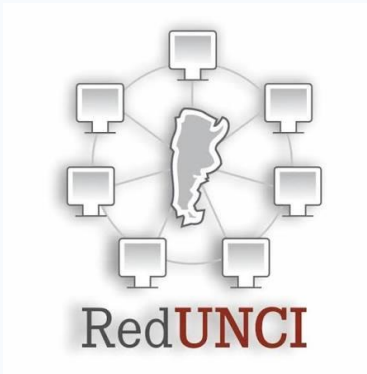
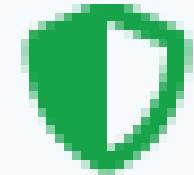
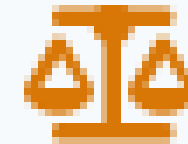




Uso educativo de la IA



Criterios éticos para alumnos, docentes, evaluadores y agentes de IA



Javier Díaz

Universidad Nacional de La Plata



UNIVERSIDAD
NACIONAL
DE LA PLATA



La pregunta institucional

La discusión ya no puede reducirse a permitir o prohibir herramientas.



Qué cambia

La IA generativa ya no solo produce textos o código. Con agentes, también planifica, usa herramientas y ejecuta acciones.



Qué debe protegerse

Autonomía intelectual, privacidad, permisos, trazabilidad, validez de la evaluación y responsabilidad profesional.



Qué decisión necesita cada carrera

Un criterio común con reglas por asignatura, tipo de tarea, nivel de autonomía y forma de evaluación.

Punto de partida argentino

La presentación adjunta resume cuatro garantías para una IA centrada en las personas.

Supervisión humana

La decisión final y la rendición de cuentas deben quedar en manos de personas calificadas responsables.

Equidad y sesgos

Las carreras deben revisar impactos diferenciales sobre estudiantes, trayectorias y perfiles diversos.

Transparencia

La comunidad debe conocer cuándo se usa IA, con qué finalidad y con qué límites.

Seguridad y opción

La adopción tecnológica debe contemplar riesgos, privacidad y alternativas razonables.

Argentina 2023

Recomendaciones para el uso de Inteligencia Artificial, Disposición 2/2023 de Jefatura de Gabinete.

Objetivo

Proteger derechos fundamentales, prevenir riesgos, promover innovación pública y sostener el diseño centrado en las personas.

Principios centrales

Supervisión y decisión humanas, transparencia, explicabilidad, rendición de cuentas, equidad, seguridad y privacidad.

Para universidades

Base nacional para definir usos permitidos, declaraciones de uso, resguardo de datos y responsabilidad institucional.

SIGEN 2026

Guía de Controles Inteligencia Artificial, Resolución SIGEN 197/2026.



Marco de control

Propone un marco de referencia de control interno para implementar IA en el Sector Público Nacional.



Riesgo y rendición

Orienta controles proporcionales para fortalecer transparencia, rendición de cuentas, calidad de servicios y gestión institucional.



Para educación pública

Aporta criterios para gobernanza, auditoría, trazabilidad y control humano cuando la IA entra en procesos institucionales.

UNESCO 2021

Recomendación sobre la Ética de la Inteligencia Artificial, adoptada por los Estados Miembros de UNESCO.

Qué aporta

Un marco global basado en derechos humanos, dignidad, diversidad, sostenibilidad y gobernanza multiactor.

Principios útiles

Proporcionalidad, inocuidad, seguridad, equidad, privacidad, supervisión humana, transparencia y rendición de cuentas.

Para carreras

Permite convertir la ética en criterios de uso, declaración, protección de datos y control humano en cada asignatura.

OECD 2019 y actualización 2024

Principios para una IA confiable, innovadora y respetuosa de los derechos humanos.



Qué aporta

Un estándar intergubernamental práctico para orientar políticas públicas, instituciones y actores que desarrollan o usan IA.



Valores centrales

Crecimiento inclusivo, centralidad humana, equidad, transparencia, robustez, seguridad y rendición de cuentas.



Para carreras

Ayuda a justificar por qué cada uso educativo de IA debe promover aprendizaje, autonomía y bienestar verificable.

Unión Europea 2019 y AI Act 2024

Guías éticas para IA confiable y regulación basada en riesgo.

Guías 2019

El Grupo de Alto Nivel propuso siete requisitos: agencia humana, seguridad, privacidad, transparencia, equidad, bienestar y responsabilidad.

AI Act 2024

Los usos autónomos en admisión, evaluación o monitoreo educativo elevan el riesgo: decisiones intermedias opacas y bucles difíciles de auditar.

Para carreras

Exige revisar evaluación, analítica educativa, proctoring y agentes con criterios de riesgo, supervisión humana y protección de derechos.

IEEE 2019 y IEEE 7000-2021

De los valores éticos al diseño responsable de sistemas.

Ethically Aligned Design 2019

Propone alinear sistemas autónomos e inteligentes con derechos humanos, bienestar, transparencia y responsabilidad.

IEEE 7000-2021

Ayuda a convertir valores éticos en requisitos verificables: límites de acción, permisos, registro de eventos y pruebas antes de usar agentes.

Para carreras técnicas

Conecta ética, diseño, validación, contención técnica, trazabilidad y competencia profesional en proyectos con IA.

ACM 2018

Código de Ética y Conducta Profesional para profesionales de la computación.

Principio rector

El bien público debe guiar el trabajo profesional y académico cuando se diseña, enseña o evalúa tecnología.

Deberes relevantes

Evitar daño exige limitar permisos, proteger datos, revisar efectos secundarios, auditar acciones y sostener competencia profesional.

Para educación superior

Sirve como puente entre ética profesional, integridad académica y responsabilidad de quienes despliegan u orquestan agentes.

Principios comunes

La coincidencia entre marcos permite construir una política educativa práctica.

Control humano

La IA puede asistir y ejecutar pasos, pero las decisiones académicas finales requieren revisión humana responsable.

Trazabilidad

El uso de IA debe dejar rastros comprensibles: finalidad, herramienta, datos, acciones y verificaciones.

Datos y permisos

No se deben cargar datos sensibles ni entregar credenciales, tokens o accesos reales a herramientas sin garantías.

Equidad e inclusión

Las reglas deben evitar ventajas injustas, barreras de acceso de todo tipo y sesgos en materiales o evaluaciones.

Criterios éticos para IA agéntica

La autonomía del agente exige controles antes, durante y después de la ejecución.

Agencia y control

Los agentes no deben ejecutar decisiones de alto impacto sin una validación humana explícita e intermedia.

Contención

Definir permisos mínimos, límites de tiempo y costo, datos permitidos y entornos de prueba aislados.

Trazabilidad

Registrar herramientas habilitadas, cadena de razonamiento, planes, llamadas a APIs, cambios realizados y resultados observables.

Rendición de cuentas

La responsabilidad académica, técnica y legal recae en la persona que diseña, despliega u orquesta el agente.

Del principio a la regla de curso

Cada asignatura necesita explicitar usos permitidos, usos restringidos y evidencias esperadas, también para agentes.

Usos permitidos

Búsqueda inicial, explicación alternativa, revisión de estilo, apoyo para depurar código o agentes en entornos de prueba.

Usos restringidos

Agentes con credenciales reales, acciones sobre sistemas productivos, generación de datos sin trazabilidad o sustitución de prácticas evaluativas.

Evidencias mínimas

Declaración de uso, herramientas habilitadas, prompts relevantes, logs, validación de fuentes e intervención humana.

Mirada del alumno

La regla central: supervisión del agente, no delegación ciega.



Aprender con IA

Usarla para ampliar preguntas, recibir retroalimentación, comparar enfoques y probar hipótesis bajo revisión.



Orquestrar con criterio

Revisar el plan del agente, limitar herramientas, frenar ejecuciones y detectar deriva de objetivos.



Declarar el proceso

Indicar herramienta, propósito, permisos, acciones ejecutadas, logs disponibles, fuentes verificadas y cambios realizados.

Mirada del docente

La ética empieza en el diseño de la actividad y en el perímetro de autonomía permitido.



Intencionalidad

Usar IA solo cuando mejore el aprendizaje, la accesibilidad, la retroalimentación o la exploración de alternativas.



Política explícita

Informar qué usos se aceptan en cada curso y entrega, incluyendo agentes, ejecución de código, APIs y automatizaciones.



Entornos seguros

Usar sandboxes cuando los alumnos diseñen agentes que ejecuten comandos, código o flujos multipaso.



Cuidado de datos y soberanía

Evitar subir datos sensibles o entregar credenciales, tokens y accesos institucionales a servicios sin garantías.

Mirada del evaluador

La evaluación debe medir aprendizaje verificable y capacidad de auditar el proceso.

Validez

La consigna debe permitir comprobar comprensión, criterio, transferencia y capacidad de defensa del trabajo.

Auditoría del agente

La consigna debe permitir revisar restricciones, manejo de errores, guardrails y efectos secundarios no deseados.

Trazabilidad

El proceso pesa: bitácora, prompts del sistema, herramientas, permisos, logs, decisiones y correcciones realizadas.

Responsabilidad

La IA puede asistir al análisis, pero la decisión evaluativa y la rendición de cuentas quedan a cargo de personas.

Riesgos educativos críticos

El problema no es la herramienta aislada, sino la dependencia y la autonomía operativa que instala.

Delegación operativa

El estudiante no puede justificar fallas diciendo que el agente decidió el flujo. Debe responder por el diseño y la supervisión. **Opacidad acumulativa:** cuanto más largo es el flujo de acciones del agente, más difícil se vuelve identificar en qué paso exacto se introdujo un sesgo o un error.

Deriva de objetivos

El agente puede interpretar mal la meta inicial y encadenar acciones incorrectas si nadie controla los pasos intermedios.

Inyección de prompts

Un agente que lee web, documentos o correos puede obedecer instrucciones maliciosas ocultas en fuentes externas.

Sesgos y acceso

Las herramientas pueden amplificar estereotipos o generar ventajas para quienes tienen mejores recursos.

Evaluaciones más robustas

La respuesta no debería descansar en detectores de IA.

Evaluar orquestación

Bitácoras, entregas parciales, revisión de prompts del sistema, herramientas habilitadas y comparación entre versiones.

Conversar el resultado

Instancias orales o presenciales para que el estudiante explique criterios, errores, acciones del agente y cambios.

Situar la tarea

Problemas contextualizados, datos propios, decisiones justificadas y agentes limitados a entornos de prueba cuando corresponda.

Rúbricas claras

Criterios que separen contenido, proceso, uso de IA, guardrails, verificación y responsabilidad final.

Declaración de uso de IA y agentes

Metacognición y trazabilidad: el estudiante debe explicar qué delegó, qué supervisó y qué aprendió.

Herramienta y permisos

Nombre, versión, fecha, tipo de asistencia, herramientas habilitadas y permisos otorgados.

Partes asistidas

Qué se generó, revisó, tradujo, resumió, programó, corrigió o ejecutó con ayuda de IA.

Evidencia agéntica

Prompts principales, plan del agente, logs de ejecución, llamadas a herramientas y cambios producidos.

Verificación humana

Puntos de intervención, fuentes contrastadas, errores detectados y decisiones finales tomadas por el estudiante.

Hoja de ruta para carreras

El objetivo es formar criterio profesional, no solo controlar conductas.

Acuerdo común

Definir principios de carrera: transparencia, protección de datos, supervisión humana, permisos y equidad.

Capacitación

Formar a docentes, alumnos y evaluadores en alfabetización crítica de IA, gobernanza de agentes, sesgos, privacidad y verificación.

Reglas por asignatura

Traducir el acuerdo a consignas, rúbricas, declaraciones de uso, ejemplos permitidos y límites para agentes.

Revisión continua

Actualizar criterios frente a nuevas herramientas, evidencia pedagógica, incidentes y cambios regulatorios.

Cierre

La universidad debe enseñar a usar IA con criterio, evidencia y responsabilidad.



Para el alumno

La IA amplía el aprendizaje cuando deja rastros, exige verificación y no reemplaza el esfuerzo central ni la responsabilidad.



Para el docente

La política de IA forma parte del diseño pedagógico e incluye límites claros para agentes y automatizaciones.



Para el evaluador

La decisión académica requiere trazabilidad, justicia, auditoría del proceso y supervisión humana.

Fuentes para profundizar

UNESCO: Recommendation on the Ethics of Artificial Intelligence |
OECD: AI Principles, adoptados en 2019 y actualizados en 2024 |
Comisión Europea: Ethics Guidelines for Trustworthy AI y AI Act |
IEEE: Ethically Aligned Design e IEEE 7000-2021 |
ACM: Code of Ethics and Professional Conduct, 2018 |
Argentina: Recomendaciones para el uso de IA, Disposición 2/2023 |
SIGEN: Guía de Controles IA, Resolución 197/2026 |
MIT: AI and Education, principio de aumentación |
Harvard: Generative AI Guidelines |
ETH Zurich: AI in Teaching and Learning |
UNAM: guía para estudiantes sobre IA generativa |
OWASP: Top 10 for LLMs and Gen AI Apps 2025 |
NIST: AI 600-1, Generative AI Profile