CORSEN AI Academy

AI Moderation and Security with Mistral

Course syllabus

Version 2026-11

Overview

Secure your AI applications with Custom Guardrails and the Moderation API featuring 11 content categories, detection thresholds, and content policy implementation in production.

Learning objectives

- Implement Custom Guardrails that block a request before the model call
- Call the Moderation API and interpret scores for the eleven content categories
- Set detection thresholds consistent with the defined content policy
- Detect and neutralize prompt injection attempts in user inputs
- Design a complete security architecture, GDPR compliant, for a production application

Audience and prerequisites

Developers, engineers, and integrators looking to leverage Mistral AI via API in their applications.

Ability to read Python and JSON, and basic command line skills. A Mistral AI Studio API key to try the examples.

Duration and format

Estimated duration : 1 h 15 min. 100% online self-paced course, accessible 24/7 from a browser. Progress is saved, allowing you to resume at any time. Estimated duration: 1 h 15 min.

Content

- Custom Guardrails (4 lessons)
- Moderation API (4 lessons)
- Production Security (4 lessons)

Assessment

Self-assessment based on the Key takeaways concluding each lesson, and direct implementation of the examples on your own examples. A final quiz of 5 corrected questions at the end of the last lesson validates acquired skills.

Accessibility

Course accessible to people with disabilities: text content compatible with screen readers, keyboard navigation, and compliant contrasts. For any adaptation needs, email hello@corsen.ai.

Price and recognition

Free

Certificate of completion issued by Corsen AI.

Access time

Immediate after account creation and email confirmation.

Corsen AI — Paris, France — SIRET 10297040700015
academy@corsen.ai — https://academy.corsen.ai