Willow Learn Ltd – Security & Safeguarding Measures

Last Updated: July 30, 2025

1. Overview

Willow is designed to protect the privacy, safety, and wellbeing of students and educators. Our platform includes security and safeguarding mechanisms that align with the UK GDPR, the Children's Code, and statutory obligations under the Keeping Children Safe in Education (KCSIE 2023) guidance.

2. Role-Based Access Control (RBAC)

- All users are assigned access roles: Administrator, Teacher, or Student.
- Teachers and administrators can view and manage student activity; students have restricted access.
- Access to backend systems is strictly limited to authorised Willow staff.
- Permissions are reviewed monthly and revoked when no longer required.

3. Logging and Audit Trails

- All user activity is logged and timestamped, including task creation, AI interactions, and moderation events.
- Logs distinguish between student and teacher actions.
- Access logs are securely stored and reviewed quarterly for anomalies.
- DSLs (Designated Safeguarding Leads) may access activity logs for compliance purposes.

4. Content Moderation and AI Filters

- Willow filters all AI-generated responses in real time to block:
  - Violent or abusive content
  - Self-harm or suicide-related responses
  - Hate speech, discriminatory language, and bias
  - Misinformation or unsafe advice
- AI outputs are never used to make automated decisions about students.
- Moderation tools are updated regularly based on emerging risks and educator feedback.

5. Safeguarding Alerts and Risk Categories

Willow flags high-risk student inputs in near real-time and routes alerts to teacher dashboards.
Alerts may cover:
- Self-harm or suicidal ideation
- Abuse or grooming indicators
- Bullying, threats, or peer conflict
- Substance use or violence
- Hate speech or identity-based harassment

Each alert includes:
- Timestamp and triggering text
- Risk category
- User role and session context

6. Safeguarding Escalation Policy

Escalation operates on a 3-tier framework:
- Tier 1: Teacher monitors flagged content and addresses low-risk events.
- Tier 2: DSL (Designated Safeguarding Lead) reviews moderate to high-risk alerts using full audit logs.
- Tier 3: Willow's safeguarding contact (dpo@willowlearn.com) is notified in cases of systemic risk or false-negative alerts.

Named Willow contact: Designated Safeguarding Lead – Jamie Munro dpo@willowlearn.com

7. Real-Time Monitoring Details

- Most alerts are generated and routed within 5–30 seconds.
- During outages or degraded performance, alerts are queued and delivered as soon as service is restored.
- Willow performs periodic QA of flagged prompt accuracy and response time.

8. Staff Training and Access Governance

- Willow staff with access to production data undergo annual safeguarding training.
- All access is controlled by RBAC and reviewed monthly.
- Internal access to flagged content is strictly limited to technical and safeguarding staff.

9. Review and Compliance

- Security and safeguarding policies are reviewed quarterly.
- This document supports school compliance with the UK GDPR, the Children's Code, and

KCSIE 2023.
- All updates are version-controlled and shared with partner schools proactively.