

# Desafíos éticos para programadores en la era de la IA

## Material de lectura para las clases 3 y 4

*Ética Profesional 2025*

*Universidad Nacional de Lomas de Zamora*

*Facultad de Ingeniería - Tecnicatura en Programación de Computadores*

*Autor: Prof. Damián A. Iusem*

*Con la colaboración de Claude 3.7 Sonnet (Anthropic) y Gemini 2.0*

---

## 1. Introducción: La IA como fenómeno transformador

El cine ha explorado frecuentemente las implicaciones de la inteligencia artificial avanzada. Películas como "Her" (2013) nos invitan a reflexionar sobre un futuro que parece estar llegando en el que las máquinas podrían desarrollar formas de interacción tan sofisticadas que desafían nuestra comprensión de lo que significa ser inteligente o incluso consciente. Estas obras, lejos de ser mera ciencia ficción, plantean preguntas fundamentales que ya comenzamos a enfrentar en el mundo real.

La inteligencia artificial (IA) representa una de las transformaciones tecnológicas más significativas de nuestra era, un cambio de paradigma comparable a la revolución industrial o la invención de la imprenta. Estamos presenciando una revolución donde las máquinas son capaces de realizar tareas que solían requerir inteligencia humana: reconocer patrones complejos, procesar lenguaje natural, tomar decisiones basadas en grandes cantidades de datos y, más recientemente, generar contenido creativo que resulta prácticamente indistinguible del producido por humanos.

**La IA no es solo una herramienta técnica, sino un fenómeno sociotécnico** con profundas implicaciones para todos los aspectos de la vida cotidiana. Desde la forma en que buscamos información, recibimos recomendaciones de contenido, obtenemos préstamos bancarios, participamos en procesos de selección laboral, o incluso cómo nos

diagnostican enfermedades, los sistemas algorítmicos han pasado a mediar numerosas interacciones sociales.

En este contexto, el rol del programador adquiere una dimensión ética fundamental. Ya no son simples "implementadores técnicos" de requisitos funcionales, sino **arquitectos de sistemas sociotécnicos complejos** que moldean cómo las personas interactúan con instituciones, acceden a oportunidades y comprenden el mundo que les rodea.

## 2. Sistemas autónomos y algoritmos de decisión: Riesgos éticos fundamentales

Los sistemas de IA y algoritmos de decisión actuales plantean riesgos éticos significativos que debemos comprender como programadores. Entre estos, la opacidad algorítmica representa uno de los desafíos más complejos. Muchos sistemas de IA, especialmente aquellos basados en aprendizaje profundo (deep learning), funcionan como "cajas negras" donde incluso sus diseñadores no pueden explicar completamente cómo llegan a determinadas conclusiones. Esta opacidad plantea serios problemas cuando estos sistemas toman decisiones que afectan a derechos fundamentales de las personas.

**El derecho a una explicación** emerge como un principio ético fundamental: las personas afectadas por decisiones automatizadas deberían poder comprender, al menos en términos generales, cómo y por qué se tomó una decisión que les afecta. Como programadores, deberán priorizar la "explicabilidad" en sus diseños, lo que implica preferir modelos interpretables cuando sea posible, implementar técnicas de explicabilidad cuando utilicen modelos más complejos, documentar cuidadosamente los supuestos del modelo, y diseñar interfaces que comuniquen adecuadamente tanto las capacidades como las limitaciones del sistema.

Otro aspecto crítico concierne a la **autonomía humana y agencia**. Los sistemas autónomos plantean preguntas profundas sobre el grado de control que delegamos en las máquinas. Cuando un sistema de navegación decide nuestra ruta, un algoritmo de recomendación determina qué información consumimos, o un asistente virtual programa nuestras reuniones, estamos cediendo parcelas de nuestra autonomía decisoria.

La preservación de la agencia humana significativa debe ser un valor central en el diseño de estos sistemas. Esto requiere garantizar que sean los humanos quienes mantengan el control final sobre las decisiones importantes, diseñar interfaces que promuevan la reflexión

en lugar de la aceptación pasiva, ofrecer opciones genuinas y no meramente ilusorias, y evitar diseños manipulativos ("dark patterns") que exploten sesgos cognitivos para dirigir el comportamiento de manera inadvertida.

La **responsabilidad y rendición de cuentas** representa otro desafío fundamental. La autonomía de los sistemas de IA genera lo que algunos filósofos han denominado "brecha de responsabilidad": cuando un sistema automatizado causa un daño, no siempre resulta claro quién debe asumir la responsabilidad. ¿Es el programador? ¿La empresa que lo implementa? ¿El usuario que lo opera? ¿O incluso el propio sistema?

Esta difusión de la responsabilidad puede llevar a situaciones donde nadie se considera verdaderamente responsable por los resultados adversos. Como profesionales deben implementar sistemas de trazabilidad y auditoría rigurosos, establecer estructuras claras de responsabilidad en equipos de desarrollo, aplicar marcos de gobernanza algorítmica apropiados, y fomentar una cultura de responsabilidad proactiva que anticipe problemas potenciales en lugar de simplemente reaccionar a las crisis.

### 3. Sesgos algorítmicos: Identificación y mitigación

Uno de los problemas éticos más significativos de los sistemas de IA es su propensión a reproducir y amplificar sesgos existentes en la sociedad. Es crucial comprender que **los algoritmos no son inherentemente objetivos**; en realidad, pueden consolidar y magnificar prejuicios presentes en los datos con los que se entrenan o en las decisiones de diseño de sus creadores.

Los sesgos algorítmicos se manifiestan de diversas formas. Cuando los datos de entrenamiento reflejan inequidades históricas, como puede ocurrir con datos de contratación que reflejan discriminación racial, hablamos de sesgos históricos. Cuando ciertos grupos están subrepresentados en los datos, como sucede con sistemas de reconocimiento facial entrenados principalmente con rostros caucásicos, enfrentamos sesgos de representación. Si utilizamos indicadores indirectos que miden inadecuadamente el fenómeno real, como usar el "código postal" como indicador de solvencia crediticia (que puede correlacionar con factores raciales o socioeconómicos no relevantes), estamos ante sesgos de medición. Al aplicar conclusiones generales a individuos sin considerar particularidades relevantes, incurrimos en sesgos de agregación. Y cuando las decisiones sobre cómo se despliega el sistema en contextos reales introducen distorsiones, hablamos de sesgos de implementación.

Como programadores, tendrán la responsabilidad de implementar estrategias efectivas para identificar y mitigar estos sesgos. Esto incluye realizar auditorías algorítmicas que evalúen sistemáticamente el comportamiento del sistema con diferentes subgrupos poblacionales; asegurar la diversidad y representatividad en los conjuntos de datos de entrenamiento; incluir perspectivas diversas en los procesos de diseño y evaluación mediante equipos interdisciplinarios; implementar métricas para medir y garantizar la equidad entre diferentes grupos; y aplicar técnicas algorítmicas específicas para reducir los sesgos detectados.

El caso del sistema COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), utilizado en Estados Unidos para predecir la reincidencia criminal, ilustra perfectamente estos problemas. Un análisis de ProPublica en 2016 reveló que el sistema asignaba sistemáticamente puntuaciones de riesgo más altas a personas afroamericanas que a personas blancas con historiales similares. Este caso demuestra cómo un algoritmo aparentemente "neutral" puede perpetuar inequidades sistémicas si no se evalúa rigurosamente su impacto diferenciado en diversos grupos sociales.

## **4. Inteligencia Artificial Generativa: nuevos desafíos éticos**

Las recientes IA generativas como DALL-E y Midjourney y los Modelos de Lenguaje de Gran Escala (LLMs) como GPT-4, Claude o Llama, han introducido una nueva dimensión de desafíos éticos que merecen atención especial. Estas tecnologías no son simples herramientas; representan una frontera donde la creatividad humana y la sintética comienzan a entrelazarse de maneras inéditas.

La **generación de contenido falso y desinformación** representa uno de los riesgos más inmediatos. Los modelos generativos actuales pueden crear contenido falso virtualmente indistinguible del contenido auténtico, lo que plantea la posibilidad de inundar plataformas con desinformación a escala, crear deepfakes que suplantan identidades en videos, audios o imágenes, y en última instancia erosionar la confianza social al dificultar la distinción entre lo real y lo sintético. El desafío para los programadores en este ámbito incluye implementar marcadores de contenido AI-generado (metadatos invisibles, marcas de agua digitales, etiquetas visibles, etc.), desarrollar sistemas robustos de detección de contenido sintético, y diseñar con salvaguardas que limiten los usos malintencionados sin restringir los beneficios legítimos de estas tecnologías.

Las cuestiones de **propiedad intelectual y creatividad** constituyen otro campo de debate intenso. Las IA generativas se entrenan con enormes corpus de contenido creativo humano, planteando preguntas complejas: ¿Deberían los creadores originales recibir compensación cuando su trabajo se usa para entrenar IAs? ¿Cómo podemos atribuir correctamente contenido generado que "aprende" de múltiples fuentes? ¿Cuándo el contenido AI-generado constituye una obra nueva y cuándo debe considerarse derivativa? El marco legal está en plena evolución, con diferentes jurisdicciones adoptando aproximaciones divergentes. En el caso puntual de Argentina, aún a kilómetros luz de contar con definiciones legales. Como programadores, deben mantenerse informados sobre estos desarrollos y participar activamente en el debate.

La **amplificación de patrones problemáticos** es otra preocupación central. Las IA generativas, entrenadas en datos de internet, pueden reproducir y amplificar patrones sociales problemáticos, incluyendo estereotipos dañinos sobre diversos grupos sociales, generación de discurso de odio o material ofensivo, y representaciones desequilibradas que subrepresentan a minorías o reproducen patrones de poder existentes. Los mecanismos de alineación, como el Aprendizaje por Refuerzo a partir de Retroalimentación Humana (RLHF), son estrategias técnicas para alinear estos sistemas con valores humanos, pero plantean a su vez preguntas sobre qué valores deben priorizarse y quién determina esta jerarquía.

## 5. El rol ético del programador en la era de la IA

En este contexto de rápida evolución tecnológica, el rol del programador trasciende lo puramente técnico para abarcar dimensiones éticas y sociales profundas. Son los arquitectos silenciosos del futuro social.

Los programadores ya no son simples implementadores técnicos, sino **arquitectos de sistemas sociotécnicos complejos** que moldean la realidad cotidiana. Sus decisiones definen qué aspectos de la experiencia humana son "medibles" y "optimizables", establecen los límites de lo posible y lo prohibido en entornos digitales, configuran interfaces que moldean comportamientos e interacciones sociales, e implementan valores y prioridades en código. Esta posición conlleva una responsabilidad sustancial que trasciende la mera funcionalidad técnica y los sitúa como agentes éticos de primer orden.

Para asumir esta responsabilidad, los programadores necesitan desarrollar competencias específicas. Necesitan comprender cómo los sistemas técnicos interactúan con las estructuras sociales y culturales, reconociendo que **la tecnología nunca opera en un**

**vacío sino dentro de contextos humanos complejos.** El **pensamiento consecuencialista** los capacita para evaluar rigurosamente los posibles impactos de sus creaciones, incluyendo escenarios de mal uso. La comprensión de justicia algorítmica les ayuda a entender cómo **los algoritmos pueden perpetuar o desafiar inequidades existentes.** Las habilidades de diseño participativo les permiten incluir perspectivas diversas en el proceso de desarrollo. Y la capacidad de reflexión ética aplicada resulta fundamental para identificar y resolver dilemas éticos en contextos prácticos de desarrollo.

Diversos marcos éticos han surgido para guiar el desarrollo responsable de sistemas de IA. Los principios de alineación buscan asegurar que los sistemas de IA actúen de acuerdo con las intenciones humanas, evitando comportamientos no deseados. La ética de la virtud aplicada a la IA enfatiza el desarrollo de sistemas que promuevan el florecimiento humano y valores como la autonomía y el bienestar. La perspectiva consecuencialista evalúa sistemas según sus resultados e impactos reales en las personas y sociedades. Y la ética deontológica ayuda a identificar límites y prohibiciones absolutas en el desarrollo tecnológico. Como profesionales, los programadores deben familiarizarse con estos marcos y aplicarlos conscientemente en su práctica profesional cotidiana.

## 6. Herramientas prácticas para el desarrollo ético de IA

Existen diversas herramientas prácticas que pueden implementar para garantizar un desarrollo ético de sistemas de IA. Las **evaluaciones de impacto algorítmico (AIA)** son procesos estructurados para identificar, evaluar y mitigar riesgos asociados con sistemas automatizados. Como desarrolladores pueden implementar evaluaciones prospectivas antes del despliegue del sistema, establecer sistemas de monitoreo continuo durante la operación, y realizar auditorías retrospectivas para analizar el impacto real una vez desplegado el sistema.

El **diseño centrado en valores (VSD)** constituye una metodología que coloca consideraciones éticas en el centro del proceso de desarrollo. Este enfoque comienza con la identificación de los valores implicados en el contexto de uso; continúa con un análisis conceptual de estos valores para comprender sus dimensiones y posibles tensiones entre ellos; avanza hacia una implementación técnica que respeta dichos valores en el diseño mismo; y concluye con una verificación empírica de que el sistema efectivamente se alinea con los valores identificados inicialmente.

La **transparencia algorítmica** resulta fundamental para la rendición de cuentas en sistemas automatizados. Herramientas como Model Cards (documentación estandarizada

sobre capacidades y limitaciones del modelo), Datasheets for Datasets (documentación detallada sobre conjuntos de datos utilizados), y declaraciones de conformidad ética proporcionan marcos estructurados para documentar aspectos clave de los sistemas de IA. Esta documentación no solo facilita la evaluación por parte de terceros, sino que también fomenta una reflexión sistemática por parte del equipo de desarrollo.

## **7. Marco regulatorio emergente**

El desarrollo de sistemas de IA está experimentando una creciente regulación global que los desarrolladores deben conocer. El Reglamento de IA de la Unión Europea representa el marco regulatorio más comprensivo hasta la fecha, con un enfoque basado en riesgo que clasifica las aplicaciones de IA según su potencial impacto y establece requisitos proporcionados. Los Lineamientos NIST AI Risk Management Framework en Estados Unidos ofrecen directrices más flexibles para la gestión de riesgos asociados con sistemas de IA. Adicionalmente, están emergiendo normativas sectoriales específicas en ámbitos como salud, finanzas y transporte, que imponen requisitos adicionales para aplicaciones de IA en estos contextos sensibles.

Como reflexionó el filósofo Hans Jonas, "actúa de tal manera que los efectos de tu acción sean compatibles con la permanencia de una vida humana auténtica". Este imperativo de responsabilidad adquiere nueva relevancia en la era digital, cuando las creaciones tecnológicas tienen el potencial de transformar radicalmente la experiencia humana, para bien o para mal.

Estas regulaciones apuntarían a transformar la IA de un "salvaje oeste" tecnológico a un campo con estándares profesionales y expectativas de conducta responsable cada vez más definidos. Como programadores, deben estar familiarizados con estos marcos regulatorios no solo para garantizar el cumplimiento legal, sino también porque ofrecen orientaciones valiosas para el desarrollo ético de sistemas.

## **8. Hacia una IA potenciadora de capacidades humanas**

Más allá de identificar problemas éticos, los programadores pueden contribuir activamente a construir un futuro donde la tecnología amplifique el potencial humano en lugar de reemplazarlo. La implementación de un paradigma de "IA aumentativa" frente a la "IA sustitutiva" representa un camino prometedor. Esto implica diseñar sistemas que amplíen la creatividad, intuición y juicio humanos, en vez de intentar replicarlos o sustituirlos. En la

práctica, esto significa privilegiar interfaces que mantengan a las personas informadas y empoderadas; desarrollar herramientas que reduzcan tareas tediosas pero preserven la toma de decisiones significativas para los humanos; y adoptar metodologías de co-creación donde usuarios y desarrolladores colaboren continuamente. Ejemplos concretos incluyen sistemas médicos que mejoran la capacidad diagnóstica de profesionales sin sustituir su criterio clínico, herramientas creativas que expanden el repertorio expresivo de artistas sin homogeneizar estilos, y plataformas educativas que personalicen experiencias de aprendizaje manteniendo al docente como guía fundamental. Este enfoque requiere humildad técnica: reconocer que muchas capacidades humanas como la empatía contextual, el juicio ético situado o la creatividad disruptiva siguen siendo difíciles de formalizar algorítmicamente, y que el mayor valor social puede generarse cuando las máquinas complementan estas facultades en lugar de intentar reemplazarlas.

## 9. Conclusión: el imperativo ético en la era de la IA

La IA está transformando profundamente nuestra sociedad, y los programadores son protagonistas de esta transformación, no meros espectadores. Su responsabilidad va más allá de la implementación técnica; incluye considerar activamente cómo sus creaciones impactan en la vida de las personas, especialmente las más vulnerables. Cada línea de código es una decisión ética, cada sistema que implementan reconfigura sutilmente el tejido social.

El desarrollo ético de IA no es solo una cuestión de cumplimiento normativo, sino de **integridad profesional** y **responsabilidad social**. Implica un compromiso con la ampliación, no reducción, de la autonomía humana; la igualdad de acceso y beneficios de la tecnología; la protección de valores fundamentales como privacidad y dignidad; y la rendición de cuentas por los sistemas creados.

Como profesionales informáticos, los programadores tienen la oportunidad y la responsabilidad de asegurar que la inteligencia artificial se desarrolle como una tecnología verdaderamente al servicio del bienestar humano y social. Esto requiere no solo competencia técnica, sino también una comprensión profunda de las dimensiones éticas, sociales y políticas de su trabajo. En última instancia, se trata de reconocer que cada línea de código que escriben puede tener implicaciones significativas para la vida de las personas y para el tejido social en su conjunto.



## Fuentes de consulta adicionales

### Artículos de interés:

- Fundar. "Guía práctica para el desarrollo ético de sistemas basados en inteligencia artificial". Disponible en:  
<https://fund.ar/publicacion/guia-practica-para-el-desarrollo-etico/>
- Angwin, Julia, Larson, Jeff, Mattu, Surya y Kirchner, Lauren. (2016). "Machine Bias". ProPublica. Disponible en:  
<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- Amnistía Internacional España. "5 ejemplos de malos usos de la inteligencia artificial que afectan directamente a la gente joven". Disponible en:  
<https://www.es.amnesty.org/en-que-estamos/blog/historia/articulo/5-ejemplos-de-malos-usos-de-la-inteligencia-artificial-que-afectan-directamente-a-gente-joven/>
- "¿Hemos creado un monstruo? Cómo influyen los sesgos humanos en la IA". CyberGhost VPN. 2023. Disponible en:  
<https://www.cyberghostvpn.com/es/privacyhub/human-biases-ai/>
- "Uso del algoritmo COMPAS en el proceso penal y los riesgos a los derechos humanos". SciELO Brasil. 2023. Disponible en:  
<https://www.scielo.br/j/rbdpp/a/6W9b8CHYbXcsc6qczDxCsfr/>
- UNESCO. "Recomendación sobre la ética de la inteligencia artificial". UNESCO, 2021. Disponible en:  
<https://www.unesco.org/es/artificial-intelligence/recommendation-ethics>

### Películas y series de ficción:

- **"Ex Machina" (2015)**. Disponible en: Max
- **"Her / Ella" (2013)**. Disponible en: Prime Video
- **"Better than us / Mejores que nosotros" (Better Than Us) (Serie de TV, 2018)**. Disponible en: Netflix.
- **"Be Right Back / Enseguida vuelvo" (Black Mirror, Temporada 2, Episodio 1, 2013)**. Disponible en: Netflix.
- **"Nosedive / Caída en picada" (Black Mirror, Temporada 3, Episodio 1, 2016)**. Disponible en Netflix.

### **Documentales y Audiovisuales:**

- **"Coded Bias / Prejuicio cifrado" (2020).** Disponible en: Netflix.
- **"Vigilancia" (Conexiones / Connected, Episodio 1, 2020).** Disponible en: Netflix.

### **Sitios de Interés y Organizaciones:**

- **AI Now Institute (Universidad de Nueva York).** Disponible en: <https://ainowinstitute.org/>
- **Partnership on AI.** Disponible en: <https://www.partnershiponai.org/>
- **UNESCO - Ética de la inteligencia artificial.** Disponible en: <https://www.unesco.org/es/artificial-intelligence/recommendation-ethics>
- **OdiselA: Observatorio de Impacto Social y Ético de la Inteligencia Artificial.** Disponible en: <https://www.odiseia.org/>