**Students are to answer 1 of the 2 cases presented. See below for essay length and formatting requirements. The included article gives relevant background information. Students can use any resource given in class or found online. Online resources should be evidence-based on peer-reviewed sources.**

## Case 1: Shared Mental Models in a Human–AI Security Operations Centre

A national cyber security operations centre (SOC) has deployed an AI-based decision support system that correlates logs, prioritises alerts, and recommends response actions in major incidents. After several serious attacks, reviews show that analysts, incident managers, and the AI system often operate with incompatible understandings of what is happening, what the priorities are, and who should do what. Some analysts treat the AI's risk scores as final decisions, others largely ignore them; incident commanders complain that they cannot see "what the AI thinks is going on", while the AI frequently recommends actions that do not match existing playbooks or the commander's intentions. Debriefs repeatedly reveal automation surprises, conflicting interpretations of the same data, and confusion about goals, roles, and plans. Management believes that the core problem is not only trust in automation, but misaligned shared mental models and shared situation awareness within the human–AI team.

Drawing on the attached journal article on situation awareness and shared mental models in human–AI systems, analyse the SOC's breakdowns, explain how gaps in individual and team situation awareness contribute to these problems, and show how more aligned shared mental models could improve performance. Propose and justify concrete changes to interface design, team roles and procedures, and training or exercises that could help the human analysts and the AI system develop, maintain, and update shared mental models during dynamic incidents. Critically evaluate the likely benefits and limitations of your proposal, including possible unintended consequences and outline how you would empirically assess whether shared situation awareness and team performance have actually improved.

Endsley, M. R. (2023). Supporting Human-AI Teams: Transparency, explainability, and situation awareness. *Computers in Human Behavior*, *140*, 107574.

**Case 2:**

**Exam Task: Organisational Culture, Mental Health, and Insider Threats in a Technology Firm**

A technology company providing cloud services has experienced several near-miss security incidents involving privileged employees. In one case, a system administrator copied large amounts of production data to a personal device shortly before resigning. An internal investigation found no malicious intent but revealed high levels of stress, burnout, and perceived injustice among technical staff. Employee surveys show low psychological safety, weak leadership support for mental health, and a belief that "security is mostly about tools, not people." At the same time, the security team has started to view dissatisfied employees primarily as potential insider threats, which has damaged trust even further.

The board has concluded that the organisation's culture and leadership practices may be increasing both psychosocial risk and security risk. You are brought in as an external consultant to design a comprehensive, human-centred insider threat mitigation strategy that does not rely solely on surveillance or technical controls.

Your task is to propose an integrated programme that addresses organisational culture, leadership behaviour, mental health, and security practices as interconnected elements. Use relevant theories and empirical research on insider threats, occupational stress and burnout, organisational justice, and safety/security climate. Outline concrete interventions at the individual, team, and organisational levels and critically evaluate ethical tensions between privacy and monitoring, and explain how you would assess the effectiveness and potential unintended effects of your proposed programme over time.

Khan, N., J. Houghton, R., & Sharples, S. (2022). Understanding factors that influence unintentional insider threat: a framework to counteract unintentional risks. *Cognition, Technology & Work*, *24*(3), 393-421.

**Formatting of Essay**

The essay can be **3000 ± 10% words** in main body (abstract and references do not count) in 12-point Times New Roman font. Your essay should be typed and double-spaced on standard-sized paper (A4) with 1" margins (standard in Word Document) on all sides. Include a page header (also known as the "running head") at the top of every page. The running head is a shortened version of your paper's title and cannot exceed 50 characters including spacing and punctuation. Not meeting this requirement can result in a lower grade. Word count does not include pictures with no text. Tables will count towards the final word count. If pictures are used for tables, they will count as 500 words. The essay shall have a title page with Title and word count listed, abstract, main body, and references. This essay is an academic text.

**See https://taltech.ee/en/formatting-guidelines for more information**

**For this essay, APA 7th is to be used (see: https://apastyle.apa.org/style-grammar-guidelines/references/examples)**

**Other resources for APA 7th (I personally use this one):**
**https://owl.purdue.edu/owl/research_and_citation/apa_style/apa_style_introduction.html**

## Grading

See https://taltech.ee/en/grading-system for Taltech Grading policy

**§ 14. Assessment of academic performance**
(1) The methods and criteria of assessment defined in syllabus shall be available to students before the commencement of studies and they must not be changed during a semester. The assessment methods define the manner of attesting the acquisition of knowledge and skills (e.g. an oral or written examination, pass/fail assessment, an essay, a report, group work, a questionnaire). If various methods are used for the assessment of learning outcomes, their relevant weights in determining the final grade shall be specified in the syllabus. An assessment criterion shall specify the expected level and scope of knowledge which can be proved by the assessment methods.

(3) **In case of graded assessment, the achievement of learning outcomes is assessed based on the following scale**:
A (5) – "excellent" – outstanding and particularly profound achievement of learning outcomes, along with creativity and consummate proficiency in applying skills and knowledge;
B (4) – "very good" – very good achievement of learning outcomes, along with proficiency in applying skills and knowledge in a targeted and creative manner. Some details of knowledge and skills may exhibit errors which are neither substantive nor serious;
C (3) – "good" – good achievement of learning outcomes, along with proficiency in applying skills and knowledge in a relevant manner. A certain imprecision and uncertainty are apparent in the depth and detail of knowledge and skills;
D (2) – "satisfactory" – sufficient achievement of learning outcomes, along with application of knowledge and skills in a typical manner; in atypical situations both, uncertainty as well as lack of knowledge and skills are apparent.
E (1) – "poor" – minimum acceptable achievement of the most important learning outcomes along with limited application of knowledge and skills in typical situations; in atypical situations both, considerable uncertainty as well as lack of knowledge and skills are apparent;
F (0) – "failed" – achievement in knowledge and skills below the minimum standard.