



D4.2 Specific Curriculum



DATA ANALYST EQF 6



Co-funded by
the European Union

Copyright © 2024 Artificial Intelligence Skills Alliance. The project resources contained herein are publicly available under the [Creative Commons license 4.0 B.Y.](https://creativecommons.org/licenses/by/4.0/)

Disclaimer

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.



Co-funded by
the European Union

Project information

The Artificial Intelligence Skills Alliance (ARISA) fast-tracks the upskilling and reskilling of employees, job seekers, business leaders, and policymakers into AI-related professions to open Europe to new business opportunities. It is a four-year transnational project funded under the EU's Erasmus+ programme. For more information, contact info@aiskills.eu | aiskills.eu

Project Partners



List of abbreviations

AI	Artificial Intelligence
ARISA	Artificial Intelligence Skills Alliance
EQF	European Qualification Framework
EU	European Union

Table of contents

1. GENERAL INFORMATION 4

2. DESCRIPTION OF THE STRUCTURE 5

3. OVERVIEW OF LEARNING UNITS..... 7

4. DETAILS OF LEARNING UNITS 8

4.1. UNSUPERVISED LEARNING 8

4.2. SUPERVISED LEARNING – CLASSIFICATION 10

4.3. SUPERVISED LEARNING – REGRESSION..... 12

4.4. MINING OF UNSTRUCTURED DATA: TEXT AND IMAGE ANALYTICS..... 15

4.5. NEURAL NETWORKS FROM GROUND-UP 17

4.6. GENERATIVE AI - LARGE LANGUAGE MODEL FROM GROUND-UP..... 19

4.7. DATA SCIENCE, AI, AND SOCIETY..... 22

1. General information

Name	Data Analyst
EQF level	EQF 6 for all units with exception on neural networks units where EQF level is 7 and 8
Goals	<ol style="list-style-type: none"> 1. Master Machine Learning Techniques: Equip learners with essential skills in supervised and unsupervised learning, including clustering, classification, and regression. 2. Develop Data Analysis Skills: Enable proficient exploratory data analysis, visualization, and implementation of data analytic workflows. 3. Design and Build Advanced Neural Networks: Foster comprehensive knowledge in constructing, optimizing, and implementing neural networks and deep learning models from scratch. 4. Apply Practical Machine Learning Tools: Train learners to effectively use machine learning tools, recognize opportunities for their application, and implement solutions in various real-world scenarios. 5. Ensure Ethical AI Deployment: Instill understanding of ethical considerations, explainability, and best practices for responsible AI implementation.
Scope	<p>Scope of the Curriculum: The curriculum provides a comprehensive and practical education in machine learning, data science, and AI, covering a broad range of topics from foundational concepts to advanced techniques. It encompasses unsupervised and supervised learning methods, neural network construction, text and image analytics, and ethical AI deployment. The training is designed to be accessible, requiring no prior knowledge of math or statistics for most learning units, and leverages advanced training techniques and modern visual analytics software to facilitate rapid and effective learning.</p> <p>Target Groups: The curriculum is tailored for employees in the industry and public sector who aim to delve deeply into machine learning, data science, and AI. It is ideal for professionals seeking to enhance their skills and apply cutting-edge technologies in their work. The pedagogical approach ensures that learners can quickly grasp complex concepts using advanced training techniques and state-of-the-art visual analytics tools, enabling them to achieve competency efficiently and effectively.</p>
Entry requirements	No prior knowledge of mathematics, statistics, computer science, or programming is required. Exceptions are "Neural Networks from Ground-Up", which requires basic knowledge of Python and its programming environment (editing code, basic debugging, running Python applications),

	and "Generative AI - Large Language Model from Ground-Up", which requires solid Python programming skills.
Programme learning outcomes (PLOs)	<ul style="list-style-type: none"> 1 - Deep Learning (EQF 6) 2 - PLO AI Technologies (EQF 6) 3 - ML Ops (EQF 6) 4 - Machine Learning (EQF 6) 5 - Explainable AI (EQF 6) 6 - Big Data & Data Analytics (EQF 6) 7 - Human-Centered AI (EQF 6) 8 - AI Ethics (EQF 6) 9 - AI Futures and Innovation (EQF 6) 10 - Business Intelligence (EQF 6) 11 - AI Awareness (EQF 6) 13 - Generative AI (EQF 6) 14 - Soft Skills (EQF 6)

2. Description of the structure

The curriculum is structured into distinct yet interconnected learning units, each focusing on a specific aspect of data science, machine learning, and AI, emphasizing a modular approach for a comprehensive understanding. The "Unsupervised Learning" unit lays a foundation in data exploration and clustering techniques, essential for understanding data patterns without labeled outcomes. This is complemented by the two "Supervised Learning" units, which train the classification and regression techniques, providing a detailed study of predictive modeling, accuracy assessment, overfitting prevention, and model explanation and understanding. The unit on "Mining of Unstructured Data" extends the knowledge to text and image analytics, introducing practical applications in unstructured data through deep neural network embedding and combining the data analysis techniques from the previous learning units. Again, this unit emphasizes pattern discovery, interpretation, and data understanding.

While these introductory machine learning units emphasize an intuitive understanding of machine learning techniques and utility in practical cases, the "Neural Networks from Ground-Up" offers an in-depth mathematical and practical understanding of neural network construction and optimization, paving the way for the "Generative AI - Large Language Model from Grounds-Up," where students replicate GPT-2, enhancing their skills in implementing and optimizing large-scale AI models. These two units were designed for those who want to gain a deep understanding of AI and state-of-the-art methods and require prior knowledge of Python programming.

The curriculum concludes with "Data Science, AI, and Society," which integrates ethical considerations, explainability, and industry practices, ensuring a holistic education that prepares

students for real-world applications and responsible AI deployment. Like other units, the training focuses on examples, case studies, and practical work.

Each unit in this Data Analyst curriculum builds on the previous ones, ensuring a seamless transition from fundamental to advanced topics while also highlighting the practical interconnections and applications of the concepts learned.

3. Overview of Learning Units

Learning unit title	Hours/ECTS	EQF level	Assessment
Unsupervised Learning	50 / 2	6	Quizzes, Project Report
Supervised Learning - Classification	50 / 2	6	Quizzes, Project Report
Supervised Learning - Regression	50 / 2	6	Quizzes, Project Report
Mining of Unstructured Data: Text and Image Analytics	50 / 2	6	Quizzes, Project Report
Neural Networks from Ground-Up	50 / 2	6	Quizzes, Project Report
Generative AI - Large Language Model from Grounds-Up	50 / 2	7	Quizzes, Project Report
Data Science, AI and Society	50 / 2	8	Quizzes, Project Report

4. Details of Learning Units

4.1. Unsupervised Learning

Description of “Unsupervised Learning”

This unit delves into unsupervised learning techniques, starting with data representation and exploratory data analysis. It covers statistical fundamentals such as distributions and correlations, complemented by extensive data visualization methods. The curriculum includes an introduction to data analysis workflows and visual programming, with a focus on various clustering methods such as hierarchical, k-means, DBSCAN, and Gaussian mixture modelling using the expectation maximization algorithm. Dimensionality reduction techniques such as principal component analysis, multidimensional scaling, and t-SNE are explained. The unit also explores distance measurement, silhouette scoring for cluster validation, and techniques for explaining point-based visualizations.

No prior knowledge of programming, computer science, statistics, or mathematics is required.

Related Programme Learning Outcomes

- 2 - PLO AI Technologies (EQF 6)
- 3 - ML Ops (EQF 6)
- 4 - Machine Learning (EQF 6)
- 5 - Explainable AI (EQF 6)
- 6 - Big Data & Data Analytics (EQF 6)
- 7 - Human-Centered AI (EQF 6)
- 10 - Business Intelligence (EQF 6)
- 11 - AI Awareness (EQF 6)
- 14 - Soft Skills (EQF 6)

Unit learning outcomes

Program-specific learning outcomes will enable course participant to acquire the following skills:

- Communicates technical details and project outcomes related to AI technologies
- Implements AI projects by applying foundational ML Ops principles
- Adapts machine learning models to evolving data and requirements
- Communicates effectively with stakeholders involved in AI projects
- Applies fundamental machine learning concepts and algorithms
- Evaluates machine learning models using established metrics and validation techniques
- Designs data pre-processing and feature engineering strategies
- Communicates machine learning findings effectively
- Develops AI models using explainable AI techniques to ensure transparency and interpretability

- Implements strategies to improve the explainability of existing AI models
- Creates effective data visualizations that clearly communicate analytical findings
- Enhances human-computer interaction through the design of intuitive and accessible AI interfaces
- Designs and implements data visualization techniques to effectively communicate
- Understands the basic concepts and technologies underlying artificial intelligence
- Solves problems creatively and efficiently
- Thinks critically, analysing situations, evaluating diverse perspectives

Learning unit-specific outcomes related to gaining an intuitive and operational understanding of AI methods include acquiring knowledge and skills in:

- Feature-based data representation
- Exploratory data analysis methods
- Data distributions, correlations
- Various data visualization techniques, including box plots, distribution graphs, and scatter plots
- Data analysis workflows and visual programming in Orange Data Mining
- Measuring distances between data instances and between data clusters
- Hierarchical clustering method
- Cluster explanation
- Silhouette Scoring
- Complexity of clustering, clustering of large data sets, k-means clustering
- Other types of clustering, including DBSCAN and Gaussian Mixture Modeling, expectation-maximization algorithm
- Dimensionality reduction techniques: projections and embedding
- Principal component analysis, multidimensional scaling, and t-SNE
- Explanation of point-based visualizations

Delivery methods

- On-site lectures with practical demonstrations of methods, tools, and cases
- Self-study assisted with pre-recorded video material and interactive lecture notes with quizzes and challenges
- Project-based study with structured delivery of findings
- On-line discussion of project results and quizzes
- On-line forum, discussions and assistance on Discord channel

Material

- Lecture notes with practical examples, cases, and demonstrations of how to use data mining tool Orange, delivered electronically to in-class participants
- Quizzes and interactive lecture notes in a mini-MOOC environment
- About 15 short educational videos
- Rescha S (2024) Machine Learning Q & AI: 30 Essential Questions and Answers on Machine Learning and AI, No Starch Press (optional reading, selected chapters)
- Kneusel RT (2023) How AI Works, No Starch Press (optional reading, selected chapters)

4.2. Supervised Learning – Classification

Description of “Supervised Learning – Classification”

This unit introduces the fundamentals of supervised learning with a focus on classification. It begins with data representation techniques specific to classification problems and an introduction to classification trees, discussing their interpretation and accuracy assessment. The unit covers crucial concepts such as overfitting, accuracy estimation through cross-validation, and the use of diverse classifiers including random forests, k-nearest neighbours, gradient boosting, logistic regression, and the naive Bayesian classifier. It also explores model performance metrics like sensitivity, specificity, and the area under the ROC curve. Additionally, the curriculum addresses the explainability of complex models using SHAP values.

No prior knowledge of programming, computer science, statistics, or mathematics is required. The successful completion of “Unsupervised Learning” from this Data Science curricula is strongly advised.

Related Programme Learning Outcomes

- 2 - PLO AI Technologies (EQF 6)
- 3 - ML Ops (EQF 6)
- 4 - Machine Learning (EQF 6)
- 5 - Explainable AI (EQF 6)
- 6 - Big Data & Data Analytics (EQF 6)
- 7 - Human-Centered AI (EQF 6)
- 10 - Business Intelligence (EQF 6)
- 11 - AI Awareness (EQF 6)
- 14 - Soft Skills (EQF 6)

Unit learning outcomes

Program-specific learning outcomes will enable course participant to acquire the following skills:

- Analyses the performance and efficiency of AI models and technologies
- Implements AI projects by applying foundational ML Ops principles
- Adapts machine learning models to evolving data and requirements
- Communicates effectively with stakeholders involved in AI projects
- Applies fundamental machine learning concepts and algorithms
- Evaluates machine learning models using established metrics and validation techniques
- Designs data pre-processing and feature engineering strategies
- Communicates machine learning findings effectively
- Develops AI models using explainable AI techniques to ensure transparency and interpretability
- Analyses existing AI models to assess and measure their explainability

- Develops data science solutions to tackle specific analytical challenges
- Creates effective data visualizations that clearly communicate analytical findings
- Incorporates human-centered design principles in the development of AI systems
- Understands the basic concepts and technologies underlying artificial intelligence
- Solves problems creatively and efficiently
- Cultivates an innovative mindset, embracing and fostering creativity
- Communicates effectively across a variety of platforms and media
- Thinks critically, analysing situations, evaluating diverse perspectives

Learning unit-specific outcomes related to gaining an intuitive and operational understanding of AI methods include acquiring knowledge and skills in:

- Classification and data representation for classification problems
- Classification trees
- Interpretation and accuracy of classification trees
- Accuracy on training and test data and effects of overfitting
- Estimation of accuracy, cross-validation
- Other classifiers, including random forests, k-nearest neighbors, gradient boosting, logistic regression
- Nomograms
- Model accuracy scoring through estimation of sensitivity, specificity, area under ROC
- Explainability of complex models and SHAP values

Delivery methods

- On-site lectures with practical demonstrations of methods, tools, and cases
- Self-study assisted with pre-recorded video material and interactive lecture notes with quizzes and challenges
- Project-based study with structured delivery of findings
- On-line discussion of project results and quizzes
- On-line forum, discussions and assistance on Discord channel

Material

- Lecture notes with practical examples, cases, and demonstrations of how to use data mining tool Orange, delivered electronically to in-class participants
- Quizzes and interactive lecture notes in a mini-MOOC environment
- About 15 short educational videos
- Rescha S (2024) Machine Learning Q & AI: 30 Essential Questions and Answers on Machine Learning and AI, No Starch Press (optional reading, selected chapters)
- Kneusel RT (2023) How AI Works, No Starch Press (optional reading, selected chapters)

4.3. Supervised Learning – Regression

Description of “Supervised Learning – Regression”

This unit explores regression analysis within supervised learning, starting with the basics of data representation for regression problems. It covers linear regression and extends to more complex forms such as polynomial regression, discussing key concepts like overfitting and regularization. The unit examines regularization techniques, including L1 and L2, and the impact of regularization on accuracy in training and test datasets. It also introduces other regression methods like random forests, k-nearest neighbors, and gradient boosting. Additionally, the curriculum delves into the selection and tuning of meta-parameters in machine learning, including strategies for searching and estimating accuracy through methods like RMSE and R^2 .

No prior knowledge of programming, computer science, statistics, or mathematics is required. The successful completion of “Unsupervised Learning” and “Supervised Learning – Classification” from this Data Science curricula is strongly advised.

Related Programme Learning Outcome

- 2 - PLO AI Technologies (EQF 6)
- 3 - ML Ops (EQF 6)
- 4 - Machine Learning (EQF 6)
- 5 - Explainable AI (EQF 6)
- 6 - Big Data & Data Analytics (EQF 6)
- 7 - Human-Centered AI (EQF 6)
- 10 - Business Intelligence (EQF 6)
- 11 - AI Awareness (EQF 6)
- 14 - Soft Skills (EQF 6)

Unit learning outcomes

Program-specific learning outcomes will enable course participant to acquire the following skills:

- Analyses the performance and efficiency of AI models and technologies
- Adapts existing AI models to new contexts and problems
- Communicates technical details and project outcomes related to AI technologies
- Implements AI projects by applying foundational ML Ops principles
- Adapts machine learning models to evolving data and requirements
- Communicates effectively with stakeholders involved in AI projects
- Evaluates machine learning models using established metrics and validation techniques
- Designs data pre-processing and feature engineering strategies
- Communicates machine learning findings effectively

- Develops AI models using explainable AI techniques to ensure transparency and interpretability
- Analyses existing AI models to assess and measure their explainability
- Implements strategies to improve the explainability of existing AI models
- Develops data science solutions to tackle specific analytical challenges
- Interprets complex data sets to identify trends, patterns, and insights that inform business strategies
- Designs and implements data visualization techniques to effectively communicate
- Understands the basic concepts and technologies underlying artificial intelligence
- Solves problems creatively and efficiently
- Cultivates an innovative mindset, embracing and fostering creativity
- Communicates effectively across a variety of platforms and media
- Thinks critically, analysing situations, evaluating diverse perspectives

Learning unit-specific outcomes related to gaining an intuitive and operational understanding of AI methods include acquiring knowledge and skills in:

- Regression in general and data representation for regression problems
- Linear regression, criteria function
- Regression accuracy scoring (RMSE, R2)
- Polynomial regression, overfitting
- Regularization
- Effects of regularization on training and test set accuracy
- L1 and L2 regularization, model complexity
- Regularization path
- Other regression techniques: random forests, k-nearest neighbors, gradient boosting
- Meta parameters in machine learning and strategies of meta-parameter search and accuracy estimation

Delivery methods

- On-site lectures with practical demonstrations of methods, tools, and cases
- Self-study assisted with pre-recorded video material and interactive lecture notes with quizzes and challenges
- Project-based study with structured delivery of findings
- On-line discussion of project results and quizzes
- On-line forum, discussions and assistance on Discord channel

Material

- Lecture notes with practical examples, cases, and demonstrations of how to use data mining tool Orange, delivered electronically to in-class participants
- Quizzes and interactive lecture notes in a mini-MOOC environment
- About 15 short p videos
- Rescha S (2024) Machine Learning Q & AI: 30 Essential Questions and Answers on Machine Learning and AI, No Starch Press (optional reading, selected chapters)

- Kneusel RT (2023) How AI Works, No Starch Press (optional reading, selected chapters)

4.4. Mining of Unstructured Data: Text and Image Analytics

Description of “Mining of Unstructured Data: Text and Image Analytics”

This unit provides an overview of mining unstructured data, focusing on text and image analytics. It starts with a recap of logistic regression and an intuitive introduction to neural networks, followed by deep convolutional neural networks and the principles of image embedding. Techniques such as image clustering and classification through transfer learning are explored. The curriculum also introduces large language models and text embedding, discussing how to mine text through clustering, classification, and regression. Additionally, it covers alternative classical text mining approaches, including the bag-of-words model and text preprocessing techniques like TF-IDF.

No prior knowledge of programming, computer science, statistics, or mathematics is required. The successful completion of “Unsupervised Learning”, “Supervised Learning – Regression”, and “Supervised Learning – Classification” from this Data Science curricula is strongly advised.

Related Programme Learning Outcome

- 1 - Deep Learning (EQF 6)
- 2 - PLO AI Technologies (EQF 6)
- 3 - ML Ops (EQF 6)
- 4 - Machine Learning (EQF 6)
- 5 - Explainable AI (EQF 6)
- 6 - Big Data & Data Analytics (EQF 6)
- 7 - Human-Centered AI (EQF 6)
- 10 - Business Intelligence (EQF 6)
- 11 - AI Awareness (EQF 6)
- 14 - Soft Skills (EQF 6)

Unit learning outcomes

Program-specific learning outcomes will enable course participant to acquire the following skills:

- Develops AI models using chosen frameworks and libraries
- Adapts existing AI models to new contexts and problems
- Communicates technical details and project outcomes related to AI technologies
- Implements AI projects by applying foundational ML Ops principles
- Adapts machine learning models to evolving data and requirements
- Communicates effectively with stakeholders involved in AI projects
- Applies fundamental machine learning concepts and algorithms
- Designs data pre-processing and feature engineering strategies
- Communicates machine learning findings effectively

- Implements data analytics methodologies to derive actionable insights from complex and unstructured datasets
- Develops data science solutions to tackle specific analytical challenges
- Interprets complex data sets to identify trends, patterns, and insights that inform business strategies
- Designs and implements data visualization techniques to effectively communicate
- Understands the basic concepts and technologies underlying artificial intelligence
- Knowledge of current trends in AI technology
- Engages in continuous learning to keep pace with rapid advancements in AI technology
- Solves problems creatively and efficiently
- Cultivates an innovative mindset, embracing and fostering creativity
- Communicates effectively across a variety of platforms and media
- Thinks critically, analysing situations, evaluating diverse perspectives

Learning unit-specific outcomes related to gaining an intuitive and operational understanding of AI methods include acquiring knowledge and skills in:

- Logistic regression, and intuitive understanding of neural networks
- Deep convolutional neural networks and image embedding
- Image clustering and classification through transfer learning and image representation in the embedded space
- Utility of large language models for document embedding
- Text mining through clustering, classification and regression using text embedding
- Dimensionality reduction, image and text maps
- Text and image classification
- Alternative classical approaches to text mining using bag-of-words representation, and text pre-processing (including TF-IDF)

Delivery methods

- On-site lectures with practical demonstrations of methods, tools, and cases
- Self-study assisted with pre-recorded video material and interactive lecture notes with quizzes and challenges
- Project-based study with structured delivery of findings
- On-line discussion of project results and quizzes
- On-line forum, discussions and assistance on Discord channel

Materials

- Lecture notes with practical examples, cases, and demonstrations of how to use data mining tool Orange, delivered electronically to in-class participants
- Quizzes and interactive lecture notes in a mini-MOOC environment
- About five short educational videos
- Rescha S (2024) Machine Learning Q & AI: 30 Essential Questions and Answers on Machine Learning and AI, No Starch Press (optional reading, selected chapters)
- Kneusel RT (2023) How AI Works, No Starch Press (optional reading, selected chapters)

4.5. Neural Networks from Ground-Up

Description of “Neural Networks from Ground-Up”

This unit offers a deep dive into the fundamental concepts and mathematics behind neural networks, starting with functions, derivatives, and gradient descent. Participants will learn the principles of backpropagation and engage in hands-on development of a Python-based framework for specifying and training models from scratch. The course includes building our own implementations of linear regression, logistic regression, softmax regression, a basic neural network, and a convolutional neural network. Additionally, the curriculum features a comparative analysis with state-of-the-art deep learning toolboxes like PyTorch, highlighting differences and advantages.

Trainees in this course are required to have basic knowledge of programming in Python and knowledge about Python programming infrastructure (editors, debugging, running Python programs).

Related Programme Learning Outcome

- 1 - Deep Learning (EQF 6)
- 2 - AI Technologies (EQF 6)
- 3 - ML Ops (EQF 6)
- 4 - Machine Learning (EQF 6)
- 5 - Explainable AI (EQF 6)
- 6 - Big Data & Data Analytics (EQF 6)
- 7 - Human-Centered AI (EQF 6)
- 9 - AI Futures and Innovation (EQF 6)
- 10 - Business Intelligence (EQF 6)
- 11 - AI Awareness (EQF 6)
- 13 - Generative AI (EQF 6)
- 14 - Soft Skills (EQF 6)

Unit learning outcomes

Program-specific learning outcomes will enable course participant to acquire the following skills:

- Analyzes the fundamental principles of neural networks
- Designs deep learning models by selecting appropriate architectures
- Applies deep learning techniques to solve problems
- Analyses the performance and efficiency of AI models and technologies
- Communicates technical details and project outcomes related to AI technologies

- Implements AI projects by applying foundational ML Ops principles
- Adapts machine learning models to evolving data and requirements
- Applies fundamental machine learning concepts and algorithms
- Designs data pre-processing and feature engineering strategies
- Identifies suitable applications for deep and shallow neural architectures
- Designs and implements data visualization techniques to effectively communicate
- Thinks critically, analysing situations, evaluating diverse perspectives

Learning unit-specific outcomes related to gaining an intuitive and operational understanding of the AI methods and following topics that include:

- Functions and derivatives
- Gradient descent
- Backpropagation learning
- Development of Python-based framework for model specification and training through back-propagation (from zero!)
- Trainee's own implementation of linear regression, logistic regression, and softmax regression
- Trainee's own implementation of a neural network
- Trainee's own implementation of a convolutional neural network
- Comparison of trainee's own implementation with state-of-the-art toolbox PyTorch for Deep Learning

Delivery method

- On-site lectures with practical demonstrations of methods, tools, and cases
- Self-study assisted with pre-recorded video material and interactive lecture notes with quizzes and challenges
- Project-based study with structured delivery of findings
- On-line discussion of project results and quizzes
- On-line forum, discussions and assistance on Discord channel

Material

- Lecture notes with practical examples, delivered electronically to in-class participants
- Quizzes and interactive lecture notes in a mini-MOOC environment
- Rescha S (2024) Machine Learning Q & AI: 30 Essential Questions and Answers on Machine Learning and AI, No Starch Press (optional reading, selected chapters)
- Kneusel RT (2023) How AI Works, No Starch Press (optional reading, selected chapters)
- Andrej Karpathy's videos on YouTube from playlist "[Neural Networks: Zetto To Hero](#)"

4.6. Generative AI - Large Language Model from Grounds-Up

Description of “Generative AI - Large Language Model from Grounds-Up”

This intensive unit guides participants through reproducing the GPT-2 (124M) model from scratch. Starting with an overview of the model's architecture and parameters, students will implement the neural network module, forward pass, and sampling mechanisms within a Python framework. The course covers detailed aspects of optimization, including overfitting a batch, implementing efficient data loaders, and parameter sharing techniques. Acceleration techniques using GPUs, mixed precision, and kernel fusion are taught to enhance training speed. Advanced topics such as hyperparameter tuning, gradient accumulation, and distributed data parallel setups are also explored. The unit concludes with evaluating the model's performance, comparing it with benchmarks, and enjoying generated text samples.

Participants in this course are required to have good knowledge of programming in Python.

Related Programme Learning Outcome

- 1 - Deep Learning (EQF 6)
- 2 - AI Technologies (EQF 6)
- 3 - ML Ops (EQF 6)
- 4 - Machine Learning (EQF 6)
- 5 - Explainable AI (EQF 6)
- 6 - Big Data & Data Analytics (EQF 6)
- 7 - Human-Centered AI (EQF 6)
- 9 - AI Futures and Innovation (EQF 6)
- 10 - Business Intelligence (EQF 6)
- 11 - AI Awareness (EQF 6)
- 13 - Generative AI (EQF 6)
- 14 - Soft Skills (EQF 6)

Unit learning outcomes

Program-specific learning outcomes will enable course participant to acquire the following skills:

- Analyzes the fundamental principles of neural networks
- Designs deep learning models by selecting appropriate architectures
- Implements deep learning models using relevant frameworks and libraries
- Applies deep learning techniques to solve problems
- Critiques the current trends and advancements in deep learning
- Collaborates effectively in teams to design, implement, and evaluate deep learning projects
- Develops AI models using chosen frameworks and libraries

- Analyses the performance and efficiency of AI models and technologies
- Communicates technical details and project outcomes related to AI technologies
- Implements AI projects by applying foundational ML Ops principles
- Applies fundamental machine learning concepts and algorithms
- Evaluates machine learning models using established metrics and validation techniques
- Designs data pre-processing and feature engineering strategies
- Implements strategies to improve the explainability of existing AI models
- Implements data analytics methodologies to derive actionable insights from complex and unstructured datasets
- Engages in continuous learning and professional development related to big data
- Identifies suitable applications for deep and shallow neural architectures
- Engages in continuous learning and professional development to stay at the forefront of AI research and innovation
- Understands the basic concepts and technologies underlying artificial intelligence
- Knowledge of current trends in AI technology
- Designs generative AI models to create novel content
- Implements Large Language Models (LLMs) in generative AI projects
- Evaluates the performance of generative AI systems
- Engages in continuous learning and professional development
- Solves problems creatively and efficiently
- Cultivates an innovative mindset, embracing and fostering creativity
- Communicates effectively across a variety of platforms and media
- Thinks critically, analysing situations, evaluating diverse perspectives

Learning unit-specific outcomes related to gaining an intuitive and operational understanding of AI methods include acquiring knowledge and skills in:

- Introduction to Generative AI and Language Models
- Understanding and Implementing GPT-2 Architecture
- Training Neural Networks Efficiently
- Data Preparation and Loading
- Optimization and Loss Functions
- Enhancing Training with GPUs and Precision Techniques
- Tuning Hyperparameters for Optimal Performance
- Model Evaluation and Performance Metrics

Delivery method

- On-site lectures with demonstrations of methods, tools, and cases
- Self-study assisted with interactive lecture notes with quizzes and challenges
- Project-based study with structured delivery of findings
- On-line discussion of project results and quizzes
- On-line forum, discussions and assistance on Discord channel

Material

- Lecture notes with practical examples, delivered electronically to in-class participants

- Quizzes and interactive lecture notes in a mini-MOOC environment
- Rescha S (2024) Machine Learning Q & AI: 30 Essential Questions and Answers on Machine Learning and AI, No Starch Press (optional reading, selected chapters)
- Kneusel RT (2023) How AI Works, No Starch Press (optional reading, selected chapters)
- Andrej Karpathy's videos on YouTube from playlist "[Neural Networks: Zetto To Hero](#)"

4.7. Data Science, AI, and Society

Description of "Data Science, AI, and Society"

This unit examines the societal implications of data science and AI, focusing on the interpretability of machine learning models and the evolution of explainability. It discusses the accuracy of machine learning, the trade-offs between explainability and accuracy, and reviews simpler, more explainable models. The unit delves into the explainability of deep neural networks, SHAP values, and addresses data and model biases, including quantitative measures of bias. It covers the AI and data science development cycle, best practices for industry application, and reviews the EU's AI Act.

No prior knowledge required.

Related Programme Learning Outcome

- 1 - Deep Learning (EQF 6)
- 2 - AI Technologies (EQF 6)
- 3 - ML Ops (EQF 6)
- 4 - Machine Learning (EQF 6)
- 5 - Explainable AI (EQF 6)
- 6 - Big Data & Data Analytics (EQF 6)
- 7 - Human-Centered AI (EQF 6)
- 8 - AI Ethics (EQF 6)
- 9 - AI Futures and Innovation (EQF 6)
- 10 - Business Intelligence (EQF 6)
- 11 - AI Awareness (EQF 6)
- 13 - Generative AI (EQF 6)
- 14 - Soft Skills (EQF 6)

Unit learning outcomes

Program-specific learning outcomes will enable course participant to acquire the following skills:

- Identifies suitable applications for deep and shallow neural architectures
- Discusses the ethical implications of deploying deep learning models
- Critiques the current trends and advancements in deep learning
- Discusses the ethical implications of deploying AI technologies
- Critiques the current trends and advancements in AI
- Demonstrates an ongoing commitment to advancing skills and knowledge
- Applies AI quality control measures throughout the machine learning project lifecycle
- Communicates effectively with stakeholders involved in AI projects

- Evaluates the ethical and social implications of deploying AI solutions
- Stays informed about emerging technologies, practices, and standards
- Evaluates machine learning models using established metrics and validation techniques
- Assesses the ethical implications of machine learning projects
- Communicates machine learning findings effectively
- Develops AI models using explainable AI techniques to ensure transparency and interpretability
- Analyses existing AI models to assess and measure their explainability
- Conveys the principles and benefits of explainable AI to stakeholders
- Navigates ethical and legal considerations associated with AI explainability
- Describes the results and implications of data analysis projects to stakeholders
- Manages AI risks by identifying potential ethical, legal, and operational issues associated with AI technologies
- Stays informed about emerging trends and best practices in human-centered AI
- Forecasts future trends in AI technology
- Knowledge of the regulations and ethical use of data and data practices
- Thinks critically, analysing situations, evaluating diverse perspectives

Learning unit-specific outcomes related to gaining an intuitive and operational understanding of AI methods include acquiring knowledge and skills in:

- Interpretation of machine learning models
- Evolution and understanding of explainability
- Assessments of accuracy of machine learning
- Trade-off between explainability and accuracy
- Review of simple and explainable machine learning models
- Explainability of deep neural networks, SHAP values
- Data and model bias, quantitative measures of bias
- AI and data science development cycle, best practices for introduction to industry
- Engineer's understanding of EU's AI Act

Delivery method

- On-site lectures with practical demonstrations of methods, tools, and cases
- Self-study assisted with pre-recorded video material and interactive lecture notes with quizzes and challenges
- Project-based study with structured delivery of findings
- On-line discussion of project results and quizzes
- On-line forum, discussions and assistance on Discord channel

Material

- Lecture notes with practical examples, cases, and demonstrations of how to use data mining tool Orange, delivered electronically to in-class participants
- Quizzes and interactive lecture notes in a mini-MOOC environment
- About 15 short educational videos

- Rescha S (2024) Machine Learning Q & AI: 30 Essential Questions and Answers on Machine Learning and AI, No Starch Press (optional reading, selected chapters)
- Kneusel RT (2023) How AI Works, No Starch Press (optional reading, selected chapters)



Artificial Intelligence Skills Alliance

www.aiskills.eu

info@aiskills.eu



Co-funded by
the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.