ClearWork Security & Data Protection

This document provides a comprehensive overview of the security architecture and data protection measures implemented in the ClearWork Automated Discovery Platform. Our architecture is designed using a defense-in-depth strategy to ensure enterprise-grade security for customer data, processes, and interview content.

Executive Summary

The ClearWork platform is built on a defense-in-depth security architecture that protects customer data across all phases: authentication, storage, network transmission, and AI processing. All security controls adhere to industry best practices and enterprise standards, providing a robust and secure environment. Key Security Highlights:

- Zero Client-Side AI API Exposure: All Artificial Intelligence (AI) calls are executed exclusively on the server side. Client applications cannot make direct AI API calls.
- No AI Training From Your Data: Data sent to third-party AI providers (e.g., Google Gemini, OpenAI) under enterprise agreements is NOT used for model training. There is no customer data retention by these providers, and processing occurs in real-time.
- Data Isolation: Customer data is strictly segregated at the organization level using unique IDs, making cross-organization data access impossible.
- Encryption In-Transit and At-Rest: Data is protected with TLS 1.3 during network transmission and is encrypted while stored in the database and file storage.
- Row-Level Security (RLS): Security policies are enforced directly at the PostgreSQL database level on every table, ensuring users only access data they are explicitly authorized to see.
- Role-Based Access Control (RBAC): Access is governed by granular roles (e.g., Super Admin, Admin, Project Manager, Contributor) with a strict permission hierarchy.
- Comprehensive Auditing: Complete logs are retained for all AI operations and administrative actions for compliance and traceability.

---

1. **Authentication & Identity Management**

1.1 User Authentication

- Provider: We utilize a PostgreSQL-backed authentication system built on SOC 2 Type II certified infrastructure.
- Method: Authentication uses JSON Web Tokens (JWT) for session management.
- Token Security: Session tokens are cryptographically signed and verified on every request. Tokens automatically expire and refresh to mitigate replay attacks.
- Access Keys: A Publishable Key is used client-side for read-only access with RLS enforcement, while a Service Role Key is strictly reserved for server-side privileged operations and is never exposed to clients.

1.2 Role-Based Access Control (RBAC)

Access to platform features and data is strictly controlled by defined roles.

Customer Administrator Has broad management Admin privileges. The Customer Administrator can access and see everything within the customer account.

Project Manager Can create and manage projects and team assignments related to projects that they create or have been assigned to. They are blocked from viewing any projects they are not a part of within their customer account.

Contributor Can participate in interviews and provide responses as well as access Collaboration Spaces in a read only manner.

---

2. **Database Security & Data Isolation**

2.1 Row-Level Security (RLS)

- Mandatory RLS: RLS policies are enabled on all database tables. This crucial security feature ensures that even if application logic were bypassed, the database itself would prevent unauthorized data access.
- Access Policy Example: Policies are written in the database to ensure users can ONLY view projects and data within their own organization ID.

2.2 Data Isolation

- Organization-Level Segregation: Every customer organization is assigned a unique org_id (UUID).
- All customer data, including projects and interviews, is linked to the organization's unique org_id.
- The combination of RLS and org_id linkage makes cross-organization data access impossible, even in the case of attempted SQL injection.

---

3. **File Storage Security**

3.1 Private Storage Buckets

Customer files, including video recordings and attachments, are stored in isolated, private storage buckets.

These buckets are configured as PRIVATE and are not publicly accessible.

3.2 Access Policies and Data Deletion

- Strict Access: File downloads require a valid authentication token, and RLS verifies user ownership to control access.
- Video Processing: Temporary files created during video processing (e.g., video frames) are automatically deleted immediately after processing is complete. The storage path for these temporary files includes the user ID to maintain segregation. Users maintain the option in admin settings to toggle automated deletion OFF. This will change all videos to be stored by default instead of deleted by default.

---

4. **AI Processing Security**

4.1 Secure Architecture and API Key Management

- Server-Side Only: AI API keys are NEVER exposed to the client. All AI-related calls and processing take place exclusively within secure, server-side Edge Functions.
- Data Flow: The architecture flow ensures that AI prompts and responses never traverse untrusted networks.
- Key Storage: All AI API keys and service keys are encrypted at rest in a secure vault and are only accessible by authenticated Edge Functions.

4.2 AI Usage Tracking & Auditing

- Complete Audit Trail: Every single AI call is logged. The audit log records details such as the User ID, Project ID, AI model used, token usage (input/output), estimated cost, duration, and success/failure status.
- Third-Party AI Privacy: Data sent to third-party AI providers (e.g., Google Gemini, OpenAI) under enterprise agreements is NOT used for model training. There is no customer data retention by these providers, and processing occurs in real-time.

---

5. **Compliance & Governance**

5.1 Transport Layer Security

- All network connections utilize TLS 1.3 for end-to-end encryption, ensuring data is fully protected while in transit.

5.2 Administrative Activity Logging

- All administrative actions are logged in a dedicated audit log. Logged actions include user role assignments, project creation/deletion, and system configuration updates.

5.3 GDPR & Privacy Compliance

The platform's design supports core GDPR principles:

- Right to Access: Users can query all their data.
- Right to Deletion: Account deletion is a cascading event, ensuring that all associated user and project data is completely and automatically removed.
- Data Minimization: Only necessary PII is collected, PII is scrubbed from AI processing, and temporary files are deleted automatically.

---

6. **Cloud Platform**

6.1 All data is stored in Supabase

- Supabase is a highly secure database and platform hosted in AWS. All security documentation can be found here: https://supabase.com/security

---

This architecture meets or exceeds security standards for handling sensitive business data and personally identifiable information (PII). The multiple layers of protection, from database-level RLS to server-side AI processing, provide robust assurance that your data is secure.