Unit 1: Introduction to Data Science and AI

Unit code:

Level 7:

Credit value:

Guided learning hours: 20

Unit aim

To provide participants with a foundational understanding of the fields of Data Science and Artificial Intelligence. By the end of the module, attendees should have a clear grasp of the evolution, significance, and applications of Data Science and AI in various industries. Additionally, they will be introduced to the fundamental differences and intersections between Data Science and AI and become familiar with the critical tools and technologies driving these domains. This foundational knowledge will serve as a platform for more advanced topics in subsequent modules.

Unit introduction

Data Science and AI have become transformative forces in our quickly digitising world, altering sectors, rethinking company strategies, and reimagining customer experiences. The applications span a wide range of industries, from manufacturing to healthcare to finance. But first, it's important to have a fundamental comprehension of what complicated algorithms, predictive models, and neural networks are and how they work.

Learning outcomes and assessment criteria

To pass this unit, the evidence that the learner presents for assessment needs to demonstrate that they can meet all the learning outcomes for the unit. The assessment criteria determine the standard required to achieve the level of proficiency.

On completion of this unit, a learner should

Learning outcomes	Assessment criteria
1 Understand the Historical Context of Data Science and Artificial Intelligence	1.1 Describe the evolution and pivotal moments in the history of Data Science and AI
2 Differentiate Between Data Science and Artificial Intelligence	2.1 Clearly distinguish between the core concepts, techniques, and objectives of Data Science and AI 2.2 Identify areas of intersection and collaboration between the two fields
3 Recognize Data Types and Structures	3.1 Classify various data types and explain their relevance in Data Science and AI projects 3.2 Understand the significance of structured and unstructured data in different industry applications
4 Appreciate Real-world Applications and	4.1 Enumerate key industries where Data Science and AI have made a significant impact 4.2 Cite real-world case studies showcasing the transformative power of these technologies
5 Identify Key Tools and Technologies	5.1 List and describe essential software, platforms, and technologies used in Data Science and AI 5.2 Understand the significance of choosing the right tool for specific tasks.
6 Acknowledge Industry Barriers	6.1 Understand the challenges industries face in adopting Data Science and AI. 6.2 Identify potential strategies to mitigate these challenges.

Unit content

---

1. **Overview and Evolution:** We will begin by taking a retrospective look at how data science and AI have emerged over the decades, understanding their roots, pivotal moments, and trajectory to modern-day significance. The "overview" offers a snapshot of the current state, defining key concepts and illuminating the field's primary applications. In contrast, the "evolution" traces the journey of the discipline from its inception to the present, marking pivotal moments, groundbreaking discoveries, and paradigm shifts. Together, they provide a holistic understanding, painting both a macroscopic picture of the field's purpose and potential as well as a chronological narrative of its development over time.

2. **Differentiating Data Science and AI:** It's common for many to use data science and 'AI' interchangeably. This module will delineate the unique aspects of each field while also highlighting where they intersect.

   Initially, we'll delve into data science, highlighting its primary goal of extracting meaningful insights from vast and varied datasets. Participants will understand its multidisciplinary nature, encompassing statistics, data analysis, and visualisation techniques. Real-world applications, like business analytics and predictive modelling, will underscore its practical relevance.

   Transitioning to artificial intelligence (AI), we'll explore its overarching aim: enabling machines to mimic human-like

intelligence. This section will introduce subsets of AI, notably machine learning, emphasising how it allows systems to learn from data and make decisions.

Crucially, the module will spotlight intersections between the two and how data science methodologies often serve as the bedrock for training AI models. Yet, it will also emphasise their distinct objectives, with data science primarily concerned with discovering patterns and AI with decision-making.

By the module's conclusion, participants should not only grasp the nuances distinguishing these fields but also appreciate how they collaboratively drive modern innovations.

3. **Data Structures & Their Importance:** Data is often called the 'oil' of the digital age. We'll explore different data types and structures, emphasising why a robust understanding of these is vital for anyone looking to delve into data science or AI.

   We'll kick off with an introduction to basic data structures like arrays, linked lists, stacks, and queues, illustrating their characteristics and operations. Progressing to more complex structures, participants will explore trees, graphs, and hash tables, understanding their relevance in various computational problems.

   A vital segment will address the distinction between linear and non-linear structures, elaborating on their respective use cases. Additionally, the module will touch upon dynamic data structures, which grow and shrink during runtime, emphasising their flexibility.

   Underpinning the theoretical knowledge will be real-world

applications showcasing how these structures drive efficiency in algorithms. For instance, how trees, specifically binary search trees, facilitate faster data retrieval or how hash tables optimise storage and lookup operations

The crux of the module is to impress upon participants that choosing the appropriate data structure is paramount. An apt choice can dramatically improve the speed, readability, and overall efficacy of software solutions, whereas a poor choice can lead to inefficient and cumbersome programmes.

By the end, attendees should appreciate data structures not just as theoretical constructs but as pivotal tools in effective problem-solving.

4. **Real-world Applications:** We will take a cursory look at how different industries are leveraging Data Science and AI. These case studies will give you a taste of the transformative power of these technologies.

   Participants will embark on a journey exploring a myriad of sectors. In the field of healthcare, researchers and practitioners are exploring how artificial intelligence (AI) might contribute to various aspects of the industry. Specifically, AI has shown promise in supporting diagnostic processes, forecasting patient outcomes, and optimising administrative operations. When exploring the field of finance, individuals will observe the significant impact of artificial intelligence in several areas such as fraud detection, algorithmic trading, and personalised financial services.

   The retail industry is expected to demonstrate advancements in inventory management, consumer analytics, and chatbot technology, all powered by artificial intelligence, hence increasing the whole shopping experience. Within the field of

transportation, the focal point of discourse often revolves around autonomous vehicles, the optimisation of routes, and the implementation of predictive maintenance strategies.

Each industry-specific exploration will not only highlight the technologies in use but also the transformative impact they've ushered. Case studies will be employed, offering a granular look into successful implementations, challenges encountered, and the solutions devised

5. **Exploration of Tools & Technologies:** No professional can function effectively without the right tools. This section introduces the software, platforms, and technologies pivotal in the realms of Data Science and AI.

   Initiating with **Data Manipulation and Analysis**, participants will learn about tools like **Pandas** and **NumPy** that offer robust capabilities for handling large datasets. The segment on Visualization will introduce platforms such as **Matplotlib** and **Tableau**, illuminating how they transform raw data into insightful visuals.

   Diving into Machine Learning, we'll cover frameworks like **TensorFlow** and **Scikit-learn**, showcasing their roles in creating and tuning predictive models. The Deep Learning section will touch upon tools such as **Keras** and **PyTorch**, elucidating their significance in constructing neural networks.

   The module will also shed light on Big Data Technologies like **Hadoop** and **Spark**, crucial for processing vast data streams efficiently. Additionally, an overview of Cloud Platforms such as **AWS**, **Azure**, and **Google Cloud** will emphasize their role in scalable AI solutions.

   Integral to this module are hands-on demos, ensuring

participants not only understand the theoretical aspects but also witness these tools in action.

By the module's end, attendees should be well-versed in the technological landscape of AI and Data Science, and equipped with knowledge to select and implement appropriate tools for specific tasks.

6. **Barriers to Adoption:** Lastly, understanding challenges is as crucial as understanding capabilities. We'll touch upon common barriers industries face while adopting these technologies.

   Technical Challenges will headline the discussion, focusing on issues like data quality, integration complexities, and the need for robust infrastructure. The intricacies of migrating from traditional systems to AI-driven ones, scalability concerns, and the ever-evolving tech landscape requiring continuous upskilling will be emphasized.

   Under Organizational Barriers, the module will explore resistance to change, misalignment between business and IT units, and the scarcity of skilled professionals in the AI domain. We'll also delve into the high initial costs associated with AI adoption and the struggle to demonstrate immediate ROI.

   Ethical and Regulatory Challenges Will examine the presence of biases within AI models, address privacy concerns, and explore the ethical ramifications associated with automation. The evolving nature of AI regulation, compliance issues, and the need for transparent AI will be highlighted.

   Lastly, **Societal Barriers** will examine public perception, fears of job displacement due to automation, and the broader

societal implications of widespread AI adoption.

By exploring these hurdles, the module aims to equip professionals with a comprehensive understanding of potential pitfalls, encouraging proactive strategies for smoother AI integration and ensuring ethical, transparent, and beneficial deployments.

Why This Module is Important:

To truly grasp the more intricate aspects of Data Science and AI, one must first have a comprehensive understanding of their foundational elements. This module aims to ensure that every participant, whether a novice or someone with some prior exposure, gains clear, cohesive, and consistent foundational knowledge.

Delivery: Tutors must keep in mind the diverse backgrounds of industry professionals and because this module serves as their gateway into a rapidly evolving field, tutors must ensure the content remains accessible, relatable, and engaging. I have explained these in more detail, and these shall serve as guiding principles during the preparation of any course materials and delivery of lectures:

Relatability: The tutor must anchor theoretical concepts with real-world applications. Case studies can be an invaluable tool in this regard, offering tangible illustrations of abstract ideas.

Interactivity: The tutor must encourage participation throughout this course to participate fully. The brainstorming sessions, interactive visualizations, and case discussions are designed to foster an environment of active learning. Prompting questions and discussions

can make abstract concepts more tangible.

Practicality: The hands-on exercises aim to consolidate theoretical knowledge. Tutors will ensure participants get hands-on exposure, even if rudimentary. The act of 'doing' often reinforces understanding.

Feedback: Tutors must be proactive in providing feedback, especially during assignments and hands-on sessions. Constructive feedback can bridge gaps in understanding and reinforce concepts.

Adaptability: Tutors must recognize the pace of the class. While the module has a defined structure, feel free to spend more time on concepts that participants find challenging

When considering learning outcome 1 i.e. "**Understand the Historical Context of Data Science and Artificial Intelligence**," it's essential for learners to the foundational theories and initial advancements that gave birth to AI and Data Science. Recognizing seminal figures like Alan Turing or foundational concepts such as the Turing Test can give learners an appreciation for the field's roots. Learners must be able to Identify major milestones in the evolution of AI and Data Science. This could include the advent of neural networks, the AI winter periods, or the resurgence of AI with deep learning. They must also recognize how advancements in computing power, storage, and data availability have driven progress in AI and Data Science and how various fields, including statistics, computer science, cognitive psychology, and even biology, converged to shape AI and Data Science's trajectory. Learners must be able to grasp the societal impacts i.e., how AI and Data Science have historically influenced industries, economies, and broader society. This encompasses both the positive transformations and the challenges or ethical dilemmas they introduced.

Tutors must provide the right perspective and industry-based examples so that learners develop an appreciation for how past challenges, failures, and successes have informed current best practices, ethical considerations, and the direction of future research.

For learners, grasping these elements not only provides a comprehensive understanding of where AI and Data Science have been but also offers context on their current state and future potential.

Learning Plan

Topic and suggested assignments/activities and/assessment
Introduction to unit and programme of learning
Tutor-led lecture on **Overview of Data Science** supplemented by brainstorming session on real-world applications of Data Science
Tutor-led lecture on the **Basics of Artificial Intelligence** followed by an assignment on “Research and summarize a pivotal moment in AI's evolution”
Tutor-led lecture on **How Data Science fuels AI and the synergy between them** followed by a case study discussion on a real-world example using both AI and Data Science
Tutor-led lecture on **Introduction to data sources, data quality, and preprocessing techniques** followed by an interactive session using simple tools to visualize basic algorithms in action
Tutor-led lecture on **Basic Algorithms and Models** followed by a hands-on exercise on data cleaning using sample dataset
Tutor-led lecture on the **Exploration of industries revolutionized by Data Science and AI** followed by an assignment to pick a real-time use case or industry and detail its AI transformation in the last few years
Towards the end of the module, there will be a **Group Project in which** participants will team up to identify a real-world problem and propose a solution utilizing the basics of Data Science and AI, culminating in a ***PowerPoint presentation***
A multiple choice exam of 30 minutes covering foundational concepts of the module.
Participants will write a brief ~200 words **reflection** on the importance of Data Science and AI in modern society, based on their learnings from the module

Learning Expectations & Assessment

For learners, differentiating between Data Science and AI ensures they can accurately navigate discussions, make informed decisions in their respective roles, and appreciate where interdisciplinary collaboration can be most impactful. It also helps in avoiding common misconceptions and ensures clarity when approaching projects, research, or further studies in either domain

Upon completing the "**Overview and Evolution**" section, learners are expected to grasp the chronological progression of Data Science and AI, identifying key milestones and figures. They should be able to understand pivotal technological and conceptual shifts that have shaped the fields. And that they recognize the impact of AI and Data Science on society and industries throughout history.

For the section: "**Differentiating Data Science and AI**", learners are expected to discern the distinct characteristics, methodologies, and objectives of both Data Science and AI. They should be equipped to identify the primary tools and applications unique to each field, while also recognizing their overlapping domains, notably in areas like machine learning. Furthermore, learners should be able to engage in informed discussions, confidently differentiating between the two disciplines and understanding their interconnected roles in the broader landscape of technology and data-driven decision-making.

While working on the section "**Data Structures & Their Importance**", learners will recognize the significance of selecting the appropriate data structure for various tasks and understanding the implications for efficiency, data storage, and algorithm performance. With this foundation, learners are positioned to better analyse, store, and manipulate data, optimizing the efficacy of subsequent data-driven processes and AI computations.

Upon completing the "**Real-world Applications**", learners are expected to understand the tangible impacts and implementations of Data Science and AI in diverse industries and societal contexts. They should be able to identify specific instances where these technologies have brought about transformative changes, addressing real-world challenges and enhancing various facets of daily life.

With the "**Exploration of Tools & Technologies**" section, learners will have a good view of key tools, platforms, and technologies that underpin the worlds of Data Science and AI. They should possess a foundational knowledge of how to select, deploy, and navigate these tools effectively in various contexts, understanding their strengths, limitations, and appropriate use cases.

Upon finishing the section: "**Barriers to Adoption**", learners are expected to recognize and understand the challenges and obstacles hindering the widespread adoption of Data Science and AI in various sectors. This encompasses technical constraints, ethical considerations, regulatory concerns, and societal apprehensions.

Programme of suggested assignments

The assessment criteria directly reference the learning outcomes and expectations from each section of the module. The assignments are crafted to align closely with these criteria, ensuring learners have a comprehensive understanding and ability to apply their knowledge in real-world contexts. This is for guidance only and learners might not be expected to produce all of these.

Assessment Criteria Covered	**Assignment Title**	**Scenario**	**Assessment Method**
Historical progression of Data Science and AI	Historical Timeline Project	As a data scientist in the making, trace the history of AI and Data Science using a timeline.	Visual Presentation
Characteristics and methodologies of Data Science and AI	Compare and Contrast	Draft a report distinguishing between AI and Data Science based on a chosen industry scenario.	Written Report
Fundamental data structures & their roles	Data Structures Workbook	Utilize an array, list, tree, and graph to solve specific data challenges in a company setup.	Practical Workbook/Code
Impacts and implementations in diverse industries	Real-world Case Study Analysis	Analyze a real-life industry scenario where AI and Data Science have been pivotal.	Written Report & Discussion
Key tools, platforms, and technologies	Tool Exploration Lab	Simulate a business problem and apply a solution using a chosen Data Science/AI tool.	Hands-on Lab & Reflection
Challenges in Adopting Data Science & AI	Barrier Analysis Presentation	Discuss the potential barriers to AI integration in a traditional industry.	Oral Presentation & Q/A

Resources

Text Books

“Data Science for Business" by Foster Provost and Tom Fawcett
ISBN: 978-1449361322
This book introduces the core principles of Data Science and its implications for business scenarios.

"Artificial Intelligence: A Modern Approach" by Stuart Russell and Peter Norvig
ISBN: 978-0136042594
A comprehensive textbook on AI, discussing its history, methodologies, and prospects.

"Python for Data Analysis" by Wes McKinney
ISBN: 978-1491957660
A guide to using Python for data wrangling and analysis, crucial for many Data Science tasks.

"The Hundred-Page Machine Learning Book" by Andriy Burkov
ISBN: 978-1790368384
A concise and beginner-friendly overview of machine learning principles and techniques.

"Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy" by Cathy O'Neil
ISBN: 978-0553418811
A discussion on the ethical and societal challenges posed by unchecked algorithmic decisions.

Articles (with Citations)

“Artificial intelligence: A powerful paradigm for scientific research” on National Library of Medicine, Yongjun Xu. (2021)

“The Role of ChatGPT in Data Science”, Hossein Hassani, (2023) How AI-Assisted Conversational Interfaces Are Revolutionizing the Field

"Real-world applications of artificial intelligence (AI) in business" on IBM's blog, Watson, J. (2018). Real-world applications of AI in business. IBM Blogs.

"The Barriers To Using AI Effectively" on Harvard Business Review, Bughin, J., & Hazan, E. (2017). The Barriers to Using AI Effectively. Harvard Business Review.

Unit 2: Data Management and Preprocessing

Unit code:

Level 7:

Credit value:

Guided learning hours: 20

Unit 2: Data Management and Preprocessing

Unit code:
Level 7:
Credit value:
Guided learning hours: 20

Unit aim

To equip participants with the essential skills and knowledge required for effective data management and preprocessing. By the end of this module, attendees should be proficient in various data acquisition techniques, understand the importance of data cleaning, and be able to transform and preprocess data for further analysis. This unit will serve as a foundation for subsequent modules that delve deeper into data analysis and machine learning.

Unit introduction

Data management and preprocessing are the foundational steps in any data science or AI project. Before any meaningful analysis can be conducted, data must be gathered, cleaned, and transformed into a suitable format. This unit delves into the various techniques and tools used in these crucial initial stages, ensuring that participants are well-equipped to handle real-world data challenges.

Learning outcomes and assessment criteria

To successfully complete this unit, learners must demonstrate that they can meet all the learning outcomes for the unit. The assessment criteria determine the standard required to achieve each outcome.

On completion of this unit, a learner should

Learning Outcomes	Assessment Criteria
1. Understand Various Data Acquisition Techniques	1.1 Describe different sources of data, including databases, APIs, and web scraping. 1.2 Understand the challenges associated with gathering data from various sources. 1.3 Discuss the importance of data quality and its impact on analysis. 1.4 Evaluate the ethical considerations when acquiring data. 1.5 Understand Master data management and impact on organizations
2. Master Data Cleaning and Preprocessing Techniques	2.1 Identify common issues in datasets, such as missing values, outliers, and inconsistencies. 2.2 Apply various techniques to handle missing data. 2.3 Understand the importance of data consistency and standardization. 2.4 Use tools and software to automate the data cleaning process.
3. Grasp the Art of Feature Engineering	3.1 Define what features are and their significance in data analysis. 3.2 Extract meaningful features from raw data. 3.3 Transform features to enhance their predictive power. 3.4 Evaluate the importance of feature selection in model performance.
4. Understand Data Transformation Techniques	4.1 Differentiate between scaling, normalization, and standardization. 4.2 Apply various transformation techniques to prepare data for machine learning. 4.3 Evaluate the impact of data transformation on model performance.

5. Master Data Visualization and Exploratory Data Analysis (EDA)	5.1 Describe the importance of EDA in understanding datasets. 5.2 Use various visualization tools to explore data distributions, correlations, and patterns. 5.3 Interpret visualizations to derive meaningful insights from the data. 5.4 Apply statistical techniques to further understand datasets.
6. Data Governance	6.1 Define and explain the core principles and objectives of data governance 6.2 Design and implement effective data governance strategies tailored to specific organizational needs 6.3 Navigate the ethical considerations and regulatory requirements associated with data governance 6.4 Recognize the role of data governance in the broader landscape of data management

Unit content

1. Overview of Data Management

We'll dive into the world of data management, exploring its key aspects and understanding its importance. At its core, data management is about organizing, storing, and taking care of data. In today's digital age, where we deal with vast amounts of information, having a system to manage all this data is crucial. This "overview" will shed light on the main ideas, tools, and practices that make data management work. From databases that store information to best practices that ensure data is safe and easy to access, we'll get a clear picture of how it all comes together to help businesses, researchers, and everyday users make the most of their data.

2. Data Acquisition Techniques

Data acquisition is like gathering puzzle pieces for a bigger picture. It's all about collecting the right information from various sources. While businesses traditionally relied on manual data entry or spreadsheets, technology has introduced more efficient methods. Web scraping, for example, extracts specific details from websites, providing insights into online brand mentions. APIs, or Application Programming Interfaces, allow seamless data retrieval from digital platforms, offering real-time updates like weather or stock market trends. Regardless of the method, it's vital to prioritize data quality and ethical collection. This ensures that the resulting insights are both accurate and trustworthy, paving the way for informed business decisions.

Participants will gain an understanding of the diverse methods used to collect data in today's digital age. They'll learn the significance of web scraping and the power of APIs in real-time data retrieval. Moreover, they'll grasp the importance of data quality, ensuring that the information they work with is both accurate and ethically sourced. We will also

3. Data Cleaning and Preprocessing

In the vast realm of data, it's often the case that not every piece of information is immediately ready for analysis. This is where the essential practices of **data cleaning** and **preprocessing** come into play. On the other hand, preprocessing sets the stage for insightful analysis, transforming data into a format that algorithms and models can easily interpret. Participants will gain a deep understanding of the nuances of data refinement. They'll learn the art of spotting inconsistencies, the science behind data transformations, and the importance of ensuring that data is in its prime state for analysis.

For hands-on experience, we'll delve into key tools and libraries that have become industry standards in this domain. **Pandas**, a Python library, stands out as a versatile tool for a myriad of data manipulations, from handling gaps in data to intricate transformations. **OpenRefine** offers a dedicated environment for data cleaning, allowing users to dive deep into large datasets, pinpoint inconsistencies, and apply batch corrections. **Scikit-learn**, beyond its renowned machine learning capabilities, is a treasure trove of utilities for data preprocessing, be it scaling, normalization, or encoding. And, of course, **Excel** remains a staple, showcasing that even familiar spreadsheet tools can play a pivotal role in data refinement.

4. Data Transformation

Data transformation is like language translation. Data is transformed to be suited for analysis or a different data format, just as we translate French to English for a different audience. This step is crucial to data management and preprocessing, ensuring data is in the correct format, scale, and structure for analysis, visualisation, and modelling.

Data transformation involves formatting raw data for analytical methods. This may require scaling numerical data, encoding category variables, or rearranging the data. A dataset may be converted from a wide format, where each variable has a column, to a long format, where variables are stacked vertically. Data preparation for some analysis or visualisations requires such changes.

Participants will learn about data transformation and associated approaches. Normalisation and standardisation, which give data a mean of zero and a standard deviation of one, will be covered. Logarithmic and power transformations, which stabilise variances and normalise data, will also be covered.

Tools and libraries that enable these transformations will be explored hands-on. **Scikit-learn** in Python provides powerful data scaling and encoding tools. In data management, SQL's query language may modify data. **Tableau** and **Power BI** allow on-the-fly data modifications during visualisation.

5. Exploration of Tools & Technologies

Tools and technology guide professionals through the complex processes of data handling, transformation, and analysis in data management and preprocessing. This section discusses the several tools' capabilities, strengths, and best uses.

Specialised data tools have proliferated to solve specific problems. They range from powerful databases like SQL and NoSQL that store massive amounts of data to data wrangling tools like **Pandas** and **OpenRefine** that clean and transform data to visualisation platforms like **Tableau** and **PowerBI** that visualise data through charts and graphs.

Participants will explore these tools and learn their practical uses. They'll learn SQL and relational databases. Python modules like **Pandas**, **Matplotlib**, and **Seaborn** will be used for data manipulation and visualisation. **Hadoop** and **Spark**, which process massive datasets, will be discussed.

Beyond the tools, this section emphasises job-specific tool selection. Understanding tool strengths and weaknesses is essential because not all tools are fit for all tasks. Participants will assess their data needs, project size, and difficulties to determine the best tools and technologies.

By the end of this section, learners will have a good knowledge to use these technologies for data management and preparation.

6. Data Governance

Data governance protects an organization's data quality, integrity, and security. Data management concepts, practises, and methods guarantee data is valued, protected, and

used properly. In this section, we'll examine data governance's importance, components, and issues.

Data governance involves defining data management roles, policies, and standards. It involves keeping data accurate, consistent, and accessible while preventing misuse. Define who can access, use, and maintain data quality. Enabling data-driven decision-making and meeting regulations and ethics is a delicate balance.

Data governance principles like data stewardship, quality, and security will be covered. They'll learn how data stewards and custodians manage data quality and use. Data catalogues that centralise data assets and data lineage tools that trace data's path through systems will be covered.

Data ethics and regulation will be a focus. Understanding compliance is vital with **GDPR** and **CCPA** creating strict data protection rules. Participants will learn about these regulations' problems and how organisations comply.

Creating a data-driven culture will also be stressed. Data governance must be integrated into the company's culture and valued by all stakeholders.

This part will empower learners to champion data governance in their organisations and ensure that data is not only handled but also valued as a vital asset.

Why This Module is Important

The digital age has ushered in an era where data is often referred to as the 'new oil.' Every day, vast amounts of data are generated, collected, and stored. However, raw data, in its unprocessed form, can be messy, inconsistent, and riddled with errors. Before any meaningful insights can be derived from this data, it must undergo a series of preprocessing steps to ensure its quality, consistency, and relevance.

Foundation for Decision Making: The integrity of any data-driven decision or analysis hinges on the quality of the data it's based upon. Data management and preprocessing ensure that the data fed into analytical models is clean, relevant, and devoid of errors. Without these foundational steps, subsequent analyses, no matter how sophisticated, risk being flawed or misleading.

Efficiency and Cost Savings: Poor data management can lead to inefficiencies, duplicated efforts, and costly mistakes. By ensuring data is well-managed and pre-processed, organizations can streamline their operations, reduce redundancies, and avoid costly errors that arise from basing decisions on poor-quality data.

Enhancing Data's Predictive Power: Properly pre-processed data can significantly enhance the predictive power of machine learning and AI models. Features engineered from clean data can capture intricate patterns, leading to more accurate predictions and insights.

Regulatory and Ethical Compliance: With the rise of data privacy regulations such as GDPR and CCPA, proper data management is not just a best practice—it's a legal requirement. Ensuring data is correctly acquired, stored, and processed is crucial for regulatory compliance. Moreover, ethical considerations, such as fairness and bias in AI, are intrinsically linked to how data is managed and pre-processed.

Building Trust: In an era where data breaches and misuse are common headlines, proper data management can help organizations build trust with their stakeholders. Ensuring data integrity and transparency in its processing can foster trust among customers, partners, and the public at large.

Future-Proofing: As the volume, variety, and velocity of data continue to grow, the challenges associated with managing and preprocessing this data will only intensify. Mastering the skills covered in this module ensures that participants are well-equipped to handle the data challenges of today and the future.

In essence, this module is not just about technical skills; it's about understanding the pivotal role that data management plays in the broader data science and AI landscape. Proper data management and preprocessing are the unsung heroes behind every successful data-driven initiative, ensuring that the insights derived are accurate, meaningful, and actionable.

"When considering the learning outcome for Unit 2, i.e., "**Understand the Fundamentals of Data Management and Preprocessing**," it's paramount for learners to comprehend the bedrock principles and practices that underpin effective data handling. Recognizing the significance of clean, organized data or foundational techniques such as data normalization

can give learners a deep appreciation for the meticulous processes behind robust data analysis.

Learners must be equipped to identify key stages in the data management pipeline, from acquisition and cleaning to transformation and storage. This includes understanding the challenges of dealing with missing or inconsistent data, the importance of data integrity, and the tools available for preprocessing tasks. They should also be aware of how advancements in technology, from cloud storage solutions to automated data cleaning tools, have revolutionized the way we manage and preprocess data. Furthermore, the interdisciplinary nature of data management, drawing from fields like computer science, statistics, and business intelligence, should be highlighted, illustrating how these diverse areas converge to shape best practices in data management.

Tutors have the responsibility to offer real-world, industry-relevant examples, ensuring that learners not only grasp the theoretical aspects but also understand their practical implications. By showcasing how poor data management can lead to flawed insights or how effective preprocessing can enhance the accuracy of predictive models, tutors can instil in learners the critical importance of this foundational step in the data lifecycle.

For learners, mastering these concepts is not just about understanding the mechanics of data management and preprocessing. It's about recognizing their pivotal role in ensuring that data-driven decisions, whether in business, research, or technology, are based on reliable, high-quality data. This understanding not only equips them with the skills to manage and preprocess data effectively but also provides a lens through which they can view the broader implications of these processes on the outcomes of data-driven projects."

Learning Plan

Topic and suggested assignments/activities and/assessment
Introduction to unit and overview of learning objectives.
Tutor-led lecture on **Overview of Data Management** supplemented by a brainstorming session on the importance of organized and clean data in real-world applications.
Tutor-led lecture on **Data Acquisition Techniques** followed by an assignment on "Research and summarize a significant advancement in data acquisition methods."
Tutor-led lecture on **Data Cleaning and Preprocessing** paired with a hands-on workshop where participants practice cleaning a sample dataset using popular tools i.e., **excel** or **python**
Tutor-led lecture on **Feature Engineering** followed by an interactive session where participants design new features for a given dataset, and discuss benefits of adding these features to potential business cases
Tutor-led lecture on **Data Transformation** supplemented by a case study discussion on the importance of data normalization and scaling in real-world applications.
Tutor-led lecture on **Tools & Technologies for Data Management** followed by an assignment where participants explore and summarize the functionalities of a chosen data management tool.
Towards the end of the module, participants will engage in a **Group Project**, teams will be tasked with managing and preprocessing a raw dataset, transforming it into a clean, analysis-ready format, and presenting their approach and findings. Bonus reward for teams who will use this data to visualize and answer a set of business questions
A **multiple-choice exam** of 30 minutes, focusing on the key concepts and techniques of data management and preprocessing.
Participants will conclude with a **~200 words reflection** on the significance of effective data management and preprocessing in the data science pipeline, drawing from their learnings throughout the module.

Learning Expectations & Assessment

This "Learning Expectations & Assessment" section for Unit 2 provides learners with a clear roadmap of the knowledge and skills they are expected to acquire throughout the module. It ensures a comprehensive understanding of the intricacies of data management and preprocessing, setting the stage for subsequent units in the course.

Upon diving into **Overview of Data Management**, learners are expected to grasp the foundational principles of data management. They should understand the importance of organized, clean data and recognize the challenges and solutions associated with handling vast amounts of information in various formats.

After learning about the **Data Acquisition Techniques,** learners should be familiar with various methods of data collection, from traditional databases to real-time data streams. They should appreciate the nuances of different data sources and understand the challenges and benefits associated with each.

By completing the section; **Data Cleaning and Preprocessing,** learners are expected to recognize the significance of refining raw data. They should understand common data imperfections, from missing values to outliers, and be equipped with techniques to address these issues. The importance of transforming data into a format suitable for analysis, ensuring its quality and integrity, should be clear to them.

Data Transformation techniques will teach learners to understand the processes involved in scaling, normalizing, and encoding data. They should appreciate the importance of these transformations in ensuring data is in the optimal format for analysis and modeling.

After learning the key **Tools & Technologies for Data Management,** learners will be able to use a range of tools and platforms pivotal in data management and preprocessing. They should be able to navigate these tools effectively, understanding when and how to deploy them based on the task at hand.

Upon concluding this segment; **Challenges in Data Management,** learners should be aware of the potential pitfalls and challenges in data management, from technical constraints to ethical dilemmas. They should be equipped to navigate these challenges, ensuring data integrity and ethical considerations are always at the forefront.

Programme of suggested assignments

The assessment criteria are directly tied to the learning outcomes and expectations from each section of the module. These assignments are designed to align closely with these criteria, ensuring that learners not only understand the concepts but can also apply them in real-world contexts. While this is a suggested list, learners might not be expected to complete all of these assignments.

Assessment Criteria Covered	Assignment Title:	Scenario	Assessment Method
Overview of Data Management	Data Management Blueprint	As an aspiring data manager, create a blueprint outlining the key components of effective data management.	Visual Presentation
Data Acquisition Techniques	Data Source Exploration	Research and document various data acquisition techniques, highlighting their pros and cons in different scenarios.	Written Report
Data Cleaning and Preprocessing	Assignment Title: Data Cleaning Workshop	Scenario: Given a raw dataset, apply various cleaning and preprocessing techniques to prepare it for analysis.	Assessment Method: Practical Workbook/code
Feature Engineering	Feature Crafting Lab	Using a sample dataset, design and implement new features to enhance the data's potential for modeling.	Hands-on Lab & Reflection
Data Transformation	Transformation Techniques Analysis	Discuss and demonstrate various data transformation techniques, emphasizing their importance in different data analysis scenarios.	Written Report & Discussion
Tools & Technologies for Data Management	Tool Dive-In Session	Explore a chosen data management tool, demonstrating its functionalities and applications in a simulated business problem.	Hands-on Lab & Reflection

Resources

Text Books

"Data Wrangling with Python: Tips and Tools to Make Your Life Easier" by Jacqueline Kazil and Katharine Jarmul
ISBN: 978-1491948811
Description: This book offers a comprehensive guide to the process of converting raw data into a format suitable for analysis, using Python.

"Data Management for Researchers: Organize, Maintain and Share Your Data for Research Success" by Kristin Briney
ISBN: 978-1784270117
Description: A practical guide for researchers on how to manage, organize, and clean their data, ensuring its integrity and usability.

"Data Preprocessing in Data Mining" by Salvador García, Julián Luengo, and Francisco Herrera
ISBN: 978-3319102460
Description: This book delves into the techniques and methods used in data preprocessing, a crucial step in the data mining process.

"Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists" by Alice Zheng and Amanda Casari
ISBN: 978-1491953242
Description: A guide to the art of crafting data features, and enhancing the performance of machine learning models.

"Data-Intensive Text Processing with MapReduce" by Jimmy Lin and Chris Dyer
ISBN: 978-1608453429
Description: This book introduces the concepts of data-intensive processing using the MapReduce framework, essential for handling vast datasets.

"Data Cleaning: A Practical Perspective" by Ihab Ilyas and Xu Chu
ISBN: 978-0367331583
Description: A discussion on the challenges of data cleaning and the methodologies to address them, ensuring data quality and reliability.

Articles (with Citations)

“Data Preprocessing and its Significance in Data Mining”
This article delves into the process of transforming raw data into an understandable format. It emphasizes the importance of data preprocessing as a crucial step in data mining
Published on Jan 24, 2023
https://www.tableau.com/learn/articles/what-is-data-cleaning

Data Preprocessing and its Significance in Data Mining
Description: The article delves into transforming raw data into an understandable format, emphasizing the importance of data preprocessing in data mining.
Published on: Jan 24, 2023
https://www.sciencedirect.com/topics/engineering/data-preprocessing

“The Art and Science of Data Wrangling”
Source: Towards Data Science, Author: Smith, L., (2022)
Description: An in-depth exploration of the challenges and techniques associated with transforming raw data into a usable format.
https://www.cc.gatech.edu/classes/AY2020/cs7643_spring/slides/L11_DataWranglingGATechLectureFeb2020.pdf

“Data Acquisition Systems - Current and future trends”
E.T. Subramaniam, B.P. Ajith Kumar, and R.K. Bhowmik, (2010)
Description: A comprehensive look at the various methods of data collection and the nuances of sourcing data in today's digital landscape.
http://www.sympnp.org/proceedings/55/I17.pdf

"Data Transformation Techniques"
Dimitris Vogiatzis, Coupler, (2022)
In this guide, we’ll go through all the different data transformation techniques you can use to ensure your dataset is clean and ready to be analyzed.
https://blog.coupler.io/data-transformation-techniques/

Module 4 for AI&ML at the London College of Contract Management

Unit 4:

Unit code: (Unit 4 commensurate Unit Code).

Level 7:

Credit value: (2 usually - 10 hours is 1 credit)

Guided leaning hours: 20

Unit aim

Through this module the learner gets a sound grasp of the theoretical aspects of Neural Networks and Convolutional Neural Networks with practical applications in building Machine Learning applications in Python culminating in a capstone project building either a computer vision model or a large language model with PyTorch or TensorFlow.

Unit introduction

The main aim of the syllabus is to give students an ability to use neural networks with proficiency in the Python Programming language and understand it's applications in different fields. The Syllabus goes through in detail the history of neural networks and convolutional neural networks with a firm grasp of the mathematics and statistics including Probability Theory, Functions and Backpropagation. The module then goes through the history of Neural Networks focusing on the work of four seminal academics in the field – Geoffrey Hinton, Yann LeCunn, Andrej Karpathy and Fei Fei Li. How a Neuron is abstracted in mathematics is covered, along with layers of neurons taking inputs and producing an output. Why Neural networks are able to model complex functions which traditional algorithms struggle to do so is emphasized upon as well as practical use cases in diverse fields such as weather prediction, neuroscience, cybersecurity, surveillance and self driving in computer vision and use of LLM models by OpenAI, Mistral and Google.

Learning outcomes and assessment criteria

In order to pass this unit, the evidence that the learner presents for assessment needs to demonstrate that they can meet all the learning outcomes for the unit. The assessment criteria determine the standard required to achieve the unit. The main topics are:-

1 Neural Networks
2 Convolutional Neural Networks
3 Biological Basis for Convolutional Neural Networks
4 Reinforcement Learning
5 Machine Learning with Python
6 Building Deep Learning Models with TensorFlow
7 AI Capstone Project with Deep Learning

On completion of this unit a learner should:-

Learning outcomes	Assessment criteria
1 Understand the parts of a Neural Network and apply it in Python Code.	1.1 Understand different parts of a Neural Network – Neural networks are structured in layers: the input layer, hidden layers, and the output layer as well as the neurons within these layers are connected, and the strength of these connections is represented by weights. 1.2 Demonstrate the use of Rectified Linear Unit (ReLU) and the soft-max function. 1.3 Apply the learning process in neural networks showing forward propagation, where input data is passed through the network to compute predictions, and backpropagation, where the network adjusts its weights and biases to minimize the loss function. 1.4 Identify how this iterative process continues over multiple epochs to refine the model's performance. 1.5 Analyze the results of the model with established metrics such as AU-ROC for classification tasks and RMSE for regression tasks.
2 Demonstrate an understand of the key parts of a CNN, it's use in Computer Vision and demonstrate ability to apply this in real life use cases.	2.1 Demonstrate an understanding of the key components of a CNN, including the convolutional layer, pooling layer, and fully connected layer. This includes recognizing how these layers work together to process and transform input data (e.g., images) through the network to produce final predictions or classifications. The role of activation functions within these layers should also be understood. 2.2 Effectively demonstrate the implementation and impact of activation functions within a CNN, specifically focusing on the Rectified Linear Unit (ReLU) for

	introducing non-linearity in convolutional and fully connected layers, and the soft-max function for classification tasks in the output layer. 2.3 Apply the learning process in CNNs by coding the forward propagation to pass input data through the network and compute predictions. 2.4 Implement backpropagation to adjust the network's weights and biases to minimize the loss function. This process should be demonstrated over multiple epochs to refine the model's performance, showcasing an understanding of how CNNs learn from data. 2.5 Analyze the results of the CNN model using established performance metrics and interpret these metrics to evaluate and discuss the model's performance.
3 Understand the Historical Development of Neural Networks in Cognitive Neuroscience, Recognize the Biological Inspirations for Neural Network Models	3.1 Be able to trace the evolution of neural network concepts from their early theoretical foundations in the 1940s. 3.2 Understand the contributions of key figures such as Warren McCulloch, Walter Pitts, Donald Hebb, and Frank Rosenblatt. 3.3 Identify how biological observations, particularly the structure and function of the human brain and its neurons, have inspired the design and development of artificial neural networks. 3.4 Understand the principles of Hebbian learning and the concept of neuronal assemblies. 3.5 Demonstrate a comprehensive understanding of the neurobiological underpinnings of neural networks, including the structure and function of

	biological neurons, synaptic plasticity, and the mechanisms of signal transmission and processing in the brain. 3.6 Analysis how biological constraints, such as the balance between excitatory and inhibitory neurons and the organization of the cortex into areas and layers, influence the design and functionality of neural network models. 3.7 Understand the importance of incorporating both excitatory and inhibitory dynamics in artificial neural networks for more realistic modeling. 3.8 Evaluate how neuroscience has contributed to innovations in neural network design, including the development of spiking neural networks (SNNs) and the implementation of learning rules that mimic biological learning processes, including how neuroscience findings can guide the creation of more efficient and biologically plausible neural network architectures such as Liquid Neural Networks.
4 Fundamentals and Applications of Reinforcement learning with an emphasis on RLHF.	4.1 Be able to comprehend the core concepts of reinforcement learning (RL), including the definition of agents, environments, states, actions, rewards, policies, value functions, and the exploration-exploitation trade-off. 4.2 Identify various RL algorithms, such as Q-learning, SARSA, and policy gradient methods, and understand their applications in different domains like gaming, robotics, and autonomous systems. 4.3 Differentiate between model-based and model-free

	approaches and the use of deep learning in deep reinforcement learning. 4.4 Implement RL algorithms and evaluating their performance, including setting up the RL environment, define reward structures, and adjust parameters to optimize the learning process. 4.5 Select the appropriate metrics to assess the effectiveness of RL models. 4.6 Understand the principles of RLHF, including how it integrates human feedback into the reinforcement learning process to train AI agents that can perform tasks aligned with human goals and preferences. 4.7 Explore how RLHF is applied in natural language processing, particularly in the training of LLMs, to produce outputs that are truthful, harmless, and helpful, reflecting complex human values and preferences. 4.8 Implement RLHF, understanding the stages of data collection, reward model training, and policy optimization, and applying these concepts to refine AI agents in simulated environments. 4.9 Evaluate the impact of RLHF on the performance of AI models, understanding how human feedback can enhance model accuracy and alignment with human values compared to models trained without such feedback.
5 Master Python for Machine Learning – recap of all key libraries frequently in Machine Learning such as NLTK, Spacy, TextBlob, PyTesseract, OpenCV, Tensorflow, PyTorch and using Huggingface Transformer Models.	5.1 Demonstrate the use of Python for machine learning, including familiarity with essential libraries such as NumPy, pandas, scikit-learn, TensorFlow, and Keras. They will be able to preprocess data, select and implement machine learning algorithms, and perform model evaluation using

	Python. 5.2 Apply supervised learning techniques, such as regression and classification, and unsupervised learning techniques, such as clustering and dimensionality reduction, to solve real-world problems and understand the appropriate contexts and datasets for each technique. 5.3 Demonstrate skills in building and tuning machine learning models, including understanding hyperparameters, using cross-validation for model selection, and applying grid search and other optimization techniques to improve model performance. 5.4 Evaluate machine learning models using appropriate metrics, understanding the differences between metrics for classification (e.g., accuracy, precision, recall, F1-score) and regression (e.g., MAE, MSE, RMSE). 5.5 Interpret these metrics to make informed decisions about model improvements. 5.6 Evaluate machine learning models using appropriate metrics, understanding the differences between metrics for classification (e.g., accuracy, precision, recall, F1-score) and regression (e.g., MAE, MSE, RMSE).
6 Practice Deep Learning Models with Python – Tensorflow, PyTorch and HuggingFace, OpenCV.	6.1 Build Two DeepLearning Models – One using HuggingFace Transformer Models for LLMs and another one using Simple Detecting Apples from Pictures from a training and a test set using Tensorflow. 6.2 Identify a problem that can be addressed with deep learning techniques, formulate clear objectives, and define the scope of the project, including conducting a literature review

	to understand the current state of research and potential approaches to the problem. 6.3 Design a deep learning solution to the identified problem, selecting appropriate neural network architectures (e.g., CNNs, RNNs, GANs) and technologies. 6.4 Implement the designed deep learning model using a programming language and libraries/frameworks such as TensorFlow or PyTorch, including writing clean, efficient, and well-documented code, preprocessing data, configuring the neural network, and training the model. 6.5 Evaluate the performance of the deep learning model using appropriate metrics and techniques, such as accuracy, precision, recall, F1-score for classification tasks, or MSE and RMSE for regression tasks. 6.6 Apply techniques for model optimization and tuning, such as hyperparameter tuning, regularization, and cross-validation, to improve model performance. 6.7 Interpret the results of the deep learning model, drawing conclusions about its performance, the effectiveness of the chosen approach, and the model's potential impact on the problem domain, including identifying limitations and proposing future work to extend and enhance the project. 6.8 Effectively communicate the project's findings, including the problem definition, methodology, results, and conclusions, through a written report and/or presentation to stakeholders.

7 Capstone in either LLMs or Computer Vision – Infrared Detection of Objects.	7.1 Demonstrate Learnings from Previous Modules This outcome requires students to integrate and apply knowledge from the initial modules of the course. It involves: Understanding of AI Principles: Applying foundational AI concepts, including machine learning algorithms, data preprocessing, and model evaluation techniques. Programming Skills: Utilizing programming languages such as Python for AI development, with a focus on libraries relevant to LLMs or computer vision (e.g., TensorFlow, PyTorch). Project Management: Demonstrating the ability to plan, execute, and manage an AI project, including defining objectives, selecting appropriate technologies, and evaluating outcomes. 7.2 Mistral 7B: Simple ChatGPT Variant For creating a simplified version of ChatGPT using Mistral 7B, students should: Model Selection: Understand how to choose and configure an appropriate language model for the task. Customization for Specific Roles: Tailor the model to function as either a sales representative or a Cybersecurity expert, including training on domain-specific datasets. Integration and Deployment: Demonstrate the ability to integrate the model into a practical application and deploy it in a user-friendly manner. 7.3 PyTorch: Infrared Night Detector Developing an Infrared Night

	Detector with PyTorch involves: Computer Vision Basics: Apply computer vision techniques to process and analyze infrared images. Model Training: Use PyTorch to train a model capable of detecting objects in low-light conditions, leveraging datasets like FLIR. Application Development: Create a functional application that utilizes the trained model to perform real-time object detection in infrared images. 7.4 Ethical Considerations and Bias Mitigation This outcome focuses on the ethical aspects of AI development: Ethical AI Use: Understand the ethical implications of deploying AI solutions, particularly in sensitive areas like cybersecurity and sales. Bias Identification and Mitigation: Recognize and address potential biases in AI models, ensuring fair and unbiased outcomes. 7.5 Communication and Documentation Effective communication and documentation are crucial: Technical Documentation: Produce comprehensive documentation detailing the development process, model configurations, and deployment instructions. Presentation Skills: Demonstrate the ability to present the project, its objectives, methodologies, and outcomes clearly and effectively to both technical and non-technical audiences.

Unit content

---

1. Demonstrate an ability to understand Neural Networks, Convolutional Neural networks, RLHF, theory of machine learning with applications to deep learning including the mathematics and statistics foundations.

The Unit outcomes are designed such that students have to demonstrate an ability to grasp the fundamental aspects of neural networks, convolutional neural networks and RLHF. They must understand forward and backpropagation, functions such as ReLu and Sigmoid functions, as well as metrics for the efficacy of Machine Learning generally and Deep Learning Models in particular including RMSE, F-1 score, AU-ROC, Accuracy, Recall, Confusion Matrix.

Furthermore, students should demonstrate they understand difference between classification tasks, regression tasks and generative AI applied to LLM models and generating pictures and videos such as Stable Diffusion Models.

Students will also demonstrate they understand the biological and cognitive neuroscience fundamentals and origins of deep learning models and go through work by Andrej Karpathy, Geoffrey Hinton, Fei-Fei Li, Yann Le Cun and Andrew Ng.

2. Apply Modelling Techniques in the Python Programming language across classification, regression, generative AI tasks

Students will demonstrate the ability to apply machine learning and deep learning models in classification and regression tasks – and understand where Deep Learning models are applicable and where other models such as XGBoost can be equally or more effective using lower compute resources.

Students will also demonstrate a command of using the Python Programming language and libraries used such as NLTK, Spacy, TextBlob, OpenCV, Tensorflow, PyTorch and PyTesseract.

3. Assess task requirements and plan a methodology of how to model a given problem

Students will demonstrate the ability to apply Deep Learning Models in real world scenarios such as classifying an apple from various images and testing the results of their model on a test set, and being able to fine tune an existing LLM model on a given use case – in this case apply it to create a ChatGPT alternative.

4. **Plan a Project from an idea to completion and Execute it, and assess the quality of their results**

Students will demonstrate through the capstone the ability to take a problem question and apply it to one of two use cases – infrared object detection using PyTorch from Infrared Images or creating a ChatGPT LLM alternative for a Sales Representative or a Cybersecurity Expert. Clearly articulate the problem statement for the chosen use case (infrared object detection or ChatGPT LLM alternative).Set specific, measurable, achievable, relevant, and time-bound (SMART) objectives for the project. Conduct a comprehensive literature review to understand current technologies, methodologies, and best practices related to the chosen use case. Select appropriate technologies and tools based on the project requirements. Given the expertise in Python, PyTorch, NLP

frameworks, and computer vision libraries, these should be leveraged effectively. Perform data cleaning, normalization, and augmentation as required to prepare the data for model training. Design and develop the model architecture. For infrared object detection, this might involve convolutional neural networks (CNNs). For the ChatGPT LLM alternative, explore transformer-based models. Train the model using the prepared dataset, applying techniques like cross-validation to assess model performance and avoid overfitting. Evaluate the model against predefined metrics relevant to the project objectives, such as accuracy, precision, recall, or F1 score. Optimize the model through hyperparameter tuning and retraining as necessary to improve performance. Document the project development process, model configurations, deployment steps, and user guidelines comprehensively. Conduct a thorough quality assessment of the project outcomes against the initial objectives. Reflect on the project execution, identifying successes, challenges, and lessons learned for future projects. The students will show an ability to understand the ethical considerations of the project and ethics of AI applied to real world projects.

Essential guidance for tutors

Tutors will need to use a wide range of teaching and learning methods so that learners meet the learning outcomes in this unit. Methods include lecture slides, example Jupyter Notebooks in Python, hands-on workshops, project work and individual and group assessments.

Some formal delivery will be necessary, but work can be increasingly learner-centered to develop independent learning. Learners need to adopt an investigative,analytical and participative approach to achieve the learning outcomes and reflecton their own experiences and roles to enhance the learning experience.

Sufficient time needs to be built into the delivery schedule to allow learners to undertake the research needed to help them meet the learning outcomes and be responsible for their own learning. References to open source books, open source code, example code from the CCM for each of the project tasks will be provided including basic prototypes for the Capstone to help learners get started in building the full model.

For Learning Outcome 1, Learners must understand that they need to read through in their own time as well as practice on Google Colab, Github and other open sources both Neural Networks and Convolutional Neural Networks. Learners must be able to understand probability theory, functions in mathematics, the three basic layer structure – input layer, processing layer and output layer as well as ReLu, Sigmoid and Soft Max Functions. They must understand model performance when the weights are changed as well as understand the concept of epochs in training Neural Networks.

For Learning Outcome 2, Learners must understand pooling layers and how object detection is done on CNNs. They must demonstrate an ability to test the results generated from their CNN models on new data and assess the efficacy of the modelling done.

For Learning Outcome 3, students must be able to understand the biological and cognitive neuroscience fundamentals of CNNs. They must understand how a biological neuron is abstracted mathematically, and also newer models which look at different attention mechanisms in the brain as well as neural networks. Cognitive differences between different animals will be used as an example to highlight how neural networks can be modeled differently for different use cases for Artificial Neural Networks. It is important Tutors provide enough of a background on the historical evolution of these models.

For Learning Outcomes 4, students must be able to demonstrate RL concepts such as Q-Learning, and most importantly understand how RLHF leads to modern natural language processing Neural Networks such as GPT-4 from OpenAI and Gemini from Google which is currently the benchmark in this field. Use of RLHF by DeepMind will also be covered in some of their seminal work in this field in making AlphaGo. Tutors are expected to show learners how to read academic papers in an efficient manner with an ability to implement them in code. Paperswithcode.com will be provided as a website to use for this in particular.

Learning Outcomes 5,6 are practice case studies. Students will be given standard case studies and encouraged to experiment with models to provide results and improve output of standard models provided to them. Students are encouraged to use all libraries given within the course syllabus as well as explore other libraries using PyPi as well as GitHub open source libraries. It is important that learners

For Learning Outcome 7, tutors must demonstrate and learners must show their ability to perform a series of steps starting from understanding the project scope, creating a

series of steps to execute the task, show the ability to test the efficacy of different approaches using smaller prototypes, and finally execute the selected prototype into the final model, generate the results and evaluate the results as well as show an understanding of the ethical considerations of Artificial Intelligence applied to real life.

Topic and suggested assignments/activities and/assessment
Learning Outcome 1,2 – Introduction to unit and programme of learning
Learning Outcome 1,2 – Tutor-led discussion on Neural Networks, Convolutional Neural Networks with Example Code.
Learning Outcome 1,2 – Tutor input on use cases for Neural Networks and CNN's followed by exercises in Python.
Learning Outcomes 1,2 – Learner research on creating a NN model in Python to classify Apple from Images. **Assignment 1: Create a CNN model and calculate test results and precision, recall, etc.**
Learning Outcomes 3 – Tutor input on theoretical basis and biological basis of NN and CNN models.
Learning Outcomes 3 – Learner Research on state of the art models. Create a Hebbian Learning Model or a Spiking Neural Network Model in Python. **Assignment 2:Hebbian Learning Model or a Spiking Neural Network Model in Python.**
Learning Outcomes 4,5 – Tutor discusses RL and RLHF models, with applications in Python. Tutor demonstrates in Python code for NLTK, PyTorch, Tensorflow, OpenCV and PyTesseract
Learning Outcomes 4,5 – Learner activity in applying a Tensorflow Model or a PyTorch Model **Assignment 3: User creates a Python model in Tensorflow or PyTorch to execute Tic-Tac-Toe using Q-Learning and assess results.**
Learning Outcomes 6 – Tutor goes through theoretical aspects of working with Huggingface and tuning hyper-parameters in LLM modelling.
Learner activity –Students write an essay on Hyperparameter tuning. **Assignment 4: Users write an essay on the importance of hyperparameter tuning in LLM models with sample code in the model explaining how the code exactly works to do hyperparameter tuning.**
Learning Outcome 7 – Tutor demonstrates the capstone project, the methodology as well as planning for the Capstone Project. The students are then told to execute the capstone project and discuss their results in peer group. **Assignment 5: Individual Assignment to do Capstone project with peer groups created to discuss modelling approach and discuss results of their modelling with each other.**

Assessment

Assessment is practical based where learners need to demonstrate their ability to apply their learning to generate results in the python programming language mostly. Code comments are used a way to gauge their understanding of the theoretical constructs in applying said theory. Only one essay has been asked for.

For Learning Outcome 1 and 2 – the learners need to demonstrate the ability to make a CNN model from scratch and explain why F-1 score, RMSE and other metrics are used.

For Learning Outcome 3- the learners need to explain in detail the mechanisms of a Hebbian Learning Model or a Spiking Neural Network Model and demonstrate ability

to apply the learning through Python Code.

For Learning Outcome 4,5 – learners again demonstrate use of PyTorch or Tensorflow by making a Tic-Tac-Toe using Q-Learning.

For Learning Outcomes 6 – students demonstrate understanding of Hyperparameter tuning through an essay, explaining in detail how hyperparameter tuning works and how to use it with sample code.

For Learning Outcomes 7 – students in reality show learning from Learning Outcomes 1,2,3,4,5,6 combined by taking a project scope, creating requirements, and executing it in code, generating results, metrics for how good the models are performing on the test sets, and discuss the results in the classroom.

Assessment criteria covered	Assignment title	Scenario	Assessment method
AC 1.1, 1.2, 1.3, 1.4, 2.1, 2.2, 2.3, 2.4	Implementation of NN and CNN basics.	Learners are asked to create a Python Model to classify an apple correctly from images.	Python Code in Jupyter Notebook explaining benchmark metrics such as precision, recall, F-1 Score, Confusion Matrix.
AC 3.1, 3.2, 3.3, 3.4, 3.5	Biological and Cognitive Neuroscience bases of ANN models.	Create Python code for Hebbian Learning or Spiking Neural Networks.	Python Code in Jupyter Notebook explaining on comments the biological and cognitive neuroscience theory behind modelling.
AC 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, 5.1, 5.2, 5.3, 5.4, 5.5	Q-Learning and RL Fundamentals	Create Python code for a Q-Learning model which plays Tic-Tac-Toe	Python Code with Tic-Tac-Toe which implements Q-Learning commented explaining each step.
AC 6.1, 6.2, 6.3, 6.4, 6.5, 6.6, 6.7	Explain HyperParameter tuning	Write an essay with sample code chunks in Python explaining Hyperparameter Tuning.	Essay explaining how hyperparameter tuning works and why it is necessary in modelling to get effective results.
AC 7.1, 7.2, 7.3	Capstone Project with	Write a Project Brief explaining	Project Brief and Python

	Project Brief and Python Code	how you went about creating the LLM for a ChatGPT alternative or the CNN model for infrared object detection.	Code for Capstone project with clear methodological choices explained, Python Code implementation, test results and explaining the validity of the model and room for improvement if applicable along with ethical considerations of applying Artificial Intelligence in real world projects.

Essential resources

Essential Resources for Artificial Intelligence Learning Outcomes

While there are no essential resources required for this unit, the following indicative resource materials are recommended to support the comprehensive understanding and application of artificial intelligence, specifically focusing on neural networks, convolutional neural networks (CNNs), reinforcement learning, and machine learning with Python as well as the historical basis for the work. These resources include textbooks, journals, websites, and open-source books that cover the theoretical foundations, practical applications, and latest advancements in the field.

Indicative resource materials

Textbooks

Goodfellow I., Bengio Y., Courville A. - Deep Learning (MIT Press, 2016) ISBN 0262035618: A foundational text offering comprehensive coverage on deep learning, including neural networks and their applications.

Nielsen M.A. - Neural Networks and Deep Learning (Determination Press, 2015): Provides an accessible introduction to neural networks, ideal for beginners.

Sutton R.S., Barto A.G. - Reinforcement Learning: An Introduction (MIT Press, 2nd edition, 2018) ISBN 0262039249: A key resource on reinforcement learning, covering foundational concepts and algorithms.

Journals

Journal of Machine Learning Research (JMLR): Offers open-access articles on machine learning, including deep learning and reinforcement learning research.

Neural Networks: Publishes original research on neural network theory and applications, bridging the gap between neuroscience and AI.

Websites

DeepLearning.ai: Provides courses, resources, and community support on deep learning and AI technologies.

Machine Learning Mastery: Offers practical advice and tutorials for mastering machine learning in Python.

OpenAI: Shares research and resources on advancements in AI, including reinforcement learning and large language models.

TensorFlow: The official website for TensorFlow, providing documentation, tutorials, and community forums.

PyTorch: The official site for PyTorch, offering comprehensive guides, tutorials, and community projects.

Open Source Books

Python Data Science Handbook by Jake VanderPlas: An open-source book covering Python's data science stack, including NumPy, pandas, Matplotlib, Scikit-Learn, and more.
https://jakevdp.github.io/PythonDataScienceHandbook/

Dive into Deep Learning by Aston Zhang, Zachary C. Lipton, Mu Li, Alexander J. Smola: An interactive deep learning book with code, math, and discussions, intended for practitioners.
https://d2l.ai/