

A2

Evaluación del impacto de los sistemas de IA

Un método cualitativo para estimar vulnerabilidad, riesgo e impacto de forma estructurada y comparable entre proyectos.

A.2.1 La fórmula encadenada V-P-N

1 Estimar la vulnerabilidad (V)

2 Estimar la probabilidad (P)

3 Calcular el riesgo ($R = P \times V$)

4 Calcular el impacto ($I = R \times N$)

A.2.6 Factores clave para la evaluación

Evaluación del impacto de los sistemas de IA

El marco de clasificación de riesgos exige una evaluación de impacto para los sistemas de IA de riesgo alto y la recomienda para los otros casos, por lo que es importante saber cómo medir el impacto. En esta sección se propone un método cualitativo sencillo para estimar el impacto de un sistema de IA de forma estructurada, reproducible y comparable entre proyectos, sin requerir datos difíciles de conseguir o métodos estadísticos complejos.

El método descansa en tres variables encadenadas: la vulnerabilidad de las personas expuestas, el riesgo (que resulta de combinar esa vulnerabilidad con la probabilidad de que ocurra un efecto adverso) y el impacto total que resulta de extender ese riesgo al número de personas potencialmente afectadas.

Las tres variables del método

- Vulnerabilidad (V): se refiere a susceptibilidades internas de las personas usuarias del sistema (socioemocionales, socioeconómicas, de género, familiares, situacionales, etc.). A mayor vulnerabilidad, mayor riesgo.
- Riesgo (R): resultado de multiplicar la probabilidad de ocurrencia de un efecto (P) por la vulnerabilidad (V) de las personas expuestas. $R = P \times V$.
- Impacto (I): es el resultado de multiplicar el riesgo (R) por el número de personas expuestas a los efectos del sistema (N). $I = R \times N$.

A.2.1

La fórmula encadenada V-P-N

Las tres variables se combinan de forma secuencial. La probabilidad y la vulnerabilidad se multiplican para producir el riesgo; el riesgo se multiplica por el número de personas expuestas para producir el impacto total. La lógica es acumulativa: un efecto poco probable sobre personas poco vulnerables y en número reducido produce un impacto bajo; un efecto poco probable sobre personas muy vulnerables y en gran número produce un impacto crítico. El método obliga a considerar las tres dimensiones de forma explícita, evitando que una evaluación se concentre solo en la probabilidad técnica del fallo e ignore quién y cuántos lo sufren.

$$R = P \times V \rightarrow I = R \times N$$

A.2.1.1 • PASO 1

Estimar la vulnerabilidad (V)

La vulnerabilidad describe qué tan susceptibles son al daño las personas expuestas al sistema. No es una propiedad del sistema sino de la población afectada y su contexto. Se estima en una escala de cinco niveles:

TABLA A.2.1

Descripción y niveles de la vulnerabilidad

NIVEL DE VULNERABILIDAD	VALOR	DESCRIPCIÓN
Muy baja	1	Personas con amplios recursos psicoemocionales, autonomía y capacidad de autoprotegerse o revertir efectos negativos.
Baja	2	Personas con recursos suficientes y algunas salvaguardas, pero no plenas.
Media	3	Personas en situación promedio, con vulnerabilidades parciales o circunstanciales.
Alta	4	Personas con vulnerabilidades significativas (socioemocionales, económicas, de género, situacionales) que limitan su capacidad de defensa.
Muy alta	5	Grupos en situación de especial protección: menores, personas con discapacidad, en pobreza extrema o sin acceso a recursos de ningún tipo.

Fuente: Elaboración propia.

Dimensiones de vulnerabilidad a considerar

- Socioemocional: dependencia, soledad, salud mental, susceptibilidad a la manipulación.
- Socioeconómica: pobreza, precariedad laboral, exclusión financiera o digital.
- De género: exposición a discriminación o violencia por razón de género.
- Familiar y situacional: dependencia de cuidados, situación migratoria, crisis temporales.
- De acceso: falta de alfabetización digital, idioma, o acceso a mecanismos de queja.

A.2.1.2 · PASO 2

Estimar la probabilidad (P)

La probabilidad es un indicador de qué tan factible es que el efecto adverso ocurra durante la operación normal del sistema. Con el propósito de que el método sea fácil de aplicar, en lugar de estimar probabilidades reales se definen cinco niveles cualitativos:

TABLA A.2.2

Descripción y valores de la probabilidad de ocurrencia

PROBABILIDAD DE OCURRENCIA (ESTIMACIÓN CUALITATIVA)	VALOR ASIGNADO	DESCRIPCIÓN
Muy baja	1	El efecto es teóricamente posible pero altamente improbable en operación normal.
Baja	2	El efecto podría ocurrir en circunstancias excepcionales.
Media	3	El efecto ocurre ocasionalmente; es razonable esperarlo en algún momento.
Alta	4	El efecto ocurre con frecuencia o es probable en condiciones habituales.
Muy alta	5	El efecto es casi seguro o recurrente en la operación normal del sistema.

Fuente: Elaboración propia.

A.2.1.3 · PASO 3

Calcular el riesgo (R = P × V)

El riesgo combina probabilidad y vulnerabilidad. La matriz siguiente muestra el producto P × V para los 25 cruces posibles. La propuesta de usar una función multiplicativa hace que un mismo valor de riesgo pueda provenir de combinaciones distintas, por ejemplo, alta probabilidad con baja vulnerabilidad, o baja probabilidad con alta vulnerabilidad.

TABLA A.2.3

Matriz cualitativa de riesgo (R = P × V)

P \ V	1	2	3	4	5
1	1	2	3	4	5
2	2	4	6	8	10
3	3	6	9	12	15
4	4	8	12	16	20
5	5	10	15	20	25

Fuente: Elaboración propia.

A.2.1.4 · PASO 4

Calcular el impacto (I = R × N)

El impacto extiende el riesgo individual al conjunto de la población afectada y la escala se determina casuísticamente por proyecto. Por ejemplo, en el caso de un sistema que recomienda la asignación de becas la población objetivo es la población escolar que califica para ese tipo de becas, y en el caso de un sistema de detección facial masiva la población afectada puede ser la de una región o la del país entero. El número de personas expuestas (N) se clasifica en órdenes de magnitud para mantener el método manejable.

TABLA A.2.4

Clasificación cualitativa de niveles de impacto (caso genérico, de pequeña escala a población entera)

NIVEL DE IMPACTO	VALOR ASIGNADO	DESCRIPCIÓN
Muy reducido	1	Decenas de personas o menos (p. ej. un piloto acotado).
Reducido	2	Cientos de personas.
Medio	3	Miles de personas.
Amplio	4	Decenas o cientos de miles de personas.
Masivo	5	Millones de personas o población general.

Fuente: Elaboración propia.

La escala es relativa y debe ajustarse casuísticamente. En la tabla anterior es aplicable a sistemas de IA que pueden afectar desde algunas personas hasta millones de personas en un país o en varios países. Pero hay muchos casos en los que los niveles de impacto pueden ir de un puñado de personas a unas cuantas decenas o cientos, por lo que habría que redefinir los niveles para ajustarlos a esa situación.

El impacto final $I = R \times N$ se interpreta en una escala de severidad que orienta la respuesta regulatoria. Dado que R va de 1 a 25 y N de 1 a 5, el impacto teórico va de 1 a 125. La siguiente tabla traduce los rangos de impacto en niveles de severidad y los conecta con el marco de clasificación de riesgos:

TABLA A.2.5

Severidad del impacto

RANGO DE I	SEVERIDAD	RESPUESTA REGULATORIA SUGERIDA
1 – 15	Baja	Cumplimiento voluntario de buenas prácticas; revisión periódica. Compatible con cuadrante A.
16 – 40	Moderada	Obligaciones de transparencia y monitoreo; soft law reforzado. Compatible con riesgo limitado.
41 – 75	Alta	Evaluación de impacto formal, auditoría y supervisión activa (sandbox o piloto). Riesgo alto.
76 – 125	Crítica	Salvaguardas máximas: evaluación de impacto previa, criterios de parada, supervisión estricta o moratoria. Evaluar si el riesgo es inaceptable.

Fuente: Elaboración propia.

Ejemplo de estimación de impacto

Para ilustrar el método, se aplica al caso de un sistema de IA que recomienda la asignación de becas escolares en función del perfil socioeconómico de las familias solicitantes.

Cálculo paso a paso

1. Población expuesta: familias en situación de pobreza que solicitan becas → vulnerabilidad muy alta ($V = 5$).
2. Probabilidad del efecto adverso (negar indebidamente una beca por sesgo del modelo): ocurre ocasionalmente, $P = 3$.
3. Riesgo: $R = P \times V = 3 \times 5 = 15$ (riesgo muy alto en la matriz).
4. Número de personas expuestas: decenas de miles de familias a nivel estatal, $N = 4$.
5. Impacto: $I = R \times N = 15 \times 4 = 60$ → severidad ALTA.

Conclusión: el sistema requiere evaluación de impacto formal, auditoría de equidad y supervisión activa antes y durante el despliegue (cuadrante B o D según la certidumbre sobre sus efectos).

Notas sobre el uso del método

- El método es cualitativo: los números no son mediciones exactas sino una forma estructurada de comparar y justificar decisiones. Su valor está en la transparencia del razonamiento, no en la precisión aritmética.
- La estimación de cada variable debe documentarse y, cuando sea posible, validarse con las personas afectadas o con expertos del sector.
- El impacto debe recalcularse cuando cambien las condiciones: una ampliación de la población (N) o un cambio en el contexto de vulnerabilidad (V) puede elevar la severidad sin que el sistema haya cambiado técnicamente.

Cuando un mismo sistema produce varios efectos adversos distintos, se evalúa cada uno por separado y se toma el de mayor impacto como referencia para la selección de modelos.

TABLA A.2.6

Factores clave para la evaluación de impacto

#	FACTOR	PREGUNTA GUÍA	VARIABLE AFECTADA
1	Escala de despliegue	¿Cuántas personas pueden verse afectadas por los efectos del sistema?	N
2	Reversibilidad	¿Pueden revertirse los efectos negativos, o son permanentes y difíciles de corregir?	P y V
3	Autonomía del sistema	¿Opera sin supervisión humana significativa? A mayor autonomía, mayor probabilidad de que un efecto adverso se materialice sin corrección.	P
4	Grupos vulnerables	¿Afecta desproporcionadamente a menores, personas con discapacidad, en pobreza u otros grupos en situación de especial protección?	V
5	Contexto sectorial	¿El sector tiene alguna regulación específica preexistente (salud, finanzas, justicia, defensa) que eleve la gravedad de un fallo?	P y V
6	Datos utilizados	¿Procesa datos personales sensibles o datos biométricos cuyo mal uso aumentaría el daño?	V
7	Frecuencia de uso	¿Con qué frecuencia se toman decisiones con el sistema? A mayor frecuencia, mayor probabilidad acumulada de un efecto adverso.	P
8	Acceso a remedios	¿Las personas afectadas pueden detectar el daño, quejarse y obtener reparación con facilidad? La dificultad de remedio agrava la vulnerabilidad.	V

Fuente: Elaboración propia.

Cómo se conecta con el método: los factores no se puntúan por separado; sirven para razonar el valor de cada variable antes de aplicar la fórmula $I = (P \times V) \times N$. Por ejemplo, los factores 4, 6 y 8 ayudan a justificar un valor alto de vulnerabilidad (V); los factores 3 y 7, un valor alto de probabilidad (P); y el factor 1, el valor de personas expuestas (N).

V: Vulnerabilidad P: Probabilidad N: Número de personas expuestas