# Whitepaper de Segurança

Enterprise Security Overview - AI Interview Analyzer

| | |
|---|---|
| **Fornecedor:** | AI Interview Analyzer Sp. z o.o. |
| **Morada:** | ul. Jedrusik 6/53, 01-748 Warszawa, Poland |
| **NIP:** | 5253079974 |
| **REGON:** | 54402118500000 |
| **Classificação:** | PUBLIC |
| **Data:** | 24.06.2026 |

# Contents

# Whitepaper de Segurança

**Fornecedor:** AI Interview Analyzer Sp. z o.o., Warszawa, Poland

**Público-alvo:** Equipas empresariais de segurança, TI e compras

**Classificação:** Público

## 1. Resumo Executivo

AI Interview Analyzer é uma plataforma empresarial de recrutamento que grava entrevistas com o consentimento explícito do candidato, as transcreve e estrutura, e produz suporte à avaliação baseado em evidências para recrutadores. Como a plataforma trata dados pessoais de candidatos e dá suporte a processos de contratação, segurança e privacidade são tratadas como restrições primárias de conceção, não como funcionalidades adicionadas posteriormente.

Este whitepaper descreve, em termos concretos e verificáveis, como protegemos os dados de clientes e candidatos. Foi escrito para as pessoas que avaliam fornecedores: engenheiros de segurança, administradores de TI, encarregados de proteção de dados e equipas de compras. Cada valor neste documento é retirado diretamente dos nossos próprios sistemas de engenharia, e não de material de marketing.

A mensagem central é simples: **não nos limitamos a afirmar que a plataforma é segura, testamos continuamente que ela o é.** A nossa base de código contém **3,171 testes automatizados**, incluindo um conjunto dedicado de segurança que exerce autenticação, autorização, isolamento entre organizações, defesas contra injeção e eliminação de dados. Além disso, executamos um mecanismo repetível de penetration testing contra implementações ativas e produzimos relatórios de auditoria por escrito. Em sete auditorias internas de segurança em março e abril de 2026, registámos **zero critical findings**, com a nossa auditoria mais recente a encerrar com um veredito de **PASS**. (A certificação formal destes controlos por terceiros faz parte do nosso roadmap; consulte a Secção 18.)

| Característica de segurança | Resumo |
|---|---|
| Alojamento | Microsoft Azure, apenas regiões da UE |
| Modelo de rede | Private endpoints, segmentação de rede default-deny, sem base de dados pública |
| Encriptação | AES-256 em repouso, TLS 1.2 ou superior em trânsito |
| Identidade | Tokens assinados de curta duração, hash de palavras-passe com bcrypt, suporte a SSO |
| Controlo de acesso | Controlo de acesso baseado em funções com isolamento rigoroso por organização |
| Segredos | Cofre centralizado de segredos com acesso por managed identity |
| Privacidade | Consentimento explícito, retenção configurável, eliminação por unidade única |
| IA responsável | Apenas suporte à decisão, humano sempre no circuito |
| Garantia | 3,171 testes automatizados mais penetration tests e auditorias recorrentes |

### 1.1 Como Ler Este Documento

As Secções 3 a 11 descrevem os controlos que protegem os dados: arquitetura, rede, identidade, aplicação, proteção de dados, privacidade e o ciclo de vida de desenvolvimento seguro. As Secções 12 e 13 cobrem o nosso programa distintivo de testes contínuos e o nosso histórico de auditorias. As Secções 14 a 17 cobrem operações, modelação de ameaças, gestão de vulnerabilidades e mapeamento de conformidade. Os apêndices fornecem um catálogo de controlos, uma FAQ para revisores e um glossário que uma equipa de segurança pode utilizar diretamente durante uma avaliação.

# 2. Âmbito e Abordagem do Documento

## 2.1 O Que Este Documento Cobre

Este whitepaper cobre a arquitetura e as práticas de segurança do serviço AI Interview Analyzer: o ambiente de alojamento, o desenho da rede, gestão de identidade e acesso, controlos ao nível da aplicação, proteção de dados, privacidade e alinhamento regulatório, o ciclo de vida de desenvolvimento seguro e o nosso programa contínuo de testes de segurança.

## 2.2 O Que o Torna Verificável

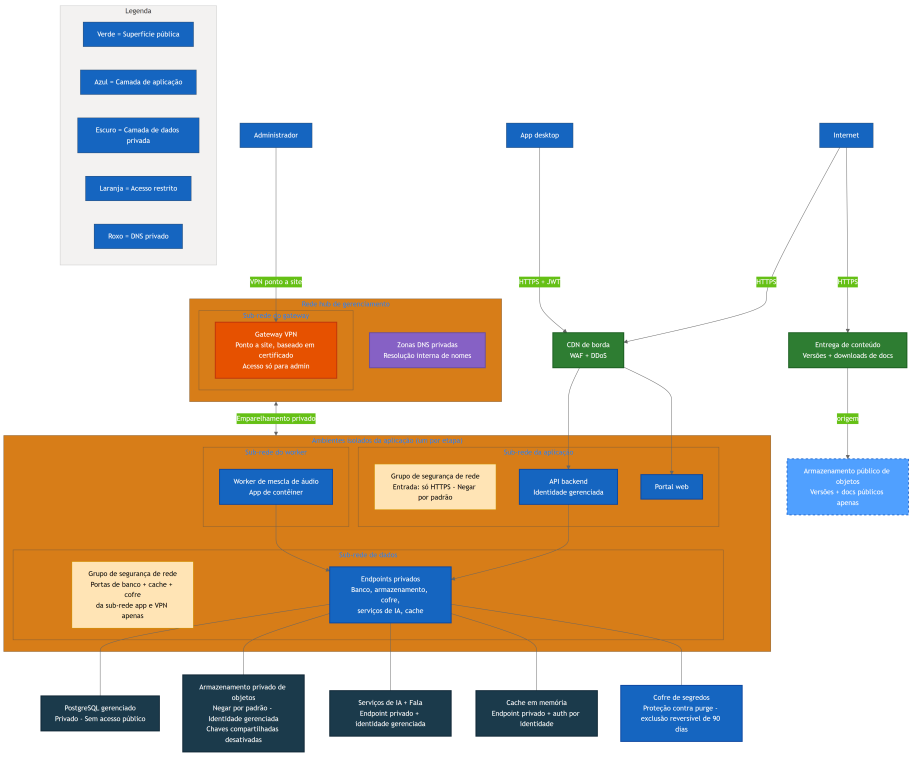
As alegações de segurança de fornecedores são fáceis de escrever e difíceis de confiar. Por isso, associámos cada alegação principal neste documento a algo concreto e quantificável dentro dos nossos sistemas de engenharia: um controlo implementado em código, um teste que prova que o controlo funciona, uma definição de infraestrutura que o aplica, ou um relatório de auditoria que regista uma verificação documentada. Quando um controlo faz parte do nosso roadmap futuro em vez de estar atualmente em produção, dizemo-lo explicitamente. Preferimos subdeclarar e ser dignos de confiança do que sobredeclarar e ser desmentidos.

## 2.3 Responsabilidade Partilhada

A plataforma é fornecida como software as a service. Operamos a infraestrutura, a aplicação, o pipeline de IA e o tratamento de dados. O cliente é responsável por gerir as suas próprias contas de utilizador e funções, configurar janelas de retenção de dados em conformidade com a sua política interna e garantir que o consentimento do candidato é obtido através do fluxo de consentimento fornecido pela plataforma. A Secção 14 descreve esta divisão com mais detalhe.

# 3. Visão Geral da Arquitetura de Segurança

A plataforma é construída como um pequeno número de serviços cooperantes em vez de um único monólito. Uma aplicação de desktop e um portal web atuam como clientes. Uma API backend central detém toda a persistência, autenticação, faturação, pipeline de IA, consentimento, email, gestão de ficheiros e dashboards. Um worker de fusão de áudio processa gravações de forma assíncrona. Todo o estado sensível vive por detrás da API backend; os clientes nunca comunicam diretamente com a base de dados, armazenamento ou serviços de IA.




O diagrama acima mostra a topologia de produção com os nomes dos recursos intencionalmente generalizados. Três princípios são visíveis nele:

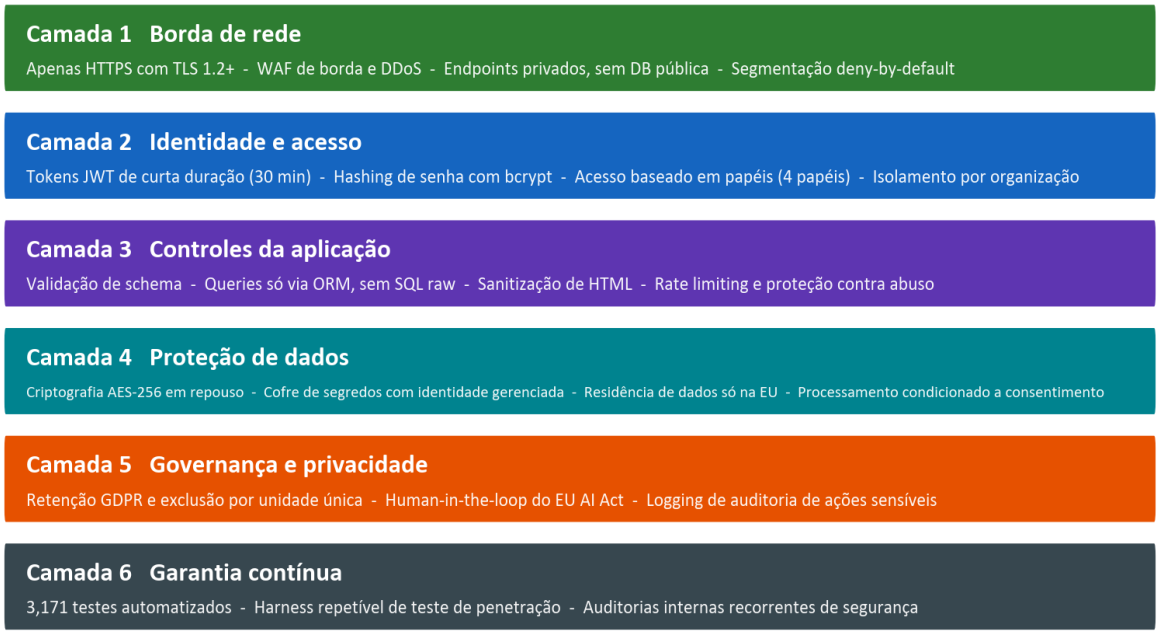
- **Sem exposição direta dos serviços de dados.** A base de dados, o armazenamento privado de objetos, os serviços de IA e a cache têm o acesso de rede pública desativado e só são alcançáveis através de private endpoints dentro de uma rede virtual isolada. O cofre de segredos é alcançado pela aplicação através de um private endpoint e é adicionalmente protegido por autenticação de identidade da plataforma e políticas de acesso de privilégio mínimo, pelo que qualquer acesso exige uma identidade válida e autorizada independentemente do caminho de rede.
- **Uma superfície pública separada.** O único armazenamento público de objetos contém downloads de versões e documentos públicos. Nunca contém dados de candidatos. O tráfego da aplicação voltado para o cliente passa por uma camada edge que fornece web application firewall, proteção contra distributed-denial-of-service e distribuição de conteúdo.
- **O acesso administrativo é controlado.** Os operadores alcançam recursos internos apenas através de uma VPN point-to-site baseada em certificado para uma rede hub de gestão, e não pela internet pública.

Cada fase de implementação (desenvolvimento e produção) é um ambiente totalmente isolado com a sua própria rede, contas de armazenamento, base de dados e segredos. Dados de produção de clientes nunca estão presentes em ambientes inferiores. Um hub de gestão partilhado contém apenas o gateway VPN e o DNS privado, emparelhado privadamente com cada ambiente.

# 4. Defesa em Profundidade

Nenhum controlo isolado é considerado suficiente para travar todos os ataques. A plataforma sobrepõe controlos independentes para que a falha de qualquer camada não exponha dados. As camadas abaixo são cada uma implementadas e, conforme descrito na Secção 12, testadas individualmente.

**Modelo de segurança em camadas: controles independentes em cada nível**




| Camada | Controlos representativos |
|---|---|
| Perímetro de rede | Transporte apenas por TLS, WAF edge e proteção DDoS, private endpoints, segmentação default-deny |
| Identidade e acesso | Tokens assinados de curta duração, hash com bcrypt, controlo de acesso baseado em funções, isolamento por organização |
| Aplicação | Validação de esquema em todas as entradas, acesso a dados apenas por ORM, codificação de saída, rate limiting |
| Proteção de dados | Encriptação em repouso, cofre de segredos com managed identity, residência de dados na UE, processamento condicionado por consentimento |
| Governação e privacidade | Retenção configurável, eliminação por unidade única, IA com humano no circuito, audit logging |
| Garantia contínua | Conjunto de testes automatizados, penetration tests repetíveis, auditorias internas de segurança recorrentes |

O restante deste documento percorre cada camada por sua vez e depois descreve como provamos, continuamente, que as camadas se mantêm.

# 5. Segurança de Rede

## 5.1 Privado por Omissão

A camada de dados é privada por construção. A base de dados PostgreSQL gerida tem o acesso de rede pública desativado e só é alcançável através de um private endpoint. O armazenamento privado de objetos está configurado para negar acesso de rede por omissão, desativa totalmente shared access keys e é acessível apenas através de managed identity a partir da subnet da aplicação. A cache, os serviços de IA e o cofre de segredos são igualmente alcançados através de private endpoints com resolução DNS privada.

Na prática, isto significa que não existe connection string exposta à internet para a base de dados nem URL pública de armazenamento para o áudio do candidato: a base de dados e o armazenamento privado têm o acesso de rede pública desativado de forma absoluta. O cofre de segredos é alcançado pela aplicação através de um private endpoint e é protegido por autenticação de identidade da plataforma e políticas de acesso de privilégio mínimo, com identidades da aplicação a receberem acesso apenas de leitura aos segredos específicos de que necessitam, pelo que os segredos não podem ser obtidos sem uma identidade válida e autorizada. A superfície de ataque a que um adversário externo pode sequer aceder está limitada aos endpoints HTTPS da aplicação por detrás da camada edge.

## 5.2 Segmentação de Rede

Cada ambiente está dividido em subnets separadas para a camada da aplicação, a camada de dados e o worker assíncrono. Cada subnet é governada por um network security group cuja regra final nega todo o tráfego de entrada. A subnet da aplicação aceita apenas tráfego HTTPS de entrada. A subnet de dados aceita apenas as portas específicas de base de dados, cache e cofre, e apenas a partir da subnet da aplicação ou da VPN administrativa. Isto significa que mesmo um atacante que de alguma forma alcançasse a camada da aplicação não poderia mover-se livremente para a camada de dados; os únicos caminhos permitidos são aqueles que a aplicação utiliza legitimamente.

## 5.3 A Camada Edge

O tráfego público da aplicação é colocado à frente de uma camada edge que fornece web application firewall, proteção DDoS e uma content delivery network. Os downloads de versões e documentos são servidos a partir de uma conta dedicada de armazenamento público através de uma front door de distribuição de conteúdo, completamente separada do armazenamento privado que contém dados de candidatos. Os dois planos de armazenamento nunca se misturam: uma configuração incorreta no plano público não pode expor dados privados de candidatos, porque são contas diferentes com regras de rede diferentes.

## 5.4 Acesso Administrativo

Não existe endpoint administrativo público para a rede privada. Os operadores ligam-se através de um gateway VPN point-to-site que utiliza autenticação baseada em certificado. O acesso administrativo à base de dados e à cache só é possível a partir do interior desse túnel, uma vez que esses serviços têm o acesso de rede pública desativado. Isto mantém as operações diárias totalmente fora da internet pública.

# 6. Gestão de Identidade e Acesso

## 6.1 Autenticação

As sessões de utilizador são estabelecidas com um token de acesso assinado válido por trinta minutos, emparelhado com um refresh token opaco e separado no lado do servidor. Os tokens de acesso são verificados em cada pedido, e o utilizador é revalidado na base de dados (incluindo uma verificação de conta ativa) em vez de se confiar apenas no conteúdo do token. Terminar sessão revoga imediatamente a sessão de refresh no lado do servidor, pelo que um refresh token roubado não pode sobreviver a um logout.

As palavras-passe nunca são armazenadas em plain text. São submetidas a hash com bcrypt usando um salt único por palavra-passe. Para organizações que preferem single sign-on, a plataforma suporta login OAuth com Microsoft e Google, caso em que nenhuma palavra-passe é armazenada.

A propriedade do email é verificada através de um link de verificação de utilização única e com tempo limitado antes de uma conta auto-registada ser tratada como verificada, e os reenvios de email de verificação estão sujeitos a rate limiting para prevenir abuso.

## 6.2 Controlo de Acesso Baseado em Funções

A autorização é aplicada através de um modelo de funções com quatro funções de privilégio crescente: interviewer, hiring manager, recruiter e administrator. O acesso a operações privilegiadas é aplicado por dependências do lado do servidor que verificam tanto a função como o estado de verificação do autor do pedido. Estas verificações de função protegem bem mais de uma centena de operações distintas da API.

| Função | Capacidades típicas |
|---|---|
| Interviewer | Conduz entrevistas atribuídas; vê apenas entrevistas que lhe foram atribuídas |
| Hiring manager | Gere recrutamentos que possui ou dos quais é membro |
| Recruiter | Gestão completa de recrutamento e candidatos dentro da organização |
| Administrator | Definições da organização, faturação, administração de utilizadores e chaves de API |

Para além das verificações grosseiras de função, a plataforma aplica regras de visibilidade ao nível dos dados. Hiring managers veem apenas os recrutamentos que criaram ou dos quais são membros; interviewers veem apenas as entrevistas que lhes foram atribuídas. O privilégio é, portanto, aplicado tanto ao nível de "que ação" como ao nível de "que registos".

## 6.3 Isolamento por Organização

A plataforma é multi-tenant, e o isolamento entre tenants é tratado como um controlo de segurança de primeira classe. Cada identidade autenticada transporta um identificador de organização, e as consultas de dados são delimitadas a essa organização. Quando um utilizador pede um registo que pertence a outra organização, a plataforma devolve uma resposta "not found" em vez de revelar que o registo existe. Identificadores internos da base de dados nunca são expostos na interface; a API apresenta identificadores de exibição e remapeia-os por pedido, o que elimina uma classe comum de ataque de enumeração entre tenants.

Isto não é apenas uma intenção de desenho. Conforme descrito na Secção 12, o nosso conjunto automatizado executa uma grande matriz entre organizações que tenta alcançar dados de uma organização usando credenciais de outra organização e verifica que todas essas tentativas falham.

## 6.4 Acesso Programático

Para integrações, organizações em planos elegíveis podem emitir chaves de API. As chaves usam um prefixo reconhecível, transportam 128 bits de entropia e são armazenadas apenas como hash; a chave bruta é mostrada uma vez na criação e nunca mais. Cada chave transporta um âmbito explícito de permissões (read, write ou integração ATS), pode ser restringida a redes de

origem específicas, pode ser revogada instantaneamente e está sujeita a limites de taxa por chave derivados do nível de plano da organização. A verificação de chaves usa uma comparação timing-safe para evitar fuga de informação através do tempo de resposta.

# 7. Segurança da Aplicação

A aplicação é escrita para remover categorias inteiras de vulnerabilidade em vez de as corrigir caso a caso.

• **Injeção.** Todo o acesso à base de dados passa por um object-relational mapper com consultas parametrizadas. A base de código não contém SQL bruto formatado por string. Isto elimina estruturalmente SQL injection.

• **Validação de entrada.** Cada corpo de pedido é validado contra um esquema rigoroso antes de chegar à lógica de negócio. Payloads excessivamente grandes são rejeitados, e os endpoints de lista são paginados para limitar o uso de recursos.

• **Codificação de saída e cross-site scripting.** Texto fornecido pelo utilizador e gerado por IA é tratado como não confiável. Quando o conteúdo tem de ser renderizado como HTML, passa por um sanitizador de allow-list no momento da escrita, e um conjunto de testes dedicado confirma que script tags, event handlers e URLs javascript são removidos.

• **Mass assignment.** As operações de atualização usam esquemas explícitos que excluem campos privilegiados como função, organização e saldo de créditos, para que um cliente não possa elevar privilégios ao enviar campos extra.

• **Rate limiting.** Os endpoints de autenticação e sujeitos a abuso são limitados por taxa usando um limitador durável baseado em base de dados que sobrevive a reinícios e funciona corretamente em múltiplas instâncias da aplicação. Login, registo, reposição de palavra-passe e reenvios de verificação têm, cada um, os seus próprios limites. A resolução de IP do cliente é reforçada contra spoofing de forwarding headers.

• **Webhooks.** Webhooks de entrada de fornecedores de pagamento e email são verificados contra assinaturas do fornecedor no corpo bruto do pedido antes de serem processados.

• **Uploads de ficheiros.** Os uploads têm limite de tamanho, são validados, armazenados sob identificadores gerados em vez de nomes fornecidos pelo utilizador, e limitados por pedido e por organização.

• **Cabeçalhos de segurança.** Em produção, as respostas transportam strict transport security, opções de content-type e frame, uma política de referrer e uma permissions policy restritiva, e ocultam banners do servidor e do framework.

# 8. Proteção de Dados

## 8.1 Encriptação

Todos os dados são encriptados em repouso usando AES-256 através das camadas de encriptação de armazenamento e base de dados da plataforma Azure. Todo o tráfego de rede é servido exclusivamente sobre HTTPS usando TLS 1.2 ou superior; HTTP em plain text é redirecionado para HTTPS em todas as camadas. Em produção, a API e o portal web emitem cabeçalhos strict transport security juntamente com um conjunto de cabeçalhos de reforço e ocultam banners de versão do servidor e do framework.

## 8.2 Gestão de Segredos

Os segredos da aplicação são mantidos num cofre centralizado de segredos com purge protection ativada e uma janela de soft-delete de noventa dias. As aplicações autenticam-se em recursos Azure usando system-assigned managed identities em vez de chaves de longa duração; por exemplo, o armazenamento privado tem shared access keys totalmente desativadas, pelo que o acesso só é possível através de atribuições de função baseadas em identidade delimitadas ao recurso individual. As políticas de acesso ao cofre concedem aos principais da aplicação acesso apenas de leitura aos segredos específicos de que necessitam, seguindo o princípio do privilégio mínimo.

## 8.3 Residência dos Dados

Todos os dados de clientes e candidatos são armazenados e processados dentro da União Europeia. O alojamento da aplicação, a base de dados, armazenamento, cache e segredos residem em West Europe, e o processamento de IA é executado em regiões da UE. O fornecedor de IA não utiliza dados de clientes para treinar os seus modelos.

## 8.4 O Ciclo de Vida de Uma Entrevista Individual

A forma mais clara de compreender os controlos de proteção de dados é seguir uma entrevista de ponta a ponta. O consentimento é capturado e registado antes de qualquer processamento. O upload é encriptado em trânsito. A transcrição e a análise decorrem em centros de dados da UE. Os resultados são escritos em armazenamento encriptado. Cada registo é então governado por um único relógio de retenção que termina numa eliminação em cascata registada. Em qualquer momento, direitos do candidato como retirada, eliminação, acesso ou portabilidade podem interromper este fluxo.

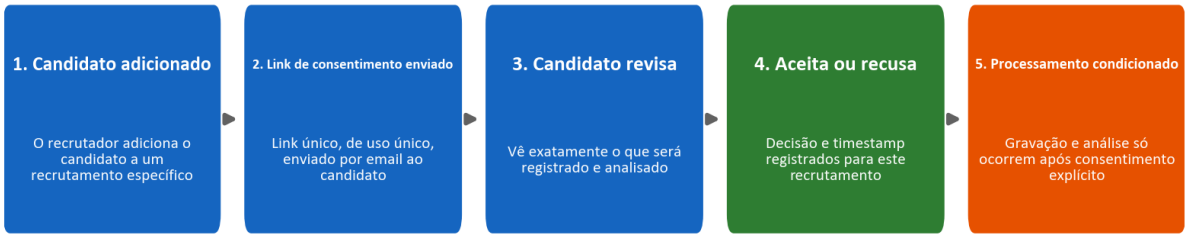
# 9. Privacidade desde a Conceção e GDPR

A privacidade está integrada no modelo de dados e no fluxo de trabalho, não apenas acrescentada por política.

## 9.1 Consentimento

Nenhuma entrevista é gravada ou analisada sem o consentimento explícito do candidato. Quando um candidato é adicionado a um recrutamento, a plataforma emite por email um link de consentimento único e de utilização única. O candidato revê o que irá acontecer e aceita ou recusa. O estado do consentimento, incluindo a hora da resposta, é registado contra esse recrutamento específico, pelo que o consentimento está sempre delimitado a um processo de contratação concreto em vez de ser concedido globalmente.

**Consentimento do candidato: explícito e registrado antes de qualquer processamento**

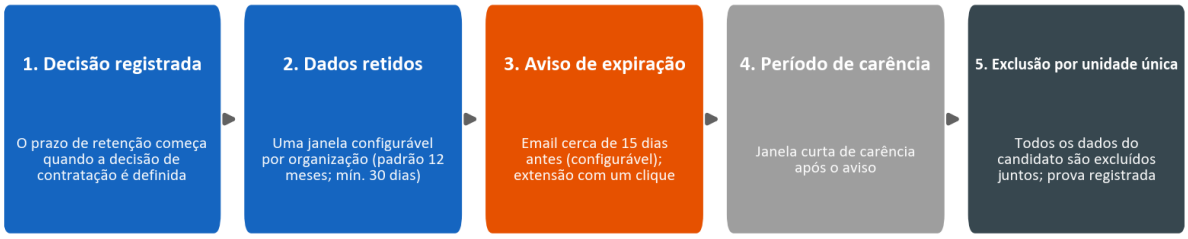



## 9.2 Retenção e Eliminação

A retenção de dados é configurável por organização, com um valor por omissão de doze meses e um mínimo configurável de trinta dias, e pode ser substituída por candidato. Existe um único relógio de retenção para os dados de um candidato, não um temporizador separado por artefacto. O relógio começa quando é registada uma decisão de contratação. Antes de os dados expirarem, a plataforma envia um aviso (por omissão cerca de quinze dias antes) e oferece uma extensão com um clique. Quando os dados são eliminados, são eliminados como uma única unidade: o registo do candidato, entrevistas, transcrições, gravações áudio, documentos e comparações são todos removidos em conjunto, e a eliminação é registada num audit log. Não há resíduos parciais nem órfãos.

O ciclo de vida abaixo mostra este relógio único e como converge para uma única eliminação em cascata com uma prova registada de apagamento.

**Retenção de dados: um prazo por candidato, exclusão por unidade única**




## 9.3 Direitos do Titular dos Dados e Sub-processors

A plataforma suporta os direitos do titular dos dados exigidos pelo GDPR, incluindo acesso, eliminação, portabilidade, oposição e explicação. O processamento é realizado ao abrigo de um data processing agreement que os clientes aceitam no registo e que é versionado por organização. Os nossos sub-processors e respetivas funções, todos dentro da UE ou sob salvaguardas apropriadas, são divulgados nesse acordo, e os clientes recebem notificação prévia de qualquer alteração. A Secção 17 contém o registo de sub-processors e o mapeamento de conformidade artigo por artigo.

# 10. IA Responsável e o EU AI Act

A plataforma enquadra-se na categoria de alto risco do EU AI Act porque apoia decisões de emprego, e tratamos essa classificação com seriedade.

A regra definidora do produto é que **a IA é suporte à decisão, não um decisor.** O sistema nunca aceita ou rejeita automaticamente um candidato. Transcreve a fala, estrutura perguntas e respostas, pontua respostas face a critérios definidos pelo recrutador e redige feedback, e um humano revê cada saída antes de ela ser utilizada. Isto mantém um humano firmemente no circuito.

Igualmente importante é aquilo que a IA não faz. Não avalia personalidade, "cultural fit", estado emocional, tom de voz, sotaque, género, idade, etnia, aparência ou linguagem corporal. A pontuação está ancorada em evidência da transcrição e em critérios definidos pelo recrutador, e os nomes dos candidatos são excluídos da entrada de avaliação para reduzir enviesamento. Publicamos um cartão de transparência, documentação do utilizador e uma declaração de conformidade descrevendo o sistema, as suas limitações e as suas salvaguardas.

| Controlo de IA responsável | Como funciona |
|---|---|
| Humano no circuito | Cada pontuação e cada elemento de feedback é revisto por um recrutador antes da utilização |
| Sem decisões automatizadas | O sistema nunca aceita nem rejeita automaticamente um candidato |
| Pontuação baseada em evidências | As pontuações referenciam evidência de suporte da transcrição |
| Conceção anti-enviesamento | Nomes excluídos da avaliação; a substância é pontuada acima do estilo |
| Limites de âmbito | Personalidade, emoção, sotaque e características protegidas nunca são avaliados |
| Segurança do feedback ao candidato | O feedback privado ao candidato passa por uma proteção de segurança de geração e validação |

Estas restrições não estão apenas declaradas na documentação; estão codificadas na camada de prompts de IA e exercidas por um programa dedicado de testes de segurança de IA descrito na Secção 12.3.

# 11. Ciclo de Vida de Desenvolvimento Seguro

A segurança é aplicada na forma como construímos e distribuímos software, não apenas no sistema em execução.

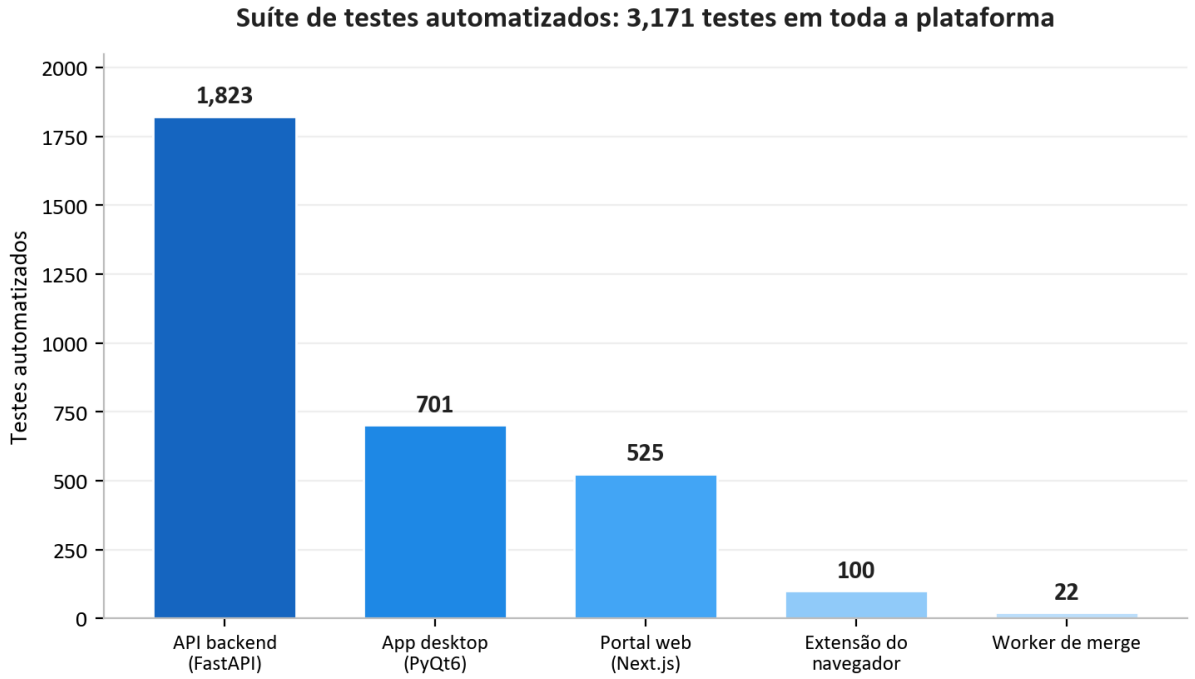
- **Separação de ambientes.** Desenvolvimento e produção são totalmente separados, cada um com a sua própria infraestrutura, contas de armazenamento, base de dados, segredos e subdomínios. Não existe estado partilhado.
- **Infrastructure as code.** Todo o ambiente cloud é definido como código e revisto como código, o que torna a postura de segurança auditável e reproduzível. Um revisor pode ler exatamente que portas estão abertas, que recursos são privados e que identidades têm que permissões.
- **Implementações fixadas e controladas.** Cada passo no pipeline de continuous-integration está fixado a uma versão exata e imutável. As implementações de produção são baseadas em tags, executadas apenas através do pipeline de produção protegido e controladas por aprovação obrigatória. O conjunto de testes automatizados é executado como gate de release: uma implementação não pode ser lançada se os testes falharem.
- **Higiene de dependências.** A monitorização automatizada de dependências propõe atualizações semanalmente no backend, desktop, web, infraestrutura e definições de pipeline, e auditorias de dependências fazem parte da nossa revisão periódica de segurança.
- **Artefactos assinados.** Os instaladores de desktop são assinados com código, para que os clientes possam verificar que o software que instalam provém genuinamente de nós.
- **Disciplina de segredos.** Os segredos vivem no cofre e nos segredos protegidos do pipeline, nunca no código-fonte.

# 12. Testes Contínuos de Segurança

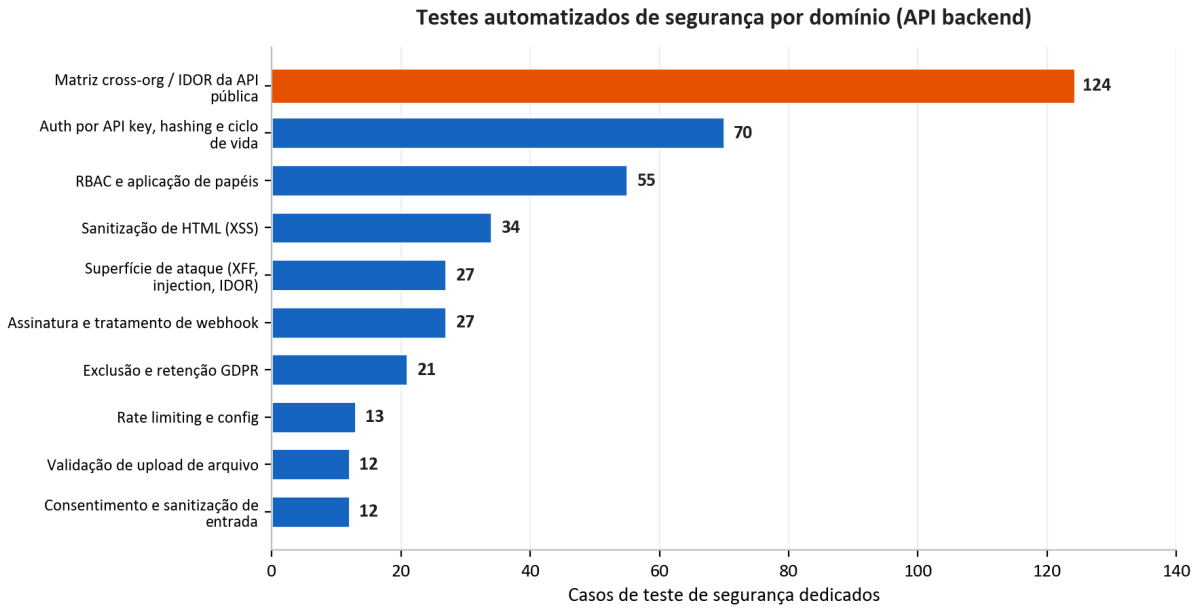
Este é o núcleo da nossa história de garantia e a parte que a maioria dos fornecedores não consegue mostrar. Tratamos a segurança como algo a ser medido continuamente, com verificações executáveis, em vez de algo afirmado uma única vez.

## 12.1 O Conjunto de Testes Automatizados

A plataforma está coberta por **3,171 testes automatizados** abrangendo a API backend, a aplicação de desktop, o portal web, a extensão de browser e o worker de fusão de áudio.




Não se trata apenas de testes funcionais. Um conjunto substancial e dedicado de segurança exerce os controlos descritos anteriormente neste documento. O gráfico abaixo decompõe os testes específicos de segurança na API backend por domínio.



Entre muitos outros, este conjunto inclui uma grande matriz de API pública que executa cada endpoint como um utilizador legítimo, como a própria chave de API da organização e como a chave de API de uma organização rival, verificando que cada tentativa entre organizações é bloqueada. Inclui dezenas de testes adversariais de superfície de ataque para spoofing de forwarding headers, header injection e fuga de identificadores, um conjunto focado de sanitização de HTML para cross-site scripting, testes de aplicação de funções para o modelo completo de funções e testes que provam que os dados do candidato são verdadeiramente eliminados como uma unidade. Como estes testes são executados como gate de release, uma regressão que enfraquecesse qualquer um destes controlos bloquearia o lançamento em vez de chegar aos clientes.

## 12.2 Live Penetration Testing

Os testes unitários automatizados provam que os controlos se comportam corretamente de forma isolada. Para provar que se mantêm em conjunto numa implementação real, mantemos uma metodologia repetível de penetration testing que executa scripts de ataque reais contra um ambiente ativo. Está organizada em seis fases:

| Fase | Foco | Exemplos do que é exercido |
|---|---|---|
| 1. Análise estática | Código-fonte | Segredos, padrões de injeção, funções perigosas, ausência de auth, HTML inseguro |
| 2. Revisão de arquitetura | Infraestrutura | Private endpoints, segmentação, TLS, configuração de segredos |
| 3. Análise de vetores de ataque | Controlo de código-fonte e cloud | Proteção de branches, âmbito de identidade, exposição pública |
| 4. Live penetration testing | Ambiente em execução | Sondagem sem autenticação, acesso entre organizações, injeção, adulteração de tokens, SSRF, bursts de rate-limit |
| 5. Enterprise scoring | Maturidade | Dezasseis categorias de segurança pontuadas face a uma baseline empresarial |
| 6. Dependências e cadeia de abastecimento | Risco de terceiros | Auditoria CVE de dependências, ações de pipeline fixadas, integridade de lock-file |

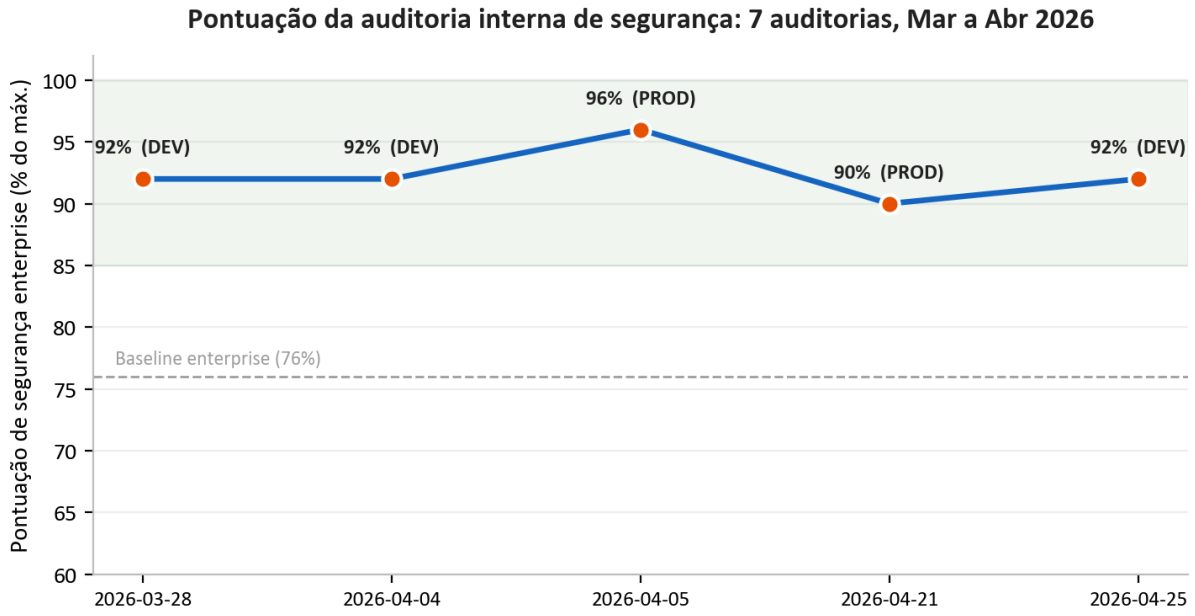
A Fase 4 é um teste adversarial genuíno contra um sistema implementado, não uma checklist. Sonda endpoints protegidos sem credenciais e confirma que recusam acesso; regista duas organizações e tenta alcançar registos de uma organização com a conta da outra; injeta payloads de cross-site-scripting e server-side-template e confirma que são neutralizados; adultera tokens de autenticação e confirma que são rejeitados; tenta server-side request forgery contra endpoints de metadados cloud; e faz bursts em endpoints de autenticação para confirmar que o rate limiting é efetivamente acionado no ambiente ativo, e não apenas em teoria.

## 12.3 Testes de Segurança do Feedback ao Candidato

Como a plataforma pode gerar feedback privado de desenvolvimento para candidatos, executamos um programa adversarial de segurança separado contra essa funcionalidade. Alimenta deliberadamente o sistema com notas de recrutadores duras e hostis e confirma que a saída voltada para o candidato nunca contém vulgaridade, nunca revela nem atribui a identidade ou opinião privada de um recrutador e nunca aplica rótulos de personalidade pejorativos. Isto protege tanto o candidato, que deve receber feedback construtivo e respeitoso, como o cliente, cuja opinião interna nunca deve vazar para o exterior.

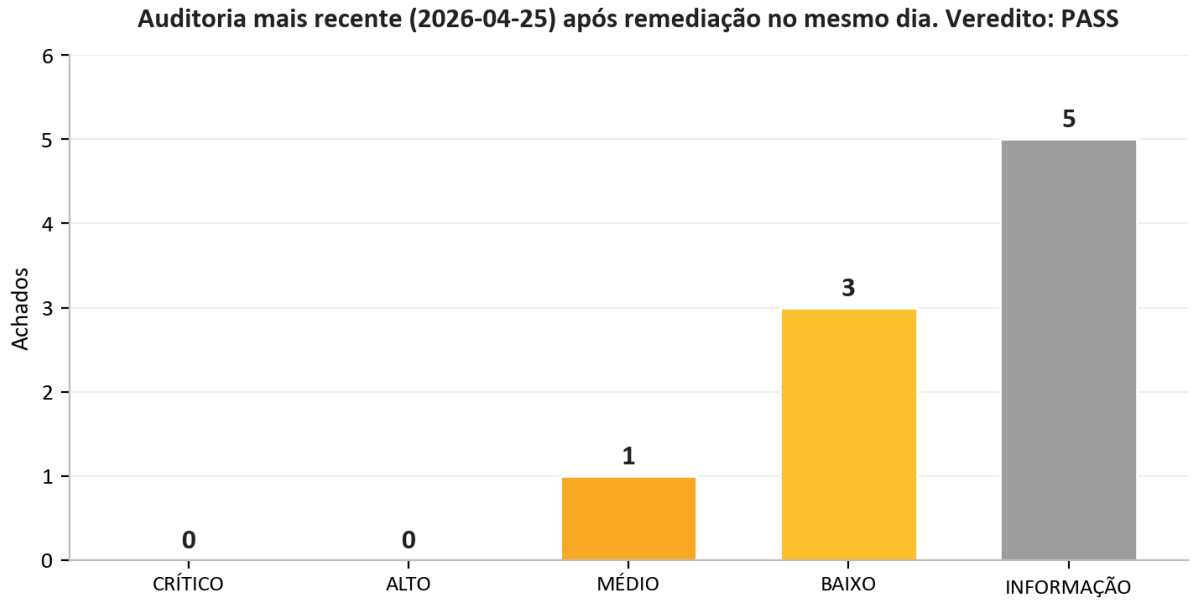
# 13. Resultados das Auditorias de Segurança

Conduzimos auditorias de segurança recorrentes usando uma metodologia estruturada e repetível de penetration testing, e documentamos cada uma como um relatório datado com findings classificados por severidade, evidência e remediação. Estas são auditorias internas executadas pelo nosso próprio processo de segurança; a certificação formal dos mesmos controlos por terceiros faz parte do nosso roadmap. Entre o final de março e o final de abril de 2026 concluímos **sete dessas auditorias** em desenvolvimento e produção.




O resultado que mais importa para um potencial cliente é a consistência: **em todas as sete auditorias houve zero critical findings.** Nas raras ocasiões em que surgiu um problema de maior severidade, este foi remediado rapidamente, muitas vezes no próprio dia, e reverificado. A rubrica de pontuação foi deliberadamente endurecida ao longo deste período (a pontuação máxima possível foi aumentada à medida que acrescentámos mais categorias a avaliar), razão pela qual a linha de pontuação normalizada se mantém elevada mesmo com a fasquia a subir.

A nossa auditoria mais recente, em 25 April 2026, ilustra como o processo funciona na prática. Foram identificados dois problemas de maior severidade, ambos corrigidos e reverificados no próprio dia, e a auditoria encerrou com um veredito de **PASS** sem problemas exploráveis remanescentes no modelo de ameaças atual.



| Auditoria | Ambiente | Critical | Veredito |
|---|---|---|---|
| 2026-03-28 | Desenvolvimento | 0 | Pronto para produção |
| 2026-04-04 | Desenvolvimento | 0 | Pronto para empresas |
| 2026-04-05 | Produção | 0 | Pronto para empresas |
| 2026-04-20 | Desenvolvimento | 0 | Pronto para produção, notas |
| 2026-04-20 | Desenvolvimento | 0 | Aprovado com notas |
| 2026-04-21 | Produção | 0 | Seguro, sem findings exploráveis |
| 2026-04-25 | Desenvolvimento | 0 | Aprovado |

O padrão nestas auditorias é a evidência mais honesta que podemos oferecer: são encontrados problemas, porque os procuramos com rigor, e são resolvidos rapidamente, porque o processo foi construído para os resolver. Um fornecedor que nunca reporta um finding é normalmente um fornecedor que não está a procurar.

---

# 14. Resiliência Operacional e Responsabilidade Partilhada

## 14.1 Monitorização e Registo

A telemetria da aplicação e da plataforma flui para um workspace centralizado de análise de logs e um serviço de monitorização de aplicações, dando-nos visibilidade sobre disponibilidade e comportamento. Ações sensíveis como eliminação de dados, aceitação de acordos legais e invocações de IA são registadas em tabelas de auditoria dedicadas, para que exista um registo durável de quem fez o quê a dados importantes.

## 14.2 Backup e Recuperação

A base de dados gerida retém backups automatizados, e o armazenamento privado é protegido por retenção soft-delete tanto em blobs como em contentores, para que eliminações acidentais ou maliciosas possam ser recuperadas dentro da janela de retenção. A infraestrutura crítica possui deletion locks para prevenir a remoção acidental de recursos de produção.

## 14.3 Resumo de Responsabilidade Partilhada

| Área | AI Interview Analyzer | Cliente |
|---|---|---|
| Infraestrutura, rede, patching | Sim | - |
| Segurança da aplicação e pipeline de IA | Sim | - |
| Encriptação, segredos, residência dos dados | Sim | - |
| Administração de utilizadores e funções | Fornece os controlos | Gere utilizadores e funções |
| Configuração da política de retenção | Fornece os controlos | Define a janela de retenção |
| Consentimento do candidato | Fornece o fluxo | Garante que é utilizado |
| Credenciais fortes de utilizador final e SSO | Suporta SSO e política | Aplica política interna |

# 15. Modelo de Ameaças e Mapeamento OWASP

Projetamos a plataforma contra um conjunto concreto de adversários: um atacante externo sem credenciais, um utilizador autenticado curioso ou malicioso de uma organização a tentar alcançar dados de outra organização, uma dependência comprometida e um erro interno. A tabela abaixo mapeia as categorias de risco amplamente utilizadas do OWASP Top 10 para os controlos específicos que as abordam nesta plataforma, cada um dos quais é exercido pelos testes descritos na Secção 12.

| Risco OWASP | Como a plataforma o mitiga |
|---|---|
| Broken access control | Controlo de acesso baseado em funções em cada endpoint privilegiado; delimitação por organização; "not found" em acesso entre organizações; remapeamento de identificadores; matriz de testes entre organizações |
| Cryptographic failures | TLS 1.2+ em trânsito; AES-256 em repouso; hash de palavras-passe com bcrypt; segredos num cofre gerido |
| Injection | Consultas parametrizadas apenas via ORM; validação rigorosa de esquema; sanitização de HTML no momento da escrita |
| Insecure design | Defesa em profundidade em camadas; modelação de ameaças e revisão de arquitetura em cada auditoria |
| Security misconfiguration | Infrastructure as code; grupos de rede default-deny; cabeçalhos de segurança; shared storage keys desativadas; esquema de API não exposto em produção |
| Vulnerable components | Monitorização automatizada semanal de dependências; auditorias CVE de dependências em revisão periódica |
| Identification and authentication failures | Tokens de curta duração; login com rate limiting; verificação de email; suporte a SSO; ausência de palavras-passe em plain text |
| Software and data integrity failures | Passos de pipeline fixados e imutáveis; instaladores de desktop assinados; verificação de assinatura de webhook; deploys de produção controlados por tag |
| Security logging and monitoring failures | Telemetria centralizada; tabelas de auditoria dedicadas para ações sensíveis |
| Server-side request forgery | Chamadas de saída restringidas a endpoints confiáveis; sondas SSRF no mecanismo de penetration testing |

Este mapeamento é a espinha dorsal do nosso argumento de garantia: para cada classe de ataque bem conhecida existe um controlo nomeado, e para cada controlo nomeado existe um teste.

# 16. Gestão de Vulnerabilidades e Responsible Disclosure

A segurança nunca está concluída, por isso operamos um ciclo contínuo de descoberta e remediação.

- **Descoberta.** As vulnerabilidades são identificadas a partir de quatro fontes: o conjunto de testes automatizados, as auditorias recorrentes de penetration testing, a monitorização automatizada de dependências e relatórios de clientes ou investigadores.
- **Triagem.** Cada finding recebe uma severidade (critical, high, medium, low ou informativo) com evidência e um responsável pela remediação, exatamente como registado nos nossos relatórios de auditoria.
- **Objetivos de remediação.** Findings critical e high são priorizados para remediação imediata; no nosso histórico de auditorias, findings de severidade superior foram tipicamente resolvidos e reverificados no próprio dia. Findings medium e inferiores são calendarizados na cadência regular de manutenção.
- **Verificação.** As correções são testadas novamente e, quando relevante, é executada uma verificação ao vivo contra o ambiente implementado para confirmar que o problema está genuinamente encerrado, e não apenas encerrado no código.
- **Divulgação.** As preocupações de segurança podem ser reportadas diretamente a nós. Acusamos receção dos relatórios, investigamos e mantemos o autor informado até à resolução.

# 17. Mapeamento de Conformidade

## 17.1 GDPR

| Área do GDPR | Implementação da plataforma |
|---|---|
| Base legal (Art. 6) | Consentimento explícito do candidato capturado antes do processamento |
| Minimização de dados e limitação de armazenamento (Art. 5) | Apenas dados relevantes para a entrevista são processados; retenção configurável com eliminação automática |
| Direito ao apagamento (Art. 17) | Eliminação por unidade única de todos os dados do candidato, com prova registada de apagamento |
| Direitos do titular dos dados (Art. 15 a 20) | Acesso, eliminação, portabilidade e oposição são suportados |
| Obrigações do processador (Art. 28) | Data processing agreement aceite no registo e versionado por organização |
| Segurança do processamento (Art. 32) | Encriptação, controlo de acesso, isolamento e testes contínuos conforme descrito neste documento |
| Transparência de sub-processors | Divulgados no data processing agreement com notificação prévia de alterações |

## 17.2 EU AI Act

A plataforma é tratada como um sistema de IA de alto risco que apoia decisões de emprego, e mantemos documentação alinhada com o regulamento, incluindo um cartão de transparência, documentação do utilizador e uma declaração de conformidade. As salvaguardas principais, supervisão humana, transparência, pontuação baseada em evidências e limites rigorosos sobre o que a IA avalia são descritos na Secção 10. Continuamos a amadurecer a nossa documentação formal de conformidade à medida que avança o calendário de implementação do regulamento.

## 17.3 Certificações de Alojamento

A plataforma é executada inteiramente em Microsoft Azure, cujos centros de dados possuem certificações independentes incluindo ISO 27001 e SOC 2. Estas certificações cobrem as camadas físicas e de plataforma sob a nossa aplicação; os controlos ao nível da aplicação são os descritos ao longo deste documento.

## 17.4 Registo de Sub-processors

| Sub-processor | Finalidade | Região |
|---|---|---|
| Microsoft Azure | Alojamento, processamento de IA e voz, armazenamento, email transacional | UE (West Europe, Sweden Central) |
| Stripe | Processamento de subscrições e pagamentos | UE (Ireland) |
| Fakturownia | Faturação | UE (Poland) |
| ATS connector (opcional) | Integração com applicant-tracking, ativada apenas a pedido | UE |

# 18. Roadmap de Segurança

Tratamos a segurança como um programa em melhoria contínua. As iniciativas atuais no nosso roadmap incluem o reforço das opções de autenticação multifator para contas administrativas, a expansão do audit logging centralizado de acesso a dados, a continuação do reforço da atualização de dependências em cadência regular e o avanço da certificação formal por terceiros dos controlos descritos neste documento. Nenhuma destas iniciativas representa uma lacuna que exponha dados de clientes hoje; cada uma é uma melhoria sobre uma postura já em camadas.

# 19. Resumo

AI Interview Analyzer protege os dados de candidatos e clientes através de uma arquitetura em camadas: uma rede privada por omissão sem serviços públicos de dados, identidade forte e isolamento por organização, código da aplicação que elimina classes inteiras de vulnerabilidades pela conceção, encriptação e residência de dados na UE, e controlos de privacidade integrados no modelo de dados. O que distingue a plataforma é a evidência por detrás destas alegações. Com 3,171 testes automatizados, uma metodologia repetível de live penetration testing, um programa dedicado de segurança de IA e um histórico de sete auditorias internas de segurança com zero critical findings, podemos demonstrar, e não apenas afirmar, que a plataforma é segura.

# Apêndice A: Catálogo de Controlos de Segurança

Uma referência condensada dos controlos principais e da evidência que sustenta cada um.

| Controlo | Mecanismo | Evidência |
|---|---|---|
| Encriptação em trânsito | Apenas HTTPS, TLS 1.2+, HTTP redirecionado | Infrastructure as code; auditoria de arquitetura |
| Encriptação em repouso | Encriptação AES-256 da plataforma em armazenamento e base de dados | Configuração da plataforma; auditoria de arquitetura |
| Proteção de palavras-passe | bcrypt com salt por palavra-passe | Controlo de código-fonte; testes de autenticação |
| Gestão de sessão | Tokens assinados de 30 minutos, refresh revogável no lado do servidor | Controlo de código-fonte; testes de autenticação |
| Autorização | Controlo de acesso de quatro funções em endpoints privilegiados | Conjunto de testes de aplicação de funções |
| Isolamento de tenant | Delimitação de consultas por organização; 404 em acesso entre organizações | Matriz de testes entre organizações |
| Segurança de chave de API | Armazenamento com hash, permissões delimitadas, rate limits por chave | Conjunto de testes de chaves de API |
| Defesa contra injeção | Consultas parametrizadas apenas via ORM | Análise estática; testes de injeção |
| Defesa contra cross-site scripting | Sanitização de HTML no momento da escrita | Conjunto de testes de sanitização de HTML |
| Rate limiting | Limitador durável baseado em base de dados em endpoints de auth | Testes de rate-limit; verificações de burst em ambiente ativo |
| Integridade de webhook | Verificação de assinatura do fornecedor no corpo bruto | Conjunto de testes de webhook |
| Gestão de segredos | Cofre gerido, purge protection, managed identity | Infrastructure as code; auditoria de arquitetura |
| Isolamento de rede | Private endpoints; segmentação default-deny | Infrastructure as code; auditoria de arquitetura |
| Eliminação de dados | Eliminação em cascata por unidade única com audit log | Conjunto de testes de eliminação GDPR |
| Cadeia de abastecimento | Passos de pipeline fixados; monitorização semanal de dependências | Configuração de pipeline; auditoria de dependências |

# Apêndice B: Perguntas Frequentes para Revisores de Segurança

**Onde estão armazenados os nossos dados?** Inteiramente dentro da União Europeia, em Microsoft Azure, em West Europe com processamento de IA em regiões da UE. Os dados de candidatos nunca saem da UE.

**Os nossos dados são usados para treinar modelos de IA?** Não. O fornecedor de IA não utiliza dados de clientes para treino.

**A base de dados é acessível a partir da internet?** Não. O acesso de rede pública está desativado e a base de dados só é alcançável através de um private endpoint dentro da rede virtual.

**Um cliente pode ver os dados de outro cliente?** Não. Cada consulta é delimitada à organização do autor do pedido, o acesso entre organizações devolve "not found", e uma matriz automatizada testa continuamente este isolamento.

**Como são armazenadas as palavras-passe?** Com hash bcrypt e um salt único por palavra-passe. Single sign-on com Microsoft e Google é suportado, caso em que nenhuma palavra-passe é armazenada.

**Suportam single sign-on?** Sim, via OAuth da Microsoft e Google.

**Durante quanto tempo os tokens de acesso são válidos?** Trinta minutos, emparelhados com uma sessão de refresh revogável no lado do servidor que é invalidada no logout.

**Como é tratado o consentimento do candidato?** Cada candidato recebe um link de consentimento único e de utilização única e tem de aceitar antes de qualquer gravação ou análise. O consentimento é registado contra o processo de contratação específico.

**Como são eliminados os dados?** Como uma única unidade cobrindo o registo do candidato, entrevistas, transcrições, áudio, documentos e comparações, segundo um calendário de retenção configurável, com prova registada de apagamento. Os candidatos também podem pedir eliminação diretamente.

**Têm um data processing agreement?** Sim, aceite no registo e versionado por organização, incluindo o registo de sub-processors.

**A IA toma decisões de contratação?** Não. Fornece apenas suporte à decisão; um humano revê cada saída e toma todas as decisões.

**Como provam as vossas alegações de segurança?** Através de 3,171 testes automatizados, incluindo um conjunto dedicado de segurança, uma metodologia repetível de penetration testing em seis fases executada contra ambientes ativos, um programa de testes de segurança de IA e relatórios de auditoria escritos e recorrentes.

**O que acontece quando encontram uma vulnerabilidade?** É-lhe atribuída uma severidade com evidência e um responsável, é remediada segundo uma prioridade definida, é reverificada incluindo verificações ao vivo quando relevante, e é registada num relatório de auditoria.

**Podemos realizar o nosso próprio penetration test?** Avaliações de segurança podem ser organizadas através do seu representante de conta sob âmbito e calendarização adequados.

---

# Apêndice C: Glossário

| Termo | Significado |
|---|---|
| AES-256 | Um padrão forte de encriptação simétrica usado para proteger dados em repouso |
| bcrypt | Uma função de hash de palavras-passe concebida especificamente para esse fim com salt por palavra-passe |
| Managed identity | Uma identidade emitida pela plataforma que permite a um serviço autenticar-se sem chaves armazenadas |
| Private endpoint | Um endereço de rede privado que mantém um serviço cloud fora da internet pública |
| Network security group | Um conjunto de regras de permissão e negação que filtra o tráfego de rede para uma subnet |
| RBAC | Controlo de acesso baseado em funções, concedendo permissões de acordo com a função de um utilizador |
| IDOR | Insecure direct object reference, uma falha de controlo de acesso contra a qual a plataforma se defende |
| SSRF | Server-side request forgery, uma classe de ataque testada nos nossos penetration tests |
| Web application firewall | Um controlo edge que filtra tráfego web malicioso |
| Data processing agreement | O contrato que rege como um processador trata dados pessoais em nome de um responsável pelo tratamento |

# Apêndice D: Contacto e Controlo do Documento

**AI Interview Analyzer Sp. z o.o.**

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

NIP: 5253079974

Para uma revisão de segurança, uma cópia do nosso data processing agreement, ou a nossa documentação de conformidade com o EU AI Act, contacte o seu representante de conta.

*Este documento descreve a postura de segurança do serviço AI Interview Analyzer na data de geração indicada no rodapé. É fornecido para fins de avaliação e não constitui parte de qualquer contrato. Compromissos contratuais específicos de segurança estão definidos no acordo aplicável e no data processing agreement.*