Twelve Labs Security FAQ

Do you have formal security policies?

Yes. We maintain a full security policy set that is reviewed and updated at least annually.

How often are audits and reviews performed?

We have established a formal audit program that includes continual, independent internal and external assessments and we undergo a SOC 2 audit annually. Management reviews the security program and conducts risk assessments at least annually.

Do you hold an independent attestation? Period covered?

Yes, we hold a SOC 2 Type II report covering April 9, 2024 to July 9, 2024 with no exception noted and are currently awaiting our completed report for 2025.

Are background checks performed before access is granted?

Yes. Background verification is performed prior to access.

Do employees and contractors sign confidentiality agreements?

All employees sign confidentiality acknowledgments during onboarding, and must acknowledge all security policies including the Code of Conduct.

Do you train employees on security?

Security awareness training is required within 30 days of hire and refreshed annually.

How is access managed and revoked?

Access is centrally managed through Okta and Google Workspace. Requests require manager approval, and all permissions are reviewed quarterly with changes tracked. Access is revoked within 24 hours of termination Access requests require manager approval.

Are unique accounts required for access?

Yes. Production access uses unique usernames or authorized SSH keys; shared logins are not used.

Is MFA required for remote or privileged access?

Yes. All production and administrative remote access requires MFA and an approved encrypted connection.

What is your password policy?

The password policy enforces length and complexity, reuse limits, and lockouts; Okta and 1Password are used. Credentials are stored in a secrets management system and protected. Passwords include lockouts for failed attempts

Where is the service hosted and who handles physical security?

AWS hosts our production environment; AWS provides physical and environmental controls; we manage logical and data-layer security.

How is your network secured?

Production systems reside in isolated virtual private clouds with firewall controls, strict security groups, and segmented networks. Firewall rule sets are reviewed at least annually and changes are tracked.

Do you employ intrusion detection and continuous monitoring?

Yes, AWS GuardDuty and other tools provide continuous detection, with centralized logging and alerting.

How is remote access secured?

Only authorized engineers can access production environments. Remote sessions require VPN and MFA; production access requires encrypted channels with MFA.

Is data encrypted at rest and in transit?

TLS is enforced for sensitive data in transit, and all stored data and backups are encrypted using industry-standard encryption.

How are keys managed?

Yes. Access to encryption keys is restricted to authorized users; keys are managed via AWS KMS.

Do you follow a data classification policy?

Data is classified by sensitivity, and handling requirements are defined for each including access and lifecycle rules.

Can customers delete or export data?

Customers can initiate deletions/exports via the product or API or upon request. Data is removed on contract termination.

Is production data used in non-production environments?

No. Non-production uses synthetic or anonymized data.

Do you notify customers of material system changes?

Yes. Commitments and material changes are communicated through public customer terms and support channels.

Are security and admin events centrally logged?

Yes, administrative and security events are logged centrally and monitored continuously, our logs capture user IDs, timestamps, actions, and results for accountability.

Is there a documented, tested incident response plan?

Yes, we have a formal plan with defined roles and escalation paths. IR is documented and tested at least annually with incidents logged and tracked.

How quickly will customers be notified of a breach?

If we discover a confirmed security incident involving customer data, we notify affected customers as soon as possible and within 72 hours.

Are penetration tests performed at least annually with remediation?

Yes, independent testing is annual and findings are tracked to closure under SLAs.

Are vulnerability scans performed on a defined cadence?

Yes, vulnerability scanning runs continuously through our CI/CD pipeline for source code and dependencies, and at least quarterly on external-facing infrastructure. Findings are triaged by severity, and remediation is tracked to closure within defined SLAs.

Do you maintain BCP/DR plans and test them at least annually?

Yes, we have documented plans that include communication, restoration, and redundancy procedures. Plans exist and are tested annually.

Are backups performed?

Yes. Backups are encrypted, controlled, monitored for completion, and periodically restored for integrity.

Do you assess vendors for security and privacy?

Yes, we evaluate all vendors with access to company or customer data before onboarding and annually thereafter.

Do contracts include confidentiality and privacy commitments?

All third-party agreements include confidentiality and data-protection requirements.

Do you comply with global privacy regulations?

Twelve Labs meets its contractual privacy obligations as a data processor/service provider, and our controls are designed to support compliance with customer requirements under frameworks such as GDPR. Formal compliance is defined by customer agreements and our published Privacy Policy.

Do you design new features with privacy in mind?

New product features go through our secure development process, which includes design reviews, change approvals, and assessment of potential security or privacy impacts before release.

Do you communicate product changes that impact customers?

Yes. Major updates or changes on how we handle data are communicated to customers.

Do you publish product and service descriptions?

Documentation is public and provided to users.

How do customers report incidents or issues?

Customers can contact our security team directly at security@twelvelabs.io or use our support portal to report issues.

Where can we read your security and privacy documentation?

Our Security and Privacy pages as well as our Trust Center are publicly available, with access to reports and documents available upon request under NDA.

AI FAQ

Please provide a system description of the AI offering, including use cases.

- Video → Embedding
- Natural language search of video library
- Text generation from video

How are model inputs and outputs processed and utilized within your system?

- Video → Embedding
- Embedding → Text generation, natural language search

Which models?

Marengo model (Video Embedding Model): Marengo analyzes video frames, their temporal relationships, speech, and sound to create "embeddings" – numerical representations that enable sophisticated information retrieval. In short, Marengo powers "any-to-any" search across text, audio, image, and video.

Pegasus model (Video Language Model): Pegasus is a video-first language model that integrates visual, audio, and speech information to achieve advanced text generation from video. It can generate concise summaries, captivating highlights, effective hashtags, and customized reports.

What types of client data does Twelve Labs store?

Twelve Labs stores client logins (name, email), along with AI outputs (embeddings, text generation).

How is your model trained, and are any datasets utilized that may have copyright, intellectual property, or licensing restrictions?

AI models are trained on massive datasets of multimodal video data. This includes not just video frames but also the corresponding audio (speech and sounds) and any associated text

(like captions or on-screen text). Twelve Labs uses video data for both the initial training of foundation models as well as the fine-tuning of trained models for specific, customer-defined applications and use cases.

Obtaining training data for foundation models that meet our methodologies' exacting specifications is a resource-intensive process. It includes obtaining third-party proprietary licensed data and, when mutually agreed upon, using customer data for fine-tuning as part of a customer agreement. Fine-tuning data, as distinct from training data, is customer-specific in that it is sourced from—and used solely for—a specific customer. In no circumstance do we use customer data to train our base foundational models or fine-tuned models without customer consent.

Data Usage & Training, Contracts & Policies

Customer Data is not used to train, fine-tune, or improve AI models. Specifically, the service undertakes not to use, in the context of tests, production data containing personal information or any other confidential information without first having:

- Obtained a Client authorization,
- Deleted or anonymized any sensitive content,
- Logged the reproduction and the use of production data in a test environment.

We implement logical network segregation between our production, test, and development environments using isolated cloud hosting accounts and or logical isolation provided by the storage systems. Contractually, our ToS/DPA prohibits the use of customer data for training AI models. We offer custom Service Level Agreements (SLAs) related to data handling and privacy.

Training Data Protection

Training data used within our services is stored following our Data management and Access control policies. We implement multiple layers of protection, including:

- Encryption in transit and at rest using industry-standard algorithms (via AES-256)
- Role-Based Access Control following the principle of least privilege
- Multifactor authentication for user access
- Audit logs monitoring all activity on the data
- Conducting regular security assessments and audits to identify and address potential vulnerabilities.
- Logical Isolation in multi-tenant environments

How do you monitor and address bias, fairness, and human oversight in your models, and how do you prevent misuse by end-users?

Twelve Labs applies a combination of human review, data diversity, and quantitative evaluation to ensure fairness, accuracy, and reliability:

- **Curate diverse datasets:** Datasets are curated to reduce over-representation of any one domain and to mitigate bias.
- **Quantitative evaluations:** We apply objective evaluations to detect bias and measure model performance across varied contexts.
- **Human data curation and validation:** Humans are involved in the initial stages of data preparation—curating datasets, labeling data, and validating quality and relevance for training—to ensure the AI learns from reliable information.
- **Model evaluation and testing:** Our AI models undergo rigorous evaluation, including manual testing by humans. This involves evaluating the AI's performance on diverse scenarios, identifying potential failure points, assessing the quality and appropriateness of its outputs, and uncovering unexpected behaviors that automated tests might miss.
- **Ongoing feedback and reporting:** We continuously monitor customer feedback and leverage user reports to identify and address incorrect, biased, or inappropriate outputs.

Additionally, Twelve Labs actively monitors usage patterns and user-reported activity to detect and mitigate potential misuse or abuse of its AI systems by end-users, ensuring responsible and ethical deployment.

Are your models subject to risk assessments, such as model inversion?

Given Twelve Lab's use case of turning videos into embeddings, an attacker would be unable to recreate any training data from the output. The risk of model inversion is minimal.

How do you protect against prompt injection attacks and adversarial inputs?

The NLP search functionality has been designed to minimize prompt-injection risk. The search query is turned into an embedding; there is no database query that occurs based on the user text input, nor is there a generative component of this process—so the risk of prompt injection is low.

To further protect against adversarial inputs and malicious prompts, multiple layers of defense are employed:

- **Input validation and sanitization:** All user-supplied prompts and video data are validated, sanitized, and checked for malicious payloads.
- **Isolation and guardrails:** Model inference runs in environments with guardrails that limit model behaviors to intended functions.

- **Rate limiting and runtime monitoring:** We combine rate limiting and continuous monitoring to detect and block adversarial patterns.
- **Ongoing security testing:** We regularly run internal tests and monitor the broader security research community to stay ahead of emerging attack vectors.

How do you address situations where decision-making processes need to be audited?

- **Auditability**: We maintain logs of the AI's inputs, internal processing steps (where feasible and relevant), and outputs for critical decision points. This creates an audit trail that can be examined.
- **Decision Provenance**: Where possible, we aim to trace back the reasoning behind a decision, linking it to the relevant data patterns learned during training and the specific input data for that instance.

What level of support and training do you offer to ensure smooth adoption and integration? What kind of customizability is available to fit our specific requirements?

Our platforms are API driven which significantly reduces the need for training. We offer support to build with our API's and basic training for users related to prompt engineering. There are multiple options for support including email, phone and slack. Each customer is unique in their requirements and we will work with you to develop a structure that meets your business requirements.

What APIs do you offer?

- **Video Indexing API**: API endpoint accepting video uploads that processes content through multimodal AI model, generating embeddings and metadata.
- **Search API**: Vector-based similarity search for query responses across high-dimensional embedding spaces.
- **Query API**: Natural language processing interface that converts user queries into embedding representations compatible with indexed video content.
- **Analyze API**: Employs our foundational Visual Language Model (VLM), which integrates a language encoder to extract multimodal data from videos and a decoder to generate concise text representations.

Is there an incident response procedure in the event of AI security being compromised?

We use the same incident management policy as the rest of our engineering systems.

What energy efficiency or sustainability metrics do you track?

While we do not have a sustainability policy at this time and plan to roll it out in 2026, we do track and optimize for:

- GPU utilization: To ensure high efficiency during training and inference.
- Realtime ratio measuring the efficiency of our video processing pipeline.

Output accountability: Who is accountable for errors or harm caused by the AI system?

How output is used by a customer is determined solely by the customer.