Whitepaper de Seguridad

Enterprise Security Overview - AI Interview Analyzer

Proveedor:	AI Interview Analyzer Sp. z o.o.
Dirección:	ul. Jedrusik 6/53, 01-748 Warszawa, Poland
NIP:	5253079974
REGON:	54402118500000
Clasificación:	PUBLIC
Fecha:	24.06.2026

Contents

Whitepaper de Seguridad

Proveedor: AI Interview Analyzer Sp. z o.o., Warszawa, Poland

Audiencia: Equipos empresariales de seguridad, TI y compras

Clasificación: Pública

1. Resumen Ejecutivo

AI Interview Analyzer es una plataforma empresarial de contratación que graba entrevistas con el consentimiento explícito del candidato, las transcribe y estructura, y produce apoyo de evaluación basado en evidencia para los reclutadores. Debido a que la plataforma maneja datos personales de candidatos y respalda procesos de contratación, la seguridad y la privacidad se tratan como restricciones primarias de diseño, no como funcionalidades añadidas posteriormente.

Este whitepaper describe, en términos concretos y verificables, cómo protegemos los datos de clientes y candidatos. Está redactado para las personas que revisan proveedores: ingenieros de seguridad, administradores de TI, delegados de protección de datos y equipos de compras. Cada cifra de este documento se extrae directamente de nuestros propios sistemas de ingeniería en lugar de material de marketing.

El mensaje central es simple: **no nos limitamos a afirmar que la plataforma es segura, la ponemos a prueba continuamente para verificarlo.** Nuestra base de código contiene **3,171 pruebas automatizadas**, incluida una suite de seguridad dedicada que ejercita autenticación, autorización, aislamiento entre organizaciones, defensas contra inyecciones y eliminación de datos. Además, ejecutamos un arnés de pruebas de penetración repetible contra despliegues en vivo y producimos informes de auditoría por escrito. A lo largo de siete auditorías internas de seguridad en marzo y abril de 2026, registramos **zero critical findings**, y nuestra auditoría más reciente concluyó con un veredicto de **PASS**. (La certificación formal por terceros de estos controles forma parte de nuestra hoja de ruta; véase la Sección 18.)

Característica de seguridad	Resumen
Alojamiento	Microsoft Azure, solo regiones de la UE
Modelo de red	Endpoints privados, segmentación de red con denegación por defecto, sin base de datos pública
Cifrado	AES-256 en reposo, TLS 1.2 o superior en tránsito
Identidad	Tokens firmados de corta duración, hash de contraseñas con bcrypt, soporte de SSO
Control de acceso	Control de acceso basado en roles con aislamiento estricto por organización
Secretos	Bóveda centralizada de secretos con acceso mediante identidad administrada
Privacidad	Consentimiento explícito, retención configurable, borrado por unidad única
IA responsable	Solo apoyo a la decisión, humano siempre en el circuito
Aseguramiento	3,171 pruebas automatizadas más pruebas de penetración y auditorías recurrentes

1.1 Cómo leer este documento

Las Secciones 3 a 11 describen los controles que protegen los datos: arquitectura, red, identidad, aplicación, protección de datos, privacidad y ciclo de vida de desarrollo seguro. Las Secciones 12 y 13 cubren nuestro distintivo programa de pruebas continuas y nuestro historial de auditorías. Las Secciones 14 a 17 cubren operaciones, modelado de amenazas, gestión de vulnerabilidades y mapeo de cumplimiento. Los apéndices proporcionan un catálogo de controles, una FAQ para revisores y un glosario que un equipo de seguridad puede utilizar directamente durante una evaluación.

2. Alcance y enfoque del documento

2.1 Qué cubre este documento

Este whitepaper cubre la arquitectura y las prácticas de seguridad del servicio AI Interview Analyzer: el entorno de alojamiento, el diseño de red, la gestión de identidad y acceso, los controles a nivel de aplicación, la protección de datos, la privacidad y alineación regulatoria, el ciclo de vida de desarrollo seguro y nuestro programa continuo de pruebas de seguridad.

2.2 Qué lo hace verificable

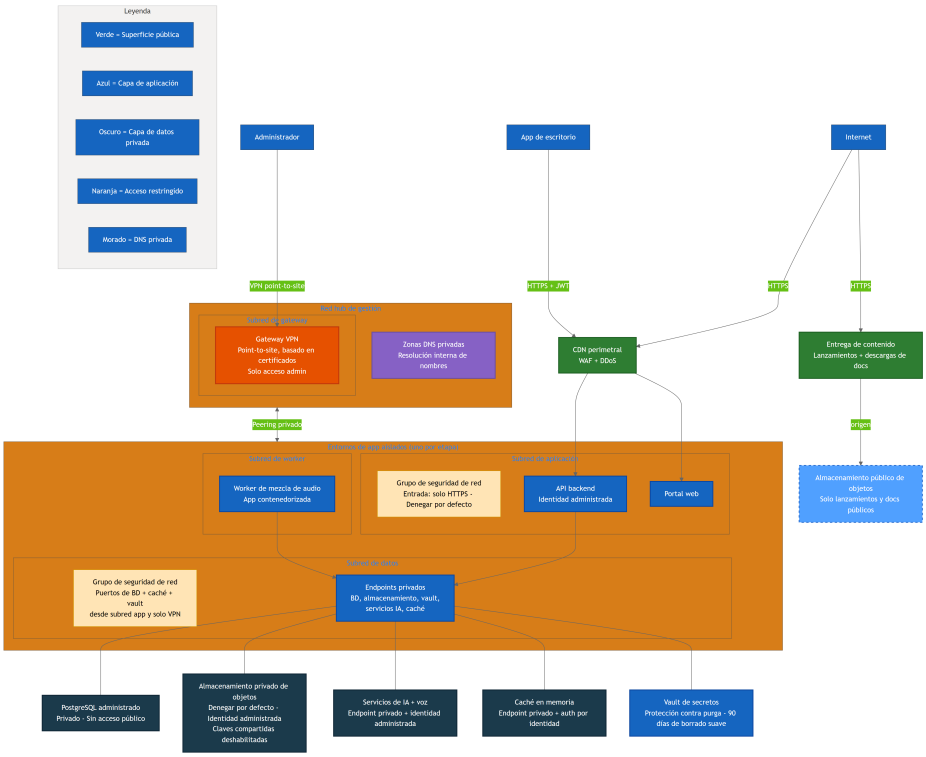
Las afirmaciones de seguridad de los proveedores son fáciles de escribir y difíciles de confiar. Por lo tanto, hemos vinculado cada afirmación principal de este documento a algo concreto y cuantificable dentro de nuestros sistemas de ingeniería: un control implementado en código, una prueba que demuestra que el control funciona, una definición de infraestructura que lo aplica o un informe de auditoría que registra una comprobación documentada. Cuando un control forma parte de nuestra hoja de ruta futura en lugar de estar implementado hoy, lo indicamos explícitamente. Preferimos afirmar menos y ser dignos de confianza que afirmar de más y ser desmentidos.

2.3 Responsabilidad compartida

La plataforma se entrega como software como servicio. Nosotros operamos la infraestructura, la aplicación, el pipeline de IA y el tratamiento de datos. El cliente es responsable de gestionar sus propias cuentas de usuario y roles, configurar las ventanas de retención de datos para que coincidan con su política interna y garantizar que el consentimiento del candidato se obtenga mediante el flujo de consentimiento que proporciona la plataforma. La Sección 14 describe esta división con más detalle.

3. Visión general de la arquitectura de seguridad

La plataforma está construida como un pequeño número de servicios cooperativos en lugar de un único monolito. Una aplicación de escritorio y un portal web actúan como clientes. Un backend API central gestiona toda la persistencia, autenticación, facturación, el pipeline de IA, consentimiento, correo electrónico, manejo de archivos y paneles. Un worker de fusión de audio procesa las grabaciones de forma asíncrona. Todo el estado sensible reside detrás del backend API; los clientes nunca se comunican directamente con la base de datos, el almacenamiento o los servicios de IA.



El diagrama anterior muestra la topología de producción con nombres de recursos intencionalmente generalizados. En él se observan tres principios:

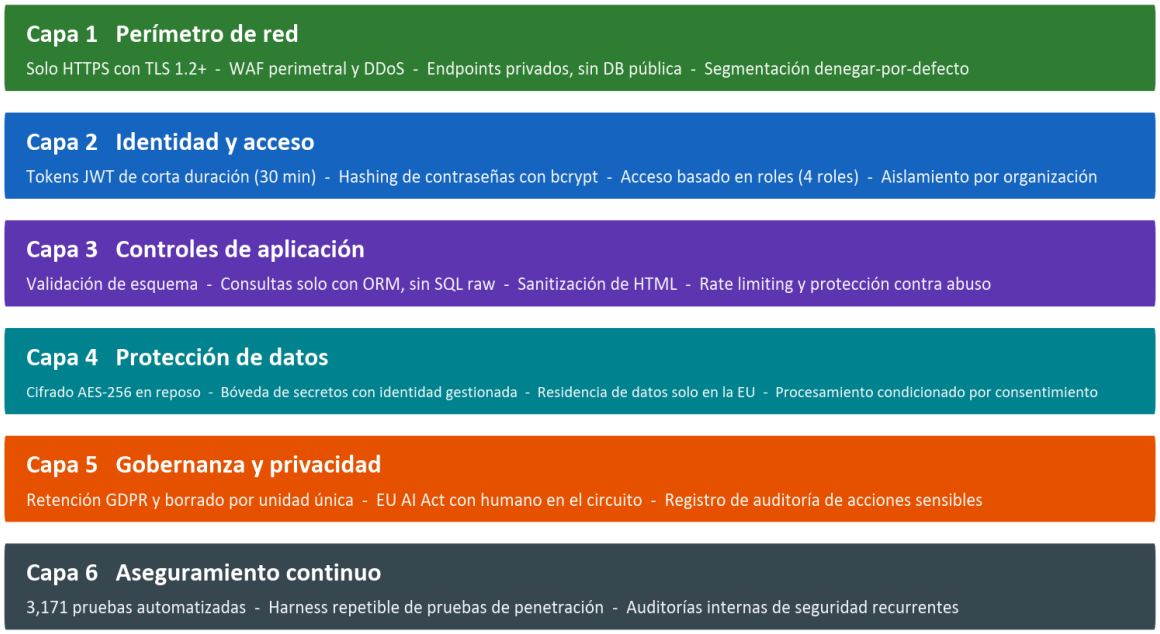
- **Sin exposición directa de los servicios de datos.** La base de datos, el almacenamiento privado de objetos, los servicios de IA y la caché tienen deshabilitado el acceso a red pública y solo son accesibles a través de endpoints privados dentro de una red virtual aislada. La bóveda de secretos es alcanzada por la aplicación a través de un endpoint privado y está además protegida por autenticación de identidad de la plataforma y políticas de acceso de privilegio mínimo, de modo que cualquier acceso requiere una identidad válida y autorizada independientemente de la ruta de red.
- **Una superficie pública separada.** El único almacenamiento público de objetos contiene descargas de versiones y documentos públicos. Nunca contiene datos de candidatos. El tráfico de la aplicación orientado al cliente pasa por una capa perimetral que proporciona web application firewall, protección contra distributed-denial-of-service y distribución de contenido.
- **El acceso administrativo está controlado.** Los operadores acceden a los recursos internos solo a través de un VPN punto a sitio basado en certificados hacia una red hub de gestión, no a través de internet pública.

Cada etapa de despliegue (desarrollo y producción) es un entorno totalmente aislado con su propia red, cuentas de almacenamiento, base de datos y secretos. Los datos de producción del cliente nunca están presentes en entornos inferiores. Un hub de gestión compartido contiene únicamente la puerta de enlace VPN y DNS privado, emparejado de forma privada con cada entorno.

4. Defensa en profundidad

No se confía en un único control para detener todos los ataques. La plataforma superpone controles independientes para que el fallo de una capa no exponga los datos. Las capas siguientes están cada una implementadas y, como se describe en la Sección 12, probadas individualmente.

Modelo de seguridad por capas: controles independientes en cada nivel



Capa	Controles representativos
Borde de red	Transporte solo con TLS, WAF perimetral y protección DDoS, endpoints privados, segmentación con denegación por defecto
Identidad y acceso	Tokens firmados de corta duración, hash con bcrypt, control de acceso basado en roles, aislamiento por organización
Aplicación	Validación de esquema en toda entrada, acceso a datos solo mediante ORM, codificación de salida, limitación de tasa
Protección de datos	Cifrado en reposo, bóveda de secretos con identidad administrada, residencia de datos en la UE, procesamiento condicionado por consentimiento
Gobernanza y privacidad	Retención configurable, borrado por unidad única, IA con humano en el circuito, registro de auditoría
Aseguramiento continuo	Suite de pruebas automatizadas, pruebas de penetración repetibles, auditorías internas de seguridad recurrentes

El resto de este documento recorre cada capa por turno y luego describe cómo demostramos, de forma continua, que las capas se mantienen.

5. Seguridad de red

5.1 Privado por defecto

La capa de datos es privada por construcción. La base de datos PostgreSQL administrada tiene deshabilitado el acceso a red pública y solo es accesible mediante un endpoint privado. El almacenamiento privado de objetos está configurado para denegar el acceso de red por defecto, deshabilita completamente las claves de acceso compartido y solo es accesible mediante identidad administrada desde la subred de la aplicación. La caché, los servicios de IA y la bóveda de secretos se alcanzan igualmente mediante endpoints privados con resolución DNS privada.

En la práctica, esto significa que no existe una cadena de conexión a la base de datos expuesta a internet ni una URL pública de almacenamiento para el audio de candidatos: la base de datos y el almacenamiento privado tienen el acceso a red pública deshabilitado por completo. La bóveda de secretos es alcanzada por la aplicación a través de un endpoint privado y está protegida por autenticación de identidad de la plataforma y políticas de acceso de privilegio mínimo, con identidades de aplicación a las que se concede acceso de solo lectura únicamente a los secretos que necesitan, por lo que los secretos no pueden recuperarse sin una identidad válida y autorizada. La superficie de ataque que un adversario externo puede siquiera tocar se limita a los endpoints HTTPS de la aplicación detrás de la capa perimetral.

5.2 Segmentación de red

Cada entorno está dividido en subredes separadas para la capa de aplicación, la capa de datos y el worker asíncrono. Cada subred está gobernada por un network security group cuya regla final deniega todo el tráfico entrante. La subred de aplicación acepta solo HTTPS entrante. La subred de datos acepta solo los puertos específicos de base de datos, caché y bóveda, y solo desde la subred de aplicación o la VPN administrativa. Esto significa que incluso un atacante que de algún modo alcanzara la capa de aplicación no podría pivotar libremente a la capa de datos; las únicas rutas permitidas son las que la aplicación utiliza legítimamente.

5.3 El borde

El tráfico público de la aplicación está al frente de una capa perimetral que proporciona web application firewall, protección DDoS y una CDN. Las descargas de versiones y documentos se sirven desde una cuenta de almacenamiento público dedicada a través de una puerta frontal de distribución de contenido, completamente separada del almacenamiento privado que contiene datos de candidatos. Los dos planos de almacenamiento nunca se mezclan: una mala configuración en el plano público no puede exponer datos privados de candidatos, porque son cuentas distintas con reglas de red diferentes.

5.4 Acceso administrativo

No existe un endpoint administrativo público hacia la red privada. Los operadores se conectan a través de una puerta de enlace VPN punto a sitio que utiliza autenticación basada en certificados. El acceso administrativo a la base de datos y la caché solo es posible desde dentro de ese túnel, ya que esos servicios tienen deshabilitado el acceso a red pública. Esto mantiene las operaciones cotidianas completamente fuera de internet pública.

6. Gestión de identidad y acceso

6.1 Autenticación

Las sesiones de usuario se establecen con un token de acceso firmado que es válido durante treinta minutos, emparejado con un refresh token separado, opaco y del lado del servidor. Los tokens de acceso se verifican en cada solicitud, y el usuario se vuelve a validar contra la base de datos (incluida una comprobación de cuenta activa) en lugar de confiar únicamente en el contenido del token. Cerrar sesión revoca inmediatamente la sesión de refresh del lado del servidor, de modo que un refresh token robado no puede sobrevivir a un cierre de sesión.

Las contraseñas nunca se almacenan en texto plano. Se aplican hashes con bcrypt utilizando una salt única por contraseña. Para las organizaciones que prefieren single sign-on, la plataforma admite inicio de sesión OAuth con Microsoft y Google, en cuyo caso no se almacena ninguna contraseña.

La titularidad del correo electrónico se verifica mediante un enlace de verificación de un solo uso y tiempo limitado antes de que una cuenta auto-registrada se considere verificada, y los reenvíos de correos de verificación están sujetos a limitación de tasa para prevenir abusos.

6.2 Control de acceso basado en roles

La autorización se aplica mediante un modelo de roles con cuatro roles de privilegio creciente: interviewer, hiring manager, recruiter y administrator. El acceso a operaciones privilegiadas se aplica mediante dependencias del lado del servidor que verifican tanto el rol como el estado de verificación del solicitante. Estas comprobaciones de rol protegen bastante más de cien operaciones API distintas.

Rol	Capacidades típicas
Interviewer	Realiza entrevistas asignadas; ve solo las entrevistas que tiene asignadas
Hiring manager	Gestiona los procesos de contratación que posee o de los que es miembro
Recruiter	Gestión completa de contrataciones y candidatos dentro de la organización
Administrator	Configuración de la organización, facturación, administración de usuarios y claves API

Más allá de las comprobaciones gruesas de roles, la plataforma aplica reglas de visibilidad a nivel de datos. Los hiring managers ven solo los procesos de contratación que crearon o de los que son miembros; los interviewers ven solo las entrevistas que tienen asignadas. Por lo tanto, el privilegio se aplica tanto al nivel de “qué acción” como al nivel de “qué registros”.

6.3 Aislamiento por organización

La plataforma es multi-tenant, y el aislamiento entre tenants se trata como un control de seguridad de primera clase. Cada identidad autenticada lleva un identificador de organización, y las consultas de datos se limitan a esa organización. Cuando un usuario solicita un registro que pertenece a otra organización, la plataforma devuelve una respuesta de “not found” en lugar de revelar que el registro existe. Los identificadores internos de la base de datos nunca se exponen en tránsito; la API presenta identificadores visibles y los vuelve a mapear por solicitud, lo que elimina una clase común de ataque de enumeración entre tenants.

Esto no es solo una intención de diseño. Como se describe en la Sección 12, nuestra suite automatizada ejecuta una gran matriz entre organizaciones que intenta alcanzar los datos de una organización usando credenciales de otra y verifica que cada uno de esos intentos falle.

6.4 Acceso programático

Para integraciones, las organizaciones de planes elegibles pueden emitir claves API. Las claves usan un prefijo reconocible, contienen 128 bits de entropía y se almacenan solo como hash; la clave en bruto se muestra una vez en el momento de su creación y nunca más. Cada clave lleva un ámbito de permisos explícito (read, write o integración ATS), puede restringirse a redes de origen específicas, puede revocarse instantáneamente y está sujeta a límites de tasa por clave derivados del nivel de plan de la organización. La verificación de claves utiliza una comparación segura en tiempo para evitar la filtración de información mediante el tiempo de respuesta.

7. Seguridad de la aplicación

La aplicación está escrita para eliminar categorías enteras de vulnerabilidad en lugar de corregirlas caso por caso.

- **Inyección.** Todo el acceso a la base de datos se realiza a través de un object-relational mapper con consultas parametrizadas. La base de código no contiene SQL en bruto formateado como cadena. Esto elimina estructuralmente la inyección SQL.
- **Validación de entrada.** Cada cuerpo de solicitud se valida frente a un esquema estricto antes de llegar a la lógica de negocio. Las cargas sobredimensionadas se rechazan, y los endpoints de lista están paginados para limitar el uso de recursos.
- **Codificación de salida y cross-site scripting.** El texto proporcionado por el usuario y generado por IA se trata como no confiable. Cuando el contenido debe renderizarse como HTML, pasa por un sanitizador de allow-list en el momento de escritura, y una suite de pruebas dedicada confirma que las etiquetas de script, los event handlers y las URL javascript se eliminan.
- **Mass assignment.** Las operaciones de actualización utilizan esquemas explícitos que excluyen campos privilegiados como rol, organización y saldo de créditos, de modo que un cliente no puede escalar privilegios publicando campos adicionales.
- **Rate limiting.** Los endpoints de autenticación y propensos al abuso están sujetos a limitación de tasa mediante un limitador duradero respaldado por base de datos que sobrevive a reinicios y funciona correctamente en múltiples instancias de aplicación. Login, registro, restablecimiento de contraseña y reenvíos de verificación tienen cada uno sus propios límites. La resolución de IP del cliente está reforzada frente a spoofing de forwarding headers.
- **Webhooks.** Los webhooks entrantes de proveedores de pago y correo electrónico se verifican frente a las firmas del proveedor sobre el cuerpo bruto de la solicitud antes de ser procesados.
- **Carga de archivos.** Las cargas tienen límite de tamaño, se validan, se almacenan bajo identificadores generados en lugar de nombres proporcionados por el usuario y están restringidas por solicitud y por organización.
- **Security headers.** En producción, las respuestas incluyen strict transport security, opciones de content-type y frame, una política de referrer y una permissions policy restrictiva, y suprimen los banners del servidor y framework.

8. Protección de datos

8.1 Cifrado

Todos los datos están cifrados en reposo usando AES-256 a través de las capas de cifrado de almacenamiento y base de datos de la plataforma Azure. Todo el tráfico de red se sirve exclusivamente sobre HTTPS usando TLS 1.2 o superior; HTTP en texto plano se redirige a HTTPS en todas las capas. En producción, el API y el portal web emiten headers de strict transport security junto con un conjunto de headers de endurecimiento, y suprimen los banners de versión del servidor y framework.

8.2 Gestión de secretos

Los secretos de la aplicación se mantienen en una bóveda centralizada de secretos con protección contra purge habilitada y una ventana de soft-delete de noventa días. Las aplicaciones se autentican frente a recursos Azure usando system-assigned managed identities en lugar de claves de larga duración; por ejemplo, el almacenamiento privado tiene las claves de acceso compartido deshabilitadas por completo, de modo que el acceso solo es posible mediante asignaciones de roles basadas en identidad limitadas al recurso individual. Las políticas de acceso a la bóveda conceden a los principales de aplicación acceso de solo lectura a los secretos específicos que necesitan, siguiendo el principio de privilegio mínimo.

8.3 Residencia de datos

Todos los datos de clientes y candidatos se almacenan y procesan dentro de la Unión Europea. El alojamiento de la aplicación, la base de datos, el almacenamiento, la caché y los secretos residen en West Europe, y el procesamiento de IA se ejecuta en regiones de la UE. El proveedor de IA no utiliza datos del cliente para entrenar sus modelos.

8.4 La vida de una sola entrevista

La forma más clara de entender los controles de protección de datos es seguir una entrevista de principio a fin. El consentimiento se captura y registra antes de que se procese cualquier cosa. La carga se cifra en tránsito. La transcripción y el análisis se ejecutan dentro de centros de datos de la UE. Los resultados se escriben en almacenamiento cifrado. Cada registro queda entonces gobernado por un único reloj de retención que termina en una eliminación en cascada registrada. En cualquier punto, los derechos del candidato como retirada, eliminación, acceso o portabilidad pueden interrumpir este flujo.

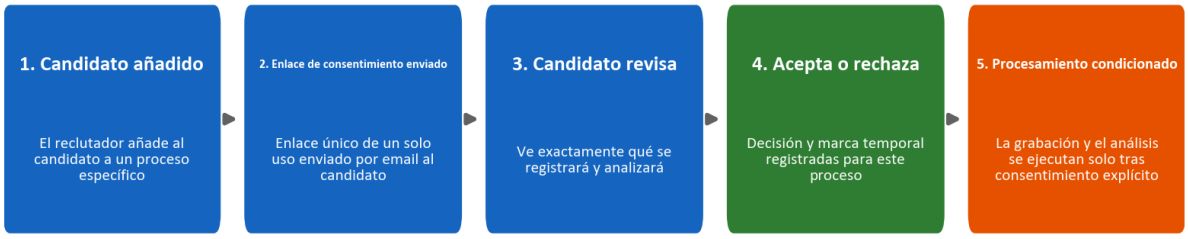
9. Privacidad desde el diseño y GDPR

La privacidad está integrada en el modelo de datos y el flujo de trabajo, no añadida únicamente mediante políticas.

9.1 Consentimiento

Ninguna entrevista se graba ni se analiza sin el consentimiento explícito del candidato. Cuando se añade un candidato a un proceso de contratación, la plataforma emite por correo electrónico un enlace de consentimiento único y de un solo uso. El candidato revisa qué ocurrirá y acepta o rechaza. El estado del consentimiento, incluida la hora de respuesta, se registra frente a ese proceso de contratación específico, de modo que el consentimiento siempre se limita a un proceso de contratación concreto en lugar de concederse globalmente.

Consentimiento del candidato: explícito y registrado antes de cualquier procesamiento



9.2 Retención y borrado

La retención de datos es configurable por organización, con un valor predeterminado de doce meses y un mínimo configurable de treinta días, y puede sobrescribirse por candidato. Existe un único reloj de retención para los datos de un candidato, no un temporizador separado por artefacto. El reloj comienza cuando se registra una decisión de contratación. Antes de que los datos expiren, la plataforma envía una advertencia (por defecto unos quince días antes) y ofrece una extensión con un solo clic. Cuando los datos se eliminan, se eliminan como una única unidad: el registro del candidato, entrevistas, transcripciones, grabaciones de audio, documentos y comparaciones se eliminan todos juntos, y la eliminación se registra en un log de auditoría. No queda residuo parcial ni huérfano.

El ciclo de vida a continuación muestra este reloj único y cómo converge en una eliminación en cascada única con una prueba registrada de borrado.

Retención de datos: un plazo por candidato, borrado por unidad única

9.3 Derechos del interesado y subencargados

La plataforma admite los derechos del interesado exigidos por el GDPR, incluidos acceso, eliminación, portabilidad, objeción y explicación. El procesamiento se lleva a cabo en virtud de un data processing agreement que los clientes aceptan en el registro y que se versiona por organización. Nuestros subencargados y sus funciones, todos dentro de la UE o bajo salvaguardas apropiadas, se revelan en ese acuerdo, y los clientes reciben aviso previo de cualquier cambio. La Sección 17 contiene el registro de subencargados y el mapeo de cumplimiento artículo por artículo.

10. IA responsable y el EU AI Act

La plataforma entra dentro de la categoría de alto riesgo del EU AI Act porque respalda decisiones de empleo, y tratamos esa clasificación con seriedad.

La regla definitoria del producto es que **la IA es apoyo a la decisión, no quien decide.** El sistema nunca acepta ni rechaza automáticamente a un candidato. Transcribe el habla, estructura preguntas y respuestas, puntúa respuestas frente a criterios que el reclutador ha definido y redacta feedback, y un humano revisa cada salida antes de que se utilice. Esto mantiene firmemente a una persona en el circuito.

Igualmente importante es lo que la IA no hace. No evalúa personalidad, "ajuste cultural", estado emocional, tono de voz, acento, género, edad, etnia, apariencia ni lenguaje corporal. La puntuación se ancla a evidencia de la transcripción y a criterios definidos por el reclutador, y los nombres de los candidatos se excluyen de la entrada de evaluación para reducir el sesgo. Publicamos una tarjeta de transparencia, documentación de usuario y una declaración de conformidad que describen el sistema, sus limitaciones y sus salvaguardas.

Control de IA responsable	Cómo funciona
Humano en el circuito	Cada puntuación y cada pieza de feedback es revisada por un reclutador antes de su uso
Sin decisiones automatizadas	El sistema nunca autoacepta ni autorechaza a un candidato
Puntuación basada en evidencia	Las puntuaciones hacen referencia a evidencia de apoyo de la transcripción
Diseño anti-sesgo	Nombres excluidos de la evaluación; se puntúa la sustancia por encima del estilo
Límites de alcance	Nunca se evalúan personalidad, emoción, acento ni características protegidas
Seguridad del feedback al candidato	El feedback privado al candidato pasa por una barrera de seguridad de generación y validación

Estas restricciones no solo se indican en la documentación; están codificadas en la capa de prompts de IA y se ejercitan mediante un programa de pruebas de seguridad de IA dedicado descrito en la Sección 12.3.

11. Ciclo de vida de desarrollo seguro

La seguridad se aplica en la forma en que construimos y entregamos software, no solo en el sistema en ejecución.

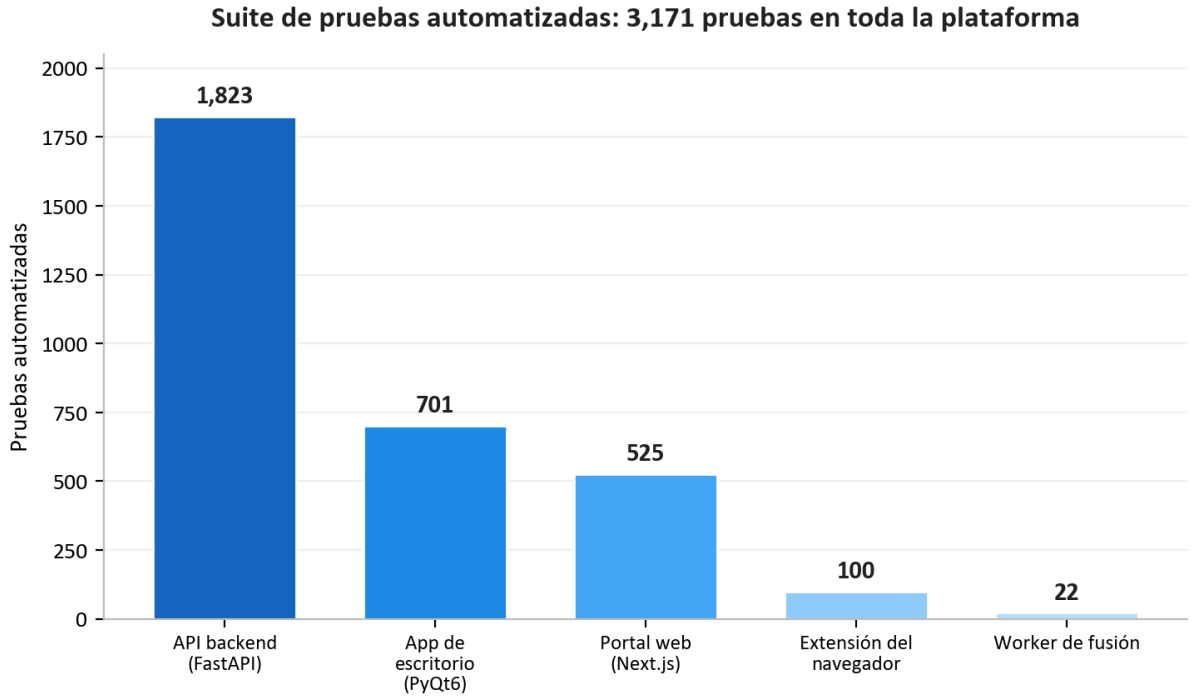
- **Separación de entornos.** Desarrollo y producción están completamente separados, cada uno con su propia infraestructura, cuentas de almacenamiento, base de datos, secretos y subdominios. No existe estado compartido.
- **Infraestructura como código.** Todo el entorno cloud se define como código y se revisa como código, lo que hace que la postura de seguridad sea auditable y reproducible. Un revisor puede leer exactamente qué puertos están abiertos, qué recursos son privados y qué identidades tienen qué permisos.
- **Despliegues fijados y controlados.** Cada paso en el pipeline de integración continua está fijado a una versión exacta e inmutable. Los despliegues de producción se basan en tags, se ejecutan solo a través del pipeline de producción protegido y están controlados por una aprobación obligatoria. La suite de pruebas automatizadas se ejecuta como control de liberación: un despliegue no puede publicarse si las pruebas fallan.
- **Higiene de dependencias.** La supervisión automatizada de dependencias propone actualizaciones semanalmente en el backend, escritorio, web, infraestructura y definiciones de pipeline, y las auditorías de dependencias forman parte de nuestra revisión periódica de seguridad.
- **Artefactos firmados.** Los instaladores de escritorio están firmados con código, para que los clientes puedan verificar que el software que instalan realmente procede de nosotros.
- **Disciplina de secretos.** Los secretos viven en la bóveda y en secretos protegidos del pipeline, nunca en el código fuente.

12. Pruebas continuas de seguridad

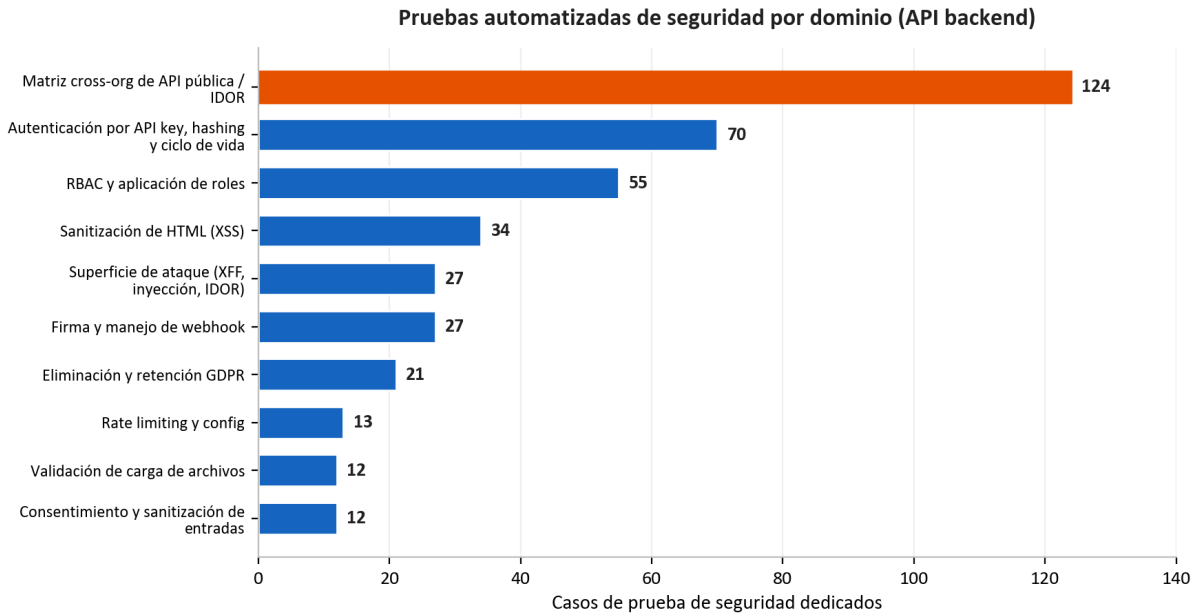
Este es el núcleo de nuestra historia de aseguramiento y la parte que la mayoría de proveedores no puede mostrar. Tratamos la seguridad como algo que debe medirse continuamente, con comprobaciones ejecutables, en lugar de afirmarse una sola vez.

12.1 La suite de pruebas automatizadas

La plataforma está cubierta por **3,171 pruebas automatizadas** que abarcan el backend API, la aplicación de escritorio, el portal web, la extensión del navegador y el worker de fusión de audio.



No se trata solo de pruebas funcionales. Una suite de seguridad sustancial y dedicada ejercita los controles descritos anteriormente en este documento. El gráfico de abajo desglosa las pruebas específicas de seguridad en el backend API por dominio.



Entre muchas otras, esta suite incluye una gran matriz de API pública que ejecuta cada endpoint como un usuario legítimo, como la propia clave API de la organización y como la clave API de una organización rival, verificando que todo intento entre organizaciones sea bloqueado. Incluye docenas de pruebas adversariales de superficie de ataque para spoofing de forwarding headers, header injection y fuga de identificadores, una suite enfocada de sanitización HTML para cross-site scripting, pruebas de aplicación de roles para el modelo completo de roles y pruebas que demuestran que los datos del candidato realmente se eliminan como una unidad. Debido a que estas pruebas se ejecutan como control de liberación, una regresión que debilitara cualquiera de estos controles detendría la liberación en lugar de llegar a los clientes.

12.2 Pruebas de penetración en vivo

Las pruebas unitarias automatizadas demuestran que los controles se comportan correctamente de forma aislada. Para demostrar que se sostienen conjuntamente en un despliegue real, mantenemos una metodología repetible de pruebas de penetración que ejecuta scripts de ataque reales contra un entorno en vivo. Está organizada en seis fases:

Fase	Enfoque	Ejemplos de lo que se ejercita
1. Análisis estático	Código fuente	Secretos, patrones de inyección, funciones peligrosas, auth ausente, HTML inseguro
2. Revisión de arquitectura	Infraestructura	Endpoints privados, segmentación, TLS, configuración de secretos
3. Análisis de vectores de ataque	Control de código fuente y cloud	Protección de ramas, alcance de identidad, exposición pública
4. Pruebas de penetración en vivo	Entorno en ejecución	Sondeo sin autenticar, acceso entre organizaciones, inyección, manipulación de tokens, SSRF, ráfagas de rate-limit
5. Puntuación empresarial	Madurez	Dieciséis categorías de seguridad puntuadas frente a una línea base empresarial
6. Dependencias y cadena de suministro	Riesgo de terceros	Auditoría de CVE de dependencias, acciones de pipeline fijadas, integridad de lock-file

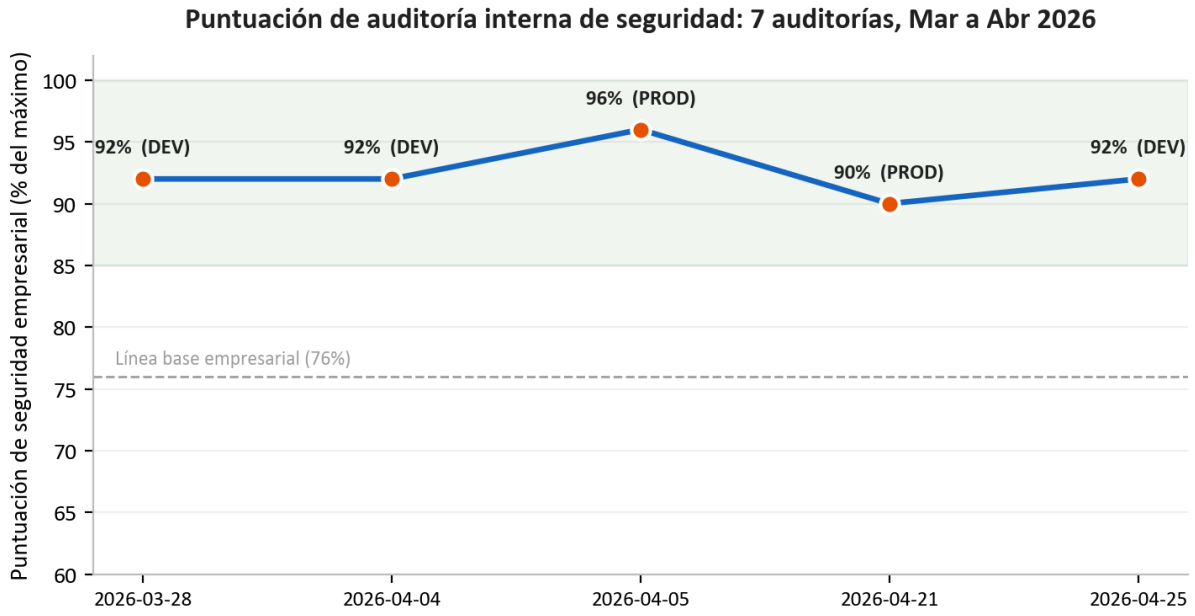
La Fase 4 es una prueba adversarial genuina contra un sistema desplegado, no una checklist. Sondea endpoints protegidos sin credenciales y confirma que rechazan el acceso; registra dos organizaciones e intenta alcanzar los registros de una organización con la cuenta de la otra; inyecta cargas de cross-site-scripting y server-side-template y confirma que son neutralizadas; manipula tokens de autenticación y confirma que son rechazados; intenta server-side request forgery contra endpoints de metadatos cloud; y lanza ráfagas contra endpoints de autenticación para confirmar que la limitación de tasa realmente se activa en el entorno en vivo, no solo en teoría.

12.3 Pruebas de seguridad del feedback al candidato

Debido a que la plataforma puede generar feedback privado de desarrollo para candidatos, ejecutamos un programa adversarial de seguridad separado contra esa funcionalidad. Alimenta deliberadamente al sistema con notas de reclutadores duras y hostiles y confirma que la salida orientada al candidato nunca contiene vulgaridades, nunca revela ni atribuye la identidad o la opinión privada de un reclutador y nunca aplica etiquetas de personalidad de carácter valorativo. Esto protege tanto al candidato, que debe recibir feedback constructivo y respetuoso, como al cliente, cuya opinión interna nunca debería filtrarse hacia el exterior.

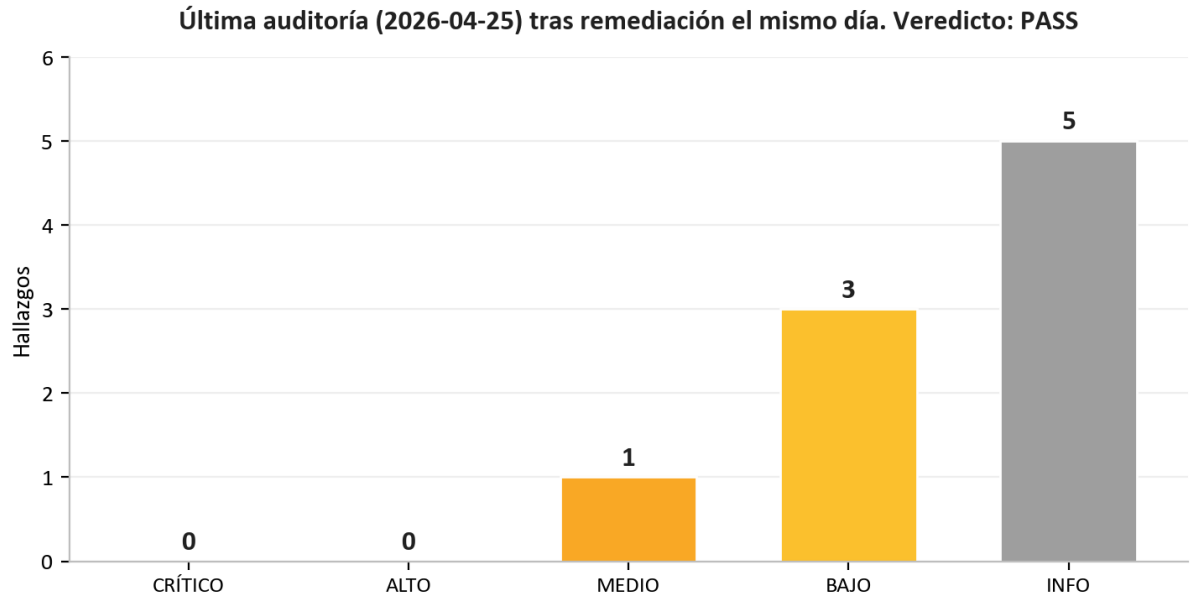
13. Resultados de auditoría de seguridad

Realizamos auditorías de seguridad recurrentes usando una metodología estructurada y repetible de pruebas de penetración, y redactamos cada una como un informe fechado con hallazgos clasificados por severidad, evidencia y remediación. Se trata de auditorías internas ejecutadas por nuestro propio proceso de seguridad; la certificación formal por terceros de los mismos controles forma parte de nuestra hoja de ruta. Entre finales de marzo y finales de abril de 2026 completamos **seven such audits** en desarrollo y producción.



El resultado que más importa a un cliente potencial es la consistencia: **across all seven audits there were zero critical findings.** En las raras ocasiones en que apareció un problema de mayor severidad, se remedió rápidamente, a menudo el mismo día, y se volvió a verificar. La rúbrica de puntuación se endureció deliberadamente durante este periodo (la puntuación máxima posible se elevó a medida que añadíamos más categorías para evaluar), razón por la cual la línea de puntuación normalizada se mantiene alta incluso cuando el listón subía.

Nuestra auditoría más reciente, el 25 April 2026, ilustra cómo funciona el proceso en la práctica. Se identificaron dos problemas de mayor severidad, ambos se corrigieron y verificaron de nuevo el mismo día, y la auditoría se cerró con un veredicto de **PASS** sin problemas explotables restantes en el modelo de amenazas actual.



Auditoría	Entorno	Critical	Veredicto
2026-03-28	Desarrollo	0	Lista para producción
2026-04-04	Desarrollo	0	Lista para empresa
2026-04-05	Producción	0	Lista para empresa
2026-04-20	Desarrollo	0	Lista para producción, notas
2026-04-20	Desarrollo	0	Aprobado con notas
2026-04-21	Producción	0	Seguro, sin hallazgos explotables
2026-04-25	Desarrollo	0	Aprobado

El patrón a lo largo de estas auditorías es la evidencia más honesta que podemos ofrecer: se encuentran problemas, porque los buscamos con rigor, y se cierran rápidamente, porque el proceso está diseñado para cerrarlos. Un proveedor que nunca informa de un hallazgo suele ser un proveedor que no está buscando.

---

14. Resiliencia operativa y responsabilidad compartida

14.1 Supervisión y registro

La telemetría de la aplicación y la plataforma fluye hacia un espacio de trabajo centralizado de análisis de logs y un servicio de supervisión de aplicaciones, lo que nos proporciona visibilidad sobre disponibilidad y comportamiento. Las acciones sensibles como eliminación de datos, aceptación de acuerdos legales e invocaciones de IA se registran en tablas de auditoría dedicadas, de modo que exista un registro duradero de quién hizo qué con datos importantes.

14.2 Copia de seguridad y recuperación

La base de datos administrada conserva copias de seguridad automatizadas, y el almacenamiento privado está protegido mediante retención de soft-delete tanto en blobs como en contenedores, por lo que la eliminación accidental o maliciosa puede recuperarse dentro de la ventana de retención. La infraestructura crítica lleva bloqueos de eliminación para prevenir el desmantelamiento accidental de recursos de producción.

14.3 Resumen de responsabilidad compartida

Área	AI Interview Analyzer	Cliente
Infraestructura, red, parcheo	Sí	-
Seguridad de la aplicación y pipeline de IA	Sí	-
Cifrado, secretos, residencia de datos	Sí	-
Administración de usuarios y roles	Proporciona los controles	Gestiona usuarios y roles
Configuración de la política de retención	Proporciona los controles	Establece la ventana de retención
Consentimiento del candidato	Proporciona el flujo	Garantiza su uso
Credenciales sólidas de usuario final y SSO	Admite SSO y política	Aplica la política interna

15. Modelo de amenazas y mapeo OWASP

Diseñamos frente a un conjunto concreto de adversarios: un atacante externo sin credenciales, un usuario autenticado curioso o malicioso de una organización que intenta alcanzar los datos de otra organización, una dependencia comprometida y un error interno. La siguiente tabla mapea las categorías de riesgo ampliamente utilizadas del OWASP Top 10 a los controles específicos que las abordan en esta plataforma, cada uno de los cuales se ejercita mediante las pruebas descritas en la Sección 12.

Riesgo OWASP	Cómo lo mitiga la plataforma
Broken access control	Control de acceso basado en roles en cada endpoint privilegiado; limitación por organización; “not found” en acceso entre organizaciones; remapeo de identificadores; matriz de pruebas entre organizaciones
Cryptographic failures	TLS 1.2+ en tránsito; AES-256 en reposo; hash de contraseñas con bcrypt; secretos en una bóveda administrada
Injection	Consultas parametrizadas solo con ORM; validación estricta de esquema; sanitización HTML en el momento de escritura
Insecure design	Defensa en profundidad por capas; modelado de amenazas y revisión de arquitectura en cada auditoría
Security misconfiguration	Infraestructura como código; grupos de red con denegación por defecto; security headers; claves compartidas de almacenamiento deshabilitadas; esquema API no expuesto en producción
Vulnerable components	Supervisión automatizada semanal de dependencias; auditorías de CVE de dependencias en revisión periódica
Identification and authentication failures	Tokens de corta duración; login con limitación de tasa; verificación de correo electrónico; soporte de SSO; sin contraseñas en texto plano
Software and data integrity failures	Pasos de pipeline fijados e inmutables; instaladores de escritorio firmados; verificación de firma de webhooks; despliegues de producción controlados por tags
Security logging and monitoring failures	Telemetría centralizada; tablas de auditoría dedicadas para acciones sensibles
Server-side request forgery	Llamadas salientes restringidas a endpoints de confianza; sondeos de SSRF en el arnés de pruebas de penetración

Este mapeo es la columna vertebral de nuestro argumento de aseguramiento: para cada clase bien conocida de ataque existe un control con nombre, y para cada control con nombre existe una prueba.

16. Gestión de vulnerabilidades y divulgación responsable

La seguridad nunca está terminada, por lo que ejecutamos un bucle continuo de descubrimiento y remediación.

- **Descubrimiento.** Las vulnerabilidades afloran a partir de cuatro fuentes: la suite de pruebas automatizadas, las auditorías recurrentes de pruebas de penetración, la supervisión automatizada de dependencias y los informes de clientes o investigadores.
- **Triaje.** A cada hallazgo se le asigna una severidad (critical, high, medium, low o informational) con evidencia y un responsable de remediación, exactamente como se registra en nuestros informes de auditoría.
- **Objetivos de remediación.** Los hallazgos critical y high se priorizan para remediación inmediata; en nuestro historial de auditoría, los hallazgos de mayor severidad normalmente se han resuelto y verificado de nuevo el mismo día. Los hallazgos medium e inferiores se programan dentro del ciclo regular de mantenimiento.
- **Verificación.** Las correcciones se vuelven a probar y, cuando corresponde, se ejecuta una comprobación en vivo contra el entorno desplegado para confirmar que el problema realmente está cerrado, no solo cerrado en código.
- **Divulgación.** Las preocupaciones de seguridad pueden notificársenos directamente. Confirmamos la recepción de los informes, investigamos y mantenemos informado al reportante hasta su resolución.

17. Mapeo de cumplimiento

17.1 GDPR

Área GDPR	Implementación en la plataforma
Base jurídica (Art. 6)	Consentimiento explícito del candidato capturado antes del procesamiento
Minimización de datos y limitación del almacenamiento (Art. 5)	Solo se procesan datos relevantes para la entrevista; retención configurable con eliminación automática
Derecho de supresión (Art. 17)	Eliminación por unidad única de todos los datos del candidato, con prueba registrada de borrado
Derechos del interesado (Art. 15 to 20)	Se admiten acceso, eliminación, portabilidad y objeción
Obligaciones del encargado (Art. 28)	Data processing agreement aceptado en el registro y versionado por organización
Seguridad del tratamiento (Art. 32)	Cifrado, control de acceso, aislamiento y pruebas continuas como se describe en este documento
Transparencia de subencargados	Divulgada en el data processing agreement con aviso previo de cambios

17.2 EU AI Act

La plataforma se trata como un sistema de IA de alto riesgo que respalda decisiones de empleo, y mantenemos documentación alineada con el reglamento, incluida una tarjeta de transparencia, documentación de usuario y una declaración de conformidad. Las salvaguardas principales, supervisión humana, transparencia, puntuación basada en evidencia y límites estrictos sobre qué evalúa la IA, se describen en la Sección 10. Seguimos madurando nuestra documentación formal de conformidad a medida que avanza la línea temporal de implantación del reglamento.

17.3 Certificaciones de alojamiento

La plataforma se ejecuta enteramente en Microsoft Azure, cuyos centros de datos cuentan con certificaciones independientes, incluidas ISO 27001 y SOC 2. Estas certificaciones cubren las capas físicas y de plataforma situadas bajo nuestra aplicación; los controles a nivel de aplicación son los descritos a lo largo de este documento.

17.4 Registro de subencargados

Subencargado	Propósito	Región
Microsoft Azure	Alojamiento, procesamiento de IA y voz, almacenamiento, correo electrónico transaccional	EU (West Europe, Sweden Central)
Stripe	Procesamiento de suscripciones y pagos	EU (Ireland)
Fakturownia	Facturación	EU (Poland)
Conector ATS (opcional)	Integración con applicant-tracking, habilitada solo bajo solicitud	EU

18. Hoja de ruta de seguridad

Tratamos la seguridad como un programa de mejora continua. Las iniciativas actuales de nuestra hoja de ruta incluyen reforzar las opciones de autenticación multifactor para cuentas administrativas, ampliar el registro centralizado de auditoría del acceso a datos, seguir endureciendo regularmente la actualización de dependencias y avanzar en la certificación formal por terceros de los controles descritos en este documento. Ninguna de estas es una brecha que exponga hoy los datos del cliente; cada una es una mejora de una postura ya estratificada.

19. Resumen

AI Interview Analyzer protege los datos de candidatos y clientes mediante una arquitectura por capas: una red privada por defecto sin servicios de datos públicos, identidad sólida y aislamiento por organización, código de aplicación que elimina clases enteras de vulnerabilidad, cifrado y residencia de datos en la UE, y controles de privacidad integrados en el modelo de datos. Lo que distingue a la plataforma es la evidencia detrás de esas afirmaciones. Con 3,171 pruebas automatizadas, una metodología repetible de pruebas de penetración en vivo, un programa dedicado de seguridad de IA y un historial de siete auditorías internas de seguridad con zero critical findings, podemos demostrar, no solo afirmar, que la plataforma es segura.

Appendix A: Catálogo de controles de seguridad

Una referencia condensada de los controles principales y la evidencia que respalda cada uno.

Control	Mecanismo	Evidencia
Cifrado de transporte	Solo HTTPS, TLS 1.2+, HTTP redirigido	Infraestructura como código; auditoría de arquitectura
Cifrado en reposo	Cifrado de plataforma AES-256 en almacenamiento y base de datos	Configuración de plataforma; auditoría de arquitectura
Protección de contraseñas	bcrypt con salt por contraseña	Control de código fuente; pruebas de autenticación
Gestión de sesiones	Tokens firmados de 30 minutos, refresh revocable del lado del servidor	Control de código fuente; pruebas de autenticación
Autorización	Control de acceso de cuatro roles en endpoints privilegiados	Suite de pruebas de aplicación de roles
Aislamiento de tenant	Limitación de consultas por organización; 404 en acceso entre organizaciones	Matriz de pruebas entre organizaciones
Seguridad de claves API	Almacenamiento con hash, permisos limitados, límites de tasa por clave	Suite de pruebas de claves API
Defensa contra inyección	Consultas parametrizadas solo con ORM	Análisis estático; pruebas de inyección
Defensa contra cross-site scripting	Sanitización HTML en el momento de escritura	Suite de pruebas de sanitización HTML
Rate limiting	Limitador duradero respaldado por base de datos en endpoints de auth	Pruebas de rate-limit; comprobaciones de ráfaga en vivo
Integridad de webhooks	Verificación de firma del proveedor sobre cuerpo bruto	Suite de pruebas de webhooks
Gestión de secretos	Bóveda administrada, protección contra purge, identidad administrada	Infraestructura como código; auditoría de arquitectura
Aislamiento de red	Endpoints privados; segmentación con denegación por defecto	Infraestructura como código; auditoría de arquitectura
Eliminación de datos	Eliminación en cascada por unidad única con log de auditoría	Suite de pruebas de eliminación GDPR
Cadena de suministro	Pasos de pipeline fijados; supervisión semanal de dependencias	Configuración de pipeline; auditoría de dependencias

Appendix B: Preguntas frecuentes para revisores de seguridad

¿Dónde se almacenan nuestros datos? Enteramente dentro de la Unión Europea, en Microsoft Azure, en West Europe con procesamiento de IA en regiones de la UE. Los datos de candidatos nunca salen de la UE.

¿Se utilizan nuestros datos para entrenar modelos de IA? No. El proveedor de IA no utiliza datos del cliente para entrenamiento.

¿Es accesible la base de datos desde internet? No. El acceso a red pública está deshabilitado y la base de datos solo es accesible a través de un endpoint privado dentro de la red virtual.

¿Puede un cliente ver los datos de otro cliente? No. Cada consulta se limita a la organización del solicitante, el acceso entre organizaciones devuelve “not found” y una matriz automatizada prueba continuamente este aislamiento.

¿Cómo se almacenan las contraseñas? Con hash usando bcrypt y una salt única por contraseña. Se admite single sign-on con Microsoft y Google, en cuyo caso no se almacena ninguna contraseña.

¿Admiten single sign-on? Sí, mediante Microsoft y Google OAuth.

¿Cuánto tiempo son válidos los tokens de acceso? Treinta minutos, emparejados con una sesión de refresh revocable del lado del servidor que se invalida al cerrar sesión.

¿Cómo se gestiona el consentimiento del candidato? Cada candidato recibe un enlace de consentimiento único y de un solo uso y debe aceptarlo antes de cualquier grabación o análisis. El consentimiento se registra frente al proceso de contratación específico.

¿Cómo se eliminan los datos? Como una unidad única que cubre el registro del candidato, entrevistas, transcripciones, audio, documentos y comparaciones, según un calendario de retención configurable, con una prueba registrada de borrado. Los candidatos también pueden solicitar la eliminación directamente.

¿Tienen un data processing agreement? Sí, se acepta en el registro y se versiona por organización, incluido el registro de subencargados.

¿La IA toma decisiones de contratación? No. Proporciona solo apoyo a la decisión; un humano revisa cada salida y toma todas las decisiones.

¿Cómo demuestran sus afirmaciones de seguridad? Mediante 3,171 pruebas automatizadas, incluida una suite de seguridad dedicada, una metodología repetible de seis fases de pruebas de penetración ejecutada contra entornos en vivo, un programa de pruebas de seguridad de IA y auditorías escritas recurrentes.

¿Qué sucede cuando encuentran una vulnerabilidad? Se le asigna una severidad con evidencia y un responsable, se remedia según una planificación priorizada, se vuelve a verificar incluyendo comprobaciones en vivo cuando corresponde y se registra en un informe de auditoría.

¿Podemos ejecutar nuestra propia prueba de penetración? Las evaluaciones de seguridad pueden organizarse a través de su representante de cuenta con el alcance y la planificación apropiados.

Appendix C: Glosario

Término	Significado
AES-256	Un sólido estándar de cifrado simétrico utilizado para proteger datos en reposo
bcrypt	Una función de hash de contraseñas diseñada específicamente para ese fin con salting por contraseña
Managed identity	Una identidad emitida por la plataforma que permite a un servicio autenticarse sin claves almacenadas
Private endpoint	Una dirección de red privada que mantiene un servicio cloud fuera de internet pública
Network security group	Un conjunto de reglas de permitir y denegar que filtra el tráfico de red hacia una subred
RBAC	Control de acceso basado en roles, que concede permisos según el rol de un usuario
IDOR	Insecure direct object reference, un fallo de control de acceso frente al que la plataforma se defiende
SSRF	Server-side request forgery, una clase de ataque sondeada en nuestras pruebas de penetración
Web application firewall	Un control perimetral que filtra tráfico web malicioso
Data processing agreement	El contrato que rige cómo un encargado trata datos personales en nombre de un responsable

Appendix D: Contacto y control del documento

AI Interview Analyzer Sp. z o.o.

ul. Jedrusik 6/53, 01-748 Warszawa, Poland

NIP: 5253079974

Para una revisión de seguridad, una copia de nuestro data processing agreement o nuestra documentación de conformidad con el EU AI Act, póngase en contacto con su representante de cuenta.

Este documento describe la postura de seguridad del servicio AI Interview Analyzer en la fecha de generación mostrada en el pie de página. Se proporciona con fines de evaluación y no forma parte de ningún contrato. Los compromisos contractuales específicos de seguridad se establecen en el acuerdo aplicable y en el data processing agreement.