Co-funded by the Erasmus+ Programme of the European Union

secureu DIGITAL EDUCATION TOOLS FOR SECURITY RISK MANAGEMENT

ERASMUS+ Strategic partnership "Digital education tools for security risk management"
Project number: 2021-1-LV01-KA220-HED-000023056

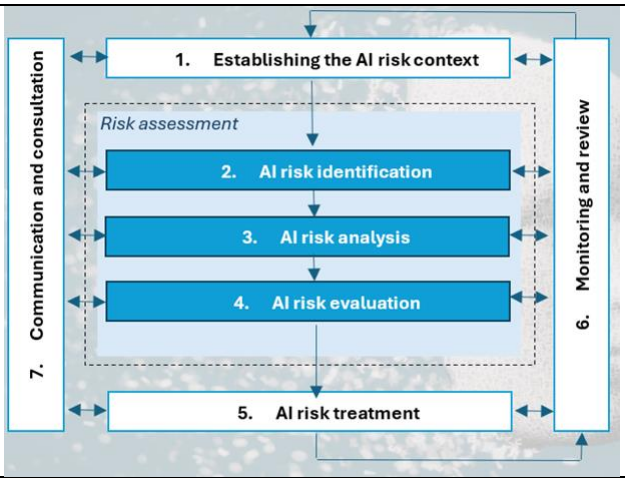
EXERCISE FOR SECURITY STUDENTS

AI Risk Assessment and Treatment Using ISO 31000

AUTHOR: Rita Lankauskienė, Kazimieras Simonavicius University, Lithuania

BACKGROUND:

The fast growth of generative artificial intelligence (AI) technology and the widespread use of creative AI solutions have led to the quick appearance of new risk types. This makes the already complicated processes of creating and implementing AI even less predictable. More and more problems are happening because of the (mis)use of AI (e.g., AI Incident Database, 2024; AIAAIC, 2024). These problems affect a lot of people, groups, and governments at all levels (national, international, and global). Using a risk management method to find risks, analyse them, rate them, and treat them is how risk management practices try to deal with core uncertainties. In this exercise, the ISO31000 standard methodology is modified and applied to assess the AI risk in real life AI cases to model the AI risk treatment scenarios.

REFERENCE TO ISO 31000 STANDRD	
The ISO 31000 standard offers a systematic methodology for risk management that may be tailored to the particular requirements of AI security, for implementing an AI risk assessment in particular circumstances.	

GOAL OF THIS EXERCISE:

The students will learn to apply the ISO31000 standard methodology-based logic to assess the AI risk in selected real life AI cases and to model the AI risk treatment scenarios.

TASK DESCRIPTION FOR STUDENTS:

1. Form the groups as instructed by teacher and allocate the responsibilities within the group:
   - Who will guide throughout the discussion process, following the task questionnaire and supportive material?
   - Who will fix the discussion summary highlights in written form?
   - Who will present the final group work outcomes to the class?

2. Familiarize yourself and your groupmates shortly with the core building blocks of risk assessment logic, embedded in the ISO31000 standard. On demand, check for additional material for the best practice guidelines, developed by the Internet of Things Security Institute (IoTSI), on how to conduct an AI security risk assessment using ISO 31000. (Available at: https://iotsecurityinstitute.com/iotsec/index.php/iot-security-institute-blog/155-conducting-an-ai-security-risk-assessment-using-iso-31000.) You may also read the material, presented in the

3. Select the case of a (mis)use of AI from AI Incident Database (2024): https://incidentdatabase.ai/. Remember to fix the general details regarding the selected case, as given in the questionnaire guidance.

4. Conduct an AI security risk assessment using ISO 31000 methodology: step-by-step implement all security assessment steps for your selected case of a (mis)use of AI. Carefully follow the questionnaire. Be flexible in forming supportive questions on demand.

5. Discuss withing a group and select one identified risk to prepare a risk treatment scenario.

6. Agree on the final outcome of your group, and prepare up to 10-minute presentation, which include the following aspects:
   - Title and source of the selected case;
   - Team members, who worked on the output;
   - The general description of the selected case of a (mis)use of AI (up to 5 sentences);
   - The internal and external context of a (mis)use of AI;
   - AI risk assessment core steps: identification, analysis and evaluation;
   - Risk treatment plan (scenario building) for selected risk: actions, resources and responsibilities.
   - Summary feedback on the most complicated and most successful stages of applying ISO31000 methodology for AI risk assessment and treatment planning.

7. Listen to other presentations of other groups. After each presentation, engage in a two-minute discussion with your group to determine whether their approach to the selected cases would have been pertinent to you as well. In your turn, articulate your views to the class.

8. Discuss in your group which of the presented approaches would be most appropriate for each case after all presentations have been completed. Please share your thoughts with the class.

TASK DESCRIPTION FOR TEACHER / TRAINER:

The teacher's tasks are as follows:

1. Estimate the number of students and the number of groups of approximately up to 5-7 students that will be formed prior to the commencement of the class.
2. Each group should have access of the supportive material (depending on the live/remote mode):
   - AI security risk assessment table, based on ISO 31000 guidelines (Annex 1);
   - ISO 31000: 2018 framework (Annex 2);
   - internet to select the case of a (mis)use of AI from AI Incident Database (available at: https://incidentdatabase.ai/) prior to the commencement of the class;
   - A1 format paper sheet, colourful post-it notes and markers (or equivalent software in a distanced mode).
3. Distribute students into groups of approximately up to 5-7 students.
4. Instruct students to familiarise with the ISO 31000:2018 standard, as well as the case on "AI Security Challenge and Risk Assessment Using ISO 31000".
5. Instruct the students how to record the outcomes of their preferred approach, following the "AI security risk assessment table, based on ISO 31000 guidelines", provided in Annex 1. Remind, that this table is elaborated from the case on "AI Security Challenge and Risk Assessment Using ISO 31000", which the students must be already familiar. Discussion outcomes might be accomplished through various methods, such as post-it notes, a PowerPoint presentation, a whiteboard, or an online environment. Students should be advised on how to prepare for the presentation of their findings. Provide instructions, what the presentation should include (see above). Remind students, that the presentation may last to a maximum of 10 minutes.
6. During the presentations, ensure that the discussion is chaired and that the groups remain within the designated time frame during presentations. The procedure is as follows:
   - One group presentation is expected to last no more than 10 minutes.
   - Following each presentation, groups should engage in a two-minute discussion within their respective groups to determine whether the approach that was presented would have been pertinent to their own approach.
   - The class is encouraged to hear the perspectives of the other groups.
   - The fundamental inquiry is: are the elaborated risk treatment plans lead to expected outcomes?
7. Follow each presentation with a discussion among all students regarding the most effective approach.

ADDITIONAL SKILLS THAT THE STUDENT ACQUIRES THROUGH THIS ASSIGNMENT:

- Working in a group
- Working under time pressure
- Giving presentation and arguing the case
- Comparison skills and critical thinking

This activity focuses on risk assessment, with the primary purpose being a comprehensive evaluator of the risks associated with the increasing adoption of AI in certain contexts. Through this process, the students will not only become acquainted with the ISO31000 standard methodology and logical framework, but will also learn to adapt it by critically evaluating the implications and effects of AI challenge in real life cases, in different spheres of human activity.

SUPPORT MATERIALS

Relevant references:

- AI Incident Database (2024). Available at: https://incidentdatabase.ai/
- IoT Analytics. State of IoT, Summer 2024. Market Report. Available at: https://iot-analytics.com/product/state-of-iot-summer-2024/
- Conducting an AI security risk assessment using ISO 31000 (2024). IoTSI. Available at: https://iotsecurityinstitute.com/iotsec/index.php/iot-security-institute-blog/155-conducting-an-ai-security-risk-assessment-using-iso-31000
- ISO 31000:2018. Risk management – Guidelines (2018). Available at: https://www.iso.org/obp/ui/en/#iso:std:iso:31000:ed-2:v1:en

Annex 1

AI security risk assessment table, based on ISO 31000 guidelines

Case data: - Title: - Link to the selected case: - Date of access:		
Group data: - Moderator: - Raporteur: - Output-generating actors:		
AI SECURITY RISK ASSESSMENT PROCEDURE		
	Contents	*Notes*
Establishing the context		
Internal context:	- Regulatory landscape; - Market and technological trends; - Threat landscape.	
External context:	- Organizational structure; - Risk management policies; - Risk appetite and tolerance.	
AI Risk Assessment		
AI risk identification	*Data risks:* - *breaches,* - *data poisoning,* - *data integrity.* *Model risks:* - *adversarial attacks,* - *model stealing,* - *model bias.* *Operational risks:* - *system failures,* - *security configuration,* *third-party risks.*	
AI risk analysis	*Impact assessment:* - *financial impact,* - *operational impacts*, - *reputational impact*, Likelihood assessment: - *historical data*, - *vulnerability analysis,* - *threat actor capability.*	
AI risk evaluation	*Risk matrix:* - High - Moderate - Low *Decision-making:* - everyone involved, - sepatare stakeholders involved, - stakehoder groups invoved.	

Risk Treatment Plan		
The *objective* - define clearly what a particular treatment is supposed to do, like lowering the risk of data leaks or weakening the effects of hostile attacks.	Objective:	
The *actions* - specify to tell the people what they need to do, like putting in place multi-factor login, encrypting data, or doing regular security checks.	Actions:	
Equip the risk treatment measures with reasonable *resources*, staff, and technology they need to be put into action.	Resources:	
Clearly assign *responsibilities* for executing the treatment plan, ensuring accountability and oversight	Responsibilities:	
Risk treatment plan example: To mitigate the risk of model bias, a treatment plan may encompass diversifying training data, applying fairness metrics, and performing regular audits to identify and rectify biases.		

Source: elaborated by author, based on IoTSI guidance (2024). Conducting an AI security risk assessment using ISO 31000. Available at: https://iotsecurityinstitute.com/iotsec/index.php/iot-security-institute-blog/155-conducting-an-ai-security-risk-assessment-using-iso-31000.

Annex 2

ISO 31000: 2018 framework

Source: ISO 31000:2018. Risk management – Guidelines (2018). Available at: https://www.iso.org/obp/ui/en/#iso:std:iso:31000:ed-2:v1:en .