

INTRODUCCIÓN

Desarrollar **sistemas basados en inteligencia artificial (SBIA)** en dominios sensibles como salud, finanzas, justicia y militar requiere integrar principios éticos para mitigar riesgos [1]. Se propone un método ágil para tratar sesgos básicos e identificar sesgos inherentes, utilizando principios ágiles para asegurar transparencia, equidad y responsabilidad en cada fase del desarrollo del SBIA. Las **historias de usuario (HU)** se emplean para formular escenarios éticos [2], definir **criterios de aceptación (CA)** basados en principios éticos, y **pruebas de aceptación (PA)** como métricas de evaluación.

OBJETIVO

Desarrollar un método para integrar principios éticos fundamentales en cada fase del desarrollo del SBIA, utilizando metodologías ágiles, a través del uso de HU.

- Integrar principios éticos fundamentales
- Detectar sesgos inherentes
- Validar resultados a través de las PA

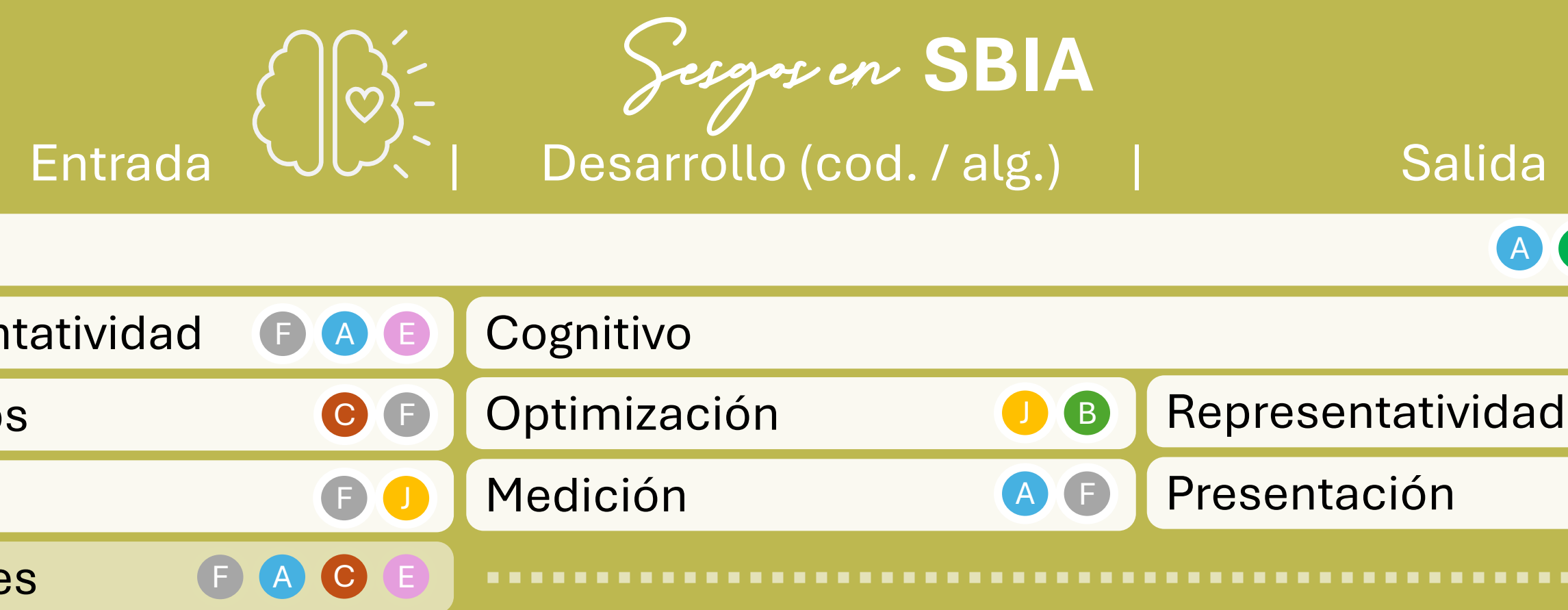
METODOLOGÍA

La propuesta se fundamenta en el marco HXAI, desarrollado en 2022 [4]. Se integraron principios éticos esenciales y se aplicó la clasificación de sesgos básicos, adicionalmente se proponen actividades para identificar sesgos inherentes y asegurar la transparencia y equidad en la toma de decisiones multidisciplinarias. Las HU y CA éticos fueron implementados como técnicas clave para una integración de los principios éticos. Se utilizó un ejemplo ilustrativo (SBIA para Análisis Genómico y Bioquímico en pacientes con artritis reumatoide) para mejorar la comprensión del método y una encuesta para identificar diferencias cognitivas y preferencias éticas entre perfiles de usuarios.



DISCUSIÓN

Los SBIA en dominios sensibles requieren una rigurosa identificación y mitigación de sesgos para asegurar decisiones éticas y responsables [1]. La creciente aplicación de la inteligencia artificial en la atención médica presenta numerosos desafíos éticos [3]. La integración de principios éticos a través de metodologías ágiles permite abordar tanto sesgos básicos como inherentes, disminuyendo los riesgos éticos y técnicos asociados [2]. Además, principios como la explicabilidad y la transparencia son esenciales para mejorar la confianza y responsabilidad en estos sistemas [4].



Ética en el Desarrollo de Sistemas Basados en IA, con metodología ágil



Principios Éticos Fundamentales

- A** **Transparencia:** La IA debe ser transversal y sus procesos explicables para generar confianza.
✦ Disponibilidad de la información, documentación, reportes, soporte, capacitaciones.
- B** **Explicabilidad:** Las decisiones de la IA deben ser entendibles para los humanos.
✦ Interfaces comprensibles e intuitivas, técnicas explainable AI, pruebas de usuario, user experience (UX), información clara.
- C** **Responsabilidad y Rendición:** Asumir las consecuencias, documentar las decisiones, tratar los riesgos.
✦ Registros, comités éticos, validación de resultados, seguimiento y métricas, revisión periódica de riesgos, políticas de acción.
- D** **Control y Regulación:** La IA debe ser supervisada para garantizar su alineación con estándares éticos y legales.
✦ Control en cada iteración, pruebas de aceptación, comité ético, auditorías internas, regulaciones y marcos legales.
- E** **Evaluación Continua:** Realizar evaluaciones periódicas para asegurar la eficacia y seguridad de la IA.
✦ Pruebas de aceptación, revisión de métricas, feedback, pruebas A/B, análisis post-despliegue, retroalimentación de usuarios.
- F** **Equidad y No Discriminación:** Equidad y justicia en los SBIA, sin sesgos ni discriminación, justicia e imparcialidad.
✦ Inclusión de perfiles diversos, herramientas de equidad, revisión de datos de entrenamiento, integración en toma de decisiones.
- G** **Autonomía del Usuario:** Los usuarios deben poder tomar decisiones informadas, manteniendo el control sobre la IA.
✦ Diseño de interfaces comprensibles, opción de personalización en interfaces, botones de consentimiento informado.
- H** **Privacidad y Protección de Datos:** Datos personales protegidos y manejados de forma segura.
✦ Contratos de confidencialidad, anonimización de datos, regulaciones, encriptación de datos, control de accesos.
- I** **Seguridad y Robustez:** La IA debe proteger a los usuarios, evitando daños y manteniendo la integridad del sistema.
✦ Pruebas de seguridad, control de accesos, mecanismos de detección de ataques, pruebas de estrés, manejo de fallos, monitoreo continuo.
- J** **Beneficencia:** La IA debe promover el bienestar y generar beneficios para la sociedad, evitando daños.
✦ Enfoque al impacto social, inclusión de métricas de impacto social, evaluación de riesgo social, definición de beneficios explícitos.
- K** **Sostenibilidad e Impacto Ambiental:** Minimizar el impacto ambiental y ser eficiente a largo plazo.
✦ Energías renovables, datos locales, análisis de impacto ambiental, eficiencia y optimización en los algoritmos y procesamiento de datos.

CONTRIBUCIÓN

Nuestra propuesta se integra en metodologías ágiles sin alterar sus actividades, eventos ni artefactos. Destaca por su flexibilidad y capacidad de incorporar herramientas éticas, lo que minimiza riesgos. Proporciona un estándar ético que evalúa la ética en cada iteración mediante PA, garantizando una mejora continua. Además, optimiza tiempos y recursos de desarrollo, asegurando la viabilidad del proyecto. Facilita la colaboración multidisciplinaria, asegurando que las decisiones estén alineadas con las expectativas de todos los perfiles involucrados. También promueve la confianza del usuario final al priorizar la transparencia, la explicabilidad y la equidad. Finalmente, contribuye a cumplir con normativas éticas y regulaciones emergentes, lo que es fundamental en dominios sensibles.

RESULTADOS

Los resultados indican que el 60% de bioquímicos y pacientes priorizan la privacidad y protección de datos, mientras que el 70% de médicos y genetistas destacan la importancia de la transparencia y explicabilidad. Estos hallazgos refuerzan la necesidad de ajustar los principios éticos a cada perfil de usuario, mediante el uso de HU y CA éticos para mitigar sesgos inherentes y promover la equidad en SBIA. Además, se logró desarrollar un método ético integrado en metodologías ágiles que corresponde al objetivo de esta propuesta. Cabe destacar que, aunque la investigación es de carácter básica y el ejemplo es ilustrativo, la encuesta realizada fue real.

CONCLUSIÓN

La implementación de nuestro método ético en el desarrollo de SBIA bajo metodologías ágiles asegura la integración y revisión de principios éticos en cada fase del proceso. A través de HU y CA éticos, se facilita la identificación y tratamiento de sesgos básicos e inherentes, promoviendo un desarrollo más responsable y robusto. Además, este enfoque fomenta la colaboración multidisciplinaria, esencial para la detección de sesgos inherentes, garantizando la adaptabilidad del método a distintos proyectos en dominios sensibles.

REFERENCIAS

- [1] Ehsan, U., Wintersberger, P., Liao, Q. V., Mara, M.,... & Riedl, M. O. (2021, May). Operationalizing human-centered perspectives in explainable AI. In Extended abstracts of the 2021 CHI conference on human factors in computing systems (pp. 1-6).
- [2] Vakkuri, V., Kemell, K. K., Jantunen, M., Halme, E., & Abrahamsson, P. (2021). ECCOLA—A method for implementing ethically aligned AI systems. *Journal of Systems and Software*, 182, 111067.
- [3] Bunster, C. E., Rojas, A., Arriagada, A., Salinas, R., Taboada, P., Ruiz-Esqueda, G., ... & Borja, H. (2024). Consideraciones éticas sobre la utilización de la inteligencia artificial en la atención clínica. *Cuadernos Médico Sociales*, 64(2), 53-59.
- [4] De los Angeles D. HXAI. (2024). <https://www.hxai.cl>