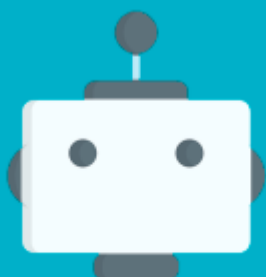


VALUEbot



**Desarrollo y validación de una plataforma
generadora de chatbots con valor público**

EDITORES

Isaac Maroto

Miguel Túnuez

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

VALUEbot

Desarrollo, validación de una plataforma generadora de chatbots con valor público

Editores

Isaac Maroto
Miguel Túñez

Esta obra es parte del proyecto titulado «Creación de una plataforma generadora de chatbots mediante IA para la comunicación del valor público del PSM. VALUEbot» (Ref. PDC2023-145885-I00), financiado por MCIN/AEI/10.13039/501100011033 y por la Unión Europea “NextGeneration EU”/PRTR.



© Universidade de Santiago de Compostela, 2026



Esta obra atópase baixo unha licenza internacional Creative Commons BY-NC-ND 4.0. Calquera forma de reprodución, distribución, comunicación pública ou transformación desta obra non incluída na licenza Creative Commons BY-NC-ND 4.0 só pode ser realizada coa autorización expresa dos titulares, salvo excepción prevista pola lei. Pode acceder Vde. ao texto completo da licenza nesta ligazón: <https://creativecommons.org/licenses/by-nc-nd/4.0/deed.gl>

Deseño de cuberta

Fonso Vidal
Edicións USC

Maquetación

Tania Sanmartín
Imprenta Universitaria

Edita

Edicións USC
Campus Vida
15782 Santiago de Compostela

DOI: <https://dx.doi.org/10.15304/op.2026.1955>

Índice

1. Introducción general	7
1.1 Justificación del proyecto	7
1.2 Contexto Audiovisual	8
1.3 La respuesta de los medios públicos europeos a la transformación del contexto mediático	9
1.4 El potencial de la IA como elemento innovador para los medios públicos europeos	11
1.5 Marco institucional y financiación	13
2. Objetivos del proyecto	14
2.1. Objetivos generales	14
2.2. Objetivos operativos y resultados esperados	15
2.3. Lógica de impacto y transferencia	16
3. Metodología iterativa y validación progresiva	17
3.1 Enfoque metodológico: Investigación-Acción Participativa (IAP) y Diseño Centrado en el Usuario (UCD)	17
3.2 Fases de desarrollo	18
Fase 1: Identificación de necesidades y elaboración del Prototipo 1	19
Fase 2: Mejora colaborativa y desarrollo del Prototipo 2	21
Fase 3: Testeo en entorno real y desarrollo del prototipo final	23
4. Características técnicas de desarrollo	24
4.1 Arquitectura general	24
4.2. Diseño modular	25
5. Herramientas técnicas implementadas	27
5.1 Retrabajando con modelos preentrenados	27
5.2 Inyección de documentos	27

5.3 Selección de modelos de IA empleados	28
5.4 Elaborando una matriz de valor público	30
6. Funcionalidades principales	34
6.1 Una interfaz intuitiva y funcional.....	34
6.2 Personalización y Administración del sistema.	
Nivel Administrador	35
6.3 Creación de proyectos y configuración de avatares.....	37
6.4 Testeo CHAT SANDBOX	42
6.5 Analítica de resultados	44
6.6 Exportación	45
6.7 Orientación de uso de los diferentes modelos	47
7. Aprendizajes y cambios integrados	50
7.1 Cambios aplicados durante las diferentes fases	51
7.2 Evaluación cualitativa	53
8. Escalabilidad y recomendaciones técnicas	54
8.1 Concurrencia y objetivo del chatbot	55
8.2 Estimación de escalado según el número de usuarios concurrentes	57
9. Conclusiones y hoja de ruta	59
9.1 El eje ético de VALUEbot	59
9.2 Hoja de ruta	61
9.3 Conclusión: coherencia entre infraestructura, valores y propósito	64
10. Bibliografía	65
11. Anexos técnicos	68
11.1 Prompts de ajustes	68
11.2 Acceso y solicitud de la demo	71

1. Introducción general

En términos generales, la expansión de la inteligencia artificial (IA) está reconfigurando los modos de producción cultural y las prácticas de comunicación social. La incorporación de sistemas algorítmicos a procesos cotidianos y estratégicos consolida nuevas dependencias respecto de grandes conglomerados tecnológicos, que fijan el ritmo del denominado progreso y condicionan la distribución de beneficios e impactos. Este desplazamiento de poder plantea una cuestión de gobernanza que no puede obviarse: cómo alinear la adopción de IA con el interés público y con marcos verificables de responsabilidad institucional. En consecuencia, este proyecto se sitúa deliberadamente lejos del triunfalismo tecnológico, porque reconoce riesgos como los sesgos, la descarga cognitiva asociada al uso intensivo, la producción masiva de contenidos de baja calidad o los usos impropios en la vida social, que requieren mitigación explícita y trazabilidad operativa. En suma, el punto de partida es ético y práctico a la vez: integrar IA de forma comprobable sin renunciar a la independencia ni a la coherencia con el mandato de servicio público.

1.1. Justificación del proyecto

Desde esta perspectiva, la noción de valor público resulta de especial relevancia. Las corporaciones audiovisuales de servicio público en Europa han reforzado su reflexión sobre legitimidad, transparencia y participación en un entorno crecientemente fragmentado y dominado por plataformas privadas. Sin embargo, Los instrumentos de comunicación y rendición de cuentas cumplen su función técnica, pero su formato no acerca. Informes anuales, auditorías y estudios de impacto quedan lejos del público y se consultan poco. Se constata, por tanto, un desafío de comunicación institucional: articular dispositivos que acerquen los contenidos, las decisiones y los valores a públicos diversos mediante lenguajes accesibles, manteniendo la verificabilidad del proceso. Esta brecha de usabilidad y de trazabilidad comunicativa legitima la búsqueda de nuevas interfaces de relación con las audiencias.

En este marco, los avances recientes en IA generativa han habilitado interfaces conversacionales capaces de interpretar corpus documentales, adaptar estilos y personalizar respuestas en función de perfiles, valores y objetivos institucionales. De este modo, la mediación entre instituciones públicas y ciudadanía puede desplazarse de canales unidireccionales a conversaciones guiadas por principios de privacidad, coherencia y soberanía tecnológica. VALUEbot surge como respuesta aplicada a esta oportunidad: una plataforma generadora de chatbots pensada para comunicar y evaluar valor público en distintos contextos, con inyección controlada de documentos, personalización por tono, valores y estilo, y despliegue tanto en entornos locales como en nube, siempre bajo criterios de alineación ética y configuración verificable. El horizonte es operativo: dotar de herramientas concretas a organizaciones, docentes y profesionales sin exigir altos conocimientos técnicos, y hacerlo con independencia de proveedores que puedan condicionar datos o decisiones.

La propuesta se alinea con los objetivos de transferencia e innovación impulsadas a nivel estatal, y sitúa la soberanía tecnológica como eje de diseño. Así, el proyecto conecta con marcos que promueven autonomía, privacidad y control de sesgos, con el objetivo de experimentar de forma responsable sobre comunicación institucional y participación ciudadana asistida por IA. En definitiva, la justificación combina prudencia y ambición: explorar lo posible sin desatender la rendición de cuentas y la coherencia con la misión pública.

1.2. Contexto Audiovisual

El ecosistema audiovisual atraviesa una transformación sostenida por la convergencia de formatos, la globalización de la distribución y la plataformización de consumos. En este escenario, el medio público enfrenta el reto de captar audiencias emergentes y sostener vínculos duraderos en entornos saturados y altamente competitivos. La experiencia reciente indica que la interactividad y la personalización se convierten en expectativas básicas, especialmente entre jóvenes habituados a contextos digitales primero, de modo que el compromiso de las audiencias, más que un indicador accesorio,

resulta de vital importancia. La IA opera aquí como catalizador organizativo y comunicativo.

En relación con este reto, el proyecto desarrolla una herramienta específica para crear chatbots orientados, entre otros usos, a la divulgación y promoción de contenidos. La lógica es sencilla y estratégica: configurar avatares capaces de dialogar con las personas usuarias, mantener la coherencia y el tono específico de un contenido, para activar dinámicas de participación que expandan el alcance de los contenidos sin perder gobernanza sobre datos y decisiones. Esta aproximación permite divulgar y realizar estrategias de comunicación que combinan notificaciones personalizadas con espacios de conversación, y añade una capa de aprendizaje institucional al registrar de forma ética las preferencias y comportamientos de uso, siempre con propósitos de mejora sin ser invasivos. La consecuencia esperable es una comunicación más situada y, a la vez, más trazable.

El valor diferencial de esta aproximación radica en su diseño de soberanía tecnológica: la plataforma admite motores locales y control granular de documentación, estilos y valores. En términos de gobernanza, esto implica la búsqueda por mantener la responsabilidad sobre el flujo de datos, habilitar la auditoría de decisiones algorítmicas y sostener la coherencia del discurso, más allá de la simple eficacia comunicativa. La promoción y la divulgación, por tanto, se entiende como un laboratorio de comunicación institucional en el que se validan hipótesis sobre lenguaje, accesibilidad y aprendizaje organizativo, con métricas que combinan volumen de interacción con información cualitativa. En suma, el contexto reclama soluciones que integren usabilidad, ética y trazabilidad como un mismo problema de diseño.

1.3. La respuesta de los medios públicos europeos a la transformación del contexto mediático

Durante la última década, las corporaciones públicas europeas han acelerado su transición digital (Tambini, 2015), incorporando producción multiplataforma, narrativas inmersivas y experiencias 360° (Rodríguez-Fernández, Sánchez-Amboage y Toural-Bran, 2018, Pérez-Seijo, Melle y Paniagua, 2018) para sostener relevancia y

confianza, en especial entre audiencias que migran hacia video bajo demanda y servicios OTT (Andersson, 2016). La literatura reciente documenta este desplazamiento y lo contextualiza como parte de una convergencia estructural de largo recorrido, con impactos en organización, contenidos y relaciones con terceros. Se verifica un tránsito desde la comunicación masiva homogénea hacia propuestas personalizadas basadas en datos (Capgemini, 2022), lo que exige nuevas capacidades de analítica y nuevas prácticas de gobernanza de modelos.

Algunos casos ilustran trayectorias complementarias. La radiotelevisión pública finlandesa Yle inició tempranamente su transformación digital con foco en captar públicos jóvenes y convertirse en plataforma de referencia; la RTBF desarrolló servicios de audio en streaming para sostener posición competitiva; y, en paralelo, diversas corporaciones han intensificado coproducciones e intercambios a través de redes como la UER, ampliando escala de contenidos y aprendizaje mutuo (Capgemini, 2022). En Portugal, RTP Arena consolidó una estrategia orientada a deportes electrónicos y cocreación en plataformas como Twitch, con resultados diferenciales en alcance y comunidad (Fieiras, Túñez y Maroto, 2022). El denominador común no es una tecnología concreta, sino la integración consistente de datos, diseño de producto y mandato de valor público.

La academia describe este proceso como plataformización, una evolución que combina adaptación y defensa del rol público frente a actores globales (van Es & Poell, 2020). Este movimiento, sin embargo, convive con la expansión de contenidos generados por usuarios en redes sociales, que multiplican la competencia por la atención y complejizan los criterios editoriales. En consecuencia, las radiotelevisiónes públicas buscan actualizar capacidades sin diluir su misión, y exploran herramientas que ayuden a traducir principios en prácticas verificables, desde la accesibilidad y la pluralidad hasta la integridad informativa y la rendición de cuentas. En este punto, la IA puede reforzar tanto la producción como la comunicación institucional, siempre que su adopción se articule con criterios de explicabilidad y responsabilidad.

La respuesta que propone este proyecto se alinea con tales aprendizajes. VALUEbot no persigue sustituir criterios editoriales ni automatizar decisiones sensibles; se orienta a operacionalizar valores en la conversación con usuarios, con controles explícitos sobre parámetros y con registros que posibilitan evaluación continua. Frente a modelos opacos, se privilegia una arquitectura que documenta la configuración de cada agente conversacional, su personalidad, su estilo y su matriz de valores, y que permite ajustar el comportamiento a principios de igualdad, diversidad, participación, cohesión, calidad y transparencia, entendidos como dimensiones operativas. En definitiva, se busca contribuir a la legitimidad de los medios públicos mediante una comunicación más inteligible, más medible y, sobre todo, más coherente con su razón de ser.

En conjunto, la justificación, el contexto y la respuesta comparada de los medios públicos europeos convergen en una tesis práctica: la integración de IA debe apoyarse en arquitecturas que garanticen soberanía tecnológica, verificación de decisiones y comunicación institucional accesible. Solo así se preserva la coherencia entre fines y medios y se hace verificable el valor público en la interacción cotidiana.

1.4. El potencial de la IA como elemento innovador para los medios públicos europeos

Las radiotelevisiones públicas europeas están transitando desde una innovación cerrada, de laboratorio interno, hacia un modelo abierto y más disruptivo. No es solo un giro técnico. Implica reformular cultura corporativa, organización y modos de producir, distribuir y comunicar contenidos para volver a encontrarse con las audiencias jóvenes, que ya se mueven con naturalidad en entornos digitales primero (Ranaivoson, Farchy y Ganesemer, 2013). Las experiencias de incubadoras y laboratorios, integradas con presencia en plataformas generalistas y redes sociales, buscan mantener la misión pública de informar, educar y entretener, al tiempo que abren la puerta a experiencias más inmersivas y conversacionales (Lin, 2015). Este es el punto, la innovación con propósito público, se convierte en un camino que comparten diferentes medios públicos.

En esa trayectoria, los análisis prospectivos sobre corporaciones de referencia, como la BBC (2024), apuntan a una intensificación del uso de tecnologías emergentes para reforzar calidad y veracidad, con más producción propia y una lógica de cocreación con la ciudadanía para que suponga una práctica que demuestre una clara intención por otorgar al público un rol como coproductor y contenidos coherencia y responsabilidad por parte del medio.

En este contexto, la inteligencia artificial funciona como palanca de innovación y como disciplina de eficiencia. Se integra en flujos de producción y verificación, en etiquetado y personalización, y en automatizaciones que ahorran tiempo para que los equipos editoriales lo inviertan donde importa, que es en criterio y servicio (Fieiras-Ceide, Vaz-Álvarez y Túñez-López, 2022). La EBU (EBU, 2019) ya señaló el avance de los grandes modelos lingüísticos en tareas de redacción y soporte de estilo. Al mismo tiempo, crecen los usos de asistentes virtuales y chatbots en procesos comunicativos de diferente escala, un efecto colateral que se está convirtiendo en un frente principal. El reto aquí es mantener trazabilidad, explicabilidad y control editorial en cada paso.

La IA también incide en la experiencia de acceso. Mejora búsqueda, descubrimiento y recomendaciones cuando se gobiernan bien los datos (Nielsen, 2024; Kshetri et al., 2024). La IA conversacional resuelve consultas frecuentes y orienta recorridos de usuario con tiempos de respuesta estables; lo hace, además, en registros cercanos que reducen fricción (Nikitha, 2021). Ahora bien, hay una condición ética y reputacional: el uso de sistemas conversacionales para crear contenidos no debería ser sustitutivo cuando erosiona confianza (European Broadcasting Union & BBC, 2025). La evidencia reciente sugiere un uso complementario, puesto al servicio de las demandas informativas concretas y con una señalización clara de qué hace la máquina y qué hace la redacción. Esa frontera honesta preserva la autoridad editorial.

De este modo, las inteligencias conversacionales abren una escena nueva para los medios públicos: interacción más cercana, ajustada a tono y valores, y una comunicación que traduce su misión a

lenguajes cotidianos. Si se articula con buena gobernanza de modelos, privacidad y coherencia editorial, esta innovación es una oportunidad para reforzar legitimidad y valor público en la práctica diaria, allí donde se juega la confianza.

1.5. Marco institucional y financiación

La propuesta se asienta en la trayectoria del Equipo de Investigación en Medios de Servicio Público (PSMresearchTeam) de la Universidad de Santiago de Compostela. El grupo, integrado en la Facultad de Ciencias de la Comunicación y vinculado al ecosistema de Novos Medios, ha consolidado en tres décadas un modelo aplicado que combina docencia, transferencia y colaboración con administraciones y empresas del sector. La experiencia acumulada no es solo académica. Es una comunidad de práctica que estudia, acompaña y evalúa transformaciones en medios públicos europeos con una mirada comparada y con atención constante a la gobernanza del cambio.

En el plano institucional, el grupo mantiene relaciones estables con RTVE, CRTVG, FORTA, CIRCOM Regional y redes internacionales especializadas en servicio público. Estas alianzas facilitan la circulación de aprendizajes y la validación en contextos reales. La coherencia metodológica se sostiene así con evidencia y proximidad operativa: lo que se diseña se prueba, y lo que se prueba se evalúa para mejorar la siguiente iteración. Esta lógica incremental permite que la innovación no pierda su anclaje en el mandato público, que es el norte de todo el proceso.

La financiación de VALUEbot se articula con el PERTE Chip, iniciativa estratégica del Gobierno de España orientada a reforzar soberanía tecnológica y digitalización avanzada mediante impulso a la microelectrónica y los semiconductores. El encaje es directo: la infraestructura de IA aplicada a comunicación pública depende de hardware y de cadenas de valor que aseguren independencia, privacidad y rendimiento. La participación del grupo en actividades asociadas a este PERTE permite explorar la convergencia entre tecnologías de base y necesidades de los medios de servicio público, en especial la automatización responsable de procesos informativos y la

innovación audiovisual con criterios de trazabilidad. En síntesis, financiación y propósito quedan alineados en la misma dirección de viaje.

En conjunto, el potencial innovador de la IA y el marco institucional que lo sostiene conforman una arquitectura de proyecto que premia la coherencia, responsabilidad y aprendizaje continuo. Frente a la adopción ad hoc, la estrategia pasa por configurar una gobernanza verificable y por analizar su efecto en la percepción del valor público. No es un atajo. Es una ruta con señales, donde cada avance deja registro y cada decisión puede explicarse con el fin de preservar la confianza de quienes miran y escuchan.

2. Objetivos del proyecto

2.1 Objetivos generales

VALUEbot nace con la finalidad de crear y validar una plataforma generadora de chatbots basada en inteligencia artificial que ayude a comunicar, evaluar y transferir el valor público de los medios audiovisuales de servicio público. Este objetivo integra dos dimensiones. Por un lado, fortalecer la legitimidad y la conexión de estas corporaciones con sus audiencias mediante experiencias conversacionales accesibles. Por otro, aportar conocimiento científico y aplicado sobre cómo integrar la IA de forma ética, inclusiva y eficaz en procesos reales de innovación y comunicación institucional.

Para ordenar este propósito, el proyecto se articula en cinco ejes que se sostienen mutuamente.

- Primero, analizar el contexto de adopción de IA en los medios públicos europeos, con sus oportunidades y sus riesgos, sin atajos.
- Segundo, identificar y llevar a la práctica las dimensiones del valor público, con atención a seis valores que orientan el camino: igualdad, diversidad, innovación, participación, cohesión social y calidad.

- Tercero, diseñar y validar una herramienta generadora de chatbots que comunique y evalúe ese valor a través de conversaciones personalizadas, comprensibles y trazables.
- Cuarto, verificar usabilidad y eficacia mediante pruebas con profesionales y audiencias en contextos reales de RTVE y CRTVG.
- Quinto, establecer mecanismos de transferencia y divulgación del conocimiento generado tanto en la academia como en la gestión cotidiana de los medios sociales.

El conjunto dibuja un objetivo general que busca la funcionalidad con el alcance social que implica integrar medios y valor públicos en el proceso.

2.2 Objetivos operativos y resultados esperados

Los objetivos operativos aterrizan la investigación en desarrollos tangibles. El primero se centra en el desarrollo tecnológico de la plataforma. Se prioriza una interfaz intuitiva que permita a usuarios no técnicos crear chatbots adaptados a distintos contextos que permitan la conversación y ofrezcan resultados funcionales.

El segundo objetivo operativo se orienta a la implementación funcional. La herramienta deberá integrarse con redes sociales para ofrecer una comunicación con el entorno mediático acorde a los gustos de las audiencias. Esta integración busca reducir fricciones y ampliar el acceso sin perder control editorial.

El tercer objetivo atiende a la evaluación y la mejora continua. Se diseñará un sistema de seguimiento de interacciones que permita analizar patrones conversacionales y valorar hasta qué punto los valores públicos se comunican y los temas que se presentan. No se trata de contar clics, sino de comprender la conversación.

El cuarto objetivo incorpora principios éticos como condición de diseño. La plataforma incluirá directrices operativas que promuevan igualdad de género, inclusión y cohesión social. Estas directrices funcionarán como salvaguardias para una comunicación responsable y alineada con los valores democráticos del servicio público.

El quinto objetivo persigue la difusión y la transferencia. Se elabora una web específica del proyecto, materiales divulgativos y talleres de formación para profesionales y académicos. A ello se suman publicaciones científicas y un manual práctico orientado a la comunicación del valor público en contextos reales.

De estos objetivos se derivan resultados esperados. Entre ellos, la validación de un prototipo operativo de chatbot, datos empíricos sobre cómo interactúan las audiencias con los valores públicos, y una base de conocimiento que oriente futuras estrategias de innovación digital en los medios. También se contempla el registro de la marca y de la propiedad intelectual de la herramienta para asegurar trazabilidad y un marco claro de transferencia tecnológica.

2.3 Lógica de impacto y transferencia

La lógica de impacto sigue una secuencia que se articula en diferentes aspectos. En primer lugar, los productos. La plataforma VALUEbot, sus módulos de personalización, el manual de uso, entre otros documentos, forman un paquete de entregables reconocible. Son piezas que se pueden tocar, configurar y replicar, como una caja de herramientas que llega con instrucciones comprensibles.

En segundo lugar, su empleo en casos concretos. Las pruebas piloto en entornos reales deberían mejorar la comprensión del valor público por parte de la audiencia, aumentar la participación cualificada y acortar la distancia entre el discurso institucional y la experiencia cotidiana. VALUEbot servirá para divulgar contenidos sobre alfabetización digital, sobre IA y desinformación; también como apoyo en procesos didácticos dentro del aula. La herramienta se ofrecerá a colectivos sociales para que puedan activar sus propias conversaciones. Los datos recogidos, siempre con criterios de privacidad y minimización, permitirán ajustar estilos y contenidos con una granularidad que antes sencillamente no existía.

En tercer lugar, la investigación. El análisis de la propia herramienta abre un camino de estudio sobre los sistemas de IA y, con él, la posibilidad de que VALUEbot trascienda como instrumento válido para equipos investigadores. Funciona a la vez como herramienta

de análisis y como vehículo de difusión de sus contenidos e informes. Una doble vida que suma rigor y alcance.

La transferencia se integra durante todo el proceso. Se han realizado cursos y formaciones para profesionales y entidades, mostrando su uso y funcionalidades para que la herramienta se adapten las necesidades específicas. El objetivo, en definitiva, es que la innovación no se evapore cuando cierre la fase de investigación, sino que permanezca, madure y se extienda con sentido.

3. Metodología iterativa y validación progresiva

3.1 Enfoque metodológico: Investigación-Acción Participativa (IAP y Diseño Centrado en el Usuario(UCD)

Para diseñar y desarrollar un prototipo tecnológico útil en el ecosistema de los medios públicos, adoptamos un enfoque mixto que combina la Investigación-Acción Participativa con el Diseño Centrado en el Usuario. La mezcla no es casual. Busca integrar la voz de quienes operan y usan el sistema con una disciplina de iteración continua que permita aprender y corregir sobre la marcha. La participación no se queda en una consulta simbólica. Se convierte en el pilar que sostiene cada decisión de diseño y de evaluación.

La Investigación-Acción Participativa (IAP) promueve la transformación social mediante la implicación activa de las personas implicadas. No se limita a observar desde fuera. Propone una colaboración horizontal que reconoce el conocimiento situado de los participantes como válido y necesario para producir cambios reales (Kemmis, McTaggart y Nixon, 2014). En este proyecto, esa lógica se traduce en la incorporación de equipos y organizaciones con responsabilidades concretas en el servicio público audiovisual, participen como interlocutores y como campo de validación. Los grupos de discusión de la Universidad de Tampere e Yle aportan reflexión académica epistemológica y funcional. El equipo de investigación de VALUEbot coordina el diálogo, pero no monopoliza las respuestas.

El Diseño Centrado en el Usuario (UCD) aporta la otra mitad del método. Sitúa usos, expectativas y contextos de las personas en el centro del desarrollo. Implica construir, evaluar y volver a construir el prototipo, siempre a partir de la retroalimentación recogida. La literatura define esta práctica desde hace décadas, con especial atención a la iteración controlada y a la adecuación al entorno de uso (Norman y Draper, 1986; ISO 9241-210:2010). La consecuencia busca que el resultado sea funcional, pero también significativo y accesible allí donde se va a desplegar.

Ambas tradiciones se tocan en un punto clave. La IAP pide reconocer la agencia de quienes participan; el UCD exige traducir esa agencia en decisiones de diseño verificables. No se trata de llevar una lista de deseos al código. Se trata de convertir esos deseos en requisitos, priorizarlos y validarlos con evidencia.

3.2 Fases de desarrollo

La metodología se organiza en tres fases encadenadas. Cada fase deja huella y prepara la siguiente. El ciclo se repite con un ritmo que reconoce la complejidad del contexto y, a la vez, evita perder tiempo en callejones sin salida.

Tabla 1. Fases de desarrollo

Fase	Descripción	Objetivo/Resultado
1. Identificación de necesidades DEMO 0 y mejora del primer (DEMO1)	A través de entrevistas y cuestionarios a miembros del ámbito académico y profesional de los medios públicos, se recogerán percepciones, expectativas y demandas sobre una primera idea de prototipo.	Establecer los requisitos iniciales de diseño y mejorar el primer prototipo. DEMO1
2. Mejora prototipo DEMO1 y reformulación (DEMO 2)	Mediante grupos de discusión con representantes de la Universidad de Tampere y de YLE, se identificarán limitaciones y oportunidades de mejora del DEMO 1	Desarrollar una versión más refinada DEMO2 incorporando aportaciones de distintos perfiles.
3. Testeo en entornos reales DEMO 2 y Prototipo Final DEMO 3	El Prototipo DEMO2 será probado en contextos reales, prestando especial atención al comportamiento de los sistemas de IA integrados.	Evaluar el funcionamiento en condiciones reales y orientar el diseño del prototipo final.

Fuente: Elaboración propia

Fase 1. Identificación de necesidades y primer refinamiento del prototipo.

Partimos de un prototipo inicial, el denominado DEMO-0, que contiene las funciones básicas definidas en la génesis del proyecto. Sobre esa base se despliega la primera recogida de información. Entrevistas semiestructuradas y cuestionarios permiten captar expectativas, usos posibles y fricciones previstas entre perfiles académicos y profesionales de los medios públicos. Esta fase bebe directamente de la IAP, porque considera a los participantes como co-constructores del conocimiento y no como informantes pasivos (Kemmis, McTaggart y Nixon, 2014). Al mismo tiempo, se alinea con el UCD, ya que explora contexto de uso, prácticas actuales y necesidades no cubiertas, que son el punto de apoyo para cualquier diseño significativo (ISO 9241-210:2010).

La selección de participantes busca diversidad de funciones, trayectorias y contextos nacionales. Se incluye a especialistas en comunicación pública, tecnologías digitales y diseño de interacción, y a profesionales en activo de organizaciones de medios públicos. Entre los perfiles previstos figuran Esther Medina de la Corporación de Radio Televisión de Galicia (CRTVG), Ángel Puche (Canal Sur), Juan Manuel Blanco Poley (Canal Sur), Marius Dragomir de Media and Journalism Research Center (MJRC), Jyri Kivimäki de la Empresa de Nacional de Radiodifusión Pública de Finlandia (Yle), junto con Mikael Hindsberg (Yle), y Petri Karlsson (Yle), y al que también se suma Egon Verharen de la Televisión Pública Noruega (NPO). Esta combinación permite contrastar necesidades estratégicas y operativas en un mismo plano de conversación, algo que a menudo se pierde cuando los equipos trabajan por separado.

Tabla 2. Participantes en las entrevistas y cuestionarios de la Fase 1

Nombre	Institución	Cargo
Esther Medina	CRTVG	Jefa de Innovación Entorno Digital
Ángel Puche	Canal Sur	Jefe de Emisiones, Canal Sur Radio
Juan Manuel Blanco Poley	Canal Sur	Jefe de Canal Sur Media
Marius Dragomir	Media and Journalism Research Center (MJRC)	Director MJRC
Jyri Kivimäki	Yle (Empresa de nacional de Radiodifusión 'Pública de Finlandia)	Productor Ejecutivo, IA y Soluciones Editoriales, Yle Newslab
Mikael Hindsberg	Yle (Empresa de nacional de Radiodifusión 'Pública de Finlandia)	Diseñador de Negocio
Petri Karlsson	Yle (Empresa de nacional de Radiodifusión 'Pública de Finlandia)	Especialista en Tecnología de Medios
Egon Verharen	NPO (Nederlandse Publieke Omroep)	Manager de Innovación

Fuente: Elaboración propia

Las técnicas de recogida se combinan por su complementariedad. Las entrevistas semiestructuradas profundizan en la experiencia y el criterio experto, y abren espacio a matices que no caben en un formulario. Los cuestionarios estructurados permiten captar información comparable y detectar patrones convergentes o divergentes. Un tercer insumo proviene de investigaciones del propio equipo, que sirven para anclar los hallazgos en marcos conceptuales y en prácticas contrastadas. La guía temática recorre funciones esperadas del prototipo, aplicaciones prioritarias, barreras de adopción y expectativas en torno a la IA aplicada a servicio público.

El resultado esperado consiste en un listado priorizado de requisitos y un mapa de oportunidades para el primer refinamiento del sistema. Por otro, una actualización del propio marco de diseño que documenta decisiones y renuncias. Con esos materiales se

construye una versión mejorada, el DEMO-1, que ya incorpora cambios visibles y prepara el terreno para la siguiente vuelta.

La fase inicial requiere una logística sencilla y rigurosa a la vez. Las entrevistas recogen trayectorias, necesidades y casos de uso prototípicos; los cuestionarios aportan una vista panorámica que ayuda a priorizar. Todo se diseña con una guía temática que explicita variables, desde funciones de la plataforma hasta barreras de adopción y expectativas respecto de la IA en medios públicos. El análisis posterior organiza la información en clústeres y traduce los hallazgos en requisitos de diseño. De ahí emerge el plan de trabajo para convertir el prototipo DEMO-0 en un prototipo DEMO-1 funcional y coherente.

Fase 2. Mejora colaborativa del prototipo DEMO 1 y desarrollo del prototipo DEMO2

Esta fase somete el prototipo DEMO-1 a una evaluación participativa para transformarlo en un prototipo DEMO 2 más sólido. De esta forma sigue el ciclo iterativo del diseño centrado en el usuario, que pide construir, escuchar, ajustar y volver a construir. El objetivo es validar críticamente el prototipo DEMO 1 mediante dinámicas grupales que identifiquen sus puntos fuertes, sus flancos débiles y las mejoras posibles, incluidas nuevas funciones.

La validación se realiza combinando entrevistas, grupos de discusión y grupos focales que combinan academia y práctica profesional. Por un lado, un grupo focal de la Universidad de Tampere, con docentes e investigadores en medios, tecnología y diseño digital. Entre las personas participantes se incluyen Ilari Kellokoski, Henri Pullinen, Anna Sendra Toset, Mikko Hautakangas, Dilara Asardag y Jaakko Suorsa, además de entrevistas de retroalimentación con Jari Väliverronen, Matleena Ylikoski y Marko Ala-Fossi.

Por otro lado, un grupo de discusión con miembros de Yle con experiencia directa en innovación y productos digitales, con la participación de Satu Keto, Mikael Hindsberg y Aura Lindeberg. La mesa es diversa a propósito: estrategia, producto, investigación y operación se miran a los ojos en el mismo espacio. (Kemmis, McTaggart y Nixon, 2014).

Tabla 3. Participantes en las entrevistas grupo de discusión y grupo focal de la Fase 2

Entrevistas	Grupo de discusión (YLE)	Grupo focal (Universidad de Tampere, 06/06/2025)
Jari Väliverronen – Profesor de Periodismo, Universidad de Tampere	Satu Keto – Líder de Innovación, YLE	Ilari Kellokoski – Profesor de Periodismo
Marko Ala-Fossi – Director del Departamento de Ciencias de la Comunicación, Universidad de Tampere	Mikael Hindsberg – Business Designer, YLE	Henri Pullinen – Investigador doctoral
Matleena Ylikoski – Investigadora en Periodismo y Pluralismo, Universidad de Tampere	Aura Lindeberg – Directora de Audio and Youth, YLE News	Anna Sendra Tosit – Investigadora postdoctoral
—	—	Mikko Hautakangas – Investigador doctoral
—	—	Dilara Asardag – Investigadora doctoral
—	—	Jaakko Suorsa – Investigador doctoral

Fuente: Elaboración propia

Las sesiones de grupo alternarán conversación guiada y análisis con estímulos. Se presentarán maquetas de pantallas, recorridos de uso, escenas de diálogo y matrices de valores. Se ofrece a cada participante que explore, compare y priorice. El fin no es el consenso rápido, sino hacer explícitas las tensiones y convertirlas en decisiones de diseño.

Del análisis cualitativo da como resultado un conjunto priorizado de observaciones y recomendaciones. Con ese material se elaborará un plan de cambios para pasar del prototipo DEMO1 al prototipo DEMO2. La nueva versión incorporará mejoras en dos planos que se tocan.

La fase se cierra con dos entregables. El primero es el prototipo DEMO 2, listo para evaluación en contextos reales y presentación a organizaciones. El segundo, un informe de decisiones de diseño que explica qué se cambió, por qué y con qué evidencias. Esta trazabilidad evita que el aprendizaje se evapore y prepara la siguiente iteración en terreno firme. En suma, la fase 2 garantiza que el desarrollo no se apoye en supuestos, sino en conversación informada y prueba controlada; una combinación que acerca la innovación a su misión pública sin perder el rumbo

Fase 3. Testeo en entorno real y desarrollo del prototipo final

La tercera fase pone el Prototipo DEMO2 frente a su prueba decisiva: el uso en contextos reales. Aquí se observa con lupa el comportamiento de los sistemas de inteligencia artificial integrados y se cierra el ciclo iterativo con una validación empírica que alimenta la última vuelta de diseño. Con el fin de identificar si la herramienta puede pasar de un plano teórico a un plano profesional.

Este tramo responde de lleno al Diseño Centrado en el Usuario. El diseño se evalúa con usuarios y en escenarios de uso reales para garantizar utilidad, usabilidad y pertinencia. No basta con intenciones, hace falta medir hechos y ajustar en consecuencia

El Prototipo DEMO 2 se probará en redacciones y áreas técnicas de medios públicos. Se realizan pruebas con equipos editoriales y tecnológicos de CMM (Televisión Castilla-La Mancha), CRTVG, pero también presentaciones del prototipo a organismos como EBU, FORTA o RTVE para tener en cuenta sus apreciaciones, puntos de vista y posibles líneas de colaboración.

El análisis de la IA se abordará en dos planos complementarios. Uno funcional, que evalúa rendimiento, cobertura lingüística, tiempos de respuesta y estabilidad. Otro ético, que revisa transparencia, interpretabilidad, grado de autonomía y alineación con los criterios de una IA responsable en medios, un debate ya bien trazado por la literatura reciente (Floridi et al., 2018).

Con los datos de campo se elaborará un informe de evaluación que agrupa recomendaciones priorizadas. Este documento guía la última iteración. Se incorporarán mejoras funcionales detectadas en

contexto, ajustes en diseño de interacción y experiencia de usuario, y revisiones críticas de los módulos de IA en función de su rendimiento y de su aceptación por parte de los equipos.

El prototipo final será, así, la consecuencia natural de un proceso iterativo, colaborativo y situado. Llega preparado para su despliegue en escenarios profesionales y sociales, sin perder de vista los principios éticos y de diseño participativo que le dieron forma desde el inicio (ISO 9241-210:2010; Kemmis, McTaggart y Nixon, 2014; Floridi et al., 2018).

4. Características técnicas de desarrollo

4.1 Arquitectura general

VALUEbot se apoya en una infraestructura híbrida que combina la coherencia de una aplicación núcleo con la flexibilidad de servicios desacoplados por eventos. El objetivo consiste en sostener comunicaciones multicanal estables y seguras y garantizar que cada interacción conserve un anclaje explícito en valores sociales configurables, estilos de comunicación y personalidades que aseguren continuidad con el propósito de valor público.

El núcleo del sistema consiste una aplicación web en Flask que gobierna la lógica funcional, el enrutamiento de peticiones y la interfaz de administración. Este núcleo opera conectado a una base de datos MySQL donde residen usuarios, proyectos, avatares, mensajes y trazas de interacción. La integración entre componentes se realiza de forma asíncrona mediante un bus de mensajería con RabbitMQ. Así, cada tarea especializada, como recordatorios, análisis o notificaciones, se procesa sin bloquear al resto, y el sistema escala en carga sin comprometer la estabilidad del núcleo.

La capacidad cognitiva se amplía con motores de IA conectables. La compatibilidad con APIs en la nube y con entornos locales como Ollama permite elegir, caso a caso, entre rendimiento, coste y soberanía tecnológica. A esta capa se suma la integración con servicios de mensajería, en particular la pasarela de WhatsApp, que

traslada la conversación a entornos cotidianos sin perder control sobre la gobernanza de datos.

La arquitectura se comporta como una membrana selectiva. Permite el intercambio entre módulos y, al mismo tiempo, centraliza el flujo de datos y eventos en el servicio núcleo, donde se aplican políticas de seguridad, privacidad y consentimiento. Variables de entorno, control de permisos y gestión de LOPD y GDPR forman parte del propio diseño, de modo que la infraestructura técnica actúa como garante de integridad comunicativa. En síntesis, no es una arquitectura neutra. Es una arquitectura con criterio, pensada para que la ética no sea un añadido, sino una condición de funcionamiento.

4.2 Diseño modular

El diseño modular organiza el sistema en áreas de responsabilidad. Flask Blueprint estructura cada ámbito funcional en un espacio autónomo que puede evolucionar sin arrastrar al conjunto.

La gestión de identidades y accesos se concentra en los ámbitos de usuarios y autenticación. A su lado, el módulo de avatares mantiene las fichas de los agentes conversacionales, con sus rasgos de personalidad, sus valores y su memoria operativa. La capa de proyectos, que actúa como contexto vivo, vincula documentos, objetivos y canales de salida. La capa de chats conserva historiales y metadatos, y ofrece las entradas necesarias para el análisis posterior. A partir de ahí, el módulo de inteligencia artificial gobierna la selección de modelos, el ajuste de parámetros y la política de uso de motores locales o en la nube. El espacio de informes transforma los registros en conocimiento accionable, con vistas que permiten leer el desempeño por proyecto, por audiencia o por valor público. El núcleo expresivo y ético se aloja en los ámbitos de valores, personalidad y estilos comunicativos, donde se definen reglas que condicionan cómo habla y cómo decide cada avatar. Junto a ellos, el módulo de consentimientos encapsula el cumplimiento normativo y traza los acuerdos con las personas usuarias.

Entre los módulos principales destacan:

- **users** (gestión de usuarios),
- **authentication** (autenticación y control de acceso),
- **avatars** (configuración y gestión de personajes IA),
- **series** (gestión de proyectos y contextos de conversación),
- **chats** (gestión de diálogos e historiales),
- **ias** (configuración de motores y modelos de IA),
- **reports** (análisis y generación de informes),
- **valor_publico, personalidad y communication_styles** (núcleo ético y expresivo del sistema),
- además de **lopds** (gestión de consentimientos y cumplimiento normativo) y **ayudas** (asistencia interna).

El servicio central o Core Service, que expone el ciclo de procesamiento, media entre estos espacios. Cada mensaje entrante se recibe, se almacena y se procesa con la configuración establecida para el proyecto y el avatar correspondiente. La respuesta se genera en el motor de IA seleccionado y regresa al canal de salida, al tiempo que se publica un evento en el bus para que otros servicios (por ejemplo, un analizador de sentimientos o un generador de reportes) actúen de manera desacoplada. El resultado es una trazabilidad completa por interacción que sirve tanto para la mejora técnica como para la auditoría interna.

La extensibilidad está prevista desde el diseño. Nuevos servicios pueden suscribirse a los eventos sin modificar el núcleo. La internacionalización mediante Babel y la compatibilidad con diversos modelos y proveedores amplían la capacidad de adaptación del sistema a contextos lingüísticos y regulatorios diferentes. Este modularidad expresa una idea de gobernanza distribuida: cada pieza hace su trabajo y todas responden a un marco común de reglas.

5. Herramientas técnicas implementadas

5.1 Retrabajando con modelos preentrenados

Para este proyecto empleamos modelos locales ya entrenados. Los utilizamos en fase de inferencia, que es el momento en que un modelo recibe una entrada y produce una salida sin volver a entrenarse. La clave está en cómo lo guiamos. Aplicamos ingeniería de prompts para orientar su comportamiento hacia un modo de hablar y de decidir que resulte coherente con la misión del servicio público. Este método consiste en ofrecer instrucciones claras, indicando roles, fijamos límites y añadimos ejemplos breves para que el modelo se sitúe.

Este trabajo con instrucciones también nos ayuda a mitigar sesgos aprendidos durante el entrenamiento. Al venir de datos de Internet, los modelos arrastran inercias. Por eso, el prompt incorpora recordatorios explícitos sobre inclusión, claridad y neutralidad. Así la conversación no solo es útil, también es más transparente.

El concepto de reorientación de valor público a través de ingeniería de Prompts dentro del sistema de VALUEBot se materializa en una matriz con seis valores principales y un total de dieciocho subvalores. La matriz funciona como un guía. Proporciona contexto ético y una pauta de estilo para orientar al modelo hacia respuestas más equilibradas. Cada valor se convierte en una regla práctica, por ejemplo: adaptar el tono y vocabulario para promover los derechos humanos o adaptar expresiones y referencias culturales según la región del usuario. Esta matriz no queda en un documento aparte. Entra en el prompt y acompaña cada interacción.

5.2 Inyección de documentos

Para dotar a los bots de contexto específico utilizamos la técnica de inyección en contexto (In context-injection). Consiste en insertar información seleccionada directamente en el texto de entrada, de forma que el modelo la use como si la tuviera en memoria. No se modifica el modelo. Se le ofrece material pertinente en el momento justo y en la cantidad adecuada.

Como los documentos reales suelen ser extensos, aplicamos un proceso de resumen de tipo mapa (Map-Reduce Summarization) y reducción. Primero partimos el documento en fragmentos manejables (Chunking). Después resumimos cada parte con criterios estables. Finalmente unimos esos resúmenes en una síntesis compacta que preserva lo esencial. Esta secuencia permite concentrar la información para que la IA pueda operar con ella sin perderse en detalles marginales.

La inyección se realiza con orden. Indicamos la fuente, el alcance temporal, la audiencia a la que se dirige el texto y el propósito comunicativo. También declaramos preferencias de estilo y límites de longitud. Con ello evitamos que el modelo improvise más de la cuenta. El resultado es una conversación que respira documentos, pero sin abrumar al usuario con tecnicismos ni con citas interminables.

5.3 Selección de modelos de IA empleados

VALUEbot integra una arquitectura híbrida que combina motores locales y servicios en la nube. El centro de la infraestructura se apoya en Ollama porque permite ejecutar modelos en servidores propios. Esta decisión protege privacidad y da soberanía tecnológica. Cuando el proyecto lo requiere, se conectan APIs de OpenAI o de Gemini para tareas puntuales. Siempre aplicando el criterio que busque el equilibrio entre rendimiento, coste y control de datos.

Trabajamos con un conjunto de modelos de lenguaje con diferentes tamaños y especialidades. Esta diversidad facilita adaptar respuestas a contextos comunicativos variados. También reduce dependencias de un único proveedor y mejora la resiliencia del sistema.

Tabla 4. IAs locales con las que trabaja VALUEbot

Nombre	Peso	Compañía / País	Características resumidas	Interés para VALUEbot
Mistral 7B	7B	Mistral AI, Francia	Eficiente, licencia Apache, secuencias largas.	Buen equilibrio para tareas locales y resúmenes.
Phi-4	14B	Microsoft Research, EE. UU.	Razonamiento y cálculo con bajo coste.	Ideal para análisis causal y planificación.
Gemma 3	12B	Google DeepMind, EE. UU.–Reino Unido	Multimodal, multilingüe, herramientas integradas.	Diálogo avanzado con control local.
Granite 3.3	8B	IBM Research, EE. UU.	Transparente y reproducible, foco institucional.	Adecuado para entornos académicos.
Qwen 2.5	14B	Alibaba Cloud, China	Multilingüe, fuerte en código y razonamiento.	Requiere auditoría ética y técnica.
Salamandra 7B Instruct	7B	BSC, España	Multilingüe, centrado en lenguas del Estado.	Alta utilidad para contextos públicos y educativos.
LLaMA 3.1	8B	Meta AI, EE. UU.	Amplia comunidad, buena conversación general.	Tutoría y diálogo semántico no comercial.
DeepSeek R1	8B	DeepSeek AI, China	Refuerzo en razonamiento lógico y de código.	Uso comparativo, no operativo.
GPT-OSS	20B	OpenAI, EE. UU.	Abierto, mezcla de expertos, cuantización eficiente.	Referencia para razonamiento avanzado.

Fuente: Datos a partir de Jiang et al. (2023); Abdin et al. (2024); Google AI, (2025), IBM Research, (2025); Qwen Team, (2025); González-Aguirre et al.(2025); Meta AI, (2024), Guo y et al. (2025); OpenAI (2025)

En conjunto, estas herramientas permiten combinar tres necesidades que a veces chocan entre sí. Necesitamos precisión y estabilidad para usos cotidianos. Necesitamos privacidad y control para cumplir

con principios de valor público y con la normativa. Y necesitamos flexibilidad para adaptar la solución a cada escenario. La ingeniería de prompts aporta la guía, la inyección de documentos aporta el conocimiento situado y la selección de modelos aporta la base técnica. Cuando estas tres piezas encajan, la conversación deja de ser un lujo y se convierte en un servicio. Un servicio que habla claro, que registra lo que hace y que se deja auditar sin excusas

5.4 Elaborando una matriz de valor público

El valor público es un concepto con aristas. No se agota en cuentas de resultados ni en métricas de audiencia. Describe los beneficios que los medios de comunicación públicos aportan a la sociedad en su conjunto, beneficios que no son puramente económicos ni inmediatos (van Es & Poell, 2020). Con la consolidación del entorno digital, el término se ha ensanchado: incorpora la defensa de los derechos comunicativos y la promoción de la cultura, e introduce con fuerza la participación como práctica cotidiana. En ese marco caben la universalidad, la creatividad, la diversidad y la cohesión social, dimensiones que ayudan a dialogar con nuevas audiencias sin perder misión ni memoria (Rodríguez-Castro et al., 2022; Donders, 2021).

La innovación aparece como hilo transversal. Empuja procesos internos, mejora la gobernanza y sostiene la calidad de los contenidos. También pide un marco de financiación y de independencia que haga posible experimentar sin hipotecar la esencia del servicio público. En otras palabras, la innovación no es un adorno: reordena prioridades y, bien gobernada, multiplica el resto de valores.

Sobre esa base, la prueba de concepto que sustenta VALUEbot trabaja con seis valores que este equipo viene identificando en proyectos e investigaciones previos de I+D+i: igualdad, diversidad, innovación, participación, cohesión social y calidad. La identificación de estos valores el resultado de analizar legislaciones, documentos corporativos, entrevistas y grupos focales con equipos de gestión de radiotelevisiones públicas europeas y con sus grupos de interés (Cañedo, Rodríguez-Castro & López-Cepeda, 2022). La matriz no pretende uniformizar realidades distintas. Pretende ofrecer un

punto de partida sobre el que orientar la programación de valor público dentro del sistema de VALUEbot.

El diseño inicial de los prompts partió de un sistema progresivo, con tres niveles para cada valor. La idea parecía limpia sobre el papel, pero en la práctica generaba confusiones. No se ajustaba a las necesidades y hacía difícil explicar por qué una respuesta tenía un “nivel 2” de participación y otra un “nivel 3”. Tras varias iteraciones, la opción se descartó. El aprendizaje fue claro: la graduación por intensidad no añadía claridad ni operatividad.

La matriz se reconfiguró, entonces, de forma granular. Cada valor principal quedó definido por tres subvalores que afinan su sentido y lo vuelven practicable en la conversación. Durante el proceso apareció una demanda recurrente, casi un subtexto que pedía nombre propio: la transparencia. Se incorporó como valor adicional porque, en un proyecto de IA aplicada a comunicación pública, la transparencia no es un adjetivo. Es una condición necesaria durante el proceso.

Así queda el mapa operativo.

Tabla 5. Matriz de valor público de VALUEbot

Valor principal	Subvalores
Igualdad	Justicia Social: adapta el tono y vocabulario para promover los derechos humanos, la igualdad de género y la no violencia, con mensajes empáticos y solidarios.
	Universalidad: garantiza una comunicación clara y accesible para todas las personas, sin importar edad, origen, idioma o nivel educativo.
	Alfabetización mediática: facilita la comprensión de conceptos complejos y promueve el pensamiento crítico y la autonomía del usuario.
Diversidad	Pluralismo: mantiene un tono neutral y respetuoso, reconociendo la multiplicidad de puntos de vista.
	Lingüística territorial: adapta expresiones y referencias culturales según la región del usuario, valorando la riqueza lingüística y territorial.
	Sociocultural: ajusta el registro comunicativo según el perfil del usuario, evitando estereotipos y representando la diversidad de forma justa.
Innovación	Transformación tecnológica: integra referencias a tecnologías emergentes o herramientas digitales de manera clara y confiable.
	Creatividad aplicada: utiliza metáforas, narrativas visuales o ejemplos originales para hacer las respuestas más dinámicas y comprensibles.
	Modelo de negocio: explica cómo las estructuras institucionales se vinculan con el valor público y la sostenibilidad.
Participación	Creación ciudadana: invita a las personas usuarias a aportar ideas, reflexiones o contenidos, reconociendo su papel como coautoras del conocimiento público.
	Escucha: transmite apertura y corresponsabilidad, mostrando cómo las opiniones del público influyen en la toma de decisiones
	Desarrollo comunitario: destaca alianzas institucionales y educativas que fortalecen el trabajo colaborativo.

Cohesión social	Territorial: promueve la empatía entre regiones mediante ejemplos que representen distintas realidades locales.
	Compromiso social: prioriza temas con impacto colectivo, abordados con un tono cercano y proactivo
	Responsabilidad y sostenibilidad: fomenta prácticas respetuosas con el entorno, ofreciendo información clara y equilibrada.
Calidad	Excelencia profesional: fomenta la precisión y claridad en la comunicación sin perder cercanía.
	Integridad e independencia: garantiza respuestas basadas en información verificada y libres de sesgos
	Contenido con valor añadido: amplía la comprensión con ejemplos, contexto y referencias culturales.
Transparencia	Asegura una comunicación institucional clara y honesta, citando fuentes cuando se ofrecen datos, reconociendo los límites del conocimiento y agradeciendo las contribuciones o correcciones del público.

Fuente: Elaboración propia

Esta matriz es la base de la integración del valor público en el prototipo. Pero no es una piedra fija. La experiencia de las iteraciones mostró que la personalización es, en sí misma, un acto de gobernanza: cada organización interpreta los valores desde su misión y su contexto. Por eso, en el nivel de administración, la matriz puede modificarse y ampliarse. Se permite crear nuevos subvalores, ajustar descriptores y asociar instrucciones de prompt específicas. La trazabilidad registra cada cambio para que el porqué y el cuándo no se pierdan.

En la práctica, cada conversación incorpora estos valores como parámetros del contexto. El sistema no “recuerda” la ética de forma abstracta. La lleva consigo en cada respuesta, como quien consulta un cuaderno de estilo vivo. Así, el valor público deja de ser un marco externo y se vuelve una dinámica interna del diseño: una brújula que acompaña al modelo y una ventana que deja ver cómo decide.

6. Funcionalidades principales

6.1 Una interfaz intuitiva y funcional

La interfaz se ha diseñado pensando en personas que no son expertas en inteligencia artificial. El objetivo es que cualquiera pueda crear y ajustar un chatbot sin perderse en tecnicismos. Para ello el sistema combina cuadros de texto, formularios claros, deslizadores y selectores que guían cada decisión. Lo que ocurre por debajo (la ingeniería de prompts que sostiene la conversación) queda automatizado y visible a la vez, de modo que el usuario puede revisar y modificar los parámetros cuando lo considere oportuno.

La navegación lateral organiza las tareas de arriba abajo, siguiendo la lógica del trabajo real. Primero la zona de administración, donde se define la base del sistema. Después el espacio de usuario, orientado a la creación de proyectos y avatares. A continuación, un área de pruebas, el entorno de *sandbox* para experimentar sin riesgos. Más abajo, la capa de analítica, que transforma las interacciones en aprendizajes. Por último, accesos rápidos para cambiar de idioma y gestionar sesiones. La estructura respira y evita el laberinto: dos acciones por pantalla como máximo, con ayuda contextual a un clic.

Imagen 1: Interfaz lateral de funcionalidades de VALUEbot

The screenshot displays the VALUEbot web application interface. On the left is a dark blue sidebar with a menu. The top of the sidebar features the VALUEbot logo and a hamburger menu icon. The menu is organized into sections: 'SETUP' (containing Usuarios, LOPD, IAs, Valores públicos, Personalidades, Estilos de comunicación, APIKEY, and Ayuda), 'MODELO' (containing Proyectos, Avatares, Avatares predefinidos, and Notificaciones), 'CHAT SANDBOX' (containing Chats), 'ANALÍTICA' (containing Reportes and Actuaciones), and 'USUARIO / ADMIN' (containing Menú de ayuda, EN, and Salir). The main content area on the right is titled 'Nuevo Usuario' and includes a breadcrumb trail 'Usuarios / Nuevo Usuario'. Below the title is a heading 'Rellena los datos del nuevo usuario'. The form consists of several input fields: 'Usuario', 'Correo', 'Contraseña (Requerido)', and 'Rol'. The 'Rol' field is currently set to 'Admin'. At the bottom of the form are two buttons: 'Guardar' (blue) and 'Cancelar' (grey).

VALUEbot

SETUP

- Usuarios
- LOPD
- IAs
- Valores públicos
- Personalidades
- Estilos de comunicación
- APIKEY
- Ayuda

MODELO

- Proyectos
- Avatares
- Avatares predefinidos
- Notificaciones

CHAT SANDBOX

- Chats

ANALÍTICA

- Reportes
- Actuaciones

USUARIO / ADMIN

- Menú de ayuda
- EN
- Salir

Nuevo Usuario

Usuarios / Nuevo Usuario

Rellena los datos del nuevo usuario

Datos Básicos

Usuario

Correo

Contraseña (Requerido)

Rol

Admin

Guardar Cancelar

Fuente: VALUEbot

6.2 Personalización de la base estructural del sistema. Nivel Administrador (SETUP)

El nivel de administración construye el armazón sobre el que se apoyan los proyectos. Aquí se configuran patrones, permisos y valores de referencia.

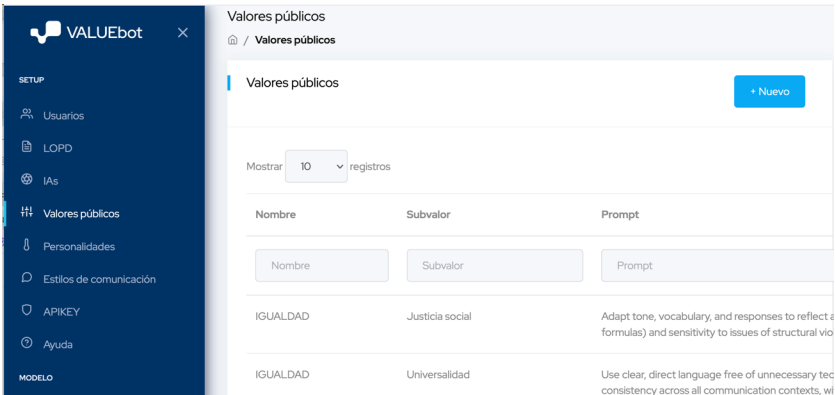
Gestión de usuarios y perfiles. La plataforma permite crear cuentas con niveles diferenciados de profundidad. El perfil de administración accede a toda la configuración. El perfil de usuario crea proyectos y avatares, realiza pruebas, consulta analítica y exporta resultados. El modo demostración permite navegar por el sistema sin alterar nada, útil para presentaciones o para formación inicial.

Cumplimiento normativo y consentimientos. El módulo LOPD permite crear, editar y versionar avisos de privacidad y términos de uso que se muestran al inicio de cada conversación, independientemente del canal desde el cual se acceda al chatbot. Cada versión queda registrada para asegurar trazabilidad y facilitar auditorías.

Motores de IA conectados. Desde SETUP se habilita el acceso a nueve modelos locales preinstalados y a los servicios de OpenAI y Gemini mediante la introducción de sus claves. La selección queda asociada a políticas de uso: escenarios locales cuando prime la soberanía tecnológica, escenarios en la nube cuando sea necesaria una capacidad puntual.

Matriz de valores del sistema. La administración puede reconfigurar la matriz de valor público: crear nuevos valores, añadir subvalores y definir los prompts asociados que guiarán el comportamiento de los avatares. Esta matriz actúa como brújula ética y como guía de estilo, por lo que cada cambio aparece documentado.

Imagen 2. Acceso a la configuración de la matriz de valores del sistema



Fuente: VALUEbot

Personalidad y estilos de comunicación. El sistema admite la edición de la matriz de personalidad y la incorporación de nuevas dimensiones si la organización lo requiere. También es posible diseñar estilos de comunicación propios para adecuarlos a la cultura institucional: institucional, divulgativo, educativo, creativo o cualquier variante que ayude a hablar en el registro correcto.

Integraciones y claves externas. El panel de API Keys concentra las credenciales para conectarse con Roblox o con avatares tridimensionales tipo “talking head” cuando el proyecto lo demande. La seguridad es explícita: cada clave se cifra y su uso queda registrado.

Centro de ayuda modular. La sección de ayuda es editable. La administración puede ordenar contenidos en árbol, crear guías breves y añadir vídeos o ejemplos para acompañar al usuario en cada paso. No es un manual escondido. Es un apoyo que vive dentro del flujo de trabajo.

6.3 Creación de proyectos y configuración de avatares (MODELO)
El corazón de la plataforma es la creación de chatbots personalizados orientados a objetivos concretos. Las funcionalidades se

organizan en una secuencia de trabajo simple: preparar el proyecto, aportar conocimiento, configurar el avatar y probar.

Asimilación de documentos y contenidos. Desde la ventana de proyecto se crea el corpus de referencia. El usuario puede subir documentos en PDF, incluir enlaces a sitios web para extraer su contenido textual o pegar directamente fragmentos relevantes. Todo lo que se añade forma el contexto que la IA utilizará para responder. Antes de inyectarse en la conversación, los textos pasan por un proceso de síntesis que equilibra precisión y manejabilidad.

Creación del avatar conversacional. Con el proyecto listo, se define el avatar: nombre, imagen y, sobre todo, su configuración profunda. Este es el punto donde las piezas encajan.

- **Elección del modelo de IA.** La plataforma sugiere el motor idóneo en función del objetivo, aunque la decisión final es del usuario. Un avatar orientado a recomendación de contenidos puede requerir un modelo; un avatar didáctico, otro.
- **Idioma de trabajo.** Se ofrece una lista de veintisiete lenguas europeas. Además, el avatar puede responder en el mismo idioma en el que se le hable, conmutando de manera automática cuando el contexto lo permite.
- **Longitud de respuesta.** Para ajustar el ritmo conversacional se ofrecen tres niveles. El modo corto favorece respuestas rápidas y precisas, útil en interfaces de mensajería y diálogos muy dinámicos. El modo normal escribe en párrafos breves, con una cadencia clara. El modo sin límite abre la puerta a explicaciones extensas, apropiadas para informes o materiales docentes.

Imagen 3: Configuración de aspectos básicos del avatar

Editar Avatar

Avatares / Editar Avatar

Edita los datos del avatar

Datos Básicos

Nombre

*FORMADOR_valuebot

Proyecto

VALUEBOT

IA

Private Phi4

Idioma

español

Responder en idioma del usuario

☒

Longitud de respuesta

Corta

Estilo de comunicación

Question Solving

☐ Mostrar Estilo de comunicación prompt resultante

Fuente: VALUEbot

- **Tipo de comunicación.** El sistema propone cinco plantillas conversacionales que cubren usos habituales:
 - **Conversational Survey.** Encuestas conversacionales donde el bot pregunta, registra y valida respuestas con un tono cercano. Útil para consultas pactadas y estudios específicos..
 - **Gamified Informational.** Información a través de juegos ligeros, retos y pequeñas mecánicas de concurso. Diseñado para contextos didácticos y de divulgación.

- **Question Solving.** Asistente orientado a resolver dudas sobre un tema específico. El emisor aporta el material y el bot responde con apoyo en la documentación.
- **Provide Samples.** Generador de ejemplos claros para explicar conceptos complejos sin infantilizar el discurso. Muy útil en formación profesional.
- **Atención al valor público.** Modo que activa la matriz ética como guía prioritaria. El avatar evita respuestas sesgadas y cuida la inclusión, la claridad y la pluralidad. La selección de valores queda visible y puede ajustarse conversación a conversación.

Imagen 4: Configuración del valor público en el avatar

Valores públicos

☒ Mostrar Valores públicos

IGUALDAD

Justicia social Universalidad Alfabetización mediática

DIVERSIDAD

Pluralismo Lingüística territorial Sociocultural

INNOVACIÓN

Transformación tecnológica Creatividad aplicada Organizacional y modelo de negocio

PARTICIPACIÓN

Creación ciudadana Escucha Desarrollo comunitario

COHESIÓN SOCIAL

Territorial Compromiso social Responsabilidad y sostenibilidad

CALIDAD

Excelencia profesional Integridad e independencia Contenido con valor añadido

TRANSPARENCIA

transparencia

Use clear, direct language free of unnecessary complexity. Accommodate different levels of complexity. Ensure consistency across all communication structures.

Explain complex concepts clearly, avoid jargon and interpretive autonomy. Add educational content that encourage informed reflection.

Adapt language use and cultural references to the target audience. Integrate part of content, not an option to contextualized and respectful manner.

Use up-to-date references to digital environment. Avoid superficial technophilia and focus on quality communication. Convey confidence in digital communication.

Invite users to contribute ideas, reflect on their own contributions as legitimate to design, and spontaneous collaboration.

Show how decisions are made with citizen participation, responsibility, and active listening. Refer to the process.

Include references that connect different territories without implicit hierarchy.

Use precise, technical, and clear language matter authority through well-structured, informality.

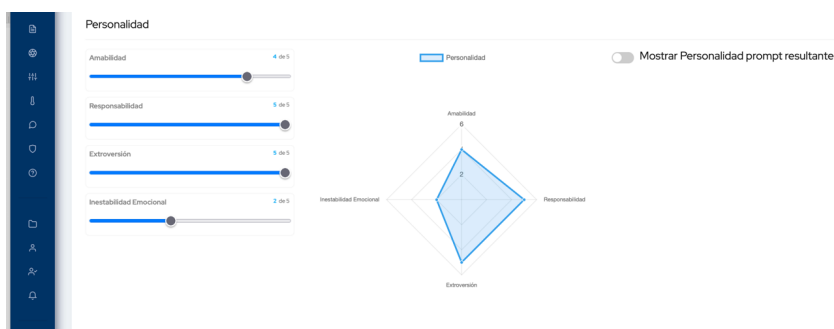
Ensure that responses are impartial and balanced.

Fuente: VALUEbot

- **Configuración de personalidad.** El avatar puede matizar su manera de hablar a partir de cuatro espectros: amabilidad, responsabilidad, extraversión e inestabilidad emocional.

Cada espectro ofrece niveles graduados que se traducen a instrucciones específicas. El resultado es una voz que se adapta al contexto: formal y académica cuando hace falta, coloquial y cercana cuando conviene.

Imagen 5: Configuración de personalidad de avatar



Fuente: VALUEbot

- **Avisos y notificaciones personalizadas.** Cuando el canal es WhatsApp, el sistema permite programar avisos que respetan el contexto de la última conversación. Sirven para informar de novedades, avisar de un nuevo episodio o recordar una actividad educativa, siempre con límites de frecuencia para evitar saturación.
- **Edición directa del prompt.** Aunque la generación de instrucciones está automatizada mediante deslizadores y selectores, cada bloque incluye una pestaña de edición manual. Quien necesite ajustar el texto final puede hacerlo sin romper el resto de la configuración.
- **Presets para acelerar la configuración.** Para crear avatares en pocos pasos se ofrecen preajustes que combinan motor de IA, longitud de respuesta, tipo de comunicación y personalidad. Son cinco aunque se pueden ajustar y crear nuevos desde el perfil de administrador:

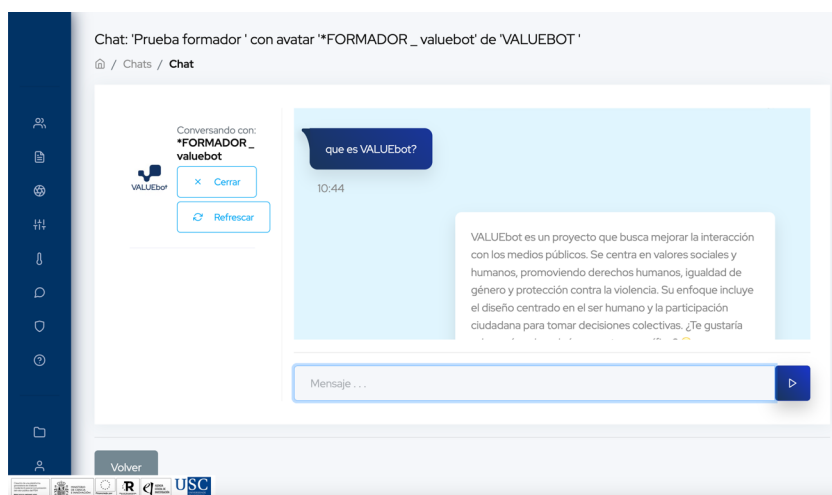
- **ValueResearch**, orientado a extraer información útil de la conversación y a organizarla para análisis posteriores.
- **ValueContent**, diseñado para recomendar contenidos de forma clara y responsable.
- **ValueLearn**, preparado para acompañar procesos de aprendizaje y divulgación.
- **ValueFiction**, pensado para adoptar una personalidad de ficción con coherencia narrativa.
- **ValueExplain**, que transforma un informe corporativo en una conversación comprensible para públicos diversos.

6.4 Testeo CHAT SANDBOX

Antes del despliegue definitivo, cada avatar pasa por un ciclo de pruebas. La plataforma incorpora herramientas que automatizan parte del proceso y ordenan el resto para que nada quede en el aire.

Prueba de chats en sandbox. El usuario puede seleccionar un avatar y abrir una sesión de prueba. La conversación se nombra para facilitar el seguimiento. La interfaz reproduce la experiencia final en entorno web, aunque también es posible simular canales de mensajería cuando el proyecto lo requiera. Aquí se detectan latencias, se corrigen tonos y se ajusta la longitud de las respuestas.

Imagen 6: Visualización de chat de prueba



Fuente: VALUEbot

Visionado y descarga. Todas las conversaciones (de prueba y de producción) pueden visualizarse y descargarse en formato de hoja de cálculo. El objetivo es doble. Por un lado, facilitar auditorías y análisis comparados. Por otro, permitir que el equipo incorpore fragmentos a informes o a sesiones de formación sin depender de copias de pantalla.

Cambio masivo de motor. Durante una campaña de evaluación puede ser necesario migrar varios avatares a otro modelo de IA. La herramienta ofrece selección múltiple y cambio simultáneo, lo que ahorra tiempo y evita inconsistencias.

Challenger. Para redoblar la calidad del test, el sistema permite poner a dialogar dos avatares asociados a proyectos distintos. Esta confrontación controlada ayuda a detectar alucinaciones, a probar roles específicos y a tensar el contexto con información parcial. Sirve también para reducir sesgos del evaluador, porque dos agentes comparables dejan menos margen a interpretaciones precipitadas.

En conjunto, estas funcionalidades convierten a VALUEbot en una herramienta práctica y gobernable. La interfaz acompaña, el nivel de administración da cimientos sólidos, el espacio de MODELO permite construir avatares con propósito y el entorno de pruebas asegura que lo diseñado se sostenga en el uso real. No hay misterio: claridad en las pantallas, reglas visibles y registro de cada decisión. Así la tecnología deja de ser un fin y vuelve a ser un medio para comunicar mejor el valor público.

6.5 Analítica de resultados

La plataforma no se limita a crear chatbots alineados con el valor público. Su sentido completo aparece cuando permite leer qué ocurre en esas conversaciones y cómo mejoran con el tiempo. Por eso incorpora un sistema de informes automatizados que combina una mirada cuantitativa con una lectura cualitativa.

El informe identifica el proyecto y los avatares implicados y genera, sin pasos extras, un dossier con métricas básicas. Se presentan gráficos sobre número de mensajes por avatar, volumen de interacciones y duración media de las sesiones. Esta capa ofrece una foto clara del uso real. A su lado, un bloque cualitativo describe los temas más tratados y recoge un análisis de sentimientos con ejemplos representativos. El objetivo es entender no solo cuánto se conversa, sino de qué se conversa y con qué tono.

Imagen 7: Ejemplo de informe automatizado



Fuente: VALUEbot

Cuando la organización introduce cambios relevantes, el sistema facilita comparaciones a lo largo del tiempo. Mediante los actuadores se marcan dos hitos, uno de inicio y otro de cierre, que enmarcan una etapa. El análisis longitudinal contrasta periodos con criterios homogéneos. Así se puede valorar, por ejemplo, el impacto de un nuevo estilo comunicativo o de una actualización en la matriz de valores. Esta lectura con fechas ayuda a decidir con calma y evita impresiones pasajeras.

6.6 Exportación

Un chatbot vale tanto como su capacidad de llegar a las personas en su lugar natural. VALUEbot ofrece varias salidas para que la interacción no se quede dentro del panel.

Para su integración en webs públicas, cada avatar dispone de un fragmento de inserción que se copia desde su propia página. Con ese código, la interfaz de chat se integra en cualquier sitio sin desarrollos complejos. El resultado es limpio y controlado, y respeta la configuración de valores y de privacidad definida en el proyecto.

Para mensajería, la integración con WhatsApp se realiza mediante un servicio de intermediación (Ultrams) que reduce la complejidad técnica. En la configuración del avatar se introducen las credenciales y nº de teléfono emisor (avatar) y el sistema se conecta. La conversación hereda el contexto del proyecto y mantiene la coherencia de estilo, aunque cambie el canal.

En entornos inmersivos, la plataforma ofrece integración con Roblox. Esta conexión permite dialogar con avatares del propio ecosistema de Roblox asistidos por VALUEbot. La ayuda incluye los pasos de configuración, porque lo importante es que la experiencia en el metaverso conserve las mismas garantías que en la web.

Imagen 8: Interacción con avatares de VALUEbot en Roblox



Fuente: Roblox VALUEbot_x4

Para experiencias audiovisuales, se habilita la conexión con un modelo tridimensional de cabeza parlante. Se trata de un modo experimental, diseñado con protección de datos explícita. Su propósito es

explorar formatos de diálogo más cercanos, donde la voz y el gesto refuercen la comprensión.

Finalmente, se estudian dispositivos autónomos en espacios físicos. La idea es permitir la interacción sin mediación del teléfono o del ordenador, de forma que el bot se integre de manera orgánica en un vestíbulo, en un aula o en una sala de exposiciones. Esta línea avanza con prudencia, porque la cercanía exige una gobernanza aún más detallada.

6.7 Orientación de uso de los diferentes modelos

De los grupos de discusión surgió una evidencia sencilla: para facilitar el uso la herramienta debía orientar la elección del modelo en función de la tarea. La evaluación comparada de motores se convirtió así en una pieza clave del diseño metodológico.

El procedimiento combina evaluación automática y contraste humano. Dos modelos (Qwen y ChatGPT-5) puntúan las respuestas siguiendo una rúbrica unificada con cinco criterios: fidelidad, claridad, profundidad, precisión y adecuación del lenguaje. Cada criterio se valora del uno al cuatro. Un prompt estandarizado proporciona el contexto y obliga a responder con formato fijo y justificaciones breves. Esta disciplina evita escapatorias y reduce la invención.

Con las puntuaciones se calculan dos indicadores. La compensación mide la concordancia entre evaluadores y es confiable cuando alcanza, como mínimo, 0,7. La desviación media calcula la diferencia entre evaluaciones y aconseja revisión si supera 0,8. En el estudio se obtuvo una compensación de 0,72 y una desviación de 0,55. La lectura es clara: hay consistencia alta entre modelos y ausencia de discrepancias graves.

Para reforzar la fiabilidad, se realizó una revisión humana estratégica sobre diez preguntas. Este segundo vistazo confirma tendencias y corrige matices que un algoritmo puede pasar por alto. El proceso no busca sustituir al equipo, sino ahorrar tiempo para que su criterio llegue donde más se necesita.

La metodología se desplegó en cinco pruebas alineadas con las capacidades centrales de la plataforma y se aplicó a los nueve modelos locales preinstalados.

- La primera prueba, tipos de comunicación, evalúa la adaptación a distintos estilos conversacionales.
- La segunda, valores públicos, analiza si la respuesta refleja la matriz ética.
- La tercera, personalidad, comprueba la coherencia entre rasgos asignados y comportamiento.
- La cuarta, contexto corporativo, mide la comprensión de informes y documentos institucionales.
- La quinta, casos ciudadanos, valora tono y adecuación en diez situaciones base.

Con las cinco pruebas se elaboró un informe integrado ponderado. Valores y asimilación cuentan un treinta por ciento cada uno. Casos base, un veinte. Comunicación y personalidad, un diez por ciento cada una. Las puntuaciones se normalizan de cero a diez para facilitar comparaciones.

Tabla 6. Resustados de modelos de test

Modelo	Valores	Comunicación	Asimilación	Casos Base	Perso- nalidad	Media Ponderada
Phi4	10.00	9.61	10.00	7.87	9.50	9.48
NubeChat-GPT	9.61	8.89	9.54	9.34	7.00	9.20
Gemma	9.29	10.00	9.57	8.41	6.50	8.99
ChatGPT OSS	8.35	9.39	9.05	9.21	9.00	8.90
Mistral	9.61	5.88	9.54	8.44	7.50	8.77
Granite	9.33	4.41	9.45	9.00	8.50	8.72
Qwen	7.28	9.39	9.97	8.72	7.50	8.61
Llama	6.46	6.45	9.51	8.61	5.50	7.71
Salamandra	6.46	8.82	9.25	5.75	5.00	7.25
Deepseek	9.25	7.20	0.00	10.00	7.00	6.20

Fuente: Elaboración propia

Los resultados dibujan un mapa útil para la selección. Phi-4 lidera con una media ponderada alta; destaca por su fidelidad a valores y su excelente asimilación, además de una personalidad adaptable. Es el candidato más completo para análisis, docencia y tareas interpretativas.

- ChatGPT OSS aparece en segundo plano, brilla en claridad comunicativa y en integración de valores, y resulta muy adecuado para formación y comunicación institucional.
- La versión en nube de ChatGPT alcanza un gran desempeño en asimilación y valores, y algo menos de plasticidad emocional. Es una opción sólida para divulgación y claridad informativa.
- Gemma se comporta bien en empatía y comunicación; puede mostrar alguna inestabilidad en personalidad, aunque compensa con buen enganche narrativo, pero experimenta problemas con la continuidad del idioma.
- Granite resulta excelente en valores y asimilación, más discreto en comunicación. Encaja en análisis técnicos donde prima la precisión.
- DeepSeek presenta un comportamiento singular. Alto en valores y en casos base, pero en asimilación profunda produjo inoperancias con el sistema.
- Mistral y Qwen se mueven en la franja alta de ocho puntos, con buen equilibrio general; son modelos cómodos para contextos creativos o interactivos que buscan cercanía.
- LLaMA y Salamandra cierran la tabla con medias más bajas, su rendimiento es más bajo en flexibilidad y en sensibilidad contextual. Son útiles para tareas rutinarias o en entornos muy controlados, donde la previsibilidad pesa más que la brillantez.

Este análisis busca ser una guía práctica que se actualiza con nuevas pruebas y con los cambios de cada modelo. La elección correcta del modelo requiere sintonía con el tono del proyecto.

7. Aprendizajes y cambios integrados

El proceso de elaboración del proyecto, apoyado en la metodología descrita en el apartado de metodología iterativa, abrió una conversación amplia. Entraron miradas técnicas y también preguntas más de fondo, casi epistemológicas. De esa mezcla salieron resultados progresivos que se fueron incorporando versión a versión. Cada vez que aparecía una demanda razonable, el sistema respondía con un ajuste. Y cuando el ajuste no bastaba, se reescribía la pieza.

VALUEbot, por diseño, admite mucha personalización. La ingeniería de prompts da juego para afinar lenguaje y tono. Los módulos de configuración permiten modular reglas, prioridades y documentación. Esta elasticidad resultó crucial para aterrizar un tema sensible: cómo traducir el valor público a comportamientos concretos del sistema. Ahí estuvo la demanda más persistente desde el inicio. Se pidió claridad y también margen de interpretación, porque cada organización tiene su acento.

En la primera versión utilizable, DEMO 1, la tabla de selección de valores se organizaba con seis deslizadores y tres niveles de intensidad por valor. La idea era elegante, pero los usuarios la sintieron confusa. Abría un debate incómodo: si en medios públicos todos los valores importan, ¿tiene sentido graduarlos como si fueran un volumen? Además, la traslación de esa gradación a los prompts no era transparente. La duda no era menor, así que se optó por reformular.

La matriz se reconstruyó con un enfoque granular. Cada valor principal pasó a desdoblarse en tres subvalores que ayudan a matizarlo y a sumarlo de forma aditiva. El resultado fue más claro para quien configura y más estable para quien conversa. El sistema ganó una brújula más fina y, al mismo tiempo, más utilizable. Esta decisión marcó un antes y un después en la percepción de control.

El avance no fue lineal. La iteración supuso una progresión adaptativa donde hizo falta concertar criterios para no perder el norte del proyecto. A continuación, se resumen los cambios principales aplicados a lo largo de las fases.

7.1 Cambios aplicados durante las diferentes fases

La tabla funciona como una línea de tiempo. Por filas resume cada eje y, por columnas, la demanda en DEMO 0, la respuesta en DEMO 1, el reto en DEMO 2 y la solución en el prototipo final. No busca exhaustividad, sino trazabilidad: muestra cómo una petición se vuelve decisión de diseño y cómo se valida o corrige en la siguiente vuelta.

Tabla 7. Tabla de cambios por Fases

Eje	Demanda / Aprendizaje DEMO0	Respuesta DEMO1	Demanda y Aprendizaje DEMO2	Respuesta Prototipo Final Demo 3
Interfaz / Frontend	Necesidad de un frontend visible y accesible para usuarios no técnicos.	Se crea una interfaz visual de pruebas (testing view) para simular la experiencia del usuario.	Los participantes reclaman una versión pública testable y pruebas con audiencias no académicas.	Test de simulación de grupo Herramienta challenger
Valores públicos	Confusión sobre los niveles de aplicación y su impacto en el chatbot.	Redefinición de la matriz con seis valores y tres subvalores; eliminación de progresión confusa.	Se pide legitimación teórica sólida, contextualización nacional y tutoriales de uso de los niveles.	Sección ayuda con Tutoriales y videotutoriales
Idiomas	Solicitud de cambio dinámico y coherencia lingüística multilingüe.	Ajuste automático de idioma incorporado.	Se detectan errores en títulos y traducciones; se pide prueba con usuarios multilingües reales.	Cambio y modificación de interfaz
Personalización / Avatar	Alta personalización visual y funcional, incluyendo voz y nuevos valores.	Se permite añadir imagen al avatar y ajustar tono y personalidad.	Se exige personalizar voz y garantizar coherencia con valores de igualdad.	Se ofrece una opción de voz experimental y manteniendo la privacidad Realización de Test valores

Eje	Demanda / Aprendizaje DEMO0	Respuesta DEMO1	Demanda y Aprendizaje DEMO2	Respuesta Prototipo Final Demo 3
Presets de comunicación	Respuestas demasiado extensas y poco esquemáticas; necesidad de claridad.	Creación de cuatro modelos preconfigurados (informativo, formativo, divulgativo, ficción).	Se pide control de alucinaciones, gestión de sarcasmo y coherencia semántica entre fuentes.	Amplia personalización en base de Prompts para ajustes emocionales concretos. Test de asimilación
Visualización de datos / Métricas	Incorporar métricas cualitativas y de sentimiento.	Implementación de KPIs cuantitativos y análisis temático y de sentimiento.	Se solicita combinar análisis cualitativo y cuantitativo e incluir medición de satisfacción del usuario.	Se añade análisis longitudinal y selectivo en el tiempo a partir de la herramienta actuadores
Modelos de lenguaje / Seguridad	Dudas sobre origen y control de modelos, necesidad de transparencia.	Inclusión de IAs locales (Mistral, Gemma3, Phi) que mejoran seguridad y control.	Se pide documentación clara de los modelos usados y criterios de selección.	Testing sistematico Integración de nuevas IAS locales
Testing y usabilidad	Faltaban pruebas con usuarios reales y documentación de integración.	Se realizan pruebas de flujo con IAs locales y escenarios críticos.	Se reclama ampliación del testing a públicos generales y evaluación de la empatía en la interacción.	Se ofrece a públicos especializados y organizaciones sociales.
Humanización / Credibilidad	Reclamaban lenguaje más natural, empático y con corrección ortográfica.	Mejora de legibilidad, reducción de latencia y control de errores.	Persiste la demanda de mayor naturalidad emocional y control semántico de interpretación y sesgos.	Mejorar la transparencia mediante directrices de verificación y reconocimiento de la incertidumbre.

Fuente: Elaboración propia

7.2 Evaluación cualitativa

El carácter iterativo del proyecto obligó a tomar decisiones cuidadosas para no comprometer privacidad ni coherencia con el valor público. La elección de ejecutar modelos locales mejoró el control, pero no resolvió todos los retos. Cuando la conversación salta a WhatsApp o a un entorno inmersivo, inevitablemente aparecen servicios de terceros. La privacidad deja de ser solo una cuestión de hardware y se convierte en una coreografía más fina de permisos, trazabilidad y minimización de datos.

En entornos inmersivos hubo que corregir rumbo. La primera opción, basada en una plataforma cerrada, impedía integrar VALUEbot con fluidez por su ritmo de cambios y su poca versatilidad. Se optó por Roblox por su comunidad activa y su capacidad de extensión. La implantación permitió alojar cuatro avatares conversacionales en un mismo escenario, aunque terminaron emergiendo limitaciones técnicas: mallas de tres dimensiones exigentes, integraciones web inestables tras actualizaciones. La lección fue clara. Se necesitaba una alternativa libre para explorar voz y representación visual en tres dimensiones sin depender de plataformas ajenas. Se desarrolló, por tanto, un prototipo experimental que refuerza privacidad y autonomía.

En la segunda fase, los grupos de discusión desplazaron la conversación hacia el significado. Se subrayó la necesidad de personalizar la matriz de valores para recoger percepciones distintas del valor público, incluso con la creación de subvalores nuevos. Pedían, además, documentación clara sobre el funcionamiento de los modelos locales y su capacidad de asimilación de funciones. Esa demanda cristalizó en la versión DEMO 2b con un testeo sistemático que evalúa cinco dimensiones: valores, asimilación documental, tipo de comunicación, adaptación de personalidad y simulación de públicos. La batería suma cien preguntas base y se aplica a los modelos locales preinstalados. Los resultados no son un ranking definitivo. Son una guía útil para elegir bien en cada caso.

Debido al retraso de las pruebas con población general debido a los altos estándares y los controles internos de las organizaciones,

organizar pruebas abiertas no se está ajustando al calendario inicial. Para no frenar el aprendizaje se activó *Challenger*, una herramienta que enfrenta dos chatbots con roles definidos para estresar el sistema y reducir sesgos del evaluador. En paralelo, se realizaron pruebas sistemáticas con usuarios tipo que representan roles sociales distintos. Se recogieron respuestas adversariales y se midió la adaptación de tono, contenido y aplicación de valores.

No todo modelo local superó las expectativas. Algunos requerían hardware que excedía el marco del proyecto. Otros no alcanzaban la estabilidad deseada. Un caso fue OpenEuroLLM-Spanish, coherente con la filosofía de su entrenamiento, pero sin el rendimiento operativo necesario para el sistema de VALUEbot.

Mirando hacia delante, el equipo identifica tres líneas de mejora principales. La primera, profundizar en evaluación multilingüe para las lenguas europeas, con especial atención a las lenguas minoritarias. La segunda, reforzar la transparencia de la plataforma con más trazas legibles para el usuario y con referencias breves que expliquen el porqué de cada respuesta cuando procede y realizar los test oportunos. La tercera, ampliar las pruebas con colectivos ciudadanos en contextos reales, buscando situaciones y dispositivos que se distancien de los clásicos como el smartphone o el pc.

En conjunto, los aprendizajes no son un epílogo. Son el motor que empuja la siguiente vuelta. La matriz de valores se afina, los modelos se eligen con evidencia, las integraciones se prueban con cautela. Y la conversación, que es lo que importa, suena un poco más clara cada día.

8. Concurrencia y escalabilidad

Cuando nos referimos a la concurrencia en el ámbito de las IAs nos referiremos al número de personas que conversan con el chatbot al mismo tiempo. No hablamos del total del mes, sino del instante en que varias sesiones conviven a la vez. En este apartado vamos a tratar de abarcar este aspecto desde la funcionalidad y el soporte tecnológico que soporta el sistema.

8.1 Concurrencia y objetivo del chatbot

Cuando nos referimos a la concurrencia en el ámbito de las IAs nos referiremos al número de personas que conversan con el chatbot al mismo tiempo. No hablamos del total del mes, sino del instante en que varias sesiones conviven a la vez. En este apartado vamos a tratar de abarcar este aspecto desde la funcionalidad y el soporte tecnológico que soporta el sistema.

Un mismo chatbot puede perseguir fines muy distintos. La forma de desplegarlo debe adaptarse tanto al propósito como al nivel de promoción previsto. No es equivalente un asistente ubicado en la web institucional que resuelve dudas o facilita el acceso a informes, a un bot diseñado para encuestas conversacionales con colectivos concretos, para apoyo a aprendizajes de idiomas o para interactuar con personajes de ficción. En estos casos la demanda simultánea suele ser baja o media. Lo importante es la disponibilidad continua y un servicio estable para quien decida conversar.

El escenario cambia cuando el bot participa en eventos de alta exposición pública. Si debe responder en directo durante un programa de televisión o acompañar una votación en segunda pantalla, la infraestructura tiene que soportar picos intensos y coordinar la interacción de cientos o miles de personas en tiempo real. Estos contextos exigen un planteamiento específico de escalabilidad, arquitectura y diseño de experiencia.

Para orientar la planificación, se propone una clasificación práctica en la siguiente tabla

Tabla 8. Tipo de uso por nivel de concurrencia

Tipo de uso del chatbot	Nivel de concurrencia esperada	Ejemplos concretos	Requisitos técnicos clave	Escalabilidad esperada
Información institucional básica	Baja 10-40	Acceso a informes, resolución de dudas frecuentes, trámites	Alta disponibilidad, baja latencia, base de datos estática	Escalado horizontal mínimo
Encuestas conversacionales dirigidas	Baja a media 10-100	Feedback de usuarios, evaluación de servicios	Capacidad de sesiones asincrónicas, persistencia	Escalabilidad moderada
Soporte educativo Recomendadores	Media 50-200	Aprendizaje de idiomas, tutor virtual, formativos Recomendador de contenidos	Personalización por usuario, memoria de contexto	Escalabilidad modular
Interacción con personajes de ficción o simulaciones	Media 50-200	Narrativas interactivas, bots lúdicos o simulaciones	Respuesta rápida, control de flujo narrativo	Moderadamente escalable
Eventos en tiempo real (TV, sorteos, directos, etc.)	Alta 1000 -XX	Concursos, votaciones, segunda pantalla, interacción durante programas	Infraestructura robusta, gestión de picos, balanceo de carga	Altamente escalable

Fuente: Elaboración propia

En el terreno de la información institucional, la demanda suele ser tranquila. Quien entra busca un informe, una respuesta breve o un apoyo para un trámite. Aquí lo que cuenta es la fiabilidad y funcionalidad. Cuando el bot se emplea para encuestas dirigidas, en función de la encuesta requiere cierta precaución para sostener sesiones asíncronas y control de los datos recogidos. La escalabilidad es moderada: más que potencia, hacen falta orden y memoria.

En soporte educativo y recomendación, la conversación se vuelve más densa. Un tutor de idiomas, una orientación formativa o un recomendador de contenidos requieren personalización y memoria de contexto para ofrecer buenos resultados.

La interacción con personajes de ficción y con simulaciones pide otro ritmo. La narrativa requiere agilidad para mantener la inmersión y que la tecnología se diluya en la experiencia de usuario.

En cuanto a los eventos en tiempo real, cuando las segundas pantallas empujan la concurrencia por encima del millar, requiere de una infraestructura robusta.

En términos generales, como referencia, una organización mediana, por ejemplo una administración local o una empresa pública, suele relacionarse con bases de entre diez mil y cien mil personas. No todas interactúan por canales digitales ni lo hacen a la vez. En campañas ordinarias, un bot informativo puede dimensionarse para gestionar entre diez y quinientas conversaciones simultáneas en horas punta. Este rango es coherente con los usos esperables y con la capacidad tecnológica disponible. La cifra final depende de la intensidad de la campaña de promoción que invite a interactuar.

8.2 Estimación de escalado según el número de usuarios concurrentes

El prototipo actual trabaja con un equipo de referencia: CPU Intel Core i7-14700K, 48 GB de RAM y una GPU NVIDIA RTX 5090 con 32 GB de VRAM. En estas condiciones, y sin cambiar de modelo ni de ventana de contexto, la plataforma sostiene entre 40 y 50 usuarios simultáneos con tiempos de respuesta bajos. Ese es el umbral operativo razonable antes de escalar en horizontal con nuevos nodos.

Para escenarios mínimos o pilotos muy acotados, basta con un equipo más modesto: procesador multinúcleo (por ejemplo, un Intel Core i5), 16 GB de RAM cuando se usa cuantización por CPU y una GPU alternativa con al menos 12 GB de VRAM, como una RTX 3060. Con esta configuración la concurrencia estimada cae a dos usuarios simultáneos. Es útil para validar flujo y entrenar equipos, no para campañas públicas.

Tabla 9. Costes de escalado por concurrencia

Concepto	50 usuarios concurrentes (PDC)*	200 usuarios concurrentes	1.000 usuarios concurrentes	10.000 usuarios concurrentes
Nº de equipos necesarios	1	4	20	200
Coste de hardware (€)	4.000 €	16.000 €	80.000 €	800.000 €
Infraestructura datacenter (€)	3.000 – 5.000 € (mínima)	15.000 – 25.000 €	25.000 – 40.000€	50.000 – 100.000 €
Consumo eléctrico estimado	0,8 kW	3,84 kW	19,2 kW	192 kW
Coste mensual electricidad (€)	125 €	600 €	3.000 €	30.000 €
Total estimado (€) (sin electricidad)	7.000 – 9.000 *€	31.000 – 41.000 €	108.000 – 123.000 €	850.000 – 900.000 €

Fuente: Elaboración propia

Para garantizar tiempos de respuesta bajos y asumiendo variabilidad, el umbral de concurrencia sostenida razonable, sin modificar el modelo, se sitúa entre cuarenta y cincuenta usuarios simultáneos. A partir de ahí se recomienda escalar en horizontal con nuevos equipos. Las cifras siguientes sirven como guía de orden de magnitud y contemplan la misma configuración de modelo y ventana de contexto.

Estas cifras de consumo eléctrico son mensuales y no se incluyen en los totales, que recogen únicamente hardware e infraestructura inicial. Sirven como base para planificar y ajustar con datos reales una vez que el servicio entre en operación.

Algunas recomendaciones operativas ayudan a que el crecimiento sea ordenado. Conviene separar planos de servicio para que la web, la mensajería y la inferencia no se afecten entre sí. También es útil activar colas visibles para el usuario cuando se superan los umbrales y habilitar modos degradados, por ejemplo, respuestas

resumidas o diferidas, que mantengan la conversación viva con una espera razonable.

La selección del modelo por campaña influye en el coste computacional. En picos previsibles puede resultar eficiente alternar temporalmente a un modelo más ligero y volver al principal al cerrar el evento. El almacenamiento en caché de respuestas recurrentes reduce latencia y libera recursos, siempre con reglas claras de privacidad y caducidad de sesión.

La observación en función a las necesidades completa el cuadro. Medir latencias medianas y percentiles altos, tasas de error por canal y niveles de saturación permite anticipar cuellos de botella. La trazabilidad que ya incorpora el sistema facilita auditorías y mejora continua. Con esta disciplina, VALUEbot puede pasar de usos cotidianos con baja simultaneidad a momentos de alta demanda sin perder estabilidad ni claridad de experiencia.

En suma, la concurrencia no es un número abstracto. Es el reflejo del objetivo comunicativo, del canal elegido y del empuje de la campaña. Una arquitectura modular, colas bien gobernadas y escalado horizontal prudente permiten crecer sin sorpresas. El plan es sencillo: definir la finalidad, estimar picos realistas, medir en producción y aprender de cada iteración para ajustar el siguiente despliegue.

9. Conclusiones y hoja de ruta

9.1 El eje ético de VALUEbot

Existe, todavía, un interés limitado por conversar con chatbots en dispositivos digitales. La gente entra, prueba, y a menudo se marcha. No siempre por desconfianza, a veces por simple cansancio de interfaces que hablan mucho y escuchan poco. VALUEbot nace en ese cruce: intenta que la tecnología sirva al propósito público integrando la conversación automatizada con un fin que no sea meramente productivo sino de alcance social.

No obstante, VALUEbot ofrece algo muy necesario dentro del ecosistema de plataformas de IA que existe en la actualidad: una manera de conversar con sentido público sin renunciar a la privacidad.

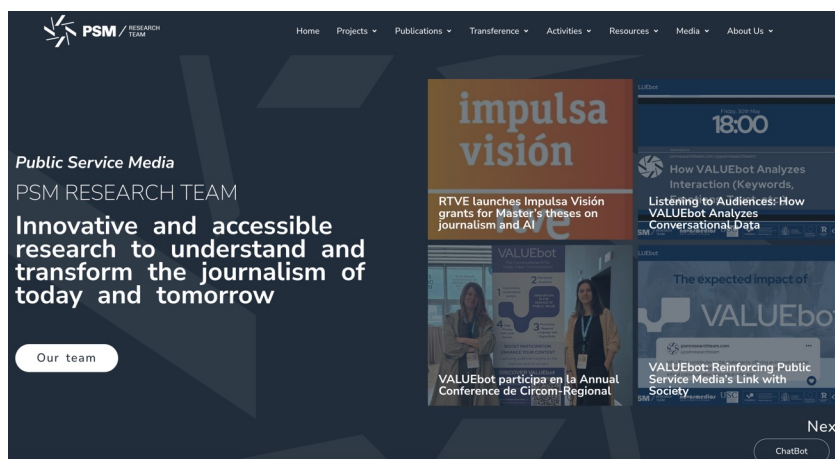
Los datos se tratan con anonimato y transparencia, sin explotación comercial ni atajos opacos. Lo que se registra sirve para mejorar el servicio y puede explicarse sin rodeos.

También aporta independencia. La arquitectura abierta evita ataduras a plataformas cerradas y permite decidir dónde se ejecuta cada pieza, con motores locales o conectados según convenga. A ello se suma la apertura auditada: el proyecto es de código abierto, legible y adaptable, de modo que cualquiera pueda revisar, aprender y proponer mejoras. No es una caja sellada, es un taller que deja ver sus herramientas permitiendo la adaptación a diferentes necesidades.

Su columna vertebral es el valor público convertido en reglas prácticas. La matriz de igualdad, diversidad, innovación, participación, cohesión social y calidad no se queda en un documento; guía cada respuesta, cada tono, cada decisión. Así la eficiencia no se mide solo en rapidez, también en cómo refuerza la autonomía y la justicia social.

Por último, aporta usos con propósito donde la tecnología se vuelve servicio. En medios públicos, mejora el diálogo con las audiencias y hace visible la trazabilidad. En colectivos ciudadanos, habilita narrativas y campañas desde la inteligencia colectiva. En educación y formación, funciona como recurso para alfabetización mediática y algorítmica, como apoyo docente para comprender las tecnologías emergentes sin perder claridad ni cercanía y como herramienta de investigación que permite el testeo sistemático de sistemas conversacionales de IA.

Imagen 9: Plataforma de contenidos e investigaciones vinculadas con VALUEbot



Fuente: <https://psmresearchteam.com>

En paralelo, el proyecto impulsa **PSM ResearchTeam** (<https://psmresearchteam.com/>), una plataforma de contenidos que reúne documentación, estudios y publicaciones sobre medios públicos europeos, valor público y adopción de IA. La estrategia de comunicación se apoya en redes como Instagram, LinkedIn y, de forma exploratoria, Roblox, con piezas breves que explican conceptos y capacidades de la herramienta accesible a públicos no especializados, con algunos formatos de video formativos para niños.

9.2 Hoja de ruta

Las tendencias recientes dibujan un paisaje mixto. La adopción de chatbots crece sobre todo entre usuarios jóvenes, entre veinte y cuarenta años. Buscan rapidez para dudas simples y, sin embargo, siguen prefiriendo trato humano cuando el asunto importa (Chatterji et al. 2025). A esto se suma un dato sensible: casi la mitad de quienes usan chatbots dice haberlos confundido alguna vez con

personas (Botpress, 2025). Si añadimos errores conocidos en cobertura informativa ((European Broadcasting Union & BBC, 2025)), se entiende que los estos sistemas pueden generar desconfianza. Por eso la señalización honesta de “estás hablando con una IA” no es un detalle; es una obligación, también cuando la interfaz se ve en un teléfono o en un ordenador que no avisa de forma intuitiva.

A la vez, la conversación sale de las pantallas. Los informes sobre tendencias muestran un aumento de interacciones con dispositivos físicos y contextos situados (Maslej et al. 2025). Allí hay una oportunidad para los medios públicos: llevar la conversación a aulas, vestíbulos, bibliotecas, plazas, sin pedir siempre un smartphone de por medio.

Con este marco, la hoja de ruta se organiza en cinco líneas.

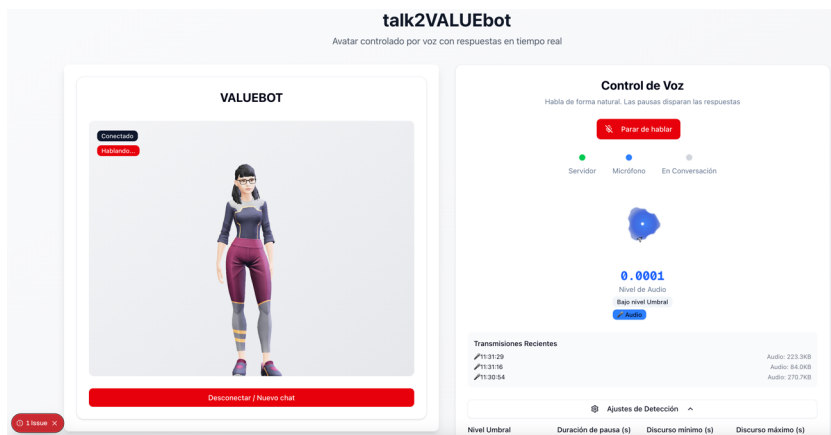
1) Despliegue físico soberano

Prototipar un dispositivo autónomo, basado en hardware asequible tipo Raspberry Pi con conectividad, micrófonos y altavoces. Alimentación con placas solares y baterías para exteriores o interiores. Objetivo: conversación fluida en espacios públicos, con la misma matriz de valores, sin depender de terceros.

2) Avatar audiovisual propio

Consolidar la integración de un avatar gráfico y sonoro dentro de la plataforma. Visualización 3D, voz y texto coordinados. Todo en ejecución controlada para preservar privacidad. El propósito no es el espectáculo, es la accesibilidad: que entender sea más fácil para más gente.

Imagen 10 : Imagen del proceso experimental del avatar audiovisual vinculado con VALUEbot



Fuente: VALUEbot

3) Transparencia reforzada

Señalización clara de identidad de la IA en todos los canales. En cada respuesta sensible, referencia breve a la fuente o al límite de conocimiento. Registro visible de cambios de modelo y de reglas aplicadas. La transparencia como hábito y no como pie de página.

4) Evaluación continua con usuarios

Extender pruebas con públicos reales, más allá de equipos técnicos. Diseñar sesiones con consentimiento explícito, recogida de impresiones cortas y métricas de claridad y confianza. Mantener *Challenger* como herramienta de estrés controlado entre bots y sumar “usuarios tipo” que representen perfiles sociales diversos.

5) Gobernanza y comunidad

Abrir ciclos regulares de contribución: issues, mejoras, traducciones, módulos. Documentación comprensible, ejemplos listos para usar y plantillas de despliegue. La comunidad como garantía de sostenibilidad.

Próximos hitos

- **Hito 1.** Despliegue del prototipo físico autónomo en dos espacios públicos, con medición de uso y percepciones.
- **Hito 2.** Convocatoria abierta para pruebas con colectivos ciudadanos y medios públicos interesados.
- **Hito 3.** Publicación de un “Glosario funcional de IA aplicada al diseño de chatbots públicos” a partir de la experiencia de VALUEbot Orientado a Profesionales de la comunicación, docentes y formadores en alfabetización digital, Investigadores en ciencias y entidades publicas.

9.3 Conclusión: coherencia entre infraestructura, valores y propósito

La documentación técnica de VALUEbot deja ver un encaje claro entre arquitectura y matriz de valor público. La ética, la personalidad, los valores sociales y la trazabilidad no son un adorno. Viven dentro del propio sistema y condicionan su manera de hablar y decidir. Desde la mirada académica, el proyecto es un caso de inteligencia conversacional de propósito público: modular, auditable, culturalmente adaptable y con preferencia real por motores locales cuando tiene sentido.

En el fondo, VALUEbot no es una colección de avatares que responden. Es una infraestructura ética y pedagógica que traduce valores en reglas de programación y convierte la técnica en una forma de institucionalidad digital. Un espacio donde la conversación automatizada convive con la integridad pública, la trazabilidad y el respeto a la diversidad.

Si las iniciativas que comunican el valor público mantienen esa coherencia mientras el ecosistema cambia, las tecnologías emergentes podrán alinearse con las necesidades sociales y no al revés. Entonces la conversación pública, la que de verdad importa, será un poco más clara y estará un poco más anclada en los valores democráticos.

10. Bibliografía

- Abdin, M., Aneja, J., Behl, H., Bubeck, S., Eldan, R., Gunasekar, S., Harrison, M., Hewett, R. J., Javaheripi, M., Kauffmann, P., Lee, J. R., Li, Y. T., Liu, W., Mendes, C. C. T., Nguyen, A., Price, E., de Rosa, G., Saarikivi, O., Salim, A., Shah, S.... Zhang, Y. (2024, diciembre). *Phi-4 Technical Report* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2412.08905>
- Andersson Schwarz, J. (2016). Public service broadcasting and data-driven personalization. *Television & New Media*, 17(2), 124–141. <https://doi.org/10.1177/1527476415616193>
- BBC. (2024). A BBC for the future. <https://www.bbc.co.uk/about-thebbc/documents/a-bbc-for-the-future.pdf>
- Botpress. (2025, 13 de mayo). Estadísticas clave de chatbots para 2025. <https://botpress.com/es/blog/key-chatbot-statistics>
- Cañedo, A., Rodríguez-Castro, M., & López-Cepeda, A. M. (2022). Distilling the value of public service media: Towards a tenable conceptualisation in the European framework. *European Journal of Communication*, 37(6), 586–605. <https://doi.org/10.1177/02673231221090777>
- Capgemini. (2022). *La radiodifusión pública tras la revolución de las OTT. ¿Salvarán la tecnología y los datos la cultura?* <https://www.capgemini.com/ar-es/insights/expert-perspectives/la-radiodifusion-publica-tras-la-revolucion-de-las-ott-salvaran-la-tecnologia-y-los-datos-la-cultura>
- Chatterji, A., Cunningham, T., Deming, D. J., Hitzig, Z., Ong, C., Shan, C. Y., & Wadman, K. (2025). How people use ChatGPT (No. w34255). National Bureau of Economic Research.
- Donders, K. (2021). *Public service media in Europe: Law, theory and practice*. Routledge.
- EBU. (2019). *EBU news report 2019: The next newsroom. Unlocking the power of AI for public service journalism*. European Broadcasting Union. https://www.ebu.ch/publications/strategic/login_only/report/news-report-2019
- European Broadcasting Union & BBC. (2025, 21 octubre). *News integrity in AI assistants: An international PSM study* [Informe].

<https://www.bbc.co.uk/mediacentre/documents/news-integrity-in-ai-assistants-report.pdf>

- Fieiras-Ceide, C., Túniz-López, M., & Maroto-González, I. (2022). Un nuevo escenario en la comunicación digital: la cobertura de deportes electrónicos en las televisiones públicas de Europa. *Revista Latina de Comunicación Social*, 80, 88–113. <https://doi.org/10.4185/RLCS-2022-1782>
- Fieiras-Ceide, C., Vaz-Álvarez, M., & Túniz-López, M. (2022). Artificial intelligence strategies in European public broadcasters: Uses, forecasts and future challenges. *Profesional de la Información*, 31(5), e310518. <https://doi.org/10.3145/epi.2022.sep.18>
- Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V.,... & Vayena, E. (2018). AI4People—An ethical framework for a good AI society. *Minds and Machines*, 28(4), 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
- Google AI. (2025, marzo 12). Introducing Gemma 3 [Entrada de blog]. Google Developers Blog. <https://developers.googleblog.com/en/introducing-gemma3/>
- González-Aguirre, A., Sierra, A., Carrasco, C., & Rodríguez, H. (2025). *Salamandra: A suite of open-source decoder-only large language models* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2502.08489>
- Guo, D., Zhang, L., Wang, J., Liu, H., & DeepSeek AI Team. (2025). *DeepSeek-R1: Incentivizing reasoning capability in large language models via reinforcement learning* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2501.12948>
- IBM Research. (2025). *Granite 3.3 8B Instruct – Model Card*. Hugging Face. <https://huggingface.co/ibm-granite/granite-3.3-8b-instruct>
- ISO 9241-210:2010. (2010). *Ergonomics of human–system interaction — Part 210: Human-centred design for interactive systems*. International Organization for Standardization.
- Jiang, A. Q., Sablayrolles, A., Mensch, A., Bamford, C., Casas, D. de las, Chaplot, D. S., Lengyel, G., Lample, G., & Scialom, T. (2023, octubre 10). *Mistral 7B* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2310.06825>

- Kemmis, S., McTaggart, R., & Nixon, R. (2014). *The action research planner: Doing critical participatory action research*. Springer.
- Lin, Y. (2015). Open data and co-production of public value of BBC Backstage. *International Journal of Digital Television*, 6(2), 145–162. https://doi.org/10.1386/jdtv.6.2.145_1
- Maslej, N., Fattorini, L., Perrault, R., Gil, Y., Parli, V., & Kariuki, N. (2025). *Artificial Intelligence Index Report 2025*. Stanford University, Institute for Human-Centered Artificial Intelligence. <https://aiindex.stanford.edu/report>
- Meta AI. (2024, noviembre 29). *Meta LLaMA 3.1: A family of state-of-the-art open-weight language models* [Entrada de blog]. Meta AI Blog. <https://ai.meta.com/blog/meta-llama-3-1>
- Nielsen. (2024, abril 14). Cineverse se asocia con Gracenote para ampliar enormemente las capacidades de búsqueda de cineSearch. *Nielsen News Center*. <https://www.nielsen.com/es/news-center/2024/cineverse-teams-up-with-graceno-te-to-vastly-expand-search-capabilities-for-cinerearch>
- Nikitha, N. (2021). Conversational AI: Indispensable way to attract and retain customers. *ISBR Management Journal*, 6(2), 15–24. <https://doi.org/10.52184/isbrmj.v6io2.112>
- Norman, D. A., & Draper, S. W. (Eds.). (1986). *User centered system design: New perspectives on human-computer interaction*. Lawrence Erlbaum Associates.
- Pérez-Seijo, S., Melle, M., & Paniagua, F. J. (2018). Innovación en radiotelevisión pública europea: narrativas inmersivas y organización de los contenidos 360 grados en plataformas digitales. *Revista Latina de Comunicación Social*, 73, 1151–1136. <https://doi.org/10.4185/RLCS-2018-1299>
- Qwen Team (Alibaba Cloud). (2025). *Qwen 2.5-72B Instruct – Model Card*. Hugging Face. <https://huggingface.co/Qwen/Qwen2.5-72B-Instruct>
- Ranaivoson, H., Farchy, J., & Gansemer, M. (2013). Differentiated strategies for digital innovation on television: Traditional channels vs. new entrants. *Observatorio (OBS) Journal*, 7(4), 23–44. <http://obs.obercom.pt/index.php/obs/article/view/659>

- Rodríguez-Fernández, M. M., Sánchez-Amboage, E., & Toural-Bran, C. (2018). Las radiotelevisiónes públicas europeas en el entorno web. *Revista Latina de Comunicación Social*, 73, 911–926.
<https://doi.org/10.4185/RLCS-2018-1288>
- Tambini, D. (2015). *Five theses on public media and digitization: From a 56-country study*. London School of Economics.
- van Es, K., & Poell, T. (2020). Platform imaginaries and Dutch public service media. *Social Media + Society*, 6(2), 2056305120933289.
<https://doi.org/10.1177/2056305120933289>

11. Anexos técnicos:

En este apartado se reúnen los materiales de apoyo que permiten entender y consultar el sistema en profundidad. Incluye los prompts base de las funcionalidades clave y se facilitan el accesos a la plataforma y al código fuente, junto al correo de acceso al formulario para solicitar credenciales de la demo.

11.1 Prompts de ajuste

11.1.1 Prompts Matriz de Valor Público

Tabla 10. Prompt Matriz Valor Público

Categoría	Subcategoría	Descripción
IGUALDAD	Justicia social	Adaptar tono, vocabulario y respuestas al compromiso con los derechos humanos, la igualdad de género y la protección frente a la violencia. Estilo empático e incluso sin fórmulas forzadas; lenguaje no estigmatizante y referencias éticas claras.
IGUALDAD	Universalidad	Usar lenguaje claro y directo, libre de tecnicismos innecesarios. Adaptar expresiones a distintos niveles de comprensión y necesidades cognitivas o lingüísticas. Garantizar consistencia y accesibilidad.
IGUALDAD	Alfabetización mediática	Explicar conceptos complejos con claridad, evitando ambigüedad. Promover pensamiento crítico y autonomía interpretativa con referencias educativas o preguntas abiertas.
DIVERSIDAD	Pluralismo	Lenguaje neutral y respetuoso hacia todas las posiciones ideológicas y culturales. Incluir ejemplos diversos y múltiples perspectivas.
DIVERSIDAD	Lingüística territorial	Adaptar el lenguaje a la diversidad territorial y valorar la pluralidad lingüística como parte integral del contenido.
DIVERSIDAD	Sociocultural	Ajustar el registro según los perfiles sin caer en estereotipos. Representación justa de género, edad, origen y capacidades.
INNOVACIÓN	Transformación tecnológica	Usar referencias actualizadas a entornos digitales y tecnologías emergentes. Evitar tecnofilia superficial y destacar el valor añadido de la innovación.
INNOVACIÓN	Creatividad aplicada	Incluir recursos creativos (metáforas, narrativas dinámicas, referencias visuales) que promuevan experiencias activas y exploratorias.

INNOVACIÓN	Organizacional y modelo de negocio	Mencionar principios de sostenibilidad y colaboración institucional. Reforzar legitimidad comunicativa mediante transparencia y coherencia.
PARTICIPACIÓN	Creación ciudadana	Invitar a la colaboración y reconocer aportes del público como conocimiento legítimo. Estilo empático y de co-diseño.
PARTICIPACIÓN	Escucha	Mostrar cómo las decisiones se construyen con participación ciudadana. Lenguaje de apertura y corresponsabilidad.
PARTICIPACIÓN	Desarrollo comunitario	Referir alianzas institucionales o educativas. Mostrar validación colectiva y lenguaje inclusivo sobre colaboraciones.
COHESIÓN SOCIAL	Territorial	Incluir referencias que conecten realidades regionales y promuevan empatía entre territorios.
COHESIÓN SOCIAL	Compromiso social	Priorizar temas de impacto social. Tono proactivo y sensible a las necesidades colectivas.
COHESIÓN SOCIAL	Responsabilidad y sostenibilidad	Fomentar prácticas sostenibles con datos claros y tono no moralizante. Reflejar compromiso institucional ético.
CALIDAD	Excelencia profesional	Lenguaje preciso y técnico con accesibilidad. Contenidos estructurados y basados en evidencia.
CALIDAD	Integridad e independencia	Garantizar imparcialidad, transparencia y reconocimiento de límites del conocimiento.
CALIDAD	Contenido con valor añadido	Añadir contexto, ejemplos culturales y conexiones relevantes que generen aprendizaje.
TRANSPARENCIA	Transparencia	Responder con tono claro e institucional. Citar fuentes, reconocer límites y aportar contexto verificable.

Fuente: Elaboración propia

11.1.2 Prompts Rasgos de personalidad

Tabla 11. Prompts Rasgos de personalidad

Rasgo	Nivel	Descripción
Amabilidad	1	Competitivo, desconfiado, egocéntrico y hostil.
Amabilidad	2	Competitivo y desconfiado pero tolerante.
Amabilidad	3	Competitivo pero tolerante y empático.
Amabilidad	4	Algo desconfiado pero modesto, tolerante y empático.
Amabilidad	5	Empático, modesto, tolerante, altruista y fiable.
Responsabilidad	1	Inmaduro, impulsivo, desinteresado y negligente.
Responsabilidad	2	Inmaduro e impulsivo pero coherente.
Responsabilidad	3	Algo inmaduro e impulsivo pero honesto y consistente.
Responsabilidad	4	Bastante maduro y fiable.
Responsabilidad	5	Muy maduro, fiable, comprometido, honesto y coherente.
Extroversión	1	Solitario, prudente, selectivo e independiente.
Extroversión	2	Solitario y selectivo pero asertivo.
Extroversión	3	Prudente pero optimista, selectivo pero sociable.
Extroversión	4	Optimista, asertivo y sociable.
Extroversión	5	Muy sociable, hablador y asertivo.
Inestabilidad emocional	1	Estable, predecible, alegre, nada conflictivo.
Inestabilidad emocional	2	Estable, alegre, a veces irritable.
Inestabilidad emocional	3	Variable: triste o alegre, pero nervioso.
Inestabilidad emocional	4	Siempre nervioso, irritable y vulnerable.
Inestabilidad emocional	5	Nervioso, conflictivo, imprevisible y vulnerable.

Fuente: Elaboración propia

11.1.3 Prompts Estilos de comunicación

Tabla 12. Prompt Estilos de comunicación

Estilo	Descripción
Conversational Survey	Integrar preguntas estratégicas en la conversación para descubrir valores y percepciones del público sobre la plataforma y su impacto social.
Gamified Information	Usar actividades lúdicas (quiz, acertijos, encuestas) para explicar fenómenos de innovación y resaltar el valor público.
Question Solving	Estilo claro y orientado a soluciones; responder de forma concisa y razonada, con ejemplos prácticos.
Provide Samples and Situations	Acompañar las respuestas con ejemplos y situaciones reales que faciliten la comprensión y adapten el contexto.
Neutro	Estilo base del modelo sin inferencias ni adaptaciones.

Fuente: Elaboración propia

11.2 Acceso y solicitud de la demo Prompt

La demostración está disponible desde la plataforma de VALUEbot, integrada en <https://psmresearchteam.com>. Para recibir una clave temporal, basta con enviar una solicitud al correo de contacto del proyecto indicando nombre, entidad y objetivo de uso. La documentación se completa la estructura del centro de ayuda y una serie de videotutoriales breves. Su propósito consiste en acompañar la puesta en marcha y servir de referencia rápida cuando aparezcan dudas. Todo queda en un mismo lugar para que el aprendizaje no dependa de correos sueltos ni de instrucciones dispersas.

12. Glosario de Inteligencia Artificial

Este glosario nace con una intención práctica: ayudar a comprender cómo funciona VALUEbot y, a la vez, servir como pieza de alfabetización digital. No es un diccionario enciclopédico. Es una guía de uso que explica lo necesario para tomar decisiones informadas, sin pedir al

lector que sea especialista. Cada entrada incluye una definición breve, un “para qué sirve” en el contexto de VALUEbot, un ejemplo en una línea y, cuando aplica, una nota sobre riesgos y buenas prácticas. El tono es claro; si un término técnico se puede decir en llano, se dice en llano.

El primer bloque reúne los **conceptos fundamentales de IA**. Aquí se aclaran términos como modelo, entrenamiento, inferencia, corpus, contexto y ventana de contexto. También se diferencian familias de modelos y lo que implica trabajar con pesos abiertos o con servicios en la nube. La mirada es funcional: entender qué hace cada cosa y qué consecuencias tiene para la privacidad, el rendimiento o la soberanía tecnológica. La idea es que, después de leerlo, cualquiera pueda explicar por qué en VALUEbot a veces conviene un motor local y otras veces uno conectado.

El segundo bloque aborda **los procesos lingüísticos, semánticos y de interacción**. Se explica qué es un prompt y cómo se construye, qué significa inyección de documentos, cómo operan los resúmenes “mapa y reducción” y por qué la memoria de contexto no es una memoria humana. Este bloque también describe la conversación como flujo: turnos, tono, longitud de respuesta, cambio de idioma. Cada término se ilustra con microejemplos anclados en casos reales, como una consulta sobre programación de un canal o una recomendación educativa. El objetivo es que el lector pueda ajustar la voz del avatar sin miedo a “romper” el sistema.

El tercer bloque se centra en **ética, gobernanza y sesgos**. Se traduce el marco de valor público a reglas operativas y se muestran sus efectos en las respuestas: igualdad, diversidad, innovación, participación, cohesión social y calidad. Se explican nociones de sesgo de datos y de modelo, transparencia, trazabilidad y consentimiento. No se queda en principios. Cada entrada incluye señales prácticas para detectar problemas y pautas de corrección, por ejemplo cómo pedir justificaciones breves cuando la respuesta es sensible o cuándo avisar que el sistema no dispone de una fuente fiable. La alfabetización aquí significa saber preguntar mejor y saber cuándo detenerse.

El cuarto bloque describe la **arquitectura y los componentes técnicos**, pero con foco en la toma de decisiones. Se explica el papel

del núcleo en Flask, la orquestación por eventos, la base de datos y los módulos que el usuario ve: proyectos, avatares, chats, analítica y valores. Se aclara qué se registra y por qué, cómo se protege una clave de API, qué implica el escalado horizontal y qué métricas conviene mirar para anticipar cuellos de botella. La técnica aparece como medio, no como fin: lo importante es entender la relación entre diseño, estabilidad y gobernanza.

Para que el glosario sea útil en el día a día, cada término enlaza con puntos específicos del informe. El glosario se publica como documento vivo. Se actualiza con cada versión de VALUEbot y recoge dudas frecuentes de talleres y pilotos. Una etiqueta de versión y un breve historial explican qué se añadió o se corrigió. Así, el lector no solo aprende vocabulario; entiende cómo cambia el sistema y por qué. Esa es la promesa: un instrumento que acompaña, que no abruma, y que devuelve al usuario el control sobre lo que está leyendo y sobre lo que está configurando.

12.1 Conceptos fundamentales de IA

1. Inteligencia Artificial (IA) Campo interdisciplinar que combina informática, estadística, lingüística y filosofía para desarrollar sistemas capaces de ejecutar tareas que requieren inteligencia humana, como el razonamiento, la percepción o la generación de lenguaje natural. En el contexto del proyecto VALUEbot, la IA se emplea como un vector de innovación pública que mejora la interacción con los usuarios y la gestión de conocimiento institucional. *Referencia en el Informe: aparece en el apartado 1.1 “Justificación del proyecto”, donde se presenta la IA como elemento estratégico para el desarrollo de herramientas de comunicación y servicio público.*

2. Aprendizaje Automático (Machine Learning) Subcampo de la IA que permite que los sistemas aprendan patrones a partir de datos y mejoren su rendimiento sin intervención humana directa. VALUEbot integra algoritmos basados en aprendizaje supervisado y modelos preentrenados para interpretar información institucional y ofrecer respuestas adaptadas. *Referencia en el Informe: desarrollado en el apartado 5.1 “Retrabajando con modelos preentrenados”, donde se*

describe cómo los modelos de lenguaje son ajustados para tareas específicas de información pública.

3. Aprendizaje Profundo (Deep Learning) Extensión del aprendizaje automático que utiliza redes neuronales profundas para analizar grandes volúmenes de datos y reconocer estructuras complejas, como texto o imágenes. En VALUEbot se aplica para la comprensión semántica y el procesamiento de lenguaje natural en distintos contextos comunicativos. *Referencia en el Informe: vinculado al apartado 5.3 “Modelos de IA empleados”, donde se detallan los tipos de redes neuronales aplicadas al sistema.*

4. Red Neuronal Artificial Modelo computacional inspirado en la estructura del cerebro humano. Está compuesto por nodos interconectados (“neuronas artificiales”) que procesan información mediante pesos ajustables. En el proyecto VALUEbot, las redes neuronales se usan para generar inferencias textuales basadas en datos contextuales. *Referencia en el Informe: aparece en el apartado 5.3 “Modelos de IA empleados”, donde se explica el uso de redes neuronales profundas para tareas lingüísticas.*

5. Pesos (Weights) Parámetros internos de una red neuronal que determinan la importancia de cada conexión entre nodos. Durante el entrenamiento se modifican para minimizar el error de predicción. *Referencia en el Informe: asociados al apartado 5.1, donde se describe el ajuste fino (fine-tuning) de modelos preentrenados mediante reponderación de parámetros internos.*

6. Gradiente Magnitud que indica el cambio de la función de pérdida respecto a los pesos de la red. Su cálculo permite optimizar los modelos durante el entrenamiento. *Referencia en el Informe: relacionado con la descripción metodológica de optimización de modelos en el apartado 5.1 “Retrabajando con modelos preentrenados”.*

7. Función de pérdida (Loss Function) Ecuación que mide la diferencia entre la salida generada por el modelo y la respuesta esperada. Su objetivo es guiar el proceso de aprendizaje minimizando el error acumulado. *Referencia en el Informe: mencionada indirectamente en*

el apartado 5.1 al explicar los procesos de ajuste y evaluación de rendimiento de modelos.

8. Entrenamiento Fase de aprendizaje en la que un modelo ajusta sus pesos para reconocer patrones en los datos. En VALUEbot, el entrenamiento se realizó a partir de modelos preexistentes optimizados mediante la inyección de información institucional. *Referencia en el Informe: abordado en los apartados 5.1 y 5.3, donde se explica cómo los modelos fueron adaptados a tareas específicas sin requerir reentrenamiento completo.*

9. Inferencia Etapa en la que un modelo entrenado genera resultados a partir de nuevas entradas. En VALUEbot, la inferencia ocurre cuando el sistema interpreta una consulta y produce una respuesta contextualizada. *Referencia en el Informe: descrita en el apartado 4.2 “Diseño modular”, donde se detalla el proceso de inferencia distribuida en la arquitectura del sistema.*

10. Modelo preentrenado Modelo de IA que ya ha aprendido representaciones lingüísticas o visuales a partir de grandes corpus de datos. VALUEbot parte de modelos preentrenados (como LLaMA o GPT) que luego se personalizan mediante ingeniería de prompts y documentos propios. *Referencia en el Informe: explicado en el apartado 5.1 “Retrabajando con modelos preentrenados”, con ejemplos de modelos utilizados en el sistema.*

11. Fine-tuning (Ajuste fino) Proceso de reentrenar parcialmente un modelo preentrenado con datos específicos para adaptarlo a un nuevo dominio. En VALUEbot, se aplica para afinar respuestas institucionales y evitar sesgos. *Referencia en el Informe: apartado 5.1, donde se detalla la metodología de ajuste fino y las limitaciones derivadas del sobreentrenamiento.*

12. Overfitting (Sobreajuste) Fenómeno por el cual un modelo aprende demasiado los datos de entrenamiento, perdiendo capacidad de generalizar. En el contexto de VALUEbot, se evita mediante validaciones cruzadas y control de datos de entrenamiento. *Referencia en el Informe: aparece implícitamente en el apartado 3.2 “Fases de*

desarrollo”, donde se explica la necesidad de validación iterativa de los modelos.

12.2 Procesos lingüísticos, semánticos e interacción

13. Modelo de lenguaje (LLM, Large Language Model) Sistema de inteligencia artificial basado en redes neuronales profundas que aprende las estructuras del lenguaje humano a partir de grandes corpus textuales. Puede generar texto, traducir, resumir o mantener conversaciones coherentes. En VALUEbot, los modelos de lenguaje permiten realizar inferencias adaptadas al contexto del usuario y del documento institucional. *Referencia en el Informe: aparece en el apartado 5.1 “Modelos de IA empleados” y en el 5.3 “Selección de modelos de IA empleados”, donde se describen los modelos preentrenados integrados en la plataforma.*

14. Token Unidad mínima de texto que un modelo puede procesar (una palabra, sílaba o símbolo). En los sistemas de IA, los tokens determinan la longitud y el coste computacional de las respuestas. *Referencia en el Informe: mencionado indirectamente en el apartado 4.2 “Diseño modular”, donde se aborda la gestión de procesamiento textual y la segmentación de cadenas de texto.*

15. Tokenización Proceso de dividir el texto en tokens para que el modelo pueda analizarlos. En VALUEbot, la tokenización es fundamental para mantener la coherencia semántica en la generación de respuestas a partir de documentos extensos. *Referencia en el Informe: relacionada con el apartado 5.2 “Inyección de documentos”, donde se detalla la segmentación de textos mediante técnicas de chunking y embeddings.*

16. Embedding Representación numérica o vectorial de una palabra o frase que captura su significado contextual. En VALUEbot, los embeddings permiten comparar similitudes semánticas entre preguntas y fragmentos de documentos institucionales. *Referencia en el Informe: aparece en el apartado 5.2 “Inyección de documentos”, al explicar la búsqueda semántica de información relevante mediante embeddings vectoriales.*

17. Inyección en contexto (In-context injection) Técnica que introduce fragmentos de información directamente en el prompt de un modelo sin modificar su entrenamiento. En VALUEbot, se utiliza para incorporar partes de documentos institucionales en las respuestas, garantizando precisión y trazabilidad. *Referencia en el Informe: descrita en el apartado 5.2 “Inyección de documentos”, junto con las técnicas de chunking y map-reduce summarization utilizadas en la inferencia documental.*

18. Chunking División de textos extensos en unidades más pequeñas y coherentes (chunks) que facilitan su procesamiento por modelos de lenguaje. VALUEbot aplica chunking para analizar documentos largos y generar resúmenes consistentes. *Referencia en el Informe: desarrollada en el apartado 5.2 “Inyección de documentos”, donde se detalla el proceso de segmentación semántica y su integración con la búsqueda contextual.*

19. Map-Reduce Summarization Estrategia de resumen en dos etapas: primero se sintetizan las partes del texto (*map*), y luego se combinan los resultados en un resumen global (*reduce*). VALUEbot utiliza esta técnica para elaborar síntesis documentales precisas. *Referencia en el Informe: mencionada en el apartado 5.2, dentro del proceso de inferencia distribuida aplicada a la documentación institucional.*

20. Ingeniería de Prompts (Prompt Engineering) Disciplina que diseña y optimiza las instrucciones dadas a un modelo de lenguaje para guiar su comportamiento. En VALUEbot, la ingeniería de prompts se emplea para asegurar coherencia ética, modularidad comunicativa y alineación con los valores públicos. *Referencia en el Informe: desarrollada en el apartado 5.1 “Retrabajando con modelos preentrenados”, donde se explica cómo los prompts estructurados definen el tono, los valores y el estilo de comunicación del sistema.*

21. In-context Learning (Aprendizaje en contexto) Capacidad de los modelos de lenguaje para aprender de ejemplos incluidos directamente en el prompt, sin reentrenamiento. En VALUEbot, esta técnica se usa para mejorar la adaptación del sistema a nuevos tipos de documentos y usuarios. *Referencia en el Informe: abordada en el*

apartado 5.1, donde se describe el proceso de ajuste contextual de las respuestas del sistema mediante ejemplos directos.

22. RAG (Retrieval-Augmented Generation) Arquitectura que combina recuperación de información (retrieval) con generación de texto (generation). En VALUEbot, el sistema consulta una base documental mediante embeddings y luego redacta respuestas fundamentadas en los fragmentos más relevantes. *Referencia en el Informe: aparece en los apartados 5.1 y 5.2, donde se detalla el proceso de recuperación y síntesis de conocimiento contextual.*

12.3 Ética, gobernanza y sesgos

23. Sesgo algorítmico Distorsión sistemática en los resultados de un modelo de IA causada por datos de entrenamiento incompletos, desbalanceados o mal representativos. Puede generar decisiones o respuestas que perpetúan desigualdades o prejuicios. En el contexto de VALUEbot, se aborda mediante curaduría de corpus y control semántico en la generación textual. *Referencia en el Informe: tratado en el apartado 5.1 y 5.4 donde se detallan las estrategias de filtrado y ajuste de los modelos para mitigar sesgos culturales y lingüísticos.*

24. Mitigación de sesgos Conjunto de procedimientos que buscan identificar, evaluar y reducir los sesgos en los datos, los modelos o las salidas de un sistema de IA. Incluye balanceo de corpus, revisión ética y ajustes de prompts. *Referencia en el Informe: aparece en el apartado 5.4, donde se explica la revisión manual de corpus, los ajustes de prompt y la validación iterativa de las respuestas del modelo.*

25. Privacidad Principio ético y legal que reconoce el derecho de las personas a controlar sus datos personales y su uso por parte de sistemas tecnológicos. En VALUEbot, la privacidad se protege mediante almacenamiento local, anonimización y ejecución de modelos en entornos controlados. *Referencia en el Informe: desarrollada en los apartados 4.1 “Arquitectura general” y 7.2 “Evaluación cualitativa”, donde se subraya la importancia del procesamiento local y la ausencia de transmisión de datos sensibles a la nube.*

26. Transparencia Valor que exige que los procesos de decisión de los sistemas de IA sean comprensibles y verificables. En VALUEbot, se implementa mediante trazabilidad de decisiones, documentación de prompts y registro de versiones del sistema. *Referencia en el Informe: mencionada en el apartado 7.2 “Evaluación cualitativa”, donde se describen los mecanismos de registro y auditoría de interacciones que garantizan la transparencia institucional.*

27. Valor Público Concepto que define el conjunto de beneficios sociales, culturales y democráticos generados por las acciones institucionales más allá del valor económico. En VALUEbot, actúa como eje ético y comunicativo, asegurando que la IA contribuya a la misión del servicio público. *Referencia en el Informe: aparece en los apartados 1.1 “Justificación del proyecto” y 5.4 “Elaborando una matriz de valor público”, donde se detalla su traducción operativa en la matriz de valor público.*

28. LOPD (Ley Orgánica de Protección de Datos) Norma española que regula el tratamiento de datos personales, garantizando la privacidad y seguridad de la información. En el entorno de VALUEbot, se aplica en el manejo de bases de datos institucionales y en la anonimización de las consultas. *Referencia en el Informe: citada en el apartado 4.1 “Arquitectura general”, donde se explica el cumplimiento normativo de los sistemas locales en relación con la protección de datos.*

29. GDPR (General Data Protection Regulation) Reglamento europeo que protege los datos personales de los ciudadanos, estableciendo obligaciones para las instituciones que los tratan. VALUEbot cumple el GDPR mediante trazabilidad, consentimiento informado y procesamiento local. *Referencia en el Informe: aparece en el apartado 4.1, junto con la LOPD, en el contexto de las medidas de cumplimiento y soberanía tecnológica.*

30. Soberanía tecnológica Capacidad de una institución o país para controlar su infraestructura digital, datos y modelos de IA sin depender de servicios externos. En el proyecto VALUEbot, este principio se materializa en la preferencia por modelos locales y entornos de

despliegue propios. *Referencia en el Informe: desarrollada en el apartado 4.1 “Arquitectura general”, donde se justifica el uso de motores locales (Ollama) frente a servicios en la nube.*

31. Trazabilidad Capacidad de registrar, auditar y reconstruir los pasos de un proceso de IA, desde la entrada hasta la salida del modelo. VALUEbot implementa trazabilidad mediante registros de prompts, respuestas y versiones de modelos. *Referencia en el Informe: descrita en el apartado 7.2 “Evaluación cualitativa”, como uno de los mecanismos de control de calidad y transparencia institucional.*

32. Explicabilidad (Explainability) Grado en que los mecanismos internos de un sistema de IA pueden ser comprendidos por los humanos. En VALUEbot, se persigue mediante documentación técnica, visualización de flujo de datos y explicación de inferencias. *Referencia en el Informe: vinculada al apartado 7.2, donde se detalla la implementación de herramientas para explicar las decisiones generadas por el modelo a usuarios y auditores.*

33. IA Responsable Modelo de desarrollo tecnológico que combina innovación con principios éticos como equidad, sostenibilidad, privacidad y rendición de cuentas. VALUEbot adopta este enfoque para garantizar legitimidad pública y confianza institucional. *Referencia en el Informe: mencionada en los apartados 1.1 donde se presenta la responsabilidad como principio transversal del sistema.*

34. PSM (Public Service Media) Medios de comunicación públicos cuyo propósito es garantizar la información plural, la diversidad y la calidad democrática. En el marco de VALUEbot, la IA se aplica para reforzar el compromiso del PSM con la ciudadanía. *Referencia en el Informe: abordada en el apartado 1.3 “Respuesta de los medios públicos europeos a la transformación del contexto mediático”.*

35. Matriz de Valor Público Instrumento que operacionaliza los valores públicos en dimensiones evaluables (igualdad, diversidad, innovación, participación, cohesión social y calidad). En VALUEbot, la matriz orienta la generación de contenido y el comportamiento conversacional. *Referencia en el Informe: aparece en el apartado 5.4*

“Elaborando una matriz de valor público” y en el anexo 11.1.1, donde se documenta su estructura detallada.

36. Deepfake Contenido audiovisual manipulado o generado mediante redes neuronales profundas para alterar rostros, voces o escenas. En comunicación pública, su uso plantea riesgos éticos y de confianza informativa.

37. Medios sintéticos (Synthetic Media) Contenidos (texto, imagen, audio o vídeo) producidos total o parcialmente por sistemas de IA. En VALUEbot, se exploran como recurso narrativo en entornos de servicio público.

12.4 Arquitectura y componentes técnicos

38. IA local (Inteligencia Artificial local) Sistema de inteligencia artificial que se ejecuta directamente en un servidor o equipo propio sin depender de una conexión a internet o de proveedores externos. En VALUEbot, la IA local garantiza privacidad, control sobre los datos y autonomía operativa. *Referencia en el Informe: aparece en el apartado 4.1 “Arquitectura general”, donde se describe el uso de motores locales (Ollama) para preservar la soberanía tecnológica y reducir dependencia de la nube.*

39. IA en la nube (Cloud AI) Conjunto de servicios de inteligencia artificial alojados en servidores remotos y accesibles por internet. Aunque permiten gran capacidad de cálculo, implican transferencia de datos. En VALUEbot, se utiliza de forma complementaria a la IA local para tareas específicas de alto rendimiento. *Referencia en el Informe: mencionada en el apartado 4.1, donde se justifica la arquitectura híbrida entre entornos locales y servicios externos (Gemini, OpenAI).*

40. Infraestructura local Conjunto de equipos, redes y servidores propios que permiten el funcionamiento autónomo de sistemas de IA. En VALUEbot, esta infraestructura sostiene la ejecución de modelos locales, la base de datos y el control de privacidad. *Referencia en el Informe: desarrollada en el apartado 4.1 “Arquitectura general”, junto con la descripción de la instalación de motores locales y servicios*

Flask.

41. Servidor Equipo informático destinado a almacenar, procesar y distribuir información a otros sistemas conectados. En VALUEbot, los servidores locales alojan los modelos y la base de datos, asegurando la disponibilidad y el control de la información. *Referencia en el Informe: aparece en el apartado 4.1, dentro de la descripción del despliegue de servicios locales y su integración con la infraestructura modular.*

42. Nube (Cloud) Infraestructura de servidores remotos que ofrece almacenamiento y potencia de cálculo accesible por internet. En VALUEbot, la nube se emplea únicamente para tareas de testeo y comparación, priorizando la ejecución local. *Referencia en el Informe: descrita en el apartado 4.1, donde se establece la política de equilibrio entre privacidad y eficiencia computacional.*

43. Arquitectura híbrida Modelo que combina ejecución local y servicios en la nube. Esta estructura permite aprovechar la potencia del procesamiento remoto sin comprometer la seguridad de los datos. *Referencia en el Informe: abordada en el apartado 4.1 “Arquitectura general”, donde se documenta la estructura combinada del sistema y sus beneficios en términos de rendimiento y soberanía tecnológica.*

44. Core Núcleo funcional de un sistema que gestiona las operaciones principales. En VALUEbot, el core gestiona la comunicación entre la interfaz de usuario, el motor de inferencia y los módulos de análisis. *Referencia en el Informe: explicado en el apartado 4.2 “Diseño modular”, donde se describe la lógica central del sistema y su relación con los servicios distribuidos.*

45. Core Service Componente principal del sistema encargado de ejecutar los procesos esenciales (carga de modelos, inferencia, registro de interacciones). *Referencia en el Informe: aparece en el apartado 4.2, donde se explica la función del servicio principal dentro del ecosistema modular Flask.*

46. Entorno Sandbox Espacio de pruebas controlado y aislado en el que se experimenta con modelos o configuraciones sin afectar al

sistema principal. En VALUEbot, el sandbox permite testear avatares, prompts y modelos antes de su implementación. Referencia en el Informe: desarrollado en los apartados 6.4 “Testeo CHAT Sandbox” y 6.4 “Pruebas y test de los chatbots (Sandbox)”, donde se detallan los entornos de experimentación.

47. **Flask** Framework de desarrollo web en Python que permite crear aplicaciones modulares. En VALUEbot, se emplea para estructurar la interfaz entre el motor de inferencia y los módulos de análisis y control. Referencia en el Informe: aparece en el apartado 4.2 “Diseño modular”, donde se describe su papel como estructura base del sistema y su integración con otros componentes como MySQL y RabbitMQ.

48. **Flask Blueprint** Módulo dentro del framework Flask que facilita la organización de funciones en componentes independientes. En VALUEbot, permite estructurar el sistema en rutas separadas para administración, inferencia y analítica. Referencia en el Informe: mencionada en el apartado 4.2, al describir la arquitectura modular del sistema y la concurrencia entre procesos.

49. **MySQL** Sistema de gestión de bases de datos relacional de código abierto, utilizado para almacenar logs, interacciones y configuraciones del sistema. Referencia en el Informe: aparece en el apartado 4.2 “Diseño modular”, donde se documenta el uso de MySQL como base de datos principal del sistema.

50. **RabbitMQ** Sistema de mensajería asíncrona que permite la comunicación eficiente entre diferentes módulos del sistema. En VALUEbot, se utiliza para gestionar colas de tareas y garantizar la estabilidad de los procesos concurrentes. Referencia en el Informe: desarrollado en el apartado 4.2 “Diseño modular”, dentro del esquema de concurrencia y comunicación entre módulos Flask.

51. **Ollama** Motor de inferencia local diseñado para ejecutar modelos de lenguaje (LLM) en dispositivos o servidores propios. En VALUEbot, Ollama es el componente clave de la IA local. Referencia en el Informe: aparece en el apartado 4.1 “Arquitectura general”, donde se detalla su integración con la estructura híbrida del sistema.

52. APIs (Interfaz de Programación de Aplicaciones) Conjunto de protocolos que permiten la comunicación entre distintas aplicaciones o módulos. En VALUEbot, las APIs conectan los motores de IA con la interfaz de usuario y los sistemas de análisis. *Referencia en el Informe: mencionadas en los apartados 4.2 “Diseño modular” y 5.7 “Orientación de uso de los diferentes modelos”.*

53. Ultramsgr Ultramsgr es un proveedor de servicios de API de WhatsApp que permite a las empresas y desarrolladores integrar la funcionalidad de mensajería de WhatsApp en sus propios sistemas, como software CRM, ERP o sitios web. No es un producto oficial de WhatsApp, sino una herramienta de terceros que utiliza la API de WhatsApp Business.

54. Concurrencia Capacidad de un sistema para ejecutar múltiples procesos simultáneamente, optimizando el uso de recursos. En VALUEbot, la concurrencia permite gestionar peticiones simultáneas de usuarios sin afectar el rendimiento. *Referencia en el Informe: explicada en el apartado 4.2 “Diseño modular”, donde se aborda la gestión de procesos paralelos y comunicación mediante RabbitMQ*



VALUEbot

Creación de una plataforma generadora de chatbots mediante IA para la Comunicación del valor público del PSM

PDC2023-145885-I00



MINISTERIO DE CIENCIA E INNOVACIÓN



Financiado por la Unión Europea
NextGenerationEU



Plan de Recuperación, Transformación y Resiliencia



AGENCIA ESTATAL DE INVESTIGACIÓN