UNIVERSITY
OF SKÖVDE

School of Information Technology

WRITTEN EXAMINATION

Course: Big Data Programming

Examination: Written exam (discontinuation, opportunity 1 out of 3)

Course code: IT739A

Credits for written examination: 5.5hp

Date: August 26, 2026

Examination time: 8:15 – 12:30

Examination responsible: Richard Senington

Teachers concerned: Juhee Bae, Richard Senington, Gunnar Mathiason

Aid at the exam/appendices

Other

Instructions

- ☐ Take a new sheet of paper for each teacher.
- ☒ Take a new sheet of paper when starting a new question.
- ☒ Write only on one side of the paper.
- ☒ Write your name and personal ID No. on all pages you hand in.
- ☒ Use page numbering.
- ☒ Don´t use a red pen.
- ☒ Mark answered questions with a cross on the cover sheet.

Grade points

Examination results should be made public within 18 working days

Good luck!

UNIVERSITY
OF SKÖVDE

IT739A, Exam, August 26th, 2026, at 8:15-12:30

Instructions that must be followed:

- The max number of points listed with each question indicates how thoroughly you are expected to answer the question.
- Number your answer clearly with the same numbering as the question and its sub-questions.
- Be clear in your writing. Sometimes fewer but more concise formulations are better than writing a lot of text.
- More points assigned indicate that a more detailed answer is expected.

The grading of your answers considers:

- The **correctness** of your answer/explanation
- The **clarity and logical cohesion** of your answer/explanation
- The **conciseness** of your answer/explanation

Grading levels:

- There are 5 questions (with sub-questions), which give a maximum 10 points each.

1) You need to pass 5 points per each of the 5 questions to pass the exam (25 points).
2) Accumulated points above that threshold linearly determines the exam grading (the exam maximum is 50 points).

- Exam grading range is A to F, which also determines the final course grading:
  A: 45-50 points, B: 40-44 points, C: 35-39 points, D: 30-34 points, E: 25-29 points.

Examination goals:

This exam assesses the following Course Curriculum goals

- *critically evaluate the relevance and application of advanced AI methods,*
- *in an in-depth manner account for the development of advanced AI methods,*
- *apply and demonstrate appropriate analysis strategies using advanced AI tools and techniques,*
- *in an in-depth manner account for ethical use of AI to solve complex problems*

The exam questions begin on the next page.

UNIVERSITY
OF SKÖVDE

Exam questions:

1. Deep Learning, Keras and Tensorflow (10 points):

1. When evaluating a trained machine learning model, you want to compare your model performance towards a *baseline*. What is a baseline, and why should you use one? (2p)
2. Briefly explain two separate reasons how the *Tensorflow Data API* helps in building better data preparation pipelines than without this API? (2p)
3. When evaluating the performance of a trained model, you need to reserve some hold-out data to use as a *test set*? Why do you need to do that? (2p)
4. Explain what is meant with *leakage* of information when you try to improve a model's accuracy by tweaking the data or the data preprocessing. (2p)
5. Convolutional neural networks (CNN): What is the *underlying assumption* about the data used with CNNs, and how does that assumption help a CNN use fewer weights than a fully connected neural network? (2p)

2. Concepts and ethics in Big Data (10 points):

1. One of the 5Vs of Big Data is Variety. Please describe the Variety aspect of big data and discuss the issues that arise from this property with respect to usage in Neural Networks. (3p)
2. One of the 5Vs of Big Data is Volume. Please describe the Volume aspect and describe 1 storage technique that can help to resolve it. You must be exact about how the technology you choose works and resolves this particular issue. (3p)
3. Discuss ethical issues related to the training, deployment and usage of LLMs. Consider the potential benefits of LLMs, the impacts of using them, applications they might be put to and whether it is right to train and deploy them. (4p)

3. Anomaly Detection (10 points):

1. How do auto-encoders work, and how do Denoising auto-encoders vary the basic design. Why might it help anomaly detection. (4p)
2. Describe a GAN. Explain the different parts of the GAN, how they work together and how you train it. (3p)
3. Explain how you can use a GAN for anomaly detection. (1p)
4. Explain how PADIM makes use of transfer learning. (2p)

4. Recurrent Neural Networks (10 points):

1. Describe what a recurrent neural network is, how it differs from a standard neural network and why it is effective for working with data series. (3p)

2. Describe a trainable neural network gateway and how it is intended to help in a network. (3p)
3. What is the vanishing gradient problem and how is it different in the context of recurrent networks? (2p)
4. Describe a GRU network and explain why it is superior to a basic recurrent network. (2p)

5. Large Language Models (LLM) (10 points):

1. Explain how a Transformer determines the attention score between two tokens. What does this attention score represent? (2p)
2. State whether the statement is true or false, and if it is false, explain why. Each question is worth 1 point, and an incorrect True/False answer automatically receives 0 points regardless of any justification. (4p)
   a) Self-attention allows each token to directly attend to all other tokens in the sequence. True or False? Justify if false.
   b) Without positional encoding, a Transformer can still distinguish word order in a sequence. True or False? Justify if false.
   c) Attention scores are computed using dot products between query and value vectors. True or False? Justify if false.
   d) Each token in self-attention shares the same query, key, and value vectors. True or False? Justify if false.
3. Why can LoRA adapt large language models without updating all parameters, and what is the main limitation of this approach? (2p)
4. Why is correct retrieval not sufficient to guarantee correct answers in RAG systems? Provide at least two reasons. (2p)