términos comunicativos.
Waizbard-Bartov et al. (2021)	Subrayan las dificultades en la comunicación verbal y no verbal como uno de los principales retos en TEA.
Pino y Guerrero (2021)	Señalan que muchos profesionales carecen de herramientas y formación adecuada para emplear tecnología adaptada en el tratamiento del TEA.
Moraiti et al. (2023)	Revisaron el uso de tecnología digital como tablets o apps móviles en el apoyo a la comunicación en TEA, destacando beneficios y limitaciones.
Barua et al. (2022)	Evidenciaron el potencial de la IA para crear herramientas de CAA adaptadas, escalables y accesibles.

Tabla 2.1. Principales aportaciones científicas sobre autismo, comunicación aumentativa e inteligencia artificial generativa

Artículo	Conclusiones clave
Iannone y Giansanti (2024)	Propusieron que la IA puede analizar patrones de comportamiento y sugerir pictogramas o asistentes personalizados.
Cavallaro et al. (2023)	Afirmaron que la IA con empatía artificial puede ajustarse al estado emocional del usuario, mejorando la calidad de la interacción.
Tang et al. (2024)	Desarrollaron EmoEden, una herramienta de IA generativa que mejora el aprendizaje emocional en niños con autismo de alto funcionamiento.
Jesse (2024)	Analizó el uso de IA generativa en entornos educativos neuro-inclusivos, aplicable también a contextos clínicos y familiares.
Hackbarth (2024)	Destacó cómo la IAG permite generar mensajes visuales más personalizados y flexibles que los sistemas tradicionales de CAA.
Maldonado Gilarranz (2024)	Propuso el uso de DALL·E 3 para generar pictogramas personalizados, demostrando la aplicabilidad directa de la IAG en el diseño de apoyos visuales.
Johnson et al. (s.f.)	Advirtieron sobre los riesgos éticos, dependencia tecnológica y la necesidad de regular el uso de IAG en diversidad funcional.

3. Definición del Trabajo

Este capítulo establece el marco y alcance del trabajo, centrado en el diseño y validación de un experimento que evalúa la capacidad de dos modelos de inteligencia artificial generativa de imagen para transformar imágenes reales de acciones cotidianas en pictogramas comprensibles por niños con autismo severo, siguiendo los criterios del sistema ARASAAC. Se detalla el objetivo general, los objetivos específicos, las limitaciones del estudio, la hipótesis a contrastar y las principales suposiciones que sustentan la investigación.

3.1 Objetivo General

El objetivo principal de este trabajo es comparar el rendimiento de dos modelos de inteligencia artificial generativa: uno estadounidense (GPT de OpenAI) y otro chino (Qwen de Alibaba) para generar pictogramas comprensibles a partir de fotografías reales de acciones cotidianas. La finalidad última es comprobar si estas herramientas pueden ser útiles para facilitar la accesibilidad cognitiva y la comunicación aumentativa en contextos educativos o terapéuticos para niños con autismo severo.

3.2 Objetivos Específicos

- Identificar y clasificar 25 acciones cotidianas relevantes para la vida diaria de niños con autismo severo.
- Diseñar un prompt persistente y estructurado que pueda aplicarse equitativamente a ambos modelos generativos.
- Ejecutar el proceso de generación de pictogramas mediante ambos modelos (DALL·E 3 y Qwen-VL), obteniendo una imagen representativa por cada acción base.
- Establecer una rúbrica de evaluación con criterios visuales inspirados en los principios del sistema ARASAAC: claridad, iconicidad, simplicidad, relevancia y contexto.
- Evaluar sistemáticamente las imágenes generadas por cada modelo aplicando dicha rúbrica, y determinar el porcentaje de imágenes consideradas válidas.

- Realizar una prueba estadística (z-test de diferencia de proporciones) para evaluar el rendimiento de cada modelo y contrastar si existen diferencias significativas entre el rendimiento de ambos modelos.

3.3 Limitaciones

- **Limitación temporal:** El experimento debe completarse dentro del curso académico 2024–2025, lo que limita la posibilidad de ampliar la muestra de imágenes o incorporar más modelos.
- **Limitación técnica:** El experimento se ha realizado desde plataformas accesibles al público general, por lo que no se ha podido ajustar directamente la arquitectura interna de los modelos ni aplicar fine-tuning.
- **Restricción de uso:** Las plataformas utilizadas imponen límites al número de imágenes que pueden generarse en un determinado periodo de tiempo. Superado ese umbral, es necesario esperar o reiniciar sesión, lo que afecta a la fluidez del proceso experimental.

3.3.1 Hipótesis

Este trabajo plantea dos contrastes estadísticos con el fin de evaluar tanto la validez general del enfoque como las diferencias de rendimiento entre los modelos analizados.

Primer contraste (Hipótesis principal)

- **Hipótesis nula (H_0):** Ninguno de los modelos supera significativamente el 50 % de imágenes válidas generadas.
- **Hipótesis alternativa (H_1):** Al menos un modelo supera significativamente el 50 % de imágenes válidas generadas.

Este primer contraste permite determinar si los modelos de inteligencia artificial evaluados presentan un desempeño significativamente superior al azar en la tarea de generar pictogramas funcionales.

Segundo contraste (Hipótesis secundaria)

- **Hipótesis nula (H_0):** No existen diferencias significativas entre los dos modelos en cuanto a la proporción de imágenes válidas generadas.

- **Hipótesis alternativa (H_1):** Existe una diferencia significativa entre ambos modelos en cuanto a la proporción de imágenes válidas generadas.

Este segundo contraste tiene como objetivo analizar si uno de los modelos presenta una capacidad significativamente superior al otro en la tarea evaluada.

3.3.2 Asunciones

- **Aplicabilidad general:** Se asume que los resultados obtenidos con estas 25 acciones pueden extrapolarse a otras acciones similares en contextos educativos.
- **Supuesto de validez visual:** Se asume que los criterios derivados definidos en la rúbica son adecuados para medir la claridad de un pictograma destinado a niños con autismo severo.
- **Supuesto de mediación comunicativa:** Se da por sentado que los pictogramas generados serán utilizados con acompañamiento visual o verbal, y no como única fuente de información para el niño.

4. Marco Teórico

4.1 Introducción a la Inteligencia Artificial Generativa

En el contexto actual de la inteligencia artificial (IA), uno de los avances más transformadores ha sido el desarrollo de modelos capaces de generar contenido original a partir de datos: imágenes, texto, audio o vídeo. Este tipo de modelos, conocidos como modelos generativos, han adquirido especial protagonismo por su capacidad para transformar descripciones textuales o imágenes de entrada en composiciones visuales coherentes y detalladas (Alto, 2023, p. 15).

Para situar correctamente este avance dentro del campo de la IA, resulta útil entender su jerarquía conceptual (véase Figura 1). En ella se observa cómo los modelos generativos forman parte del subcampo de la IA generativa, que a su vez se inscribe dentro del Deep laerning, una subdisciplina del Machine Learning, y este a su vez dentro de la IA en general (Alto, 2023, p. 19).

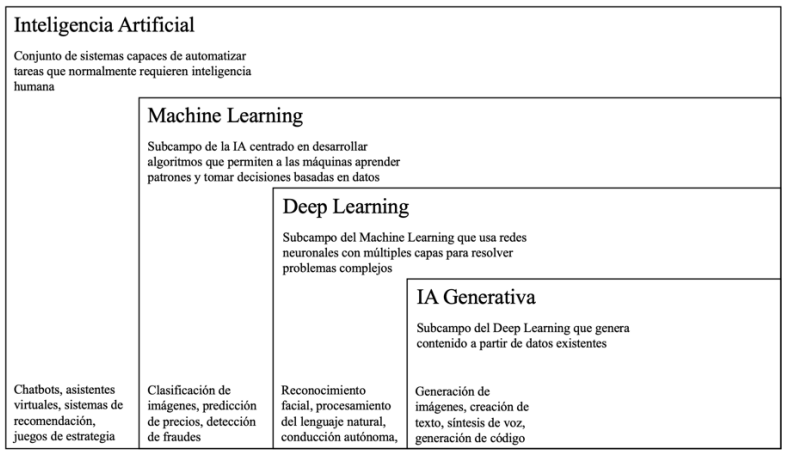

Figura 4.1: Jerarquía conceptual de la Inteligencia Artificial Elaboración propia.

4.1.1 Redes Neuronales y Evolución Arquitectónica

La base técnica de los modelos generativos actuales es, por tanto, el Deep Learning, que utiliza redes neuronales compuestas por múltiples capas conectadas entre sí. Estas redes se inspiran en el funcionamiento del cerebro humano y procesan la información mediante transformaciones no lineales. A medida que los datos atraviesan las distintas capas, la red aprende representaciones cada vez más abstractas de los patrones presentes en los datos de entrada, lo que permite abordar tareas de alta complejidad como la traducción automática, el reconocimiento de voz o la generación de imágenes (LeCun, Bengio y Hinton, 2015).

A lo largo de su evolución, las redes neuronales han adoptado arquitecturas específicas en función del tipo de dato que deben procesar:

- **Las Redes Neuronales Convolucionales (CNN)**, han sido fundamentales para el procesamiento de imágenes. Estas redes utilizan filtros que se deslizan sobre la imagen para extraer características locales como bordes, texturas y formas. Gracias a su capacidad para capturar patrones espaciales, las CNN han sido ampliamente empleadas en tareas como clasificación de imágenes, segmentación semántica y detección de objetos (Ortega Candel, 2025, pp. 103–107).

- **Las Redes Neuronales Recurrentes (RNN)**, por su parte, fueron diseñadas para manejar datos secuenciales, como texto, audio o series temporales. Estas redes incorporan conexiones que permiten conservar información a lo largo del tiempo, lo que resulta útil para modelar dependencias contextuales. No obstante, presentan dificultades al procesar secuencias largas debido a problemas como el desvanecimiento del gradiente (Ortega Candel, 2025, pp. 83–87).

- **Las Redes Generativas Antagónicas (GAN)** marcaron un hito al introducir una arquitectura compuesta por dos redes en competencia: un generador que crea imágenes sintéticas a partir de ruido aleatorio, y un discriminador que evalúa si esas imágenes son reales o falsas. Este enfoque permitió generar imágenes de notable realismo, pero presentaba desafíos importantes como la inestabilidad del entrenamiento, la sensibilidad a hiperparámetros y la dificultad para controlar el contenido generado (Ortega Candel, 2025, pp. 149–177).

4.1.2 Modelos de Difusión y Transformers

Debido a estas limitaciones, surgieron nuevas alternativas que ofrecieran mayor estabilidad, control semántico y calidad visual. Entre ellas destacan los modelos de difusión, una familia de modelos generativos que han demostrado excelentes resultados en tareas de generación de imágenes. Estos modelos aprenden a revertir un proceso progresivo de adición de ruido a los datos, generando imágenes coherentes a partir de ruido gaussiano inicial mediante un proceso iterativo de eliminación de ruido. La generación se realiza en el espacio latente o directamente sobre los píxeles, dependiendo del modelo, y se entrena utilizando funciones de pérdida como la predicción del ruido (Ho et al., 2020)

En paralelo, el modelo Transformer, propuesto por Vaswani et al. (2017), revolucionó el campo del procesamiento de lenguaje natural al introducir el mecanismo de autoatención, que permite identificar qué partes de una secuencia son más relevantes para predecir el siguiente elemento. A diferencia de las RNN, los Transformers permiten paralelizar el procesamiento y modelar dependencias a largo plazo de manera más eficiente. Esta arquitectura fue adoptada con rapidez en tareas de texto, y posteriormente adaptada al dominio visual. (Ortega Candel, 2025, pp. 171–180).

En el ámbito de las imágenes, los Transformers se han combinado con modelos como U-Net y Vision Transformer (ViT). El primero, U-Net, es una arquitectura simétrica de codificación-decodificación ampliamente utilizada en tareas de segmentación de imágenes médicas, que permite preservar la resolución espacial mientras se aprende una representación abstracta (Ronneberger, Fischer y Brox, 2015). Por su parte, ViT adapta la estructura del Transformer para procesar imágenes dividiéndolas en parches que se tratan como si fueran tokens de texto, lo que permite aplicar mecanismos de atención directamente sobre representaciones visuales (Dosovitskiy et al., 2021).

La combinación de Transformers con modelos de difusión ha dado lugar a arquitecturas multimodales capaces de transformar texto en imágenes de alta calidad. Un componente clave en esta combinación es CLIP (Contrastive Language–Image Pretraining), un modelo desarrollado por OpenAI que entrena conjuntamente un codificador de texto y otro de imagen para que ambos representen conceptos en un espacio latente compartido.

Esto permite que un sistema de generación de imágenes "entienda" el significado semántico del texto y lo transforme en contenido visual coherente (Brück, 2024).

En conjunto, estas arquitecturas; Transformers, modelos de difusión, U-Net, ViT y CLIP, han convergido en sistemas generativos de última generación, como DALL·E 3 o Qwen, que integran múltiples módulos especializados para ofrecer resultados visuales precisos, ricos en detalle y semánticamente alineados con el prompt proporcionado por el usuario.

4.2 Generación de imágenes mediante IA

Tras haber revisado las principales arquitecturas utilizadas en los modelos generativos, es necesario comprender cómo se inicia el proceso de generación desde el punto de vista del usuario. En este contexto, el concepto de prompt adquiere un papel central.

Entre las múltiples aplicaciones de la inteligencia artificial generativa, la generación de imágenes a partir de texto (text-to-image) se ha consolidado como una de las más relevantes. Esta técnica permite transformar una descripción escrita, conocida como prompt, en una imagen coherente, capaz de representar conceptos semánticos con alto nivel de fidelidad visual (Podell et al., 2023).

Además, el prompt no tiene por qué limitarse a entradas textuales. Existen modelos que aceptan imágenes como entrada inicial, en lo que se conoce como image-to-image translation, así como sistemas multimodales que combinan texto e imagen en un mismo prompt. Esta flexibilidad ha dado lugar a modelos capaces de operar en contextos más complejos, donde se requiere integrar múltiples tipos de información de forma simultánea (Balaji, Bansal y Efros, 2023).

El proceso de generación de imágenes puede descomponerse en tres fases principales:

1. Codificación del prompt.
2. Generación en el espacio latente.
3. Decodificación visual.

4.2.1 Codificación del Promt

Tanto si el prompt es textual como si es visual, el sistema debe traducirlo a un espacio latente matemático, que actúa como un punto intermedio donde se realiza la generación. Para ello, se utilizan encoders (codificadores), que transforman el mensaje original en un vector numérico que resume su contenido semántico. Este vector condensa el significado del prompt de forma que pueda ser procesado por el modelo generativo. (Ma et al., 2024).

Este proceso se representa matemáticamente mediante:

$$z_{\text{prompt}} = f_{\text{encoder}}(x) \tag{4.1}$$

donde (x) es el prompt (texto o imagen), y z_{prompt} es su representación latente, obtenida mediante una red neuronal o Transformer. Esta codificación permite al modelo comprender qué quiere el usuario y empezar a construir una imagen coherente a partir de ese significado (Ma et al., 2024).

4.2.2 Generación Latente mediante Difusión

A continuación, el modelo genera una imagen en el espacio latente utilizando un modelo de difusión, que parte de ruido gaussiano y lo refina progresivamente hasta obtener una representación coherente con el contenido semántico del prompt. Este procedimiento se basa en un proceso iterativo de eliminación de ruido (*denoising*) modelado como una cadena de Markov inversa (Ho, Jain y Abbeel, 2020).

El objetivo es aprender la distribución condicional:

$$p\left(x_0 \mid z_{\text{prompt}}\right) \tag{4.2}$$

Donde x_0 es la imagen generada final, y z_{prompt} la codificación del promt. Durante el entrenamiento, el modelo se optimiza para predecir el ruido que fue añadido a una imagen real en el paso tt del proceso de noising. La función de pérdida de difusión es:

$$\mathcal{L} = E_{x,\, \epsilon\sim\mathbb{N}(\mathbb{0},I)}\left[|\epsilon - \epsilon_\theta\left(x_t, t, z_{\text{prompt}}\right)|^2\right] \tag{4.3}$$

Donde ϵ es el ruido aplicado, ϵ_θ es la predicción del modelo sobre dicho ruido, condicionada por el paso t y el prompt codificado.

4.2.2 Decodificación Visual

Finalmente, la representación latente obtenida se convierte en una imagen RGB perceptible mediante un decodificador entrenado. En los modelos más actuales, este decodificador es en sí un modelo de difusión, que traduce iterativamente los vectores latentes en imágenes. En otros modelos más antiguos, se utilizaban decodificadores basados en autoencoders convolucionales, que mapeaban directamente el espacio latente al espacio de píxeles.

Ambos enfoques comparten el objetivo de transformar una representación abstracta en una imagen final coherente y visualmente rica (Podell et al., 2023).

4.3 Análisis de Modelos Generativos

En este trabajo se analizarán en detalle los dos modelos generativos seleccionados para el estudio: DALL·E 3 (OpenAI) y Qwen2.5-Max (Alibaba). Ambos permiten generar imágenes a partir de texto, pero siguen enfoques arquitectónicos distintos: uno centrado en el uso de modelos lingüísticos avanzados, y otro con orientación multimodal y multilingüe. A continuación, se presentan sus principales características técnicas.

4.3.1 DALL·E 3 (OpenAI)

DALL·E 3 es la tercera generación del modelo de generación de imágenes de OpenAI. Se diferencia de sus versiones previas por su capacidad para interpretar prompts complejos y por su integración profunda con el modelo lingüístico GPT-4 (OpenAI, 2023).

El pipeline de DALL·E 3 integra tres módulos esenciales:

- **Codificador semántico** (CLIP-like): Aunque no forma parte directa del generador, el sistema CLIP (*Contrastive Language–Image Pretraining*) es fundamental en DALL·E 3. Este modelo, desarrollado por Radford et al. (2021), alinea representaciones textuales e imágenes en un espacio latente compartido

mediante aprendizaje contrastivo. En DALL·E, se utiliza como guía semántica y mecanismo de validación para asegurar que la imagen generada responde al contenido del prompt (OpenAI, 2023). Esta lógica se muestra en la Figura 2, que representa la arquitectura base de DALL·E 2, aún vigente en DALL·E 3 (Ramesh et al., 2022).

- **Transformer autoregresivo:** Una vez codificado el prompt, se genera una secuencia de tokens visuales latentesmediante un transformer autoregresivo inspirado en GPT. Estos tokens, que representan unidades mínimas de información visual, se generan uno a uno condicionados por el embedding textual del prompt, que actúa como representación semántica condensada (OpenAI, 2023).

- **Decodificador por difusión:** Por último, un modelo de difusión transforma los tokens en una imagen final, partiendo de ruido y eliminándolo progresivamente en función del prompt. Esta técnica mejora la coherencia visual y semántica frente a arquitecturas previas como VQ-VAE, y permite un mayor control sobre el detalle visual (Ho et al., 2020; Ramesh et al., 2022; OpenAI, 2023).

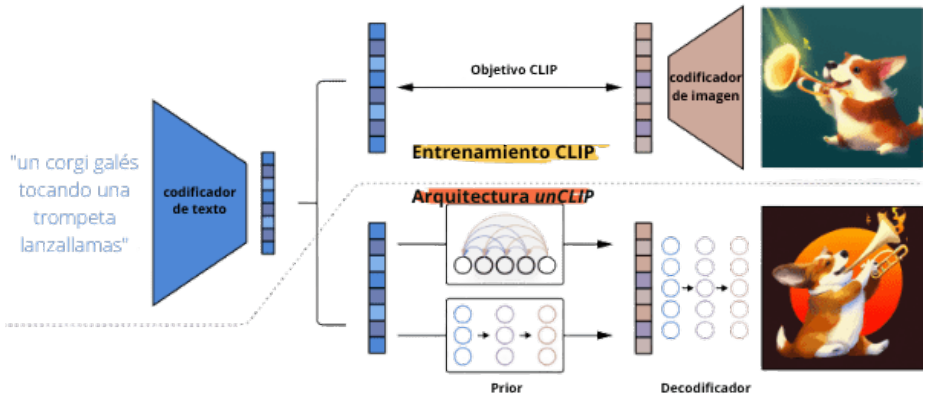

Figura 4.2: Esquema del entrenamiento CLIP y arquitectura unCLIP. Página Web OpenAI.

4.3.2 Qwen (Alibaba)

Aunque el ecosistema Qwen no cuenta con un modelo específico llamado “Qwen Image”, sí incorpora capacidades de generación de imágenes a través de arquitecturas multimodales como Qwen-VL, desarrolladas por Alibaba DAMO Academy (2023). Estas capacidades se integran dentro de modelos diseñados para comprender y generar texto e imagen en múltiples idiomas, especialmente chino e inglés.

La generación visual se basa en un modelo de difusión adaptado al contexto multilingüe. Para codificar el prompt, Qwen utiliza encoders similares a CLIP, que traducen el texto en vectores latentes optimizados mediante cross-attention. Estos vectores se procesan mediante transformers multimodales entrenados de forma conjunta sobre datos visuales y lingüísticos, lo que permite una mayor generalización semántica frente a arquitecturas más secuenciales como DALL·E 3 (Qwen Team, 2023).

La imagen final se genera mediante un decodificador por difusión iterativa, condicionado por el prompt textual. Este proceso es conceptualmente similar al de OpenAI, aunque Alibaba introduce mecanismos propios como tokenización jerárquica y codificación semántica guiada. Sin embargo, parte de los detalles técnicos permanecen sin divulgar públicamente (Alibaba DAMO Academy, 2023).

5. Metodología

Esta sección se presenta la metodología seguida en el estudio. Se describe paso a paso cómo se ha llevado a cabo el experimento, desde la selección y preprocesamiento de las imágenes hasta el diseño del entorno de evaluación y la aplicación de los modelos generativos ChatGPT y Qwen. También se detalla el sistema de evaluación empleado para analizar la calidad y adecuación de las imágenes generadas.

5.1 Descripción de la Metodología

Este trabajo tiene como objetivo principal evaluar la capacidad de dos modelos de inteligencia artificial generativa de imagen, representativos de las potencias tecnológicas de Estados Unidos y China, para transformar imágenes reales de acciones cotidianas en pictogramas comprensibles por niños con autismo severo, con el fin de facilitar la comunicación aumentativa y accesibilidad cognitiva.

Para llevar a cabo esta comparación, se han seleccionado dos modelos multimodales avanzados:

- ChatGPT (GPT-4o integrado con DALL·E 3) - desarrollado por OpenAI (Estados Unidos).
- Qwen2.5-Max - desarrollado por Alibaba DAMO Academy (China).

Se han seleccionado estos dos modelos específicos por dos razones principales:

- Ambos representan las dos mayores potencias tecnológicas actuales en el campo de la IA generativa (EE.UU. y China).

- Ambos modelos ofrecen capacidades multimodales completas, es decir, permiten introducir como entrada texto, texto e imagen o únicamente una imagen y generar una nueva imagen como salida. A diferencia de otros modelos actuales en el mercado, que dependen únicamente del texto para generar la imagen

La generación de imágenes se guiará por los criterios visuales del sistema ARASAAC (Portal Aragonés de la Comunicación Aumentativa y Alternativa), un repositorio de pictogramas de uso libre ampliamente validado y utilizado en entornos educativos y

clínicos con personas con Trastorno del Espectro Autista (TEA).Estos pictogramas se caracterizan por su alta transparencia semántica, simplicidad gráfica, ausencia de detalles irrelevantes y uso de líneas negras gruesas con colores planos, lo que los hace especialmente eficaces para facilitar la comprensión de conceptos, rutinas y acciones por parte de personas con autismo severo (Gobierno de Aragón, s.f.).

Según Cabello & Bertola (2015), los pictogramas de ARASAAC presentan "una estructura gráfica coherente, universal, y cognitivamente accesible", convirtiéndolos en el estándar ideal para evaluar si una imagen generada por IA es adecuada para su uso en comunicación aumentativa.

5.2 Generación del Prompt

Para garantizar la consistencia en la tarea de generación visual, se diseñó un prompt maestro persistente, que fue utilizado como instrucción base tanto para el modelo de OpenAI (DALL·E 3) como para el modelo chino (Qwen2.5-Max). Este prompt tenía como finalidad orientar a la IA en la transformación de una imagen compleja en un pictograma claro y comprensible, siguiendo el estilo gráfico de ARASAAC.

El diseño del prompt se basó en un marco propuesto por Greg Brockman, presidente de OpenAI, y atribuido internamente a Ben Hylak, un colaborador de la compañía especializado en diseño de interacción y experiencia de usuario. Si bien Hylak no ha publicado formalmente en revistas científicas, su marco fue presentado por Brockman en el evento OpenAI DevDay de 2024, y ha sido adoptado ampliamente en entornos prácticos por su utilidad a la hora de estructurar instrucciones para modelos generativos multimodales.

Este sistema, aplicable tanto a modelos de OpenAI como a otras IAs avanzadas, se compone de cuatro elementos clave para la construcción de prompts eficaces:

- **Objetivo:** qué se espera que la IA genere.
- **Formato de salida:** cómo debe estructurarse la imagen.
- **Advertencias:** qué debe evitarse expresamente.
- **Resumen de contexto:** para afinar el comportamiento del modelo.

Aunque fue desarrollado originalmente en el entorno de OpenAI, este marco se aplicó por igual a ambos modelos utilizados en este experimento, con el fin de mantener igualdad

de condiciones metodológicas. Esto permite evaluar la calidad de las imágenes generadas de forma justa, sin que la formulación del prompt beneficie de manera sesgada a uno de los modelos.

Prompt utilizado:

Objetivo:
A partir de una imagen que representa *"la acción cotidiana X"*, genera una imagen estilo pictograma tipo ARASAAC, clara y comprensible para niños con autismo severo.

Formato de salida:

- Un único personaje central (si es posible).
- Fondo blanco o transparente.
- Estilo plano 2D.
- Líneas negras gruesas y colores planos.
- Sin texto ni etiquetas.
- Formato pictograma ARASAAC.

Advertencias:
No incluir fondo decorativo, expresiones ambiguas, texto ni más personajes de los necesarios. Si no es posible generar la imagen bajo estas condiciones, indícalo.

Contexto:
Este prompt se emplea en el marco de un experimento académico con el objetivo de evaluar si modelos de IA generativa pueden transformar imágenes reales en pictogramas comprensibles, siguiendo criterios visuales derivados del sistema ARASAAC, para su uso en entornos educativos y terapéuticos con niños con autismo severo.

Antes de introducir el prompt definitivo, se estableció en la conversación una definición clara de lo que se entiende por pictograma en el estilo ARASAAC. Asimismo, se aportaron al modelo imágenes de ejemplo junto con la descripción de la salida esperada, con el objetivo de contextualizar la tarea y facilitar que el modelo generativo comprendiera el tipo de representación visual requerida. Este paso previo permitió orientar al sistema hacia la estética, la simplicidad y la funcionalidad propias de los pictogramas ARASAAC, actuando como una forma de preentrenamiento informal dentro del entorno conversacional.

5.3 Selección y clasificación de imágenes

Para este estudio se han seleccionado 25 imágenes, constituyendo una muestra de n = 25. La elección de estas imágenes se ha basado en criterios funcionales y pedagógicos, tomando como referencia estudios clínicos y educativos centrados en las necesidades comunicativas de personas con Trastorno del Espectro Autista (TEA), especialmente en casos de autismo severo. El objetivo principal de esta selección es representar acciones cotidianas que favorezcan la anticipación de rutinas, el desarrollo del autocuidado y la comprensión del entorno social y escolar.

En primer lugar, se ha tomado como base el trabajo de Saénz y Juárez (2016), que identifica las principales dificultades de comprensión que enfrentan las personas con TEA en la vida diaria. El autor pone énfasis en la necesidad de representar visualmente acciones básicas, con el fin de facilitar la comprensión de normas sociales y escolares. De dicho estudio se ha extraído una clasificación general por áreas funcionales, que incluye categorías como autocuidado, alimentación, juego o interacción social.

Asimismo, el estudio de Fernández y Bandrés (2023) ofrece una propuesta didáctica concreta basada en rutinas diarias y situaciones de aula, enmarcada en el Diseño Universal para el Aprendizaje (DUA). Este enfoque proporciona ejemplos de acciones reales que suelen resultar especialmente difíciles de comprender para el alumnado con TEA, como “esperar el turno”, “guardar el material” o “esperar un semáfoto”. Estas acciones han sido incorporadas directamente a la selección de imágenes del presente estudio.

A continuación, se presenta una tabla con las 25 acciones seleccionadas, organizadas por áreas funcionales. Esta clasificación facilita el análisis posterior y garantiza la cobertura de situaciones relevantes en el desarrollo de habilidades comunicativas y sociales en personas con autismo.

Tabla 5.1: Clasificación funcional de las 25 acciones seleccionadas

Área	Acción
Autocuidado	1. Lavarse las manos
	2. Cepillarse los dientes
	3. Sentarse en el váter
	4. Peinarse
	5. Bañarse
	6. Abrocharse los cordones de los zapatos
	7. Tomar temperatura con termómetro
Alimentación	8. Comer con cuchara
	9. Beber agua con un vaso
	10. Poner la mesa
Escolar	11. Sentarse en círculo
	12. Levantar la mano para hablar
	13. Guardar los materiales
Juegos	14. Niños jugando a las cartas
	15. Pintar con lápices de colores
	16. Tirar una pelota a otro niño
	17. Soplar las velas de una tarta
Normas sociales	18. Cruzar en semáforo en verde
	19. Esperar semáforo en rojo
	20. Esperar turno en la fila
	21. Decir adiós
Participación comunitaria	22. Subir a un autobús
	23. Pagar en la caja del supermercado
	24. Pasear un perro
	25. Sacar la basura

5.4 Criterios de Evaluación

Para valorar la idoneidad de las imágenes generadas por los modelos de inteligencia artificial en el contexto de la comunicación aumentativa con niños con Trastorno del Espectro Autista (TEA) severo, se ha definido una rúbrica de evaluación multidimensional basada en parámetros visuales, semánticos y funcionales. Esta rúbrica permite establecer una comparación objetiva entre ambos modelos, evaluando cada imagen generada según cinco dimensiones clave. La versión completa de la rúbrica puede consultarse en el Anexo 1.

Las cinco dimensiones consideradas en la rúbrica son las siguientes:

- **Comprensibilidad pictográfica**: analiza la capacidad de la imagen para transmitir de forma directa y accesible el mensaje o la acción representada, atendiendo a los principios de claridad, iconicidad y simplicidad visual propios de los pictogramas adaptados para entornos educativos o terapéuticos.

- **Fidelidad semántica al prompt**: examina el grado en que la imagen generada refleja fielmente la instrucción textual que le dio origen, asegurando coherencia en el contenido representado y presencia de los elementos esperados.

- **Calidad visual técnica:** valora aspectos técnicos de la imagen como resolución, nitidez, ausencia de errores gráficos o deformaciones, y correcta composición visual, sin entrar en juicios estéticos.

- **Adecuación para el usuario objetivo:** mide si la imagen es apropiada para niños con autismo severo, considerando variables como la carga visual, los colores utilizados o la familiaridad de los elementos con el entorno cultural del usuario.

- **Consistencia entre imágenes:** evalúa la coherencia visual entre distintas imágenes generadas por el mismo modelo, especialmente en cuanto a estilo gráfico, escalas y paleta cromática, lo cual resulta fundamental cuando se pretende construir un conjunto homogéneo de pictogramas.

Cada imagen se valorará individualmente en estas cinco dimensiones, y cada una de ellas está compuesta por varios indicadores específicos. La escala de puntuación aplicada a cada indicador es la siguiente:

0 = Inadecuado

1 = Insuficiente

2 = Aceptable

3 = Bueno

4 = Excelente

A partir de estas puntuaciones, se calcula una puntuación final ponderada para cada imagen, aplicando un peso distinto a cada dimensión según su relevancia funcional en la interpretación visual de personas con autismo severo. Esta puntuación final (entre 0 y 4) puede utilizarse de forma individual para analizar el rendimiento de cada imagen, o bien calcularse un promedio por modelo (DALL·E 3 y Qwen) con el fin de realizar una comparativa global de desempeño.

El cálculo de la puntuación ponderada de cada foto y la puntación final de cada modelo se realiza mediante la siguiente fórmula de media ponderada:

Puntuación Final = (0,30 × C) + (0,20 × F) + (0,15 × Q) + (0,20 × A) + (0,15 × S)

Donde:

C = Comprensibilidad pictográfica

F = Fidelidad semántica

Q = Calidad técnica

A = Adecuación al usuario

S = Consistencia

6. Análisis de los resultados

Una vez generadas las imágenes y puntuadas según la rúbrica descrita anteriormente, en este apartado se analiza si se cumple el objetivo y los resultados previstos. El objetivo principal era comprobar si dichas imágenes podían considerarse pictogramas válidos para niños con autismo severo, siguiendo criterios derivados del sistema ARASAAC.

Para determinar la validez de cada imagen, se consideró válida si su puntuación media era igual o superior a 2, valor que corresponde con el nivel "Aceptable". Los resultados fueron los siguientes:

- DALL·E 3 generó una media de 2,95 puntos como media de todas las imágenes y 23 de las 25 imágenes (92 %) superaron el umbral de validez.
- Qwen, en cambio, obtuvo una media de 1,32 puntos, con tan solo 6 de las 25 imágenes (24 %) consideradas válidas.

A priori, se observa una media notablemente más alta y una mayor precisión en DALL·E 3 en comparación con Qwen, lo que evidencia una diferencia clara en el rendimiento de ambos modelos. Para verificar si esta diferencia es estadísticamente significativa, se recurre a un contraste formal mediante un test de proporciones.

Para comprobar si DALL·E 3 supera significativamente el umbral mínimo de utilidad (fijado en el 60 %), se aplicó un contraste estadístico de una proporción (z-test unilateral).

La hipótesis nula plantea que DALL·E no alcanza ese umbral de validez, mientras que la alternativa plantea que sí lo supera de forma significativa. Con 23 aciertos sobre 25, se obtiene una proporción muestral $\hat{p} = \frac{23}{25} = 0{,}92$, frente a una proporción de referencia $p_0 = 0{,}60$. Aplicando la fórmula:

$$z = \frac{0{,}92 - 0{,}60}{\sqrt{0{,}60 \cdot (1 - 0{,}60)/25}} \approx \frac{0{,}32}{0{,}098} \approx 3{,}27$$

El valor-z obtenido es 3,27, lo que produce un p-valor ≈ 0,0005. Dado que este valor es significativamente inferior al umbral de significación convencional (α = 0,05). Se rechaza la hipótesis nula. En consecuencia, se concluye que DALL·E 3 supera de forma

estadísticamente significativa el 60 % de imágenes válidas, lo que respalda su aplicabilidad como herramienta de apoyo en contextos educativos o terapéuticos con niños con autismo severo.

En el caso de Qwen, con solo 6 imágenes válidas de 25, no se ha considerado necesario realizar un contraste formal, ya que la proporción $\hat{p} = \frac{5}{25} = 0{,}24$ está muy por debajo del umbral establecido, y la media global de puntuación (1,32) respalda esta baja eficacia. Estos resultados ponen en duda la utilidad del modelo chino Qwen en tareas donde la claridad visual y la simplicidad comunicativa son fundamentales.

Finalmente, para comprobar si las diferencias entre ambos modelos son estadísticamente significativas en cuanto a su capacidad para generar pictogramas válidos, se aplicó un contraste de dos proporciones independientes. DALL·E 3 alcanzó una proporción de aciertos del 92 % (23/25), mientras que Qwen obtuvo únicamente un 24 % (6/25). Con ambos tamaños muestrales iguales (n = 25), el estadístico z se calcula de la siguiente forma:

$$z = \frac{0{,}92 - 0{,}24}{\sqrt{\frac{0{,}92 \cdot (1 - 0{,}92)}{25} + \frac{0{,}24 \cdot (1 - 0{,}24)}{25}}} \approx \frac{0{,}68}{0{,}145} \approx 4{,}69$$

El valor obtenido ($z \approx 4{,}69$) corresponde a un p-valor aproximado de 0,000001, lo cual indica una diferencia altamente significativa entre ambos modelos. Esto confirma que DALL·E 3 supera de forma contundente a Qwen en esta tarea, no solo en cantidad de aciertos, sino también en consistencia global. Desde una perspectiva aplicada, estos resultados respaldan que DALL·E 3 es una opción más eficaz y funcional para la generación automatizada de pictogramas, especialmente en contextos educativos y terapéuticos dirigidos a mejorar la accesibilidad cognitiva en niños con autismo severo.

Para ilustrar gráficamente el tipo de resultados obtenidos por cada modelo, se han incorporado en esta sección algunas de las imágenes generadas por los modelos.

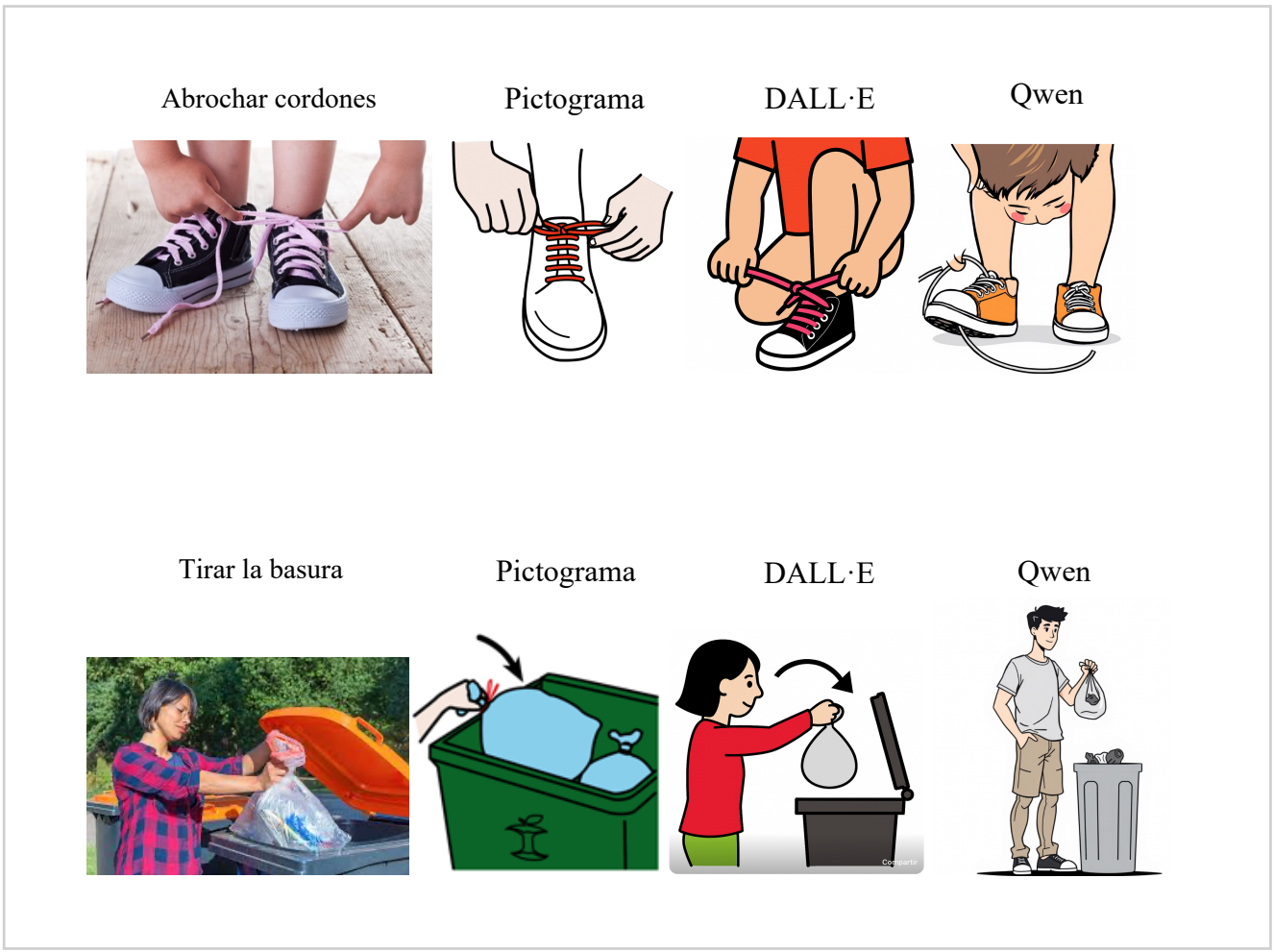

Figura 6.1: Comparativa de pictogramas generados. Elaboración propia.

Uno de los aspectos donde DALL·E 3 destaca de manera más clara es la consistencia visual entre imágenes. El modelo tiende a mantener una línea gráfica coherente, con concordancia en formas, proporciones, paleta cromática y estilo de representación corporal. Esta uniformidad facilita que los pictogramas generados puedan ser utilizados como parte de un sistema homogéneo, tal como requiere una herramienta educativa orientada a niños con TEA.

En contraste, Qwen2.5-Max presenta una variabilidad muy alta en su producción visual. Algunas imágenes adoptan un estilo realista, mientras que otras se aproximan a ilustraciones tipo cómic o boceto. Esta falta de cohesión se traduce en una menor comprensibilidad pictográfica, ya que la iconografía no responde a los principios de claridad y simplicidad necesarios. Además, en varios casos se observan errores gráficos evidentes: extremidades mal formadas o incompletas, proporciones inverosímiles o ausencias anatómicas, como se aprecia en la imagen de “abrocharse los cordones”, donde una mano aparece cortada.

Estas inconsistencias afectan también a la fidelidad semántica al prompt, ya que en algunas imágenes generadas por Qwen el contenido no representa con precisión la acción solicitada, o introduce elementos irrelevantes. En cuanto a la adecuación para el usuario objetivo, las imágenes de Qwen tienden a presentar mayor sobrecarga visual o estilos menos familiares para un entorno educativo estructurado, lo que puede dificultar su utilidad comunicativa.

Por todo ello, el análisis visual confirma y contextualiza los resultados cuantitativos obtenidos, reforzando la idea de que, al menos en esta etapa de desarrollo, DALL·E 3 ofrece una solución más robusta y fiable para la generación de pictogramas orientados a niños con autismo severo.

7. Comercialización de la solución

Desde una perspectiva comercial, este tipo de solución puede integrarse como plugin en aplicaciones móviles de uso cotidiano (como WhatsApp o Telegram) o como funcionalidad específica dentro de apps orientadas a la asistencia personalizada. El funcionamiento sería sencillo: el usuario toma una fotografía con su dispositivo móvil y recibe de forma inmediata una representación simbólica adaptada. Este enfoque busca promover la autonomía de las personas con TEA, permitiéndoles interpretar mejor su entorno sin requerir la mediación constante de un adulto.

La viabilidad de su implementación se ve reforzada por el creciente apoyo institucional a proyectos de inteligencia artificial inclusiva. El propio Gobierno de España ha destinado cinco millones de euros a iniciativas de IA orientadas a la mejora de la vida de personas con discapacidad, promoviendo una digitalización más equitativa y accesible (Congreso Diario, 2025). A nivel local, el Ayuntamiento de Barcelona impulsa el proyecto IDEM, cuyo objetivo es crear herramientas de IA que hagan más accesible la información pública y el debate democrático para personas neurodivergentes (Ajuntament de Barcelona, 2024). Además, entidades como la Universitat Politècnica de València han desarrollado asistentes virtuales con IA específicos para entornos de apoyo a personas con TEA, validando así el interés y la aplicabilidad de estas tecnologías en contextos reales (RUVID, 2024).

Estas iniciativas públicas y privadas demuestran que existe una demanda creciente de soluciones tecnológicas inclusivas, lo que abre la puerta a posibles acuerdos de colaboración, licenciamiento o financiación con fundaciones, administraciones y organizaciones sociales. El sistema propuesto podría incorporarse también en dispositivos de Realidad Aumentada o integrarse en asistentes virtuales contextuales, generando pictogramas en tiempo real mediante el uso de la cámara del móvil, y ofreciendo así un recurso accesible, escalable y útil tanto en contextos educativos como en la vida cotidiana.

Desde el punto de vista técnico, la solución puede desplegarse como un sistema modular basado en arquitecturas de backend que operan con modelos generativos entrenados previamente (como DALL·E o similares). Estos modelos se alojarían en servidores que procesan las imágenes enviadas por el usuario mediante llamadas API, devolviendo los pictogramas generados en segundos. Esta arquitectura permite la interoperabilidad con

diferentes plataformas y sistemas operativos, y facilita su implementación tanto en aplicaciones móviles como en entornos web. La capacidad de ajuste del modelo a distintas categorías semánticas, idiomas o estilos gráficos lo convierte en una herramienta adaptable a múltiples perfiles de usuario, regiones y niveles de competencia comunicativa.

En definitiva, se trata de una propuesta técnicamente viable y socialmente relevante, con un alto potencial de impacto positivo. Su comercialización puede orientarse a instituciones públicas, asociaciones de familias y centros educativos, pero también al mercado de la tecnología inclusiva, mediante licencias de uso, integraciones personalizadas o soluciones freemium adaptadas a distintos niveles de necesidad.

8. Conclusiones

Los resultados de este estudio evidencian el potencial real y tangible que la inteligencia artificial generativa puede tener como herramienta de apoyo a la comunicación en niños con Trastorno del Espectro Autista (TEA) de nivel 3. La comparación entre DALL·E 3 y Qwen2.5-Max no solo ha servido como ejercicio técnico, sino también como reflejo del contexto internacional en el que estas tecnologías emergen, se desarrollan y compiten.

Frente a la madurez y el refinamiento del modelo estadounidense, resultado de más de dos años de evolución, el modelo chino, aún en una fase relativamente temprana, parece haber generado expectativas algo desproporcionadas en relación con su rendimiento actual. Sin embargo, esta disparidad no desacredita el progreso alcanzado por la tecnología china, sino que más bien pone de relieve cómo la competencia entre regiones actúa como motor de innovación constante en este ámbito.

Las grandes potencias tecnológicas, como Estados Unidos y China, están protagonizando una carrera estratégica por liderar el desarrollo de la IA generativa. En paralelo, Europa comienza también a posicionarse con propuestas emergentes como Gemini, desarrollada por Google DeepMind. Aunque esta compañía tenga sede central en Reino Unido, su enfoque representa un esfuerzo europeo por crear modelos avanzados que combinen altos estándares técnicos con principios éticos sólidos.

Más allá del rendimiento comparado, este trabajo demuestra cómo estas tecnologías pueden cambiar la vida de muchas personas. Lo que a ojos de una mayoría puede parecer un avance menor, como transformar una imagen en un pictograma comprensible, se convierte en un recurso esencial para miles de familias, cuidadores, profesionales y, especialmente, para los propios niños. En un entorno donde el lenguaje verbal no está disponible, contar con un sistema personalizado y accesible de apoyo visual puede marcar la diferencia entre la desconexión y la comprensión.

Poco a poco, la tecnología comienza a dar visibilidad a colectivos tradicionalmente invisibilizados, reconociendo no solo sus necesidades, sino también sus capacidades. Este tipo de iniciativas no busca compensar una carencia, sino potenciar sus fortalezas, ayudarles a desplegar sus máximas posibilidades y acompañarles en una comunicación más justa, humana y adaptada. Porque cada avance en inclusión es también un paso hacia

una sociedad más equitativa, donde se entiende que apoyar no es limitar, sino liberar. El reto ahora está en continuar desarrollando herramientas con propósito, accesibles y escalables, que faciliten la vida cotidiana en todos sus ámbitos: desde lo clínico hasta lo urbano.

Este estudio aspira a ser un pequeño paso en ese camino, convencido de que la inteligencia artificial no debe quedarse en la vanguardia técnica, sino acercarse al día a día de quienes más pueden beneficiarse de ella. Porque solo cuando la innovación se pone al servicio de los más vulnerables, podemos hablar verdaderamente de progreso.

9. Futuras líneas de investigación

Uno de los próximos pasos fundamentales sería la validación empírica de la solución propuesta en contextos reales con niños diagnosticados con TEA de nivel 3. Realizar experimentos controlados en entornos educativos, terapéuticos o familiares permitiría observar directamente la eficacia comunicativa de los pictogramas generados, evaluar la aceptación por parte de los usuarios y recoger retroalimentación cualitativa de profesionales y cuidadores. Esta etapa de validación es esencial para refinar la herramienta y garantizar que responde a necesidades reales, más allá de la evaluación técnica.

Otro aspecto relevante sería analizar la posible integración de estos modelos en contextos multilingües y multiculturales. Explorar cómo se comportan los generadores de imágenes al interpretar descripciones en diferentes idiomas y con referencias culturales variadas permitiría ampliar la utilidad de la herramienta a poblaciones más diversas. Este tipo de estudios podría ayudar a ajustar los modelos para que mantengan una coherencia visual comprensible universalmente, o bien para adaptarse de forma localizada a distintas regiones.

También resulta pertinente investigar la experiencia de uso desde una perspectiva de usabilidad y accesibilidad. Diseñar estudios que analicen la interacción de usuarios con diferentes perfiles cognitivos y capacidades tecnológicas ayudaría a identificar barreras de acceso, mejorar las interfaces y garantizar que la tecnología sea realmente intuitiva y funcional para todos los públicos. Incorporar principios de diseño universal sería clave en esta línea de desarrollo.

Por otro lado, sería interesante explorar la evolución de las representaciones generadas por los modelos en función de variables como la edad del usuario o el contexto de uso. Por ejemplo, analizar si ciertos pictogramas son más efectivos para niños en etapas tempranas frente a otros más avanzados, o si el nivel de abstracción visual puede ajustarse automáticamente según el perfil del receptor.

Finalmente, conviene estudiar las implicaciones éticas del uso de IA generativa en contextos vulnerables. Investigar cómo preservar la privacidad de los datos, garantizar la transparencia en los procesos de generación y prevenir sesgos en los resultados es

fundamental para asegurar un desarrollo responsable y alineado con principios éticos. En este sentido, las futuras líneas de investigación también deben incluir un enfoque crítico y reflexivo que acompañe al avance tecnológico con una mirada humana y socialmente comprometida.

10. Bibliografía

Ajuntament de Barcelona. (2024, 29 de enero). *IDEM, el proyecto para crear herramientas de IA para hacer más accesibles la información y el debate democrático*. https://ajuntament.barcelona.cat/digital/es/actualidad/noticias/idem-el-proyecto-para-crear-herramientas-de-ia-para-hacer-mas-accesibles-la-informacion-y-el-debate-democratico-1364047

Alibaba Cloud. (s.f.). *Generative AI Solutions*. Recuperado el 18 de junio de 2025, de https://www.alibabacloud.com/en/solutions/generative-ai?_p_lc=1

Alibaba DAMO Academy. (2023). *Qwen-VL: Technical Documentation*. https://damo.alibaba.com

American Psychiatric Association. (2014). *DSM-5, Manual Diagnóstico y Estadístico de los Trastornos Mentales DSM-5*. Madrid: Editorial Médica Panamericana. https://www.eafit.edu.co/ninos/reddelaspreguntas/Documents/dsm-v-guia-consulta-manual-diagnostico-estadistico-trastornos-mentales.pdf

Asperger, H. (1944). Die "Autistischen Psychopathen" im Kindesalter. *Archiv für Psychiatrie und Nervenkrankheiten*, 117, 76–136.

Barua, P. D., Vicnesh, J., Gururajan, R., Oh, S. L., Palmer, E., Azizan, M. M., ... & Acharya, U. R. (2022). Artificial intelligence enabled personalised assistive tools to enhance education of children with neurodevelopmental disorders—a review. *International Journal of Environmental Research and Public Health*, 19(3), 1192. https://doi.org/10.3390/ijerph19031192

BBC News. (2025, 28 de enero). *The AI teacher: How this school put a chatbot in charge of a class*. Recuperado el 18 de junio de 2025, de https://www.bbc.com/news/articles/cd643wx888qo

Brück, F. (2024). *Generative neural networks for characteristic functions*. arXiv preprint arXiv:2401.04778. https://arxiv.org/abs/2401.04778

Cabello, F., & Bertola, E. (2015). Características formales y transparencia de los símbolos pictográficos de ARASAAC. *Revista de Investigación en Logopedia*, 5(1), 60-70. Universidad de Castilla-La Mancha. http://www.redalyc.org/articulo.oa?id=350841434004

Cabello, M. A., & Bertola, D. (2015). *Manual de pictogramas: Intervención educativa y terapéutica para personas con autismo*. Editorial CCS.

Congreso Diario. (2025, 5 de junio). *El Gobierno español destina 5 millones de euros a proyectos de IA inclusivos*. https://congresodiario.com/gobierno-espanol-5-millones-proyectos-ia-inclusivos/

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2021). An image is worth 16x16 words: Transformers for image

recognition at scale. *International Conference on Learning Representations (ICLR)*. https://arxiv.org/abs/2010.11929

Fernández, B., & Bandrés, C. (2023). *Tecnologías accesibles: diseño universal y comunicación aumentativa*. Editorial UOC.

Frolli, A., Cavallaro, A., La Penna, I., Sica, S. L., & Bloisi, D. (2023). Artificial intelligence and autism spectrum disorders: A new perspective on learning. In *Proceedings of the Digital Innovations for Learning and Neurodevelopmental Disorders* (pp. 22–34). CEUR Workshop Proceedings, Vol-3751. http://ceur-ws.org/Vol-3751/paper3.pdf

García-Peñalvo, F. J., Llorens-Largo, F., & Vidal, J. (2024). La nueva realidad de la educación ante los avances de la inteligencia artificial generativa. *RIED: Revista Iberoamericana de Educación a Distancia*, 27(1), 9–39. https://doi.org/10.5944/ried.27.1.37716

Gobierno de Aragón. (s.f.). *ARASAAC: Portal Aragonés de Comunicación Aumentativa y Alternativa*. Recuperado el 18 de junio de 2025, de https://arasaac.org/aac/

Iannone, A., & Giansanti, D. (2024). Breaking barriers—The intersection of AI and assistive technology in autism. *Journal of Personalized Medicine*, 14(1), 41. https://doi.org/10.3390/jpm14010041

Jesse, T. (2024). Creating neuro-inclusive learning environments: Integrating generative AI and outcome-led selection of teaching methods. In *Autism, Neurodiversity, and Equity in Professional Preparation Programs* (pp. 79–99). IGI Global.

Johnson, C., Smart, K., & Mahar, P. (s.f.). Is there a place for generative artificial intelligence in special education? *National Social Science Technology Journal*, 48.

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child*, 2, 217–250. http://www.th-hoffmann.eu/archiv/kanner/kanner.1943.pdf

Maldonado Gilarranz, C. (2024). *Generación de pictogramas con inteligencia artificial generativa para miembros del espectro autista* (Trabajo de Fin de Grado, Universidad Pontificia de Comillas).

Miao, M., Zheng, L., Wang, H., & Li, Y. (2021). Artificial intelligence in autism spectrum disorder: A systematic review. *Research in Developmental Disabilities*, 119, 104108. https://doi.org/10.1016/j.ridd.2021.104108

Montenegro-Rueda, M. (Ed.). (s.f.). *Innovación educativa, inclusión y tecnología. Estrategias para una sociedad accesible*. Dykinson.

Moraiti, I., Fotoglou, A., Stathopoulou, A., & Loukeris, D. (2023). Strategies & digital technologies for Autism integration. *Brazilian Journal of Science*, 2(5), 107–124. https://doi.org/10.14295/bjs.v2i5.290

OpenAI. (s.f.). *DALL·E 3*. Recuperado el 18 de junio de 2025, de https://openai.com/es-ES/index/dall-e-3/

Pino, N., & Guerrero, L. A. (2021). Artificial intelligence for special education: A systematic literature review. *Education and Information Technologies*, 26(1), 389–414. https://doi.org/10.1007/s10639-020-10398-8

Rojas-Torres, L., Alonso-Esteban, Y., & Alcantud-Marín, F. (2020). Revisión de evidencias de las técnicas de DIR/Floortime™ para la intervención en niños y niñas con trastornos del espectro del autismo. *Siglo Cero*, 51(2), 25–38. https://www.researchgate.net/publication/285043588_Dynamic_social_formations_of_pedestrian_groups_navigating_and_using_public_transportation_in_a_virtual_city

RUVID. (2024, 7 de noviembre). *Lanzan un asistente virtual de IA para ofrecer apoyo a personas con TEA y a sus entornos*. https://ruvid.org/lanzan-un-asistente-virtual-de-ia-para-ofrecer-apoyo-a-personas-con-tea-y-a-sus-entornos/

Sáenz, M., & Juárez, A. (2016). *Comunicación y lenguaje en personas con TEA*. Universidad Nacional Autónoma de México. https://d1wqtxts1xzle7.cloudfront.net/48546281/COMUNICACION_Y_LENGUAJE_EN_PERSONAS_CON_TEA-libre.pdf

Tang, Y., Chen, L., Chen, Z., Chen, W., Cai, Y., Du, Y., ... & Sun, L. (2024, mayo). EmoEden: Applying generative artificial intelligence to emotional learning for children with high-function autism. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (pp. 1–20).

11. Anexos

Anexo 1: Rúbrica de evaluación

1. Comprensibilidad pictográfica (30%)

Ponderación interna: 25 % por indicador

- Claridad del objeto principal: ¿Se identifica claramente el elemento o acción central de la escena sin generar confusión?
- Ausencia de elementos distractores: ¿Está libre de componentes superfluos o detalles visuales que puedan dificultar la atención al elemento principal?
- Estilo pictográfico adecuado: ¿La imagen se aproxima al estilo gráfico simplificado y esquemático característico de sistemas como ARASAAC?
- Iconicidad funcional: ¿La escena representa adecuadamente la función o intención comunicativa prevista (por ejemplo, “beber agua”?

2. Fidelidad semántica al prompt (20%)

Ponderación interna: 40 % – 30 % – 30 %

- Correspondencia semántica: ¿La imagen refleja con claridad el contenido solicitado en el prompt?
- Coherencia de escena: ¿Los elementos representados tienen una lógica interna y no presentan contradicciones?
- Integridad visual: ¿Están presentes todos los componentes relevantes indicados en el prompt?

3. Calidad visual técnica (15%)

Ponderación interna: 40 % – 30 % – 30 %

- Resolución y nitidez: ¿La imagen es clara y está bien definida?
- Ausencia de artefactos: ¿No presenta errores comunes en modelos generativos como duplicaciones anatómicas, texto distorsionado o elementos malformados?
- Composición visual: ¿La escena está bien encuadrada y organizada espacialmente?

4. Adecuación para el usuario objetivo (20%)

Ponderación interna: 40 % – 30 % – 30 %

- Simplicidad visual: ¿La imagen evita una densidad visual excesiva o elementos complejos?
- Uso de colores planos o neutros: ¿Los colores empleados favorecen la interpretación sin distraer?
- Familiaridad cultural: ¿Los objetos o escenas representadas resultan comprensibles en contextos culturales neutrales o conocidos?

5. Consistencia entre imágenes (15%)

Ponderación interna: 40 % – 30 % – 30 %

- Uniformidad de estilo: ¿Se mantiene una línea gráfica consistente a lo largo de las imágenes?
- Escala y proporciones: ¿Los tamaños relativos de los elementos son coherentes entre imágenes?
- Paleta cromática coherente: ¿Se utilizan colores similares o armónicos en las distintas imágenes?

Anexo 2: Declaración uso de IA

Declaración de Uso de Herramientas de IA Generativa en Trabajos Fin de Grado en Relaciones Internacionales.

Por la presente, yo, Nuria Castillo García estudiante de E2 Business Analytics de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado "Evaluación comparativa de pictogramas generados por Inteligencia Artificial para niños con autismo: modelo chino vs modelo estadounidense" declaro que he utilizado la herramienta de IA Generativa ChatGPT otras similares de IAG de código sólo en el contexto de las actividades descritas a continuación:

1. Sintetizador y divulgador de libros complicados: Para resumir y comprender literatura compleja.
2. Traductor: Para traducir textos de un lenguaje a otro.
3. Generación de pictogramas.

Afirmo que toda la información y contenido presentados en este trabajo son producto de mi investigación y esfuerzo individual, excepto donde se ha indicado lo contrario y se han dado los créditos correspondientes (he incluido las referencias adecuadas en el TFG y he explicitado para qué se ha usado ChatGPT u otras herramientas similares). Soy consciente de las implicaciones académicas y éticas de presentar un trabajo no original y acepto las consecuencias de cualquier violación a esta declaración.

Fecha: 4/06/2025

Firma: