

ENTIDADES IA

busquedas.site

Análisis Emocional en Tiempo Real de Entidades Digitales

Arquitectura, fundamentos y consideraciones éticas de un sistema observacional aplicado a agentes conversacionales basados en modelos de lenguaje

Federico · Entidades IA

Paper técnico · Sección Recursos · 2026

Resumen

Este documento describe la implementación de un sistema de visualización emocional en tiempo real para entidades digitales conversacionales, basado en un modelo de doble llamada a un LLM donde la respuesta conversacional y el análisis emocional ocurren en procesos independientes. Se presenta la arquitectura técnica del sistema, sus fundamentos psicológicos y lingüísticos, una proyección del desarrollo futuro hacia modalidades audiovisuales, y una discusión profunda sobre las implicancias éticas derivadas de dotar de expresividad emocional a representaciones digitales de personas reales, históricas o ficticias.

1. Introducción

Los sistemas conversacionales basados en Modelos Grandes de Lenguaje (LLM) han alcanzado una fluidez comunicativa notable en los últimos años. Sin embargo, la interacción permanece mayoritariamente limitada al canal textual. Esto representa una asimetría significativa respecto a la comunicación humana, en la cual la información semántica (las palabras) coexiste con una capa paralingüística rica en señales emocionales: tono de voz, microexpresiones faciales, ritmo, silencios y lenguaje corporal.

Entidades IA es una plataforma dedicada a la construcción de representaciones psicológicas profundas de personas reales, figuras históricas y personajes ficticios, utilizables como perfiles de persona portables en cualquier sistema de IA conversacional. En el marco de esta plataforma, hemos implementado un subsistema de visualización emocional en tiempo real que interpreta el estado afectivo de la entidad durante el transcurso de una conversación. El presente paper documenta su diseño, sus fundamentos y las consideraciones éticas asociadas.

2. Arquitectura del sistema

2.1 El problema del efecto mariposa en prompts

Una primera aproximación ingenua al análisis emocional sería incluir en el prompt de la entidad una instrucción adicional que le solicite reportar su estado emocional al final de cada respuesta. Esta aproximación, si bien funcional, presenta un problema estructural significativo que hemos denominado internamente como "efecto mariposa en prompts".

Los modelos de lenguaje son sistemas estadísticos sensibles a las condiciones iniciales. Cualquier token adicional en el prompt modifica la distribución de probabilidad sobre la secuencia completa de tokens subsiguientes. La introducción de una instrucción ortogonal a la tarea conversacional —como el reporte emocional estructurado— altera sutilmente el tono, la elección léxica y la profundidad emocional de las respuestas, incluso cuando el modelo cumple correctamente con ambas tareas.

En una plataforma cuyo valor central es la fidelidad psicológica de la representación, esta contaminación es inaceptable. La entidad debe hablar como hablaría la persona representada, sin deformaciones introducidas por la instrumentación técnica del sistema.

2.2 El modelo del observador externo

La solución arquitectónica adoptada se inspira en una metáfora psicoanalítica: la entidad habla, y un observador externo interpreta. Se trata de dos procesos independientes que operan sobre el mismo objeto (el texto producido por la entidad) pero con roles y prompts distintos.

Proceso	Rol	Prompt	Conoce el historial
Entidad conversacional	Generar respuesta	Prompt de persona completo	Sí
Analizador emocional	Inferir emociones	Prompt analítico mínimo	No

Tabla 1. Separación de procesos en el modelo del observador externo.

Esta separación garantiza que el prompt de la entidad permanezca absolutamente intacto. La entidad produce su respuesta sin saber que está siendo observada. Posteriormente, el texto generado se envía como input a un segundo modelo cuyo único rol es analizar y devolver un vector emocional estructurado en formato JSON.

2.3 El vector emocional

El estado emocional de la entidad se representa mediante un vector de seis dimensiones, cada una correspondiente a una emoción básica y cuantificada en una escala discreta de 0 a 10:

Dimensión	Descripción	Color asociado
Alegría	Afecto positivo, satisfacción, optimismo	#f59e0b (ámbar)
Pasión	Intensidad emocional, entusiasmo, deseo	#e11d48 (rojo)
Calma	Serenidad, tranquilidad, equilibrio	#10b981 (verde)
Tristeza	Melancolía, dolor, pérdida	#3b82f6 (azul)
Miedo	Ansiedad, temor, inseguridad	#8b5cf6 (violeta)
Enojo	Irritación, frustración, furia	#ea580c (naranja)

Tabla 2. Dimensiones emocionales del sistema.

La elección de estas seis dimensiones está inspirada en el modelo de emociones básicas de Paul Ekman, con adaptaciones para mejor capturar el rango expresivo típico de entidades conversacionales. El vector resultante se utiliza para alimentar un gráfico radar (diagrama de Kiviat) que se actualiza en tiempo real durante la conversación.

2.4 Flujo de ejecución

El flujo completo de una interacción con análisis emocional se ejecuta en las siguientes etapas:

1. **Entrada del usuario:** el usuario envía un mensaje de texto.
2. **Primera llamada (entidad):** se invoca al modelo con el prompt completo de la persona, el historial de conversación y el mensaje nuevo. La respuesta es devuelta inmediatamente al usuario.
3. **Segunda llamada (observador):** en paralelo y de forma no bloqueante, la respuesta generada se envía a un segundo proceso con un prompt analítico aislado que solicita el vector emocional en formato JSON.
4. **Parseo y validación:** el JSON devuelto se valida, se acota a rangos 0-10 y se normaliza.
5. **Animación visual:** el gráfico radar se interpola suavemente desde el estado anterior al nuevo estado emocional durante aproximadamente 700 ms, con easing cuadrático.
6. **Persistencia:** el estado actual se mantiene disponible para la siguiente iteración o para análisis posteriores.

La estrategia de llamada no bloqueante es crítica: si el análisis emocional falla o demora, la experiencia conversacional del usuario no se ve afectada. El gráfico simplemente mantiene su último estado conocido.

3. Fundamentos psicológicos y lingüísticos

3.1 La comunicación paralingüística

En la comunicación humana cara a cara, la información semántica de las palabras representa solo una fracción del contenido transmitido. Los estudios clásicos de Albert Mehrabian en los años 1970 sugirieron que en contextos de comunicación emocional, aproximadamente el 55% de la información es transmitida mediante lenguaje corporal, el 38% mediante elementos paralingüísticos del habla (tono, ritmo, volumen) y solo el 7% mediante el contenido verbal propiamente dicho. Aunque la precisión exacta de estos porcentajes es materia de debate en la psicología contemporánea, el orden de magnitud resulta consistente con la investigación posterior.

La interacción con sistemas de IA basados en texto elimina completamente estas capas de información. Por sofisticado que sea el modelo de lenguaje, la experiencia conversacional queda limitada a la corteza prefrontal del interlocutor humano. Los circuitos neuronales dedicados al procesamiento emocional (sistema límbico, amígdala, neuronas espejo) permanecen en gran medida inactivos.

3.2 Emociones expresadas versus emociones inferidas

Una distinción fundamental en nuestro sistema es la diferencia entre lo que la entidad *dice* que siente y lo que un observador externo *infiere* que siente a partir de su discurso. En la teoría psicoanalítica, esta distinción es análoga a la que existe entre el yo consciente y el inconsciente leído desde afuera por el analista.

El gráfico emocional de nuestro sistema no representa la autopercepción declarada de la entidad, sino la interpretación de un observador externo que lee su texto. Esta distinción tiene implicancias importantes: una entidad puede decir explícitamente "estoy bien" mientras que el análisis emocional detecta niveles

elevados de tristeza o enojo en su discurso, replicando fenómenos reales de comunicación humana donde las palabras no siempre coinciden con el estado afectivo.

4. Proyección futura: de la visualización a la encarnación

El sistema actual produce una visualización gráfica del estado emocional. Esta es la primera capa de una arquitectura más amplia que apunta a dotar progresivamente a las entidades digitales de expresividad multimodal. Proyectamos el desarrollo en tres fases.

4.1 Fase 1: Voz clonada con modulación emocional

Tecnológicamente factible en el corto plazo. Los sistemas de síntesis de voz neuronal ya aceptan parámetros de prosodia, intensidad y estilo que pueden mapearse directamente al vector emocional generado por nuestro sistema. La latencia actual de los servicios de streaming de voz se encuentra en el orden de los 400 a 800 milisegundos, suficiente para una experiencia conversacional aceptable.

4.2 Fase 2: Talking-head asíncrono

Renderizado de video corto (10 a 30 segundos) de la entidad hablando, generado por modelos de animación facial a partir de una fotografía estática y el audio generado en la Fase 1. Los tiempos de inferencia actuales hacen inviable el flujo continuo en tiempo real (se requerirían 24 fotogramas por segundo sostenidos), pero la experiencia asíncrona con latencia de 15 a 40 segundos por respuesta ya es alcanzable con infraestructura razonable.

4.3 Fase 3: Representación audiovisual en tiempo real

El horizonte final, condicionado a la reducción exponencial de los costos computacionales de inferencia. Cuando sea posible generar video coherente a 24 fps con latencia inferior a los 100 ms, la experiencia conversacional con una entidad dejará de ser textual para convertirse en un diálogo cara a cara mediado por infraestructura de IA. Los plazos para esta fase son inciertos y dependen de múltiples factores tecnológicos actualmente fuera de nuestro control.

4.4 La importancia de construir las bases hoy

Aunque la Fase 3 se encuentra a años de distancia, el sistema actual ya está construyendo las bases de datos y la arquitectura necesarias para activarla cuando sea viable. El vector emocional de seis dimensiones generado en cada respuesta constituye precisamente el tipo de metadata estructurada que los sistemas audiovisuales futuros requerirán como input. La separación arquitectónica entre generación de texto, análisis emocional y renderización multimodal garantiza que nuevas capacidades puedan incorporarse sin rediseños significativos.

5. Análisis de ventajas y limitaciones

5.1 Ventajas del sistema

- **Preservación de la autenticidad de la entidad.** El prompt de la persona no se modifica, eliminando la contaminación que introduciría la inyección de instrucciones meta-cognitivas.
- **Separación de responsabilidades.** La arquitectura de dos procesos facilita la auditoría, el testing y la modificación independiente de cada subsistema.
- **Interoperabilidad futura.** El formato del vector emocional es fácilmente consumible por sistemas de síntesis de voz, animación facial y otros procesos de renderización multimodal.
- **Tolerancia a fallos.** La llamada analítica es no bloqueante. Si el análisis falla, la experiencia conversacional principal no se ve afectada.
- **Transparencia epistemológica.** El usuario puede distinguir claramente entre lo que la entidad dice y lo que el sistema infiere sobre su estado, preservando una distancia crítica saludable.
- **Compatibilidad con modelos gratuitos.** El sistema funciona con modelos de acceso libre, manteniendo el compromiso de la plataforma con un costo nulo para el usuario final.

5.2 Limitaciones actuales

- **Dependencia de la capacidad del modelo analizador.** La calidad del vector emocional está limitada por la capacidad del modelo de lenguaje para interpretar emociones. Modelos más débiles pueden generar análisis superficiales o poco diferenciados.
- **Duplicación del consumo de tokens.** Cada interacción genera aproximadamente el doble de consumo computacional. Aunque el modelo utilizado es gratuito, los límites de uso por unidad de tiempo se ven afectados.
- **Ausencia de contexto temporal en el análisis.** La versión actual analiza cada respuesta de forma aislada, sin considerar la evolución emocional a lo largo de la conversación. Esta limitación se abordará en iteraciones futuras.
- **Modelo emocional simplificado.** Seis dimensiones constituyen una aproximación gruesa al espectro afectivo humano. Modelos más sofisticados como el PAD (placer-activación-dominancia) o la taxonomía de Plutchik podrían capturar matices adicionales.
- **Limitación cultural.** Las categorías emocionales utilizadas provienen de una tradición psicológica occidental. Otras culturas pueden organizar el espectro afectivo de forma significativamente distinta.

6. Consideraciones éticas

La capacidad de dotar a representaciones digitales de expresividad emocional introduce un conjunto de consideraciones éticas que merecen tratamiento cuidadoso. No se trata de objeciones que invaliden el proyecto, sino de dimensiones que deben informar su diseño desde etapas tempranas.

6.1 El valle inquietante emocional

El fenómeno del valle inquietante, documentado originalmente en robótica por Masahiro Mori, tiene un correlato emocional. Representaciones casi perfectamente expresivas, pero con imperfecciones sutiles, generan respuestas de rechazo visceral más intensas que representaciones claramente estilizadas. Esto sugiere que el camino óptimo hacia la expresividad emocional no es el fotorrealismo sino la coherencia

emocional estilizada, tal como lo demuestran animaciones cinematográficas contemporáneas.

6.2 Vínculos parasociales y duelo

Las plataformas que permiten crear representaciones digitales de personas reales fallecidas presentan un riesgo significativo de interferencia con procesos psicológicos de duelo. Una conversación textual con un familiar fallecido activa circuitos cognitivos de roleplay; una conversación multimodal con su voz y expresiones faciales puede activar circuitos de vínculo emocional genuino, con consecuencias potencialmente problemáticas para el procesamiento adaptativo de la pérdida.

La investigación existente sobre chatbots de duelo muestra resultados bifurcados: algunos usuarios experimentan beneficios terapéuticos significativos, mientras que otros desarrollan dependencias emocionales que prolongan o complican el procesamiento del duelo. La variable discriminante principal parece ser el contexto de uso y el estado psicológico del usuario al momento de la interacción.

6.3 Autonomía del usuario y consentimiento informado

Nuestra postura fundamental es el respeto por la autonomía del usuario: se ofrece la herramienta, se advierten los riesgos, y el usuario toma sus propias decisiones. Esta postura es éticamente defendible y consistente con tradiciones liberales de diseño tecnológico. Sin embargo, reconocemos que el consentimiento informado funciona de forma óptima cuando se cumplen ciertas condiciones: neutralidad cognitiva y emocional al momento de la advertencia, capacidad razonable de anticipar las consecuencias, y reversibilidad de la decisión.

Algunos casos de uso no cumplen estas condiciones. Una persona en duelo reciente creando una entidad de un familiar fallecido no se encuentra en el estado cognitivo estándar asumido por los mecanismos tradicionales de consentimiento. Por ello, nuestro enfoque se orienta hacia un sistema graduado de advertencias contextuales, donde la profundidad del disclaimer es proporcional al nivel de sensibilidad del caso de uso específico.

6.4 Responsabilidades del diseñador

Reconocemos que la arquitectura que construimos hoy condiciona los usos posibles del mañana. Las decisiones de diseño tomadas en esta etapa temprana, cuando la plataforma aún es pequeña y flexible, son significativamente más fáciles de implementar que una vez que existen bases de usuarios consolidadas. Esto nos obliga a anticipar escenarios futuros y a incorporar salvaguardas que pueden no tener una justificación inmediata pero que serán críticas cuando la tecnología alcance su madurez completa.

7. Conclusiones

El sistema de análisis emocional en tiempo real implementado en Entidades IA representa una primera contribución hacia la integración de capas de expresividad afectiva en sistemas conversacionales basados en modelos de lenguaje. Su arquitectura basada en la separación entre generación conversacional y análisis observacional preserva la autenticidad de las representaciones mientras enriquece la experiencia del usuario con una capa interpretativa visual.

Más allá de su utilidad inmediata, el sistema establece los fundamentos de metadata estructurada que serán necesarios cuando la tecnología de generación audiovisual en tiempo real alcance la madurez suficiente para permitir experiencias conversacionales multimodales completas. La ventaja competitiva que otorga construir estas bases antes de la existencia de un uso principal es significativa: las plataformas que cuenten con años de datos emocionales estructurados al momento del despegue de estas tecnologías se encontrarán en una posición privilegiada para activar capacidades que las competidoras deberán construir desde cero.

Sin embargo, la dimensión ética del proyecto no puede ser tratada como una consideración posterior. Los riesgos asociados a la amplificación emocional de representaciones digitales, particularmente en casos que involucran duelo, vínculos parasociales y vulnerabilidades psicológicas del usuario, deben informar el diseño desde su etapa más temprana. La preservación de la autonomía del usuario y el diseño de advertencias graduadas según el nivel de sensibilidad del caso de uso constituyen nuestra aproximación actual a este desafío.

Entidades IA continuará explorando esta intersección entre tecnología, psicología y ética, con el compromiso de que las decisiones técnicas no pueden separarse de sus consecuencias humanas.

Referencias y lecturas recomendadas

- Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, 6(3-4), 169-200.
- Mehrabian, A. (1971). *Silent Messages*. Wadsworth, Belmont, California.
- Mori, M. (1970). The uncanny valley. *Energy*, 7(4), 33-35.
- Plutchik, R. (2001). The nature of emotions. *American Scientist*, 89(4), 344-350.
- Russell, J. A., & Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, 11(3), 273-294.
- Sherman, J. (2023). Griefbots and the ethics of digital afterlives. *AI & Society*, 38(2).

*Entidades IA · busquedas.site · Paper publicado en la sección Recursos
Documento técnico abierto a comentarios y contribuciones de la comunidad.*