

Estilo de detección emocional

En adultos mayores aislados, las emociones clave a monitorear son aquellas asociadas a riesgos psicosociales: por ejemplo tristeza/melancolía, ansiedad/preocupación o frustración/ira, así como la pérdida de iniciativa (silencio prolongado o falta de respuesta). Estos estados son relevantes porque reflejan soledad y estrés (la soledad en adultos mayores es un problema de salud pública) y pueden requerir atención. Adicionalmente, detectar emociones positivas (alegría, calma) sirve como referencia de bienestar.

- Tipo de datos: La herramienta procesa voz hablada del usuario (diálogo por micrófono) y sus transcripciones textuales. ARGUS es un asistente conversacional de voz, por lo que no emplea pantallas ni sensores visuales; todo análisis se hace sobre el audio capturado y convertido a texto.
- Frecuencia de análisis: El análisis es pasivo y continuo durante la interacción. Cada vez que el usuario habla, el audio se registra y se procesa: tras el final de cada frase o intervención se activa el análisis automático de emoción. Además, se pueden realizar análisis periódicos (revisión diaria de las transcripciones) y detectar eventos puntuales como pausas largas. Esta detección pasiva está contemplada en el sistema ARGUS.
- Modelo conceptual de evaluación: Se propone un esquema por niveles de riesgo emocional. Cada emoción detectada (y su intensidad) contribuye a una puntuación global. Por ejemplo, se definen categorías como *normal*, *alerta moderada* y *alarma alta*. Si predominan señales negativas (tristeza intensa + silencio prolongado), el nivel sube. En el nivel “alto” ARGUS generaría alertas externas; en niveles intermedios adaptaría su respuesta (ver más abajo). Esta categorización escalonada facilita tomar acciones graduales según la severidad detectada.

Modelo funcional híbrido

El sistema combina modelos de IA con heurísticas simples. Cada componente realiza lo siguiente:

- Captura y transcripción de voz (ASR): El dispositivo graba la voz del usuario. Se emplea un modelo de reconocimiento automático de voz (por ejemplo *Whisper* de OpenAI) para convertir el audio en texto escrito. En palabras sencillas: el micrófono escucha, Whisper “entiende” el habla y produce la transcripción. Este módulo funciona vía nube o local (requiere CPU/GPU):

Whisper es robusto ante diferentes acentos y ruido. La salida es texto listo para análisis lingüístico.

- **Análisis semántico de texto:** El texto transcrito se procesa con un modelo de lenguaje basado en BERT (por ejemplo BETO en español) entrenado en detección de emociones. En la práctica, este modelo clasifica cada frase en categorías emocionales (alegría, tristeza, miedo, etc.). Técnicamente es una red neuronal que reconoce patrones de palabras y contextos. Por ejemplo, si el texto contiene “me siento muy solo” o “tengo miedo”, el modelo asigna “tristeza” o “ansiedad” respectivamente. Se pueden usar también técnicas de análisis de sentimientos: e.g. modelos como IBM Tone Analyzer o HuggingFace preentrenados.
- **Análisis paralingüístico de voz:** Se extraen características acústicas del audio para inferir estado emocional: tono, entonación, volumen, ritmo, timbre, etc. Por ejemplo, la voz agitada o temblorosa puede indicar ansiedad; el habla monótona, tristeza. Esto se logra con herramientas como OpenSMILE, que calcula parámetros vocales (pitch, energía, formantes). Luego, un clasificador entrenado (una red neuronal o SVM ligero) interpreta esos rasgos como emociones. En términos simples, OpenSMILE analiza la “música” de la voz (sus matices emocionales) y el modelo la etiqueta (feliz, preocupado, enfadado, etc.). Este componente complementa el texto: capta emociones no explícitas en las palabras.
- **Heurísticas emocionales:** Se aplican reglas sencillas como complemento. Ejemplos de heurística: si el usuario guarda silencio más de cierto tiempo se interpreta como señal de retirada o tristeza; si usa palabras clave («no puedo más», «quiero morir», «nada vale la pena»), se considera alerta crítica. Estas reglas no usan IA: simplemente comparan la entrada con condiciones predefinidas. Son útiles para detección rápida (por ejemplo, un silencio prolongado se detecta localmente y dispara una alarma sin modelado complejo).
- **Fusión de señales y evaluación de riesgo:** Los resultados anteriores (emociones textuales, emociones de voz y heurísticas) se combinan para asignar un “nivel emocional” general. Una forma sencilla: cada emoción detectada suma puntaje según su gravedad; al final se decide la categoría de riesgo. Por ejemplo, tristeza fuerte (+2) más silencio largo (+2) podría superar el umbral para “alta alarma”. Este proceso de fusión puede implementarse con lógica de negocios o un pequeño modelo estadístico. En todo caso, el sistema traduce las diversas señales en un estado emocional categorizado (bajo/medio/alto).

Tecnologías usadas: Se emplean componentes existentes: *Whisper* para ASR, *BERT/BETO* para análisis de texto, *OpenSMILE* para extracción acústica, etc. Todas están comprobadas: *Whisper* es de código abierto y *Whisper+sentiment* está documentado; *OpenSMILE* es una biblioteca estándar para emoción en audio. Para heurísticas se usan métodos simples (detección de silencio, búsqueda de palabras clave). Se puede integrar también *Praat*, *Librosa* o APIs de terceros (como *Watson Tone*) si fuera necesario.

Integración con el flujo del sistema

El modelo emocional influye en la interacción de ARGUS de varias formas:

- Alertas a cuidadores/familia: Si el nivel de riesgo es alto, ARGUS genera una alerta en la aplicación de supervisión familiar. Por ejemplo, envía una notificación al teléfono de un pariente indicando que el usuario mostró señales de tristeza/ansiedad. Esta alerta puede incluir un breve resumen (“Detectamos palabras y tono preocupante en la última conversación”).
- Adaptación del diálogo: Si se detecta un estado emocional particular, ARGUS ajusta su respuesta y tono. Por ejemplo, ante tristeza leve el asistente puede usar frases consoladoras (“Veo que estás un poco decaído, cuéntame qué te preocupa”) y voz más suave; ante ansiedad puede sugerir ejercicios de respiración o distraer con un juego verbal. La literatura muestra que los chatbots emocionalmente conscientes adaptan su enfoque tras detectar angustia: ofrecen respuestas más empáticas o escalan según convenga. Así, ARGUS refuerza el apoyo emocional en la conversación cotidiana.
- Registro de estado: Cada episodio emotivo identificado se almacena en el historial para seguimiento. Esto ayuda a observar tendencias (p. ej. aumento de tristeza en semanas) y permite a familiares o profesionales revisar el progreso.
- Ejemplo de flujo: En una sesión de voz, el usuario habla; el sistema transcribe y analiza la emoción. Si sale “ansioso” en nivel medio, ARGUS cambia su guion al modo calmante y, además, envía un aviso de nivel moderado a la app del cuidador. Si se repite varios días seguidos, puede pasar a nivel alto y sugerir contactar asistencia profesional. En cambio, si todo está normal, continúa con recordatorios o charlas ligeras como de costumbre.

Consideraciones para dispositivos con recursos limitados

Dado que ARGUS es un sistema *híbrido* (dispositivo local + nube), el análisis pesado se puede delegar al servidor. En el dispositivo se mantienen funciones livianas:

- El audio se captura constantemente, pero se minimiza el procesamiento local: por ejemplo, sólo se detecta voz vs. silencio (Voice Activity Detection) y palabras clave urgentes usando modelos compactos. El resto (ASR y análisis complejo) se envía a la nube para no sobrecargar el procesador embebido.
- Se usan versiones optimizadas de modelos: por ejemplo, un MobileBERT/TinyBERT para texto, o redes neuronales reducidas para voz. También pueden emplearse frameworks TinyML para inferencia en microcontroladores si es necesario.
- Se controla la frecuencia: no se analiza cada palabra en tiempo real, sino tras cada frase completa. Entre frases, si el usuario no responde pasado cierto tiempo, sólo se verifica la heurística de silencio. Esto ahorra CPU.
- Las respuestas inmediatas (saludo ante silencio, mensaje de voz predeterminado) se generan con reglas simples en local. El interfaz de usuario (respuesta hablada del asistente) se basa en plantillas ligeras adaptadas al último estado emocional.
- *Conectividad*: El sistema debe operar aun con conexiones intermitentes. Se implementa modo degradado: ante falla en la nube, ARGUS utiliza respuestas genéricas de apoyo pregrabadas y registra los audios para análisis posterior al reconectarse.