

INTRODUCCIÓN A LA IA GENERATIVA

Material introductorio

Objetivos de la clase

Al terminar esta clase, vas a poder:

- explicar qué es la IA generativa y en qué se diferencia de otros usos de la inteligencia artificial;
- reconocer por qué su uso plantea cuestiones éticas, y describir tres ejes para analizarlas: transparencia, responsabilidad y equidad;
- identificar los principales riesgos asociados a la información que se introduce en estas herramientas y al contenido que producen;
- aplicar criterios básicos de privacidad, propiedad intelectual y verificación en situaciones concretas;
- distinguir qué tareas pueden delegarse en una herramienta generativa y cuáles requieren siempre el juicio de una persona.

1. ¿Qué es la IA generativa?

De analizar a crear

En la clase anterior vimos que la inteligencia artificial aprende patrones a partir de datos y los usa para realizar tareas que normalmente requieren inteligencia humana. Muchas de esas tareas consisten en **analizar** información que ya existe: clasificar un correo como spam, estimar la demanda de un producto o reconocer un rostro en una foto.

La IA generativa da un paso distinto: en lugar de solo analizar o clasificar, **produce contenido nuevo**.

IA generativa: tipo de inteligencia artificial que, a partir de los patrones aprendidos en grandes conjuntos de datos, produce contenido nuevo —texto, imágenes, audio, video o código— en respuesta a una indicación de la persona usuaria.

El Instituto Nacional de Estándares y Tecnología de los Estados Unidos (NIST) la describe, en términos similares, como una clase de modelos que imitan la estructura y las características de los datos con los que fueron entrenados para generar contenido sintético derivado de ellos (NIST AI 600-1, 2024).

Dos ideas de esta definición son importantes para lo que sigue:

- **El contenido es nuevo, pero no surge de la nada.** Se construye a partir de patrones aprendidos. Por eso se parece a los ejemplos del entrenamiento sin ser, en principio, una copia de ellos. Como veremos, esa relación con los datos de origen es la fuente de varios de los problemas éticos de esta tecnología.
- **Responde a una indicación.** La persona escribe un pedido —por ejemplo, "redactá un correo para reprogramar una reunión"— y la herramienta genera una respuesta. Esa indicación suele llamarse *prompt*. Cómo formular buenas indicaciones es un tema propio, que se trabajará en módulos posteriores.

Qué puede generar

Tipo de contenido	Ejemplos cotidianos	Ejemplos profesionales
Texto	Redactar un mensaje, resumir un artículo, traducir una receta	Preparar el borrador de un informe, responder consultas frecuentes, adaptar un texto a distintos públicos
Imágenes	Crear una ilustración para una invitación	Producir bocetos para una campaña, generar variantes de un diseño

Tipo de contenido	Ejemplos cotidianos	Ejemplos profesionales
Audio	Convertir un texto en voz para escucharlo	Locutar un video institucional, crear música de fondo
Video	Animar una foto	Producir piezas breves de capacitación o de comunicación
Código	Crear una fórmula para una planilla	Sugerir funciones, explicar un error, documentar un programa

Analizar y generar: una comparación

Imaginemos una tienda en línea que recibe miles de opiniones de clientes.

- Un sistema de IA **no generativo** puede leerlas y clasificarlas como positivas o negativas, o detectar cuáles mencionan demoras en la entrega. El resultado es una etiqueta, un número o una categoría.
- Un sistema de IA **generativo** puede, a partir de esas opiniones, redactar un resumen de los principales reclamos o escribir un borrador de respuesta para cada cliente. El resultado es un contenido nuevo que antes no existía.

En la práctica, ambos usos suelen combinarse: primero se analiza la información y después se genera algo a partir de ella.

Por qué importa

La IA generativa se difundió con enorme rapidez porque se usa con lenguaje cotidiano, sin conocimientos técnicos, y porque puede ahorrar mucho tiempo en tareas de redacción, diseño o búsqueda de ideas. Esa misma facilidad es la que obliga a pensar con cuidado cómo se usa: una herramienta que produce contenido convincente en segundos también puede producir errores, prejuicios o problemas legales a la misma velocidad.

Idea clave: la IA generativa no busca una respuesta guardada; construye un contenido nuevo a partir de patrones aprendidos. Eso explica tanto su utilidad como sus riesgos.

Actividad 1: ¿generativa o no?

Indicá si en cada caso se trata de un uso generativo de la IA o de otro tipo de uso.

1. Una aplicación bancaria marca una compra como posible fraude.
2. Un asistente redacta tres opciones de título para una nota de prensa.
3. Una plataforma de música sugiere canciones parecidas a las que escuchás.
4. Una herramienta crea una imagen de un aula del futuro a partir de una descripción.
5. Un sistema detecta patentes de autos en las cámaras de un estacionamiento.
6. Un programa convierte el guion de un curso en una locución con voz sintética.

Respuestas orientativas: son generativos 2, 4 y 6. Los casos 1, 3 y 5 analizan, clasifican o recomiendan a partir de información existente, pero no crean contenido nuevo.

2. ¿Por qué la IA generativa plantea cuestiones éticas?

Beneficios y riesgos al mismo tiempo

La IA generativa puede aportar mucho: ayuda a superar la página en blanco, acelera tareas repetitivas, facilita la traducción y la accesibilidad, y permite explorar muchas alternativas en poco

tiempo. Pero, por la forma en que funciona, también puede producir resultados problemáticos. Entre los más habituales:

- **Información incorrecta.** Un texto fluido y seguro que contiene datos falsos.
- **Contenido dañino.** Lenguaje ofensivo, discriminatorio o que puede causar perjuicios.
- **Sesgos.** Resultados que favorecen o perjudican de forma injusta a ciertos grupos.
- **Problemas de privacidad.** Datos personales o confidenciales que se exponen o se usan sin control.
- **Conflictos de propiedad intelectual.** Contenidos que reproducen o se parecen demasiado a obras protegidas.
- **Resultados difíciles de comprender.** Recomendaciones o respuestas cuyo origen no puede explicarse con claridad.

Estos problemas no suelen ser intencionales. Surgen de los datos con los que se entrenó la herramienta, de cómo fue diseñada o de cómo se la utiliza. Justamente por eso requieren atención deliberada.

Qué es la ética de la IA

La ética, en sentido general, se ocupa de los criterios que nos permiten distinguir entre lo correcto y lo incorrecto al actuar. Cuando esas acciones se apoyan en sistemas de IA, aparecen preguntas nuevas: ¿quién responde si la herramienta se equivoca?, ¿cómo sabemos si trata a todas las personas por igual?, ¿con qué datos se entrenó?

Ética de la IA: campo multidisciplinario que estudia cómo aprovechar al máximo los beneficios de la inteligencia artificial y, al mismo tiempo, reducir sus riesgos y sus efectos adversos.

Es **multidisciplinario** porque ninguna disciplina puede resolver estas cuestiones por sí sola: intervienen la informática, el derecho, la filosofía, la sociología, la gestión de organizaciones y el conocimiento propio de cada área de aplicación, como la salud o la educación. Y abarca **todo el ciclo de vida** de un sistema: su diseño, su entrenamiento, su puesta en funcionamiento y su uso cotidiano.

Tres ejes para analizar

Distintos organismos internacionales han propuesto principios para orientar una IA confiable. La Recomendación sobre la Ética de la Inteligencia Artificial de la UNESCO (2021), los Principios de IA de la OCDE (2019, actualizados en 2024) y el Marco de Gestión de Riesgos de IA del NIST (2023) coinciden, con matices, en ideas como la transparencia, la rendición de cuentas, la equidad, la privacidad, la seguridad y la supervisión humana.

En esta clase vamos a organizar el análisis en torno a tres ejes centrales, que permiten ordenar los demás temas:

Eje	Pregunta que orienta
Transparencia	¿Podemos saber que estamos usando IA, qué puede hacer y cuáles son sus límites?
Responsabilidad	¿Quién responde por los resultados y quién los revisa?
Equidad	¿La herramienta trata de manera justa a todas las personas y grupos?

Esta selección es una herramienta didáctica, no una lista cerrada. La privacidad, la propiedad intelectual y la prevención de contenido dañino, que veremos más adelante, se relacionan con los tres ejes a la vez.

Idea clave: los riesgos de la IA generativa no son razones para no usarla, sino razones para usarla con criterio.

Actividad 2: identificar el problema

Una organización cultural usa una herramienta generativa para redactar la biografía de una artista local que va a exponer en su sala. El texto queda muy bien escrito y se publica en la web. Días después, la artista avisa que la biografía le atribuye premios que nunca recibió y una formación académica que no tiene.

¿Qué problema ético aparece? ¿Con cuál de los tres ejes se relaciona más directamente? ¿Qué debería haberse hecho antes de publicar?

Orientación: aparece información incorrecta presentada como cierta. Se relaciona sobre todo con la responsabilidad, porque nadie verificó el texto antes de publicarlo; también con la transparencia, si la organización no sabía que la herramienta podía inventar datos.

3. Transparencia

Qué significa

Transparencia en IA: posibilidad de acceder a información suficiente para comprender cómo fue creado un sistema de IA, qué puede y qué no puede hacer, y cómo llega a sus resultados.

La transparencia no significa que cualquier persona deba entender los detalles matemáticos de un modelo. Significa que cada persona debería contar con la información que necesita, según su rol, para usar la herramienta con criterio o para evaluar una decisión que la afecta. En la práctica, abarca varios niveles.

Saber que se está interactuando con IA. Una persona debería poder saber cuándo está interactuando con un sistema de IA y, cuando resulte relevante, si una imagen, un audio o un video fueron generados artificialmente. Esto importa porque cambia la forma en que interpretamos lo que recibimos: no evaluamos igual la respuesta de un profesional que la de un sistema que genera texto probable. Además, algunas jurisdicciones establecen obligaciones legales específicas de información; por ejemplo, el Reglamento de Inteligencia Artificial de la Unión Europea (2024) exige informar en determinados casos que se interactúa con un sistema de IA y señalar ciertos contenidos generados artificialmente. Las obligaciones concretas varían según el país.

Conocer capacidades y limitaciones. Quien usa una herramienta debería saber para qué fue diseñada, en qué tareas funciona bien y en cuáles puede fallar. Por ejemplo, que su información llega hasta cierta fecha, que puede inventar datos o que no fue pensada para dar consejo médico.

Conocer el origen de los datos, cuando sea posible y relevante. Saber, al menos en términos generales, con qué tipo de información se entrenó un sistema ayuda a anticipar sus puntos débiles. Una herramienta entrenada con textos de un solo país o de un solo idioma probablemente funcione peor fuera de ese contexto.

Poder comprender por qué se produjo un resultado. Esto suele llamarse **explicabilidad**. En tareas de alto impacto, no alcanza con recibir una recomendación: hace falta saber en qué se basó.

Transparencia y confianza

La confianza en una herramienta debería ser proporcional a lo que sabemos de ella. Si desconocemos sus límites, tendemos a confiar demasiado o demasiado poco. Ambos extremos son problemáticos: el exceso de confianza lleva a aceptar errores sin revisarlos; la desconfianza total lleva a desaprovechar una herramienta útil.

La transparencia es especialmente importante cuando la IA interviene en decisiones que afectan la vida de las personas:

- **Salud.** Si un sistema sugiere priorizar a ciertos pacientes en una lista de espera, el equipo médico necesita saber qué factores consideró para poder evaluarlo y, eventualmente, corregirlo.

- **Finanzas.** Si una persona recibe el rechazo de un crédito en el que intervino un sistema automático, es razonable que pueda conocer los motivos principales y pedir una revisión.
- **Educación.** Si una plataforma recomienda que un estudiante repita un curso o tome un apoyo adicional, docentes y familias necesitan entender en qué se basó esa sugerencia.
- **Recursos humanos.** Si una herramienta ordena candidaturas para un puesto, quienes seleccionan deben saber qué criterios aplica, porque un error puede dejar afuera a personas calificadas.

En todos estos casos, un resultado que no puede explicarse tampoco puede discutirse ni corregirse.

Los límites de la transparencia

Conviene ser realistas: los sistemas de IA generativa actuales son muy complejos, y ni siquiera quienes los desarrollan pueden explicar en detalle cada resultado. Además, hay información que las empresas no publican por razones comerciales o de seguridad. Por eso, la transparencia se entiende como un objetivo y una práctica —documentar, informar, advertir limitaciones— más que como una condición absoluta. Cuando una herramienta no puede explicar sus resultados, su uso en decisiones importantes exige más revisión humana, no menos.

Idea clave: una herramienta es más confiable cuando sabemos qué es, qué puede hacer, qué no puede hacer y en qué se basa.

Actividad 3: comparar dos situaciones

Situación A. El sitio de un municipio ofrece un asistente virtual para consultas sobre trámites. Al abrirlo aparece el mensaje: "Soy un asistente automático basado en IA. Puedo orientarte sobre requisitos generales, pero la información oficial está en la página de cada trámite. Para casos particulares, podés comunicarte con una persona del área".

Situación B. Otro municipio ofrece un servicio similar, con el nombre y la foto de una supuesta empleada, "Laura". No aclara que se trata de un sistema automático ni menciona sus limitaciones.

¿Cuál es más transparente? ¿Qué riesgos concretos podría generar la situación B?

Orientación: la situación A informa que se trata de IA, delimita su alcance y ofrece una vía de contacto humano. En la B, las personas pueden creer que hablan con alguien del municipio y tomar sus respuestas como información oficial y definitiva.

4. Responsabilidad y supervisión humana

La IA no elimina la responsabilidad

Cuando algo sale mal con una herramienta de IA, es tentador decir "fue el sistema". Pero un sistema no puede responder por sus actos: no tiene intenciones ni obligaciones. Detrás de cada herramienta hay personas y organizaciones que tomaron decisiones.

Rendición de cuentas (responsabilidad) en IA: principio según el cual las personas y organizaciones que desarrollan, implementan o utilizan sistemas de IA responden por sus resultados, incluidas las consecuencias no previstas.

Esas decisiones humanas están presentes en todo el recorrido:

- **quienes desarrollan** eligen los datos de entrenamiento, definen qué se considera un buen resultado y deciden qué controles incorporar;
- **quienes implementan** deciden para qué se usará la herramienta, en qué procesos y con qué supervisión;
- **quienes usan** deciden qué información introducen, qué resultados aceptan y cómo los aplican.

Un sistema que parece objetivo y automático sigue siendo el resultado de muchas decisiones humanas. Por eso la responsabilidad se distribuye entre distintos actores, cada uno en su nivel.

No existe todavía una legislación universal sobre IA. Distintos países y regiones están desarrollando normas propias, con enfoques diferentes. Pero, más allá de lo que exija cada ley, el principio ético es el mismo: alguien tiene que poder responder por lo que la herramienta produce.

Supervisión humana y verificación

La forma más directa de ejercer la responsabilidad es la **supervisión humana**: que una persona con conocimiento del tema revise los resultados antes de que tengan efectos. Esta idea suele llamarse "humano en el circuito" (*human-in-the-loop*).

La intensidad de la supervisión debería ajustarse al riesgo. No requiere el mismo control una lista de ideas para el nombre de un evento que un informe técnico, un contrato o una comunicación a pacientes. Cuanto mayores sean las consecuencias de un error, más cuidadosa debe ser la revisión.

La responsabilidad también incluye el seguimiento posterior. Una organización que pone en marcha un asistente automático debería revisar periódicamente sus respuestas, recoger los reclamos de quienes lo usan y estar preparada para corregirlo, limitarlo o suspenderlo si produce resultados inadecuados.

Tres riesgos a tener en cuenta

Contenido tóxico. Las herramientas pueden producir, aun sin que se lo pidan, lenguaje de odio, insultos, obscenidades o mensajes hostigadores. En un servicio de atención al público, una sola respuesta de este tipo puede dañar a una persona y a la reputación de la organización.

Información falsa o engañosa. La herramienta puede generar contenidos falsos que parecen verdaderos. Si se difunden sin revisión, las personas pueden tomarlos como ciertos y basar sus decisiones en ellos. El problema no es solo el error puntual, sino la pérdida de capacidad de las personas para decidir con información correcta.

Divulgación de información confidencial. Si un sistema fue entrenado o alimentado con información reservada sin los resguardos adecuados, puede reproducirla en sus respuestas. Es una forma de fuga de datos, que retomaremos en la sección de privacidad.

Los límites de la automatización

Automatizar una tarea no significa automatizar la responsabilidad. Hay decisiones que no deberían delegarse por completo en una herramienta, especialmente las que afectan derechos, la salud, el empleo, la situación económica o la seguridad de las personas. En esos casos, la IA puede ofrecer información, borradores u opciones, pero la decisión final debe quedar en manos de alguien que pueda justificarla y responder por ella.

Idea clave: un resultado generado por IA es un insumo de trabajo, no un producto terminado. Quien lo utiliza es responsable de revisarlo.

Actividad 4: repartir tareas

Una estudiante que trabaja en un estudio contable usa una herramienta generativa para preparar un correo a un cliente explicándole los vencimientos impositivos del mes y el monto estimado que deberá pagar.

Distinguí:

- ¿Qué parte del trabajo puede hacer razonablemente la herramienta?
- ¿Qué debe revisar la estudiante antes de enviar el correo?
- ¿Qué decisión no debería delegarse?

Orientación: la herramienta puede proponer la estructura y la redacción del mensaje. La persona debe verificar fechas, montos, normativa vigente y datos del cliente. La decisión sobre qué informar y el envío final son responsabilidad del estudio.

5. Equidad, sesgos y representación

Qué es un sesgo

En lenguaje cotidiano, un sesgo es una inclinación que distorsiona un juicio: una tendencia a favorecer o desfavorecer algo sin una razón justificada. Las personas tenemos sesgos, muchas veces sin darnos cuenta. Como los sistemas de IA aprenden de datos producidos por personas y son diseñados por personas, pueden incorporar esos sesgos, reproducirlos e incluso amplificarlos.

Equidad en IA: objetivo de que un sistema trate de manera justa a las personas y grupos, minimizando los sesgos y favoreciendo una representación inclusiva.

Dónde puede aparecer el sesgo

Para esta introducción es útil distinguir tres momentos en los que puede aparecer un sesgo:

En los datos. Si los ejemplos con los que se entrena una herramienta reflejan desigualdades históricas o representan mal a ciertos grupos, la herramienta aprende esa distorsión. Un sistema de reconocimiento de voz entrenado mayormente con grabaciones de un acento determinado, por ejemplo, puede funcionar peor con personas de otras regiones.

En las decisiones. Cuando la herramienta participa en decisiones —aprobar, rechazar, priorizar— el sesgo puede traducirse en ventajas injustas para un grupo y desventajas para otro. Un sistema que evalúa solicitudes de crédito a partir del historial de aprobaciones de los últimos veinte años puede repetir criterios que en su momento ya eran injustos.

En los resultados generados. El contenido que produce una herramienta generativa puede representar a ciertos grupos de forma estereotipada. Si al pedir imágenes de "una persona que dirige una empresa" o "una persona que cuida niños" la herramienta devuelve de forma sistemática personas de un mismo género o de un mismo origen, está reforzando estereotipos, aunque nadie lo haya pedido.

Esta distinción es una forma didáctica de ordenar el problema, no una clasificación única. En la literatura especializada existen otras; el NIST, por ejemplo, agrupa los sesgos en sistémicos, estadísticos o computacionales, y humanos (NIST SP 1270, 2022). Lo importante es comprender que el sesgo puede ingresar en cualquier etapa: al elegir los datos, al procesarlos, al diseñar el sistema y al interpretar sus resultados.

Por qué los datos históricos pueden producir injusticias

Una herramienta de IA no sabe si un patrón es justo o injusto: solo detecta que se repite. Si durante años ciertos grupos tuvieron menos acceso a un tipo de empleo, a la educación universitaria o al crédito, los datos históricos mostrarán ese patrón, y una herramienta entrenada con ellos puede concluir que ese es el comportamiento "normal". Así, una desigualdad del pasado se convierte en una regla automática del presente.

El problema se agrava porque estos resultados parecen neutrales: provienen de una máquina y se expresan en números o recomendaciones. Esa apariencia de objetividad puede hacer que el sesgo pase inadvertido.

Buenas prácticas

No hay una fórmula que elimine todos los sesgos, pero existen prácticas que ayudan a reducirlos:

- **Representación diversa.** Verificar que los datos incluyan de manera adecuada a los distintos grupos de la población a la que la herramienta va a servir.
- **Documentación de fuentes.** Registrar de dónde provienen los datos, cómo se seleccionaron, depuraron o modificaron y quién tiene acceso a ellos. Sin esa documentación, es muy difícil detectar el origen de un problema y corregirlo.

- **Equipos interdisciplinarios y diversos.** Un equipo que reúne distintas disciplinas, experiencias y perspectivas tiene más posibilidades de reconocer sesgos que otro más homogéneo pasaría por alto. Idealmente, incluye también la mirada de quienes serán afectados por la herramienta.
- **Evaluación y monitoreo.** Probar la herramienta con casos reales y distintos grupos antes de usarla, y seguir controlándola después, porque ningún sistema está "terminado". Las evaluaciones realizadas por equipos independientes aportan una mirada adicional.
- **Revisión humana.** Mantener personas que revisen las recomendaciones antes de que produzcan efectos, sobre todo en decisiones importantes.

Cuando se diseña y se controla con cuidado, la IA también puede ayudar a detectar sesgos humanos y a tomar decisiones más consistentes. La equidad no es un resultado automático, sino una meta que requiere trabajo deliberado.

Idea clave: una herramienta de IA no es neutral por ser una máquina. Aprende de datos humanos y puede repetir sus desigualdades si nadie lo controla.

Actividad 5: detectar un posible sesgo

Una universidad quiere identificar tempranamente a estudiantes con riesgo de abandonar la carrera para ofrecerles tutorías. Para ello, entrena una herramienta con datos de los últimos diez años. El sistema empieza a señalar como "de alto riesgo", en una proporción mucho mayor, a estudiantes que trabajan más de treinta horas semanales y a quienes viven lejos de la sede.

- ¿Podría haber un sesgo? ¿De qué tipo?
- ¿Qué consecuencias podría tener si la universidad usara esta clasificación para decidir becas o admisiones, en lugar de tutorías?
- ¿Qué haría falta revisar?

Orientación: el sistema detecta una correlación real en los datos históricos, pero puede estar reflejando falta de apoyos institucionales más que una menor capacidad del estudiantado. Usarlo para ofrecer ayuda puede ser razonable con revisión humana; usarlo para excluir o reducir oportunidades reproduciría la desigualdad. Habría que revisar los datos, los criterios de uso y los efectos sobre cada grupo.

6. Privacidad y confidencialidad

Tres tipos de información

Antes de usar una herramienta de IA, conviene distinguir qué tipo de información se está manejando:

Estas categorías sirven como guía práctica y pueden superponerse: que un dato sea públicamente accesible no significa que deje de ser un dato personal.

Tipo	Qué es	Ejemplos
Pública	Información que ya está disponible para cualquier persona de forma legítima	Una noticia publicada, el horario de atención de un organismo, una norma oficial
Personal	Información que identifica o permite identificar a una persona	Nombre y apellido, número de documento, dirección, teléfono, correo, historial médico, datos bancarios
Confidencial	Información reservada de una organización o de sus clientes, aunque no sea personal	Presupuestos no publicados, estrategias comerciales, contratos, código fuente, resultados internos

Algunos datos personales reciben una **protección reforzada** por su especial sensibilidad, por ejemplo ciertos datos relativos a la salud, el origen, las convicciones, las opiniones, las afiliaciones o la vida privada. La definición jurídica exacta y las categorías protegidas dependen de la legislación de cada país.

Riesgos de introducir información en herramientas de IA

Cuando escribimos algo en una herramienta generativa, esa información sale de nuestro control directo. Según el servicio y sus condiciones de uso, puede ser almacenada, revisada por personas del proveedor o utilizada para mejorar el sistema. Muchas personas desconocen estas condiciones y pegan en la herramienta documentos completos con datos de clientes, pacientes o estudiantes.

Los principales riesgos son:

- **Pérdida de control sobre los datos.** La organización puede no saber dónde queda almacenada la información ni cómo se usará.
- **Fuga de datos.** Si un sistema fue entrenado o alimentado con información personal o confidencial sin resguardos, puede reproducirla en respuestas a otras personas. Por ejemplo, un asistente de atención al cliente entrenado con correos reales sin anonimizar podría mencionar datos de una compra ajena.
- **Uso sin autorización.** Procesar datos de personas con herramientas de IA sin una base adecuada —por ejemplo, sin su consentimiento cuando este es necesario— puede constituir una infracción a la normativa de protección de datos.

Principios generales de protección

Existen principios de protección de datos ampliamente reconocidos a nivel internacional, como los que recogen las directrices de privacidad de la OCDE y la mayoría de las leyes nacionales. Aplicados a la IA, se traducen en prácticas como estas:

- **Minimización.** Usar solo los datos estrictamente necesarios para la tarea. Si hace falta resumir un reclamo, no es necesario incluir el nombre, el documento ni el teléfono de quien lo hizo.
- **Anonimización o seudonimización.** Eliminar o reemplazar los datos que permiten identificar a una persona ("Cliente A" en lugar del nombre real). Hay que tener en cuenta que a veces una persona puede ser identificada por la combinación de otros datos aparentemente inocentes.
- **Control de acceso.** Limitar quién puede ver y usar cada conjunto de datos.
- **Seguridad.** Proteger la información con medidas técnicas, como el cifrado, y organizativas, como procedimientos y capacitación.
- **Plazos de conservación.** No guardar los datos más tiempo del necesario.
- **Consentimiento e información, cuando corresponda.** Informar a las personas cómo se usarán sus datos y obtener su consentimiento cuando la normativa lo exige.
- **Uso de herramientas aprobadas.** En el ámbito laboral, utilizar las herramientas y configuraciones autorizadas por la organización, que suelen ofrecer condiciones de privacidad distintas de las versiones públicas y gratuitas.

Estas prácticas son principios generales. Las obligaciones legales concretas dependen de cada jurisdicción: en Argentina rige la Ley 25.326 de Protección de los Datos Personales; en la Unión Europea, el Reglamento General de Protección de Datos; y otros países tienen normas propias. Ante una duda en un contexto profesional, corresponde consultar las políticas de la organización y, si es necesario, asesoramiento especializado.

Idea clave: antes de escribir algo en una herramienta de IA, conviene preguntarse: ¿estaría dispuesto o dispuesta a que esta información salga de mi control?

Actividad 6: ¿qué no conviene introducir?

Indicá cuáles de estos datos sería prudente **no** introducir en una herramienta pública y gratuita de IA, y por qué. Para los que sí podrían usarse, pensá si hace falta alguna precaución.

1. El texto de una ley publicada en el boletín oficial.
2. Una planilla con nombres, documentos y diagnósticos de pacientes.
3. El borrador de una presentación para un congreso, ya aprobada para su difusión.
4. La lista de precios que una empresa va a negociar con un proveedor la semana próxima.
5. Un reclamo de un cliente, con su nombre, dirección y número de pedido.
6. Una consulta general sobre cómo organizar una agenda semanal.

Orientación: no convendría introducir 2 (datos personales sensibles), 4 (información confidencial) ni 5 tal como está (datos personales). El reclamo podría usarse si se eliminan los datos identificatorios. Los casos 1, 3 y 6 no presentan, en principio, riesgos de privacidad.

7. Propiedad intelectual y derechos de uso

No todo lo disponible es de libre uso

Que un texto, una imagen o una canción estén en internet no significa que cualquiera pueda usarlos libremente. La **propiedad intelectual** protege las creaciones —obras literarias y artísticas, marcas, invenciones, diseños, programas informáticos— y reconoce a sus autores y titulares derechos sobre su uso.

Para usar legítimamente una obra ajena, en general hace falta que se cumpla alguna de estas condiciones: que la obra esté en el dominio público, que exista una autorización o **licencia** que permita ese uso, o que el uso esté comprendido en alguna excepción prevista por la ley, como la cita con fines educativos, que tiene requisitos y límites que varían según el país. Las licencias pueden ser muy distintas: algunas permiten cualquier uso con solo mencionar la autoría; otras prohíben el uso comercial o la modificación. También los términos de servicio de las plataformas pueden restringir cómo se usa su contenido.

Riesgos en la información que se introduce

Al trabajar con herramientas generativas, la propiedad intelectual puede verse afectada en dos direcciones.

Propiedad intelectual propia. Introducir en una herramienta externa el código fuente de un programa, el diseño de un producto no lanzado o un manual interno puede exponer información valiosa de la organización, por las mismas razones vistas en la sección de privacidad.

Propiedad intelectual de terceros. Cargar en una herramienta obras ajenas completas —un libro, un conjunto de fotografías, un curso de otra institución— para reelaborarlas o reutilizarlas puede implicar un uso no autorizado, según las condiciones de la herramienta y la legislación aplicable.

Riesgos en los resultados generados

Como la IA generativa aprende de grandes cantidades de obras existentes, en ocasiones puede producir resultados muy parecidos a material protegido: un párrafo casi idéntico a un artículo publicado, una imagen que reproduce elementos protegidos de una obra concreta o un diseño que puede confundirse con una marca existente. Si ese resultado se publica o se usa comercialmente, puede generar reclamos legales y daños a la reputación.

Al mismo tiempo, no sería correcto afirmar que todo contenido generado por IA infringe derechos, ni que todo puede usarse libremente. La situación depende del caso concreto y de la jurisdicción. Además, varias cuestiones siguen abiertas: si el uso de obras protegidas para entrenar sistemas de IA es legítimo, o si el contenido generado por IA puede protegerse con derechos de autor y quién sería su titular. La Organización Mundial de la Propiedad Intelectual (OMPI) señala que estas preguntas están en debate y que las respuestas varían de un país a otro (OMPI, *Generative AI: Navigating Intellectual Property*, 2024).

Buenas prácticas

- Revisar las condiciones de uso de la herramienta, incluidas las que se refieren a la propiedad de los resultados.
- Usar obras ajenas solo cuando se cuente con la licencia o autorización correspondiente, y respetar la atribución exigida.
- No introducir propiedad intelectual propia o de terceros en herramientas no autorizadas por la organización.
- Revisar los resultados antes de publicarlos o usarlos profesionalmente, prestando atención a similitudes sospechosas con obras, marcas o textos conocidos.
- Evitar pedir expresamente que una herramienta reproduzca o imite de cerca una obra protegida.
- Ante usos comerciales o dudas relevantes, consultar a especialistas.

Idea clave: que una herramienta pueda generar algo no significa que ese contenido pueda usarse libremente. La verificación antes de publicar es parte del trabajo.

Actividad 7: ¿qué hay que verificar?

Un emprendimiento de repostería prepara su nueva página web. Para ahorrar tiempo:

- copia las descripciones de productos de la página de una pastelería reconocida y le pide a una herramienta generativa que las "reescriba un poco";
- usa como fondo fotografías encontradas en un buscador de imágenes;
- genera con IA un logotipo que, al mirarlo con atención, se parece bastante al de una cadena de cafeterías conocida.

¿Qué debería verificar antes de publicar la página? ¿Qué alternativas tendría en cada caso?

Orientación: las descripciones parten de textos ajenos y una reescritura superficial puede seguir siendo una copia; conviene redactarlas desde cero a partir de los productos propios. Las fotografías requieren verificar su licencia o reemplazarlas por imágenes propias o de bancos con licencia adecuada. El logotipo puede generar conflictos con una marca registrada; conviene descartarlo y buscar un diseño claramente distinto.

8. Contenido dañino, información incorrecta y alucinaciones

Contenido dañino e incorrecto

Contenido dañino: material generado por IA —texto, imágenes, audio o video— que ofende, discrimina, hostiga, engaña o puede causar perjuicios a las personas.

Incluye, entre otros, los discursos de odio (que buscan humillar o atacar a un grupo), el lenguaje abusivo (que intenta intimidar o degradar), las obscenidades, las instrucciones que podrían facilitar daños y las imágenes o audios falsos de personas reales creados para engañar.

Contenido incorrecto: información falsa, inexacta o desactualizada presentada como si fuera verdadera.

Qué es una "alucinación"

En el ámbito de la IA generativa, se llama **alucinación** a una respuesta que contiene información falsa o inventada presentada con la misma seguridad que un dato correcto. El NIST utiliza para este fenómeno el término **confabulación** (NIST AI 600-1, 2024). Algunos ejemplos típicos:

- una cita textual atribuida a una persona que nunca la dijo;

- una referencia bibliográfica con autor, título y año que no existe;
- un artículo de una ley con un número y un contenido inventados;
- un dato estadístico preciso que no proviene de ninguna fuente.

En herramientas de visión artificial también pueden producirse errores de este tipo, cuando el sistema "ve" objetos o patrones que no están en la imagen.

Por qué una respuesta convincente puede ser falsa

Como vimos en la clase anterior, una herramienta generativa construye su respuesta eligiendo, paso a paso, el contenido más probable según los patrones que aprendió. Su objetivo es producir un texto plausible, no comprobar que sea verdadero. Por eso:

- puede completar con datos inventados cuando no tiene información suficiente;
- escribe con el mismo tono seguro cuando acierta y cuando se equivoca;
- puede basarse en información desactualizada;
- tiende a responder aunque la pregunta esté mal planteada o no tenga respuesta.

La fluidez del texto genera una impresión de autoridad. Justamente por eso, la calidad de la redacción no es un indicio de que el contenido sea correcto.

Riesgos de usar resultados no verificados

Las consecuencias dependen del contexto. En una lista de ideas para un evento, un error es irrelevante. Pero en otros ámbitos puede ser grave: una indicación equivocada sobre un medicamento, un dato falso en un informe que se entrega a un cliente, una referencia inexistente en un trabajo académico, un plazo legal mal informado o información errónea difundida durante una emergencia. Además de los daños directos, estos errores deterioran la confianza en las personas y organizaciones que los difunden.

Estrategias para prevenirlo

Algunas medidas corresponden a las organizaciones que ponen en marcha estas herramientas; otras, a cada persona que las usa.

- **Definir claramente el propósito.** Establecer para qué se usará la herramienta y para qué no. Un asistente pensado para responder sobre horarios y trámites no debería dar consejos médicos o legales.
- **Limitar el ámbito de uso.** Restringir los temas y las fuentes que la herramienta puede usar, por ejemplo, que responda solo a partir de documentos oficiales de la organización.
- **Usar instrucciones y formatos estructurados.** Indicar con precisión qué se espera, en qué formato y con qué límites; por ejemplo, pedir que la herramienta indique cuando no tiene información suficiente en lugar de completar con suposiciones.
- **Filtrar contenido dañino.** Incorporar filtros que detecten y bloqueen lenguaje ofensivo o peligroso antes de que llegue a las personas usuarias.
- **Probar antes de usar.** Evaluar la herramienta con muchos casos, incluidos casos difíciles, antes de ponerla en funcionamiento.
- **Monitorear.** Revisar periódicamente las respuestas y los reclamos, y ajustar o reentrenar la herramienta cuando sea necesario.
- **Verificar con fuentes fiables.** Contrastar datos, citas, cifras y referencias con fuentes originales y reconocidas.
- **Recurrir a supervisión humana.** Que una persona con conocimiento del tema revise los resultados, especialmente cuando las consecuencias de un error son importantes.

Idea clave: una respuesta bien escrita no es una respuesta verificada. Cuanto más importante es la decisión, más necesaria es la verificación.

Actividad 8: antes de usar la respuesta

Un estudiante le pregunta a una herramienta generativa por estudios sobre los efectos del uso del celular en el rendimiento escolar. La herramienta responde con un párrafo muy claro que menciona "un estudio de 2021 de la Universidad de Toronto, publicado en una revista científica, según el cual el uso del celular en clase reduce el rendimiento un 23 %".

¿Qué pasos debería seguir antes de incluir ese dato en su trabajo?

Orientación: buscar el estudio en fuentes académicas o en el sitio de la universidad; comprobar que existe y que dice lo que la herramienta afirma; leer al menos el resumen para entender el contexto del dato; citar la fuente original y no la herramienta. Si no puede encontrarlo, no debe usar el dato.

9. Uso responsable de la IA generativa

Integrar lo aprendido

Los temas de esta clase pueden resumirse en una idea central:

La IA puede generar opciones, acelerar tareas y organizar información, pero las personas conservan la responsabilidad de aportar contexto, evaluar los resultados y decidir cómo utilizarlos.

Una forma práctica de aplicar esta idea es pensar el uso de la herramienta en tres momentos.

Antes de usarla

- ¿Esta tarea es adecuada para una herramienta generativa, o requiere un juicio que no debería delegarse?
- ¿La herramienta está autorizada para este uso en mi organización?
- ¿Voy a introducir información personal, confidencial o protegida? ¿Puedo evitarlo o anonimizarla?

Mientras la uso

- ¿Di instrucciones claras sobre qué necesito y con qué límites?
- ¿Aporté el contexto que la herramienta no tiene?

Después de obtener el resultado

- ¿Verifiqué datos, cifras, citas y referencias?
- ¿El resultado contiene sesgos, estereotipos o lenguaje inadecuado?
- ¿Se parece demasiado a una obra ajena?
- ¿Corresponde informar que se usó IA en este trabajo?
- ¿Estoy en condiciones de responder por este resultado?

Qué hace la herramienta y qué hace la persona

La IA puede...	La persona debe...
Generar borradores, textos, imágenes o código	Revisar su exactitud, calidad y adecuación
Sugerir opciones o recomendaciones	Decidir cuál aplicar y asumir la decisión
Resumir documentos extensos	Verificar que el resumen sea fiel y no omita lo importante
Identificar patrones en la información	Interpretar su significado en el contexto real
Producir muchas alternativas en poco tiempo	Evaluar sus consecuencias para las personas afectadas

La división de tareas no es fija: depende del riesgo y del contexto. Pero hay algo que no cambia: la herramienta no puede asumir la responsabilidad por lo que produce. Esa responsabilidad sigue siendo humana.

Actividad 9: reflexión sobre una tarea propia

Elegí una tarea de tu estudio, tu trabajo o tu vida cotidiana en la que podrías usar una herramienta de IA generativa (por ejemplo, preparar un informe, responder correos, organizar apuntes o diseñar una presentación). Luego respondé:

1. ¿Qué parte de la tarea podría hacer la IA?
2. ¿Qué debería revisar o verificar la persona?
3. ¿Qué información no convendría introducir en la herramienta?
4. ¿Qué decisión no debería delegarse automáticamente, y por qué?

Síntesis

La IA generativa es un tipo de inteligencia artificial que produce contenido nuevo —texto, imágenes, audio, video o código— a partir de los patrones aprendidos en grandes volúmenes de datos. A diferencia de los usos que solo analizan o clasifican información, crea resultados que antes no existían. Esa capacidad la vuelve muy útil para acelerar tareas y explorar alternativas, pero también explica sus riesgos: puede producir información incorrecta, contenidos dañinos, sesgos, problemas de privacidad y conflictos de propiedad intelectual.

La ética de la IA es el campo multidisciplinario que busca aprovechar los beneficios de esta tecnología y reducir sus riesgos. En esta clase la analizamos a partir de tres ejes. La **transparencia** permite saber que se interactúa con IA, conocer sus capacidades y limitaciones y comprender en qué se basan sus resultados, lo que es especialmente importante en ámbitos de alto impacto. La **responsabilidad** recuerda que la IA no elimina la obligación de responder por los resultados: quienes desarrollan, implementan y usan estas herramientas deben revisarlas, supervisarlas y corregirlas. La **equidad** exige atender a los sesgos que pueden aparecer en los datos, en las decisiones y en los contenidos generados, mediante representación diversa, documentación, equipos interdisciplinarios, monitoreo y revisión humana.

A estos ejes se suman cuidados concretos. En materia de **privacidad**, conviene limitar y proteger los datos personales y confidenciales, anonimizar cuando sea posible y respetar la normativa de cada jurisdicción. En **propiedad intelectual**, no todo contenido disponible es de libre uso, y los resultados generados deben revisarse antes de publicarse. Frente al **contenido dañino y las alucinaciones**, es necesario definir el propósito de la herramienta, limitar su alcance, probarla, monitorearla y verificar sus resultados con fuentes fiables.

En conjunto, el uso responsable de la IA generativa se basa en una idea sencilla: la herramienta puede proponer, pero la persona evalúa, verifica y decide.

Actividad de repaso

1. Explicá con tus palabras en qué se diferencia un uso generativo de la IA de un uso que solo analiza información. Da un ejemplo de cada uno.
 2. ¿Por qué el hecho de que la IA generativa aprenda de grandes cantidades de datos existentes se relaciona con problemas de sesgo y de propiedad intelectual?
 3. Un servicio de atención al cliente usa un asistente automático con nombre y foto de una persona, sin aclarar que es IA. ¿Qué principio se ve afectado y qué cambiarías?
 4. Una clínica usa una herramienta que sugiere qué pacientes deben tener prioridad para un turno, pero nadie puede explicar en qué se basan sus sugerencias. ¿Por qué es un problema? ¿Qué medidas propondrías?
 5. "Si la IA se equivocó, la culpa es del sistema." Analizá esta afirmación a partir del concepto de rendición de cuentas.
 6. Una herramienta genera imágenes para una campaña sobre profesiones, y todas las ingenieras e ingenieros aparecen como varones jóvenes. ¿Qué tipo de sesgo aparece y qué consecuencias podría tener?
 7. Una docente quiere usar una herramienta pública de IA para redactar informes sobre sus estudiantes. ¿Qué precauciones de privacidad debería tomar?
 8. ¿Por qué no es correcto afirmar ni que todo contenido generado por IA infringe derechos de autor ni que puede usarse libremente? Explicá qué conviene verificar antes de publicarlo.
 9. Una herramienta te entrega un resumen de una norma con un número de artículo y una fecha de entrada en vigencia. ¿Qué harías antes de usar esa información en un documento profesional y por qué?
 10. Pensá en una decisión importante de tu ámbito de estudio o de trabajo. ¿Por qué no debería delegarse por completo en una herramienta de IA? ¿Qué papel sí podría cumplir la herramienta?
-

Fuentes de referencia

Las siguientes fuentes se utilizaron para complementar, generalizar o actualizar los contenidos del material de base:

- NIST (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, NIST AI 100-1. Instituto Nacional de Estándares y Tecnología, Estados Unidos.
- NIST (2024). *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile*, NIST AI 600-1.
- NIST (2022). *Towards a Standard for Identifying and Managing Bias in Artificial Intelligence*, NIST Special Publication 1270.
- OCDE (2019, actualizada en 2024). *Recomendación del Consejo sobre Inteligencia Artificial* (Principios de IA de la OCDE).
- UNESCO (2021). *Recomendación sobre la Ética de la Inteligencia Artificial*.
- OMPI/WIPO (2024). *Generative AI: Navigating Intellectual Property*.
- Unión Europea (2024). Reglamento (UE) 2024/1689, de Inteligencia Artificial (mencionado solo como ejemplo de regulación de una jurisdicción).
- Argentina. Ley 25.326 de Protección de los Datos Personales (mencionada solo como ejemplo de normativa nacional).