



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

UNIVERSIDAD DE LAS FUERZAS ARMADAS - ESPE

Trabajo de integración curricular, previo a la obtención del título de Ingeniero
en Mecatrónica

“Diseño e implementación de un robot interactivo como herramienta tecnológica de apoyo para el desarrollo de habilidades lingüísticas en niños de 4 a 5 años, mediante actividades interactivas de comprensión de cuentos, formación de oraciones y corrección de pronunciación.”

AUTORES: Jefferson Javier Cabezas Guevara
 Moreno Guijarro Kevin Mateo

DIRECTOR: Ing. Mayra Elizabeth Comina Tubón, Ph. D.





- 1. Objetivos**
- 2. Introducción**
- 3. Metodología**
- 4. Pruebas**
- 5. Resultados**
- 6. Conclusiones**
- 7. Recomendaciones**



Objetivo General

Diseñar e implementar un robot interactivo de apoyo para el desarrollo de habilidades lingüísticas en niños de 4 a 5 años, mediante comprensión de cuentos, formación de oraciones y corrección de pronunciación.

Objetivos Específicos



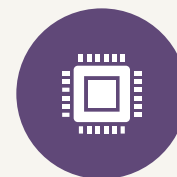
Estado del arte



Diseño
mecánico,
electrónico y de
control



Interfaz gráfica
con juegos



Implementación
del prototipo



Validación y
pruebas

La robótica social ya demuestra impacto en el lenguaje infantil

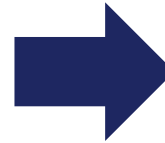
Tema	Evidencia	Fuente
Vocabulario y narrativa	Más vocabulario y patrones narrativos al jugar	Kory & Breazeal (2014)
Imitación lingüística	Imitación del habla del robot; mejoras medibles	Kory-Westlund & Breazeal (2019)
Reconocimiento de habla	Whisper: transcripción infantil en tiempo real	Janssens, et al (2024)

Vacío identificado: la investigación se concentra en primaria o lenguas extranjeras; falta evidencia en preescolares hispanohablantes.

¿De qué manera un robot educativo interactivo puede mejorar las habilidades lingüísticas de niños de 4 a 5 años?

PROBLEMÁTICA ACTUAL

- Sin robots para lenguaje oral en preescolares
- Materiales tradicionales poco motivadores
- Ningún dispositivo narra, escucha y corrige a la vez
- **Laboratorios de la sede matriz sin estas herramientas**



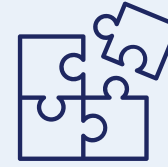
OPORTUNIDAD

- Reconocimiento de voz infantil poco explorado
- Robot interactivo motiva el lenguaje (4–5 años)
- Narrar, escuchar, corregir y participar en un solo robot
- Validado por expertos en pedagogía e ingeniería mecatrónica y mecánica

Un proyecto interdisciplinario con impacto técnico, educativo y social



Interacción natural



**Enfoque
interdisciplinario**



**Desarrollo
socioemocional**



**Accesible y
pertinente**

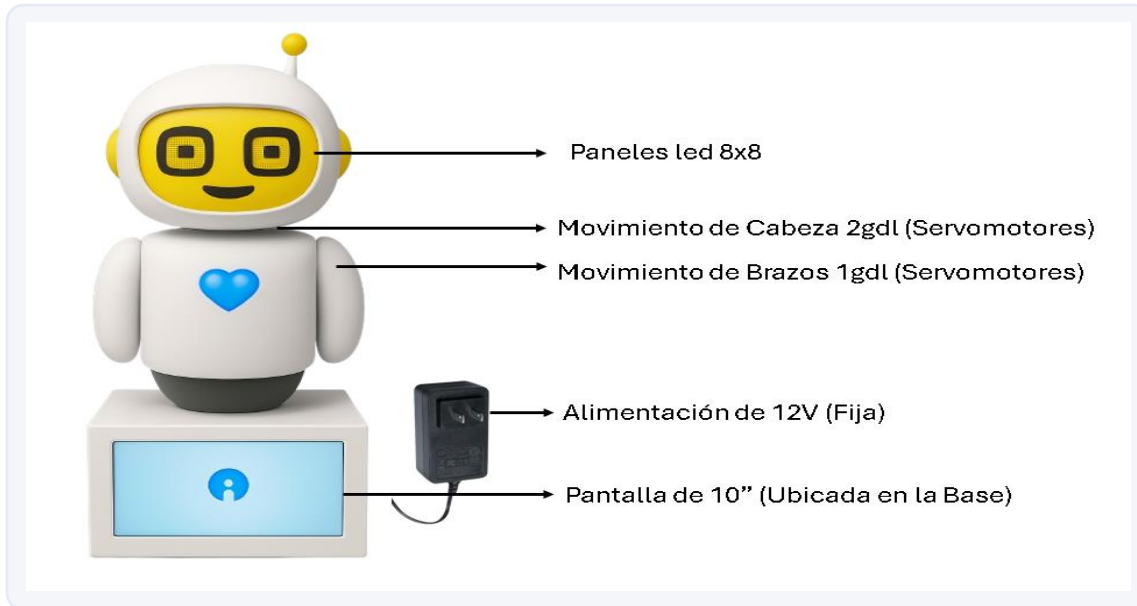
Seis fases, desde la investigación hasta la validación



Diez necesidades clave definen el diseño del robot

Necesidad	Interpretación
Aspecto amigable	Formas y colores que generen confianza y conexión emocional.
Brazos y cabeza móviles	Movilidad para simular gestos y mantener la atención.
Pantalla de 10" a color	Visualización gráfica de tamaño adecuado para el niño.
Interacción táctil	Sistema táctil en pantalla para navegación intuitiva.
HMI interactivo y sencillo	Interfaz clara que evite la saturación de información.
Voz: parlante y micrófono	Emitir indicaciones y reconocer respuestas orales.
Fácil de transportar y armar	Sin piezas sueltas; montaje simple.
Juegos de lenguaje	Aplicaciones educativas específicas a programar.

Dos conceptos de diseño evaluados antes del prototipo final



Concepto 1 — Ganador

- Pantalla 10.1" en base frontal, inclinada 30°
- Raspberry Pi + ESP32
- Brazos SG90 (1 GDL), cabeza SG995 (2 GDL)
- Fuente AC/DC estacionaria



Concepto 2

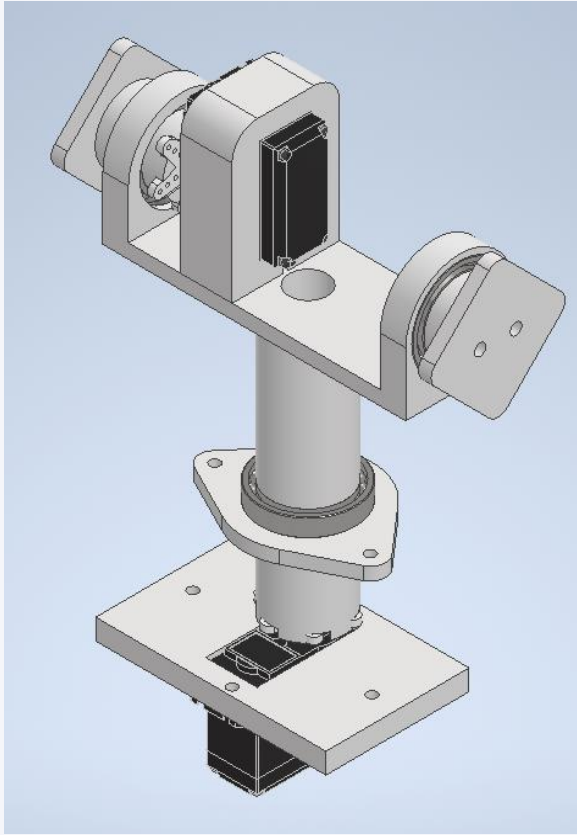
- Pantalla 10" vertical en el torso
- Jetson Nano + ESP32
- Brazos SG995 (2 GDL), cabeza SG90 (1 GDL)
- Batería LiPo recargable

El Concepto 1 obtiene la mejor valoración ponderada

Criterio	Valor	Concepto 1	Concepto 2
Visualización	20	8	9
Procesamiento	10	3	7
Durabilidad	20	9	3
Transportabilidad	5	7	5
Construcción	5	7	8
Estética de diseño	20	9	8
Costo	20	9	3
TOTAL	100	52	43

El Concepto 1 fue seleccionado por su ventaja en costo y durabilidad, atribuida a la ubicación baja de la pantalla.

El eje de la cabeza soporta todo el conjunto superior

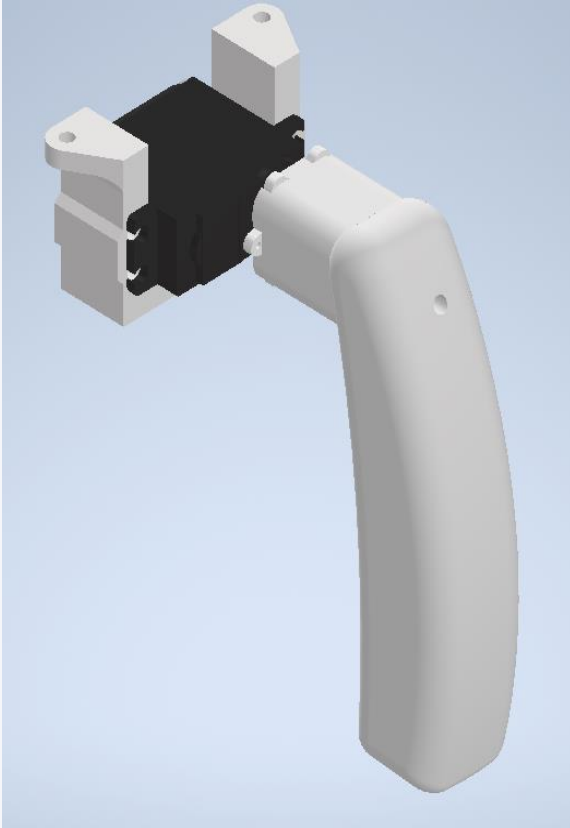


CÁLCULO

$\tau_{\text{inercial}} \approx 0,0181 \text{ N}\cdot\text{m}$
 $\tau_{\text{gravedad}} \approx 0,0491 \text{ N}\cdot\text{m}$
 $\tau_{\text{total}} \approx 0,7 \text{ kg}\cdot\text{cm}$
 $\sigma_{\text{máx}} = 0,366 \text{ MPa}$

4,9 kg·cm
Torque de diseño de la cabeza
(con FS > 7)

Dimensionamiento del servomotor de brazo



CÁLCULO

$$\tau_{\text{inercial}} \approx 0,0298 \text{ N}\cdot\text{m}$$
$$\tau_{\text{gravedad}} \approx 0,0288 \text{ N}\cdot\text{m}$$
$$\tau_{\text{total}} \approx 0,6 \text{ kg}\cdot\text{cm}$$
$$FS = 7 \rightarrow \tau_{\text{diseño}}$$

4,2 kg·cm
Torque de diseño del brazo
(con FS = 7)

Motores MG996R soportan 7.5kg.cm

El torso concentra los esfuerzos principales del robot

Impresión 3D por FDM en PLA (Inventor → OrcaSlicer), con esfuerzos muy por debajo del límite del material.

0,138 MPa

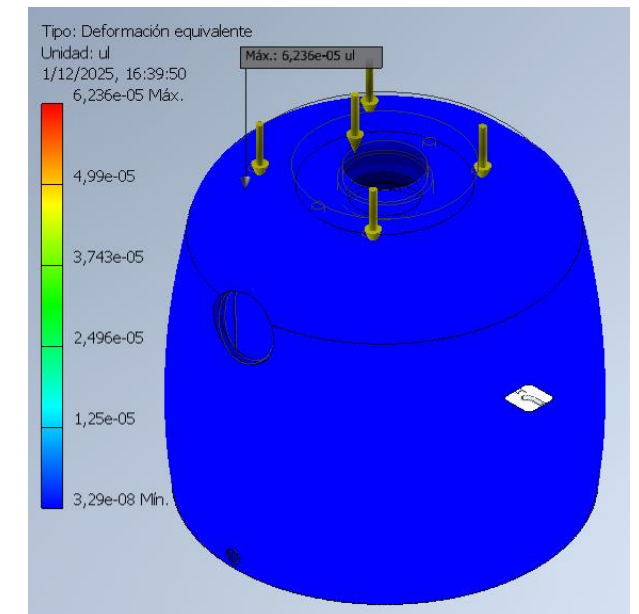
Esfuerzo máximo en el
torso

47 MPa

Límite de tracción del
PLA

CS = 15

Coeficiente de
seguridad obtenido en
el torso



Fuente de 5 V / 15 A con margen de seguridad ante consumo máximo



CÁLCULO

$$I_{\text{nominal}} = 2,72 \text{ A}$$

$$I_{\text{máximo}} = 14,5 \text{ A}$$

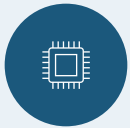
$$FS_{\text{nominal}} = 15\text{A} / 2,72\text{A} = 5,51$$

$$FS_{\text{máximo}} = 15\text{A} / 14,5\text{A} = 1,03$$

5 V / 15 A

Fuente seleccionada — margen $\approx 500 \text{ mA}$
en demanda máxima

Arquitectura de doble unidad de control



Raspberry Pi 5 — Unidad central

Procesamiento de voz, interfaz gráfica (GUI) y lógica pedagógica; salida micro HDMI y USB-C.

UART · 3,3 V



ESP32 — Control auxiliar

Coordina 4 servomotores MG996R y 2 matrices LED 8×8, enlazado por UART a 3,3 V con la Raspberry Pi.



Entradas y salidas

Entradas

Micrófono USB · Pantalla táctil 10.1" (1024×600 px)

Salidas

Servomotores · Matrices LED 8×8 · Bocinas

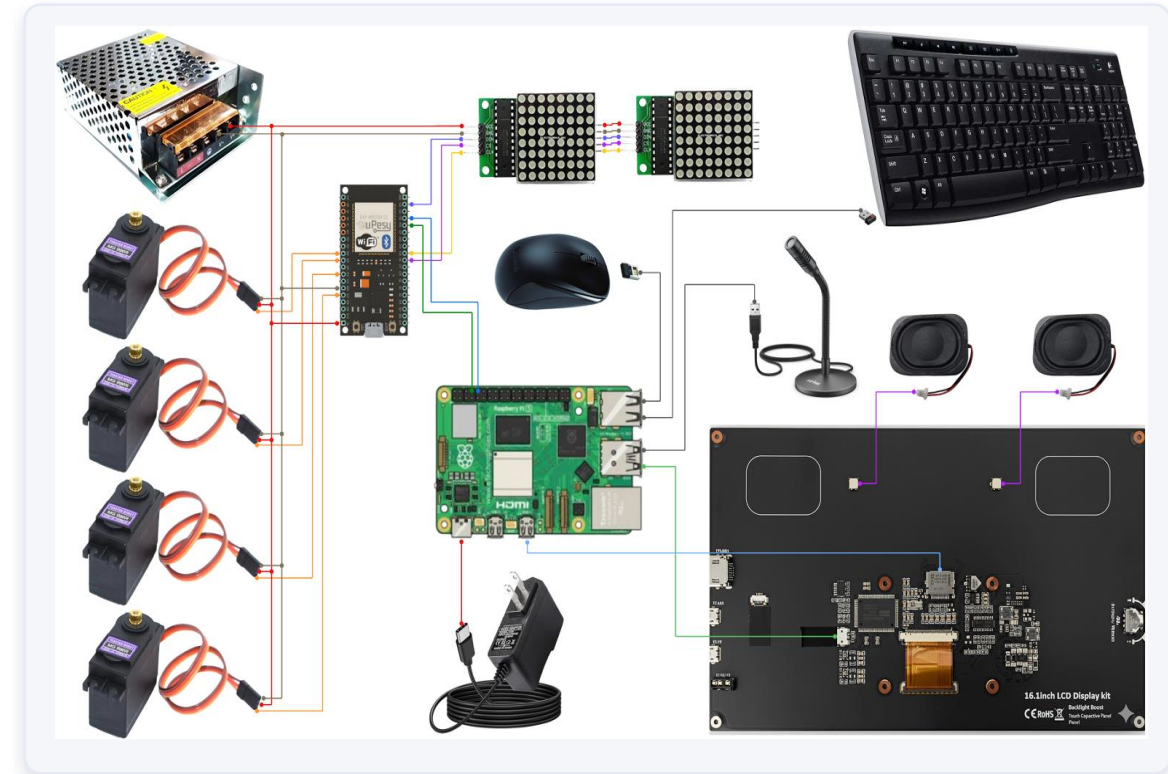


Diagrama general de conexiones

PCB dedicada a la etapa de potencia de la ESP32 y actuadores

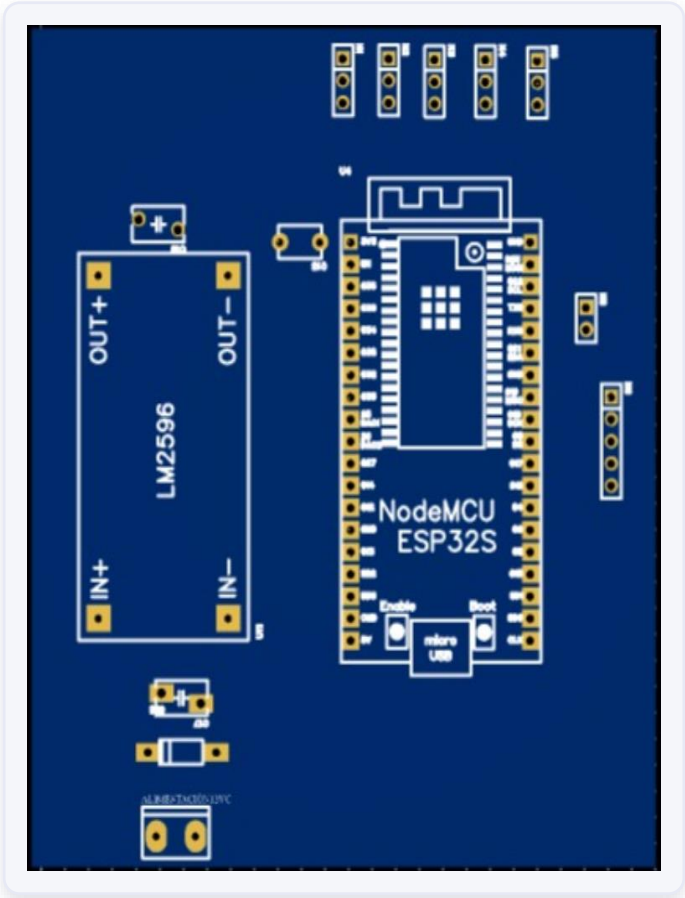


Regulador de voltaje integrado que alimenta la ESP32 a 3,3 V.



Ancho de pista según la corriente máxima ($\Delta T = 15\text{ }^{\circ}\text{C}$, cobre 1 oz).

Etapa	Corriente	Ancho de pista
Señal	40 mA	0,15 mils
Alimentación ESP32	500 mA	4,90 mils
Potencia · 4 servos + 2 matrices LED	14 A	485,38 mils
Fuente principal	14,5 A	509,45 mils



Diseño de la PCB de potencia

Reconocimiento híbrido: Google en línea o Vosk local, según conectividad

1

Captura de voz

El micrófono USB capta la voz del niño y la digitaliza con la librería sounddevice.



2

Reconocimiento híbrido

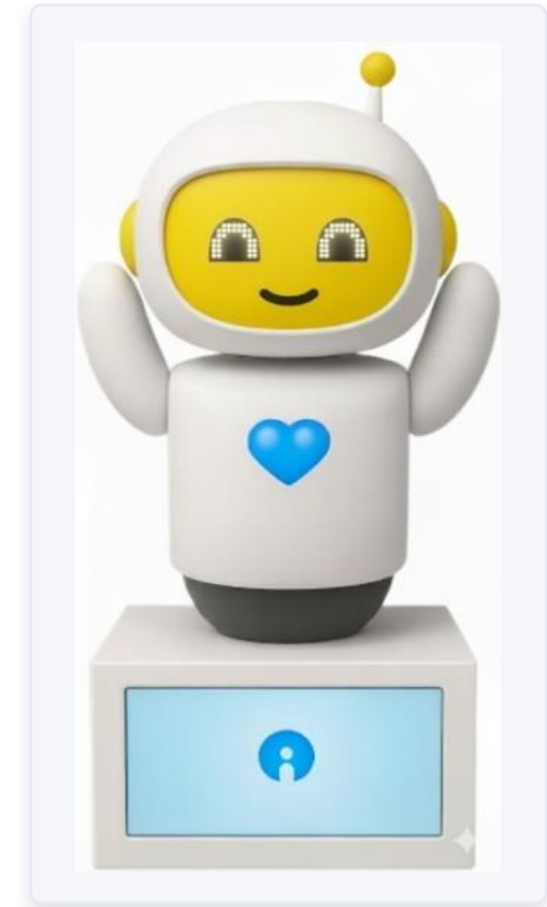
Verifica la conexión a internet y usa Google (mayor precisión) en línea o Vosk de forma local, según disponibilidad.



3

Selección automática de motor

Todo se gestiona desde la Raspberry Pi 5, que mantiene el reconocimiento funcional con o sin conexión a internet.



Piper: síntesis de voz neuronal en tiempo real, sin locuciones pregrabadas

1

Motor Piper

Modelo neuronal de síntesis de voz que genera el audio en tiempo real, sin necesidad de locuciones pregrabadas.



2

Generación en tiempo real

El habla se genera y reproduce al instante durante la interacción, frase por frase.



3

Ejecución en vivo

Durante la operación, la Raspberry Pi ejecuta el modelo Piper en vivo para cada frase, sin depender de audios pregrabados.



4

Resultado

Respuestas casi inmediatas, voz natural sin locuciones fijas y sincronización estable entre voz, pantalla y movimientos.



Pantallas interactivas desarrolladas en Qt Designer y PyQt6



Pantalla de menú de juegos de la HMI



Arquitectura modular: la capa visual (Qt) está separada de la lógica de control y el hardware.



**Cuenta
Cuentos**



**Ordena
Oraciones**



Pronunciación



Cada juego integra video/audio, preguntas de comprensión y pantallas de retroalimentación (felicitaciones o reintento).



Diseño pensado para niños de 4-5 años: navegación simple, iconografía clara y sin saturación de información.

Un cuarto módulo: charla libre con IA, gratuita gracias a OpenRouter

1

Charla libre en español con botón
“pulsar para hablar”

2

Al soltar el botón, el audio se transcribe
automáticamente

3

La IA responde en frases cortas y apropiadas
para niños

4

Enrutado por OpenRouter — Llama 3.3 70B,
gratuito

5

La respuesta se convierte en voz y el robot la
reproduce



Administrador de Contenido: la profesora amplía el material sin programar

1

App de escritorio: agregar, editar o eliminar sin código

2

3 pantallas: Cuentos, Oraciones y Pronunciación

3

Editar cuento: título, portada, video y preguntas

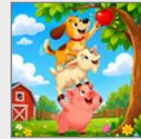
4

Imágenes recortadas y redimensionadas automáticamente (máx. 600 px)

5

Confirma antes de eliminar y avisa cambios sin guardar

Título:

Portada: 

Video actual: assets/stories_video/videos/001.mp4

Preguntas del cuento:

1. ¿Qué encontró el perro en el árbol?
2. ¿Quién intentó tirar la manzana usando sus cuernos?
3. ¿Quién tuvo la idea para alcanzar la manzana?

Manufactura aditiva FDM con parámetros diferenciados por tipo de pieza



Modelado en Autodesk Inventor → exportado a STL → laminado en OrcaSlicer.



Piezas internas Estructurales

- Altura de capa: 0,2 mm
- 2 perímetros
- Relleno 10 %
- Orientadas para resistir tensión y torsión



Piezas externas Estéticas

- Altura de capa: 0,12 mm
- Velocidad 60 mm/s
- 3 perímetros
- Mejor acabado superficial



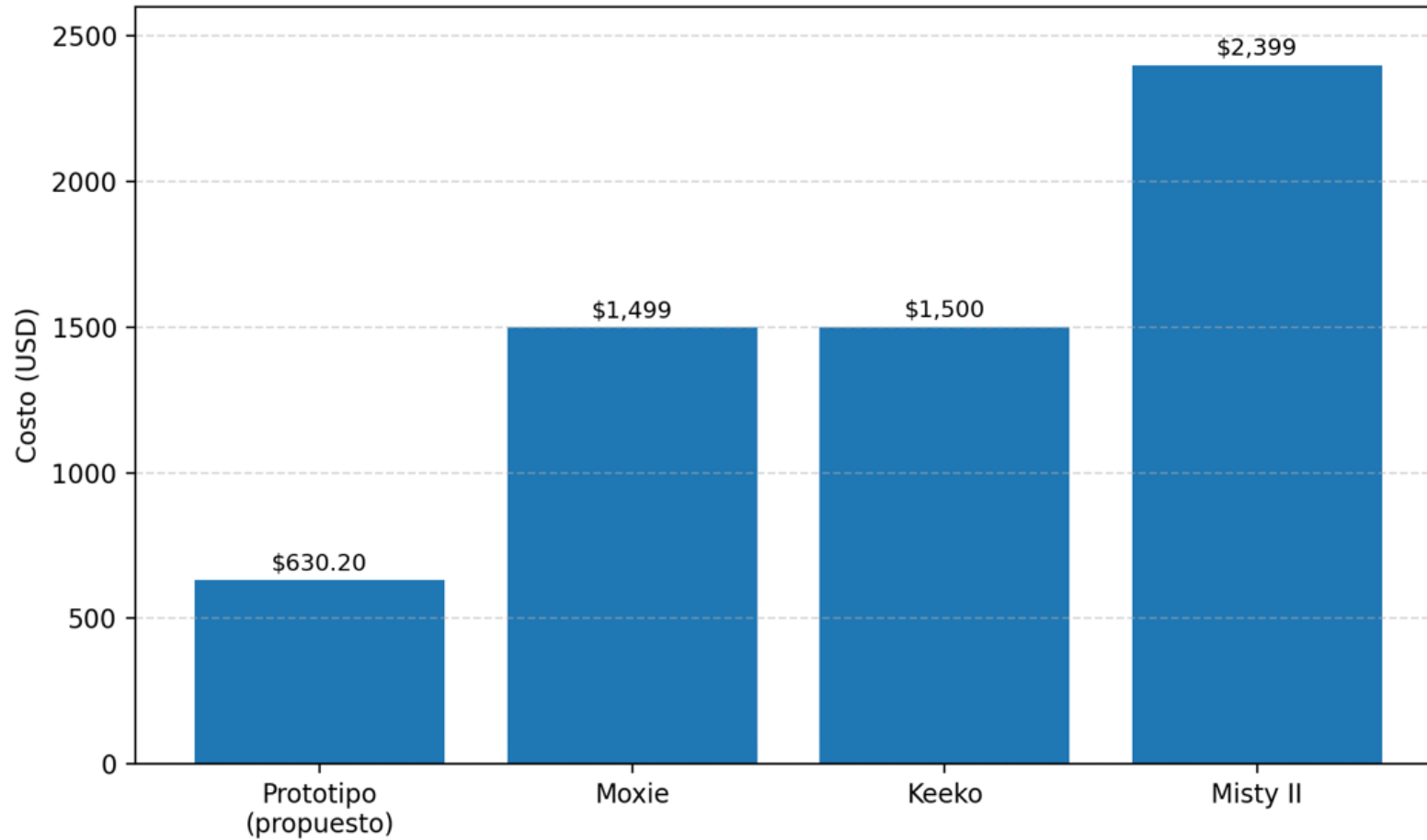
Post-procesado: lijado, aplicación de primer gris y pintura metalizada final.



Diseño final del robot tras acabado y pintura

Costo: prototipo vs. robots comerciales

Comparativa de costos del prototipo frente a robots comerciales



20 repeticiones por expresión, comparando ángulo programado vs. ejecutado

Expresión	Movimientos evaluados	Repeticiones	Cabeza / Secuencia
Felicidad	Brazos: $0^{\circ} \rightarrow (60^{\circ} \rightarrow 30^{\circ}) \times 4 \text{ ciclos} \rightarrow 0^{\circ}$	20	Cabeza inmóvil
Tristeza	Brazos: $0^{\circ} \rightarrow 40^{\circ}$ · Cabeza: $0^{\circ} \rightarrow 30^{\circ}$	20	Movimiento conjunto



100 % de confiabilidad en la ejecución de secuencias de expresión

20/20

Secuencias correctas
Felicidad

20/20

Secuencias correctas
Tristeza

Errores angulares dentro de un rango de precisión aceptable

Expresión	Posición evaluada	Promedio medido	Error porcentual
Felicidad	60° (brazos)	60,2°	0,75 %
Felicidad	30° (brazos)	30,25°	2,5 %
Tristeza	40° (brazos)	40°	0,75 %
Tristeza	30° (cabeza)	30,25°	2,5 %

 Todos los errores se mantienen dentro del rango aceptable, validando la precisión del sistema

0,75 % – 2,5 %

Rango de error
porcentual

$\leq 0,25^\circ$

Desviación angular
promedio máxima
medida

Tiempo de carga medido sobre 20 repeticiones por pantalla

PRUEBA



- Se midió el tiempo de carga de las pantallas de la HMI en 20 pruebas consecutivas.
- Los tiempos se registraron entre la selección del botón y la visualización completa de la nueva pantalla.
- Objetivo: respuesta ágil para mantener la atención de niños de 4 a 5 años.

» Carga de pantallas ágil y estable en las 20 mediciones

0,81 s

Tiempo mínimo de carga

0,86 s

Tiempo promedio de carga

0,91 s

Tiempo máximo de carga

n = 20 mediciones. Un tiempo de respuesta por debajo de 1 s favorece la atención sostenida en niños de 4 a 5 años.

1000 evaluaciones distribuidas en 10 categorías de vocabulario infantil



- 100 palabras en 10 categorías: Animales, Frutas, Colores, Números, Partes del cuerpo, Familia, Alimentos, Juguetes, Transporte y Verbos.
- Cada palabra se sometió a 10 intentos de pronunciación.
- Se registró el número de reconocimientos correctos por palabra y categoría.
- Objetivo: validar la precisión del ASR ante voz en vocabulario pedagógico.

1000

Evaluaciones totales

957

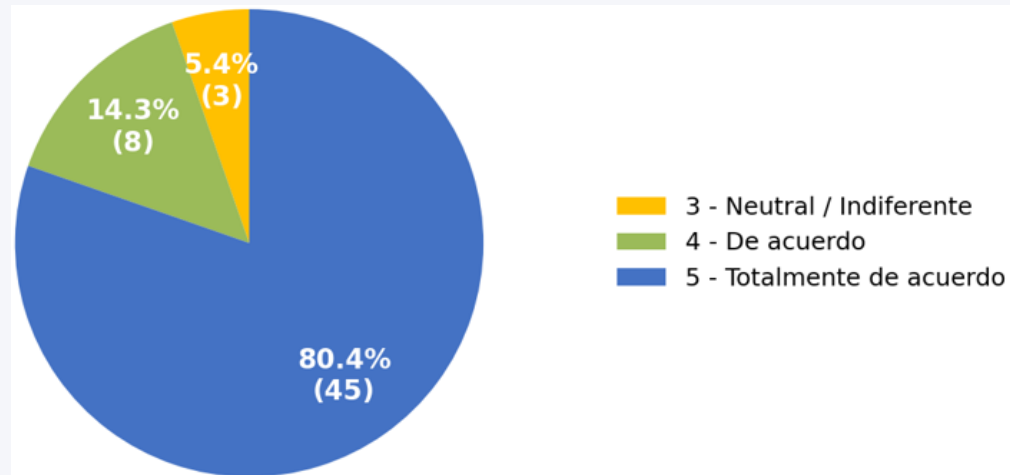
Reconocimientos correctos

95,7 %

Precisión global del sistema

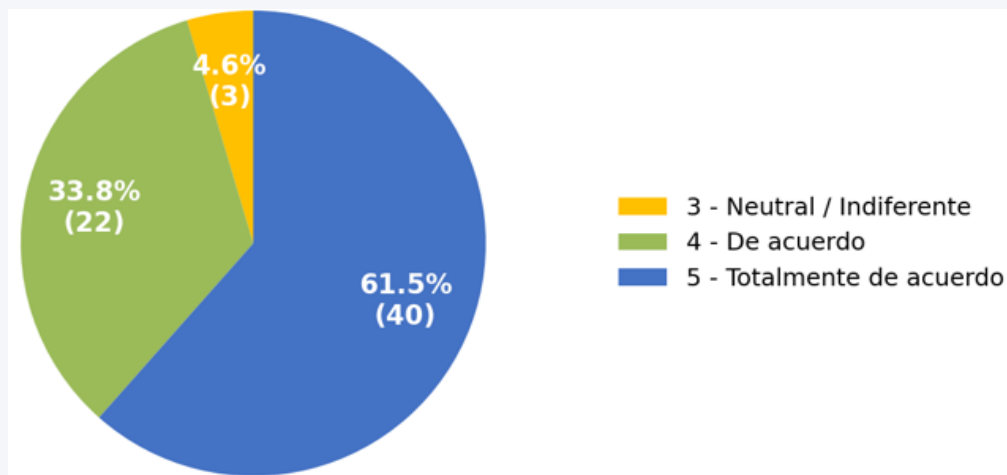
100 palabras × 10 intentos, distribuidas en 10 categorías de vocabulario infantil.

Validación con expertos: dos encuestas, un mismo patrón de respuesta



Educación Inicial — 4,75 / 5

- 4 docentes encuestados, 14 criterios en 5 dimensiones.
- Mejor calificada: personalización y ampliación de contenido (5,0).
- Menor calificación, aun así positiva: interfaz y experiencia del niño (4,5).
- 80,4 % de sus respuestas fue “totalmente de acuerdo”.



Ingeniería (Energía y Mecánica) — 4,57 / 5

- 5 docentes encuestados, 13 criterios en 4 dimensiones.
- Mejor calificada: escalabilidad y gestión de contenido (4,8).
- Menor calificación, aun así positiva: movimiento y expresividad física (4,4).
- 61,5 % de sus respuestas fue “totalmente de acuerdo”.

Conclusiones

1

Estado del arte

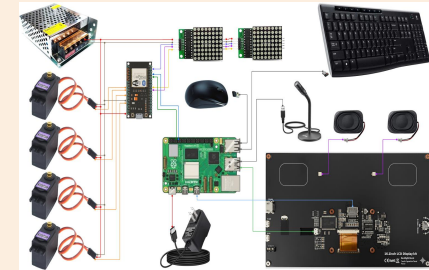
35+ fuentes · Ningún sistema integra cuentos, oraciones y pronunciación



2

Diseño mecánico, electrónico y de control

FS mecánico $\geq 7 \cdot \sigma_{\text{máx}}$ 0,366 MPa frente a 47 MPa del PLA · FS eléctrico de 1,03 a 5,51.



3

Interfaz gráfica (HMI)

14 pantallas interactivas · 3 módulos · Carga promedio 0,86 s

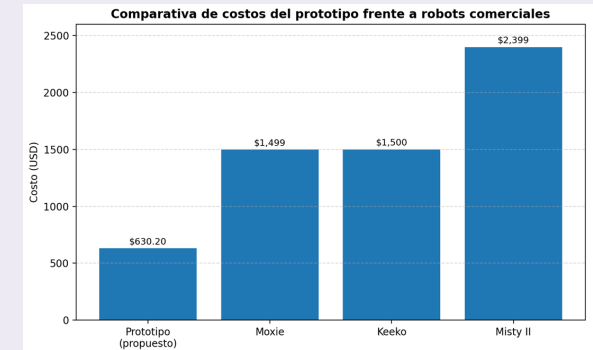


Conclusiones

4

Implementación del prototipo

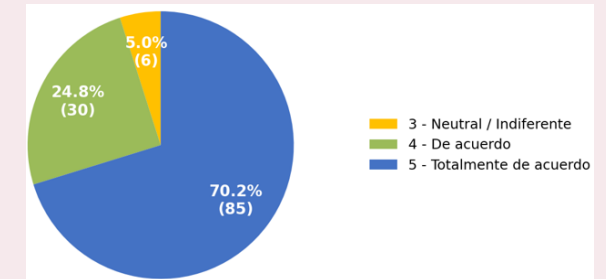
\$630,20 total · HW \$457,20 · mecánica \$90 · electrónica \$38 · PCB \$45 · software \$0 · 2,4–3,8× más barato que Misty II, Keeko y Moxie



5

Validación y pruebas de funcionamiento

100 % de confiabilidad en gestos · 95% de confiabilidad en reconocimiento de voz · 9 expertos: pedagogía + mecatrónica y mecánica · 4,75/5 y 4,57/5 · 95 % favorable



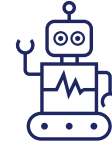
Los cinco objetivos específicos se cumplieron satisfactoriamente.

Recomendaciones Operativas

- 1 Espacio de Trabajo libre para los movimientos
- 2 Mejor desempeño en entornos con bajo ruido
- 3 Adulto presente todo el tiempo
- 4 Mantenerlo alejado del sol y la humedad



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA



¡Gracias por su atención!

Robot social para el desarrollo del lenguaje en niños de 4 a 5 años

