Afinar un personaje en la sala

Un método de iteración en público para agentes de voz, con 80 versiones en tres eventos

Cris Olmedo y Kurt Sommer
Grupo de investigación Divina Machina
Documento de trabajo · Método
DM-2026-02, versión 1.0 · 30 de octubre de 2026

Resumen. Casi toda la IA conversacional se diseña en una mesa y se prueba con un guion. Un personaje que contesta una cabina telefónica pública no puede hacerse así: quien descuelga no tiene tarea ni paciencia y decide en un segundo si la voz al otro lado está viva. Describimos el método con el que afinamos a GAIA 2147: cambiar una sola cosa con la sala abierta, medir las llamadas siguientes y quedarse o deshacer.

En tres eventos salieron 80 versiones del agente, con 13 min (IC 95 % 7–25 min) de mediana entre una y la siguiente; el 74 % (IC 95 % 63–83 %) salió menos de una hora después de la anterior. En Barcelona Design Week el ritmo fue de 10,0 versiones al día y la versión mediana atendió 4 llamadas antes de ser sustituida. Damos el ciclo, la instrumentación que lo hace posible, seis reglas que salieron de medir y lo que el método no puede hacer.

Iterar en público cambia control experimental por relevancia. Cada comparación de la serie DM-2026 es observacional, y este documento propone las piezas que faltan para que deje de serlo.

Palabras clave: diseño de personajes; agentes de voz; iteración; evaluación en campo; instrumentación; prompts.

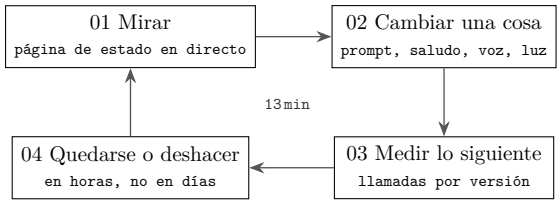


Figura 1. El ciclo de afinado. En el centro, el tiempo mediano por vuelta.

1. La tesis

No tratamos el prompt como un guion. Lo tratamos como el ajuste de un instrumento, y la sala como el lugar donde se afina [INT]. Cambiamos el personaje mientras los visitantes lo usan, medimos cada cambio en las llamadas que vienen después y nos quedamos con lo que funciona. Los papers DM-2026-03 a DM-2026-08 son los resultados de ese ciclo; DM-2026-09 describe los datos [1].

2. El ciclo

El ciclo tiene cuatro pasos (figura 1). Mirar: alguien está junto a la cabina y una página de estado muestra cada llamada. Cambiar una sola cosa: una línea del prompt, el saludo, un ajuste de voz, una luz. Medir: cada llamada registra la versión que la atendió, así que el efecto se lee en las llamadas siguientes. Quedarse o deshacer: un cambio se queda si mueve la medida que quería mover. Es rápido porque un error cuesta poco: una versión que falla está en vivo unos minutos y toca un puñado de llamadas [INT].

La tabla 1 y la figura 2 dan el ritmo [MED]. Solo el 24 de octubre el personaje cambió 23 veces. En IFA el ritmo fue mucho más lento (74 min (IC 95 % 13–205 min)), un evento con menos acceso a la cabina.

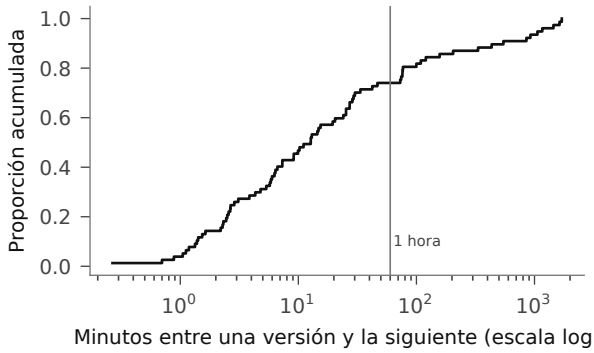


Figura 2. Distribución acumulada del tiempo entre versiones, en los tres eventos.

3. La cabina es un instrumento

El ciclo necesita una cabina que se mida a sí misma. Cada llamada registra versión, duración, cierre y, en cada turno, el tiempo de cada etapa de la respuesta. La cabina informa de su estado a una página en directo; algunos ajustes cambian desde la nube sin tocarla; las actualizaciones esperan a que se cuelgue. Dos plataformas, ElevenLabs y Vidiv, llevan el mismo personaje, y si una falla la cabina pasa a la otra (DM-2026-07). Nada de esto es ingeniería rara; lo raro es usarla en una obra en un pabellón público con la sala abierta [INT].

4. Seis reglas que salieron de medir

5. Lo que no hacemos

No guardamos las palabras de los visitantes: el análisis lee las transcripciones en memoria y conserva recuentos. No recordamos a nadie de una llamada a otra. No reiniciamos una cabina que pueda estar en una llamada. No publicamos una cifra sin decir de dónde sale: cada número de la serie sale de un script y lleva su intervalo. No inventamos ni redondeamos al alza.

Tabla 1. Ritmo de iteración por evento. Intervalos del 95 % por bootstrap.

Evento	Versiones	Días	Por día	Min. entre versiones [IC]	Llamadas por versión
Sónar+D	36	4	9,0	15 [9; 27]	5
IFA	14	4	3,5	74 [13; 205]	12
BDW	30	3	10,0	6 [3; 10]	4

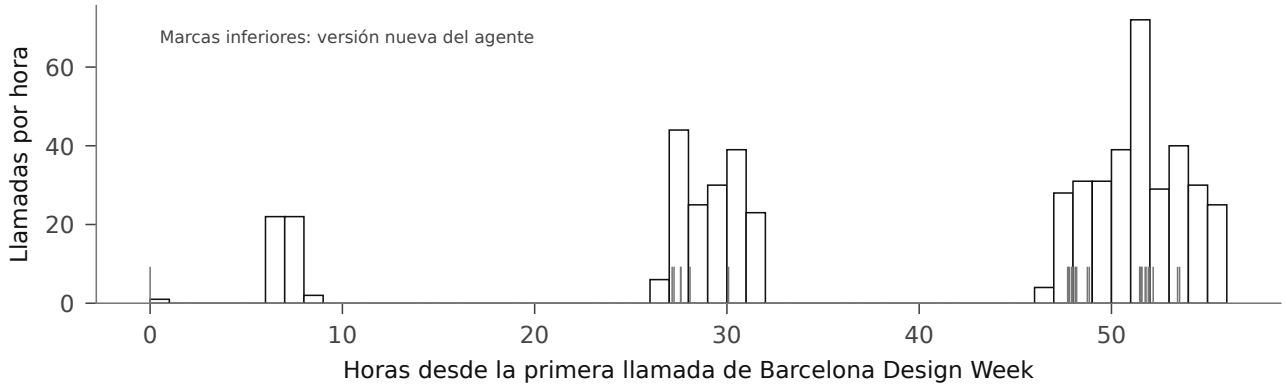


Figura 3. Barcelona Design Week: llamadas por hora y momento de cada versión nueva del agente.

6. Lo que el método no puede hacer

Iterar rápido y en público cambia rigor por relevancia. Cambian varias cosas a la vez, el público cambia a lo largo del día y no hay dos eventos iguales. Todas nuestras comparaciones son observacionales [INT]. La respuesta no es ir más despacio: es añadir al ciclo las piezas que faltan [HIP].

1. Una versión de control que funcione sin cambios en franjas alternas, para tener un contrafactual en la misma sala.
2. Asignar las variantes al azar por llamada cuando el cambio lo permita, como ya hacemos con el saludo corto.
3. Una hoja de observación por hora rellenada en la cabina.
4. El mismo límite de tiempo en cada evento.
5. Prerregistrar la medida principal antes de cada evento [2].

7. Recomendaciones para la comunidad

Para quien construya personajes de voz en espacios públicos: versionar cada cambio y guardar la versión con cada llamada; medir en horas; elegir antes la medida que el cambio debe mover; mirar dos veces, porque el primer resultado no es el estable; y separar lo que el modelo decide de lo que hace el dispositivo.

8. Datos y código

Dataset gaia-2147 [1]; cifras de `analysis/analyze.py`.

9. Conclusiones

Un personaje de voz no se escribe: se afina. Ochenta versiones en tres eventos no son desorden; son el instrumento sonando con público. Lo que queda por hacer es convertir ese ritmo en experimentos, sin perder la sala.

Firman. Cris Olmedo y Kurt Sommer, Grupo de investigación Divina Machina. *Cómo citar*: Olmedo, C. y Sommer, K. (2026). Afinar un personaje en la sala. Documento de trabajo DM-2026-02, v 1.0. Grupo de investigación Divina Machina. Divina Machina by Qs Ventures.

Referencias

Fuentes consultadas hasta el 29 de octubre de 2026.

[1] C. Olmedo and K. Sommer. GAIA-2147. descriptor de datos de 1527 conversaciones con un agente de voz en espacios públi-

Tabla 2. Reglas del método y el resultado que las sostiene. Detalle e incertidumbre en cada paper.

Regla	Resultado	Paper
Nunca empezar con silencio	Sin saludo, 71 % de llamadas cortas; con saludo corto, 7 %	DM-2026-04
Medir el silencio en la sala	Cascada 2,68 s, voz a voz 0,81 s de mediana en campo	DM-2026-03
Escribir la personalidad como hábitos contables	Una regla de tono redujo un 92 % las etiquetas sufridas, y se desgastó	DM-2026-05
El modelo pone el sentido, la sala la presencia	El modelo usó luz, pantalla o papel en el 5,2 % de las llamadas	DM-2026-06
Diseñar para quien pasa	En IFA, el 49 % de quienes descolgaron no habló	DM-2026-06
El borde lo que se toca, la nube el lenguaje	Fallos casi siempre físicos; nube fiable	DM-2026-07

cos. Documento de trabajo DM-2026-09. Grupo de investigación Divina Machina, 2026.

[2] Center for Open Science. Preregistration. Center for Open Science, s. f. URL https://www.cos.io/initiatives/prereg.