DEPUTACIÓN DE LUGO

A intelixencia artificial aplicada á Administración Pública

Unidade 4.3

Do cumprimento ao uso responsable: supervisión, riscos e responsabilidade

Unidade 4.3. Do cumprimento ao uso responsable: supervisión, riscos e responsabilidade

Punto de partida

A Axencia Española de Supervisión da Intelixencia Artificial (AESIA) é unha entidade de dereito público estatal con personalidade xurídica propia. Forma parte da estrutura española de gobernanza da IA e exerce funcións de supervisión, inspección, apoio e difusión para favorecer un uso seguro, responsable e fiable destes sistemas.

As guías da AESIA son recursos de apoio para aplicar e cumprir o Regulamento europeo de intelixencia artificial. Desenvolvéronse durante o piloto español do sandbox regulatorio de IA, coa colaboración das entidades participantes, asistencias técnicas, posibles autoridades nacionais competentes e o grupo asesor de persoas expertas do sandbox. A experiencia deste piloto permitiu trasladar os requisitos normativos a recomendacións, comprobacións e exemplos de uso.

Estas guías non teñen carácter vinculante, non substitúen o RIA nin desenvolven por si mesmas a normativa aplicable. Ofrecen recomendacións prácticas aliñadas cos requisitos regulatorios mentres se completan as normas harmonizadas e as directrices europeas. Publícanse sen prexuízo das guías técnicas que elabora a Comisión Europea para favorecer unha aplicación coherente do RIA en todos os Estados membros. Segundo explica a páxina oficial, o traballo español tamén serve de base ao grupo da Comisión encargado de preparar as guías europeas.

Os documentos están sometidos a avaliación e revisión permanentes. A AESIA prevé actualizacións periódicas conforme avancen os estándares, as directrices da Comisión Europea e as modificacións do marco normativo. A páxina oficial anuncia, en particular, unha actualización cando se aprobe o Ómnibus dixital que modifica o RIA. Por iso, as guías deben consultarse sempre na súa versión vixente e non como documentos pechados.

O conxunto organízase en catro bloques:

- Guías 01 e 02: introdución ao ámbito, aos papeis, aos riscos e aos casos de uso.
- Guías 03 a 15: orientación técnica sobre os requisitos aplicables aos sistemas de alto risco.
- Guía 16: manual para utilizar as listas de comprobación.
- Arquivo complementario: compendio de listas de comprobación e exemplos en formato ZIP.

Esta é a organización editorial da páxina oficial. Para comprender o ciclo de cumprimento, a unidade reorganiza as dezaseis guías en cinco bloques funcionais: orientación, entrada no mercado, deseño

e control, evidencia e funcionamento continuado. Non se engade unha categoría nova; cámbiase o mapa para relacionar cada documento cun momento do sistema.

As 16 guías non son dezaseis normas novas. Moitas céntranse nas obrigas dos provedores, pero unha administración que compra e utiliza un sistema necesita coñecelas para saber que documentación debe esixir, que condicións debe comprobar e que controis debe organizar durante o despregamento.

Esta unidade conecta esas orientacións co RIA e co RXPD mediante un protocolo de traballo. Trátanse tres preguntas:

- Que documentación e control necesita unha entidade?
- Que significa supervisar de verdade?
- Quen responde cando interveñen provedor, organización, persoal técnico e órgano decisor?

A mensaxe non será “a persoa que asina é sempre a única responsable”. A responsabilidade pode distribuírse segundo o caso. O que non cambia é que unha ferramenta non elimina as obrigas da entidade nin pode asumir unha competencia pública.

1. As guías da AESIA nun único mapa

Bloque funcional	Momento do ciclo	Pregunta para a entidade
Orientación inicial	Antes de definir ou adquirir	Que sistema, finalidade, papel e risco estamos a analizar?
Entrada no mercado	Antes da posta en servizo	Que avaliación, declaracións, instrucións e sistema de calidade corresponden ao provedor?
Deseño e control	Preparación e uso	Como se xestionan riscos, datos, precisión, solidez, seguridade e supervisión?
Evidencia e trazabilidade	Durante todo o ciclo	Que documentación e rexistros permiten reconstruír o funcionamento?
Funcionamento continuado	Operación, cambio e retirada	Como se detectan cambios, erros e incidentes e quen debe actuar?

Orientación inicial

As guías 01 e 02 explican ámbito, papeis, riscos e casos. Serven para entrar no RIA e identificar que documentación se necesita. Non permiten certificar un sistema nin substitúen a lectura do Regulamento.

Entrada no mercado

A guía 03 trata a avaliación de conformidade. O provedor debe demostrar que o sistema de alto risco cumpre antes da comercialización ou posta en servizo. Segundo o tipo, pode empregar control interno ou necesitar un organismo notificado.

A guía 04 desenvolve o sistema de xestión da calidade do provedor: políticas, responsabilidades, deseño, datos, probas, documentación, cambios e vixilancia.

Para unha entidade compradora, as preguntas son:

- Está claramente identificado o sistema e a versión?
- Recibimos declaración, marcado e documentación cando corresponda?

- As instrucións cobren o noso uso?
- Quen comunica cambios e incidentes?
- A personalización cambia a finalidade ou o papel da entidade?

Deseño e control

As guías 05 a 11 forman o núcleo técnico.

Guía	Pregunta que debe entender a entidade
Xestión de riscos	Que danos se identificaron, como se probaron e que risco queda?
Vixilancia ou supervisión humana	Quen pode interpretar, corrixir e deter?
Datos e gobernanza	De onde proceden os datos e son pertinentes e representativos?
Transparencia	Que sabe o usuario sobre finalidade, límites e precisión?
Precisión	Que métricas serven para esta tarefa e como se manteñen?
Solidez	Que ocorre ante erro, fallo, cambio ou entrada imprevista?
Ciberseguridade	Como se protexe fronte a manipulación, acceso e ataques propios da IA?

A precisión media pode ocultar erros graves nun grupo. A robustez non significa que o sistema nunca falle, senón que se probaron condicións e existen medidas de recuperación. A ciberseguridade non se limita a contrasinais; inclúe datos, modelo, dependencias e entradas manipuladas.

Evidencia e trazabilidade

A guía 12 trata rexistros. Permiten investigar que versión se utilizou, cando, con que entrada e que saída se produciu. Deben ser útiles e estar protexidos; gardar todo sen límite crea riscos de datos.

A guía 15 trata documentación técnica: finalidade, deseño, datos, métricas, riscos, supervisión e cambios. Corresponde ao provedor, pero o responsable do despregamento necesita información suficiente para usar, supervisar e explicar.

Funcionamento continuado

A guía 13 aborda a vixilancia posterior á comercialización. O provedor debe recoller e analizar información durante a vida do sistema. A entidade usuaria debe comunicar erros e patróns, non esperar a un dano grave.

A guía 14 explica incidentes graves. O provedor ten obrigas de notificación; o responsable do despregamento debe vixiar, informar e suspender cando considere que existe un risco nas condicións do artigo 26.

A guía 16 ofrece unha lista de comprobación para o diagnóstico e a madurez. É unha ferramenta de traballo, non unha certificación.

Quen debe facer que

Materia	Provedor	Responsable do despregamento
Finalidade e instrucións	Define a finalidade prevista e documenta capacidades e límites	Comproba que o uso propio encaixa e organiza instrucións internas
Riscos e rendemento	Deseña, proba e documenta segundo as obrigas aplicables	Avalía o contexto, vixía resultados e comunica problemas
Datos	Cumpre os requisitos de datos que lle correspondan no deseño	Controla pertinencia e representatividade dos datos de entrada baixo a súa responsabilidade
Supervisión	Deseña medidas e información que permitan intervir	Designa persoas competentes, con tempo, información e autoridade
Rexistros e incidentes	Xera, conserva e notifica segundo o seu papel	Conserva os rexistros baixo o seu control, informa e suspende cando proceda

Un contrato pode concretar tarefas, pero non borrar as obrigas legais de cada actor. A entidade debe comprobar o que entrega o provedor e, ao mesmo tempo, organizar aquilo que só pode decidir no seu contexto: persoas usuarias, datos, procedemento, efecto e capacidade de detención.

2. Que debe facer o responsable do despregamento

Se unha entidade local usa un sistema de alto risco, o artigo 26 esixe, entre outras actuacións:

- Seguir as instrucións mediante medidas técnicas e organizativas.
- Encargar a supervisión a persoas competentes, formadas e con autoridade.
- Controlar pertinencia e representatividade dos datos de entrada baixo a súa responsabilidade.
- Vixiar o funcionamento.
- Informar de riscos e incidentes e suspender cando corresponda.
- Conservar os rexistros baixo o seu control polo período aplicable.
- Informar ao persoal e aos seus representantes se o sistema se usa no traballo.
- Cumprir rexistro e información ás persoas afectadas nos supostos previstos.

Non se deduce que todo uso dun asistente xeral sexa de alto risco. Primeiro hai que clasificar. Pero seguir instrucións, asignar responsable, comprobar entradas e poder deter son prácticas útiles tamén noutros usos.

3. Avaliación de impacto en dereitos fundamentais

Antes de determinados usos de alto risco do anexo III, os organismos de dereito público deben avaliar o impacto sobre dereitos fundamentais.

A avaliación describe:

- Proceso no que se empregará.
- Período e frecuencia.
- Persoas e colectivos afectados.
- Riscos específicos.
- Medidas de supervisión.
- Resposta se o risco se materializa.

Non é o mesmo ca unha avaliación de impacto en protección de datos. A do RXPD céntrase no tratamento de datos e no alto risco para dereitos e liberdades. A do RIA considera o sistema e un conxunto máis amplo de dereitos no contexto do uso. Poden coordinarse e complementarse.

O certificado do provedor non substitúe esta análise. O provedor avalía o produto e a entidade avalía o seu despregamento concreto.

Cuestión	Avaliación de impacto en protección de datos	Avaliación de impacto en dereitos fundamentais
Marco	RXPD, especialmente artigo 35	RIA, artigo 27
Obxecto	Operacións de tratamento de datos persoais	Uso concreto de determinados sistemas de alto risco
Foco	Risco para dereitos e liberdades derivado do tratamento	Efectos do sistema sobre os dereitos fundamentais no contexto de despregamento
Resultado	Medidas, modificación ou decisión de non tratar	Medidas de supervisión, gobernanza, resposta e decisión sobre o uso
Relación	Pode fornecer información e medidas á outra avaliación	Pode realizarse xunto coa avaliación de datos cando ambas procedan

No caso transversal, a avaliación de protección de datos examina bases, categorías, accesos, conservación e decisión automatizada. A avaliación de dereitos fundamentais amplía a mirada: acceso efectivo á prestación, igualdade, explicación, reclamación, impacto en colectivos vulnerables e consecuencias dunha prioridade incorrecta. Coordinalas evita duplicación, pero non permite substituír unha pola outra.

4. Transparencia en capas

“Ser transparente” pode referirse a obrigas distintas:

1. O provedor informa ao responsable do despregamento sobre finalidade, capacidades, límites, precisión e supervisión.
2. A entidade informa as persoas cando un sistema de alto risco se usa para adoptar ou axudar a adoptar decisións relacionadas con elas.
3. Determinados sistemas deben revelar interacción con IA ou contido sintético.

4. O RXPD esixe información sobre o tratamento de datos.
5. O procedemento administrativo mantén motivación, acceso e garantías.

Publicar “usamos IA” non explica que datos, que efecto, que revisión nin que dereitos existen. Tampouco se debe expoñer información técnica que comprometa a seguridade. A transparencia debe ser comprensible e útil para a persoa destinataria.

5. Supervisión humana que serve

A guía 06 emprega a expresión vixilancia humana. Nesta unidade tamén se utiliza supervisión humana para describir a mesma función do artigo 14 do RIA: permitir que persoas competentes comprendan, revisen, corrixan e, cando proceda, deteñan o funcionamento ou o uso do sistema.

Hai supervisión efectiva cando se cumpren cinco condicións.

Competencia

A persoa coñece finalidade, límites, indicadores e erros previsibles. Unha formación xenérica sobre prompts non abonda para supervisar unha puntuación de risco.

Información

Pode acceder a fontes, instrucións, explicación e rexistros necesarios. Se só ve unha puntuación sen procedencia, a supervisión queda limitada.

Tempo

Pode comprobar antes de que a saída produza efecto. Revisar centos de decisións en segundos non é unha intervención realista.

Autoridade

Pode corrixir, ignorar, deter ou escalar. Se a organización penaliza apartarse do resultado, aumenta a dependencia.

Deseño

A interface mostra límites e permite intervención. Unha pantalla que presenta unha recomendación como resultado definitivo favorece o sesgo de automatización.

O sesgo de automatización consiste en conceder máis peso á saída porque procede dun sistema. Pode aparecer mesmo en persoal experto, sobre todo baixo presión de tempo. Mostrar unha persoa ao final do fluxo non resolve o problema se non cumpre estas condicións.

Aplicación ao caso transversal

A pantalla presenta unha puntuación de 0 a 100 e recomenda atender primeiro as solicitudes superiores a 75. Para que exista supervisión real, a persoa revisora debe coñecer que significa a puntuación, que datos a sustentan, que erros son previsibles e como consultar o expediente. Debe dispoñer de tempo para revisar, poder cambiar a orde e deixar constancia do motivo.

Se a interface oculta os factores, se o volume obriga a confirmar en segundos ou se apartarse da recomendación require unha autorización excepcional, a supervisión é aparente. A medida correcta non é engadir unha sinatura: hai que modificar información, carga, interface ou procedemento, e deter o uso se non se pode garantir unha intervención efectiva.

6. Alucinacións: fluidez sen garantía

Un modelo de linguaxe xera texto probable. Non comproba necesariamente que unha norma, sentenza, cifra ou nome exista. Pode:

- Inventar unha norma con título verosímil.
- Citar un artigo real cun contido incorrecto.
- Mesturar dúas versións dunha disposición.
- Crear unha cifra para completar un oco.
- Atribuír unha conclusión á fonte que non a sostén.

A busca e a recuperación documental reducen algúns erros, pero non os eliminan. Poden seleccionar unha fonte desactualizada ou un fragmento fóra de contexto.

Unha ligazón non é verificación. Hai que abrir a fonte e comprobar o fragmento.

Requiren comprobación directa:

- Normas, artigos, prazos e importes.
- Nomes de órganos, persoas e procedementos.
- Cifras, porcentaxes e datas.
- Citas e referencias.
- Feitos que condicionan dereitos ou decisións.

7. Sesgos e efectos desiguais

O sesgo non aparece só nos datos de adestramento. Pode entrar por:

- Selección das fontes.
- Etiquetas e categorías.
- Definición do problema.
- Variables utilizadas.
- Métrica elixida.
- Prompt.
- Interface.
- Forma de usar a saída.

Un código postal pode actuar como substituto da renda ou da orixe. Un historial de inspeccións pode reflectir onde se inspeccionou máis, non onde había máis incumprimento. Unha media de precisión pode ocultar máis falsos positivos nun grupo.

Preguntas de revisión:

- Quen está representado e quen falta?
- Quen soporta os falsos positivos e negativos?
- Usamos unha variable innecesaria?

- Cambiaría a saída cunha característica protexida irrelevante?
- Existe unha alternativa menos intrusiva?
- Pode a persoa obter explicación e revisión?

Problema observado	Risco administrativo	Control verificable
Norma ou cifra inventada	Motivación baseada nun feito inexistente	Contraste coa fonte oficial e rexistro da corrección
Fonte recuperada fóra de contexto	Interpretación errónea do expediente	Mostrar fragmento, versión e localización exacta
Erro máis frecuente nun colectivo	Efecto desigual ou discriminación indirecta	Medir por grupos pertinentes e revisar variables e limiares
Aceptación sistemática da puntuación	Delegación material da decisión	Mostra de discrepancias, tempo de revisión e autoridade para apartarse
Falta de versión ou rexistro	Imposibilidade de reconstruír	Identificador de sistema, entrada, saída, revisión e decisión

O control debe deixar evidencia. Unha recomendación como revisar con coidado non permite saber se a revisión ocorreu nin se podía detectar o erro.

8. Responsabilidade por capas

A ferramenta non é un órgano administrativo. Non asina, non recibe a competencia e non motiva por si soa.

Poden concorrer:

- Provedor: deseño, documentación e cumprimento que lle corresponda.
- Entidade: contratación, configuración, organización, formación e control.
- Equipo técnico: integración, operación, seguridade e cambios segundo funcións.
- Persoa usuaria ou supervisora: uso conforme, revisión e comunicación de erros.
- Órgano competente: acto, decisión e motivación.
- Responsabilidade patrimonial, disciplinaria ou contractual segundo o caso.

Non todo recae sempre na persoa que asina. Tampouco pode esa persoa escusarse en que “o dixo a IA”. O expediente debe permitir coñecer fontes, intervencións, correccións e fundamento.

9. Protocolo antes de usar unha saída

1. Define finalidade, destinatario e efecto.
2. Identifica sistema, versión, fontes e datos.
3. Separa feitos, interpretacións, propostas e decisións.
4. Marca normas, cifras, nomes, prazos e citas.
5. Abre as fontes e comproba cada elemento crítico.
6. Busca omisións e compara coa entrada.
7. Revisa grupos afectados e efectos desiguais.
8. Confirma autorización de datos e contorno.
9. Comproba que unha persoa competente pode corrixir e deter.
10. Pide segunda revisión ou consulta se hai efecto xurídico ou risco elevado.
11. Rexistra aceptar, corrixir, escalar ou descartar.
12. Comunica erros repetidos, riscos e incidentes.

Caso aplicado

Un sistema propón priorizar inspeccións. Mostra unha puntuación e tres factores, pero non informa da calidade dos datos nin permite conservar unha explicación estable. O persoal acepta a orde porque o provedor anuncia alta precisión.

Problemas:

- Non se coñece se os datos representan todo o territorio.
- A precisión publicitada pode non medir falsos positivos por zona.
- A persoa non pode reconstruír a puntuación.
- Non hai rexistro suficiente.
- A interface favorece aceptación.
- Non está definida a autoridade para suspender.

A actuación non é “revisar máis”. Hai que pedir documentación, avaliar impacto, definir supervisión e deter se non se pode controlar o risco.

Recursos complementarios

- Páxina web da AESIA: https://aesia.digital.gob.es/es
- Portal das guías AESIA: https://aesia.digital.gob.es/es/guias
- Guías AESIA 4.0: https://guiasaesia.vercel.app/
- Regulamento (UE) 2024/1689: https://eur-lex.europa.eu/eli/reg/2024/1689/oj
- Guía AESIA sobre vixilancia humana: https://aesia.digital.gob.es/storage/media/06-guia-vigilancia-humana.pdf
- Manual de checklist AESIA: https://aesia.digital.gob.es/storage/media/16-manual-de-checklist-de-guias-de-requisitos.pdf