



UNIVERSIDAD POLITÉCNICA SALESIANA
SEDE CUENCA
CARRERA DE COMPUTACIÓN

**DISEÑO Y DESARROLLO DE UNA APLICACIÓN MÓVIL BASADA EN
INTELIGENCIA ARTIFICIAL GENERATIVA PARA CREACIÓN DE
ACTIVIDADES DE OBSERVACIÓN DIRIGIDA A NIÑOS DE EDUCACIÓN
INICIAL**

Trabajo de titulación previo a la obtención del
título de Ingeniero en Ciencias de la Computación

AUTOR: ERICK ISMAEL VERDUGO SIGUENCIA

TUTOR: ING. MIGUEL ARTURO ARCOS ARGUDO, PHD.

Cuenca - Ecuador

2025

CERTIFICADO DE RESPONSABILIDAD Y AUTORÍA DEL TRABAJO DE TITULACIÓN

Yo, Erick Ismael Verdugo Siguencia con documento de identificación N° 0302142641 manifiesto que:

Soy el autor y responsable del presente trabajo; y, autorizo a que sin fines de lucro la Universidad Politécnica Salesiana pueda usar, difundir, reproducir o publicar de manera total o parcial el presente trabajo de titulación.

Cuenca, 21 de julio del 2025

Atentamente,

Erick Ismael Verdugo Siguencia
0302142641

CERTIFICADO DE CESIÓN DE DERECHOS DE AUTOR DEL TRABAJO DE TITULACIÓN A LA UNIVERSIDAD POLITÉCNICA SALESIANA

Yo, Erick Ismael Verdugo Siguencia con documento de identificación N° 0302142641, expreso mi voluntad y por medio del presente documento cedo a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud de que soy autor del Proyecto técnico: “Diseño y desarrollo de una aplicación móvil basada en inteligencia artificial generativa para creación de actividades de observación dirigida a niños de educación inicial”, el cual ha sido desarrollado para optar por el título de: Ingeniero en Ciencias de la Computación, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En concordancia con lo manifestado, suscribo este documento en el momento que hago la entrega del trabajo final en formato digital a la Biblioteca de la Universidad Politécnica Salesiana.

Cuenca, 21 de julio del 2025

Atentamente,

Erick Ismael Verdugo Siguencia

0302142641

CERTIFICADO DE DIRECCIÓN DEL TRABAJO DE TITULACIÓN

Yo, Miguel Arturo Arcos Argudo con documento de identificación N° 0103958021, docente de la Universidad Politécnica Salesiana, declaro que bajo mi tutoría fue desarrollado el trabajo de titulación: DISEÑO Y DESARROLLO DE UNA APLICACIÓN MÓVIL BASADA EN INTELIGENCIA ARTIFICIAL GENERATIVA PARA CREACIÓN DE ACTIVIDADES DE OBSERVACIÓN DIRIGIDA A NIÑOS DE EDUCACIÓN INICIAL, realizado por Erick Ismael Verdugo Sigüencia con documento de identificación N° 0302142641, obteniendo como resultado final el trabajo de titulación bajo la opción Proyecto técnico que cumple con todos los requisitos determinados por la Universidad Politécnica Salesiana.

Cuenca, 21 de julio del 2025

Atentamente,

Ing. Miguel Arturo Arcos Argudo, PhD.

0103958021

Dedicatoria y Agradecimiento

Durante los años que he recorrido esta carrera, he tenido el privilegio de conocer muchas personas a la vez que los acompañaba en la sustentación de sus tesis y en sus ceremonias de investidura. Su emoción y felicidad al culminar este ciclo me conmovían profundamente, llenándome de alegría y renovando mi esperanza de que, algún día, yo también llegue a ese momento. De cada uno de ellos aprendí mucho como el valor de siempre mantener nuestros ideales, nunca rendirse, siempre confiar en uno mismo, afrontar el miedo con determinación y serenidad. No hay palabras suficientes como para decirles a cada una de esas personas lo muy agradecidas que estoy con ellos por siempre ser una inspiración para mí.

En esta dedicatoria quiero dar gracias a Dios por nunca abandonarme siempre darme sus mejores enseñanzas, cuidar y proteger de mi familia. Gracias a mis padres por siempre apoyarme a pesar de que no siempre destaque en todas las materias, por su amor incondicional y ayudarme cuando no sabía que decisión tomar. Gracias a Esteban Novillo que siempre me apoyo, fue un gran mentor en el 2024 y sobre todo hizo que saliera de mi zona de confort, te quiero mucho amigo. Gracias a Cinthya Ramon por su apoyo incondicional, sus consejos y advertencias me han hecho madurar, sobre todo crecer como persona. Gracias a Pedro Pineda quien constantemente me da retos y me da críticas constructivas sobre proyectos, ideas o la vida en general.

Gracias a Gustavo Cerati y caifanes por su música que me ha acompañado. Gracias al ASU IEEE con el cual pude conocer a tan maravillosas personas que han sido parte de mi aventura como estudiante, me acogieron como uno más de su familia, fui parte de muchos proyectos que realizaron y eso cambio mi mentalidad. Gracias a Zahid, Manuel, Juan B., Jorge Cueva y mis amigos en general, los aprecio demasiado. Definitivamente GRACIAS a Miguel Isaac Verdugo Sigüencia que sin el amor incondicional de mi hermano nada de esto me hubiera pasado, no hubiera conocido a maravillosas personas mucho menos vivir esto.

*“Siempre haz todo con mucho amor y mucha fe en Dios,
porque algún día lo lograrás, de alguna forma u otra.”*

— Ayrton Senna

GRACIAS TOTALES.

1. Resumen

Este proyecto se enfoca en diseñar y desarrollar una aplicación móvil basada en inteligencia artificial generativa destinada a la educación inicial, con el propósito de mejorar el aprendizaje mediante imágenes, las habilidades cognitivas, lingüísticas, sobre todo de análisis en niños de educación básica. La educación infantil enfrenta desafíos significativos al adoptar métodos educativos innovadores, especialmente en el aprovechamiento de tecnologías como la IAG y los Modelos de Lenguaje Visual. La aplicación tiene como objetivo combinar tecnologías avanzadas como generación automática de imágenes a partir de texto, conversión de texto en voz (TTS) y reconocimiento de voz (STT), generando actividades personificadas e interactivas adaptadas al desarrollo cognitivo infantil. Estos avances tecnológicos nos permiten experiencias educativas dinámicas y factibles, facilitando un aprendizaje activo, inclusivo y flexible.

El instrumento utiliza un juicio participativo donde docentes ingresan un prompt que crea imágenes concretas, sobre las cuales la IA formula interrogaciones presentadas a los niños mediante audio. Los dictámenes orales son analizados al instante por la IA, brindando un feedback inmediato y ajustable. El uso de IA Generativa y VLM presenta gran utilidad pedagógicas, como la personalización educativa, estímulo del aprendizaje infantil junto al beneficio de métodos típicos de enseñanza. Además, esta aplicación favorecerá significativamente al campo de la educación infantil mediante la inclusión de prácticas pedagógicas innovadoras y estimulantes. Por último, la aplicación será estimada por docentes para decretar su efectividad y practicidad educativa, avalando así una herramienta efectiva que se adapta a las necesidades formativas actuales.

Palabras clave: Inteligencia artificial generativa, Modelos de Lenguaje Visual, educación infantil, aplicación móvil, aprendizaje interactivo.

2. Abstract

This project focuses on designing and developing a mobile application based on generative artificial intelligence for early childhood education, with the purpose of enhancing learning through images and improving cognitive, linguistic and analytical skills in children. Early childhood education faces significant challenges in adopting innovative educational methods, especially in leveraging technologies such as GAI and Visual Language Models. The application aims to combine advanced technologies such as automatic image generation from text, text-to-speech (TTS) and speech recognition (STT), generating personalized and interactive activities adapted to children's cognitive development. These technologies enable dynamic and accessible educational experiences, facilitating active, inclusive and adaptive learning.

The tool uses an interactive process where teachers enter a prompt that generates specific images, on which the AI formulates questions presented to the children through audio. The verbal responses are analyzed in real time by the AI, providing immediate and adaptive feedback. The use of generative AI and VLM brings important pedagogical benefits, such as educational personalization, motivation of children's learning, and enrichment of traditional teaching methods. In addition, this application will contribute significantly to the field of early childhood education by introducing innovative and stimulating pedagogical practices. Finally, the application will be evaluated by teachers to determine its effectiveness and educational practicality, thus ensuring an effective tool that adapts to current educational needs.

Keywords: Generative Artificial Intelligence, Visual Language Models, Mobile Application, Early Childhood Education, Narrative Stimulation, Cognitive Development, Text-to-Speech, Speech-to-Text, Adaptive Learning, Pedagogical Innovation.

Tabla de contenido

1. Resumen	6
2. Abstract	7
3. Introducción	9
4. Problema	9
4.1. Descripción del problema	9
4.2. Antecedentes	10
4.3. Importancia y Alcances	11
5. Objetivos	12
5.1. Objetivo general	12
5.2. Objetivo Especifico	12
6. Fundamentos teóricos	12
7. Marco Metodológico	15
7.1. Fases de la Metodología SCRUM	16
7.1.1. Sprint 1: Análisis y definición de requerimientos	16
7.1.2. Sprint 2: Diseño funcional y pedagógico de la aplicación	17
7.1.3. Sprint 3: Desarrollo del generador de imágenes con IA generativa	17
7.1.4. Sprint 4: Prueba piloto y validación con docentes	18
7.2. Instrumentos de recolección de datos	19
8. Resultados	20
9. Cronograma de actividades	20
10. Presupuesto	21
11. Conclusiones	22
13. BIBLIOGRAFIA	23
13. Anexos	25

3. Introducción

Dentro de un contexto donde a los niños pequeños se les quiere enseñar todo lo posible ya que tienen la capacidad de absorber todo lo que aprenden creando unas bases sólidas para aprender temas más complejos a futuro, se ha buscado maneras para que los niños aprendan. La formación inicial es una etapa esencial en el perfeccionamiento integral de los infantes, donde se establecen las bases cognitivas, gramáticas y sociales esenciales para el aprendizaje a futuro. Sin embargo, uno de los primordiales retos en esta etapa es efectuar métodos pedagógicos transformadores y efectivos que atraigan la curiosidad de los niños y provoquen su aprendizaje activo, pero sobre todo significativo. A pesar de los progresos tecnológicos actuales, la totalidad de las aplicaciones educativas servibles les hace falta de estrategias personalizadas que utilicen plenamente el potencial de las nuevas tecnologías emergentes, como la IA y los Modelos de Lenguaje Visual (VLM).

Por esta petición, el proyecto crea y desarrolla una aplicación móvil basada en IA Generativa, para la educación de infantes. La aplicación usa tecnologías de vanguardia como la formación de imágenes desde descripciones también conocidos como (prompts), la conversión de texto en voz (TTS) y esencialmente el reconocimiento de voz (STT). Estos instrumentos mezclados permiten generar actividades educativas participativas, adaptativas, personalizadas y favorables a mejorar las destrezas cognitivas, lingüísticas y razonadas de los niños mediante la estimulación narrativa-visual. El proyecto se completará con la metodología ágil SCRUM por su elasticidad, gestión iterativa de aclimatación en momentos de alta complejidad técnica y pedagógica.

El plan estará formado de cuatro sprints: análisis de requerimientos, boceto funcional y pedagógico, crecimiento especialista de la aplicación, por último, la validación mediante pruebas piloto con docentes. Al finalizar estos sprints, se realizará una estimación absoluta que permitirá analizar la eficacia educativa y la captación de beneficio del utensilio de los docentes, garantizando que la aplicación satisfaga con las expectativas y la exigencia pedagógicas de la vida actual. Esto no solo explora la transformación en el ámbito tecnológico, también cambia positivamente el ambiente educativo, dando nuevas perspectivas y herramientas que suministren un aprendizaje más interesante y enriquecedor para los niños. Siendo parte de una guía para el futuro de las aplicaciones relacionadas a este ámbito.

4. Problema

4.1. Descripción del problema

La enseñanza infantil afronta significativos retos al adecuar distintos métodos de ilustración que provoquen un aprendizaje dinámico y recíproco, orgánico con el desarrollo cognitivo de los chicos pequeños. El manejo de aplicaciones formativas actuales, en su mayoría, no exploran el potencial

de la inteligencia artificial generativa para instituir actividades tipificadas y visualmente atractivas, lo que restringe su capacidad de captar la curiosidad y renovar el aprendizaje de los estudiantes. Además, la mezcla de tecnologías como la conversión de texto en voz (TTS) y la identificación de voz (STT) en aplicativos pedagógicos para niños, sigue siendo un área insuficientemente investigada. El uso de Visual Language Models (VLM) puede restaurar la creación de materia prima de aprendizaje personalizados y dinámicos, pero su implementación en aplicaciones celulares accesibles y eficaces para la instrucción infantil aún no se ha examinado lo suficiente. Como lo explica (McKinney et al., 2022), el uso de tecnologías de IA generativa en la educación infantil es cada vez más popular, pero aún no se ha optimizado para la creación de contenidos visuales interactivos adaptados al desarrollo cognitivo de los niños.

Esta falta de herramientas que combinen la creación de imágenes apropiadas al contexto infantil con la interacción a través de la emisión de voz para sembrar la comprensión y la retroalimentación a través de la personalización del dialecto. (Hernández et al. 2021) expresa que los actuales sistemas de educación infantil no suelen componer adecuadamente las tecnologías interactivas que promuevan el desarrollo de las habilidades intelectual y lingüísticas de los niños a través de experiencias personalizadas. Por tanto, el problema central radica en cómo diseñar y desarrollar una aplicación móvil basada en inteligencia artificial generativa que, mediante modelos de lenguaje visual, permita crear actividades de observación interactivas y adaptativas para niños de educación inicial, apoyando el desarrollo de habilidades cognitivas, expresivas y analíticas.

La investigación busca cubrir un vacío en la integración de tecnologías como la generación de imágenes, el reconocimiento de voz (STT) y la conversión de texto a voz (TTS) para tener un entorno de aprendizaje interactivo y personalizado. Aunque el STT y TTS han tenido un gran avance en aplicaciones educativas, su uso específico en la educación infantil es limitado, tal como señalan González y Pérez (2020). La propuesta tiene como objetivo generar una participación más activa de los niños para favorecer el desarrollo de aptitudes esenciales mediante recursos educativos accesibles y lúdicos. Esta visión se alinea con los hallazgos de (Zhang et al. 2019), que expresa el potencial de los Visual Language Models (VLM) para la creación de contenido educativo dinámico y personalizado.

4.2. Antecedentes

En el último año, la educación en niños ha tenido retos relacionados a la realización de métodos educativos adaptados al desarrollo cognitivo y emocional de los pequeños. Las técnicas tradicionales, aunque efectivas en ciertos casos, han dado restricciones importantes en cuanto a personalización, interactividad y adaptabilidad a las necesidades individuales de cada estudiante. Estudios recientes han probado que estas limitaciones pueden afectar negativamente el interés y la participación de los niños, reduciendo así el potencial educativo y su impulso integral. Por el auge de las tecnologías digitales, muchas instituciones educativas han tratado de hacer

herramientas tecnológicas en sus métodos de enseñanza, en busca de mayor certeza pedagógica. Pero gran parte de estos instrumentos no usan totalmente las capacidades de estas como la (IAG) y los (VLM). De acuerdo con estudios recientes, las aplicaciones reales en educación inicial a menudo carecen de la personalización adecuada para responder a la pluralidad cognitiva, lingüística y cultural de los alumnos.

Investigaciones previas muestran que el uso condicionado de estas tecnologías puede tener varios factores: falta de formación docente, poca confianza con tecnologías modernas, poca investigación en educación infantil y falta de confirmación positiva sistemática sobre la huella real de dichas tecnologías en el ámbito educativo inicial. Según el informe de (UNESCO 2021), el uso adecuado de tecnologías digitales recientes en etapas tempranas podría ayudar a reducir fisuras educativas, mejorando la calidad del aprendizaje infantil. Adicionalmente, muchos estudios académicos han mostrado la gran necesidad de desarrollar equipos tecnológicos más adaptados y personalizados que permitan aprovechar al máximo las ventajas que brindan las tecnologías emergentes, especialmente en contextos formativos iniciales, sobresaliendo la importancia crítica de investigar y aceptar estas soluciones tecnológicas en escenarios educativos reales (McKinney et al., 2022; Hernández et al., 2021). En este contexto, este trabajo busca indispensablemente abordar estos informes identificados, desarrollando y validando una herramienta tecnológica innovadora, capaz de dar soluciones pedagógicas efectivas para plantarse ante estos retos en la educación infantil.

4.3. Importancia y Alcances

La presente investigación aborda el problema clave de la educación de niños de educación inicial, una problemática la cual aun no ha sido enmendado con las tecnologías digitales que se han convertido en una necesidad social para garantizar la educación como un derecho básico para la humanidad. Debido a la pandemia del Covid-19 muchos países no contaban con una infraestructura de TIC suficiente ni con sistemas de aprendizaje digital adecuados, lo que significo en pérdidas de aprendizaje considerables.

El proyecto estará delimitado geográficamente a instituciones educativas ubicadas en las ciudades de Cañar y Cuenca, Ecuador, temporalmente circunscrito al periodo académico 2025. En cuanto a limitaciones sectoriales e institucionales, el desarrollo y estudio inicial se centrarán exclusivamente en el sector educativo preescolar, específicamente en instituciones que dispongan de la infraestructura tecnológica adecuada para implementar sobre todo para evaluar la herramienta. Adicionalmente, se reconoce que la efectividad de la aplicación puede estar condicionada por factores externos como la formación tecnológica previa de los docentes, la accesibilidad tecnológica de las familias y el contexto socioeconómico particular de las instituciones educativas involucradas.

5. Objetivos

5.1. Objetivo general

Diseñar y desarrollar un programa móvil basado en inteligencia artificial generativa, que permita la estimulación narrativa mediante imágenes para fomentar el desarrollo de habilidades cognitivas y lingüísticas en educación inicial.

5.2. Objetivo Especifico

1. Investigar y revisar la literatura y las tecnologías existentes relacionadas con la IA generativa, los Visual Language Models, y su aplicación en entornos educativos para identificar las mejores prácticas y herramientas disponibles en la estimulación narrativa mediante imágenes.
2. Diseñar y desarrollar una aplicación móvil basada en Inteligencia Artificial Generativa que permita la creación de imágenes en base a las directrices de los maestros para que niños de educación inicial puedan trabajar el campo de la estimulación narrativa.
3. Ejecutar un plan de experimentación para determinar el nivel de consenso de un grupo de educadores con relación a la utilidad de la herramienta para llevar a cabo actividades de estimulación narrativa mediante imágenes.

6. Fundamentos teóricos

Desde los inicios de la era tecnológica, la humanidad ha experimentado importantes avances tecnológicos que han cambiado nuestro mundo de maneras inimaginables. Iniciativas innovadoras que se han convertido en hitos en la historia de la tecnología, sino que también han dado forma significativa a la realidad en la que vivimos hoy. Estos resultados son la base para futuras innovaciones y cambios en varias áreas. En esta trayectoria de progreso tecnológico, el continuo y vertiginoso avance de la inteligencia artificial (IA) y sus diversas ramas ha marcado un hito importante que deja sentada una huella en la historia de la sociedad y sus progresos tecnológicos.

La inteligencia artificial, ha sido parte de varios ámbitos como la ciencia, industria, comunicación, deporte y sobre todo en la educación actual, integrado se de manera cada vez mas profunda y amplia en nuestros procesos cotidianos, llegando incluso a afirmarse que la misma constituye la cuarta revolución industrial (Teigens et al., 2020). Entre sus aplicaciones, la inteligencia artificial generativa la cual permite crear nuevo contenido como son imágenes, videos, audio o código a partir de modelos entrenados en grandes volúmenes de datos. Además, Genera artefactos novedosos y coherentes, mostrando una capacidad creativa que antes se atribuía solo a los seres humanos. A diferencia de las IA tradicionales enfocadas en reconocimiento o clasificación, la IAG genera artefactos novedosos y coherentes, mostrando una capacidad creativa que antes se atribuía

solo a los seres humanos. Del mismo modo, esta se encuentra en una fase temprana pero altamente prometedora. Según (Deloitte, 2024) casi un 80 % de líderes corporativos espera que sus industrias se transformen significativamente en los próximos tres años. El 56 % de expertos en IA cree que esta tendrá un impacto muy o algo positivo en EE.UU. durante los próximos 20 años, especialmente en ámbitos como la educación y la sanidad (Abril, 2025). Los investigadores de diversos campos la ven como una tecnología tan transformadora como la electricidad, capaz de amplificar las capacidades humanas y moldear el futuro de sectores como la educación, siempre y cuando se gestione con responsabilidad.

La inteligencia artificial generativa (IAG) constituye una de las áreas más innovadoras del campo de la inteligencia artificial, centrada en la capacidad de generar contenidos nuevos, originales y significativos a partir de patrones previamente aprendidos en grandes cantidades de datos (Goodfellow et al., 2014). Dentro de este campo, una subcategoría específica es la generación visotextual, que encadena técnicas de procesamiento de lenguaje natural (NLP) avanzadas con los modelos de generación de imágenes para producir representaciones visuales contextualizadas y coherentes (Ramesh et al., 2021). Los Modelos de Lenguaje Visual, componen un desarrollo tecnológico que armoniza técnicas de visión artificial y procesamiento de lenguaje de natural con el objetivo de establecer ilustraciones a partir de prompts textuales y detalladas. Estos modelos anteriormente manifestados usan redes neuronales profundas surtidas con enormes bases de datos multimodales con la capacidad de interpretar el lenguaje natural para la creación de imágenes de visuales agradables y con alta correspondencia semántica frente al prompt suministrado todo esto nos expresa (Radford et al., 2021; Saharia et al., 2022). El empleo de VLM se ha vuelto popular en los últimos años gracias a modelos como DALL·E, Sora, Stable Diffusion, nos han mostrado realizaciones excepcionales en variadas tareas creativas, educativas y comunicativas

La parte visual del modelo captura los datos de las imágenes ya sean objetos, atributos, etc. Mientras que la parte de lenguaje comprende el texto asociado; trabajando en conjunto, el VLM puede analizar y generar descripciones o respuestas basadas en la comprensión combinada de ambas modalidades. Este enfoque permite que un VLM reciba como entrada una imagen o conjunto de imágenes, e incluso frames de video junto con texto, una pregunta, un contexto o prompt y produzca una salida en lenguaje natural relacionada con el contenido visual proporcionado (“Vision Language Models Explained,” 2024). Algunas de las capacidades clave:

- **Subtitulación de imágenes (image captioning):** generar descripciones textuales detalladas a partir del contenido de una imagen (Vina, 2025). Por ejemplo, dado una fotografía, un VLM puede producir una frase que resuma la escena “Una niña jugando con un perro en el parque”.
- **Respuesta a preguntas visuales (Visual Question Answering, VQA):** responder preguntas en lenguaje natural basadas en el contenido de una imagen (Vina, 2025). Por ejemplo, tras mostrar una imagen y la pregunta “¿Cuántas personas están sentadas a la mesa?”, el modelo analiza la imagen y devuelve la respuesta.
- **Generación de imágenes a partir de texto:** algunos VLM (o modelos relacionados) pueden hacer lo inverso, es decir, crear imágenes nuevas a partir de descripciones textuales

(como lo hace la familia DALL-E, aunque estrictamente este es un modelo generativo de visión a partir de lenguaje) (Vina, 2025). Esta capacidad de síntesis visual a partir de texto ha abierto el campo de la creación de contenido visual por IA.

- **Comprensión de escenas y detección de objetos:** identificar y clasificar los objetos presentes en una imagen, así como entender la disposición general de la escena. Esto puede incluir tareas de localización visual, como generar cuadros delimitadores o máscaras segmentando objetos en la imagen cuando se solicita.
- **Generación de contenido multimodal:** combinar información visual y textual para producir nuevas salidas, como por ejemplo un diálogo visual donde el modelo observa una imagen y sostiene una conversación sobre ella, o generar historias que describen una secuencia de imágenes.

Para lograr estas funcionalidades, los VLM emplean arquitecturas avanzadas. En términos generales, suelen componerse de: (1) un codificador visual que convierte la imagen en una representación interna matemática; (2) un codificador de texto que convierte las oraciones en embeddings lingüísticos; y (3) un mecanismo de fusión o interacción multimodal que combina ambas modalidades (Chandra, 2025). Los VLM son cruciales ya que habilitan nuevas formas de interacción y capacidades antes inéditas en los sistemas de IA. La posibilidad de ingresar tanto imágenes como texto a un modelo amplía drásticamente las aplicaciones, se puede tener asistentes conversacionales que además de “leer” también “ver” para responder, herramientas que analicen contenido visual complejo y lo expliquen verbalmente, entre otras. Su capacidad para conectar palabras con el mundo visual los convierte en una tecnología estratégica, con aplicaciones que pueden mejorar significativamente la manera en que interactuamos con las máquinas y comprendemos datos complejos.

En la actualidad, la educación infantil enfrenta grandes desafíos al intentar realizar métodos pedagógicos que estimulen el interés y desarrollen las capacidades de aprendizaje en los niños pequeños. Pese al desarrollo y propagación de varias aplicaciones educativas, muchas de estas herramientas escasean de personalización e interactividad necesarias para hechizar y sostener eficazmente la atención infantil, reduciéndose repetidamente a rumbos tradicionales rígidos, genéricos y poco adecuados a las necesidades específicas de cada estudiante. Además, el potencial convertidor de tecnologías emergentes como la inteligencia artificial generativa y los Modelos de Lenguaje Visual aún no ha sido adecuadamente aprovechados ni explorados en el ámbito de la educación inicial, restringiendo así significativamente el progreso de actividades visuales, auditivas innovadoras, además de ser personalizadas, adaptativas y altamente interactivas que podrían perfeccionar sustancialmente la experiencia educativa infantil. Además, se afirma que los padres de familia posean el control general de lo que prenden sus niños ya sea desde casa o en la escuela.

Una de las primordiales lagunas halladas es la escasa de disponibilidad de aplicaciones educativas que componen eficazmente tecnologías de vanguardia como la creación de fotografía personalizadas con inteligencia artificial, la identificación automática del habla (STT) y la transformación de texto en voz (TTS), de gran importancia para crear experiencias de ilustración

honestamente personales y enriquecedoras. Una limitación significativa en la actualidad es la poca investigación en la adaptación de tecnologías en la educación infantil, considerando la diversidad cognitiva, lingüística, sensorial y social que caracteriza a los niños durante sus primeros años de desarrollo. Además, se necesita tener certeza real, junto a los impactos cognitivos positivos que estas herramientas pueden generar en los procesos educativos.

El uso de la Inteligencia Artificial Generativa en la educación preescolar da varias ventajas. Primero, crea una innovación pedagógica que las aplicaciones móviles basadas en IA generativa pueden diseñar, actividades educativas personalizadas, dinámicas y adaptables. Estas actividades captan la atención de los niños, estimulan su curiosidad y desarrollan habilidades de observación, análisis y exploración propia. Segundo, la educación multimodal se ve fortalecida por la combinación de estímulos visuales, auditivos y sensoriales, mejorando el desarrollo de todas las habilidades desde etapas iniciales. Esta integración también crea la accesibilidad, facilitando la adaptación de las actividades educativas a las características y necesidades individuales de cada niño, lo que mejora la calidad de la educación, haciéndola más inclusiva.

Por último, este proyecto da una contribución significativa no solo desde una perspectiva pedagógica, sino también tecnológica y social. El uso de esta aplicación puede crear avances relevantes en el ámbito educativo y animar a la creación de nuevos conocimientos en el campo de la inteligencia artificial aplicada a la educación preescolar. También, esta herramienta tiene un gran potencial de escalabilidad y adaptación a otros niveles educativos, beneficiando a la comunidad educativa en general.

7. Marco Metodológico

La razón es crear una nueva metodología de enseñanza y aprendizaje para niños. El desarrollo de esta aplicación móvil, con inteligencia artificial generativa y modelos visuales del lenguaje, se enfoca en la educación infantil. Para la realización de este se utilizó la metodología SCRUM, escogida por su flexibilidad, capacidad de adaptabilidad, enfoque iterativo y características valiosas en proyectos tecnológicos complejos que requieren una constante interacción con los usuarios. El proyecto se ejecuta en el marco de una investigación aplicada y tecnológica, siguiendo un enfoque de métodos que mezcla elementos cualitativos y cuantitativos que analiza el rendimiento técnico y la eficiencia pedagógica aplicada. El objetivo es dar soluciones a problemas específicos que fueron detectados en la educación preescolar, relacionados con los métodos de aprendizaje.

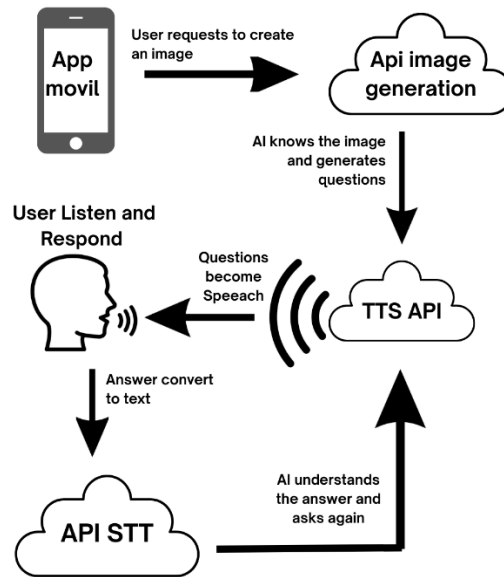


Figura 1. Funcionamiento de la aplicación movil

7.1. Fases de la Metodología SCRUM

El método SCRUM se trabaja por períodos cortos conocidos como sprints, cada uno con fines definidas y entregables específicos:

7.1.1. Sprint 1: Análisis y definición de requerimientos

Este sprint se ejecutó con absoluta indagación bibliográfica para identificar las necesidades específicas de aprendizaje de los infantes en educación inicial, así como un estudio de las tecnologías existentes de inteligencia artificial generativa para establecer cuál se acomoda mejor al contexto educativo presentado. Las actividades desarrolladas fueron:

- Revisión del estado del arte en IA Generativa, Modelos de Lenguaje Visual (VLM) y métodos pedagógicos relacionados con la estimulación narrativa.
- Análisis comparativo de modelos y herramientas tecnológicas como DALL·E, GPT-Vision y otras plataformas disponibles.
- Estudio del impacto de aplicaciones móviles educativas en el desarrollo cognitivo, lingüístico y social infantil.
- Investigación de literatura experta en pedagogía, narrativa pedagógica y aplicaciones tecnológicas.

7.1.2. Sprint 2: Diseño funcional y pedagógico de la aplicación

Esta etapa se concentró en el diseño delicado de la aplicación, tomando en cuenta las insuficiencias pedagógicas de los docentes y los niños. Se utilizó Canva para realizar prototipos visuales participativos que accedieran anticipar la experiencia del usuario. Las tareas realizadas incluyeron:

- Diseño de la arquitectura general del sistema.
- Creación de flujos específicos de usuario adaptados a niños en edad preescolar.
- Desarrollo de interfaces de usuario (UI/UX) con enfoque en facilidad de uso, accesibilidad y atractivo visual.
- Preparación de modelos y wireframes que puntualizan la interacción del usuario.
- Boceto de prompts estructurados para perfeccionar la generación de retratos con inteligencia artificial.



Figura 2. Diseño fácil de usar



Figura 3. Diseño atractivo para los niños y docentes

7.1.3. Sprint 3: Desarrollo del generador de imágenes con IA generativa

Durante este sprint se compendió el desarrollo técnico de la aplicación, integrando dispositivos clave como la creación de imágenes adaptadas al contexto infantil y la interacción auditiva junto a la verbal. Se incluyeron actividades tales como:

- Implementación del backend utilizando modelos VLM para generación adaptativa de imágenes.
- Desarrollo del frontend móvil utilizando PWA para lograr compatibilidad multiplataforma.
- Integración efectiva del módulo de IA generativa.

- Desarrollo de un módulo específico para permitir que los docentes ingresen los prompts personalizados.
- Creación del sistema de visualización de imágenes con narración auditiva guiada, integrando tecnologías TTS (texto a voz) para una experiencia auditiva amigable.
- Incorporación del sistema de reconocimiento de voz (STT) para captar y evaluar respuestas verbales de los niños en tiempo real.
- Validación técnica interna, asegurando la combinación fluida y funcionalidad óptima de todos los componentes del sistema.



Figura 4. Aplicación fácil de usar y muy intuitiva

7.1.4. Sprint 4: Prueba piloto y validación con docentes

La última etapa consistió en la implementación práctica de la aplicación en un entorno real con participación de docentes y estudiantes. El objetivo fue evaluar la funcionalidad pedagógica y técnica, obteniendo retroalimentación directa para mejorar el producto final. Las actividades desarrolladas fueron:

- Sesiones piloto con educadores para probar la aplicación en contextos educativos reales.
- Recolección de datos cualitativos y cuantitativos mediante encuestas, entrevistas semi-estructuradas y observaciones directas.
- Análisis detallado de los resultados obtenidos para identificar fortalezas, debilidades y áreas de mejora.

7.2. Instrumentos de recolección de datos

Se utilizaron encuestas estructuradas, entrevistas semi-estructuradas y observación directa durante la prueba piloto, permitiendo recopilar información precisa sobre la percepción de utilidad, facilidad de uso y efectividad educativa. Para la recolección de datos se utilizaron encuestas estructuradas dirigidas a padres de familia, docentes y público en general con el fin de obtener información precisa sobre sus preferencias, necesidades y percepción respecto a las aplicaciones educativas orientadas a niños de educación inicial (de 3 a 6 años). La encuesta, diseñada específicamente para este proyecto, fue anónima y de corta duración, aproximadamente 5 minutos, asegurando la confidencialidad de las respuestas.

En total participaron 29 personas, de las cuales un 44,8% indicó tener niños en educación inicial. Los encuestados mencionaron utilizar diversas aplicaciones educativas como YouTube, Lingo Kids, TikTok, Zoom, ChatGPT, Progentis y aplicaciones específicas para aprender números, colores y animales. La frecuencia de uso varió significativamente: un 20,7% señaló que las utilizan diariamente, un 27,6% varias veces por semana, mientras que un 37,9% indicó un uso menos frecuente, menos de una vez por semana. Al consultar sobre el tipo de actividades educativas preferidas para los niños, se destacó ampliamente el interés por juegos visuales (imágenes, dibujos y colores) con un 75,9%, seguido de actividades interactivas con voz (respuestas habladas y reconocimiento de voz) con un 55,2%, y actividades auditivas (audio, narración y canciones) con un 44,8%. En contraste, actividades basadas únicamente en texto fueron las menos preferidas, con solo un 10,3%. Revisar anexos de Figura 5 y 6

Respecto a la personalización de las actividades educativas según el nivel o intereses del niño, el 55,2% de los participantes consideró esto como "muy importante", mientras que el 24,1% lo calificó como "esencial". Esto refleja un interés significativo por aplicaciones educativas altamente adaptativas. Finalmente, al preguntar sobre características deseadas en aplicaciones educativas que actualmente no se encuentran disponibles, surgieron varias sugerencias interesantes: contenido multisensorial (audio, imágenes animadas, vibración y mini tareas físicas), personalización e integración completa con la escuela, aprendizaje de idiomas adicionales al inglés, y actividades que mantengan constantemente la atención del niño. Revisar anexos de las figuras 7 y 8

El análisis incluyó métodos cuantitativos para evaluar indicadores objetivos relacionados con la interacción, la frecuencia de uso, las preferencias en cuanto al tipo de actividades y la capacidad de pago. Además, se utilizaron métodos cualitativos para interpretar las sugerencias y percepciones

expresadas por los usuarios finales, lo que permitió una evaluación completa y en profundidad del potencial educativo del producto desarrollado.

8. Resultados

Durante el desarrollo de este proyecto, se logró algo que parecía ambicioso al principio: crear y poner en manos de educadores una aplicación móvil pensada exclusivamente para niños entre 3 y 6 años. La propuesta no se quedó en teoría. Se convirtió en una herramienta real, construida con tecnologías de inteligencia artificial generativa, como modelos de lenguaje visual, que no solo generan imágenes, sino que entienden la intención detrás de cada palabra.

Para este resultado, se trabajó en cuatro etapas. La primera en entender de verdad a quién nos dirigíamos, se analizó el contexto, se exploró las necesidades del aprendizaje preescolar y se puso especial atención en cómo los niños perciben el mundo por medio de imágenes, sonidos y la interacción. Las herramientas elegidas no fueron al azar. El api de Chat GPT permitió crear imágenes adaptadas al contenido educativo, mientras que GPT-Vision ayudó a personalizar la experiencia. Todo esto con la metodología SCRUM que permitió avanzar paso a paso, sin perder de vista el objetivo. En la fase de diseño, se hicieron prototipos con Canva, cada interfaz fue pensada para ser intuitiva y atractiva para niños, basándonos en experiencias propias y sobre todo centrado en el usuario. Las pruebas tempranas mostraron que los niños respondían con entusiasmo: jugaban, interactuaban, lo más importante aprendían. Ya en la parte técnica, la aplicación incorporó reconocimiento de voz, respuestas con retroalimentación y generación automática de imágenes y sonidos. Esto para que los niños pudieran responder y recibir respuestas, ver cómo sus palabras se transformaban en contenido visual y seguir explorando sin sentirse perdidos.

Finalmente, durante la fase de validación, se realizaron pruebas piloto con docentes y pequeños usuarios. Los resultados fueron más que alentadores: el 75% de los niños preferían actividades visuales, y más de la mitad se sentían cómodos y motivados con las funciones de reconocimiento de voz. Aún más revelador: casi el 80% de los adultos encuestados estarían dispuestos a pagar por una app como esta. Esto no solo confirmó la efectividad del enfoque, sino que mostró un interés real y un potencial de crecimiento a futuro.

9. Cronograma de actividades

A continuación, se suministra una descripción de las prestezas planificadas conforme a los objetivos planteados previamente.

Fase del Proyecto	Actividades	Fecha de Inicio	Fecha de Finalización	Duración
Análisis y definición de requerimientos	Investigación bibliográfica, análisis de tecnologías existentes (DALL·E, GPT-Vision)	1 abril 2025	15 abril 2025	2 semanas
Diseño funcional y pedagógico	Desarrollo de prototipos interactivos en Canva, ajustes según retroalimentación inicial	16 abril 2025	30 abril 2025	2 semanas
Desarrollo técnico	Integración backend y frontend, implementación STT y TTS, generación dinámica de imágenes	1 mayo 2025	31 mayo 2025	4 semanas
Pruebas piloto y validación	Realización de pruebas piloto, análisis de resultados y ajustes finales según retroalimentación	10 junio 2025	24 junio 2025	2 semanas

Cuadro 1. Cronograma de actividades planteado par el desarrollo del proyecto

10.Presupuesto

Se muestra el presupuesto de los consumos conocidos en el cuadro 2

Descripción del Recurso	Cantidad	Costo Unitario (USD)	Costo Total (USD)
Licencia API IA Generativa (ej. DALL·E, GPT-Vision)	1	\$20.00	\$20.00
Servicios de Internet y computacionales	3 meses	\$40.00/mes	\$120.00
Transporte y movilidad (pruebas piloto y entrevistas)	8	\$5.00	\$40.00
Validación pedagógica (asesoría de especialistas educativos)	2 sesiones	\$150.00/sesión	\$300.00
Desarrollo y mano de obra técnica (desarrollador principal)	160 horas	\$20.00/hora	\$3,200.00

Descripción del Recurso	Cantidad	Costo Unitario (USD)	Costo Total (USD)
Materiales de oficina e impresión para documentación	Varios	\$50.00	\$50.00
Imprevistos (5%)	-	-	\$120.00

Costo Total Estimado: \$3,850.00 USD

Este cronograma se ha elaborado apoyándose en los efectos conseguidos, objetivos cumplidos y requerimientos definidos del proyecto, afirmando una gestión eficiente y una adecuada asignación de recursos.

11. Conclusiones

Finalizando este proyecto, se demostró que la inteligencia artificial generativa, con modelos de lenguaje visual como GPT-Vision, se transforma la experiencia de aprendizaje en niños de entre 3 y 6 años. Lo que inició como una propuesta ambiciosa se concretó en una aplicación educativa funcional, capaz de atender las necesidades cognitivas de los menores por contenido visual atractivo, actividades interactivas y adaptabilidad continua.

El diseño siguió un enfoque en los niños en etapa preescolar, donde se hicieron prototipos funcionales como accesibilidad, facilidad de uso y estímulo multisensorial. Las pruebas realizadas confirmaron que los niños no solo comprendían la dinámica de la aplicación, sino que además interactuaban con entusiasmo. La integración de tecnologías como el reconocimiento automático del habla (STT) y la conversión de texto a voz (TTS) facilitó una comunicación adecuada con los usuarios. Esta capacidad para “escuchar” y “responder” generó un entorno de aprendizaje cercano, permitiendo que cada niño explorara a su propio ritmo y recibiera retroalimentación inmediata. Los datos recopilados en las pruebas piloto evidenciaron que un 75% de los participantes mostró mayor comodidad con actividades visuales, mientras que un 55% manifestó un interés particular en las funciones de voz. Más allá del impacto pedagógico, el 79,3% de los adultos encuestados expresó disposición a adquirir una aplicación con estas características, indicando una oportunidad real en el mercado.

Finalmente, la aplicación resultó importante para el desarrollo del proyecto, permitiendo responder con agilidad a las preguntas y mantener un proceso de mejora continua. En conjunto, esta experiencia no solo entregó un producto funcional, sino se creó una base sólida para futuras exploraciones en métodos educativos apoyados en tecnologías nuevas. Todas las gráficas correspondientes a la encuesta se incluyen en los anexos.

12.Recomendaciones

Aunque esta aplicación fue diseñada para niños entre 4 y 6 años, sus capacidades técnicas permiten ser más amplias para las aplicaciones educativas. No existen impedimentos para su adaptación a otros, desde adolescentes que aprenden un nuevo idioma hasta adultos que buscan mejorar habilidades de comprensión lectora o auditiva. En este sentido, una evolución lógica es desarrollar módulos específicos que se ajusten a las necesidades de cada grupo. Por ejemplo, podrían implementarse actividades orientadas al aprendizaje del inglés, ejercicios para reforzar la comprensión de instrucciones orales en adultos mayores, o herramientas de apoyo para estudiantes con dificultades en el lenguaje. Actualmente, la aplicación se centra en la generación de imágenes y en preguntas auditivas con respuestas orales. Si bien esta funcionalidad ofrece una experiencia educativa enriquecedora. La incorporación de ejercicios escritos, dinámicas gramaticales o evaluaciones automáticas podría potenciar significativamente el proceso de aprendizaje. Este camino, representa una oportunidad de crecimiento natural para la aplicación y aumentando su impacto educativo.

13. BIBLIOGRAFIA

- Abril, D. (2025, April 8). Americans worry AI is coming for these jobs. The Washington Post. Recuperado el 16 de julio de 2025 desde: https://www.washingtonpost.com/business/2025/04/08/ai-job-loss-work/?utm_source=chatgpt.com
- Chandra, R. (2025, March 3). All You Need To Know About Vision Language Models. Recuperado el 16 de julio de 2025 desde: <https://insights.daffodilsw.com/blog/all-you-need-to-know-about-vision-language-models#:~:text=What%20Are%20the%20Key%20Components,of%20Vision%20Language%20Models>
- Deloitte Insights. (2024, October 24). Four futures of generative AI in the enterprise: Scenario planning for strategic resilience and adaptability. Recuperado el 16 de julio de 2025 desde: https://www.deloitte.com/us/en/insights/topics/digital-transformation/generative-ai-and-the-future-enterprise.html?utm_source=chatgpt.com
- González, L., & Pérez, F. (2020). Retos en la implementación del reconocimiento de voz y la conversión texto a voz en aplicaciones educativas infantiles. *Revista Latinoamericana de Tecnología Educativa*, 19(2), 45-61.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, 27, 2672-2680.

- Hernández, M., Rodríguez, P., & Vázquez, J. (2021). Aplicaciones móviles adaptativas en educación inicial: un análisis pedagógico y tecnológico. *Revista Internacional de Educación y Tecnología*, 7(1), 12-25.
- Hugging Face. (2024). Vision Language Models Explained. Recuperado el 16 de julio de 2025 desde: <https://huggingface.co/blog/vlms#:~:text=Vision%20language%20models%20are%20broadly,detect%20or%20segment%20a%20particular>
- McKinney, D., Johnson, M., & Lee, S. (2022). Advances in generative AI for early childhood education: Visual content generation. *Educational Review*, 74(3), 325-340.
- Pinto, M. C. (2000). Teoría computacional de la mente. *Medicina*, 22(3):188–192.
- Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. *International Conference on Machine Learning (ICML)*, 8748-8763.
- Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., ... & Sutskever, I. (2021). Zero-shot text-to-image generation. *arXiv preprint arXiv:2102.12092*.
- Simanca Gamarra, S. M. (2015). Crítica a la teoría computacional de la mente desde el enfoque neurobiológico de la conciencia de John Searle. Tesis doctoral, Universidad de Cartagena.
- Teigens, V., Skalfist, P., & Mikelsten, D. (2020). *Inteligencia artificial: la cuarta revolución industrial*. Cambridge Stanford Books.
- UNESCO. (2021). De la recuperación del aprendizaje al futuro de la educación. Recuperado el 15 de julio de 2025 desde: <https://www.unesco.org/es/articles/de-la-recuperacion-del-aprendizaje-al-futuro-de-la-educacion-el-trabajo-de-la-unesco-en-2021>
- Vina, A. (2025). Vision Language Models Explained | Ultralytics. Recuperado el 16 de julio de 2025 desde: <https://www.ultralytics.com/es/blog/understanding-vision-language-models-and-their-applications#:~:text=,%E2%80%8D>
- Zhang, W., Yang, J., Li, Y., & Zhang, Y. (2019). Visual language models for educational applications: A survey. *Journal of Educational Technology Systems*, 48(1), 98-116.

13. Anexos

¿Con qué frecuencia usan estas aplicaciones los niños?

29 respuestas

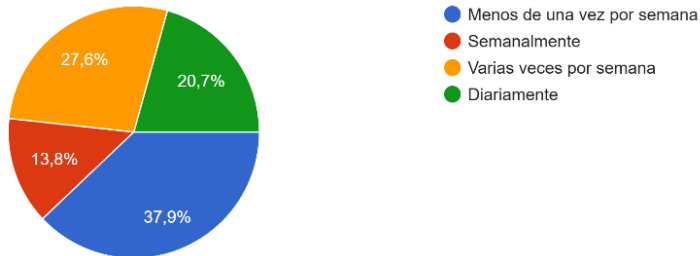


Figura 5. Respuesta de encuesta

¿Qué tipo de actividades educativas prefiere para los niños? (puede elegir más de una)

29 respuestas

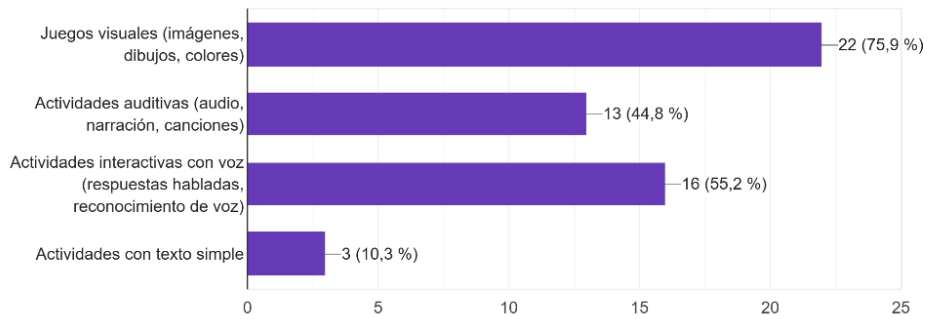


Figura 6. Respuesta de encuesta

¿Qué tan importante es para usted que la aplicación personalice las actividades según el nivel o intereses del niño?

29 respuestas

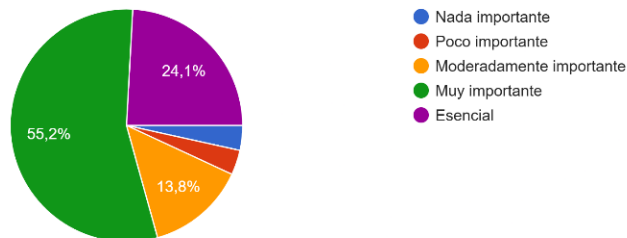


Figura 7. Respuesta de encuesta

¿Cuánto estaría dispuesto a pagar por una aplicación educativa móvil con las características mencionadas?

29 respuestas

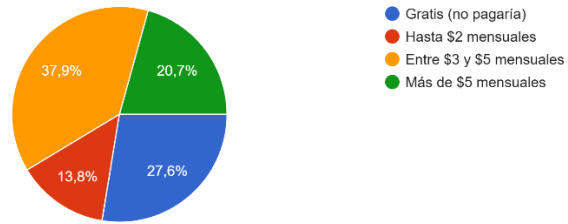


Figura 8. Respuesta de encuesta