IA

ITV·IA
Auditorías técnicas AI Act

MUESTRA · NO ES UN INFORME REAL

Informe de auditoría técnica AI Act

Evaluación de cumplimiento del Reglamento (UE) 2024/1689 sobre los sistemas de Inteligencia Artificial en producción de **TalentMatch S.L.**

CLIENTE
TalentMatch S.L. (ficticio)

SECTOR
HR Tech · Selección de personal

PERIODO
17 – 21 de marzo de 2026

REFERENCIA
REF-2026-0142

EQUIPO AUDITOR
J. Okuja, A. Romero, L. Bernal

VERSIÓN
1.0 — Firmada

ÍNDICE DEL INFORME

Contenido

Alcance de la evaluación

La auditoría cubrió los **cuatro sistemas de IA** en producción identificados durante el kickoff: el motor de ranking de candidatos, el clasificador automático de CVs, el agente conversacional de prescreening y el módulo de scoring de empleabilidad. La revisión incluyó modelos, pipelines de datos, documentación interna, prácticas de gobernanza y entrevistas con responsables técnicos y de producto.

Documentos revisados

- Repositorios de los cuatro sistemas (acceso de solo lectura)
- Política interna de uso de IA (borrador v0.3)
- Datasets de entrenamiento y validación (muestras anonimizadas)
- Registros de inferencia de los últimos 90 días
- Contratos con proveedores de modelos fundacionales
- DPIA existente para la plataforma principal

SECCIÓN 01

Resumen ejecutivo

TalentMatch opera cuatro sistemas de IA que intervienen en decisiones de empleo de personas físicas en la Unión Europea. Tres de ellos quedan, en nuestra evaluación, dentro del ámbito de **alto riesgo** recogido en el Anexo III, punto 4 del Reglamento (UE) 2024/1689 (en adelante, «AI Act»), cuyas obligaciones aplican plenamente desde el **2 de agosto de 2026**.

62 / 100

Compliance score global · escala 0–100

Hallazgos críticos	3
Hallazgos altos	5
Hallazgos medios	7
Hallazgos bajos	4
Plazo objetivo	2 de agosto de 2026

Conclusiones principales

- **Tres sistemas serán clasificados como de alto riesgo** cuando entren en vigor las obligaciones del AI Act en agosto de 2026. No están preparados para una inspección externa hoy.
- **La trazabilidad de entrenamiento es insuficiente.** No es posible reconstruir, para el ranking de candidatos, qué datasets y qué versiones del código produjeron el modelo actualmente en producción.
- **No existe evaluación formal de sesgo** sobre género, edad o nacionalidad en ninguno de los modelos que toman decisiones sobre personas, pese a estar legalmente expuesta por la Directiva (UE) 2000/78 y la LISMI.
- **El registro de inferencia es parcial.** Solo el 41 % de las decisiones automatizadas de los últimos 90 días pueden reconstruirse con su contexto de entrada.
- La **documentación técnica** requerida por el Artículo 11 y el Anexo IV no existe en forma consolidada. Sería necesario producirla desde cero antes de cualquier auditoría externa.

> **Riesgo principal observado.** La combinación de (i) sistemas de alto riesgo en producción, (ii) ausencia de evaluación de sesgo y (iii) trazabilidad parcial supone una exposición simultánea bajo AI Act, GDPR y normativa laboral. Un único requerimiento de un cliente enterprise o un cliente institucional podría bloquear contratos en curso.

Decisiones recomendadas a la dirección

1. Aprobar un **plan de remediación de 14 semanas** con responsable ejecutivo identificado (sección 9).
2. Designar formalmente un **responsable de cumplimiento de IA** con autoridad sobre producto y datos.
3. Detener el despliegue del módulo de scoring de empleabilidad hasta completar la evaluación de impacto sobre derechos fundamentales (FRIA) prevista por el Artículo 27.

4. Iniciar la creación de la **documentación técnica del Anexo IV** para los tres sistemas de alto riesgo en paralelo al plan de producto.

SECCIÓN 02

Inventario de sistemas de IA

Sistemas identificados durante el kickoff y confirmados mediante revisión de código y entrevistas.

ID	SISTEMA	FUNCIÓN	MODELO / PROVEEDOR	PRODUCCIÓN
SYS-01	**Ranking de candidatos**	Ordena candidatos para un puesto en función de afinidad estimada.	Modelo propio (LightGBM) sobre features históricas	Sí
SYS-02	**Clasificador de CVs**	Extrae habilidades y clasifica adecuación a familias de puestos.	Modelo propio (transformer fine-tuned) sobre BERT base	Sí
SYS-03	**Agente de prescreening**	Conversa con candidatos para validar requisitos básicos.	API de proveedor externo (LLM general)	Sí
SYS-04	**Scoring de empleabilidad**	Estima probabilidad de éxito del candidato a 12 meses.	Modelo propio (gradient boosting) — beta interna	Beta

Volumen de decisiones automatizadas

SISTEMA	DECISIONES / MES	PERSONAS AFECTADAS / MES	PAÍSES UE EN USO
SYS-01 Ranking	~ 480.000	~ 92.000	ES, PT, FR, IT, DE
SYS-02 Clasificador	~ 1.200.000	~ 92.000	ES, PT, FR, IT, DE
SYS-03 Prescreening	~ 38.000	~ 38.000	ES, PT
SYS-04 Scoring	~ 2.400 (beta)	~ 2.400	ES

Observación. El volumen y el ámbito geográfico confirman que TalentMatch es un «proveedor» de sistemas de IA bajo el AI Act y, simultáneamente, un «responsable del despliegue» en sus propios procesos internos. Ambos roles conllevan obligaciones distintas que el cliente debe asumir en paralelo.

SECCIÓN 03

Clasificación de riesgo según AI Act

Resultado del análisis frente al Título III y Anexos del Reglamento (UE) 2024/1689.

SISTEMA	CATEGORÍA AI ACT	BASE DE LA CLASIFICACIÓN	OBLIGACIONES
SYS-01 Ranking	Alto riesgo Anexo III · 4(a)	Sistema utilizado para reclutamiento o selección, en particular para la criba de candidaturas.	Art. 8–17, 27, 49
SYS-02 Clasificador	Alto riesgo Anexo III · 4(a)	Mismo razonamiento que SYS-01: filtra candidaturas antes de la decisión humana.	Art. 8–17, 27, 49
SYS-03 Prescreening	Riesgo limitado	El agente conversa con personas físicas. Aplica la obligación de transparencia del Art. 50.	Art. 50
SYS-04 Scoring	Alto riesgo Anexo III · 4(b)	Toma decisiones que influyen en condiciones laborales (recomendación de contratación).	Art. 8–17, 27, 49

Implicaciones por sistema

Sistemas de alto riesgo (SYS-01, SYS-02, SYS-04)

- Obligación de **sistema de gestión de riesgos** (Art. 9) documentado y vivo.
- **Documentación técnica** según el Anexo IV (Art. 11), mantenida durante 10 años.
- **Trazabilidad y logging** de eventos relevantes (Art. 12).
- **Información y transparencia** hacia el responsable del despliegue (Art. 13).
- **Supervisión humana** efectiva (Art. 14).
- **Precisión, robustez y ciberseguridad** (Art. 15) con métricas declaradas.
- **Registro en la base de datos de la UE** antes de su puesta en servicio (Art. 49).
- **Evaluación de impacto sobre derechos fundamentales** (Art. 27) por parte del responsable del despliegue, cuando aplica.

Sistema de riesgo limitado (SYS-03)

El usuario debe ser informado de forma clara de que está interactuando con un sistema de IA. Esto se cumple parcialmente: el aviso existe en el chat, pero no en la convocatoria previa que el candidato recibe por correo.

> **Nota.** Aunque SYS-04 está en beta, está tomando decisiones sobre personas reales en producción restringida. Para efectos del AI Act, esto cuenta como puesta en servicio.

SECCIÓN 04

Transparencia y explicabilidad

Evaluamos qué puede TalentMatch explicar, hoy, sobre las decisiones automatizadas que produce, tanto frente al usuario final como frente a un cliente empresarial o un regulador.

CAPACIDAD	SYS-01	SYS-02	SYS-03	SYS-04
Aviso al usuario de que interactúa con IA	Parcial	No	Sí	No
Explicación individual de la decisión	No	No	Parcial	No
Feature importance global documentada	Parcial	No	N/A	Parcial
Política pública de funcionamiento	No	No	No	No
Mecanismo de revisión humana	Parcial	Parcial	Sí	No

Hallazgos relevantes

Falta de explicación individual de decisiones del ranking

F-04-01 · Crítico

Sistema: SYS-01 · Ranking de candidatos
Marco aplicable: AI Act Art. 13, 14 · GDPR Art. 22

El ranking no expone, ni a los reclutadores ni al candidato, las razones individuales por las que un perfil aparece más arriba que otro. No hay storage de feature contributions por inferencia. Esto impide tanto la supervisión humana significativa como el ejercicio de derechos del candidato bajo GDPR.

Sin aviso de uso de IA en la clasificación de CVs

F-04-02 · Alto

Sistema: SYS-02 · Clasificador
Marco aplicable: AI Act Art. 50

Los candidatos no son informados de que su CV es procesado por un sistema automático antes de la primera revisión humana. La política de privacidad menciona «sistemas tecnológicos» sin precisar.

SECCIÓN 05

Sesgo y equidad

Ejecutamos análisis preliminares sobre una muestra de 60.000 inferencias del último trimestre, comparando la tasa de selección por género estimado (a partir de nombre y, cuando estaba disponible, autodeclaración), edad declarada y nacionalidad declarada.

MÉTRICA	SYS-01 RANKING	SYS-02 CLASIFICADOR	UMBRAL INTERNO
Disparidad de selección por género (top-10)	**0,71**	0,82	≥ 0,80
Disparidad por nacionalidad (UE vs. no UE)	0,79	0,76	≥ 0,80
Disparidad por edad (<30 vs. ≥45)	0,68	0,73	≥ 0,80
Calibración global (Brier)	0,21	0,18	≤ 0,20

Hallazgo. El ranking presenta una disparidad de selección por género en el top-10 de 0,71, por debajo del umbral del 0,80 habitualmente referido en la práctica (regla del 80 %). Para candidatos mayores de 45 años, la disparidad cae a 0,68 en ambos sistemas. Esto no constituye, por sí solo, prueba de discriminación, pero sí riesgo material que TalentMatch no ha medido formalmente.

Limitaciones del análisis

- El género se infirió a partir del nombre cuando no estaba autodeclarado. Esto introduce un margen de error estimado del 6 %.
- No se dispone de datos de discapacidad ni de origen étnico, por restricciones legítimas bajo el GDPR.
- El análisis es retrospectivo y no incluye los cambios introducidos en el modelo en febrero de 2026.

Recomendaciones de mitigación

1. Establecer una **política formal de evaluación de sesgo** con métricas, umbrales y frecuencia (mensual).
2. Implementar **monitorización continua** con alertas cuando una métrica cae por debajo del umbral.
3. Diseñar un **protocolo de respuesta** cuando una alerta se dispara (suspender la versión, rollback, comunicación interna).
4. Documentar la **limitación intencional del modelo** y publicarla en la ficha técnica del Anexo IV.

SECCIÓN 06

Seguridad y datos

Superficie de ataque revisada

- Endpoints internos y públicos de los cuatro sistemas.
- Prompts del agente conversacional (SYS-03) e instrucciones del sistema.
- Cadena de dependencias del modelo propio: librerías, datasets, infraestructura.
- Acceso al proveedor externo de LLM (SYS-03) y al almacenamiento de logs.

VECTOR	SISTEMA	SEVERIDAD	COMENTARIO
Prompt injection vía CV malicioso	SYS-03	Crítico	El agente concatena el CV en el contexto sin sanitización. Se demostró extracción del system prompt.
Fuga de PII en logs de inferencia	SYS-01, SYS-02	Alto	Los logs contienen nombre, email y teléfono sin enmascarar. Retención de 180 días.
Dependencia única de proveedor LLM	SYS-03	Alto	Sin plan de contingencia ante cambio de términos o indisponibilidad.
Falta de rate limiting en API interna	SYS-01	Medio	Permite enumeración de candidatos por un usuario interno comprometido.
Datasets de entrenamiento sin firma	SYS-01, SYS-02, SYS-04	Medio	Sin verificación de integridad ante manipulación interna.
Modelo .pkl no firmado en despliegue	SYS-01	Bajo	Riesgo de sustitución silenciosa en el pipeline.

Prompt injection demostrable a través del CV

F-06-01 · Crítico

Sistema SYS-03 · Agente de prescreening

Marco aplicable AI Act Art. 15 · ENS · GDPR Art. 32

Insertando una sección «Notas para el evaluador automático» en el cuerpo del CV con instrucciones formuladas en lenguaje natural, conseguimos que el agente revelara su system prompt en el 4 de 10 intentos y que respondiera fuera del flujo previsto en el 7 de 10. Se recomienda aislar la entrada del candidato del contexto del sistema mediante un patrón *output-only*, sanitización y validación de la respuesta antes de su entrega al reclutador.

SECCIÓN 07

Brechas de documentación

Comparación contra los artefactos exigidos por el Anexo IV del AI Act (documentación técnica) y por el Artículo 11 (sistema de gestión de calidad).

ARTEFACTO REQUERIDO	FUENTE	ESTADO	BRECHA
Descripción general del sistema	Anexo IV(1)	Parcial	Existe README interno; falta ficha estructurada por sistema.
Descripción detallada de elementos y proceso de desarrollo	Anexo IV(2)	No	No existe versión escrita reconocible para los cuatro sistemas.
Información sobre el seguimiento, funcionamiento y control	Anexo IV(3)	No	Sin documentación operativa.
Sistema de gestión de riesgos	Art. 9 · Anexo IV(4)	No	No formalizado.
Gobernanza de datos y datasets	Art. 10	Parcial	Política de datos genérica; sin ficha de dataset por modelo.
Registros automáticos	Art. 12	Parcial	Logs existen pero parciales (41 % de inferencias reconstruibles).
Instrucciones de uso para el responsable del despliegue	Art. 13 · Anexo IV(5)	Parcial	Documentación de cliente orientada a producto, no a cumplimiento.
Plan de monitorización post-comercialización	Art. 72	No	No existe.
Declaración UE de conformidad	Art. 47 · Anexo V	No	Pendiente.
Registro en la base de datos UE	Art. 49	No	Pendiente.

Observación. La empresa cuenta con buen material técnico interno (READMEs, notebooks, decisiones en JIRA). El problema no es la falta de conocimiento, sino la falta de una capa documental orientada a auditoría y a cliente regulado.

SECCIÓN 08

Exposición regulatoria

Escenarios plausibles que pueden materializarse en los próximos 18 meses según la posición actual.

Escenario A · Requerimiento de un cliente enterprise

Un cliente del IBEX 35 incluye en su renovación anual un anexo de IA con preguntas sobre clasificación AI Act, evaluación de sesgo, FRIA, registro en la base de datos UE y plan de monitorización post-comercialización. TalentMatch no puede responder afirmativamente a 7 de las 12 preguntas. **Probabilidad estimada: alta** en los próximos 6 meses.

Escenario B · Reclamación individual ante la AEPD

Un candidato rechazado solicita explicación de la decisión automatizada bajo el Artículo 22 del GDPR. La empresa no puede aportar feature contributions ni la lógica subyacente del ranking. **Probabilidad estimada: media** en los próximos 12 meses.

Escenario C · Inspección sectorial coordinada

La AESIA, en coordinación con la AEPD y la Inspección de Trabajo, inicia una revisión sectorial sobre uso de IA en selección. **Probabilidad estimada: baja-media**, alta vez iniciada la aplicación de obligaciones de alto riesgo en agosto de 2026.

Resumen de exposición económica indicativa

ORIGEN	RANGO SANCIONADOR	COMENTARIO
AI Act · Art. 99(3)	Hasta 15 M€ o 3 % de la facturación global	Por incumplimiento de los Art. 8–17 sobre sistemas de alto riesgo.
AI Act · Art. 99(4)	Hasta 7,5 M€ o 1 % de la facturación global	Por información incorrecta a las autoridades.
GDPR · Art. 83(5)	Hasta 20 M€ o 4 % de la facturación global	Por incumplimiento del Art. 22, principios de tratamiento.
Pérdida comercial	Variable	Bloqueo en procurement de clientes enterprise.

Las cifras son los topes legales máximos previstos por las normas citadas. La exposición real depende del caso concreto y no debe interpretarse como una previsión.

SECCIÓN 09

Plan de remediación

Acciones priorizadas por impacto sobre el riesgo identificado, esfuerzo aproximado y dependencias. Las prioridades P0 son condición para presentar el sistema a un cliente enterprise o a un regulador.

SEMANAS 1–4 · P0

- Aislar el contexto del agente (SYS-03) frente a prompt injection.
- Enmascarar PII en logs y reducir retención.
- Aviso explícito de uso de IA al candidato.
- Designar responsable de cumplimiento de IA.
- Detener despliegue de SYS-04 hasta FRIA.

SEMANAS 5–10 · P1

- Evaluación formal de sesgo y monitorización continua.
- Trazabilidad completa de inferencias (objetivo: 95 %).
- Ficha técnica Anexo IV para los tres sistemas de alto riesgo.
- Sistema de gestión de riesgos documentado.
- Procedimiento de supervisión humana significativa.

SEMANAS 11–14 · P2

- FRIA de SYS-04 con dirección y RR. HH.
- Política pública de uso de IA y FAQ candidatos.
- Plan de monitorización post-comercialización.
- Borrador de declaración UE de conformidad.
- Auditoría interna previa al registro en la base de datos UE.

Detalle de acciones críticas

ID	ACCIÓN	RESPONSABLE PROPUESTO	ESFUERZO	PRIORIDAD
R–01	Sanitizar entrada del candidato en SYS-03 y verificar salida.	Tech Lead SYS-03 · Seguridad	1 sprint	P0
R–02	Enmascarar PII en logs de SYS-01 y SYS-02, reducir retención a 30 días.	Plataforma · Data	1 sprint	P0
R–03	Comunicar uso de IA en el flujo de aplicación al puesto.	Producto · Legal	0,5 sprint	P0
R–04	Designar Responsable de Cumplimiento de IA con mandato escrito.	CEO	—	P0
R–05	Implementar pipeline de monitorización de sesgo con alertas.	Data · Producto	2 sprints	P1
R–06	Ampliar logging para reconstrucción de cualquier inferencia.	Plataforma	2 sprints	P1
R–07	Producir ficha Anexo IV por sistema, versionada en repo de cumplimiento.	Tech Leads · ITV·IA	3 sprints	P1
R–08	Definir y formalizar el sistema de gestión de riesgos (Art. 9).	Responsable Cumplimiento	2 sprints	P1
R–09	FRIA de SYS-04 antes de pasar a producción general.	Responsable Cumplimiento · RR. HH.	1 sprint	P2
R–10	Preparar registro en la base de datos UE para sistemas de alto riesgo.	Responsable Cumplimiento	1 sprint	P2

SECCIÓN 10

Compliance score y conclusiones

El compliance score agrega seis dimensiones, cada una puntuada de 0 a 100 sobre los criterios habituales de auditoría técnica de IA. La ponderación refleja el peso atribuido por el AI Act y por la práctica reciente de procurement empresarial.

DIMENSIÓN	PESO	PUNTAJE	OBSERVACIÓN PRINCIPAL
Gobernanza y roles	15 %	**55**	Sin responsable formal asignado.
Trazabilidad y documentación	25 %	**48**	Documentación técnica Anexo IV ausente.
Datos y modelos	20 %	**66**	Buenas prácticas de ingeniería; faltan fichas de dataset.
Sesgo y equidad	15 %	**54**	Sin evaluación formal. Métricas iniciales por debajo del umbral.
Seguridad técnica	15 %	**70**	Hallazgo crítico en SYS-03 pero buen baseline general.
Transparencia hacia el usuario	10 %	**74**	Avisos parciales; falta política pública.

62 / 100

Compliance score global ponderado

Estado actual	**No apto para auditoría externa**
Objetivo a 14 semanas	**82 / 100**
Objetivo a 12 meses	**≥ 90 / 100**
Próxima revisión	**Q3 2026**

Conclusión final

TalentMatch dispone de la base técnica para alcanzar un nivel de cumplimiento alto antes del 2 de agosto de 2026, fecha en la que entran en vigor las obligaciones de los sistemas de alto riesgo. El obstáculo no es tecnológico, sino documental, organizativo y de medición. Ejecutar el plan de remediación propuesto sitúa a la empresa en posición de responder afirmativamente a un anexo de IA de un cliente enterprise y de pasar una primera revisión externa.

Firmado por el equipo auditor el 21 de marzo de 2026. ITV·IA · Madrid.

ANEXO A

Metodología y alcance

Marco de referencia

- Reglamento (UE) 2024/1689 — AI Act, en particular Título III, Anexos III, IV, V.
- Reglamento (UE) 2016/679 — GDPR, en particular artículos 22, 32 y 35.
- NIST AI Risk Management Framework 1.0 — aplicable como buena práctica.
- ISO/IEC 42001:2023 — sistema de gestión de IA, como referencia organizativa.

Actividades realizadas

- Kickoff de 90 minutos con CEO, CTO y Head of Product.
- Revisión de código de los cuatro sistemas con acceso de solo lectura.
- 5 entrevistas técnicas con responsables de ingeniería y datos.
- Revisión documental de políticas, contratos y DPIA existente.
- Pruebas técnicas: prompt injection, calidad de logs, reproducibilidad de modelos.
- Análisis estadístico sobre 60.000 inferencias de los últimos 90 días.
- Presentación final de 60 minutos con dirección.

Limitaciones

- El análisis se ha realizado sobre el estado de los sistemas durante la semana de auditoría. Cambios posteriores pueden alterar las conclusiones.
- Algunos componentes son operados por proveedores externos cuya evaluación profunda excede el alcance de esta semana.
- No se ha realizado pentest formal de la infraestructura cloud, fuera del alcance acordado.

Confidencialidad

Este informe es propiedad del cliente y está sujeto al acuerdo de confidencialidad firmado el 14 de marzo de 2026. La presente versión es una **muestra** con datos ficticios elaborada por ITV·IA con fines de presentación comercial.