

HUMAN LAYER

White Paper

Arquitectura para la Coexistencia Humano-IA

Autor: Edgar Prieto Sarrat

Barcelona, España

2026

1. Introducción

La inteligencia artificial se está convirtiendo rápidamente en un componente central de los sistemas tecnológicos modernos.

En múltiples sectores, los sistemas de inteligencia artificial participan cada vez más en procesos de decisión que tradicionalmente estaban gestionados exclusivamente por seres humanos. Estos procesos incluyen ámbitos como los mercados financieros, los sistemas de salud, las infraestructuras de transporte, la gestión corporativa y la gobernanza pública.

A medida que las capacidades de la inteligencia artificial continúan ampliándose, también aumenta el nivel de influencia que estos sistemas ejercen sobre entornos complejos de toma de decisiones.

Gran parte del debate público y académico sobre inteligencia artificial se ha centrado en cuestiones como:

- la precisión de los modelos algorítmicos
- la seguridad de los sistemas automatizados
- la alineación ética de las tecnologías emergentes.

Estas cuestiones son fundamentales. Sin embargo, existe otra dimensión del problema que ha recibido relativamente menos atención: la forma en que los sistemas de inteligencia artificial influyen en la estructura de los entornos de decisión humanos.

A medida que los sistemas inteligentes se integran más profundamente en los procesos operativos, existe el riesgo de que la capacidad humana para comprender, evaluar y tomar decisiones en esos mismos entornos se vea gradualmente alterada.

Este fenómeno no es necesariamente el resultado de fallos tecnológicos.

Más bien surge como una consecuencia natural de la interacción entre sistemas altamente automatizados y procesos cognitivos humanos.

Cuando los sistemas automatizados proporcionan recomendaciones cada vez más sofisticadas, los participantes humanos pueden tender, con el tiempo, a confiar en estas recomendaciones de forma creciente, reduciendo gradualmente su propio nivel de implicación analítica.

Este proceso puede conducir a una situación en la que los humanos siguen siendo formalmente responsables de las decisiones, pero su participación cognitiva real se vuelve progresivamente más limitada.

Este documento introduce el concepto de Human Layer como un marco conceptual diseñado para abordar este desafío emergente.

Human Layer propone la introducción de una capa estructural dentro de los sistemas socio-técnicos cuyo objetivo es preservar la capacidad humana de decisión en entornos cada vez más automatizados.

En lugar de centrarse únicamente en la regulación del comportamiento de los sistemas de inteligencia artificial, el marco de Human Layer examina cómo deben diseñarse los entornos de interacción humano-máquina para garantizar que los humanos sigan siendo participantes activos, conscientes y responsables dentro de los procesos de decisión.

El objetivo no es limitar el progreso tecnológico, sino garantizar que el aumento de las capacidades tecnológicas no debilite involuntariamente las capacidades cognitivas humanas que sostienen el funcionamiento de las sociedades complejas.

A medida que los sistemas de inteligencia artificial continúan evolucionando, la arquitectura de la interacción humano-máquina puede convertirse en uno de los factores más importantes para garantizar la estabilidad, la resiliencia y la responsabilidad de los sistemas de decisión del futuro.



2. El problema estructural de los sistemas humano-IA

La integración de la inteligencia artificial en entornos de decisión está transformando la arquitectura de los sistemas socio-técnicos modernos.

En muchos ámbitos, los sistemas de IA participan ya en procesos que históricamente eran realizados completamente por humanos. Estos incluyen actividades como

decisiones de inversión financiera, apoyo al diagnóstico médico, optimización logística, evaluación de riesgos y análisis de políticas públicas.

La suposición dominante que guía esta transformación es que los sistemas tecnológicos deben optimizarse para lograr eficiencia, precisión predictiva y rendimiento.

Aunque estos objetivos son legítimos, tienden a centrarse principalmente en las capacidades de la máquina, más que en el funcionamiento a largo plazo de los agentes humanos que interactúan con ella.

Este desequilibrio revela un problema estructural en el diseño de los sistemas humano-IA contemporáneos.

La mayoría de los enfoques actuales de desarrollo y gobernanza de la inteligencia artificial se centran en tres dimensiones principales:

- la fiabilidad técnica de los sistemas de IA
- la seguridad y robustez de sus resultados
- la posibilidad de supervisión humana.

Sin embargo, la interacción entre la cognición humana y los sistemas inteligentes introduce dinámicas adicionales que rara vez se abordan explícitamente en el diseño de los sistemas.

Estas dinámicas incluyen fenómenos como:

- sesgo de automatización (automation bias), donde los humanos tienden a confiar excesivamente en recomendaciones algorítmicas
- externalización cognitiva (cognitive offloading), donde las personas transfieren progresivamente el esfuerzo de razonamiento a sistemas automatizados
- fragmentación de la atención, causada por la interacción constante con interfaces tecnológicas complejas
- complacencia en la toma de decisiones, donde los humanos mantienen la responsabilidad formal mientras se desvinculan gradualmente del proceso activo de decisión.

Estos fenómenos no son necesariamente fallos de la tecnología.

Más bien son propiedades emergentes de sistemas complejos en los que la cognición humana y la inteligencia de máquina interactúan de forma continua.

En otras palabras, incluso sistemas de inteligencia artificial que funcionan perfectamente pueden producir efectos a largo plazo sobre la manera en que los humanos piensan, evalúan el riesgo y toman decisiones.

Esta interacción introduce un fenómeno sistémico más amplio que puede describirse como la erosión gradual de la capacidad humana de decisión en entornos altamente automatizados.

Cuando los sistemas de decisión se optimizan cada vez más para el rendimiento algorítmico, los participantes humanos pueden seguir siendo formalmente responsables mientras pierden progresivamente la implicación cognitiva activa en los procesos subyacentes.

Esta dinámica tiene precedentes históricos.

En la aviación, la introducción de sistemas avanzados de piloto automático mejoró significativamente la seguridad y la eficiencia. Sin embargo, también introdujo nuevos desafíos relacionados con la implicación del piloto y la conciencia situacional.

De forma similar, en los mercados financieros, los sistemas de trading algorítmico aumentaron la eficiencia del mercado mientras amplificaban simultáneamente ciertas formas de volatilidad sistémica.

Estos ejemplos ilustran un principio crítico:

Los sistemas tecnológicos no operan de forma aislada. Transforman los patrones cognitivos y conductuales de los humanos que interactúan con ellos.

A medida que la inteligencia artificial se integra cada vez más en los procesos de decisión cotidianos, el diseño de esta interacción se convierte en un desafío central para las sociedades tecnológicas modernas.

La cuestión clave, por tanto, no es únicamente cuán inteligente puede llegar a ser un sistema, sino cómo ese sistema afecta a la capacidad humana de comprender, evaluar y decidir dentro de ese entorno.

Human Layer propone que abordar este desafío requiere introducir una capa arquitectónica explícita dedicada a regular la interacción entre la cognición humana y los sistemas inteligentes.

Esta capa está diseñada para garantizar que el aumento de las capacidades tecnológicas no debilite involuntariamente las capacidades humanas que permiten el funcionamiento de sistemas sociales complejos.

3. Limitaciones de los marcos actuales de gobernanza de la IA

En los últimos años, gobiernos, instituciones internacionales y organismos reguladores han avanzado de forma significativa en el desarrollo de marcos de gobernanza para la inteligencia artificial.

Estos marcos tienen como objetivo garantizar que los sistemas de IA se desarrollen y desplieguen de forma segura, transparente y alineada con los derechos fundamentales.

Ejemplos de estas iniciativas incluyen el enfoque regulatorio de la Unión Europea a través del AI Act, así como diversos debates políticos y regulatorios que tienen lugar en organizaciones internacionales, gobiernos nacionales y organismos industriales.

La mayoría de los modelos actuales de gobernanza se centran en tres principios principales: 1. Seguridad y mitigación de riesgos, asegurando que los sistemas de IA no produzcan resultados dañinos o imprevisibles. 2. Transparencia y explicabilidad, permitiendo que los usuarios, reguladores y personas afectadas puedan comprender cómo los sistemas generan sus resultados. 3. Supervisión humana, garantizando que los humanos mantengan la capacidad de supervisar e intervenir en el funcionamiento de los sistemas automatizados.

Estos principios representan un paso importante y necesario en la gobernanza responsable de las tecnologías emergentes.

Sin embargo, se centran principalmente en el comportamiento de los propios sistemas de inteligencia artificial, más que en las dinámicas más amplias de interacción entre humanos y máquinas a lo largo del tiempo.

Por ejemplo, la supervisión humana suele asumir que el actor humano que supervisa el sistema mantiene intacta su capacidad cognitiva para ejercer un juicio significativo.

En la práctica, la interacción prolongada con sistemas automatizados puede alterar gradualmente la forma en que los humanos participan en los procesos de decisión.

Con el tiempo, los individuos pueden volverse menos atentos a los mecanismos subyacentes, más dependientes de las recomendaciones algorítmicas o menos inclinados a cuestionar los resultados generados por el sistema.

Este fenómeno no implica necesariamente negligencia o pérdida formal de responsabilidad.

Más bien refleja un cambio estructural en el funcionamiento de los entornos de decisión cuando humanos y sistemas inteligentes operan conjuntamente.

Como consecuencia, los marcos actuales de gobernanza pueden estar pasando por alto una dimensión crítica de la integración de la inteligencia artificial: la preservación a largo plazo de la agencia cognitiva humana.

Las regulaciones existentes garantizan que los humanos puedan intervenir en los sistemas de IA desde un punto de vista teórico.

Sin embargo, no siempre garantizan que los humanos sigan siendo prácticamente capaces de hacerlo de forma significativa.

Esta distinción adquiere cada vez mayor importancia a medida que los sistemas de IA se integran profundamente en las operaciones diarias de sectores como:

- la sanidad
- las finanzas
- el transporte
- la educación
- la gobernanza pública.

En estos entornos, la responsabilidad humana puede mantenerse formalmente intacta mientras la capacidad práctica de comprender y evaluar decisiones se reduce gradualmente.

Human Layer propone que los futuros marcos de gobernanza podrían necesitar incorporar una perspectiva adicional:

no solo cómo regular el comportamiento técnico de los sistemas de inteligencia artificial, sino también cómo diseñar entornos socio-técnicos en los que la capacidad humana de decisión permanezca activa, informada y resiliente.

En lugar de sustituir los mecanismos de gobernanza existentes, Human Layer puede entenderse como un marco conceptual complementario que amplía el foco de la gobernanza de la IA hacia la sostenibilidad de la participación cognitiva humana dentro de sistemas tecnológicos complejos.



4. La arquitectura Human Layer

Human Layer se propone como un marco conceptual y operativo diseñado para regular la interacción entre la toma de decisiones humanas y los sistemas de inteligencia artificial cada vez más capaces.

En lugar de centrarse en limitar las capacidades tecnológicas de los sistemas de IA, el enfoque de Human Layer se centra en preservar y fortalecer la capacidad humana de comprender, evaluar y asumir la responsabilidad de las decisiones dentro de entornos tecnológicos complejos.

En este sentido, Human Layer no funciona como una restricción a la inteligencia artificial, sino como una capa arquitectónica que garantiza la sostenibilidad de la agencia humana en sistemas altamente automatizados.

El concepto puede entenderse como una capa adicional dentro de la arquitectura de los sistemas socio-técnicos.

La arquitectura tradicional de los sistemas suele componerse de tres elementos principales:

- la capa de datos, donde se recopila y almacena la información
- la capa algorítmica, donde los modelos procesan y analizan los datos
- la capa de ejecución, donde se generan acciones basadas en los resultados del sistema.

Human Layer introduce una capa estructural adicional situada entre los resultados algorítmicos y la ejecución final.

Esta capa actúa como una interfaz reguladora entre la inteligencia de máquina y la toma de decisiones humana.

Su función no consiste simplemente en proporcionar supervisión humana, sino en configurar activamente las condiciones bajo las cuales los humanos interactúan con sistemas inteligentes.

En términos prácticos, la arquitectura Human Layer se basa en varias funciones fundamentales.

En primer lugar, mantiene la implicación cognitiva humana en los procesos de decisión. En lugar de permitir que los humanos aprueben pasivamente las recomendaciones algorítmicas, el sistema fomenta la interpretación activa y la comprensión contextual.

En segundo lugar, garantiza conciencia contextual. Muchos sistemas de IA operan optimizando objetivos específicos definidos dentro de sus parámetros de

entrenamiento. Human Layer introduce mecanismos que animan a los responsables humanos de las decisiones a considerar variables contextuales más amplias que pueden no estar capturadas por el modelo.

En tercer lugar, preserva la responsabilidad en la toma de decisiones. En entornos altamente automatizados, la responsabilidad puede volverse difusa o ambigua. Human Layer refuerza el principio de que los humanos siguen siendo responsables de las decisiones finales y que los sistemas deben apoyar esta responsabilidad en lugar de diluirla.

En cuarto lugar, promueve el aprendizaje adaptativo entre humanos y sistemas. Los entornos Human Layer están diseñados para crear bucles de retroalimentación en los que tanto los participantes humanos como los sistemas tecnológicos mejoran continuamente sus patrones de interacción.

Por último, protege la interpretabilidad humana de procesos complejos. Incluso cuando los algoritmos operan a un nivel de complejidad que supera la capacidad de replicación directa por parte de los humanos, la arquitectura del sistema debe garantizar que los humanos mantengan la capacidad de comprender la lógica estructural de las decisiones.

En conjunto, estas funciones definen Human Layer como una forma de infraestructura cognitiva integrada dentro de los sistemas tecnológicos.

No sustituye a la inteligencia artificial ni reduce sus capacidades.

En cambio, crea las condiciones estructurales necesarias para que la inteligencia humana y la inteligencia de máquina operen de forma estable y mutuamente reforzada.

A medida que la inteligencia artificial se integra cada vez más en funciones críticas de la sociedad, el diseño de esta capa de interacción puede llegar a ser tan importante como el desarrollo de los propios algoritmos.

Human Layer representa, por tanto, un principio arquitectónico para la próxima fase de integración entre humanos y sistemas de inteligencia artificial.



5. Principios fundamentales de Human Layer

El marco de Human Layer se basa en un conjunto de principios fundamentales diseñados para guiar la interacción entre los responsables humanos de la toma de decisiones y los sistemas tecnológicos inteligentes.

Estos principios no pretenden limitar la innovación ni restringir el desarrollo de la inteligencia artificial. Su objetivo es garantizar que el progreso tecnológico refuerce, en lugar de debilitar, el funcionamiento a largo plazo de los sistemas cognitivos humanos.

Cada principio aborda un aspecto estructural de cómo humanos y máquinas interactúan dentro de entornos cada vez más automatizados.

En conjunto, forman la base conceptual de la arquitectura Human Layer.



5.1 Preservación de la agencia humana

Los sistemas basados en Human Layer deben preservar la capacidad de los humanos para participar activamente en los procesos de decisión.

Los sistemas de inteligencia artificial pueden proporcionar análisis, predicciones y recomendaciones, pero la estructura del sistema debe garantizar que los participantes humanos permanezcan cognitivamente implicados.

La agencia humana implica más que la capacidad formal de aprobar o rechazar los resultados generados por el sistema.

Requiere que los humanos mantengan una comprensión significativa del contexto de decisión y de las consecuencias de sus acciones.

Los sistemas que convierten gradualmente a los responsables humanos de las decisiones en validadores pasivos de recomendaciones algorítmicas debilitan este principio.

Human Layer promueve, por tanto, modelos de interacción que fomentan la interpretación, la reflexión y la responsabilidad.



5.2 Sostenibilidad cognitiva

La capacidad humana de decisión es un recurso limitado y valioso.

Los entornos altamente automatizados pueden erosionar involuntariamente esta capacidad al fomentar una dependencia excesiva de los sistemas automatizados.

Human Layer introduce el concepto de sostenibilidad cognitiva, que se refiere a la preservación a largo plazo de las capacidades de razonamiento humano dentro de sistemas tecnológicos complejos.

De la misma manera que la sostenibilidad ambiental busca proteger los ecosistemas naturales, la sostenibilidad cognitiva busca proteger el ecosistema cognitivo humano dentro de entornos de decisión cada vez más complejos.

Los sistemas deben diseñarse, por tanto, de forma que refuercen las capacidades analíticas humanas en lugar de sustituirlas progresivamente.



5.3 Conciencia contextual

Los sistemas de inteligencia artificial operan dentro de los límites definidos por sus datos, modelos y objetivos de optimización.

Sin embargo, muchas decisiones del mundo real implican variables contextuales que pueden quedar fuera del alcance de los modelos algorítmicos.

Human Layer reconoce la importancia del juicio contextual humano.

Los participantes humanos poseen una capacidad única para integrar consideraciones cualitativas, éticas y situacionales que pueden no estar presentes en los modelos computacionales.

Los sistemas Human Layer están diseñados para que el análisis algorítmico y la interpretación contextual humana se complementen.



5.4 Integridad de la responsabilidad

En sistemas socio-técnicos complejos, la responsabilidad puede volverse difusa.

Cuando múltiples capas tecnológicas y procesos automatizados participan en la toma de decisiones, puede resultar difícil identificar quién es responsable de los resultados.

Human Layer refuerza el principio de integridad de la responsabilidad, garantizando que la autoridad final de decisión permanezca claramente identificada.

La arquitectura del sistema debe permitir que los responsables humanos de las decisiones comprendan, expliquen y asuman la responsabilidad de los resultados.

Este principio resulta especialmente importante en sectores donde las decisiones tienen consecuencias sociales, financieras o éticas significativas.



5.5 Complementariedad humano-máquina

La inteligencia humana y la inteligencia de máquina poseen fortalezas diferentes.

La inteligencia artificial destaca en el procesamiento de grandes volúmenes de datos, la identificación de patrones y la realización de cálculos complejos.

La cognición humana destaca en el razonamiento contextual, la evaluación ética y la interpretación de información ambigua o incompleta.

Los sistemas Human Layer están diseñados para maximizar esta complementariedad.

En lugar de intentar sustituir la toma de decisiones humana, la arquitectura busca alinear las fortalezas de ambas formas de inteligencia.

El objetivo no es generar competencia entre humanos y máquinas, sino crear sistemas cooperativos en los que cada uno aporte sus capacidades únicas.



Estos principios constituyen la base conceptual del marco Human Layer.

Establecen las condiciones bajo las cuales la inteligencia artificial puede evolucionar sin comprometer la estabilidad de la capacidad humana de decisión.

A medida que los sistemas tecnológicos se vuelven más avanzados, preservar este equilibrio puede convertirse en uno de los desafíos centrales de la gobernanza de los ecosistemas humano-IA.



6. Modelo operativo de los sistemas Human Layer

Aunque el marco de Human Layer es de naturaleza conceptual, está diseñado para poder implementarse dentro de sistemas socio-técnicos reales. Su propósito no es únicamente ofrecer una reflexión teórica, sino organizar de manera práctica entornos de decisión en los que humanos y sistemas de inteligencia artificial interactúan de forma continua.

En términos operativos, Human Layer funciona como una arquitectura de interfaz de decisión que regula cómo fluyen la información, las recomendaciones y las acciones entre los sistemas inteligentes y los participantes humanos.

El modelo puede describirse como un ciclo estructurado de interacción compuesto por varias fases.



6.1 Detección de señales

Todo sistema Human Layer comienza con la detección de señales dentro de un entorno determinado.

Estas señales pueden originarse en distintas fuentes:

- análisis algorítmico de grandes volúmenes de datos
- modelos predictivos que identifican patrones emergentes
- observaciones humanas o aportaciones contextuales
- cambios en el entorno externo.

Los sistemas de inteligencia artificial desempeñan un papel importante en esta fase, ya que son capaces de procesar grandes cantidades de información e identificar correlaciones que pueden no ser inmediatamente visibles para los observadores humanos.

Sin embargo, el resultado de esta fase no desencadena automáticamente una acción.

En su lugar, genera señales estructuradas que pasan al proceso de interacción Human Layer.



6.2 Capa de interpretación

Una vez generadas las señales, el sistema introduce una fase de interpretación en la que los participantes humanos evalúan la información proporcionada por los sistemas de IA.

Esta etapa es fundamental para mantener la implicación cognitiva activa.

En lugar de limitarse a aprobar resultados automatizados, los responsables humanos de la decisión interpretan el significado, la relevancia y las posibles implicaciones de las señales.

Este paso interpretativo permite integrar factores que pueden no estar presentes en los datos utilizados por el algoritmo.

Entre estos factores pueden encontrarse:

- información contextual
- restricciones institucionales

- consideraciones éticas
- implicaciones estratégicas a largo plazo.

De esta manera, Human Layer garantiza que los resultados generados por la IA se traten como inputs para el razonamiento, no como instrucciones automáticas.



6.3 Formación de la decisión

Después de la fase de interpretación comienza a formarse una decisión.

En esta etapa, los responsables humanos combinan varios elementos:

- recomendaciones algorítmicas
- interpretación contextual
- experiencia profesional
- evaluación del riesgo.

La inteligencia artificial puede asistir en la simulación de posibles resultados o en la presentación de escenarios alternativos, pero la responsabilidad de integrar estos elementos en una decisión coherente sigue siendo humana.

La fase de formación de la decisión representa, por tanto, el punto en el que convergen la inteligencia de máquina y el juicio humano.



6.4 Ejecución

Una vez formada la decisión, el sistema procede a su ejecución.

La ejecución puede implicar procesos automatizados, acciones humanas o una combinación de ambos.

Es importante destacar que la arquitectura Human Layer garantiza que la ejecución de las decisiones sea trazable y comprensible.

El sistema preserva la capacidad de reconstruir el proceso de razonamiento que condujo a la acción final.

Esta trazabilidad refuerza la transparencia, la responsabilidad y el aprendizaje dentro de la organización.



6.5 Retroalimentación y aprendizaje

La fase final del modelo operativo es la retroalimentación.

Después de ejecutar una decisión, el sistema evalúa los resultados obtenidos.

Tanto los participantes humanos como los sistemas de inteligencia artificial aprenden a partir de esta retroalimentación.

En el caso de los modelos de inteligencia artificial, la retroalimentación puede utilizarse para mejorar la precisión predictiva o refinar el rendimiento del modelo.

En el caso de los participantes humanos, contribuye a mejorar el juicio y la comprensión situacional.

La interacción entre estos dos procesos de aprendizaje genera un ciclo adaptativo continuo en el que tanto las capacidades humanas como las tecnológicas evolucionan con el tiempo.



Este modelo operativo demuestra cómo Human Layer puede funcionar como un sistema estructurado de interacción, más allá de ser un concepto puramente teórico.

Al introducir fases explícitas de interpretación, formación de decisión y retroalimentación, Human Layer garantiza que el creciente poder de los sistemas de inteligencia artificial refuerce, en lugar de debilitar, la capacidad humana de decisión.

En entornos altamente automatizados, esta arquitectura ayuda a mantener el equilibrio entre la eficiencia tecnológica y la responsabilidad humana.



7. Aplicaciones potenciales de Human Layer

El marco de Human Layer está diseñado como un principio arquitectónico general que puede aplicarse en múltiples sectores donde los sistemas de inteligencia artificial interactúan con responsables humanos de la toma de decisiones.

A medida que las tecnologías de IA se integran cada vez más en los procesos operativos, la necesidad de mantener una participación humana significativa en los entornos de decisión se vuelve relevante en una amplia variedad de ámbitos.

Human Layer no depende de una tecnología específica, de una industria concreta ni de una estructura organizativa determinada.

Más bien proporciona un marco conceptual para estructurar la interacción humano-IA en sistemas donde las decisiones tienen consecuencias relevantes.

Varios sectores ilustran especialmente bien cómo el enfoque Human Layer puede implementarse.



7.1 Servicios financieros y gestión patrimonial

Los entornos de decisión financiera dependen cada vez más del análisis algorítmico.

Los sistemas de inteligencia artificial se utilizan ampliamente para optimización de carteras, gestión de riesgos, análisis de mercados y cumplimiento normativo.

Aunque estas tecnologías aumentan significativamente la capacidad analítica, las decisiones financieras suelen implicar consideraciones humanas complejas que van más allá del análisis puramente cuantitativo.

Entre estas consideraciones se encuentran:

- la psicología del cliente y sus dinámicas de comportamiento
- la planificación familiar a largo plazo y la transmisión intergeneracional del patrimonio
- las consideraciones éticas en las estrategias de inversión
- la interpretación de la incertidumbre macroeconómica.

En estos entornos, los sistemas Human Layer pueden ayudar a garantizar que los conocimientos generados por la IA sigan siendo herramientas para una toma de decisiones informada, en lugar de sustituir el juicio humano.

El asesor o responsable de la decisión mantiene un papel activo en la interpretación de los resultados algorítmicos y en su integración dentro de marcos estratégicos más amplios.

Este modelo refuerza la confianza en la relación financiera al mismo tiempo que aprovecha el poder analítico de los sistemas tecnológicos avanzados.



7.2 Sanidad y apoyo a la decisión médica

La sanidad representa otro ámbito crítico en el que el juicio humano y la inteligencia artificial interactúan cada vez más.

Los sistemas de IA ya son capaces de asistir en el análisis de imágenes médicas, el apoyo al diagnóstico, la optimización de tratamientos y la monitorización de pacientes.

Sin embargo, la toma de decisiones médicas implica consideraciones éticas, preferencias del paciente y factores contextuales que no pueden capturarse completamente mediante modelos algorítmicos.

Las arquitecturas Human Layer pueden apoyar a los profesionales sanitarios garantizando que los conocimientos diagnósticos generados por la IA permanezcan integrados dentro de un proceso de razonamiento clínico más amplio.

En lugar de sustituir a los médicos, los sistemas de IA funcionan como socios analíticos que mejoran la calidad de las decisiones médicas al tiempo que preservan el papel central del cuidado humano y la responsabilidad profesional.



7.3 Sistemas de decisión corporativos

Dentro de las organizaciones, la inteligencia artificial se utiliza cada vez más para apoyar la planificación estratégica, la gestión operativa, la asignación de recursos humanos y el análisis del comportamiento de los clientes.

En entornos corporativos complejos, la interacción entre el liderazgo humano y los sistemas algorítmicos se vuelve crítica.

Las arquitecturas Human Layer pueden ayudar a las organizaciones a diseñar procesos de decisión en los que:

- los sistemas automatizados identifican patrones y oportunidades
- los líderes humanos interpretan las implicaciones estratégicas
- las decisiones finales integran evidencia analítica y contexto organizativo.

Este tipo de sistemas reduce el riesgo de sobreautomatización mientras aumenta la eficacia del liderazgo humano en entornos tecnológicamente complejos.



7.4 Gobernanza pública y diseño de políticas

Los gobiernos y las instituciones públicas están explorando cada vez más el uso de inteligencia artificial en ámbitos como la planificación urbana, la gestión de la salud pública, el análisis de políticas económicas y la supervisión regulatoria.

Los entornos de decisión pública son especialmente sensibles porque afectan a grandes poblaciones y suelen implicar consideraciones éticas, sociales y políticas complejas.

Los principios de Human Layer pueden apoyar la integración responsable de la IA en los sistemas de gobernanza garantizando que las herramientas algorítmicas asistan, pero no sustituyan, las decisiones políticas.

Los responsables públicos siguen siendo quienes interpretan las prioridades sociales, evalúan las implicaciones éticas y mantienen la responsabilidad democrática.



7.5 Entornos de decisión personal

Más allá de las aplicaciones institucionales, los principios de Human Layer también pueden aplicarse a entornos de decisión individuales.

A medida que tecnologías personales como los dispositivos wearables, los asistentes digitales y los sistemas de recomendación se vuelven más avanzados, las personas dependen cada vez más de la orientación algorítmica en ámbitos como la salud, la planificación financiera o la organización de la vida cotidiana.

Los sistemas inspirados en Human Layer podrían apoyar a los individuos ayudándoles a mantener la conciencia y el control sobre los procesos de decisión que afectan a sus vidas.

En lugar de sustituir el juicio personal, estos sistemas funcionarían como compañeros cognitivos diseñados para apoyar una toma de decisiones reflexiva.



Estos ejemplos ilustran la flexibilidad del concepto Human Layer.

El marco puede adaptarse a una gran variedad de sectores manteniendo su objetivo central: garantizar que los sistemas tecnológicos refuercen, y no reduzcan, el papel de la inteligencia humana en entornos de decisión complejos.

A medida que la inteligencia artificial continúe evolucionando, la capacidad de diseñar modelos estables y sostenibles de interacción humano-IA puede convertirse en uno de los desafíos definitorios de las próximas décadas.



8. Implicaciones para la gobernanza y la regulación de la IA

A medida que los sistemas de inteligencia artificial se integran cada vez más en infraestructuras económicas, institucionales y sociales, la cuestión de cómo gobernar estas tecnologías se convierte en un elemento central para la estabilidad de las sociedades modernas.

Los marcos actuales de gobernanza se centran principalmente en garantizar que los sistemas de inteligencia artificial sean seguros, transparentes y estén alineados con estándares legales y éticos.

Estos objetivos siguen siendo esenciales.

Sin embargo, el marco de Human Layer sugiere que las discusiones sobre gobernanza pueden necesitar considerar también una dimensión adicional: la estabilidad a largo plazo de la capacidad humana de decisión dentro de entornos mediadores por IA.

En otras palabras, la gobernanza no debería limitarse únicamente a regular lo que los sistemas de inteligencia artificial pueden hacer, sino también considerar cómo su integración afecta al papel cognitivo e institucional de los responsables humanos de las decisiones.

Human Layer propone así una perspectiva complementaria para los futuros modelos de gobernanza.



8.1 De la regulación del sistema a la regulación de la interacción

La mayoría de los marcos regulatorios actuales se centran en el comportamiento técnico de los sistemas de inteligencia artificial.

Su objetivo es garantizar que los algoritmos operen de forma segura, eviten resultados discriminatorios y ofrezcan niveles adecuados de transparencia.

Human Layer introduce la idea de que la gobernanza debería abordar también el diseño de la interacción entre humanos y sistemas de IA.

Incluso cuando un sistema funciona correctamente desde el punto de vista técnico, entornos de interacción mal diseñados pueden reducir gradualmente la implicación y la comprensión humanas.

Por ello, los enfoques regulatorios podrían beneficiarse de ampliar su foco hacia la arquitectura de los entornos de decisión, más allá de las propiedades de los propios algoritmos.



8.2 Preservar una supervisión humana significativa

Muchos marcos políticos y regulatorios actuales destacan la importancia de mantener supervisión humana sobre los sistemas de inteligencia artificial.

Sin embargo, la supervisión puede adoptar diferentes formas.

En algunos casos, los humanos conservan formalmente la autoridad sobre sistemas automatizados, pero interactúan con ellos de forma superficial.

Human Layer subraya la importancia de una supervisión humana significativa, en la que los participantes humanos sigan siendo capaces de comprender y evaluar el razonamiento detrás de los resultados generados por el sistema.

Esto requiere diseñar sistemas que fomenten la interpretabilidad, el razonamiento contextual y la implicación activa, en lugar de limitarse a la aprobación pasiva de resultados.

Los marcos de gobernanza pueden llegar a reconocer cada vez más la diferencia entre una supervisión humana simbólica y una participación humana real en la toma de decisiones.



8.3 Resiliencia institucional en sistemas altamente automatizados

A medida que los sistemas de inteligencia artificial se integran en infraestructuras críticas, las instituciones pueden volverse cada vez más dependientes de procesos algorítmicos.

Human Layer sugiere que preservar la capacidad humana de decisión dentro de estos sistemas contribuye a la resiliencia institucional.

Las organizaciones que mantienen fuertes capacidades humanas de interpretación pueden estar mejor preparadas para responder a situaciones inesperadas, fallos del sistema o entornos que cambian rápidamente.

Esta perspectiva está alineada con enfoques más amplios de gestión del riesgo, que destacan la importancia de mantener múltiples fuentes de juicio e interpretación dentro de sistemas complejos.



8.4 Alineación con los marcos regulatorios emergentes

El concepto Human Layer es ampliamente compatible con muchos de los principios regulatorios existentes.

Por ejemplo, dentro del entorno regulatorio europeo, iniciativas como el AI Act destacan la importancia de la supervisión humana, la gestión del riesgo y la responsabilidad en el despliegue de sistemas de inteligencia artificial.

Human Layer puede entenderse como un marco conceptual que ayuda a operacionalizar estos principios dentro de arquitecturas reales de sistemas.

En lugar de introducir nuevas cargas regulatorias, el marco proporciona orientación sobre cómo diseñar sistemas que apoyen de forma natural objetivos regulatorios como la transparencia, la responsabilidad y la confianza.



8.5 Una perspectiva de gobernanza a largo plazo

La inteligencia artificial seguirá siendo un campo tecnológico en rápida evolución.

Los marcos de gobernanza deberán, por tanto, adaptarse continuamente a nuevas capacidades y aplicaciones.

Human Layer introduce una perspectiva a largo plazo al centrarse no solo en los riesgos inmediatos de tecnologías específicas, sino también en las condiciones estructurales que permiten que la toma de decisiones humana siga siendo robusta dentro de sociedades tecnológicamente avanzadas.

Al integrar estas consideraciones en la arquitectura de los sistemas humano-IA, los responsables políticos y los diseñadores de sistemas pueden contribuir a garantizar que el progreso tecnológico siga estando alineado con el funcionamiento a largo plazo de las instituciones humanas.



Human Layer ofrece así una perspectiva complementaria desde la que abordar la gobernanza de la inteligencia artificial.

Invita a responsables políticos, ingenieros y organizaciones a considerar cómo la interacción entre humanos y máquinas configura los entornos de decisión a lo largo del tiempo.

Esta perspectiva puede resultar cada vez más valiosa a medida que los sistemas de inteligencia artificial se vuelven más potentes y se integran más profundamente en las estructuras que sostienen las sociedades modernas.



9. Limitaciones y futuras líneas de investigación

El marco de Human Layer se presenta como una contribución conceptual al debate en evolución sobre la integración de la inteligencia artificial en los entornos de decisión humanos.

Como ocurre con cualquier marco emergente, existen limitaciones que deben ser reconocidas y exploradas mediante futuras investigaciones.

Reconocer estas limitaciones es esencial para garantizar que el concepto permanezca abierto al refinamiento, la validación empírica y la experimentación práctica.



9.1 Naturaleza conceptual del marco

En su estado actual, Human Layer es principalmente una arquitectura conceptual, más que un protocolo técnico completamente estandarizado.

El marco define principios, componentes estructurales y modelos de interacción, pero su implementación práctica puede variar en función de las características específicas de cada sector u organización.

Los desarrollos futuros podrían centrarse en traducir el modelo conceptual en estándares operativos de diseño, guías de interfaz y métricas que permitan evaluar el nivel de implicación cognitiva humana dentro de sistemas automatizados.

Estos avances permitirían que el concepto Human Layer evolucione desde una arquitectura teórica hacia metodologías prácticas de diseño de sistemas.



9.2 Desafíos de medición

Una de las ideas centrales de Human Layer es la preservación de la implicación cognitiva humana en los procesos de decisión.

Sin embargo, medir la implicación cognitiva dentro de sistemas complejos presenta importantes desafíos metodológicos.

Algunos indicadores pueden ser observables, como:

- la frecuencia de intervención humana
- los procesos de explicación de decisiones
- los mecanismos de retroalimentación entre humanos y sistemas.

Sin embargo, otros aspectos implican dimensiones cualitativas del razonamiento humano que resultan más difíciles de cuantificar.

Investigaciones interdisciplinarias que combinen campos como:

- ciencia cognitiva
- interacción humano-computador
- economía conductual
- diseño organizativo

podrían contribuir a desarrollar marcos de medición más robustos en el futuro.



9.3 Adaptación sectorial

Los diferentes sectores presentan distintos tipos de entornos de decisión.

Por ejemplo:

- los sistemas financieros priorizan la gestión del riesgo y la planificación a largo plazo
- los entornos sanitarios enfatizan la seguridad y la responsabilidad ética
- la gobernanza pública implica rendición de cuentas democrática y legitimidad institucional
- los entornos corporativos priorizan la eficiencia operativa y la agilidad estratégica.

La implementación de los principios de Human Layer probablemente requerirá adaptaciones específicas para cada sector, con el fin de abordar las dinámicas particulares de cada contexto.

Las investigaciones futuras podrían centrarse en desarrollar modelos de aplicación detallados para dominios específicos.



9.4 Interacción con el rápido desarrollo tecnológico

Las tecnologías de inteligencia artificial continúan evolucionando con gran rapidez.

Los avances en modelos generativos, sistemas autónomos y arquitecturas multiagente pueden modificar significativamente la forma en que los humanos interactúan con los sistemas inteligentes.

Por ello, Human Layer debe mantenerse como un marco flexible y adaptable.

El concepto no debe interpretarse como un diseño estático, sino como un principio arquitectónico en evolución capaz de integrarse con futuros desarrollos tecnológicos.

El diálogo continuo entre tecnólogos, responsables políticos e investigadores será esencial para garantizar que el marco siga siendo relevante a medida que surjan nuevas capacidades.



9.5 Validación empírica

En última instancia, el valor del concepto Human Layer dependerá de su capacidad para demostrar resultados positivos cuando se implemente en sistemas reales.

Los estudios empíricos sobre arquitecturas inspiradas en Human Layer podrían explorar cuestiones como:

- si estos sistemas mejoran la calidad de la toma de decisiones humanas
- si reducen la dependencia excesiva de las recomendaciones automatizadas
- si aumentan la resiliencia institucional en entornos altamente automatizados
- si incrementan la confianza en los sistemas de decisión asistidos por inteligencia artificial.

Las implementaciones piloto en sectores como las finanzas, la sanidad o la gobernanza corporativa podrían proporcionar información valiosa sobre la eficacia práctica del marco.



Reconocer estas limitaciones no representa una debilidad del planteamiento Human Layer, sino un paso natural en el desarrollo de nuevos modelos conceptuales.

El marco pretende abrir una conversación sobre cómo el diseño de la interacción humano-IA puede influir en el funcionamiento a largo plazo de las sociedades complejas.

La investigación futura y la experimentación práctica desempeñarán un papel clave en la mejora, validación y ampliación de las ideas presentadas en este documento.



10. Conclusión

La inteligencia artificial se está convirtiendo rápidamente en un componente fundamental de los sistemas tecnológicos modernos. Desde los servicios financieros y la sanidad hasta la gobernanza pública y la gestión corporativa, los sistemas de IA influyen cada vez más en los entornos en los que se toman decisiones críticas.

La narrativa dominante en torno a la inteligencia artificial suele centrarse en la capacidad tecnológica: hasta qué punto pueden llegar a ser potentes los sistemas, con qué eficiencia pueden procesar información y con qué precisión pueden generar predicciones.

Si bien estas capacidades son importantes, representan solo una parte de la transformación más amplia que está teniendo lugar.

La integración de la inteligencia artificial no es únicamente una evolución tecnológica; es una transformación estructural de los entornos en los que se producen las decisiones humanas.

A medida que los sistemas inteligentes se vuelven más capaces, el diseño de la interacción entre humanos y máquinas se convierte en un factor determinante para la estabilidad de los sistemas socio-técnicos complejos.

Human Layer propone que mantener la eficacia a largo plazo de estos sistemas requiere prestar atención explícita a la preservación de la agencia cognitiva humana.

En lugar de considerar la inteligencia artificial únicamente como una herramienta para optimizar resultados, Human Layer plantea la integración de la IA como un desafío arquitectónico: cómo diseñar sistemas en los que la inteligencia humana y la inteligencia de máquina se refuercen mutuamente en lugar de competir o sustituirse.

El marco presentado en este documento propone que la interacción entre humanos y sistemas de inteligencia artificial debe convertirse en un dominio deliberado de diseño.

Mediante la introducción de una capa arquitectónica orientada a mantener la implicación humana, el razonamiento contextual y la integridad de la responsabilidad, Human Layer busca garantizar que el progreso tecnológico fortalezca, en lugar de erosionar, las capacidades humanas que sostienen instituciones, organizaciones y sociedades.

Es importante destacar que Human Layer no se opone al desarrollo tecnológico.

Por el contrario, parte de la premisa de que la inteligencia artificial continuará expandiendo sus capacidades y su influencia.

El objetivo del marco no es ralentizar la innovación, sino asegurar que la evolución de los sistemas inteligentes siga siendo compatible con el funcionamiento a largo plazo de las estructuras de decisión humanas.

A medida que las sociedades avanzan hacia una era de automatización generalizada y de inteligencia algorítmica, la cuestión ya no es únicamente cuán poderosas pueden llegar a ser las máquinas.

La cuestión más fundamental es cómo convivirán humanos y máquinas dentro de los mismos entornos de decisión.

Human Layer ofrece una base conceptual para abordar esta cuestión, subrayando que la sostenibilidad de la cognición humana debe seguir siendo una consideración central en el diseño de los sistemas tecnológicos del futuro.

En este sentido, Human Layer puede entenderse no como una restricción a la inteligencia artificial, sino como una arquitectura habilitadora para la próxima etapa de cooperación entre humanos y sistemas inteligentes.

