BLACKBIRD FOR TRAINING

Safe and Responsible Use of Generative AI

UK Traininig

PARTNER

www.blackbird-training.com

Safe and Responsible Use of Generative AI

Introduction

With the rapid spread of Generative AI in various work environments, it has become essential to understand the risks and challenges associated with using these advanced technologies. Employees today face new challenges related to information security, data protection, and dealing with AI-generated content. Awareness of security and responsible use of Generative AI has become an urgent necessity, not just an option.

The course "Safe and Responsible Use of Generative AI" has been designed to educate all employees, regardless of their job levels, about the security and ethical risks associated with using Generative AI tools such as Gemini. The course aims to equip participants with the knowledge and skills necessary to use these tools safely, protect confidential data, identify misinformation, and adhere to organizational policies and ethical frameworks.

This course is not just technical training; it is an investment in the organization's security culture and building collective awareness about responsible technology use, contributing to protecting the organization's reputation, data, and employees from increasing risks in the digital age.

Course Objectives

This course aims to empower all employees with an integrated set of security and ethical knowledge and skills, summarized in the following objectives:

1. Understanding the Security Risks of Generative AI: Recognizing the risks associated with using Generative AI tools, including data leakage, breaches, and misuse.
2. Familiarity with Organizational Security Policies: Understanding SPH Media's policies regarding AI use, confidential data protection, and regulatory compliance.
3. Safe Use of AI Tools: Learning how to use Gemini and other AI tools safely, including handling sensitive information and avoiding common risks.
4. Detecting Misinformation: Developing the ability to distinguish between human-generated and AI-generated content, and using source verification tools.
5. Understanding AI Ethics: Grasping the principles of transparency, fairness, and responsibility in AI use, and applying them in the workplace.
6. Building an Organizational Security Culture: Contributing to creating a safe and responsible work environment by sharing experiences and security challenges with colleagues.

Course Outlines

Level: Awareness All Levels
Duration: 2 Days 8 Training Hours
Target Audience: All Employees

Day 1: Security Fundamentals and Responsible Use

1.1 Course Opening and Introduction

Topics Covered:

- Welcoming participants and ice-breaking activities
- Reviewing course objectives and expected outcomes for all employees
- Defining Generative AI and its importance in the modern work environment
- The importance of security in the age of Generative AI
- Overview of increasing security risks
- Setting expectations and ground rules for participation
- Surveying participants' experiences with AI tools

1.2 Information Security and Privacy

Topics Covered:

- Definition of Information Security: Understanding basic concepts of information security in the context of AI
- Risks of Using Generative AI Tools:
    - Leakage of sensitive data customer data, financial information, trade secrets
    - Unknowingly sharing confidential information
    - Data storage on external servers
    - Privacy risks for employees and customers
    - Data usage for model training purposes
- Real-World Examples of Breaches and Security Incidents:
    - Data leakage cases from major companies
    - Examples of employees sharing confidential information with AI tools
    - Consequences of breaches on reputation and trust
    - Lessons learned from past incidents
- Types of Sensitive Data in the Workplace:
    - Personal data of employees and customers
    - Financial information and budgets
    - Confidential documents and contracts
    - Intellectual property and designs
    - Organizational strategies and plans
- Best Practices for Information Protection:
    - Classifying data by sensitivity level
    - Determining what can be shared with AI
    - Using secure and approved tools
    - Adhering to privacy policies
- Employee Responsibility in Information Protection:
    - Every employee is the first line of defense
    - Reporting suspected security incidents
    - Participating in building a strong security culture

1.3 Security Policies at SPH Media

Topics Covered:

UK Traininig
PARTNER

- Organizational AI Usage Policies:
    - Approved tools
    - Prohibited or restricted tools
    - Procedures for obtaining approval to use new tools
    - Policy updates and changes
- Confidential Data Protection:
    - Definition of confidential data in the organizational context
    - Procedures for handling confidential data
    - Penalties for violating data protection policies
    - The employee's role in maintaining confidentiality
- Regulatory Compliance:
    - Local and international data protection laws
    - Security standards in the printing and media sector
    - Client and partner compliance requirements
- Incident Reporting Procedures:
    - How to report a security incident
    - Approved reporting channels
    - Reporting timeframe
    - Confidentiality of reporting and whistleblower protection
- Policy Updates and Continuous Training:
    - Importance of regular policy updates
    - Continuous employee training
    - Communication channels for security awareness
- Q&A on Policies:
    - Opening discussion to clarify any policy ambiguities
    - Sharing practical examples to illustrate application

1.4 Responsible Use of Gemini

Topics Covered:

UK Traininig
PARTNER

- What is Gemini? Brief overview of Gemini and its capabilities
- How to Use Gemini Safely:
    - Not sharing personal or confidential information
    - Not uploading sensitive documents
    - Using the secure enterprise version if available
    - Logging out after use
    - Using strong passwords
- Handling Sensitive Information:
    - Identifying information that should not be shared with Gemini
    - Anonymization techniques when data sharing is necessary
    - Using synthetic data for testing
    - Documenting what has been shared with AI
- Common Mistakes to Avoid:
    - Sharing personally identifiable information
    - Uploading sensitive internal documents
    - Using Gemini for unauthorized tasks
    - Relying on outputs without verification
- Privacy and Security Settings:
    - Configuring privacy settings in Gemini account
    - Understanding data retention policies
    - Options for deleting conversations and data
- Practical Guidelines for Daily Use:
    - Checklist before using Gemini
    - Steps for safe use in daily tasks
    - Handling outputs with caution
- Case Studies:
    - Examples of safe Gemini use
    - Examples of unsafe use and consequences

1.5 Practical Exercises - Security Scenarios

Topics Covered:

UK Traininig
PARTNER

- Exercise 1: Potential Data Leak Scenario:
    - Reading a scenario about an employee who shared confidential information with Gemini
    - Analyzing the situation: What went wrong?
    - Identifying potential consequences
    - Suggesting corrective actions
- Exercise 2: Suspicious Email Scenario:
    - Analyzing an email claiming to be from Gemini
    - Identifying signs of fraud and phishing
    - Correct procedures for handling the email
- Exercise 3: Document Sharing Scenario:
    - Evaluating whether a specific document is safe to share with Gemini
    - Identifying sensitive information in the document
    - Suggesting safe alternatives for sharing
- Exercise 4: Breach Scenario:
    - What to do if you discover your Gemini account has been breached?
    - Incident response steps
    - Reporting channels and procedures
- Exercise 5: New Policy Scenario:
    - Applying a new security policy to a real work scenario
    - Identifying implementation challenges
    - Suggesting policy improvements
- Discussion of Exercise Outputs:
    - Sharing solutions with the group
    - Feedback from trainer and peers
    - Extracting lessons learned

1.6 Group Discussion - Sharing Experiences and Security Challenges

Topics Covered:

UK Traininig
PARTNER

- Sharing Personal Experiences:
    - Participants sharing experiences with AI tools
    - Security challenges they faced
    - Solutions they used
- Common Security Challenges in the Workplace:
    - Identifying recurring challenges
    - Analyzing root causes
    - Suggesting shared solutions
- Best Practices from Participants:
    - Successful tips and experiences
    - Effective preventive measures
    - Useful tools and techniques
- Hypothetical Scenarios for Discussion:
    - What if you faced a specific security situation?
    - How would you handle it?
    - What are the lessons learned?
- Building an Internal Security Network:
    - Encouraging collaboration among colleagues on security issues
    - Creating channels for sharing security information
    - Supporting each other in security challenges
- Group Recommendations:
    - Suggesting improvements to policies and procedures
    - Initiatives to strengthen security culture
    - Practical steps for improvement

1.1 Week 1 Summary and Q&A Session

Topics Covered:

- Comprehensive review of Day 1 concepts:
    - Information security and privacy risks
    - SPH Media security policies
    - Responsible use of Gemini
    - Security scenarios and challenges
- Summary of key points:
    - Every employee is the first line of defense
    - Security awareness is the best protection
    - Immediate incident reporting
- Answering participants' questions and inquiries
- Preview of Day 2 content Misinformation and Ethics
- Preparation tasks for the next day if any

UK Traininig
PARTNER

Day 2: Misinformation and Ethics

2.1 Day 1 Recap and Review

Topics Covered:

- Quick retrieval of key concepts from Day 1 through interactive questions
- Reviewing security policies and responsible use
- Sharing participants' experiences after Day 1
- Addressing any remaining queries
- Setting the stage for Day 2 topics
- Connecting Day 1 concepts to ethics and misinformation

2.2 Distinguishing Between Human and AI-Generated Content

Topics Covered:

- Understanding AI-Generated Content:
  - How AI content generation works
  - Types of generated content text, images, video, audio
  - AI capabilities in generating realistic content
- Misinformation and Deepfakes:
  - Definition of misinformation
  - Examples of deepfakes fake images and videos
  - Impact of misinformation on individuals and organizations
  - Consequences of misinformation spread
- Content Verification Tools:
  - AI-generated content detection tools
  - Image and video verification tools
  - Source and reference verification tools
  - Metadata analysis tools
- Warning Signs of Suspicious Content:
  - Grammatical or logical errors
  - Inaccurate or outdated information
  - Unreliable sources
  - Requests for personal information
  - Overly emotional content
- Information Verification Strategies:
  - Source verification
  - Reverse image search
  - Comparing information with reliable sources
  - Using specialized verification tools
  - Questioning unusual information

UK Traininig
PARTNER

- Employee Responsibility in Information Sharing:
    - Verifying before sharing
    - Not sharing unconfirmed information
    - Reporting misinformation
    - Promoting a verification culture in the workplace

2.3 Practical Exercises - Analyzing Suspicious Content

Topics Covered:

- Exercise 1: Analyzing Suspicious Text:
    - Reading AI-generated text
    - Identifying signs that it is AI-generated
    - Verifying sources and references mentioned
    - Assessing content reliability
- Exercise 2: Analyzing Suspicious Images:
    - Examining an image potentially edited or generated
    - Using reverse image search tools
    - Analyzing metadata if available
    - Determining whether the image is real or fake
- Exercise 3: Analyzing a Short Video:
    - Watching a video clip potentially a deepfake
    - Identifying manipulation signs
    - Using video verification tools
    - Assessing video credibility
- Exercise 4: Analyzing Suspicious Email:
    - Reading an email containing misinformation
    - Identifying signs of fraud
    - Verifying sender address and sources
    - Correct procedures for handling the email
- Exercise 5: Analyzing Social Media Post:
    - Analyzing a social media post
    - Verifying information accuracy
    - Identifying potential motives for the post
    - How to respond or handle the post
- Discussion of Results:
    - Sharing analysis results with the group
    - Discussing verification challenges
    - Extracting lessons learned
    - Developing better verification strategies

2.4 Generative AI Ethics

Topics Covered:

UK Traininig
PARTNER

- Core Ethical Principles:
    - Transparency: Disclosing AI use
    - Fairness: Avoiding bias and discrimination
    - Responsibility: Taking responsibility for outputs
    - Privacy: Respecting individual privacy
    - Security: Protecting data and systems
- Transparency in AI Use:
    - Disclosing the use of AI-generated content
    - Clarifying AI's role in decisions
    - Explaining how models work to users
    - Transparency with colleagues and customers
- Fairness and Avoiding Bias:
    - Understanding bias in AI models
    - How to detect bias in outputs
    - Impact of bias on decisions and employees
    - Strategies to reduce bias
- Responsibility Toward Colleagues and the Organization:
    - Ensuring accuracy of AI outputs
    - Not blindly relying on AI
    - Reporting errors and problems
    - Protecting organizational reputation
- Responsibility Toward Customers and Society:
    - Protecting customer data
    - Providing accurate and reliable information
    - Transparency with customers about AI use
    - Contributing to positive technology use
- Ethical Frameworks in the Organization:
    - AI ethics policies
    - Ethics and oversight committees
    - Ethical review procedures for projects
    - Continuous ethics training
- Ethical Case Studies:
    - Examples of ethical dilemmas in the workplace
    - Analyzing ethical decisions
    - Potential consequences of unethical decisions
    - Lessons learned from past cases

2.5 Interactive Workshop - Complex Ethical Cases

Topics Covered:

UK Traininig
PARTNER

- Workshop Structure:
    - Dividing participants into small groups
    - Assigning a complex ethical case to each group
    - Time for analysis and group discussion
    - Presenting findings and recommendations
- Sample Ethical Cases:
    - Case 1: AI in Hiring:
        - Using AI to screen job applicants
        - Risks of bias in hiring process
        - How to ensure fairness in AI-supported hiring
    - Case 2: AI-Generated Content in Marketing:
        - Using AI to generate marketing content
        - Transparency with customers about AI-generated content
        - Risks of misleading in marketing
    - Case 3: Employee Privacy:
        - Using AI to monitor employee performance
        - Balancing monitoring and privacy
        - Clear workplace privacy policies
    - Case 4: Data Security in Joint Projects:
        - Sharing data with external partners using AI
        - Protecting confidential data in collaborations
        - Security and ethical agreements
    - Case 5: Responsibility for AI Errors:
        - Who is responsible when AI makes mistakes?
        - Procedures for handling errors
        - Learning from errors and improving models
- Ethical Analysis Methodology:
    - Identifying conflicting ethical values
    - Analyzing available options
    - Evaluating potential consequences of each option
    - Choosing the most ethical solution
    - Developing an implementation plan
- Group Presentations and Discussion:
    - Each group presents their case analysis
    - Discussing different perspectives
    - Trainer and peer feedback
    - Extracting key ethical lessons
- Developing Ethical Policies:
    - Proposing ethical policies for handling similar cases
    - Recommendations for improving current practices
    - Steps to strengthen ethical culture in the organization

2.6 Course Summary, Assessment, and Conclusion

Topics Covered:

UK Traininig
PARTNER

- Comprehensive Review:
    - Summarizing Day 1 concepts Security Fundamentals
    - Summarizing Day 2 concepts Misinformation and Ethics
    - Connecting all concepts in an integrated framework
    - Re-emphasizing key points
- Key Takeaways from the Course:
    - Security is everyone's responsibility
    - Verify information before sharing
    - Transparency and ethics in AI use
    - Reporting incidents and security concerns
- Short Quiz:
    - Assessing security and ethical knowledge
    - Questions on security policies
    - Questions on detecting misinformation
    - Questions on ethics and responsible use
    - Immediate feedback and explanation of answers
- Course Evaluation Survey:
    - Collecting participant feedback on content quality, trainer performance, and materials
    - Assessing the usefulness of the course for participants
    - Suggestions for improvement
    - Identifying additional training needs
- Recommendations and Continuity:
    - Practical steps to apply what was learned
    - Resources for continuous learning on security and ethics
    - Channels for reporting security and ethical concerns
    - Encouraging participation in security initiatives
- Certificate Distribution:
    - Awarding certificates of completion to participants
    - Recognizing their commitment and efforts in enhancing security and responsibility

Conclusion

At the conclusion of this course, participants will have gained a deep understanding of the security and ethical risks associated with using Generative AI, and are now equipped with the knowledge and skills necessary to use these tools safely and responsibly in the workplace.

Through understanding information security risks, organizational security policies, and safe handling of Gemini, participants are now able to protect themselves and their organization from growing threats. Their ability to distinguish between human and AI-generated content, and their commitment to ethical principles, makes them part of the solution in facing the challenges of the digital age.

Building a strong security and ethical culture is an investment in the organization's future. Every employee is a fundamental line of defense in protecting data, reputation, and the trust that the organization has built with its customers and partners. This course is an important step in strengthening this culture and ensuring that AI is used as a positive and safe tool in achieving the organization's goals.

UK Traininig
PARTNER