



Universidad de San Carlos de Guatemala  
Facultad de Ingeniería  
Escuela de Ingeniería en Ciencias y Sistemas

**MODELOS GENERATIVOS EN CONJUNTO CON EL PROCESAMIENTO DE TEXTO  
NATURAL PARA EL APOYO DE PERSONAS CON DISCAPACIDAD EN CAMPOS COMO  
CIENCIAS Y ARTES**

**Luis Fernando Ralda Estrada**

Asesorado por el Ing. Edwin Estuardo Zapeta Gómez

Guatemala, febrero de 2024



UNIVERSIDAD DE SAN CARLOS DE GUATEMALA



FACULTAD DE INGENIERÍA

**MODELOS GENERATIVOS EN CONJUNTO CON EL PROCESAMIENTO DE TEXTO  
NATURAL PARA EL APOYO DE PERSONAS CON DISCAPACIDAD EN CAMPOS COMO  
CIENCIAS Y ARTES**

TRABAJO DE GRADUACIÓN

PRESENTADO A LA JUNTA DIRECTIVA DE LA  
FACULTAD DE INGENIERÍA

POR

**LUIS FERNANDO RALDA ESTRADA**

ASESORADO POR EL ING. EDWIN ESTUARDO ZAPETA GÓMEZ

AL CONFERÍRSELE EL TÍTULO DE

**INGENIERO EN CIENCIAS Y SISTEMAS**

GUATEMALA, FEBRERO DE 2024



UNIVERSIDAD DE SAN CARLOS DE GUATEMALA  
FACULTAD DE INGENIERÍA



**NÓMINA DE JUNTA DIRECTIVA**

DECANO	Ing. José Francisco Gómez Rivera (a. i.)
VOCAL II	Ing. Mario Renato Escobedo Martínez
VOCAL III	Ing. José Milton de León Bran
VOCAL IV	Ing. Kevin Vladimir Cruz Lorente
VOCAL V	Ing. Fernando José Paz González
SECRETARIO	Ing. Hugo Humberto Rivera Pérez

**TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO**

DECANO	Ing. José Francisco Gómez Rivera (a. i.)
EXAMINADOR	Ing. Marlon Francisco Orellana López
EXAMINADOR	Ing. Gabriel Alejandro Díaz López
EXAMINADORA	Inga. Claudia Liceth Rojas Morales
SECRETARIO	Ing. Hugo Humberto Rivera Pérez

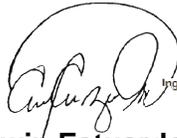
Guatemala, 24 de noviembre de 2023

Ingeniero  
**Carlos Alfredo Azurdía**  
**Coordinador de Privados y Trabajos de Tesis**  
**Escuela de Ingeniería en Ciencias y Sistemas**  
**Facultad de Ingeniería - USAC**

Respetable Ingeniero Azurdía:

Por este medio hago de su conocimiento que doy el visto bueno del trabajo de investigación del estudiante **Luis Fernando Ralda Estrada** con carné **201709065** y CUI **3004735750101** titulado **“MODELOS GENERATIVOS EN CONJUNTO CON EL PROCESAMIENTO DE TEXTO NATURAL PARA EL APOYO DE PERSONAS CON DISCAPACIDAD, EN CAMPOS COMO CIENCIAS Y ARTES”**, ya que cumple con los objetivos establecidos de manera satisfactoria.

Atentamente,



Ing. Estuardo Zapeta  
Ingeniería en Ciencias y Sistemas  
Colegiado 12767

**Ing. Edwin Estuardo Zapeta Gómez**  
Colegiado No. 12767



Universidad San Carlos de Guatemala  
Facultad de Ingeniería  
Escuela de Ingeniería en Ciencias y Sistemas

Guatemala 27 de noviembre de 2023

Ingeniero  
**Carlos Gustavo Alonzo**  
Director de la Escuela de Ingeniería  
En Ciencias y Sistemas

Respetable Ingeniero Alonzo:

Por este medio hago de su conocimiento que he revisado el trabajo de graduación del estudiante **LUIS FERNANDO RALDA ESTRADA** con carné **201709065** y CUI **3004 73575 0101** titulado **“MODELOS GENERATIVOS EN CONJUNTO CON EL PROCESAMIENTO DE TEXTO NATURAL PARA EL APOYO DE PERSONAS CON DISCAPACIDAD, EN CAMPOS COMO CIENCIAS Y ARTES”**, y a mi criterio el mismo cumple con los objetivos propuestos para su desarrollo, según el protocolo aprobado.

Al agradecer su atención a la presente, aprovecho la oportunidad para suscribirme,

Atentamente,



**Ing. Carlos Alfredo Azurdia**  
Coordinador de Privados  
y Revisión de Trabajos de Graduación

UNIVERSIDAD DE SAN CARLOS  
DE GUATEMALA



FACULTAD DE INGENIERÍA

SIST.LNG.DIRECTOR.5.EICCSS.2024

El Director de la Escuela de Ingeniería en Ciencias y Sistemas de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer el dictamen del Asesor, el visto bueno del Coordinador de área y la aprobación del área de lingüística del trabajo de graduación titulado: **MODELOS GENERATIVOS EN CONJUNTO CON EL PROCESAMIENTO DE TEXTO NATURAL PARA EL APOYO DE PERSONAS CON DISCAPACIDAD EN CAMPOS COMO CIENCIAS Y ARTES**, presentado por: **Luis Fernando Ralda Estrada**, procedo con el Aval del mismo, ya que cumple con los requisitos normados por la Facultad de Ingeniería.

“ID Y ENSEÑAD A TODOS”



Ingeniero Carlos Gustavo Alonzo  
DIRECTOR  
Escuela de Ingeniería en Ciencias y Sistemas

Guatemala, febrero de 2024

Ingeniería Civil, Ingeniería Mecánica Industrial, Ingeniería Química, Ingeniería Mecánica Eléctrica, -Escuela de Ciencias, Regional de Ingeniería Sanitaria y Recursos Hidráulicos (ERIS), Maestría en Sistemas Mención construcción y Mención Ingeniería Vial. Carreras: Ingeniería Mecánica, Ingeniería Electrónica, Ingeniería en Ciencias y Sistemas, Licenciatura en Matemática, Licenciatura en Física. Centros: de Estudios Superiores de Energía y Minas (CESEM). Guatemala, Ciudad Universitaria, Zona 12, Guatemala, Centroamérica



**USAC**  
TRICENTENARIA  
Universidad de San Carlos de Guatemala

Decanato  
Facultad e Ingeniería

24189101- 24189102

LNG.DECANATO.OIE.100.2024

El Decano de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer la aprobación por parte del Director de la Escuela de Ingeniería en Ciencias y Sistemas, al Trabajo de Graduación titulado: **MODELOS GENERATIVOS EN CONJUNTO CON EL PROCESAMIENTO DE TEXTO NATURAL PARA EL APOYO DE PERSONAS CON DISCAPACIDAD EN CAMPOS COMO CIENCIAS Y ARTES**, presentado por: **Luis Fernando Ralda Estrada** después de haber culminado las revisiones previas bajo la responsabilidad de las instancias correspondientes, autoriza la impresión del mismo.

IMPRÍMASE:

Firmado electrónicamente por: José Francisco Gómez Rivera  
Motivo: Orden de impresión  
Fecha: 08/02/2024 09:02:56  
Lugar: Facultad de Ingeniería, USAC.

Ing. José Francisco Gómez Rivera  
Decano a.i.



Guatemala, febrero de 2024

Para verificar validez de documento ingrese a <https://www.ingenieria.usac.edu.gt/firma-electronica/consultar-documento>

Tipo de documento: Correlativo para orden de impresión Año: 2024 Correlativo: 100 CUI: 3004735750101

Escuelas: Ingeniería Civil, Ingeniería Mecánica Industrial, Ingeniería Química, Ingeniería Mecánica Eléctrica, - Escuela de Ciencias, Regional de Ingeniería Sanitaria y Recursos Hidráulicos (ERIS). Postgrado Maestría en Sistemas Mención Ingeniería Vial. Carreras: Ingeniería Mecánica, Ingeniería Electrónica, Ingeniería en Ciencias y Sistemas. Licenciatura en Matemática. Licenciatura en Física. Centro de Estudios Superiores de Energía y Minas (CESEM). Guatemala, Ciudad



## **HONORABLE TRIBUNAL EXAMINADOR**

En cumplimiento con los preceptos que establece la ley de la Universidad de San Carlos de Guatemala, presento a su consideración mi trabajo de graduación titulado:

**MODELOS GENERATIVOS EN CONJUNTO CON EL PROCESAMIENTO DE TEXTO  
NATURAL PARA EL APOYO DE PERSONAS CON DISCAPACIDAD EN CAMPOS COMO  
CIENCIAS Y ARTES**

Tema que me fuera asignado por la Dirección de la Escuela de Ingeniería en Ciencias y Sistemas, con fecha 20 de agosto de 2022.



**Luis Fernando Ralda Estrada**



## **ACTO QUE DEDICO A:**

### **Mis padres**

Sergio Ralda y Flor Estrada, por su apoyo y amor incondicional. Les debo todo lo que soy.

### **Mis abuelos**

Luis Ralda, Irma Orantes y Gloria Moreno, por siempre creer en mí. Su amor lo siento todos los días.

### **Mi hermano**

Por siempre recordarme lo mucho que tengo para dar. Sé que siempre puedo contar contigo.

### **Mi novia**

Sofía García, por ser para mí, un constante apoyo y alguien en quien puedo confiar.



## **AGRADECIMIENTOS A:**

<b>Universidad de San Carlos de Guatemala</b>	Por forjar las bases de mi educación superior.
<b>Facultad de Ingeniería</b>	Por brindarme las herramientas para ser un profesional exitoso y autodidacta.
<b>Mis amigos de la Facultad</b>	Alejandro García, Max Florián, Cristian Castellanos y Paola Casiano, por ser un alivio y un soporte increíble durante mi etapa como estudiante, además de unos asombrosos amigos.
<b>Mi familia</b>	Por su invaluable apoyo y confianza.
<b>Mi jefe y un gran amigo</b>	Boris Lemus, por confiar en mí, darme la oportunidad de un gran empleo, y ser alguien en quien siempre puedo confiar.
<b>Mi asesor de tesis</b>	Ing. Estuardo Zapeta, por su invaluable apoyo y retroalimentación durante la realización de este trabajo.



## ÍNDICE GENERAL

ÍNDICE DE ILUSTRACIONES.....	V
GLOSARIO .....	VII
RESUMEN.....	IX
OBJETIVOS.....	XI
INTRODUCCIÓN .....	XIII
1. PLANTEAMIENTO DEL PROBLEMA .....	1
1.1. Personas con discapacidad y las herramientas que comúnmente utilizan para llevar a cabo sus actividades .....	1
1.2. Inteligencia artificial y su aporte en temas de discapacidad .....	2
1.3. Consecuencias colaterales de una discapacidad .....	3
1.4. Desafíos que debe enfrentar una persona con discapacidad....	4
1.5. Bondades que ofrece la inteligencia artificial para la inclusión de personas con discapacidad .....	6
1.6. Modelos generativos y procesamiento de lenguaje natural en el contexto de la discapacidad.....	7
1.7. Pregunta a la que esta investigación busca dar respuesta.....	9
2. ESTUDIO DE LAS REDES NEURONALES Y SU EVOLUCIÓN .....	11
2.1. Redes neuronales y su relación con modelos generativos.....	11
2.2. Procesamiento de texto natural, base teórica, importancia y aplicaciones.....	18
2.3. Otras aplicaciones de las redes neuronales .....	21
2.4. Profundización en el funcionamiento de soluciones ya existentes .....	22

2.4.1.	Asistentes de voz (Siri, Google Assistant y Alexa)	22
2.4.2.	Aplicaciones para usos específicos.....	24
3.	PRESENTACIÓN DE RESULTADOS.....	26
3.1.	Aplicación de dibujo por capas.....	27
3.1.1.	Descripción del funcionamiento.....	27
3.1.2.	Diseño técnico de la aplicación .....	35
3.1.3.	Análisis de riesgo .....	36
3.1.4.	Especificaciones de desarrollo .....	37
3.2.	Asistente para la movilización en la vía pública .....	37
3.2.1.	Descripción del funcionamiento.....	38
3.2.2.	Diseño técnico de la aplicación .....	44
3.2.3.	Análisis de riesgo .....	46
3.2.4.	Especificaciones de desarrollo .....	46
3.3.	Aplicación para la traducción de lenguaje de señas .....	47
3.3.1.	Descripción del funcionamiento.....	48
3.3.2.	Diseño técnico de la aplicación .....	56
3.3.3.	Análisis de riesgo .....	58
3.3.4.	Especificaciones de desarrollo .....	59
3.4.	Aplicación de asistencia para la lectura en tiempo real.....	60
3.4.1.	Descripción del funcionamiento.....	60
3.4.2.	Diseño técnico de la aplicación .....	66
3.4.3.	Análisis de riesgo .....	68
3.4.4.	Especificaciones de desarrollo .....	68
3.5.	Aplicación de comunicación para personas con discapacidad del habla.....	69
3.5.1.	Descripción del funcionamiento.....	70
3.5.2.	Análisis de riesgo .....	77

3.5.3.	Especificaciones de desarrollo .....	78
4.	DISCUSIÓN DE RESULTADOS .....	81
4.1.	Aplicación de dibujo por capas .....	81
4.2.	Asistente para la movilización en la vía pública.....	82
4.3.	Aplicación para traducción de lenguaje de señas.....	84
4.4.	Aplicación de asistencia para la lectura en tiempo real .....	85
4.5.	Aplicación de comunicación para personas con discapacidad del habla .....	87
	CONCLUSIONES .....	91
	RECOMENDACIONES .....	93
	REFERENCIAS .....	95



# ÍNDICE DE ILUSTRACIONES

## FIGURAS

<b>Figura 1.</b>	Aprendizaje supervisado.....	14
<b>Figura 2.</b>	Aprendizaje bajo refuerzo .....	15
<b>Figura 3.</b>	Interfaz aplicación de dibujo por capas .....	28
<b>Figura 4.</b>	Modo esperando instrucción .....	29
<b>Figura 5.</b>	Solicitud procesada con éxito .....	30
<b>Figura 6.</b>	Solicitud errónea .....	31
<b>Figura 7.</b>	Circulo verde en la mitad del lienzo .....	33
<b>Figura 8.</b>	Cielo y tierra .....	33
<b>Figura 9.</b>	Dibujo de un arbol .....	34
<b>Figura 10.</b>	Interfaz aplicación asistente visual.....	38
<b>Figura 11.</b>	Solicitud de permisos .....	40
<b>Figura 12.</b>	Persona siendo detectada .....	41
<b>Figura 13.</b>	Confirmación de bienestar .....	43
<b>Figura 14.</b>	Contactando a emergencias .....	44
<b>Figura 15.</b>	Solicitud de acceso a cámara .....	48
<b>Figura 16.</b>	Menú principal traductor de lenguaje de señas.....	49
<b>Figura 17.</b>	Interfaz de grabación .....	50
<b>Figura 18.</b>	Pantalla de espera y pantalla de confirmación.....	52
<b>Figura 19.</b>	Menú de retroalimentación.....	53
<b>Figura 20.</b>	Reproducción de la traducción.....	54
<b>Figura 21.</b>	Historial de grabaciones.....	55
<b>Figura 22.</b>	Grabaciones favoritas .....	56
<b>Figura 23.</b>	Solicitud de acceso a cámara .....	61

<b>Figura 24.</b>	Pantalla de inicio.....	62
<b>Figura 25.</b>	Pantalla de escaneo de texto .....	62
<b>Figura 26.</b>	Pantalla de espera de procesamiento .....	64
<b>Figura 27.</b>	Pantalla de resultados del escaneo .....	65
<b>Figura 28.</b>	Pantalla de reproducción del texto.....	65
<b>Figura 29.</b>	Solicitud de acceso a micrófono .....	70
<b>Figura 30.</b>	Apartado de práctica.....	71
<b>Figura 31.</b>	Menú principal voz a texto .....	73
<b>Figura 32.</b>	Pantalla de interpretación básica.....	74
<b>Figura 33.</b>	Pantalla interpretar en tiempo real.....	76
<b>Figura 34.</b>	Historial de interpretaciones básicas .....	77

## GLOSARIO

<b>Bison</b>	Generador de analizadores sintácticos de uso general que convierte gramáticas libres de contexto en analizadores deterministas.
<b>CUP</b>	Por sus siglas en inglés, Construction of Useful Parsers, es una herramienta de computación para la creación de analizadores sintácticos, utilizado comúnmente en el campo de los compiladores.
<b>FLEX</b>	Por sus siglas en inglés, <i>Fast Lexical Analyzer</i> , es una herramienta de computación para generar analizadores léxicos, comúnmente utilizado en el campo de los compiladores.
<b>GPS</b>	El Sistema de Posicionamiento Global, es un sistema que permite localizar cualquier objeto en la Tierra.
<b>LENSEGUA</b>	Lengua oficial para poder comunicarse en lenguaje de señas en Guatemala.
<b>Lookout</b>	Herramienta creada por Google que ayuda a personas con discapacidad visual brindando retroalimentación del entorno utilizando visión por computadora.

<b>Machine Learning</b>	Disciplina del campo de la inteligencia artificial que provee a las computadoras de la capacidad de identificar patrones y elaborar predicciones.
<b>OCR</b>	Optical character recognition, es una herramienta de la visión por computadora que permite transformar imágenes en texto.
<b>PLN</b>	Procesamiento de lenguaje natural. Área de la informática que se centra en las interacciones entre personas y computadoras utilizando el lenguaje hablado.
<b>RPN</b>	<i>Region Proposal Network</i> . Red neuronal que al mismo tiempo predice los vínculos entre objetos, así como que tan segura es la predicción de ser ese objeto.
<b><i>Speech-to-text</i></b>	Sistema informático que permite transformar automáticamente la voz humana en texto.
<b><i>Text-to-speech</i></b>	Sistema informático que permite producir voz a partir un texto escrito.
<b>Wifi</b>	Tecnología que permite la interconexión inalámbrica de dispositivos electrónicos.

## RESUMEN

Este estudio se enfoca en el potencial de los modelos generativos y el procesamiento de texto natural para mejorar la accesibilidad de personas con discapacidad. A través de cinco proyectos conceptuales, se exploran aplicaciones prácticas que abordan desafíos en comunicación, movilidad y expresión artística.

Los proyectos, que van desde herramientas de dibujo por capas hasta asistentes para la movilización en espacios públicos, ofrecen soluciones innovadoras y mejoras en la autonomía de las personas con discapacidad. Se identifican puntos fuertes y áreas de mejora en cada propuesta, destacando la necesidad de colaboraciones interdisciplinarias y la exploración de nuevas arquitecturas neuronales.

Las recomendaciones subrayan la importancia de implementar proyectos reales basados en estos modelos, enfocándose en la inclusión, estableciendo alianzas estratégicas y desarrollando programas de capacitación para su efectiva implementación.

En resumen, este estudio enfatiza el impacto positivo de la tecnología en la vida diaria de las personas con discapacidad, instando a seguir desarrollando soluciones inclusivas y adaptadas a necesidades específicas.



## **OBJETIVOS**

### **General**

Analizar tanto la base teórica de los modelos generativos como sus aplicaciones y el pragmatismo de estas para el apoyo de personas con discapacidad en sectores como las ciencias y artes

### **Específicos**

1. Estudiar la evolución de las redes neuronales y su uso para el desarrollo de modelos generativos y procesamiento de texto natural para la comprensión de estas.
2. Contrastar proyectos presentados en tiempos recientes sobre modelos generativos mediante un análisis de capacidades para determinar sus posibles aplicaciones reales
3. Presentar una guía o compendio de proyectos concepto a poder realizarse utilizando estas tecnologías, además de sus características y cómo estas pueden ser de ayuda para personas con discapacidad



## INTRODUCCIÓN

En este trabajo se explora cómo la inteligencia artificial puede ayudar a superar desafíos para las personas con discapacidad, especialmente en campos cruciales como la ciencia y el arte. Se comienza abordando el día a día de las personas con discapacidad, sus herramientas comunes y las dificultades que enfrentan. También se destaca, cómo la inteligencia artificial puede ser una gran ayuda para incluir a estas personas en la sociedad.

En el estudio de las redes neuronales, se ve cómo evolucionan e influyen en el desarrollo de modelos generativos. Se explora el procesamiento de texto natural y sus diversas aplicaciones, proporcionando la base teórica necesaria para entender estas soluciones.

En la sección de resultados, se presentan cinco proyectos concepto, diseñados para satisfacer necesidades específicas en áreas como el arte y la ciencia. Desde una aplicación de dibujo hasta un asistente para moverse en la calle, cada proyecto se explica detalladamente, desde cómo funciona hasta su diseño técnico.

La discusión de resultados evalúa críticamente cada proyecto, comparándolo con soluciones existentes y considerando su utilidad real en la vida diaria. Las conclusiones y recomendaciones resumen los hallazgos, destacando cómo la inteligencia artificial puede ser una aliada valiosa para lograr una sociedad más inclusiva y accesible para todas las personas, independientemente de sus habilidades.



# 1. PLANTEAMIENTO DEL PROBLEMA

## 1.1. Personas con discapacidad y las herramientas que comúnmente utilizan para llevar a cabo sus actividades

- Personas con discapacidad auditiva
  - Reproductores y grabadoras de sonido. En caso de tener un impedimento leve, es común el uso de reproductores para la amplificación del sonido.
  - Dispositivos *speech-to-text*: proveen de una funcionalidad que convierten el audio en texto, para poder leerlo en caso no poder escuchar.
- Personas con discapacidad del habla
  - Dispositivos *text-to-speech*: convierte texto a audio natural, perfecto para darse a entender en caso sea imposible hablar.
  - Servicio de mensajería instantánea: para los casos de uso donde el habla no es fundamental, pero es necesario comunicarse.
- Personas con discapacidad visual
  - Aplicaciones para procesar el texto de una imagen y convertirlo en audio.

- Servicios para el apoyo de personas con desordenes de lectura y escritura, como dislexia.
- Bastones, para poder localizar objetos e infraestructura cercana a ellos.
- Personas con diversas discapacidades físicas
  - Dispositivos manos libres: perfectos para la comunicación sin la necesidad de interactuar físicamente con un dispositivo.

## **1.2. Inteligencia artificial y su aporte en temas de discapacidad**

Al ver como las herramientas previamente mencionadas se comparan con herramientas que la inteligencia artificial puede brindar, se concluye lo siguiente: Como bien se mencionó antes, las herramientas al alcance de la mayoría de las personas están limitadas por las funcionalidades que ofrecen, o porque se enfocan en satisfacer una necesidad en particular.

Un dispositivo de amplificación de sonidos es perfectamente funcional, pero está limitado al rango auditivo de la persona y no provee una solución inmediata para la sordera absoluta. Por el contrario, una solución que anticipa su escalabilidad puede utilizar inteligencia artificial para mantener la funcionalidad original, como para convertir ese audio en texto, traducirlo, resumirlo, entre muchas otras.

El mismo caso puede resultar de analizar un dispositivo clásico utilizado comúnmente para ayudar a estas personas, con cualquier solución que la inteligencia artificial pueda brindar. El único aspecto donde los dispositivos

clásicos ganarían sería en el tiempo de desarrollo que, en su caso, es nulo. Pero si se ve más allá de una solución que solo sirva para pasar el rato, se puede comprender por qué las soluciones que la inteligencia artificial puede brindar son tanto mucho más personalizables como funcionales y escalables.

### **1.3. Consecuencias colaterales de una discapacidad**

Desde el año 2000, tan solo en Estados Unidos, el número de trabajadores que presentan alguna discapacidad incrementó en un 35 % de acuerdo con información provista por el gobierno. En ese mismo sentido, la salud financiera del empleado promedio ha disminuido ya que muchos de ellos tienen que requerir a utilizar ahorros de toda la vida o entrar en deuda para poder satisfacer sus necesidades tanto físicas como mentales. Aunque a simple vista no resalte, el daño en la economía que una discapacidad presenta es notable y no parece que vaya a disminuir hasta que se posea un sistema de salud público y gratuito (Namkung, 2020).

Estas limitaciones han generado una gran carga emocional sobre el individuo, especialmente si pasó parte de su vida sin esta discapacidad. Aquellos que están sanando de un golpe o lesión, especialmente cuando esta resulta en una discapacidad permanente, tienen que lidiar con el lado más emotivo al darse cuenta de que nunca podrán correr de nuevo, o jugar con sus hijos, abrazar a sus padres o siquiera caminar. Depresión, ira y culpa son todos efectos colaterales de una discapacidad y deben ser tratados con paciencia y consideración, estas emociones negativas deben ser tratadas para poder pensar en algún día volver a ser quién algún día fue el individuo.

Por último, y el que más concierne a esta investigación, consecuencias a nivel comunidad, donde el individuo ya no es visto de la misma manera por la

sociedad en la que vive, incluso por sus propios amigos. Esta discapacidad puede afectar como realiza su trabajo de toda la vida, o como una persona se le acerca a hablarle por la calle. Esto es especialmente problemático cuando el individuo se siente molesto todo el tiempo y no quiere ningún tipo de compañía. También presente cuando la impotencia que se construye conduce al individuo a un estado emocional que afecta su salud mental (Howard, 2021).

#### **1.4. Desafíos que debe enfrentar una persona con discapacidad**

Una persona que posea una discapacidad o esté lidiando con un problema de esta naturaleza enfrenta diversos retos al momento de querer ser parte activa de la sociedad; En especial cuando el formar parte involucra actividades motrices dentro de los requerimientos de cierto trabajo o actividad. Y en determinadas situaciones, son incluso descartados por su incapacidad de aportar en áreas específicas, como salud o la industria.

Primero, existen retos o barreras de carácter actitudinal. Estas son las más básicas, como por ejemplo las dificultades en acceder o entrar a cierto lugar por las condiciones del terreno o la falta de características que aumenten la accesibilidad del lugar. Estas barreras traen consigo 2 problemas principales:

- Estereotipos dañinos: donde se asume que estas personas no tienen una buena calidad de vida o que no pueden participar de algunas actividades por su discapacidad.
- Estigmas, discriminación y prejuicios: dentro de la sociedad, estas actitudes están relacionadas a la mentalidad de las personas, donde se piensa que las personas con discapacidad son una tragedia y deberían

estar alejadas del resto, aludiendo a un castigo por su falta de habilidad para realizar estas actividades.

También se tienen barreras en la comunicación. Estas son experimentadas principalmente por personas que poseen una discapacidad de carácter físico, como lo es un impedimento de audición o habla. Este tipo de personas pueden poseer problemas al momento de querer realizar ciertas acciones como:

- Visualizar carteles con información de seguridad importante, donde las personas con impedimento de vista no puedan recibir dicho mensaje.
- El uso de lenguaje técnico, que personas con impedimento del habla no puedan vocalizar por utilizar demasiadas sílabas, o porque realmente no pueden decir absolutamente nada.
- Mensajes importantes transmitidos a través de canales auditivos como parlantes no llegarán a personas con impedimento de la audición, poniéndolos en riesgo.

Por último, una barrera que es de carácter muy importante para esta investigación, las barreras creadas por políticas, estas incluyen:

- Denegar a personas calificadas el participar en programas, servicios o trabajos por el único motivo de poseer una discapacidad.
- Denegar a individuos el acceso a programas, servicios y beneficios por barreras físicas en el ambiente de desarrollo de dicho servicio.

- Denegar al individuo comodidades razonables para realizar sus funciones esenciales en el trabajo al cual han sido contratados para realizar (Howard, 2021).

### **1.5. Bondades que ofrece la inteligencia artificial para la inclusión de personas con discapacidad**

Como se mencionó en una sección anterior, la inteligencia artificial es una herramienta muy importante que desde hace tiempo forma parte de la vida de las personas con discapacidades, ofreciendo una nueva forma de realizar acciones que antes les eran imposibles, o de volver a realizar acciones que alguna vez perdieron la capacidad de realizar.

Desde traductores de lenguaje de señas hasta acompañantes para situaciones problemáticas, la inteligencia artificial no para de proveer herramientas y soluciones que apoyan el día a día de estas personas, y su mayor inclusión en una sociedad digital como la que se tiene hoy en día.

Con gran accesibilidad en todas ellas, estas herramientas buscan como propósito principal el ayudar a las personas a tener un estilo de vida digno, donde puedan ganarse la vida de la manera que la hace cualquier otra persona y no necesitar de subsidios gubernamentales, mucho menos lástima de colegas y compañeros de trabajo (Berman, 2017).

## **1.6. Modelos generativos y procesamiento de lenguaje natural en el contexto de la discapacidad**

Según análisis de las aplicaciones de la inteligencia artificial para la ayuda a personas con discapacidad, se pueden observar los siguientes resultados:

Para empezar, se tienen los asistentes virtuales, véase Alexa, Siri o Google Assistant. Estos asistentes trabajan tanto de forma textual como por voz, permitiendo un nivel de accesibilidad muchísimo mayor, en especial para personas que no cuentan con la capacidad de poder ingresar a sus teléfonos celulares y presionar botones.

El poder activar estos asistentes con tan solo decir una palabra clave es fundamental para que estas personas puedan interactuar con los sistemas, para luego hacer llamadas, programar eventos o incluso realizar tareas que les serían imposibles de otra manera como el encender las luces de la casa o cerrar las cortinas. Estos sistemas se apoyan principalmente de la domótica y el volver tu casa inteligente. Pero nada de ello sería posible para estas personas sin la capacidad de activar la funcionalidad con tan solo hablar.

Ahora bien, al ver el otro lado de la moneda, ¿Qué pasa si el sujeto si puede utilizar sus facultades físicas, pero en cambio no puede hablar? Las personas con discapacidad del habla, entonces, no serían capaces de utilizar estas funcionalidades, a menos que también puedan ser ordenadas de alguna otra manera. Y ahí es donde estos sistemas brillan, en su polivalencia, porque son capaces de funcionar tanto por comandos de voz como por texto ingresado desde algún dispositivo inteligente con dichas capacidades.

Otra aplicación es “Lookout” de Google, que tiene como objetivo principal ayudar a personas con discapacidad visual en momentos que realmente necesitan poder leer un documento, entender un texto, entre otras tareas. Esta aplicación funciona a través del análisis de texto por medio de la cámara de un dispositivo móvil. Lo único que la persona tiene que hacer es apuntar hacia el documento hasta que la aplicación tome una foto donde se distinga todo el texto, e inmediatamente recibirán retroalimentación en forma de audio con el cual pueden escuchar todo el texto escrito en el documento escaneado.

En casos más extremos estas aplicaciones también cuentan con el apoyo de voluntarios a los cuales las personas no videntes pueden llamar y preguntarles por algo que quieren ver, apuntando la cámara y recibiendo retroalimentación todo el tiempo por parte de la persona del otro lado de la aplicación. De forma que todo es muy humano y requiere de poca interacción por parte del usuario principal.

Parece que las aplicaciones que realmente son de ayuda a personas con diferentes tipos de discapacidad ofrecen un alto nivel de accesibilidad, tanto texto como voz, además, son fáciles de usar, y solventan problemas mayores para estas personas.

Aquí es donde tienen relación los modelos generativos y los sistemas de procesamiento de lenguaje natural. ¿Qué pasaría si las personas sin acceso a capacidades motrices pudiesen generar cualquier tipo de imagen para poder expresarse de manera más precisa en conversaciones? El resto de las personas poseen otro tipo de herramientas como la gesticulación o la ejemplificación, pero una persona no vidente jamás podrá tratar de explicarle a otra persona una labor que tiene que hacer si esta conlleva algo que él jamás ha visto en su vida y tan solo ha escuchado como es.

Las posibilidades son casi infinitas con la velocidad a la que estas tecnologías evolucionan, y el proveer de una nueva oportunidad o experiencia a este tipo de personas es una recompensa en sí misma. Las tecnologías están a la disposición, y en el último capítulo de esta investigación se presentará un compendio que pretende proveer con proyectos con capacidad real de implementación y un esbozo de su realización.

### **1.7. Pregunta a la que esta investigación busca dar respuesta**

La indagación central que guiará este trabajo consiste en explorar cómo la inteligencia artificial puede ser efectivamente utilizada para proporcionar apoyo a personas con discapacidad. Se examinará tanto la base teórica de los modelos generativos como sus aplicaciones prácticas, evaluando su viabilidad en distintos sectores para el respaldo a este grupo.

Para alcanzar una comprensión exhaustiva de la base teórica que sustenta a las redes neuronales y la inteligencia artificial, así como sus ramificaciones relevantes para esta investigación, se abordará la evolución de las redes neuronales.

Este análisis se centrará en su contribución al desarrollo de modelos generativos y procesamiento de texto natural, facilitando una comprensión más profunda de estos fundamentos.

Una pregunta crucial se enfocará en cómo se puede verificar la existencia de aplicaciones reales para tecnologías como los modelos generativos. Este proceso implicará contrastar proyectos recientes que han empleado modelos

generativos, evaluando sus capacidades y determinando de manera crítica sus posibles aplicaciones prácticas en diversos contextos.

Por último, si los modelos generativos son una tecnología que no está al alcance de todos y sus aplicaciones conllevan mucho tiempo y esfuerzo, se hace la pregunta.

- ¿Qué producto puede ser realizado para de igual manera proveer de algo útil y demostrativo?

Una guía o compendio de proyectos concepto para poder realizar, utilizando estas tecnologías, además de sus características y cómo estas pueden ser de ayuda para personas con discapacidad.

## 2. ESTUDIO DE LAS REDES NEURONALES Y SU EVOLUCIÓN

### 2.1. Redes neuronales y su relación con modelos generativos

Las redes neuronales reflejan el comportamiento del cerebro humano, esto permite que computadoras tengan la habilidad de reconocer patrones y resolver problemas en campos como inteligencia artificial, aprendizaje de máquina y aprendizaje profundo.

Pero, en esencia, ¿Qué es una red neuronal? Las redes neuronales, también conocidas como redes neuronales artificiales o redes neuronales simuladas, son un subconjunto del aprendizaje de máquina y son el núcleo de los algoritmos de aprendizaje profundo. Su nombre y estructura están inspirados por el cerebro humano, imitando la manera en la que biológicamente las neuronas reales se comunican unas con otras.

Para poder entender el funcionamiento a profundidad de una red neuronal es necesario que se comprenda parte de su historia y como fueron concebidas. A pesar de ser un tema bastante novedoso y del que no se tenía mucha información hasta hace pocos años, las redes neuronales vienen desde hace mucho más de lo que se puede pensar.

En *A logical calculus of the ideas immanent in nervous activity*, se pretendía entender como un ente tan complejo como lo es el cerebro humano, puede producir patrones complejos a través de la conexión de células o

neuronas. Gracias a este estudio se pudo establecer la relación del cerebro humano con la lógica computacional (Pitts & Warren, 1943).

Uno de los pilares fundamentales en el entendimiento de redes neuronales y aprendizaje de máquina se dio en *The Perceptron: a probabilistic model for information storage and organization in the brain*. En ella se discuten temas como la actividad neuronal nerviosa y provee de un factor antes no tomado en cuenta, los pesos en la ecuación de decisión, abriendo caminos nunca antes explorados (Rosenblatt, 1958).

Luego, una de las primeras aplicaciones reales fue implementada, al demostrar como el uso de restricciones podía ser utilizada para el entrenamiento de algoritmos. En 1989, el servicio postal de los Estados Unidos utilizó una red neuronal que identificaba los códigos postales escritos a mano en las cartas y paquetes, todo gracias a Yan LeCun y sus métodos de propagación inversa y la integración con la arquitectura de las redes neuronales (Lecun, 1992).

Si se piensa por un instante en como el cerebro aprende los aspectos más fundamentales de la supervivencia durante una edad temprana se ve que había un patrón sencillo: el cerebro realiza una acción y el medio ambiente devuelve un estímulo o respuesta a esa acción. Por ejemplo: un niño pequeño llora por lo que los padres de este niño se alteran y le brindan de diversas cosas para calmar su llanto, ya sea comida, cobijo, atención, entre otros.

De esta manera el cerebro genera asociaciones entre acciones y reacciones. De la misma manera que el cerebro, una red neuronal aprende como si estuviera desarrollándose, siendo imposible que esta sepa cómo realizar cierta acción sin antes haberle mostrado cómo realizar esa acción.

La forma correcta de entrenar una de estas redes neuronales es la misma que al criar un niño: adaptando el estímulo de respuesta acorde a la manera que esta se comporta con cierta acción realizada, o en este caso, un *input* o entrada en específico.

Para que una red neuronal pueda “aprender” existen 3 métodos o paradigmas por seguir:

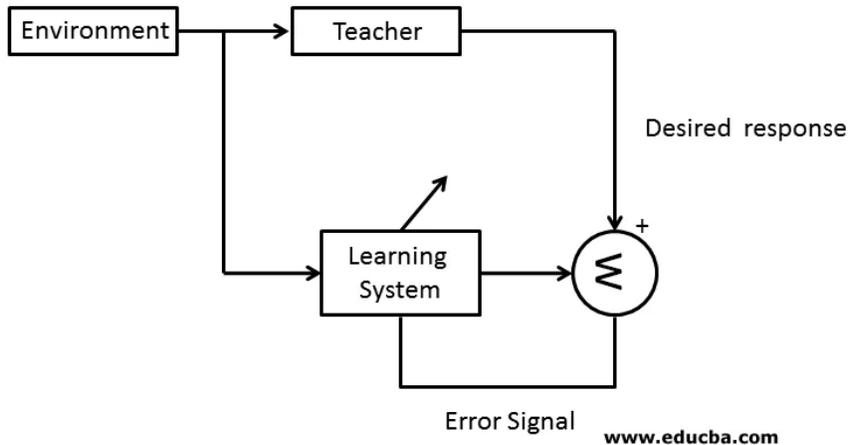
- Aprendizaje supervisado
- Aprendizaje con refuerzo
- Aprendizaje no supervisado

El aprendizaje supervisado como su nombre lo sugiere, es aquel que tiene la presencia de un instructor o maestro. Esto significa que se tiene un set de información etiquetada previamente, además de la reacción esperada a esa información. La red es luego dada nuevos sets de datos para analizar y poder producir la reacción deseada nuevamente. Es un sistema cerrado con retroalimentación, pero el ambiente no se encuentra dentro del bucle de aprendizaje.

En la Figura 1 se puede observar el diagrama del flujo de una red neuronal que trabaja con aprendizaje supervisado. Esta incluye elementos como “Environment”: el medio ambiente, “Teacher”: El instructor, o quien provee la entrada y retroalimentación, y “Learning System”: la red neuronal.

**Figura 1.**

*Aprendizaje supervisado*



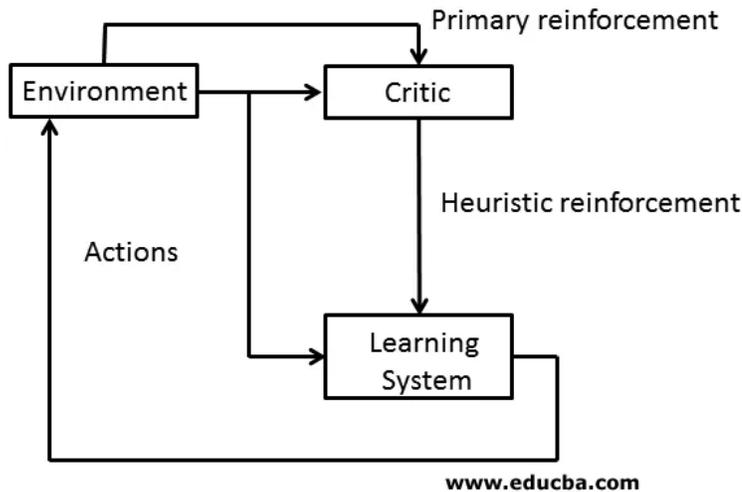
*Nota.* El diagrama muestra como se comporta una red neuronal bajo aprendizaje supervisado. Obtenido de EDUCBA (2022). *Supervised learning* [Aprendizaje supervisado]. (<https://educba.com>), consultado el 13 de marzo de 2022. De dominio público.

El aprendizaje bajo refuerzo se basa en un sistema de constante mapeo entrada-salida, realizado con la interacción del ambiente para minimizar el índice de desempeño. Aquí, en lugar de un maestro, una crítica convierte la primera señal de refuerzo, por ejemplo, la información de entrada del ambiente, en una señal de refuerzo heurística, que sería un escalar de entrada con mayor calidad de señal de refuerzo. Este tipo de aprendizaje apunta a minimizar el costo de la función de costo, que, por ejemplo, sería la suma de acciones tomadas para realizar cierta acción.

En la Figura 2 se puede observar como el sistema cambia del tradicional aprendizaje supervisado, añadiendo factores de refuerzo, tanto primario (dado por el ambiente) y heurístico (generado por la crítica).

**Figura 2.**

*Aprendizaje bajo refuerzo*



*Nota.* El diagrama muestra como se comporta una red neuronal bajo aprendizaje bajo refuerzo. Obtenido de EDUCBA (2022). *Reinforced learning* [Aprendizaje reforzado]. (<https://educba.com>), consultado el 13 de marzo de 2022. De dominio público.

El aprendizaje no supervisado, por otra parte, no posee ningún maestro o supervisor. En este caso, la información no está etiquetada ni clasificada, y no se le da ningún tipo de guía previa a la red. Por lo que la red tendrá que organizarla, dividirla y agruparla en base a similitudes, diferencias y patrones sin ningún tipo de entrenamiento provisto.

Con los tipos de entrenamiento vistos, hay que enfocarse en su funcionamiento desde un ángulo puramente matemático, que es la manera en la que estas redes trabajan debajo de su lado más superficial.

Se puede entender cada una de las neuronas dentro de la red como un nodo, y cada uno de estos nodos es su propio modelo de regresión lineal, un

tema muy conocido en estadística básica, donde se tiene la información de entrada, los pesos y un sesgo o un límite. La fórmula de toda una red neuronal se vería de esta manera:

$$\sum_{i=1}^m w_j x_j + \text{límite} = W_1 X_1 + W_{n-1} X_{n-1} + \dots + \text{sesgo}$$

Donde  $m$  sería el número de neuronas que posea la red,  $W$  los pesos,  $X$  la información y limite el sesgo.

Ahora se podría definir la función que dé el resultado de cierta neurona en la red de esta manera:

$$\text{resultado} = \begin{cases} 1 & \text{si } \text{sum}(W, X) + L \geq 0 \\ 0 & \text{en cualquier otro caso} \end{cases}$$

Una vez esta capa de entrada está definida, los pesos  $W$  son asignados. Estos pesos determinan la importancia de cualquier valor dado, con valores más largos se obtiene mayor importancia en la reacción, comparados con valores más pequeños.

Todas estas entradas son después multiplicadas por sus respectivos pesos y sumadas. Después, el resultado obtenido es ingresado en la llamada función de activación, que determina el resultado final. Si este resultado final sobrepasa cierto límite definido, dispara o activa el nodo, pasando la información hacia la siguiente capa en la red. Este proceso de paso de información entre capas es crucial para el funcionamiento.

Con esta teoría se puede empezar a pensar en casos de uso mucho más prácticos que simple algebra booleana, como identificación de imágenes,

clasificación de imágenes, todo esto utilizando aprendizaje supervisado con información etiquetada para que el algoritmo pueda aprender. Estos modelos mientras más entrenados están, mejores resultados proveen, de forma idéntica a la que el cerebro humano trabaja. Esto provee una función de costo más precisa y un trabajo mejor hecho.

Ahora bien, las redes neuronales, para los conocedores de las ramas de la inteligencia artificial podrían sonar muy parecidas al aprendizaje profundo, materia que ha provisto de utilidades como los *Deep Fakes*, entre otros. Y tendrían razón... parcialmente. Tanto aprendizaje profundo como redes neuronales son conceptos que pueden utilizarse intercambiamente en una conversación, porque el aprendizaje profundo solo es llamado así por la profundidad en las capas. Cualquier red neuronal con más de tres capas puede ser considerada un algoritmo de aprendizaje profundo, pero una red neuronal sencilla se considera con 1 o 2 capas máximo.

Para finalizar la sección, se hablará de los modelos generativos, el corazón de esta investigación.

Los modelos generativos son aquellos que contrastan con los modelos discriminativos, por lo que, en lugar de diferenciar, clasificar y dividir, estos modelos crean y generan. Pero... ¿crean y generan qué? Un modelo generativo puede generar fotos de animales que se parezcan a animales reales, paisajes basados en lugares reales, objetos abstractos, e incluso rostros u objetos.

Un modelo generativo incluye la distribución de la información como tal, y devuelve de manera cuantitativa qué tanto se parece (basado en muchos parámetros), a otra entrada dada. Un ejemplo muy conocido en el campo es el de modelos que predicen la siguiente palabra en una secuencia, típicamente

utilizados en procesadores de texto. Ningún tipo de modelo tiene que retornar un número para representar una probabilidad, como se vio en las redes neuronales más básicas. Se puede modelar la distribución de la información al imitar la propia distribución de la información.

Por ejemplo, un modelo generativo podría modelar una distribución al producir información falsa convincente que se parezca a la distribución de la cual está copiando.

Ahora bien, los modelos generativos son muy complicados. Estos modelos afrontan tareas mucho más complicadas que redes neuronales normales, a pesar de ser lo mismo a un nivel muy bajo. Un modelo generativo para imágenes puede capturar correlaciones entre la información como que, las cosas que parecen botes probablemente salen cerca de cosas que parezcan agua, o que, los ojos no aparecen comúnmente en la frente. Estas son distribuciones muy complicadas de asimilar.

Existen varios sistemas que utilizan los modelos generativos para mejorar ciertas experiencias o crear nuevas. OpenAI es una compañía que posee varios de estos modelos. Su funcionalidad abarca desde generación de texto, predicción de secuencias, generación de imágenes tanto aleatorias como con sesgo y varias más. En capítulos posteriores se analizará cómo utilizar estos modelos abiertos al público para ofrecer una mejor experiencia a sectores menos afortunados en campos que quizás nunca han tenido oportunidad de participar.

## **2.2. Procesamiento de texto natural, base teórica, importancia y aplicaciones**

El objetivo principal del procesamiento de texto natural es como el de cualquier otra rama de la inteligencia artificial: construir sistemas que entiendan el comportamiento humano y puedan responder como uno también.

En este caso en específico se busca el poder entender y responder a texto o información en formato de audio, y que esa respuesta sea con texto o en formato de audio también.

Por definición, el procesamiento de lenguaje natural se refiere a la rama de la inteligencia artificial que intenta proveer a las máquinas con la habilidad de entender texto y palabras habladas de la misma forma que el cerebro humano lo hace.

El PLN (Procesamiento de lenguaje natural), combina varias ramas de la informática, como:

- Lingüística computacional, modelado en base a reglas del lenguaje humano.
- Estadística.
- Aprendizaje de máquina.
- Modelos de aprendizaje profundo.

Así que de forma fundamental es otra aplicación de las redes neuronales, pero muy avanzada.

El PLN es el responsable de los sistemas de traducción, *text-to-speech* o *speech-to-text*, asistentes personales y resumidores de información funcionen de manera óptima, incluso en tiempo real. Es muy probable que ya se haya

interactuado con uno antes, en forma de sistemas GPS, Alexa, o incluso *chatbots* de alguna empresa nacional o internacional. Pero el PLN tiene un rol muy importante también en la construcción de soluciones a nivel empresarial para optimizar operaciones e incrementar productividad de los empleados.

Algunas tareas que desarrollan los sistemas PLN:

- Reconocimiento de voz: también llamado *speech-to-text*, que es la conversión de texto en formato de audio a texto en formato escrito. Esta característica es cada vez más requerida en sistemas que responden a preguntas o comandos, además de sistemas manos libres o de mayor accesibilidad.
- La desambiguación de palabras: que en términos sencillos es la elección del significado correcto en base al contexto de una palabra con varios significados.
- Análisis de sentimiento: el cual pretende calificar a un texto en base a cualidades como las actitudes, emociones, sarcasmo, confusión, entre otras.

Los casos de uso más frecuentes para sistemas de PLN son muchos y variados, entre ellos se tienen los siguientes:

- Detección de spam, tanto en redes sociales, como correos electrónicos, puede basarse en patrones y data aprendida o utilizando herramientas como la detección de sentimiento.

- Traducción: como el Google Translator o algunos más avanzados como Amazon Translate.
- *Chatbots* y agentes virtuales: para contactar con una empresa sin la necesidad de que un humano atienda la línea o los mensajes.
- Resumen de texto: digieren cantidades gigantescas de información y se quedan con los aspectos más importantes y fundamentales sin perder contexto o conclusiones.

### **2.3. Otras aplicaciones de las redes neuronales**

Hasta ahora se ha visto a las redes neuronales hacer tareas muy sencillas como el cálculo de cierto algoritmo del cual se sabe la respuesta, o ser utilizadas para casos de uso mucho más profesionales como lo son los modelos generativos y el procesamiento de texto natural. Pero las redes neuronales son mucho más que 2 o 3 casos de uso.

En esta sección se mostrarán varios casos de uso más y como afectan de manera positiva a empresas tan grandes como Netflix o Amazon.

- Reconocimiento de rostros: los dispositivos móviles ya cuentan con esta característica, haciendo uso de redes neuronales para realizar el proceso de comparación.
- Predicción del clima: sistemas muy importantes de predicción meteorológica en el mundo utilizan redes neuronales e inteligencia artificial para predecir el clima de los días por venir. Utilizando mucha cantidad de

información de años anteriores es posible predecir con cierta precisión como actuará la madre naturaleza próximamente.

- Recomendación de contenido personalizado: desde Netflix, hasta Spotify, pasando por Amazon, esta característica la utiliza cualquier empresa que tenga contenido como materia prima. Basándose en decisiones de compra y reproducción es posible generar una base de datos personalizada para cada usuario y con ella, recomendar contenido en base a patrones analizados dentro del banco de datos.
- Predicción de resultados deportivos: casas de apuestas importantes utilizan redes neuronales para poder predecir resultados deportivos y en base a eso definir sus tarifas de pago por resultado. Todo esto basándose en bancos de información enormes, donde es analizado todo, desde desempeño reciente, hasta cualidades como jugar en casa o de visita.

## **2.4. Profundización en el funcionamiento de soluciones ya existentes**

En este apartado, se explorará detalladamente el funcionamiento de soluciones previamente implementadas en el ámbito de la inteligencia artificial para el apoyo a personas con discapacidad. A través de un análisis exhaustivo, se desentrañarán los componentes esenciales y los algoritmos clave que respaldan estas soluciones, con el objetivo de obtener una comprensión holística de su estructura y operatividad.

### **2.4.1. Asistentes de voz (Siri, Google Assistant y Alexa)**

Un asistente de voz es una aplicación de la inteligencia artificial conversacional, una herramienta que utiliza comandos de voz como entrada y

realiza acciones varias como salida. Utilizando estas tecnologías, diversos dispositivos como bocinas dedicadas, altavoces o parlantes, teléfonos y hasta relojes pueden interactuar y responder a consultas humanas realizadas mediante lenguaje natural.

Al poseer una máquina la habilidad de comprender el lenguaje natural humano y comunicarse bidireccionalmente se abre la puerta a muchas posibilidades de uso, tanto comerciales como personales. Puede ayudar a acelerar procesos, automatizar tareas tediosas, proveer de una manera fácil para controlar otros dispositivos, aumentar la productividad en el trabajo, entre muchas otras.

Todas ellas funcionan de una manera fundamentalmente similar y es importante comprender todos los pasos que se realizan para que un comando de voz se transforme en el resultado deseado.

Existen 5 etapas generales por la que todo comando atraviesa, además de las 2 etapas obligatorias que son la entrada y la salida (voz humana y realizar la acción). Estas 5 etapas se muestran a continuación:

- Conversión de voz a texto (*speech-to-text*): una inteligencia artificial es utilizada para la separación del comando entero en palabras separadas, utilizando para ello el reconocimiento de ondas de sonido recibido como entrada.
- Removiendo sonidos no deseados (ruido): al momento de realizar un comando, quizá este pueda estar “sucio”, véase con ruido de fondo como voces de otras personas, música, entre otros. Otra inteligencia artificial se

encarga en este paso de eliminar de la mejor manera posible todo el ruido detectado, que para ella no forme parte del comando.

- **Procesamiento del lenguaje natural:** en este paso una red neuronal se encarga de encontrar la mejor solución al comando dado en base al entrenamiento de billones de parámetros realizado en su fase de entrenamiento. La inteligencia artificial intentará analizar el significado de la oración y enlazarlo con la mejor salida (resultado) posible.
- **Entendiendo el contexto:** para aplicaciones tan poderosas es necesario tomar en cuenta el contexto en el cuál se está hablando, otra inteligencia artificial se encarga de analizar de manera profunda la solución con mayor probabilidad de ser la correcta, examinando de manera cautelosa de qué manera el usuario está solicitando la consulta.
- **Evaluación de la respuesta:** en este paso la inteligencia artificial tiene que examinar la o las respuestas obtenidas en los pasos anteriores para verificar que cumpla (y tenga sentido), con lo que el usuario solicitó en el primer paso. Esto analizando cada potencial respuesta y filtrando aquellas que tengan una coincidencia perfecta.

#### **2.4.2. Aplicaciones para usos específicos**

Google Lookout y Google Lens son dos aplicaciones que utilizan visión de maquina o *computer visión* para realizar tareas de diversa índole. En el caso de Google Lookout, el enfoque está completamente dirigido hacia el apoyo a personas con discapacidad visual. Utilizando inteligencia artificial puede proveer con retroalimentación acerca de qué está viendo la cámara del dispositivo donde

se está utilizando la aplicación. O en términos más sencillos, funcionando como un sensor que avisa qué objetos tiene delante.

En el caso de Google Lens el enfoque se amplía, ya no está centrado en el apoyo de un grupo en específico, sino que provee de retroalimentación acerca de cualquier objeto que se muestre en cámara. Esto puede ser casi lo que sea, una pintura famosa devolverá el nombre y el autor, un paisaje o lugar de destino turístico famoso dirá la localización, un producto con código de barras mostrará un enlace para poder comprarlo, entre muchas otras. El objetivo principal es utilizar la inteligencia artificial de Google para detectar objetos y devolver qué son, además de información útil acerca de ellos.

Tanto Google Lookout como Google Lens funcionan tras bambalinas de la misma manera, utilizando *computer visión* en un dispositivo con cámara para entender y ofrecer opciones dependiendo del objeto. A continuación, se muestra a grandes rasgos cómo funciona esta tecnología.

Estas aplicaciones combinan 3 grandes pilares para el reconocimiento de objetos mediante imágenes: visión por computadora, aprendizaje de máquina y grafos de conocimiento. Esto combinado trabaja en una red conocida como Red de Propuestas por Región, o RPN por sus siglas en inglés (Region Proposal Network).

Esta red es una red neuronal completamente convolucional que predice los vínculos de objetos y al mismo tiempo da un resultado de qué tan seguro es que ese objeto es lo que se está prediciendo. Estas redes están entrenadas para generar propuestas por región de alta calidad, las cuales son utilizadas para una detección más rápida. Incluso se puede llevar a otro nivel como el reconocimiento de texto dentro de una imagen, siendo estas redes utilizadas dentro de redes

ópticas para poder detectar estos caracteres dentro de una imagen mediante un reconocimiento por líneas.

Gracias a los grafos de conocimiento, estas aplicaciones saben entender contexto mediante palabras ligadas, como por ejemplo si un sustantivo debe ir con mayúscula por ser propio, además de errores en la escritura y demás. Estas aplicaciones también son capaces de detectar bloques de texto como columnas o diferentes tipos de fuentes y color dentro de cada bloque de texto diferente, utilizando señales para el alineado del texto, lenguaje o la relación geométrica entre párrafos para poder determinar si este es el párrafo final o el texto continúa semánticamente.

Por último, cabe mencionar que también se utilizan poderosos algoritmos de aprendizaje de máquina para poder realizar traducciones de texto en caso la imagen se encuentre en un idioma desconocido, siendo capaz de traducir en tiempo real, sustituyendo el texto en pantalla por el texto traducido, manteniendo gramática y dicción. Todos estos algoritmos y tecnologías juntas funcionan meticulosamente para generar herramientas tan poderosas que parecen magia.

### **3. PRESENTACIÓN DE RESULTADOS**

### **3.1. Aplicación de dibujo por capas**

Mediante una aplicación de dibujo (estilo MS Paint), se busca poder realizar trazos, insertar objetos y formas, rellenar con colores y guardar los dibujos realizados, todo únicamente utilizando comandos de voz.

Utilizando una red neuronal de procesamiento de voz se recibirán instrucciones de los trazos a poder plasmar en un lienzo, convirtiendo cada instrucción en una capa a plasmar dentro de dicho lienzo. Esta secuencia de capas podrá ser visualizada al ser plasmada una capa sobre otra secuencialmente sobre el lienzo. Para finalmente, cuando el dibujo se dé por finalizado por el usuario, poder exportar la imagen generada en cualquiera de los formatos más comunes, como: JPG, PNG, SVG, entre otros.

Se quiere presentar una aplicación sencilla que le permita a niños, jóvenes y adultos sin las capacidades físicas necesarias, pero con el deseo de participar en actividades artísticas como el dibujo y la pintura, realizar sus trazos deseados sin la necesidad de interactuar físicamente con un dispositivo de ningún modo.

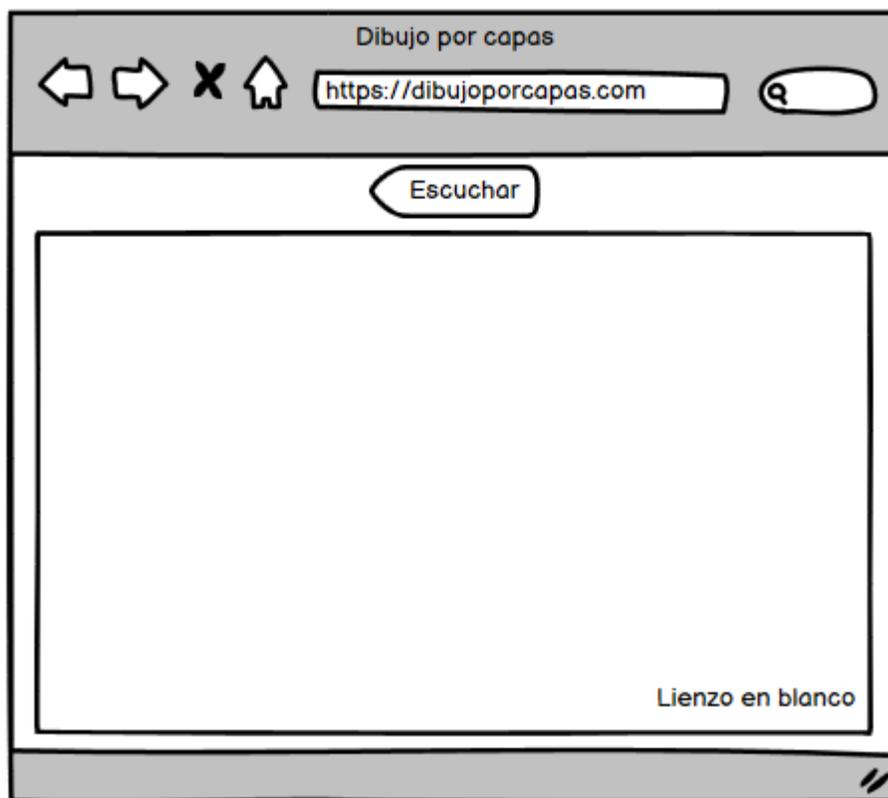
Para una utilización más sencilla, la aplicación estará restringida a trazos básicos como líneas rectas, líneas curvas, figuras geométricas como círculos o rectángulos, figuras especiales como estrellas o flechas, y el rellenado de una figura o un objeto con un color específico. La aplicación también requerirá, al inicializarse, acceso al micrófono del dispositivo donde se está utilizando, esto para facilitarle al usuario el ingreso de audio.

#### **3.1.1. Descripción del funcionamiento**

A continuación, en la Figura 3, se muestra un maquetado de la interfaz principal de la aplicación.

**Figura 3.**

*Interfaz aplicación de dibujo por capas*



*Nota.* Pantalla principal de la aplicación de dibujo por capas. Elaboración propia, realizado con Balsamiq Mockups.

La aplicación cuenta únicamente con dos elementos principales, un botón para volver a solicitar acceso al micrófono del dispositivo en caso no se haya otorgado inicialmente, y un lienzo en blanco, véase un espacio rectangular blanco sin nada encima. La interfaz es muy simple y evita complicaciones al momento

de su uso, priorizando la experiencia del usuario al evitar que tenga que interactuar con componentes físicos con los que quizás no le sea capaz tratar.

Luego de que el audio haya sido inicializado la aplicación empezará a escuchar todo lo que el usuario diga, pudiendo activarse con una palabra en específico y cambiando a modo de procesar la instrucción. Esta palabra es lo que se conoce en los asistentes de voz como *wake word*, una palabra con la que al ser escuchada por la aplicación empieza a funcionar el modo de dibujo.

Es importante mencionar que mientras el micrófono esté abierto, pero no se haya mencionado la palabra clave, la aplicación solo estará en modo de esperar dicha palabra por lo que ningún registro de audio quedará guardado durante el modo previo a su funcionamiento.

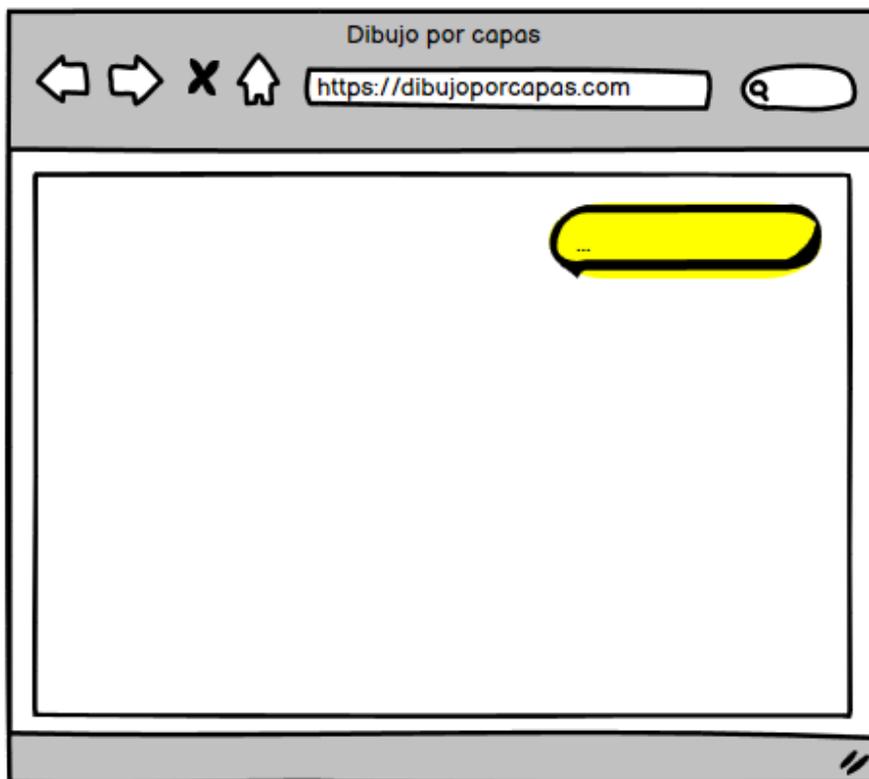
Al inicializar el modo de dibujo la aplicación estará lista para recibir una instrucción acerca del trazo a plasmar en el lienzo, esta instrucción será dictada de forma natural por el usuario, solicitando qué quiere dibujar, además de detalles sobre el trazo como tamaño, posición y ubicación dentro del lienzo.

A continuación, un ejemplo de su funcionamiento:

El usuario activa la aplicación con la palabra clave, inmediatamente la aplicación le informa al usuario que está en modo de recibir instrucción mediante un cambio leve en su interfaz como se puede observar en la Figura 4.

**Figura 4.**

*Modo esperando instrucción*



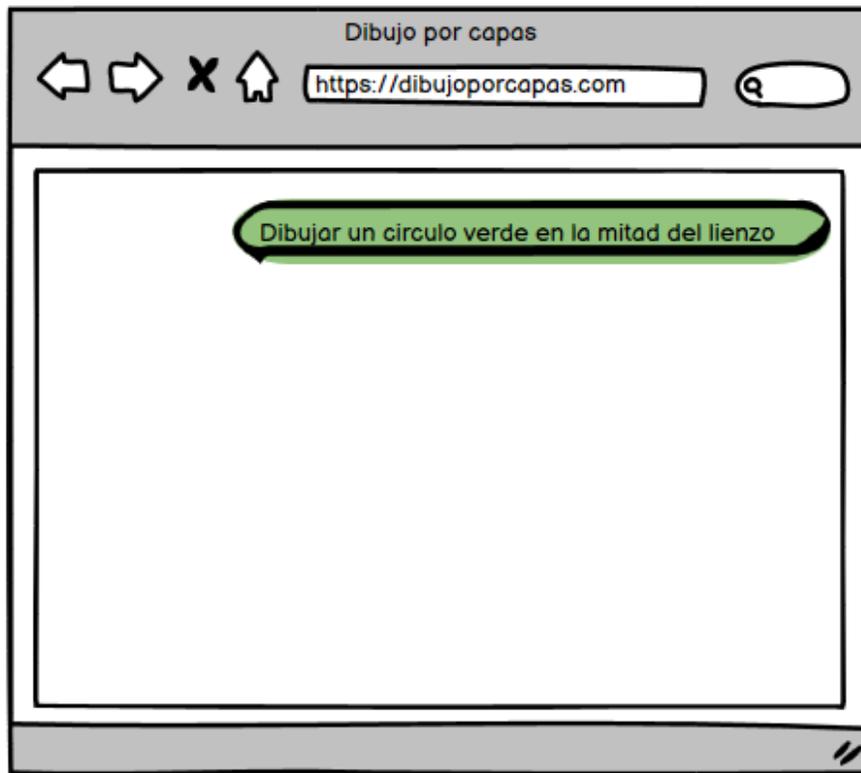
*Nota.* Pantalla de espera de instrucción de la aplicación de dibujo por capas. Elaboración propia, realizado con Balsamiq Mockups.

Como se observa, un pequeño globo de texto resaltado de amarillo aparece en el lienzo, este globo de texto irá mostrando la información que está dictando el usuario para proveer de retroalimentación acerca de si lo que se procesó fue lo mismo que el usuario dijo.

Al finalizar la instrucción el globo de texto cambiará a color verde si la instrucción fue procesada exitosamente o a rojo en caso no pudiese ser procesada. Este cambio puede observarse en las Figuras 5 y 6 a continuación.

**Figura 5.**

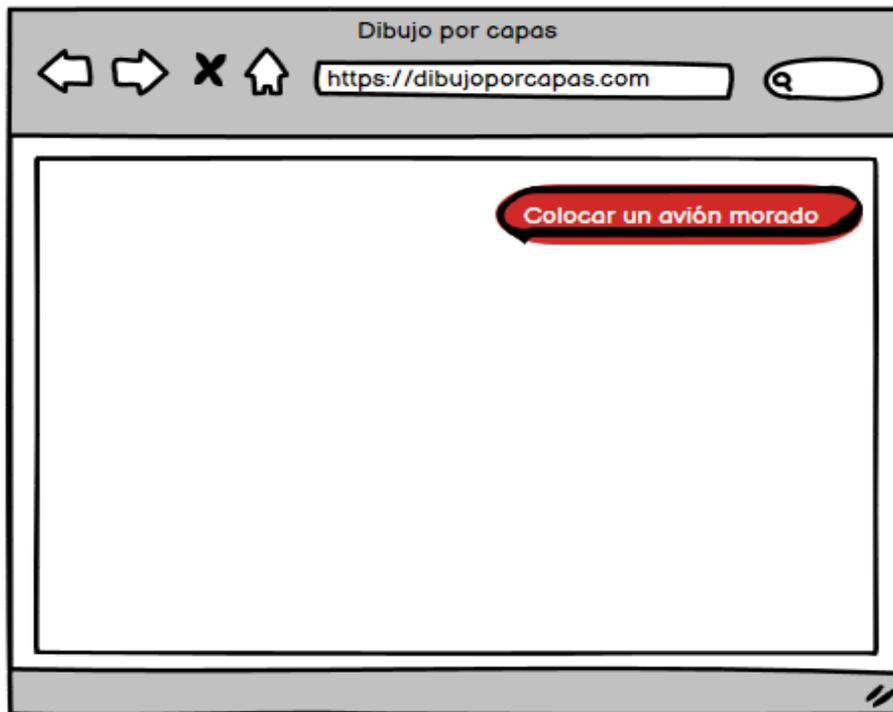
*Solicitud procesada con éxito*



*Nota.* Pantalla de solicitud realizada exitosamente de la aplicación de dibujo por capas. Elaboración propia, realizado con Balsamiq Mockups.

**Figura 6.**

### Solicitud errónea



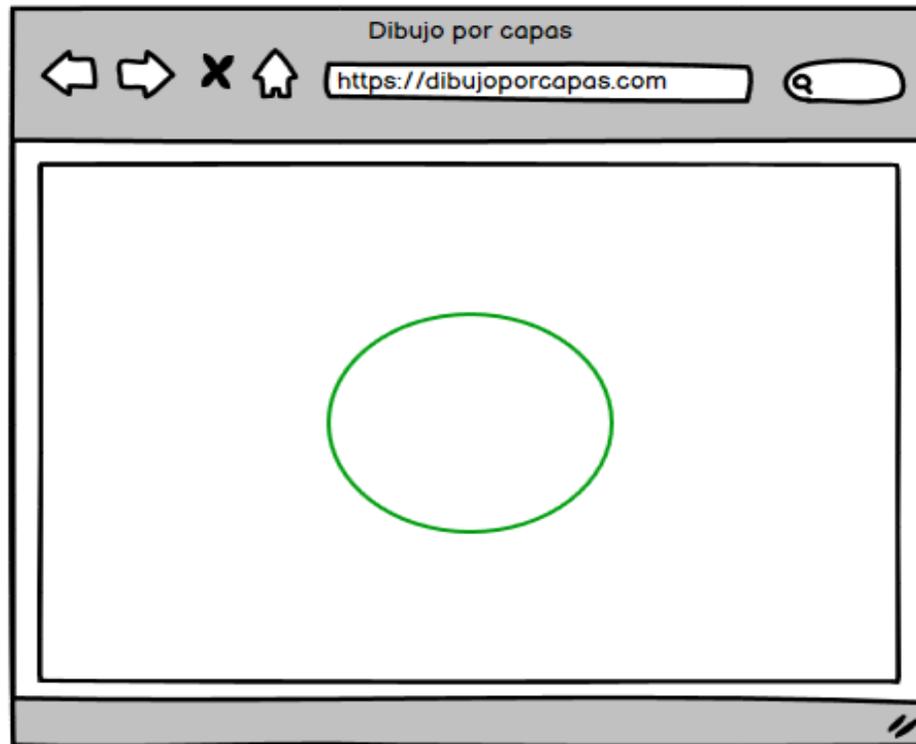
*Nota.* Pantalla de solicitud errónea de la aplicación de dibujo por capas. Elaboración propia, realizado con Balsamiq Mockups.

Si la solicitud fue exitosa la burbuja de texto desaparecerá, y el trazo aparecerá en la pantalla donde fue solicitado, en caso de no haber especificado donde trazarlo aparecerá por defecto en el centro del lienzo, si no se especifica color será negro por defecto. En caso de querer realizar modificaciones se puede solicitar la eliminación del último trazo o capa. Esto con el objetivo de simplificar el funcionamiento y poder colocar y borrar capas secuencialmente hasta obtener el resultado deseado.

En la Figura 7 se puede observar cómo después de hacer la solicitud de la Figura 5 el objeto será plasmado dentro del lienzo como se esperaba.

### Figura 7.

*Circulo verde en la mitad del lienzo*

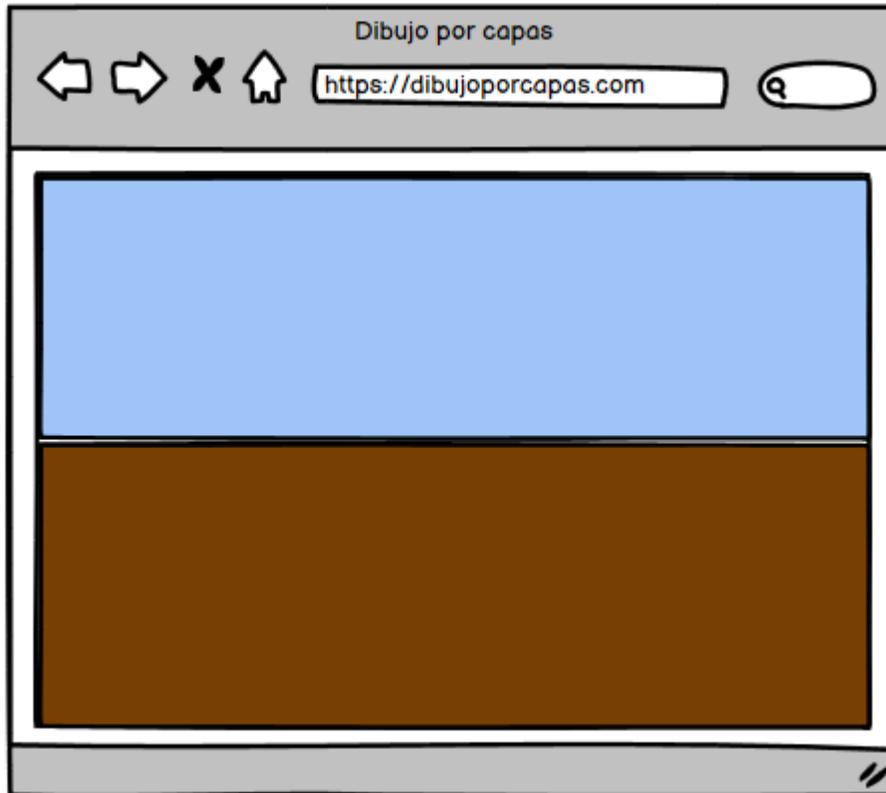


*Nota.* Pantalla de lienzo con un círculo verde dibujado de la aplicación de dibujo por capas. Elaboración propia, realizado con Balsamiq Mockups.

El sistema de capas es muy útil, incluso cuando se quiere hacer dibujos más complejos, como por ejemplo un paisaje. Un paisaje sencillo puede ser dibujado fácilmente mediante el trazo de varias capas secuenciales, por ejemplo: para dibujar un bosque primero se realizaría la solicitud de cubrir la mitad superior del lienzo con azul, simulando el cielo. Luego cubrir la mitad inferior de café, simulando tierra. Los comandos serían: “Dibujar un rectángulo celeste en la mitad superior del lienzo” y “Dibujar un rectángulo café en la mitad inferior del lienzo”. En la Figura 8 se puede observar el resultado de estos dos comandos.

### Figura 8.

## Cielo y tierra

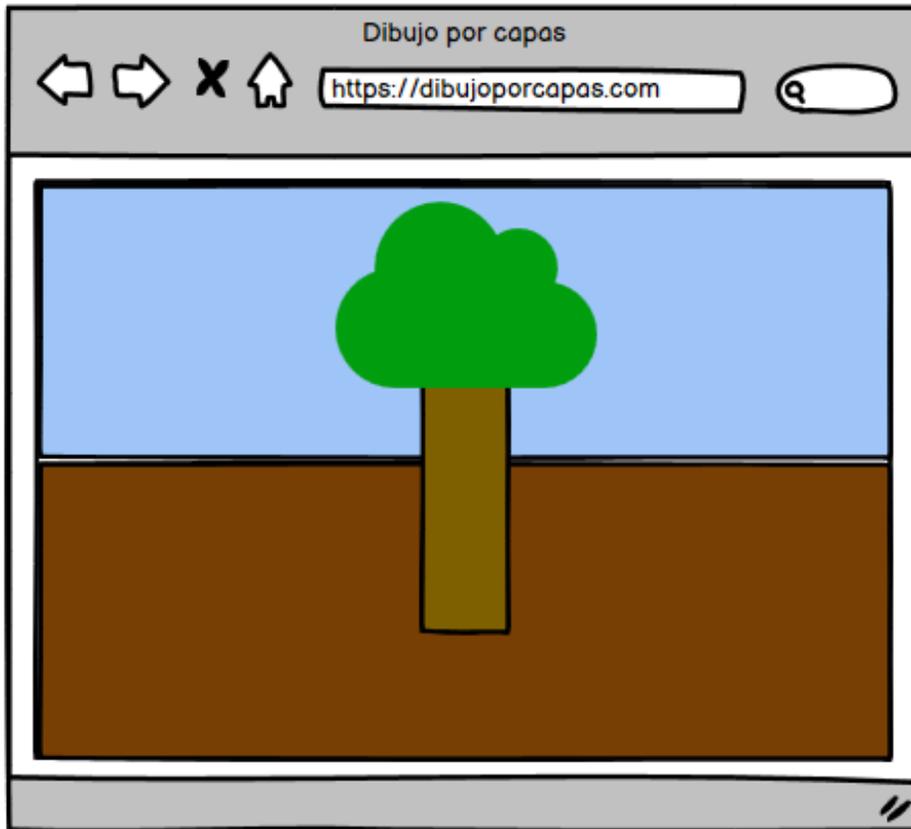


*Nota.* Pantalla de lienzo simulando un horizonte de la aplicación de dibujo por capas. Elaboración propia, realizado con Balsamiq Mockups.

Luego para dibujar los árboles se haría la solicitud de dibujar un rectángulo vertical que cubra 75 px de alto y 20 px de ancho, relleno de color café oscuro, ubicado en el centro del lienzo. Y para formar el follaje se solicitaría dibujar una nube verde 75 px arriba de la mitad del lienzo. En la Figura 9 se observa el resultado final.

### **Figura 9.**

*Dibujo de un árbol*



*Nota.* Pantalla de lienzo con un árbol dibujado de la aplicación de dibujo por capas. Elaboración propia, realizado con Balsamiq Mockups.

Después de esto es solo cuestión de repetir pasos y mover objetos a gusto del usuario, demostrando el alcance de la aplicación y delegando el límite en la creatividad del usuario.

### **3.1.2. Diseño técnico de la aplicación**

La aplicación está compuesta por 3 subsistemas, dividiendo el trabajo en:

- Procesamiento de texto y lenguaje natural

- Conversión de texto procesado a capas y trazos
- Dibujo o plasmado de capas y trazos

En la primera etapa se encuentra una red neuronal para poder procesar audio como entrada y generar texto como salida. Durante el procesamiento de audio es importante convertir el lenguaje natural en instrucciones claras para la interfaz de conversión del segundo paso, es decir remover partes innecesarias y solo dejar claramente la figura, la forma, el tamaño, la ubicación, el color y demás detalles necesarios para poder realizar el trazo.

Esta red neuronal a su vez debe estar compuesta por varias subredes que realicen el trabajo parte por parte. Primero removiendo todo ruido y sonido innecesario de la instrucción dada por el usuario, segundo convirtiendo el sonido filtrado sin ruido a texto para que en el tercer paso el procesamiento del lenguaje natural se lleve a cabo, teniendo en cuenta el contexto y definiendo cuál sería la mejor acción para tomar de acuerdo con lo dicho por el usuario.

En la segunda etapa, el texto generado antes tiene que ser transformado a un objeto o interfaz entendible para la computadora, ya que no se estará utilizando más redes neuronales o inteligencia artificial a partir de aquí. El formato para transformar no es relevante, toda vez se especifiquen de forma clara todos los atributos y detalles de la capa o trazo a realizar en el lienzo como se mencionó antes. Esta lista de objetos o interfaces será luego trasladada a la tercera etapa que se encargará de dibujarlo.

La tercera etapa se encarga de trazar los objetos solicitados por el usuario una vez estos hayan sido procesados. Esto se puede realizar de distintas maneras y variará acorde a como se desarrolle.

### **3.1.3. Análisis de riesgo**

Esta aplicación no presenta riesgo alguno a al ser plenamente recreacional y no involucrar de ninguna manera la seguridad o bienestar del usuario.

#### **3.1.4. Especificaciones de desarrollo**

El sistema donde se pueda desarrollar puede ser:

- Aplicación Web
- Aplicación Desktop
- Aplicación Mobile

El funcionamiento se mantendrá sin importar que tipo de sistema se seleccione, ya que todos cuentan con excelentes marcos de trabajo y herramientas para llevarlo a cabo. Entre ellos:

- Python para las redes neuronales
- JavaScript si la aplicación será web
- C# / C++ para desarrollo en desktop
- Java / Kotlin para desarrollo en Android

La infraestructura, alojamiento, mantenimiento y control son factores a tener en cuenta, así como disponibilidad, accesibilidad y complejidad de desarrollo. Por ser una aplicación con fines de recreación y de complejidad sencilla se recomendaría mantener un alcance local y propio de cada dispositivo.

### **3.2. Asistente para la movilización en la vía pública**

Por medio de una aplicación móvil se busca el proveer de apoyo para personas con discapacidad visual cuándo quieran o necesiten movilizarse por la vía pública, brindando retroalimentación auditiva acerca de sus alrededores.

Utilizando redes neuronales aplicadas en la visión por computadora se detectarán objetos en la proximidad del usuario, utilizando para esto la cámara de su dispositivo móvil. Estos objetos luego serán procesados en tiempo real, determinando sus características y definiendo posibles acciones a tomar. Al finalizar el procesamiento se brindará retroalimentación al usuario acerca de los posibles objetos, personas o riesgos que se encuentren próximos a él.

Para su mejor funcionamiento, la aplicación estará limitada a la detección de objetos estáticos y con movimiento uniforme y despacio, debido a la dificultad de poder proveer, con suficiente tiempo de antelación al usuario, de riesgos por objetos en movimiento acelerado; como pueden ser un carro cruzando a gran velocidad la calle o una pelota cruzándose la acera a gran velocidad frente al usuario.

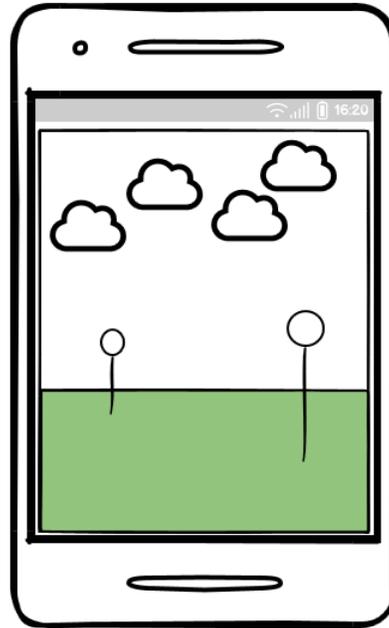
Por obvias razones la aplicación necesitará del permiso de acceso a la cámara del dispositivo en todo momento, además de estar enfocada en la misma dirección en la que el usuario estará caminando, con un buen ángulo de visión.

### **3.2.1. Descripción del funcionamiento**

En la Figura 10 se observa un maquetado de la interfaz principal de la aplicación.

#### **Figura 10.**

*Interfaz aplicación asistente visual*



*Nota.* Pantalla principal de la aplicación de asistente para la movilización en la vía pública. Elaboración propia, realizado con Balsamiq Mockups.

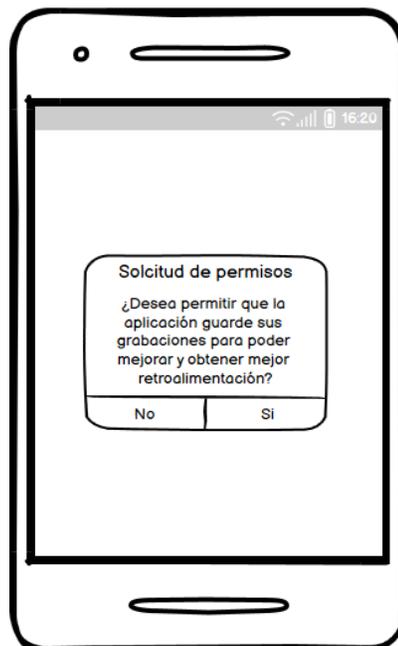
La aplicación cuenta únicamente con un espacio donde será visible la cámara del dispositivo móvil, en este espacio se mostrará en tiempo real todo durante el trayecto completo. Mientras personas, objetos y riesgos potenciales vayan saliendo en la cámara del dispositivo, la aplicación los procesará y le brindará al usuario la retroalimentación necesaria para poder tomar acción en caso fuera necesario.

Esta retroalimentación será en formato de audio, recomendando al usuario utilizar audífonos para escuchar de mejor manera. Además, para el primer uso de la aplicación se le solicitará al usuario permitir acceso a su cámara mientras la aplicación está funcionando, asimismo el aceptar las políticas de uso como el poder utilizar las grabaciones que la aplicación realice para poder mejorar el algoritmo de visión por computadora.

En la Figura 11 se puede observar un maquetado de cómo podría lucir la solicitud de permiso al usuario:

**Figura 11.**

*Solicitud de permisos*



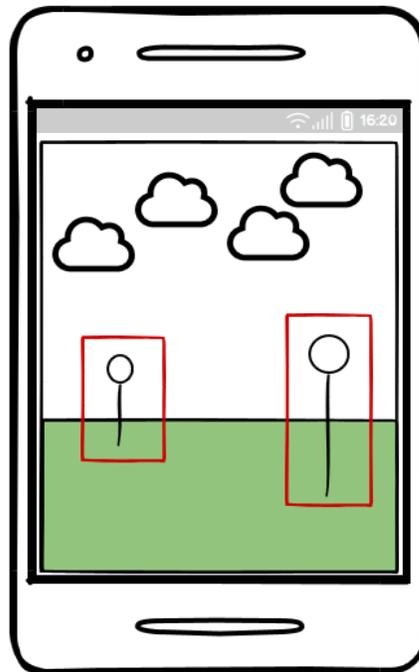
*Nota.* Pantalla de solicitud de permisos de la aplicación de asistente para la movilización en la vía pública. Elaboración propia, realizado con Balsamiq Mockups.

Una vez las políticas hayan sido aceptadas la aplicación estará lista para funcionar, y a partir de ahora no se volverá a solicitar ningún tipo de acceso. Al inicializar cada trayecto la aplicación brindará un pequeño sonido en forma de retroalimentación para dejar saber al usuario que el trayecto ha iniciado. Además, es posible que la aplicación solicite al usuario colocar la cámara en un ángulo más elevado o bajo, dependiendo de si esta lo cree conveniente.

Una vez iniciado el trayecto, la aplicación estará pendiente todo el tiempo de novedades que puedan aparecer en cámara, en la Figura 12 se puede observar como la aplicación detectaría una persona en proximidad del usuario.

**Figura 12.**

*Persona siendo detectada*



*Nota.* Pantalla de detección de objetos de la aplicación de asistente para la movilización en la vía pública. Elaboración propia, realizado con Balsamiq Mockups.

Un rectángulo rojo rodeará objetos, personas y riesgos en la proximidad del usuario, y después de cada detección el usuario recibirá una alerta auditiva explicando qué se está observando, a qué distancia y el nivel de riesgo detectado. Por ejemplo, si la aplicación detecta una persona a 2 metros del usuario caminando por su derecha, el usuario recibiría una alerta, utilizando lenguaje natural, que diría algo como:

*“Persona detectada a dos metros sobre la derecha, movilizándose en dirección norte. Nivel de riesgo bajo”*

Otro ejemplo donde se demuestra de mejor manera la prevención de riesgos es en el momento que se detecta potenciales peligros mayores como alcantarillas abiertas, objetos prominentes o punzantes, la finalización de la acera para dar paso al paso de vehículos, entre muchos otros. Por ejemplo, la aplicación detectaría una fosa o alcantarilla abierta en el trayecto del usuario, esto generaría una alerta de alto riesgo, dándose prioridad sobre cualquier otro aviso que fuese detectado antes y avisando inmediatamente al usuario de una manera como esta:

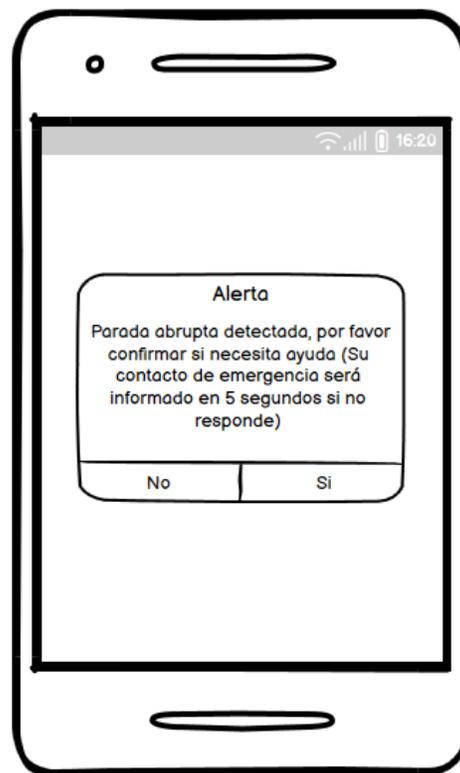
*“Riesgo alto detectado, movilizarse con cuidado. Fosa abierta en el trayecto, tomar precauciones”*

Por último, la aplicación también será capaz de detectar accidentes o cuando el usuario se detenga de manera abrupta, preguntando primero si se encuentra a salvo o si prefiere que se llame a un contacto de emergencia y se envíe su ubicación. Esta última opción será la usada por defecto en caso el usuario no conteste la solicitud en un tiempo corto definido.

En la Figura 13 se puede observar como la aplicación lanzará un estado de alerta solicitando al usuario confirmar su bienestar. Caso contrario, en la Figura 18 se muestra como el teléfono inmediatamente le escribirá a su contacto de emergencia con la ubicación.

**Figura 13.**

*Confirmación de bienestar*



*Nota.* Pantalla de confirmación del bienestar del usuario de la aplicación de asistente para la movilización en la vía pública. Elaboración propia, realizado con Balsamiq Mockups.

El contacto de emergencia podrá ser configurado por el usuario en caso querer a alguien en específico. El contacto de emergencia por defecto serán los servicios de emergencia del 911.

## Figura 14.

*Contactando a emergencias*



*Nota.* Pantalla de contacto a emergencias de la aplicación de asistente para la movilización en la vía pública. Elaboración propia, realizado con Balsamiq Mockups.

Al finalizar el trayecto, el usuario solo deberá cerrar la aplicación y está por defecto dejará de tomar control sobre la cámara y la ubicación del dispositivo. Entrando a modo de espera hasta su siguiente inicio.

### 3.2.2. Diseño técnico de la aplicación

La aplicación está compuesta por 3 partes principales, siendo estas:

- Visión por computadora para la detección de riesgos

- Sistema de texto a voz utilizando lenguaje natural para la retroalimentación.
- Servicio de contacto a emergencias en caso de detección de accidente.

El primer subsistema es el más complejo y el encargado del funcionamiento principal de la aplicación. Mediante inteligencia artificial, aprendizaje de máquina y redes neuronales se tendrá un sistema de visión por computadora capaz de detectar riesgos en la vía pública. Estos riesgos van desde transeúntes, objetos prominentes, objetos cortopunzantes, fosas, alcantarillas, hasta paredes o el fin de la acera. Este subsistema está también encargado de convertir estos objetos detectados en tiempo real, en información entendible por el segundo subsistema.

Este segundo subsistema recibirá los objetos detectados por la visión de computadora y los almacenará en una cola de prioridad a medida que van llegando en tiempo real. Esta cola de prioridad funcionará permitiendo que alertas de nivel de alto riesgo no tengan que esperar a las demás alertas para poder informar al usuario, previniendo posibles daños o perjuicios.

En caso no existan alertas prioritarias, las demás alertas seguirán el orden de una cola cualquiera, siendo informadas al usuario de manera paulatina y cuando se tenga una distancia considerablemente corta entre ella y el usuario. La forma en la que se le informará al usuario será utilizando tecnologías de texto a voz, aprovechándose del procesamiento a lenguaje natural, siendo amigable con el usuario con alertas entendibles y nada complicadas.

Por último, el tercer subsistema es el encargado de tomar las riendas en caso el usuario posiblemente haya tenido un accidente y sea incapaz de poder

ubicarse o solicitar ayuda por su cuenta. Mediante el uso de muchos sensores como giroscopio y acelerómetro se generará un posible choque o detenimiento abrupto, con el cuál se procederá inmediatamente a contactar con el contacto de emergencia o las líneas de emergencia locales. Proveyendo de la ubicación del dispositivo, activando la localización mediante GPS.

### **3.2.3. Análisis de riesgo**

Esta aplicación posee un nivel de riesgo moderado a alto. Es fundamental tener una red neuronal entrenada a la perfección para la detección de amenazas al usuario, en especial para dos casos en específico:

- No informarle al usuario de un riesgo alto
- Informarle al usuario de manera errónea

Con una red muy bien entrenada el nivel de riesgo baja considerablemente, especialmente teniendo en cuenta que la aplicación está lista para poder informar en caso haya sucedido un accidente.

### **3.2.4. Especificaciones de desarrollo**

Debido a la naturaleza de la aplicación y la necesidad de estar en movimiento para poder ser de utilidad, la aplicación está restringida a ser utilizada en un dispositivo móvil, véase Android o iOS.

Cualquiera de las dos plataformas permite el uso de la cámara y la ubicación por lo que los requerimientos no presentan un reto. Además de que los resultados serán satisfactorios en cualquiera de las dos plataformas por la variedad de librerías y marcos de trabajo que ofrecen.

Algunas de las herramientas o lenguajes de programación a utilizar son:

- Python para las redes neuronales y sistemas de PLN
- Java / Kotlin para el desarrollo de la aplicación en Android
- Swift / Objective C para el desarrollo de la aplicación en iOS
- JavaScript / HTML / CSS si se piensa realizar multiplataforma

La aplicación necesitará una conexión estable a la red celular, ubicación mediante GPS y una conexión a internet mediante datos o wifi para funcionar, pero no es necesario ninguna conexión que demande una infraestructura más pesada. Por lo que es recomendable mantener el desarrollo local y sin implementación de servidores propios.

### **3.3. Aplicación para la traducción de lenguaje de señas**

Por medio de una aplicación móvil se busca proveer de apoyo a las personas con discapacidad auditiva al momento de querer comunicarse con alguien que no posee conocimientos de lenguaje de señas.

Utilizando redes neuronales aplicadas en la visión por computadora, poder detectar los elementos del lenguaje de señas, y luego estos elementos poder ser compartidos al intérprete de lenguaje de señas, también a desarrollar. Este intérprete recibirá los elementos procesados por la red neuronal como entrada, y devolverá el significado de la expresión, en texto, como salida.

Por último, un subsistema de *text-to-speech* le brindará al usuario la traducción de la frase ingresada al inicio utilizando lenguaje natural para su mejor entendimiento.

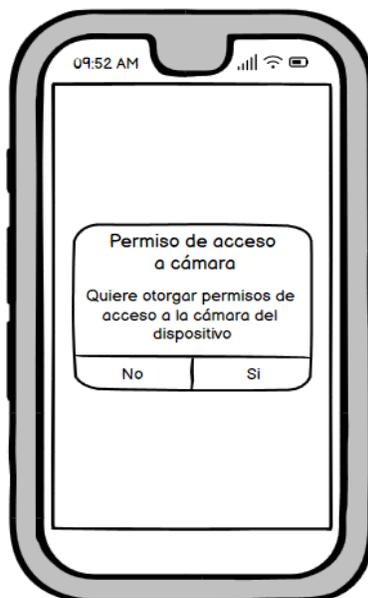
La aplicación requerirá el acceso a la cámara del dispositivo del usuario para poder grabar las expresiones en lenguaje de señas. Además, estará limitada a la detección del LENSEGUA, o lenguaje de señas guatemalteco. Dado que el lenguaje de señas posee diferencias dependiendo de donde se aprenda y se hable, sería imposible poder reconocer modismos o expresiones nativas de algún lugar en específico.

### 3.3.1. Descripción del funcionamiento

La primera pantalla que el usuario verá la primera vez que esta abra la aplicación es donde se le solicita acceso a su cámara, a continuación, en la Figura 15 se muestra un maquetado de esta:

**Figura 15.**

*Solicitud de acceso a cámara*



*Nota.* Pantalla de solicitud de permisos de la aplicación para la traducción de lenguaje de señas. Elaboración propia, realizado con Balsamiq Mockups.

Una vez los permisos hayan sido acreditados, el menú principal de la aplicación será mostrado.

Este menú contendrá 3 opciones:

- Grabar una frase nueva
- Historial de grabaciones
- Frases favoritas

Estas 3 opciones encapsulan el uso general de la aplicación, permitiéndole al usuario poder realizar una grabación en ese momento para traducir, revisar el historial de sus últimas 10 grabaciones para utilización repetitiva si fuera necesario, y poder ver una lista con sus 5 frases favoritas personalizadas para poder reproducir de manera sencilla sin la necesidad de traducir todo de nuevo.

En la Figura 16 se observa el maquetado del menú principal de la aplicación.

**Figura 16.**

### Menú principal traductor de lenguaje de señas

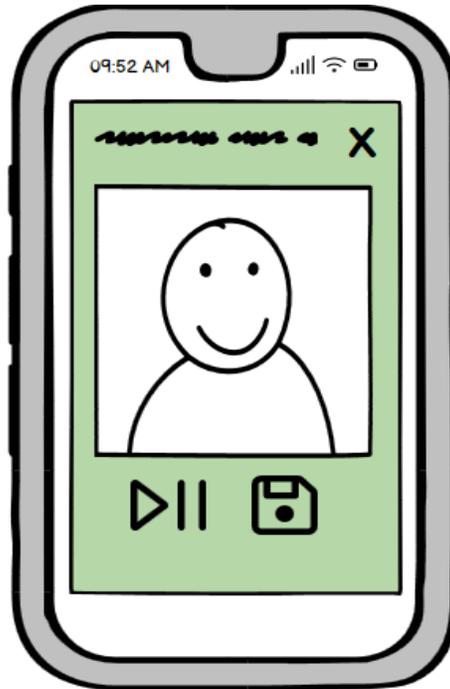


*Nota.* Pantalla principal de la aplicación para la traducción de lenguaje de señas. Elaboración propia, realizado con Balsamiq Mockups.

Al ingresar en el primer botón, el usuario estará dentro del apartado de poder grabar una frase nueva. Este apartado abrirá inmediatamente la cámara del dispositivo con una interfaz propia de la aplicación, esta interfaz es sencilla y busca evitar dificultad al momento de utilizar la aplicación. Solo contiene un botón para grabar o detener la grabación, un botón para guardar la grabación y seguir con el siguiente paso, y un botón en la esquina superior derecha en forma de X para salir del apartado. En la Figura 17 se muestra esta interfaz principal de grabación.

**Figura 17.**

## Interfaz de grabación



*Nota.* Pantalla de grabación de la aplicación para la traducción de lenguaje de señas. Elaboración propia, realizado con Balsamiq Mockups.

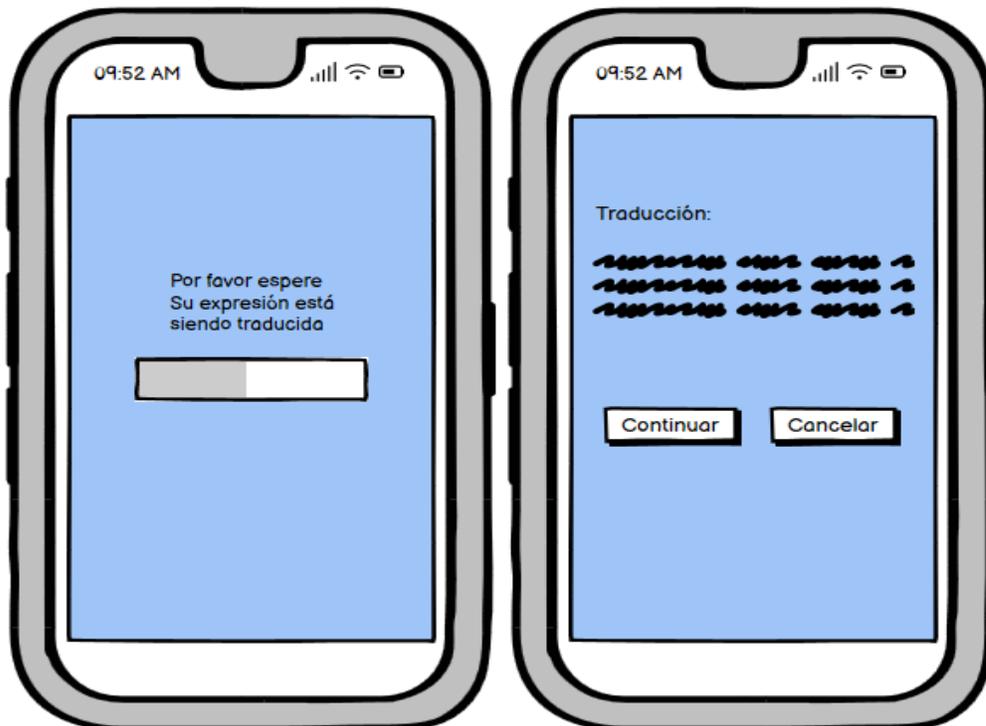
Una vez que el usuario esté satisfecho con su grabación y pulse el botón de guardar, será redireccionado al apartado de traducción, este mostrará un breve mensaje indicándole al usuario que espere hasta que la traducción haya sido realizada por completo para luego mostrarle, en texto, la frase traducida que ingresó grabando sus expresiones en lenguaje de señas.

En caso el usuario no esté de acuerdo con la traducción propuesta, el usuario deberá indicar que eso no es lo que quiso decir, utilizando el botón indicado para ello. Luego será ubicado en otra pantalla que le pedirá retroalimentación de lo que quiso decir para poder mejorar la aplicación con su uso, para finalmente reproducir por el altavoz el mensaje que quiso traducir

mientras es devuelto al menú principal. Todo este proceso se muestra en las figuras 18 y 19.

**Figura 18.**

*Pantalla de espera y pantalla de confirmación*



*Nota.* Pantallas de espera y de confirmación de la aplicación para la traducción de lenguaje de señas. Elaboración propia, realizado con Balsamiq Mockups.

**Figura 19.**

*Menú de retroalimentación*



*Nota.* Pantalla de retroalimentación de la aplicación para la traducción de lenguaje de señas. Elaboración propia, realizado con Balsamiq Mockups.

Si el usuario decide retroalimentar al sistema, este se comunicará con el servidor para poder ingresar la nueva información obtenida y así poder ir mejorando con cada uso correcto o incorrecto. En caso de no querer enviar la información corregida, el usuario simplemente será dirigido al menú principal.

En caso la traducción de la frase si haya sido correcta y el usuario confirme la traducción, esta empezará a sonar por el altavoz del dispositivo, mostrando en pantalla un ícono de altavoz únicamente, además de la opción de al poder

reproducir de nuevo el audio. En la Figura 20 se ve la última pantalla de este segmento de la aplicación.

**Figura 20.**

*Reproducción de la traducción*

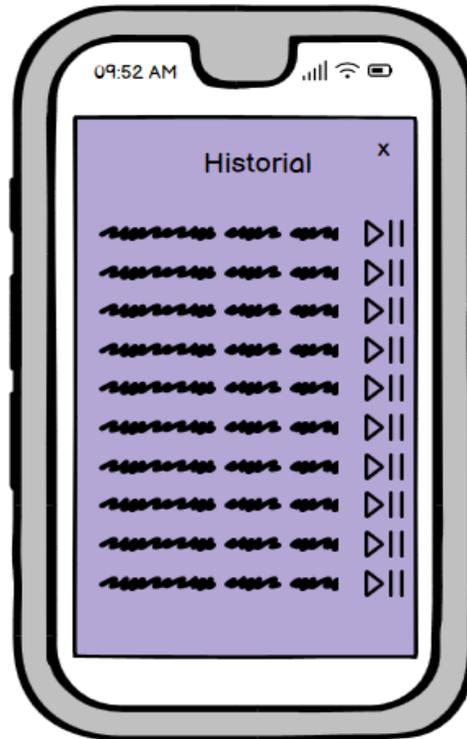


*Nota.* Pantalla de reproducción de la traducción de la aplicación para la traducción de lenguaje de señas. Elaboración propia, realizado con Balsamiq Mockups.

El siguiente apartado de la aplicación muestra el historial de las últimas 10 grabaciones en formato de lista, además de un botón a su lado para poder reproducir. Como siempre, un botón para poder salir del apartado y regresar al menú principal. Este historial no es modificable y se actualiza automáticamente después de cada grabación exitosa. En la Figura 21 se puede observar esta pantalla.

**Figura 21.**

*Historial de grabaciones*



*Nota.* Pantalla de historial de grabaciones de la aplicación para la traducción de lenguaje de señas. Elaboración propia, realizado con Balsamiq Mockups.

En el último apartado el usuario tendrá la posibilidad de reproducir, agregar o eliminar 5 de sus frases más utilizadas o favoritas. Para agregar una frase favorita lo puede hacer desde el apartado de grabación después de realizar una grabación exitosa, marcando como favorita la frase.

Para visualizar y poder reproducir sus frases favoritas es tan sencillo como ingresar a este apartado, identificar qué frase quiere reproducir dentro de la lista de favoritas, y proceder a reproducirla mediante el botón indicado a su derecha.

Además, por debajo de cada elemento aparecerá un botón para poder eliminarlo de la lista de favoritos. En la Figura 22 se observa el menú de favoritos.

**Figura 22.**

*Grabaciones favoritas*



*Nota.* Pantalla de grabaciones favoritas de la aplicación para la traducción de lenguaje de señas. Elaboración propia, realizado con Balsamiq Mockups.

### **3.3.2. Diseño técnico de la aplicación**

La aplicación está compuesta por 4 partes principales, siendo estas:

- Visión por computadora para la grabación del lenguaje de señas
- Intérprete de las expresiones de lenguaje de señas a texto natural

- Sistema de reproducción de audio utilizando lenguaje natural
- Servicio de historial y grabaciones favoritas

En el primer apartado se encuentra el corazón de la aplicación, además de ser el sistema más complejo de realizar. Para el que hay que desarrollar una red neuronal aplicada en visión por computadora que identifique con gran precisión cada una de las expresiones existentes dentro del diccionario del lenguaje de señas guatemalteco.

Es importante definir bien este alcance y marcar los límites para evitar ambigüedades y poder optimizar la velocidad de traducción. Esta red está encargada también de enviar información al segundo apartado, esta información debe estar estructurada de una manera que sea entendible ya sea para un intérprete u otra red encargada de interpretar.

El segundo apartado es el intérprete para la traducción de lenguaje de señas a lenguaje natural (español). Este apartado puede ser desarrollado de dos maneras, desarrollando un compilador o intérprete de lenguaje de señas, teniendo en cuenta que cada expresión de la frase ingresada estará previamente estructurada para su plena comprensión por este apartado.

Existen muchas herramientas y marcos de trabajo para la realización de un compilador, como por ejemplo *Flex*, *Cup*, *Bison*, entre otros. Esto provee de un sistema fuerte y eficaz, pero trae consigo la dificultad de desarrollar un compilador, que puede ser demasiado pensando en el resto de los subsistemas y la dificultad que trae consigo.

La otra opción para este apartado es la utilización de otra red neuronal que reciba la información de la primera; esta segunda red estará entrenada para la

interpretación de lenguaje de señas, utilizando las propiedades previamente estructuradas como entrada. En este caso se tiene un alcance mucho más experimental y de dificultad menor, aunque se necesitaría de mucho entrenamiento para poder alcanzar el nivel de eficiencia esperado de este tipo de proyectos. No importa la opción que se elija, se puede desarrollar un intérprete poderoso y eficiente, con el suficiente tiempo y empeño.

El tercer subsistema es más sencillo de desarrollar, dadas las circunstancias y todas las herramientas a disposición del desarrollador. Este será el encargado de obtener la traducción realizada por el apartado anterior, en formato de texto, y reproducirla en cualquiera de todos los menús que lo necesiten. Para este desarrollo puede utilizarse una red neuronal propia o utilizar uno de los muchos servicios de text-to-speech en la nube, como el que ofrece *Amazon Web Services: Polly*.

El último subsistema no requiere de prácticamente la utilización de ningún servicio externo, más que una base de datos donde poder almacenar información del usuario, así como una lista de sus frases favoritas. Este apartado puede incluso ser local si así se requiere, ya que el espacio a almacenar será muy pequeño, dado que solo se guardará texto.

### **3.3.3. Análisis de riesgo**

Esta aplicación posee un nivel de riesgo bajo a nulo. Ya que el usuario debe confirmar todo el tiempo si lo que está por ser reproducido es lo que el quiso decir desde un principio. Es importante hacer consciencia al usuario de las frases que piensa utilizar y el contexto en el que éstas serán utilizadas para evitar inconvenientes o confusiones que puedan llevar a problemas más grandes. No se recomienda la utilización de esta aplicación para aspectos legales.

### **3.3.4. Especificaciones de desarrollo**

Para esta aplicación poder ser útil es fundamental que su desarrollo se lleve a cabo en plataformas móviles, como Android o iOS. A pesar de que podría ser funcional como aplicación web, el tiempo de reacción es fundamental en estos casos por lo que un desarrollo nativo es la mejor opción.

Ambas plataformas proveen suficientes librerías para la utilización de la cámara del dispositivo, además de poder conectarse a internet en caso el desarrollo involucre la conexión con un servidor en la nube. El resultado será satisfactorio en cualquiera de las plataformas escogidas, siendo estas muy flexibles al momento de permitir al desarrollador utilizar diversos marcos de trabajo como se verá a continuación.

Algunas de las herramientas o lenguajes de programación a utilizar son:

- Python para las redes neuronales y sistemas de PLN
- Java / Kotlin para el desarrollo de la aplicación en Android
- Swift / Objective C para el desarrollo de la aplicación en iOS
- JavaScript / HTML / CSS si se piensa realizar multiplataforma

La aplicación necesitará una conexión estable a la red celular y una conexión a internet mediante datos o Wifi para funcionar. En este caso si es recomendable realizar un desarrollo pensando en escalabilidad y acceso a base de datos y herramientas en la nube, ya que es un proyecto bastante más complejo que los demás presentados hasta ahora. Además, la conexión con una base de datos es casi indispensable y utilizar la memoria del dispositivo del usuario solo es recomendable para hacer un caché de la información.

### **3.4. Aplicación de asistencia para la lectura en tiempo real**

Por medio de una aplicación móvil se busca empoderar a las personas con discapacidad visual al proporcionarles acceso instantáneo y autónomo a información escrita en tiempo real. Esto incluye la capacidad de leer libros, revistas, señalizaciones públicas y más, lo que fomenta su independencia y participación en la sociedad al eliminar las barreras de la comunicación escrita.

Esta aplicación utiliza la cámara del dispositivo móvil para escanear el texto y lo convierte en audio a través de la síntesis de voz. Esto permite a los usuarios acceder a la información de manera independiente y rápida, mejorando su capacidad para leer y comprender el mundo que les rodea.

Además de su capacidad de reconocimiento de texto en tiempo real, la aplicación ofrece una experiencia de navegación táctil que permite a los usuarios explorar el texto escaneado mediante toques en la pantalla. Esto facilita la navegación por páginas, capítulos o secciones del contenido con precisión y comodidad.

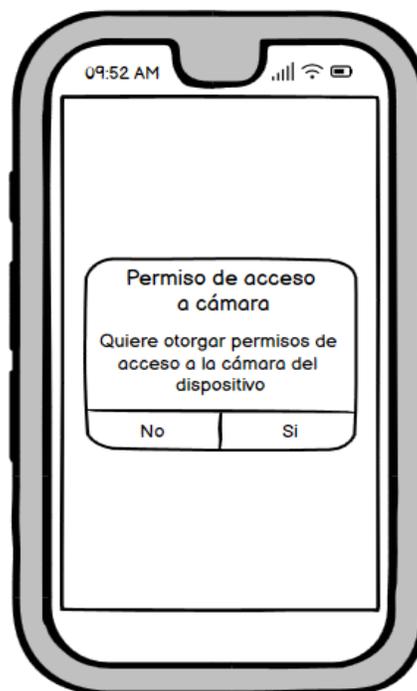
La aplicación requerirá el acceso a la cámara del dispositivo del usuario para escanear el texto impreso en tiempo real utilizando la cámara del dispositivo.

#### **3.4.1. Descripción del funcionamiento**

La primera pantalla que el usuario verá la primera vez que esta abra la aplicación es donde se le solicita acceso a su cámara, a continuación, en la Figura 16 se muestra un maquetado de esta:

**Figura 23.**

*Solicitud de acceso a cámara*



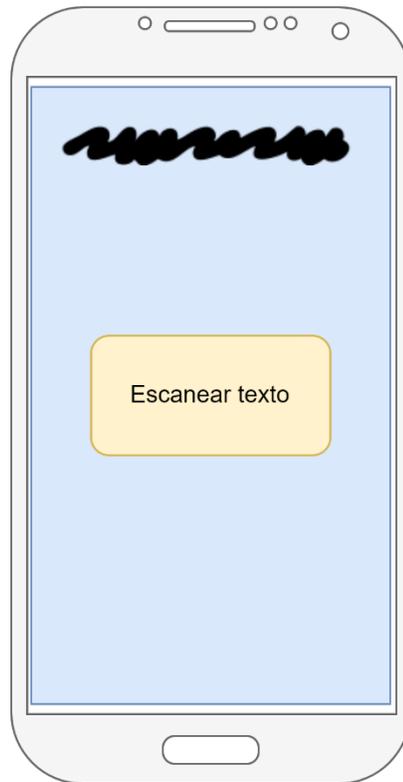
*Nota.* Pantalla de solicitud de permisos de la aplicación de asistencia para la lectura en tiempo real. Elaboración propia, realizado con draw.io.

Una vez los permisos hayan sido acreditados, el menú principal de la aplicación será mostrado. Este menú solo contendrá la opción de escanear el texto. Esta única opción provee de una forma sencilla y directa de acceder al propósito principal de la aplicación.

En la Figura 24 se muestra la pantalla principal de la aplicación:

**Figura 24.**

*Pantalla de inicio*



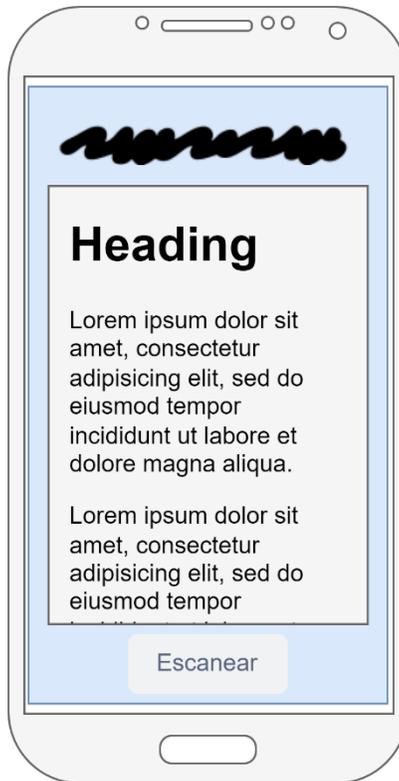
*Nota.* Pantalla principal de la aplicación de asistencia para la lectura en tiempo real. Elaboración propia, realizado con draw.io.

En la siguiente pantalla, la cámara del dispositivo se activa y permite mostrar en tiempo real el objetivo enfocado. Además, un botón para escanear el texto visible en la pantalla de la cámara.

En la Figura 25 se observa la pantalla de escaneo:

**Figura 25.**

*Pantalla de escaneo de texto*



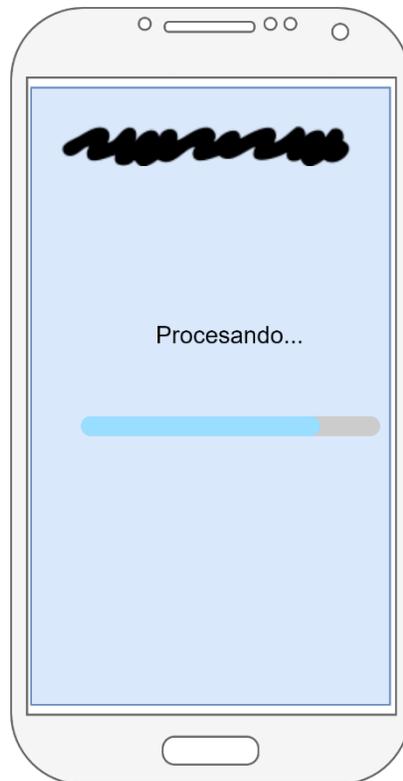
*Nota.* Pantalla de escaneo de texto de la aplicación de asistencia para la lectura en tiempo real. Elaboración propia, realizado con draw.io.

La imagen digitalizada es ahora convertida en texto mediante el uso de inteligencia artificial, más específicamente utilizando OCR (*optical character recognition*), una rama de las redes neuronales en la aplicación de visión por computadora.

Mientras que el algoritmo procesa y convierte la imagen a texto, al usuario le será mostrada una pantalla de espera con una barra de carga que informa del avance del proceso. Dicha pantalla de espera se muestra a continuación en la Figura 26.

**Figura 26.**

Pantalla de espera de procesamiento



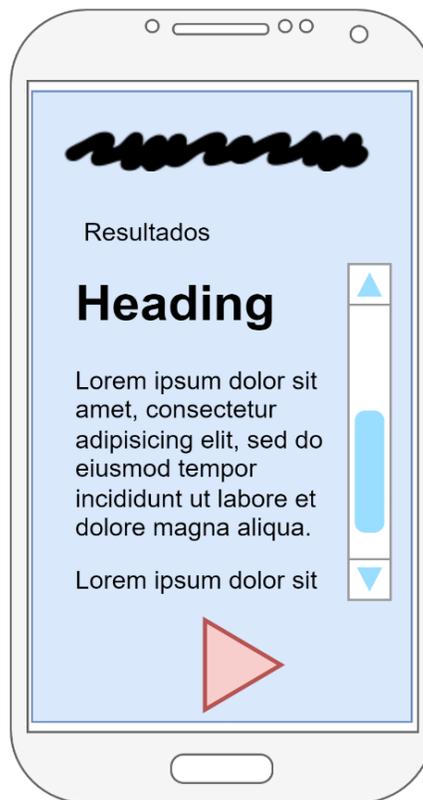
*Nota.* Pantalla de espera del procesamiento de la aplicación de asistencia para la lectura en tiempo real. Elaboración propia, realizado con draw.io.

Una vez la espera ha finalizado se muestra la pantalla de resultados del escaneo, en la cual la aplicación muestra el texto digitalizado en la pantalla con un formato legible. En esta pantalla se muestra también el botón para la reproducción del texto escaneado para convertir el texto en audio y poder ser evaluado por el usuario. Al presionarlo la aplicación comenzará a reproducir en voz alta el texto convertido utilizando una síntesis de voz natural.

En la pantalla de reproducción de texto se le permitirá al usuario ajustar la velocidad de lectura o pausar la lectura en cualquier momento con gestos intuitivos. Las pantallas de resultado del escaneo, así como de reproducción de texto se muestran a continuación en las figuras 27 y 28:

**Figura 27.**

Pantalla de resultados del escaneo



*Nota.* Pantalla de resultados del escaneo de la aplicación de asistencia para la lectura en tiempo real. Elaboración propia, realizado con draw.io.

**Figura 28.**

*Pantalla de reproducción del texto*



*Nota.* Pantalla de reproducción del texto de la aplicación de asistencia para la lectura en tiempo real. Elaboración propia, realizado con draw.io.

### **3.4.2. Diseño técnico de la aplicación**

La aplicación está compuesta por 3 partes principales, siendo estas:

- Captura y digitalización del texto en imagen
- Procesamiento de imagen a texto
- Sistema de Reproducción de Audio con Lenguaje Natural

La primera parte de la aplicación se centra en la adquisición de texto impreso a través de la cámara del dispositivo móvil. Cuando el usuario abre la

aplicación y activa la función de escaneo, la cámara se inicia para mostrar una vista en tiempo real. El usuario dirige la cámara hacia el texto que desea leer, como un libro o una señalización. La aplicación utiliza tecnología avanzada para capturar la imagen y luego procesarla.

En la segunda etapa, la cuál es la más crítica, la aplicación utiliza tecnología de OCR para transformar la imagen digital del texto en texto legible por máquina. La tecnología OCR es capaz de identificar las letras, palabras y estructura del texto en la imagen y convertirla en un formato de texto digital que la aplicación puede interpretar. Este proceso de conversión es esencial para que el texto impreso se vuelva accesible a las personas con discapacidad visual. Una vez que el OCR ha realizado su trabajo, la aplicación está lista para procesar y presentar el texto convertido al usuario de manera legible y comprensible.

La última parte de la aplicación se centra en transformar el texto convertido en una experiencia auditiva de alta calidad. Cuando el usuario decide iniciar la reproducción de voz, la aplicación utiliza un sistema de síntesis de voz natural. Esta tecnología permite que el texto sea leído en voz alta de manera fluida y realista. Los usuarios pueden ajustar la velocidad de lectura y pausar la reproducción en cualquier momento para adaptarse a sus preferencias. Además, la aplicación puede ofrecer opciones de configuración, como la elección de voces y la selección de idiomas, para personalizar aún más la experiencia auditiva.

Este sistema de reproducción de audio con lenguaje natural garantiza que los usuarios con discapacidad visual puedan acceder al contenido impreso de manera efectiva y cómoda, lo que mejora significativamente su capacidad para interactuar con el mundo que les rodea.

En conjunto, estas tres partes esenciales de la aplicación trabajan en armonía para ofrecer a las personas con discapacidad visual una experiencia de lectura en tiempo real accesible y enriquecedora, permitiéndoles acceder a la información impresa de manera autónoma y efectiva.

### **3.4.3. Análisis de riesgo**

Esta aplicación posee un nivel de riesgo bajo a moderado. Ya que el usuario debe entender el riesgo de utilizar esta aplicación para la lectura de información sensible o de carácter peligroso, como operatividad de maquinaria. Es importante hacer consciencia al usuario de la importancia de poder confirmar con otra persona si el texto convertido es importante, o su mala interpretación pone en riesgo su salud o la de los demás. No se recomienda la utilización de esta aplicación para aspectos legales, operatividad de maquinaria, y demás tópicos sensibles.

### **3.4.4. Especificaciones de desarrollo**

Para que esta aplicación pueda ser efectiva, es esencial que su creación se realice en entornos móviles como Android o iOS.

Tanto Android como iOS ofrecen amplias bibliotecas que permiten el uso de la cámara del dispositivo, además de facilitar la conectividad a internet en caso de que el desarrollo requiera interacciones con un servidor en la nube. El resultado será satisfactorio en cualquiera de las plataformas elegidas, ya que ambas son altamente flexibles y permiten a los desarrolladores utilizar una variedad de marcos de trabajo, como se detallará a continuación.

Algunas de las herramientas o lenguajes de programación a utilizar son:

- Python para las redes neuronales y sistemas de PLN
- Java / Kotlin para el desarrollo de la aplicación en Android
- Swift / Objective C para el desarrollo de la aplicación en iOS
- JavaScript / HTML / CSS si se piensa realizar multiplataforma.

La aplicación necesitará una conexión estable a la red celular, o una conexión a internet mediante datos o Wifi para funcionar, pero no es necesario ninguna conexión que demande una infraestructura más pesada. Por lo que es recomendable mantener el desarrollo local y sin implementación de servidores propios.

### **3.5. Aplicación de comunicación para personas con discapacidad del habla**

Por medio de una aplicación móvil se busca proveer a las personas con discapacidad del habla una forma efectiva de comunicarse en situaciones sociales y cotidianas. La aplicación ofrece dos funcionalidades clave: la conversión del habla del usuario en texto y la síntesis de voz para transformar texto en voz sintetizada. Esto permite una comunicación más fluida y accesible, brindando a las personas con discapacidad del habla la capacidad de expresarse y comprender a los demás de manera independiente.

Esta aplicación ofrecerá una funcionalidad esencial: permitirá la conversión automática del habla del usuario en texto en tiempo real, eliminando las barreras de comunicación cuando el habla sea complicada de comprender. Además, la aplicación registrará las conversaciones anteriores en forma de texto, lo que simplificará la revisión y referencia en el futuro. Al integrarse con aplicaciones de mensajería instantánea y proporcionar una interfaz de usuario

intuitiva, la aplicación se convertirá en una herramienta esencial que empoderará a las personas con discapacidad del habla, permitiéndoles comunicarse de manera autónoma y efectiva en una variedad de situaciones sociales y cotidianas.

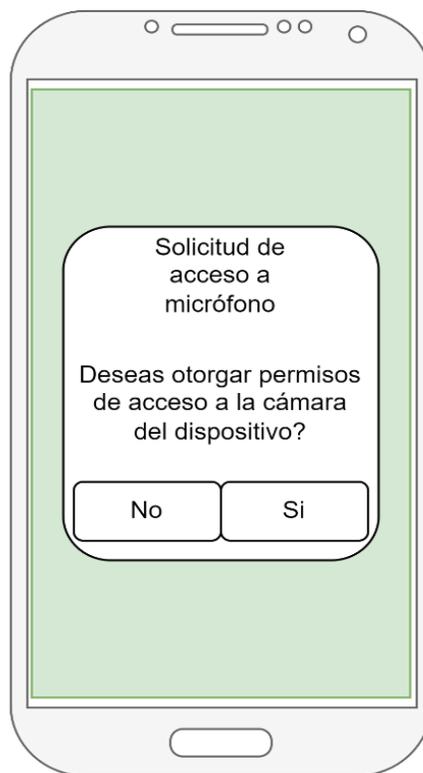
La aplicación se enfocará en aprender la manera de hablar del usuario, sus debilidades y sus intenciones, ofreciéndole un segmento únicamente para practicar, donde será posible que el repita textos mostrados en pantalla para poder aprender y mejorar el algoritmo de *machine learning* que mejorará con el uso y con la práctica. Este segmento de práctica será el más importante, ya que de aquí la aplicación se irá volviendo cada vez más inteligente y podrá proveer al usuario de cada vez mejores interpretaciones.

### **3.5.1. Descripción del funcionamiento**

La primera pantalla que el usuario verá la primera vez que esta abra la aplicación es donde se le solicita acceso a su micrófono, a continuación, en la Figura 29 se muestra un maquetado de esta:

#### **Figura 29.**

*Solicitud de acceso a micrófono*



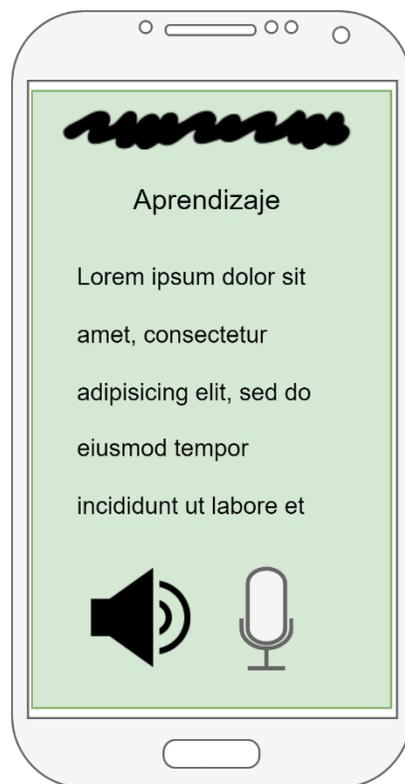
*Nota.* Pantalla de solicitud de permisos de la aplicación de comunicación para personas con discapacidad del habla. Elaboración propia, realizado con draw.io.

La aplicación tendrá un apartado de práctica, en el cual al usuario le serán mostradas ciertas frases para que él lea lo que está escrito y la aplicación aprenda patrones y formas únicas que el usuario tiene para comunicarse. Este apartado será lo primero que el vea en el primer inicio de la aplicación, para la aplicación poder proveer de un mejor servicio desde la primera interpretación.

A continuación, en la Figura 30 se muestra un maquetado del apartado de práctica:

**Figura 30.**

*Apartado de práctica*



*Nota.* Pantalla de práctica de la aplicación de comunicación para personas con discapacidad del habla. Elaboración propia, realizado con draw.io.

Múltiples pantallas como esa serán mostradas al usuario en sucesión hasta obtener culminado el proceso de aprendizaje, este proceso puede que tarde una cantidad considerable de tiempo dada la importancia del aprendizaje del algoritmo.

Una vez el primer aprendizaje haya concluido, y cada vez que el usuario inicie de nuevo la aplicación, le será mostrada la pantalla de menú principal, la cual cuenta con 4 apartados:

- Conversión de habla a texto: apartado donde el usuario habla al micrófono del dispositivo y esta es convertida a texto para luego ser compartido o copiado al portapapeles.
- Conversión en tiempo real: pensada como un modo para que alguien más vea la pantalla de la aplicación donde se irá mostrando en tiempo la conversión de voz a texto proveniente del usuario.
- Historial de conversiones: historial de las últimas 10 conversiones de voz a texto obtenidas del primer apartado.
- Apartado de práctica: para poder mejorar el algoritmo y obtener predicciones más precisas.

A continuación, en la Figura 31, se muestra el menú principal con estos 4 apartados:

**Figura 31.**

*Menú principal voz a texto*



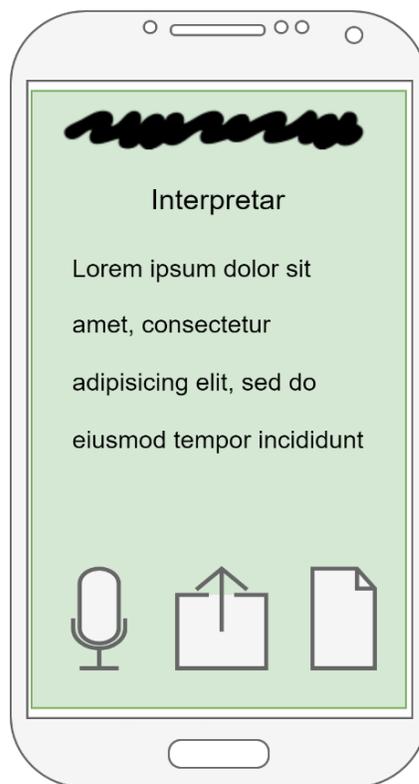
*Nota.* Pantalla principal de la aplicación de comunicación para personas con discapacidad del habla. Elaboración propia, realizado con draw.io.

Luego, en el módulo de interpretación básica, se muestra una pantalla con un cuadro de texto que contiene el texto interpretado, un botón de micrófono para activar la recepción del audio, un botón de compartir para enviar como texto a cualquier servicio de mensajería o aplicación de terceros, así como un botón de copiar al portapapeles para poder pegarlo donde deseen.

En la Figura 32 se muestra un maquetado de esta pantalla de interpretación básica:

**Figura 32.**

*Pantalla de interpretación básica*



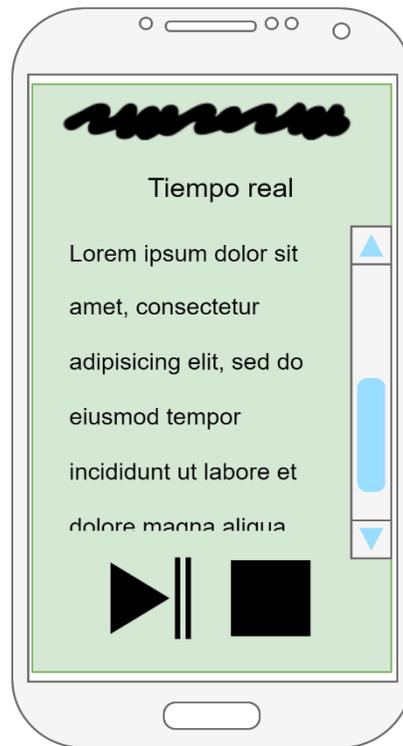
*Nota.* Pantalla de interpretación básica de la aplicación de comunicación para personas con discapacidad del habla. Elaboración propia, realizado con draw.io.

En el siguiente apartado, interpretar en tiempo real, la aplicación provee de una forma para que el usuario comparta su teléfono con alguien más (pensado para la interacción en persona) y esta pueda ver como en un cuadro de texto el habla interpretada del usuario va mostrándose en tiempo real, como si se tratase de un teleprompter. Muy útil en casos como monólogos o presentaciones ante muchas personas, en el cual se podría compartir la pantalla del dispositivo y utilizar esta funcionalidad como subtítulos integrados.

A continuación, en la Figura 33, se muestra un maquetado de la pantalla de interpretar en tiempo real:

**Figura 33.**

*Pantalla interpretar en tiempo real*



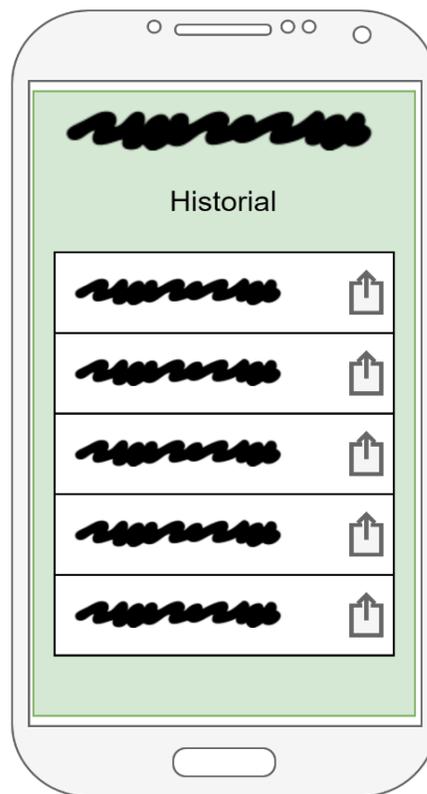
*Nota.* Pantalla de interpretación en tiempo real de la aplicación de comunicación para personas con discapacidad del habla. Elaboración propia, realizado con draw.io.

En el último apartado, historial de interpretaciones básicas, el usuario tendrá acceso a sus últimas interpretaciones para poder compartir recientes. Estas estarán mostradas en formato lista, indicando el texto interpretado más un botón para compartirlas de manera sencilla y práctica. Este apartado es especialmente útil si el usuario es alguien que manda mensajes complejos y largos.

A continuación, en la Figura 34, se muestra un maquetado de la pantalla de historial de interpretaciones básicas.

**Figura 34.**

*Historial de interpretaciones básicas*



*Nota.* Pantalla de historial de interpretaciones básicas de la aplicación de comunicación para personas con discapacidad del habla. Elaboración propia, realizado con draw.io.

### **3.5.2. Análisis de riesgo**

Esta aplicación posee un nivel de riesgo bajo a moderado. Ya que el usuario siempre debe corroborar lo que intentó decir contra lo que la aplicación interpretó, para evitar caer en malentendidos.

Es importante hacer consciencia al usuario de la importancia de esta comprobación en todo momento, en especial si se trata de ambientes profesionales o críticos. No se recomienda la utilización de esta aplicación para aspectos legales, operatividad de maquinaria, y demás tópicos sensibles.

### **3.5.3. Especificaciones de desarrollo**

Para que esta aplicación pueda ser efectiva, puede estar disponible tanto en dispositivos móviles (Android o iOS), como en formato web. Podría ofrecer cierta funcionalidad como una aplicación web, y la capacidad de respuesta en tiempo real enfocada en la aplicación móvil.

Optar por un desarrollo nativo en dispositivos móviles se considera la opción más adecuada para garantizar una experiencia óptima, mientras que la versión web complementaria puede ampliar su accesibilidad.

Tanto plataformas móviles como web ofrecen amplias bibliotecas que permiten el uso del micrófono del dispositivo, además de facilitar la conectividad a internet en caso de que el desarrollo requiera interacciones con un servidor en la nube. El resultado será satisfactorio en cualquiera de las plataformas elegidas, ya que son altamente flexibles y permiten a los desarrolladores utilizar una variedad de marcos de trabajo, como se detallará a continuación.

Algunas de las herramientas o lenguajes de programación a utilizar son:

- Python para las redes neuronales y sistemas de PLN
- Java / Kotlin para el desarrollo de la aplicación en Android
- Swift / Objective C para el desarrollo de la aplicación en iOS
- JavaScript / HTML / CSS si se piensa realizar Web

La aplicación necesitará una conexión estable a la red celular, o una conexión a internet mediante datos o Wifi para funcionar, pero no es necesario ninguna conexión que demande una infraestructura más pesada. Por lo que es recomendable mantener el desarrollo local y sin implementación de servidores propios.



## **4. DISCUSIÓN DE RESULTADOS**

### **4.1. Aplicación de dibujo por capas**

La aplicación de dibujo por capas representa un avance significativo hacia la inclusión en el ámbito artístico para aquellos con limitaciones físicas. Esta solución utiliza comandos de voz para dirigir un lienzo digital, ofreciendo a personas de todas las edades, especialmente a aquellos con discapacidades físicas, una forma accesible de expresar su creatividad a través del arte visual.

La principal fortaleza de esta aplicación radica en su enfoque intuitivo y accesible. Al utilizar la voz como medio de instrucción, elimina la barrera física de la interacción directa con un dispositivo, permitiendo que niños, jóvenes y adultos con diversas capacidades participen activamente en actividades artísticas. El diseño minimalista de la interfaz se enfoca en maximizar la experiencia del usuario al evitar la complejidad innecesaria.

Existen soluciones similares en el mercado, como VOPA, que también utiliza comandos de voz para dibujar, pero tienen la limitación de basarse en un menú de comandos de voz para dibujar. La aplicación propuesta en la descripción del proyecto tiene varias ventajas sobre las soluciones existentes. En primer lugar, es más sencilla de usar, ya que se limita a trazos básicos.

Esto facilita el aprendizaje y la utilización de la aplicación por parte de personas con discapacidad motriz. En segundo lugar, la aplicación utiliza una red neuronal de procesamiento de voz, lo que permite una mayor precisión en la

interpretación de los comandos. Esto reduce la frustración del usuario y mejora la calidad de los dibujos.

El análisis de riesgo confirma que esta aplicación no presenta peligros relevantes para los usuarios, ya que su propósito recreativo no involucra aspectos relacionados con la seguridad o el bienestar. Además, la especificación de desarrollo destaca la adaptabilidad de la aplicación a diferentes plataformas y su implementación mediante lenguajes de programación populares.

Si bien la aplicación aborda la inclusión y la accesibilidad en el ámbito artístico, es importante reconocer sus posibles limitaciones. Aunque se enfoca en trazos básicos y elementos visuales simples, podría enfrentar desafíos al intentar capturar instrucciones más complejas o detalladas. La capacidad de interpretar y procesar instrucciones más elaboradas puede ser un área de Mejora Futura.

#### **4.2. Asistente para la movilización en la vía pública**

La aplicación diseñada para asistir a personas con discapacidad visual en la movilización por espacios públicos representa un hito significativo en la integración y seguridad de este grupo demográfico. Al emplear tecnologías de visión por computadora y procesamiento de audio en tiempo real, la aplicación brinda información auditiva sobre el entorno del usuario, alertando sobre posibles riesgos y facilitando la navegación.

Su función principal radica en detectar objetos y riesgos cercanos a través de la cámara del dispositivo móvil del usuario, transformando esta información visual en una retroalimentación auditiva. El enfoque se centra en proporcionar alertas sobre objetos estáticos y en movimiento uniforme, garantizando así una mejor prevención de riesgos potenciales en entornos urbanos.

La interfaz minimalista de la aplicación se orienta hacia la experiencia del usuario, con un diseño que prioriza la retroalimentación auditiva y la utilización de la cámara en tiempo real para ofrecer información detallada sobre el entorno. Los pasos de inicio, que incluyen la solicitud de permisos de cámara y la aceptación de políticas de uso, aseguran el correcto funcionamiento y la mejora continua del algoritmo de visión por computadora.

Existen soluciones similares en el mercado, como BlindSquare o Seeing AI. Estas aplicaciones también utilizan la visión por computadora para detectar objetos, pero tienen algunas limitaciones. Por ejemplo, EyeSee está limitada a la detección de objetos estáticos, mientras que BlindSquare y Seeing AI, pueden detectar objetos en movimiento, pero pueden ser menos precisas en estos casos. La aplicación propuesta tiene varias ventajas sobre las soluciones existentes. En primer lugar, utiliza redes neuronales, que permiten una mayor precisión en la detección de objetos.

Esto reduce el riesgo de que la aplicación falle a la hora de detectar objetos peligrosos, como automóviles o peatones. En segundo lugar, la aplicación proporciona retroalimentación auditiva, lo que facilita a las personas con discapacidad visual la comprensión de la información que les proporciona la aplicación.

Si bien la aplicación demuestra un gran potencial para mejorar la seguridad y la independencia de las personas con discapacidad visual, existen desafíos relacionados con la precisión de la detección y la gestión de riesgos en entornos dinámicos. La capacidad para anticipar y alertar sobre objetos en movimiento rápido podría ser una mejora futura para abordar riesgos más complejos en entornos urbanos.

### **4.3. Aplicación para traducción de lenguaje de señas**

La aplicación propuesta presenta un enfoque innovador y significativo para asistir a las personas con discapacidad auditiva en la comunicación con individuos no familiarizados con el lenguaje de señas. Su objetivo principal radica en utilizar la visión por computadora y las redes neuronales para detectar y traducir elementos del lenguaje de señas guatemalteco (LENSEGUA), a texto natural comprensible por cualquier persona.

El funcionamiento de la aplicación se divide en varios segmentos, comenzando con la solicitud de acceso a la cámara del dispositivo para grabar las expresiones en lenguaje de señas. El menú principal ofrece opciones claras para grabar una nueva frase, acceder al historial de grabaciones y administrar frases favoritas, lo que mejora la usabilidad y la facilidad de acceso para el usuario.

El proceso de grabación se simplifica mediante una interfaz clara que contiene botones para iniciar, detener y guardar la grabación. Una vez guardada, la aplicación procede a traducir la expresión a texto, presentando al usuario el resultado para su confirmación o corrección. Esta retroalimentación del usuario se considera esencial para mejorar continuamente la precisión de la traducción y el entrenamiento del sistema.

Existen soluciones similares en el mercado, como MIVOS, Hand Talk y Deaf Sign Language App. Estas aplicaciones también utilizan la visión por computadora para detectar los elementos del lenguaje de señas, pero tienen algunas limitaciones. Por ejemplo, MIVOS está limitada a la detección de expresiones básicas, mientras que Hand Talk y Deaf Sign Language App pueden

detectar expresiones más complejas, pero pueden ser menos precisas en estos casos.

La aplicación propuesta en la descripción del proyecto tiene varias ventajas sobre las soluciones existentes. En primer lugar, utiliza redes neuronales, que permiten una mayor precisión en la detección de elementos del lenguaje de señas. Esto reduce el riesgo de que la aplicación falle a la hora de traducir expresiones complejas. En segundo lugar, la aplicación proporciona una traducción en lenguaje natural, lo que facilita a las personas con discapacidad auditiva la comprensión de la información que les proporciona la aplicación. Por último, esta aplicación se basa en el LENSEGUA, siendo más precisa al utilizarse el lenguaje de señas oficial en Guatemala.

El análisis de riesgos señala un nivel bajo a nulo de riesgo, ya que la aplicación depende en gran medida de la confirmación del usuario para cada traducción. Sin embargo, se destaca la importancia de considerar el contexto y la precisión en las frases utilizadas para evitar malentendidos o confusiones, especialmente en situaciones críticas o legales.

#### **4.4. Aplicación de asistencia para la lectura en tiempo real**

Este proyecto de aplicación móvil busca proporcionar a las personas con discapacidad visual acceso inmediato a información escrita mediante el uso de la cámara del dispositivo y la síntesis de voz. Su objetivo principal es eliminar las barreras de la comunicación escrita, permitiendo a los usuarios leer libros, revistas, señalizaciones y más, promoviendo su independencia y participación en la sociedad.

El funcionamiento de la aplicación se inicia con la solicitud de acceso a la cámara del dispositivo. Una vez otorgados los permisos, la aplicación muestra un menú principal con una única opción: escanear el texto. Esta simplicidad proporciona un acceso directo al propósito central de la aplicación, lo que facilita su uso para personas con discapacidad visual.

Al seleccionar la opción de escaneo, la cámara se activa para mostrar en tiempo real el texto objetivo. La interfaz permite al usuario realizar el escaneo del texto visible en la pantalla de la cámara. Posteriormente, se procesa la imagen digitalizada utilizando tecnología OCR (reconocimiento óptico de caracteres), basada en redes neuronales de visión por computadora.

Durante el proceso de conversión de la imagen a texto, se muestra una pantalla de espera con una barra de carga que indica el progreso del proceso. Una vez completado, la aplicación muestra el texto digitalizado en una pantalla legible y ofrece la opción de reproducción de voz utilizando síntesis de voz natural. Esta reproducción de audio permite al usuario escuchar el texto convertido y ajustar la velocidad de lectura según sus preferencias.

Existen soluciones similares en el mercado, como Be My Eyes, Seeing AI y OCR-Scanner. Estas aplicaciones también utilizan la cámara del dispositivo móvil para escanear el texto, pero tienen algunas limitaciones. Por ejemplo, Be My Eyes conecta a las personas con discapacidad visual con voluntarios videntes que pueden ayudar a leer el texto, mientras que Seeing AI y OCR-Scanner utilizan reconocimiento óptico de caracteres (OCR), para convertir el texto en texto digital, que luego puede ser leído por un lector de pantalla. Siendo esta última la más parecida a la aplicación propuesta.

La aplicación propuesta tiene varias ventajas sobre las soluciones existentes. En primer lugar, proporciona acceso instantáneo y autónomo a la información escrita, lo que facilita a las personas con discapacidad visual la lectura y comprensión del mundo que les rodea. En segundo lugar, ofrece una experiencia de navegación táctil que facilita la navegación por el contenido.

En cuanto al análisis de riesgos, la aplicación presenta un nivel bajo a moderado de riesgo. Se destaca la importancia de que los usuarios entiendan los posibles riesgos asociados con la interpretación incorrecta del texto convertido, especialmente en situaciones críticas o peligrosas. Se recomienda la confirmación con otra persona en casos de información sensible o peligrosa.

#### **4.5. Aplicación de comunicación para personas con discapacidad del habla**

El propósito de esta aplicación móvil es facilitar la comunicación de las personas con discapacidad del habla, ofreciendo dos funcionalidades esenciales: convertir el habla del usuario en texto y sintetizar texto en voz. Estas características permiten una comunicación más fluida y accesible, empoderando a las personas con discapacidad del habla para expresarse y comprender a los demás de manera independiente.

La aplicación se inicia con una solicitud de acceso al micrófono del dispositivo para capturar el habla del usuario. Luego, ofrece un segmento de práctica donde el usuario lee ciertas frases para que la aplicación aprenda patrones únicos en su forma de comunicarse. Este aprendizaje es clave para mejorar la interpretación del algoritmo de machine learning de la aplicación.

Una vez concluido el proceso de aprendizaje, la aplicación presenta un menú principal con cuatro apartados: conversión de habla a texto, conversión en tiempo real, historial de conversiones, y un apartado de práctica para mejorar el algoritmo.

El apartado de conversión de habla a texto permite al usuario hablar al micrófono para que la aplicación convierta su habla en texto. El apartado de conversión en tiempo real muestra en tiempo real la conversión de voz a texto, ideal para compartir la pantalla en interacciones presenciales.

El historial de conversiones muestra las últimas interpretaciones de voz a texto realizadas por la aplicación, facilitando la revisión y compartición de conversaciones recientes.

Existen soluciones similares en el mercado, como Proloquo2Go y Talkie. Estas aplicaciones también ofrecen la conversión del habla en texto y la síntesis de voz, pero tienen algunas limitaciones. Por ejemplo, Proloquo2Go y Talkie pueden convertir habla compleja, pero pueden ser más difíciles de usar.

La aplicación propuesta tiene varias ventajas sobre las soluciones existentes. En primer lugar, ofrece una funcionalidad esencial: la conversión automática del habla del usuario en texto en tiempo real. Esto elimina las barreras de comunicación cuando el habla es complicada de comprender. En segundo lugar, la aplicación registrará las conversaciones anteriores en forma de texto, lo que simplifica la revisión y referencia en el futuro. En tercer lugar, la aplicación se integra con aplicaciones de mensajería instantánea, lo que facilita la comunicación con otras personas.

En cuanto al análisis de riesgos, la aplicación tiene un nivel de riesgo bajo a moderado, ya que se enfatiza la importancia de que el usuario verifique lo interpretado por la aplicación para evitar malentendidos, especialmente en entornos críticos o profesionales.



## CONCLUSIONES

1. El análisis profundo de los modelos generativos y el procesamiento de texto natural ha revelado un espectro amplio de aplicaciones beneficiosas para personas con discapacidad. Estas tecnologías emergentes muestran un gran potencial para apoyar sectores como las ciencias y las artes, proporcionando herramientas innovadoras que mejoran la accesibilidad y la participación en la sociedad de las personas con discapacidad.
2. El estudio detallado de la evolución de las redes neuronales ha revelado un progreso significativo en el desarrollo de modelos generativos y el procesamiento de texto natural. Se ha observado una transición desde modelos básicos hacia arquitecturas más complejas, como las redes neuronales recurrentes (RNN) y las redes neuronales convolucionales (CNN), mejorando la comprensión y generación de texto de manera notable.
3. La revisión de proyectos actuales ha demostrado avances notables en modales generativos, resaltando capacidades impresionantes como la creación de texto, imágenes y hasta la síntesis de voz. Estos avances muestran un potencial real en aplicaciones prácticas para la asistencia a personas con discapacidad, especialmente en la comprensión y generación de contenido accesible.
4. Se ha elaborado una guía exhaustiva que detalla diversos proyectos concepto utilizando estas tecnologías. Estos proyectos se han identificado

como herramientas valiosas para asistir a personas con discapacidad en áreas como la lectura en tiempo real para discapacidad visual o la comunicación efectiva para aquellos con discapacidad del habla. Cada proyecto se ha descrito detalladamente, resaltando sus características y el potencial impacto positivo en la vida diaria de las personas con discapacidad.

## RECOMENDACIONES

1. Proponer colaboraciones interdisciplinarias entre expertos en tecnología y profesionales en ciencias y artes, para identificar necesidades específicas de personas con discapacidad. Esto permitirá desarrollar soluciones innovadoras que integren modelos generativos y procesamiento de texto natural para mejorar la accesibilidad en estos sectores.
2. Fomentar la investigación en arquitecturas neuronales emergentes para comprender su impacto en la mejora de modelos generativos. La exploración de nuevas técnicas y enfoques en redes neuronales, como las redes adversariales generativas (GAN), puede impulsar la generación y comprensión de texto de manera aún más efectiva para personas con discapacidad.
3. Colaborar con organizaciones y comunidades de personas con discapacidad para implementar proyectos reales basados en modelos generativos. Esto implica la creación de aplicaciones específicas que utilicen la capacidad de estos modelos para facilitar la comunicación, la lectura en tiempo real y otras necesidades específicas, maximizando así su impacto en la vida diaria de las personas con discapacidad.
4. Enfocar futuros desarrollos en tecnologías que promuevan la inclusión y la autonomía, priorizando soluciones que aborden necesidades específicas en la comunicación, lectura o interacción social. Además de establecer programas de divulgación y capacitación para la

implementación de los proyectos concepto identificados. Buscando alianzas estratégicas con instituciones para su implementación práctica y efectiva.

## REFERENCIAS

- Berman, S. (28 de julio de 2017) *Artificial intelligence as a tool for people with disabilities* [La inteligencia artificial como herramienta para personas con discapacidad]. Equal Entry. <https://equalentry.com/artificial-intelligence-as-a-tool-for-people-with-disabilities/>.
- Howard, J. (16 de noviembre de 2021) *Obstacles & challenges people with physical disabilities face in digital spaces* [Obstáculos y desafíos que enfrentan las personas con discapacidad física en los espacios digitales]. InclusionHub. <https://www.inclusionhub.com/articles/obstacles-people-with-physical-disabilities-face>. Accessed 20 February 2022.
- Lecun, Y. (1992) *Reverse TDNN: an architecture for trajectory generation* [TDNN inversa: una arquitectura para la generación de trayectorias]. New Jersey, Estados Unidos. *AT&T Bell Laboratories*. <https://proceedings.neurips.cc/paper/1991/hash/fde9264cf376ffe2ee4ddf4a988880d-Abstract.html>
- Namkung, E. H., & Carr, D. (2020). The psychological consequences of disability over the life course: assessing the mediating role of perceived interpersonal discrimination [Las consecuencias psicológicas de la discapacidad a lo largo de la vida: evaluación del papel mediador de la discriminación interpersonal percibida]. *Journal of health and social behavior* [Revista de Salud y Comportamiento Social], 61(2), 190–2. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7450392/>

Pitts, W., & Warren, S. (1943). A logical calculus of the ideas immanent in nervous activity [Un cálculo lógico de las ideas immanentes a la actividad nerviosa]. *Bulletin of Mathematical Biophysics* [Boletín de Biónica Matemática], 5.  
<https://home.csulb.edu/~cwallis/382/readings/482/mccolloch.logical.calculus.ideas.1943.pdf>

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain [El perceptrón: un modelo probabilístico para el almacenamiento y organización de la información en el cerebro]. *Psychological Review* [Revisión Psicológica], 65(6), 386–408.  
<https://psycnet.apa.org/record/1959-09865-001>