



D4.2 Specific Curriculum



DATA SCIENTIST 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 4

3. OVERVIEW OF LEARNING UNITS..... 6

4. DETAILS OF LEARNING UNITS 7

4.1. DATA ANALYTICS 7

4.2. SIGNAL PROCESSING AND TIME SERIES 8

4.3. MACHINE LEARNING: SUPERVISED 9

4.4. MACHINE LEARNING: UNSUPERVISED..... 10

4.5. NEURAL NETWORKS..... 11

4.6. DEEP LEARNING 12

4.7. GENERATIVE AI..... 13

4.8. FOUNDATIONAL MODELS AND LLMs 14

4.9. HUMAN-CENTERED AI..... 15

4.10. AI PLATFORMS & ARCHITECTURES 16

4.11. DATA SECURITY..... 17

4.12. PROFESSIONAL ABILITIES..... 18

1. General information

Name	Data Scientist
EQF level	EQF 6
Goals	The goal of the Data Science program is to provide a solid foundation in core data science techniques, develop advanced machine learning and AI skills, and ensure the ability to build and manage ethical and secure AI systems. These goals aim to produce graduates who are ready to lead and innovate in the data science field, capable of tackling complex problems and contributing to advancements in technology and society. The program is designed to ensure that graduates are proficient in both theoretical and practical aspects of data science and capable of applying their knowledge to real-world problems.
Scope	<i>The course is aimed at people who intend to become Data Scientist/EQF 6. It can also be used for upskilling or reskilling of individuals that already have a background in computer science and want to specialize in Data Science.</i>
Entry requirements	Generic requirements: programming (python), statistics, algebra, calculus, probability, data bases, data structures.
Programme learning outcomes (PLOs)	<ul style="list-style-type: none"> 1 - Deep Learning (EQF 6) 2 - AI Technologies (EQF 6) 3 - Machine Learning (EQF 6) 4 - Explainable AI (EQF 6) 6 - Big Data & Data Analytics (EQF 6) 7 - Human-Centered AI (EQF 6) 8 - AI Ethics (EQF 6) 9 - Business Intelligence (EQF 6) 10 - AI Awareness (EQF 6) 11 - Cyber and Data Security (EQF 6) 12 - Generative AI (EQF 6) 13 - Change Management (EQF 6) 14 - Soft Skills (EQF 6)

2. Description of the structure

The curriculum presents 12 learning Units(LU) which can be grouped in four main areas: (1) Basic Data Science, which includes “Data Analytics” and “Signal Processing and Time Series”; (2)

Machine Learning which includes “Machine Learning: Supervised”, “Machine Learning: Unsupervised” and “Neural Networks”; (3) Deep Learning which includes “Deep Learning”, “Generative AI” and “Foundation Models and LLMs” and (4) AI Infrastructure, comprising of “Human-Centered AI”, “AI Platforms & Architectures” and “Data Security”.

These four areas capture the breadth of knowledge and skills, from fundamental concepts to the latest advancements, that a Data Scientist should be familiar with. The curriculum begins with essential knowledge areas like "Data Analytics" and "Signal Processing and Time Series" under the Basic Data Science category. These units ensure that students have a strong grasp of the foundational techniques used to analyze and interpret data. Skills such as statistical analysis, data visualization, and understanding temporal data are crucial for any data scientist, providing the groundwork for more advanced studies. These LUs can be thought sequentially or in parallel but should be introduced before the other 3 areas.

As students progress, they delve deeper into Machine Learning, with units on both supervised and unsupervised learning, as well as neural networks. These units cover a broad spectrum of machine learning techniques, from regression and classification in supervised learning to clustering and dimensionality reduction in unsupervised learning. By understanding neural networks, students can begin to explore more complex models and algorithms that underpin many modern AI systems. This comprehensive approach ensures that students develop a robust understanding of both the theoretical and practical aspects of machine learning. These LUs can be taught sequentially or in parallel but should be taught before any LU of the third area.

The curriculum then advances into the specialized area of Deep Learning, where students learn about sophisticated deep network architectures and their applications. Furthermore, the inclusion of topics such as Generative AI and Foundation Models and LLMs (Large Language Models) introduces students to the forefront of Generative AI. These units cover innovative techniques for creating new data samples and utilizing large-scale models for natural language processing, ensuring that students are well-versed in the latest trends and technologies. These LU should be taught sequentially and only if the LUs of areas 1 and 2 have already been introduced.

Lastly, the AI Infrastructure area ensures that students are equipped with the skills necessary to build and deploy AI systems effectively and responsibly. Units such as "Human-Centered AI" emphasize the importance of designing AI systems that are ethical, transparent, and user-friendly. Topics in "AI Platforms & Architectures" focus on the infrastructure needed to deploy and scale AI solutions, including cloud platforms and distributed computing. The curriculum also addresses critical aspects of data privacy and security through the "Data Security" unit, ensuring that students understand how to protect data and comply with relevant regulations. These LUs can be taught in parallel or sequentially, and after the LU of areas 1, 2 and 3 have already been introduced.

There is also an extra Learning Unit, “Professional Abilities” whose goal is to equip students with the essential non-technical skills and behaviours that are critical for success in professional environments. These skills complement the technical knowledge and ensure that students can effectively apply their data science expertise in real-world settings. This LU can be taught in parallel with areas 2, 3 and 4, but after the LU of area 1.

3. Overview of Learning Units

Learning unit title	Hours/ECTS	EQF level	Assessment(s)
Data Analytics	6 ECTS	EQF 6	Exam + Practical Assignment
Signal Processing and Time Series	6 ECTS	EQF 6	Exam+Practical Assignment
Machine Learning: Supervised	6 ECTS	EQF 6	Exam+Practical Assignment
Machine Learning: Unsupervised	6 ECTS	EQF 6	Exam+Practical Assignment
Neural Networks	6 ECTS	EQF 6	Exam+Practical Assignment
Deep Learning	6 ECTS	EQF 6	Exam+Practical Assignment
Generative AI	6 ECTS	EQF 6	Exam+Practical Assignment
Foundational Models and LLMs	6 ECTS	EQF 6	Exam+Practical Assignment
Human-Centered AI	6 ECTS	EQF 6	Exam+Practical Assignment
AI Platforms & Architecture	6 ECTS	EQF 6	Exam+Practical Assignment
Data Security	6 ECTS	EQF 6	Exam+Practical Assignment
Professional Abilities	6 ECTS	EQF 6	Practical Assignment

4. Details of Learning Units

4.1. Data Analytics

Description

The Data Analytics learning unit is designed to provide students with a comprehensive understanding of the fundamental techniques and tools used to analyze and interpret data. It comprises 4 modules: 1. Python for Data Analytics, 2. Data Collection and Data Cleaning, 2. Data Visualization, 3. Exploratory Data Analysis, 4. Introduction to SQL.

Related Programme Learning Outcome(s)

- 1 - AI Technologies (EQF 6)
- 2 - Machine Learning (EQF 6)
- 3 - Big Data & Data Analytics (EQF 6)
- 4 - Business Intelligence (EQF 6)
- 5 - AI Awareness (EQF 6)

Unit learning outcomes

- Selects appropriate AI frameworks and libraries for specific project needs
- Designs data pre-processing and feature engineering strategies
- 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 AI
- Identifies key AI applications in the programming application market

Delivery method(s)

Face-to-face; Online synchronous learning

Materials

Lecturer Materials: slides and Notebooks

Suggested Readings:

- SQL QuickStart Guide: The Simplified Beginner's Guide to Managing, Analyzing, and Manipulating Data With SQL
- Python for Data Analysis 3e: Data Wrangling with pandas, NumPy, and Jupyter

4.2. Signal Processing and Time Series

Description

The "Signal Processing and Time Series" learning unit is designed to introduce students to the fundamental concepts and techniques used in analyzing and interpreting time-dependent data. This unit covers the principles of signal processing and the methodologies for handling and extracting meaningful insights from time series data. It contains the following modules: (1) Basic Signal Processing techniques, (2) AR, MA, ARMA, ARIMA, SARIMA models, (3) Spectral Analysis and Fourier transform and (4) Time Series Libraries: prophet & darts.

Related Programme Learning Outcome(s)

- 1 - AI Technologies (EQF 6)
- 2 - Machine Learning (EQF 6)
- 3 - Big Data & Data Analytics (EQF 6)
- 4 - Business Intelligence (EQF 6)
- 5 - AI Awareness (EQF 6)

Unit learning outcomes

- Selects appropriate AI frameworks and libraries for specific project needs
- Designs data pre-processing and feature engineering strategies
- 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
- Identifies key AI applications in the programming application market

Delivery method(s)

Face-to-face

Online synchronous learning

Materials

Lecturer Materials: slides and Notebooks

Suggested Readings:

- Handbook of Time Series Analysis, Signal Processing, and Dynamics

4.3. Machine Learning: Supervised

Description

The "Machine Learning: Supervised" learning unit is designed to provide students with a deep understanding of supervised learning techniques. Students will learn how to develop, evaluate, and optimize predictive models to solve various classification and regression problems. Modules: (1) k-NN, Linear and Logistic regression, (2) Decision Trees & SVM, (3) Model Evaluation and Validation, (4) Hyperparameter Tuning and Model Optimization and (5) AutoML.

Related Programme Learning Outcome(s)

- 1 - AI Technologies (EQF 6)
- 2 - Machine Learning (EQF 6)
- 3 - AI Awareness (EQF 6)

Unit learning outcomes

- Selects appropriate AI frameworks and libraries for specific project needs
- Develops AI models using chosen frameworks and libraries
- Applies fundamental machine learning concepts and algorithms
- Evaluates machine learning models using established metrics and validation techniques
- Understands the basic concepts and technologies underlying artificial intelligence
- Identifies key AI applications in the programming application market

Delivery method(s)

Face-to-face

Online synchronous learning

Materials

Lecturer Materials: slides and Notebooks

Suggested Readings:

- Introduction to Machine Learning with Python: A Guide for Data Scientists

4.4. Machine Learning: Unsupervised

Description

The "Machine Learning: Unsupervised" learning unit is designed to introduce students to the techniques and algorithms used in unsupervised learning. Students will learn how to apply various unsupervised learning methods to identify clusters, reduce dimensionality, and extract useful features from data. Modules: (1) Introduction to Unsupervised Learning, (2) K-means and Hierarchical Clustering, (3) DBSCAN, GMM, (4) Dimensionality Reduction: PCA, t-SNE and UMAP, (5) Anomaly Detection

Related Programme Learning Outcome(s)

- 1 - AI Technologies (EQF 6)
- 2 - Machine Learning (EQF 6)
- 3 - AI Awareness (EQF 6)

Unit learning outcomes

- Selects appropriate AI frameworks and libraries for specific project needs
- Develops AI models using chosen frameworks and libraries
- Applies fundamental machine learning concepts and algorithms
- Evaluates machine learning models using established metrics and validation techniques
- Understands the basic concepts and technologies underlying artificial intelligence
- Identifies key AI applications in the programming application market

Delivery method(s)

Face-to-face
Online synchronous learning

Materials

Lecturer Materials: slides and Notebooks

Suggested Readings:

- Introduction to Machine Learning with Python: A Guide for Data Scientists

4.5. Neural Networks

Description

The "Neural Networks" learning unit is focused on introducing students to the fundamentals of neural networks, a cornerstone of modern artificial intelligence and deep learning. This unit covers the architecture, functioning, and practical applications of different types of neural networks. Modules: (1) Introduction: Perceptrons and MLP, (2) Frameworks: TensorFlow, PyTorch, Keras, (3) Activation Functions, (4) Feedforward NN, (6) Training NN: Backpropagation and Gradient Descent and (5) Optimization Techniques.

Related Programme Learning Outcome(s)

- 1 - Deep Learning (EQF 6)
- 2 - AI Technologies (EQF 6)
- 3 - Machine Learning (EQF 6)
- 4 - AI Awareness (EQF 6)

Unit learning outcomes

- Analyzes the fundamental principles of neural networks
- Identifies suitable applications for deep and shallow neural architectures
- Selects appropriate AI frameworks and libraries for specific project needs
- Develops AI models using chosen frameworks and libraries
- Evaluates machine learning models using established metrics and validation techniques
- Understands the basic concepts and technologies underlying artificial intelligence
- Identifies key AI applications in the programming application market

Delivery method(s)

Face-to-face

Online synchronous learning

Materials

Lecturer Materials: slides and Notebooks

Suggested Readings:

- Neural networks for pattern recognition
- Neural Networks and Deep Learning: A Textbook

4.6. Deep Learning

Description

The "Deep Learning" learning unit delves into the advanced aspects of neural networks and their extensive applications across various fields, including computer vision, natural language processing, and audio recognition. Modules: (1) CNNs, (2) RNNs, (3) LSTM, (5) Autoencoders, (6) Transfer Learning and Fine-Tuning and (7) Applications: NLP and vision.

Related Programme Learning Outcome(s)

- 1 - Deep Learning (EQF 6)
- 2 - AI Technologies (EQF 6)
- 3 - Machine Learning (EQF 6)

Unit learning outcomes

- Identifies suitable applications for deep and shallow neural architectures
- Designs deep learning models by selecting appropriate architectures
- Implements deep learning models using relevant frameworks and libraries
- Selects appropriate AI frameworks and libraries for specific project needs
- Develops AI models using chosen frameworks and libraries
- Evaluates machine learning models using established metrics and validation techniques

Delivery method(s)

Face-to-face

Online synchronous learning

Materials

Lecturer Materials: slides and Notebooks

Suggested Readings:

- Deep Learning with Python
- Neural Networks and Deep Learning: A Textbook
- "Deep Learning" by Ian Goodfellow, Yoshua Bengio, and Aaron Courville

4.7. Generative AI

Description

The "Generative AI" learning unit focuses on creating new data instances that resemble training data. This area includes techniques that allow models to generate novel images, texts, music, and other forms of media, simulating human-like creativity. Modules: (1) Deep Learning for Generative AI: GANs and VAEs, (2) Natural Language Generation, (3) Image and Video Generation, (4) Audio Generation.

Related Programme Learning Outcome(s)

1 - AI Technologies (EQF 6)

2 - Generative AI (EQF 6)

Unit learning outcomes

- Selects appropriate AI frameworks and libraries for specific project needs
- Develops AI models using chosen frameworks and libraries
- Designs generative AI models to create novel content
- Integrates generative AI into diverse applications

Delivery method(s)

Face-to-face

Online synchronous learning

Materials

Lecturer Materials: slides and Notebooks

Suggested Readings:

- "Generative Deep Learning: Teaching Machines to Paint, Write, Compose, and Play" by David Foster
- "GANs in Action: Deep Learning with Generative Adversarial Networks" by Jakub Langr and Vladimir Bok

4.8. Foundational Models and LLMs

Description

The "Foundational Models and Large Language Models (LLMs)" learning unit explores the rapidly evolving area of foundational models, particularly focusing on large-scale language models that are transforming natural language processing, and beyond. Modules: (1) Introduction to Foundational Models, (2) Transformers, (3) Training LLMs, (4) Fine Tuning and (5) Integration of foundational models.

Related Programme Learning Outcome(s)

1 - AI Technologies (EQF 6)

2 - Generative AI (EQF 6)

Unit learning outcomes

- Selects appropriate AI frameworks and libraries for specific project needs
- Develops AI models using chosen frameworks and libraries
- Develops prompt engineering skills
- Implements Large Language Models (LLMs) in generative AI projects
- Integrates generative AI into diverse applications

Delivery method(s)

Face-to-face

Online synchronous learning

Materials

Lecturer Materials: slides and Notebooks

Suggested Readings:

- Natural Language Processing with Transformers: Building Language Applications With Hugging Face
- Generative AI with LangChain: Build large language model (LLM) apps with Python, ChatGPT and other LLMs

4.9. Human-Centered AI

Description

The "Human-Centered AI" learning unit focuses on the integration of humanistic principles into the design, development, and deployment of artificial intelligence systems. This unit explores how AI can be developed to be more interpretable, ethical, and supportive of human needs and values. Students will learn about the importance of designing AI systems that are not only technically proficient but also socially responsible and aligned with human goals. Modules: (1) Introduction to Human-Centered AI, (2) Ethical Considerations: Fairness, Explainability and Robustness, (3) AI Governance and Regulation, (4) AI Ethics in Generative AI and (5) Design for User Experience: Inclusivity and accessibility.

Related Programme Learning Outcome(s)

- 1 - Explainable AI (EQF 6)
- 2 - Human-Centered AI (EQF 6)
- 3 - AI Ethics (EQF 6)

Unit learning outcomes

- Develops AI models using explainable AI techniques to ensure transparency and interpretability
- Analyses existing AI models to assess and measure their explainability
- Incorporates human-centered design principles in the development of AI systems
- Manages AI risks by identifying potential ethical, legal, and operational issues associated with AI technologies
- Identifies ethical considerations and challenges in AI development and deployment
- Implements strategies to mitigate ethical risks in AI applications
- Knowledge of industry-specific laws (national and international)

Delivery method(s)

Face-to-face
Online synchronous learning

Materials

Lecturer Materials: slides and Notebooks

Suggested Readings:

- Human-Centered AI, Ben Shneiderman

4.10. AI Platforms & Architectures

Description

The "AI Platforms & Architectures" learning unit focuses on the technical underpinnings and infrastructure necessary for developing and deploying AI systems effectively. Students will learn about cloud computing services, edge computing, the importance of hardware accelerators like GPUs and TPUs, and software frameworks that are crucial for AI development. Modules: (1) Cloud Computing for AI, (2) Edge Computing and AI, (3) Hardware for AI, (4) Containers and (5) MLOps.

Related Programme Learning Outcome(s)

- 1 - AI Technologies (EQF 6)
- 2 - Big Data & Data Analytics (EQF 6)
- 3 - Business Intelligence (EQF 6)
- 4 - Cyber and Data Security (EQF 6)

Unit learning outcomes

- Selects appropriate AI frameworks and libraries for specific project needs
- Utilizes big data technologies to efficiently process and analyse large volumes of data
- Utilizes business intelligence tools and software to collect, process, and analyse large datasets
- Integrates data warehousing and data management practices
- Designs security architectures for information systems

Delivery method(s)

Face-to-face

Online synchronous learning

Materials

Lecturer Materials: slides and Notebooks

Suggested Readings:

- Cloud Computing: Concepts, Technology & Architecture
- <https://aws.amazon.com/es/ai/>
- Azure.microsoft.com
- Data Science on AWS: Leveraging AWS for Advanced Data Science Solutions
- Fundamentals of Data Engineering: Plan and Build Robust Data Systems

4.11. Data Security

Description
<p>The "Data Security" learning unit presents the concepts regarding protection of data integrity, confidentiality, and availability in the context of AI and data science projects. Modules: (1) Introduction to Data Security, (2) Secure Data Management, (3) Security in Cloud Computing, (4) Data Governance and Data protection laws and (5) AI Robustness.</p>
Related Programme Learning Outcome(s)
<p>1 - Cyber and Data Security (EQF 6)</p>
Unit learning outcomes
<ul style="list-style-type: none"> • Identifies a variety of cybersecurity threats and vulnerabilities • Implements key cybersecurity measures • Manages cybersecurity incidents by effectively deploying incident response
Delivery method(s)
<p>Face-to-face</p> <p>Online synchronous learning</p>
Materials
<p>Lecturer Materials: slides and Notebooks</p> <p>Suggested Readings:</p> <ul style="list-style-type: none"> - Practical Data Privacy: Enhancing Privacy and Security in Data - Privacy Engineering: A Runbook for Engineers

4.12. Professional Abilities

Description
<p>The "Professional Abilities" learning unit is designed to equip students with the essential non-technical skills necessary for a successful career in data science. Modules: (1) Professional skill in data Science, (2) Effective Communication and Presentation skills, (3) Collaboration and Teamwork, (4) Project Management and (5) professional writing and documentation.</p>
Related Programme Learning Outcome(s)
<p>1 - Change Management (EQF 6) 2 - Soft Skills (EQF 6)</p>
Unit learning outcomes
<ul style="list-style-type: none"> • Understands the principles and theories of change management • Communicates change effectively, using clear, persuasive messaging • Collaborates within teams, contributing positively to group efforts • Solves problems creatively and efficiently • Delivers impactful presentations • Cultivates an innovative mindset, embracing and fostering creativity • Communicates effectively across a variety of platforms and media • Thinks critically, analysing situations, evaluating diverse perspectives • Manages conflicts constructively
Delivery method(s)
<p>Face-to-face Online synchronous learning</p>
Materials
<p>Lecturer Materials: slides and Notebooks Suggested Readings:</p> <ul style="list-style-type: none"> - Business Skills for Data Scientists: Practical Guidance in Six Key Topics



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.