BCS Foundation Award Machine Learning

Syllabus V1.1

Document V1.0

*SFIA**plus***

This course provides candidates with the level of knowledge highlighted within the table, enabling them to develop the skills to operate successfully at the levels of responsibility indicated.

Level	Levels of knowledge	Levels of skill and responsibility (SFIA)
K7	Evaluate	Set strategy, inspire, mobilise
K6		Initiate and influence
K5	Synthesise	Ensure and advise
K4	Analyse	Enable
K3	Apply	Apply
K2	Understand	Assist
K1	Remember	Follow

Key topics

Syllabus area	Syllabus weighting	Question level
1. What is machine learning?	20%	K1/K2
2. Coding for machine learning	20%	K1
3. Algorithms used in machine learning	30%	K1/K2
4. Machine learning in practice	30%	K1/K2

Learning outcomes

1. What is machine learning?

1.1 Define machine learning

1.2 Explain different applications of machine learning

1.3 Describe the role of a learning agent

1.4 Explain the concept of deep learning

1.5 Describe the purpose of a neural network

1.6 Illustrate how machine learning compliments knowledge-based systems

1.7 Explain the process through which machine learning works with data

2. Coding for machine learning

2.1 Explain the use of at least one coding language used in machine learning

2.2 Identify common open source and proprietary software used in coding for machine learning

Learning outcomes

3. Algorithms used in machine learning

3.1 Explain the use of mathematics in enabling a machine to solve numerical problems

3.2 List and describe typical algorithms used in machine learning

3.3 Describe supervised, unsupervised and semi-supervised learning

4. Machine learning in practice

4.1 Describe a particular problem that can be addressed through the use of machine learning

4.2 Outline typical tasks required in the preparation of data for developing a particular application of machine learning

4.3 Explain the process of training a machine learning model

4.4 Explain the process of testing a machine learning model

4.5 Discuss how to evaluate the results of testing in order to identify the information to be shared with key stakeholders

What is machine learning?

Key topic 1

20% K1/K2

1.1 Define machine learning

It is important for learners to understand that machine learning is a subset of AI (artificial intelligence). AI itself is not a new concept; machine learning is another step in the evolution of AI. Machine learning is used within data science and is the application of algorithms to derive insight from data and Big Data.

The definitions of AI and machine learning

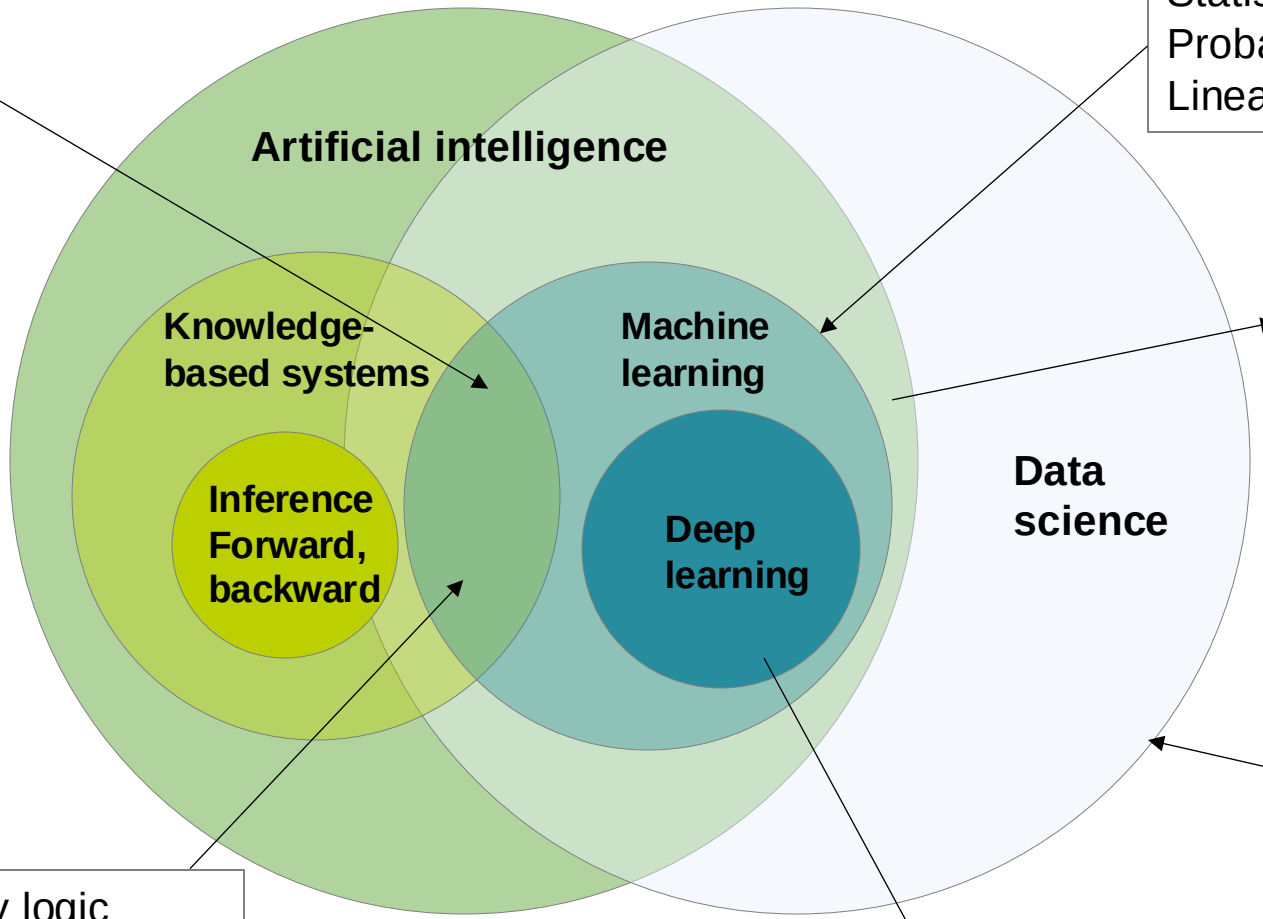
Artificial intelligence is “the science and engineering of making intelligent machines”

John McCarthy

“Machine learning (ML) is the study of computer algorithms that allow computer programs to automatically improve through experience.”

Tom Mitchell

Machine learning and AI



Data
Structured
Unstructured
Semi-structured

Mathematics
Statistical analysis
Probability
Linear algebra

ML approaches:
Structured
Unstructured
Semi-structured
Reinforcement

Data scientist tasks:
Requirements gathering
Data mining
Data cleaning
Data exploration
Feature engineering
Predictive modelling
Data visualisation

Fuzzy logic
Baysian updating

Neural networks

Diagram developed as an adaptation of Professor Adrian A. Hopgood's model as presented within his book "Intelligent Systems for Engineers and Scientists: A Practical Guide to Artificial Intelligence", 4th Edition 2022

1.2 Explain different applications of machine learning

Machine learning can be used in a number of contexts to complete different types of tasks. Learners should be encouraged to explore different examples and applications of machine learning.

Applications of machine learning

Discriminative outputs:

- Prediction or forecasting
- Object recognition
- Classification
- Clustering
- Recommendations

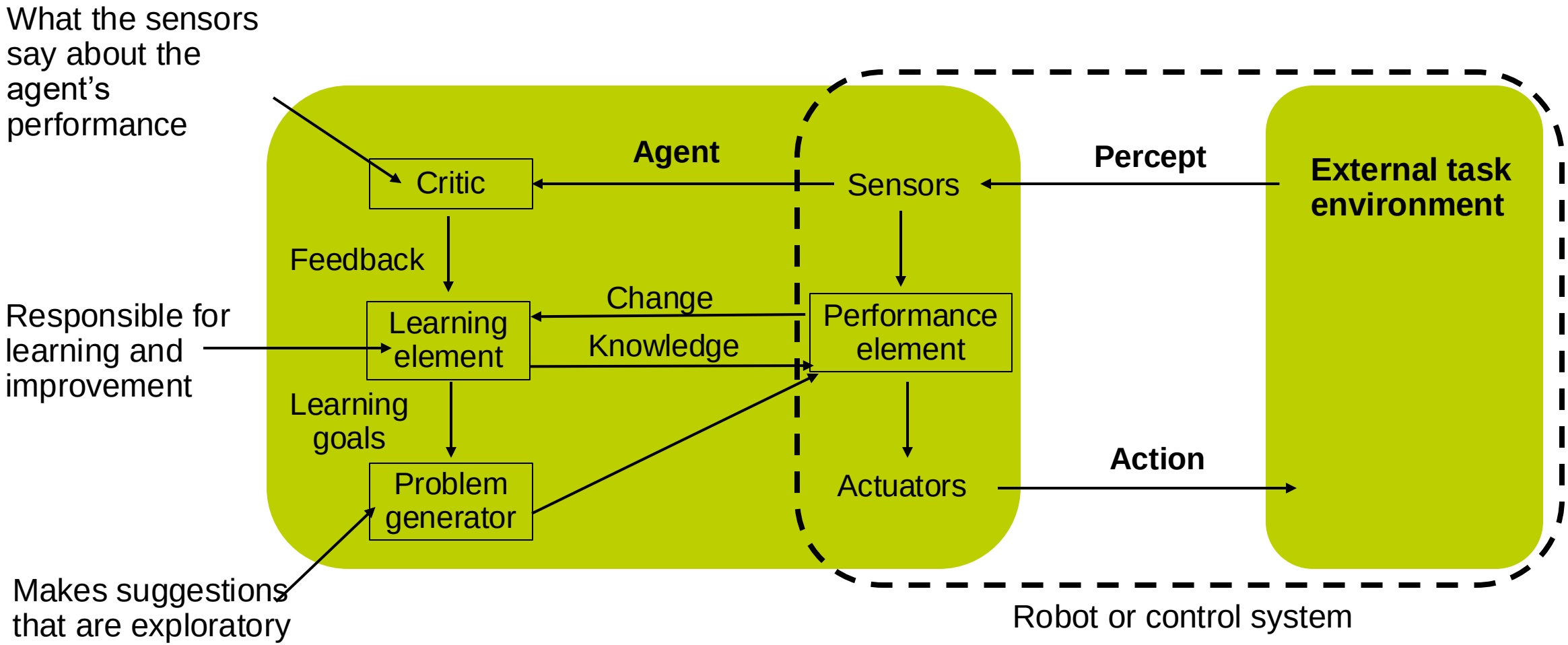
Generative outputs:

- Generating content

1.3 Describe the role of a learning agent

Learning agents are commonly used in machine learning. Each agent is designed to undertake a specific task using a given amount of data, which they undertake autonomously. Through the repetition of undertaking this task, they learn to improve each time. Examples include chatbots, driverless cars, facial recognition.

The learning agent



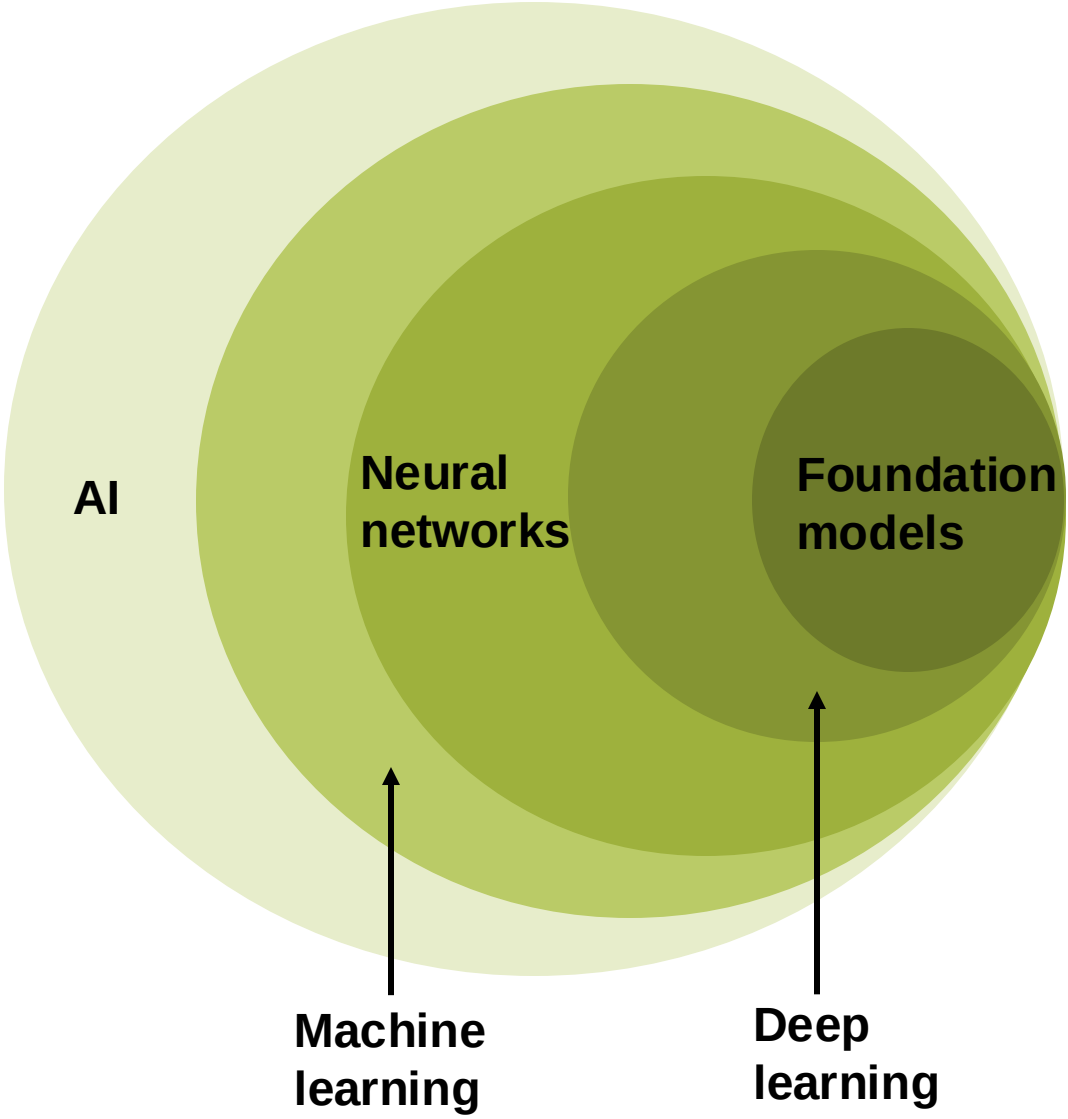
Based on the structure of agents in Russell and Norvig (2016). Adapted from Lowe, A. and Lawless, S. (2024). *Artificial Intelligence and Machine Learning Foundations*, p.36. 2nd Edition. Swindon. BCS, The Chartered Institute for IT.

1.4 Explain the concept of deep learning

The application of deep learning (a subset of machine learning) involves the training of large neural networks to process and analyse vast amounts of data to derive greater insight and to solve more complex problems.

Deep learning

- "Deep" refers to the many layers in the network, with each layer learning more complex features of the data
- A type of ML that uses multi-layered neural networks*, inspired by the human brain
- Made up of layers of nodes (neurons) connected to each other, where each layer processes information and passes it to the next
- Each layer learns a different part of the task

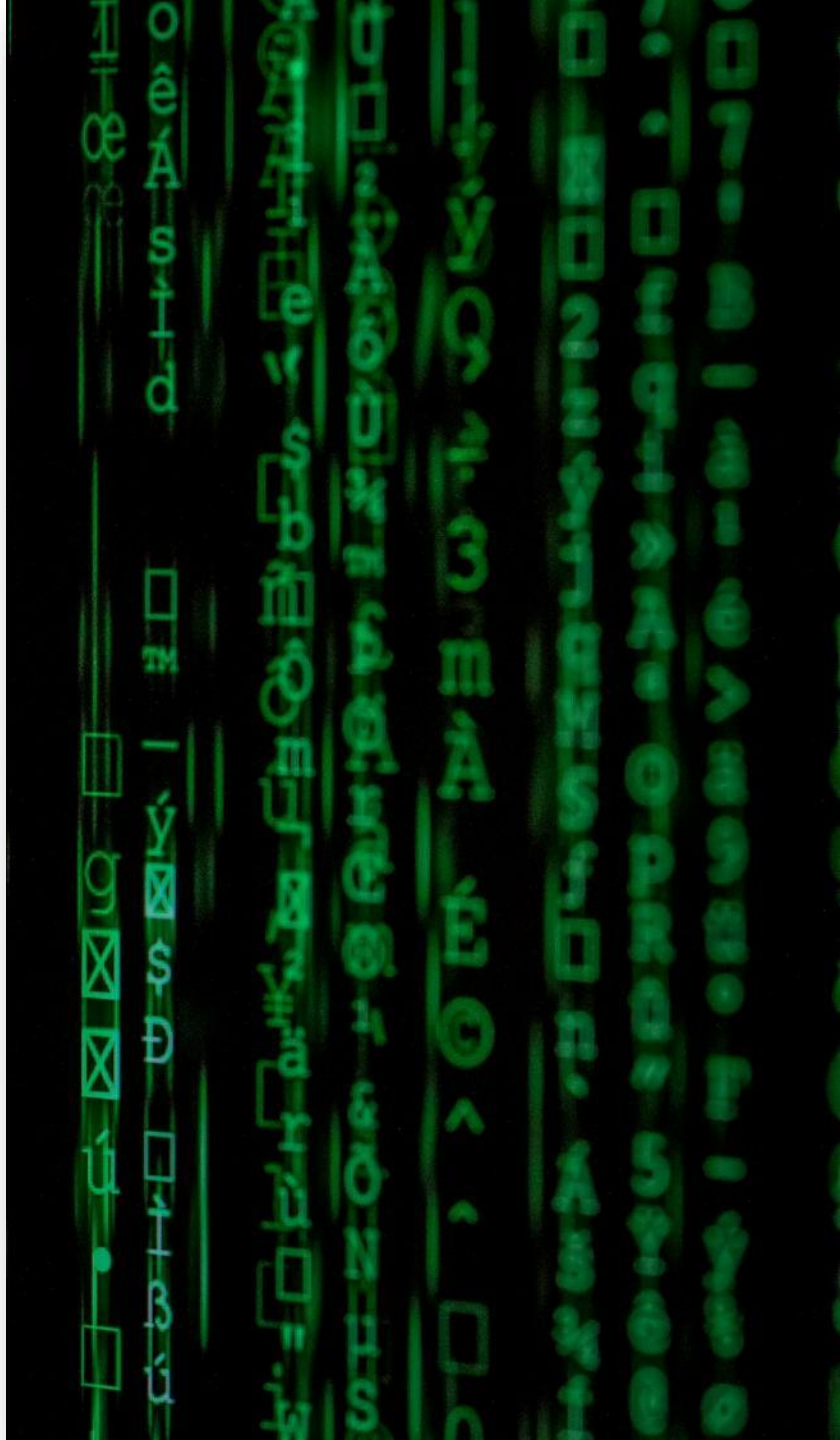


The position of deep learning with AI. Adapted from Lowe, A. and Lawless, S. (2024). *Artificial Intelligence and Machine Learning Foundations*, p.12. 2nd Edition. Swindon. BCS, The Chartered Institute for IT.

Combination with large data sets

Large data sets enhance deep learning by enabling neural networks to:

- Recognise patterns at scale
- Perform better with unseen data
- Perform better when given smaller sets of data for specific tasks
- Improve generalisation
- Uncover new insights and support data-driven discoveries



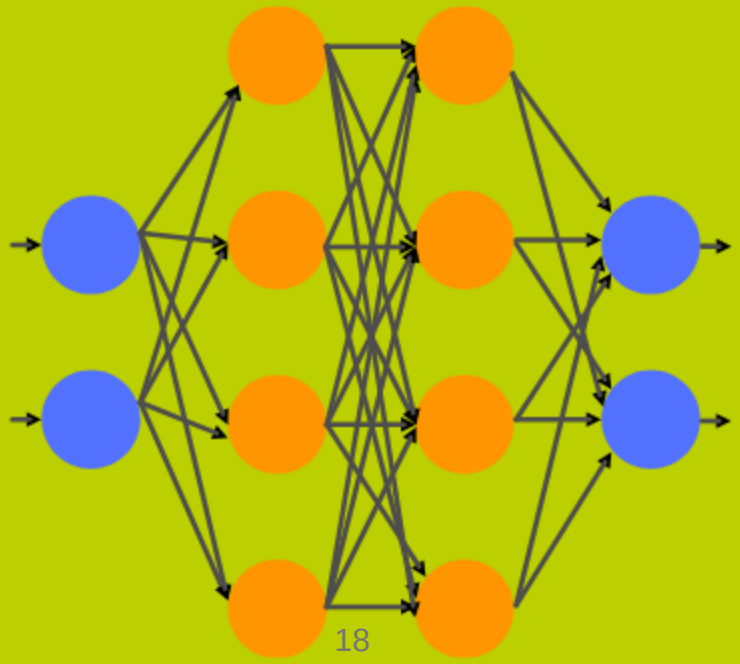
1.5 Describe the purpose of a neural network

Neural networks are commonly used in machine learning, particularly in the analysis of unstructured or unlabelled data (e.g. images, handwritten documents), whereby the input data is analysed to determine any recognisable or similar patterns against other learned bits of data in order to determine the output.

Neural networks…

…are computer programs that take an input and analyse its features using layers of interconnected units (neurons) to derive an output that's based on the likelihood that the input corresponds to the stored output.

INPUT

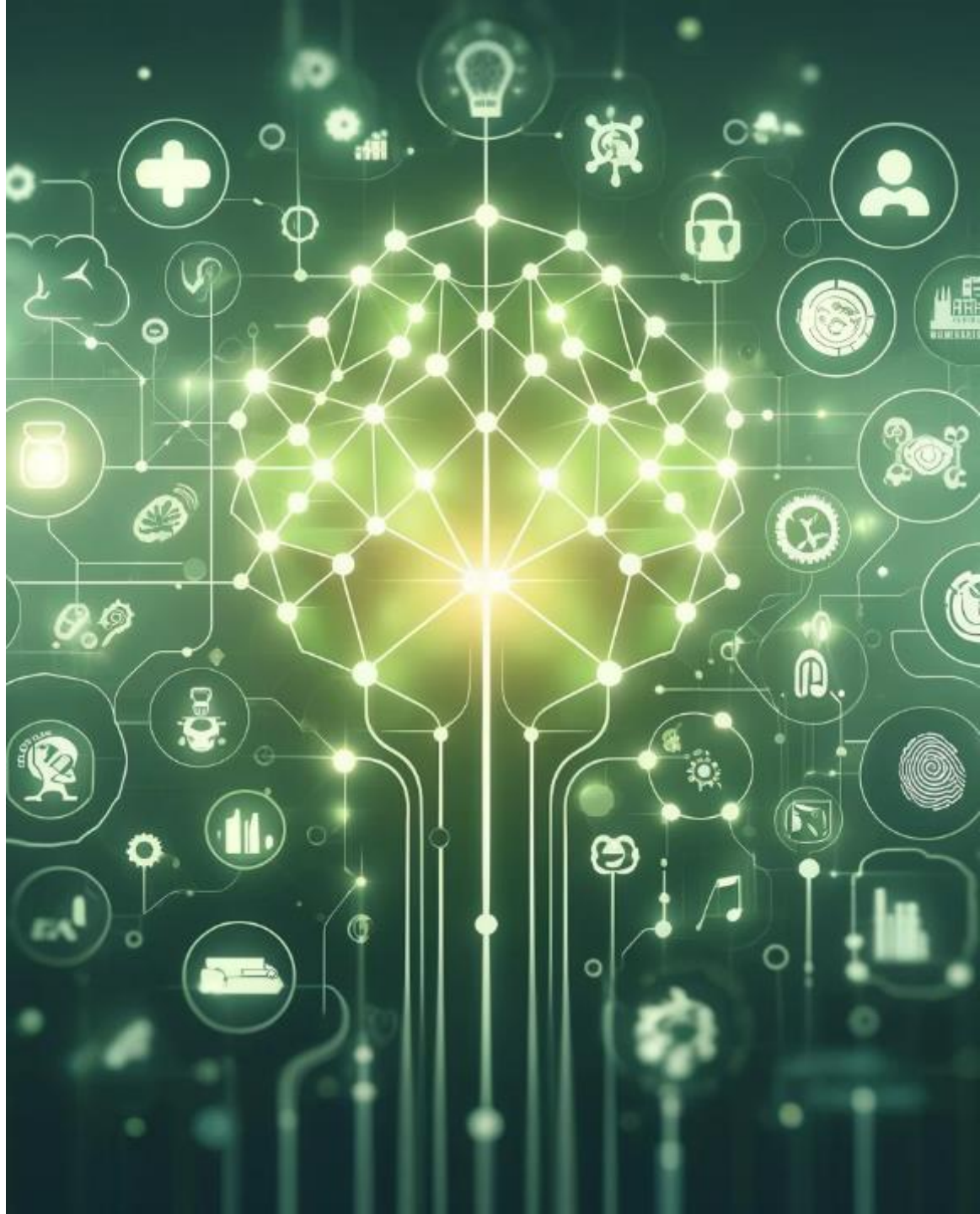


OUTPUT

Decision making

Neural networks are widely used in decision-making tasks where they:

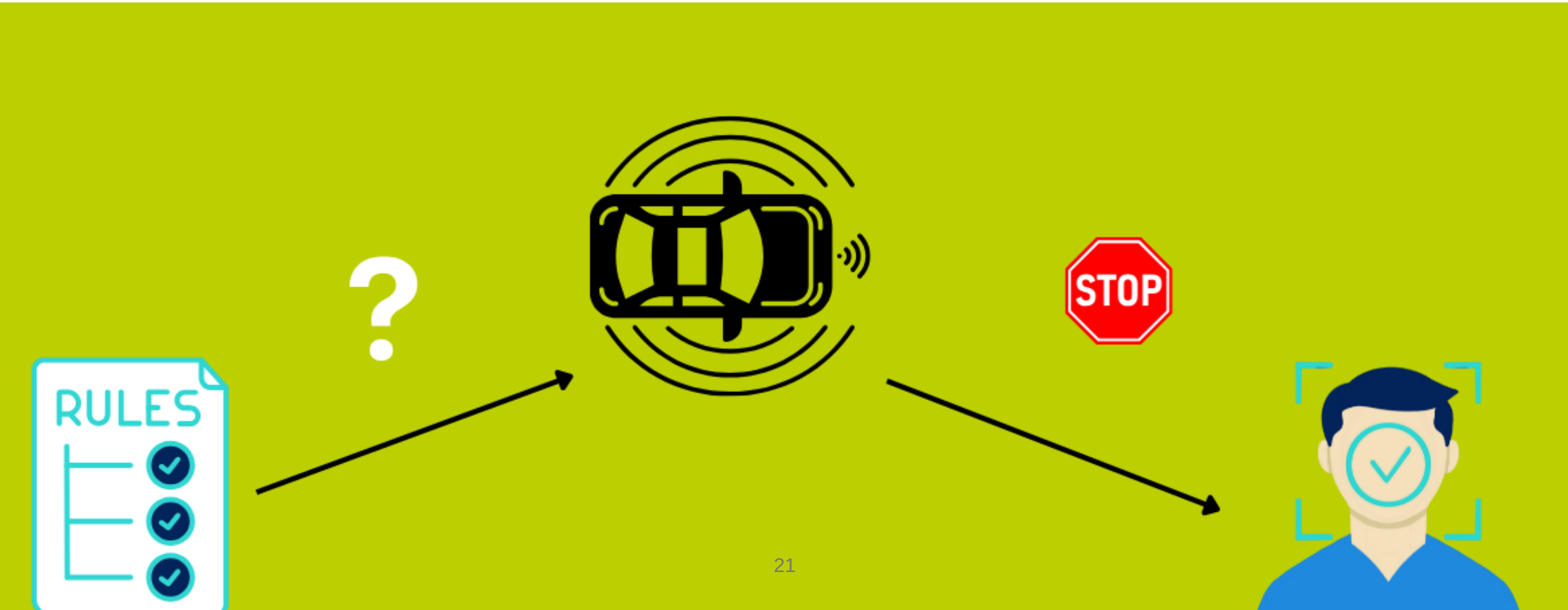
- Interpret complex, multi-dimensional data
- Predict future trends based on historical patterns
- Adapt to new data through continuous learning



1.6 Illustrate how machine learning complements knowledge-based systems

A knowledge-based system is a form of AI designed to capture human expertise/knowledge (within a knowledge base) and apply a set of rules to identify an outcome (through an inference engine). Machine learning is data-based and can derive outcomes through the use of algorithms e.g. a neural network.

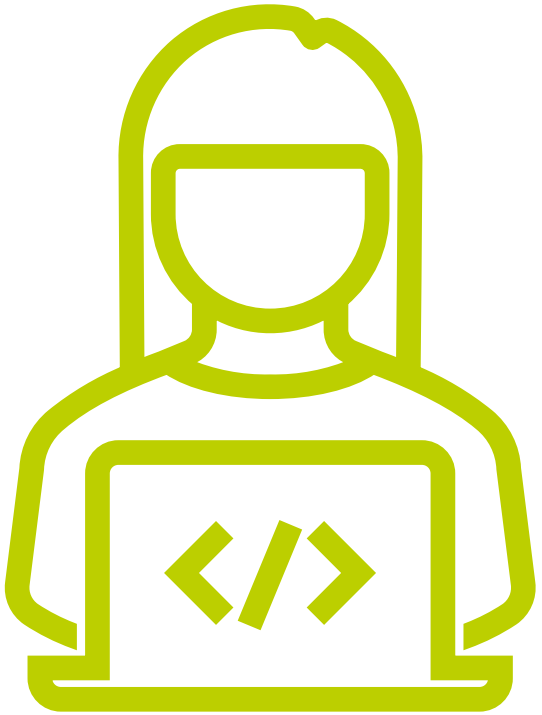
Systems that use both machine learning and knowledge-based systems (KBS)



1.7 Explain the process through which machine learning works with data

The Machine Learning Process allows us to define the solution based on the problem that has been identified through the process of data selection, pre-processing, visualisation and testing of data with specific algorithms. Once we are happy that both the data and the algorithms we have chosen to use are performing well we can deploy our model.

Machine learning:
The use of data in training AI



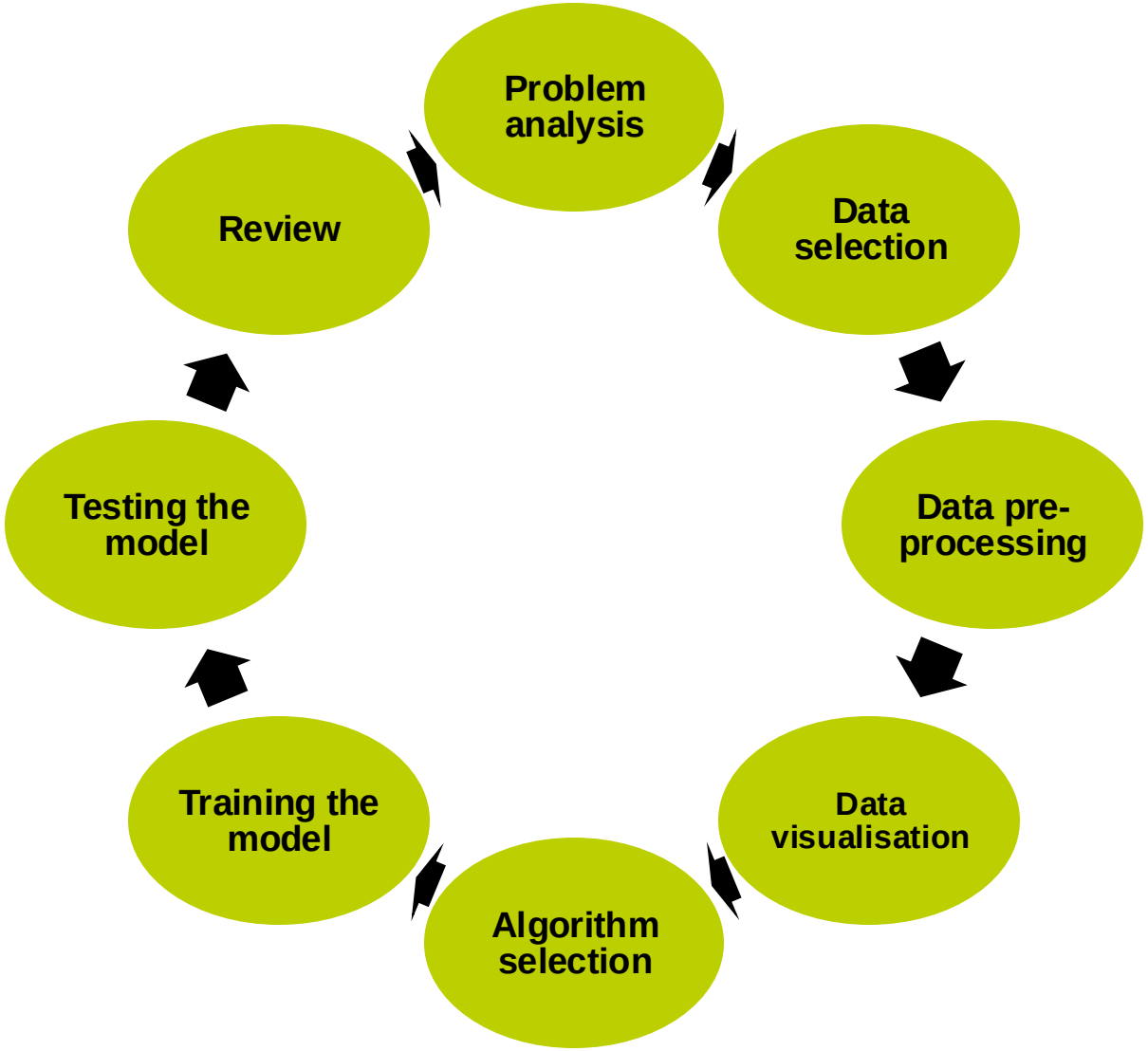
Data can be leveraged at all stages of a machine learning project to:

- Enhance understanding
- Facilitate communication
- Ensure better decision-making

Throughout the machine learning project lifecycle, data is integral to:

- Ensure clarity
- Enhance insights
- Drive informed decision-making

The machine learning process



Coding for machine learning

Key topic 2

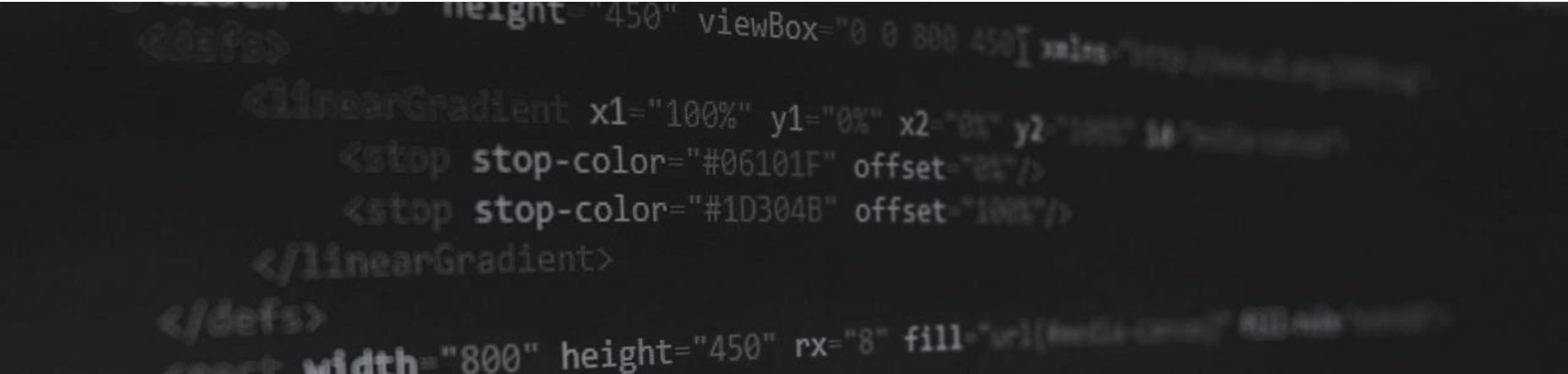
20% K1

2.1 Explain the use of at least one coding language used in machine learning

Learners should be familiar with common programming languages and their use, although it is not expected that they are fluent in using them. Python is a very popular language used in machine learning and data science. Libraries are used to bundle functions into templates that include the use of different programming languages e.g. Python.

Common programming languages

- Python
- R
- C++
- Java

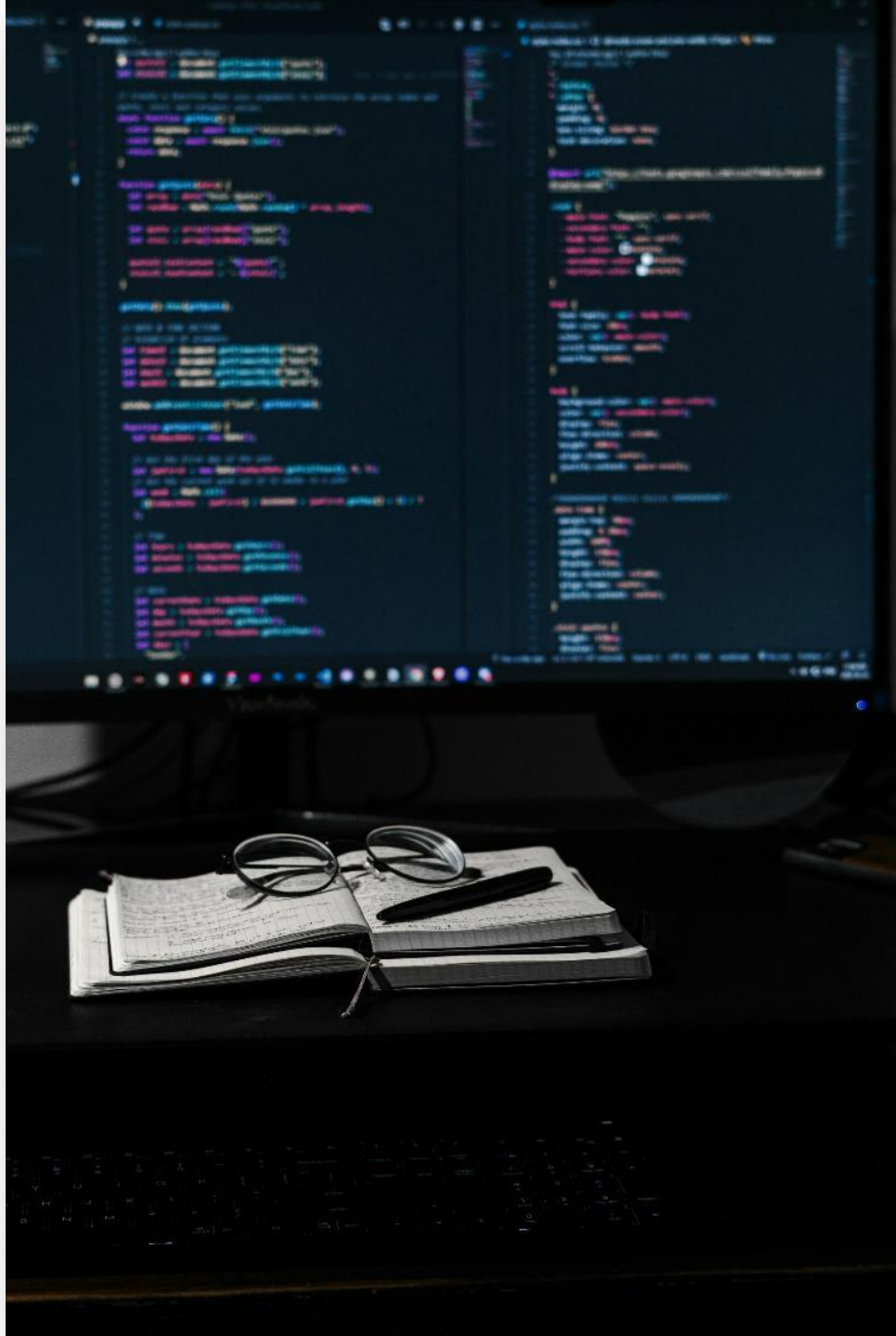


Libraries / Templates

- Provide pre-built functions and frameworks to streamline ML tasks
- Reduce the need to write algorithms from scratch, saving time and effort
- Enable developers to focus on model design and experimentation
- Help standardise processes, ensuring consistent and efficient workflows
- Often include support for data manipulation, model training, evaluation, and deployment

The benefits of Python for machine learning

- Easy to learn and use
- Faster development
- Rich library support
- Built-in utilities
- Community support
- Flexibility

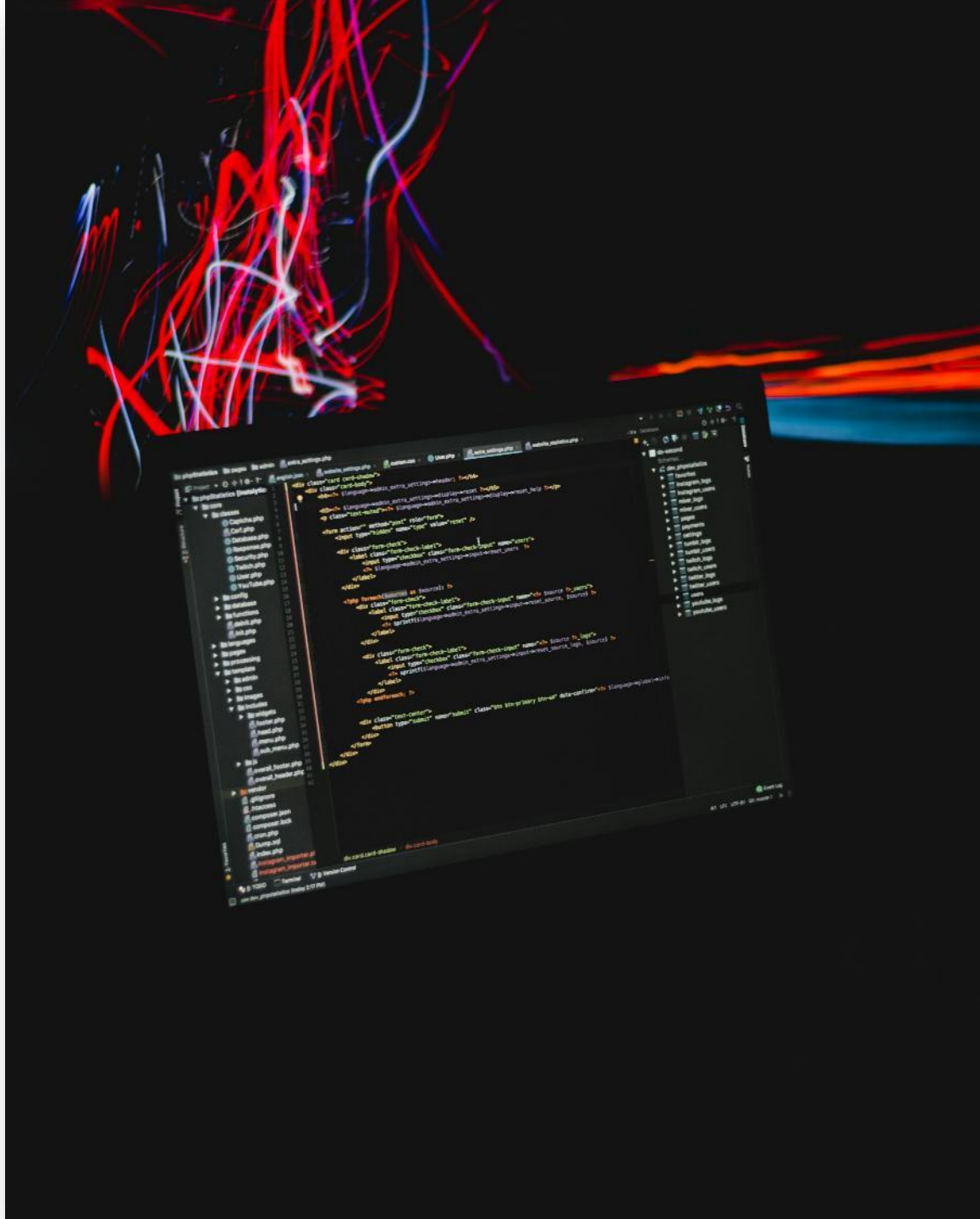


2.2 Identify common open source and proprietary software used in coding for machine learning

Learners should be encouraged to explore some of the known software and programming environments used in programming machine learning. It is not expected that they are proficient in their use however they should be familiar with at least one software.

Software used in programming ML

- Tensorflow
- R Studio
- CUDA
- Scikit-Learn
- MATLAB



Algorithms used in machine learning

Key topic 3

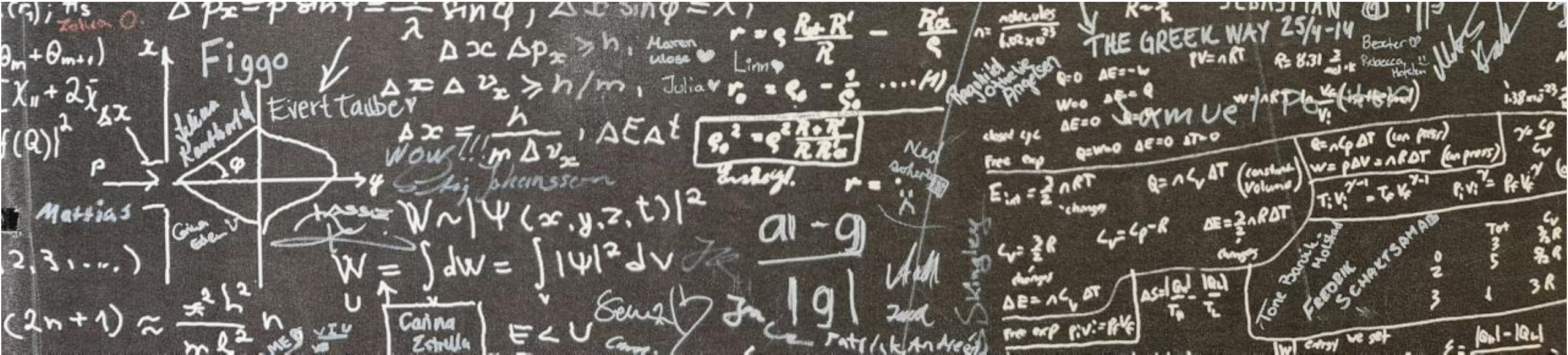
30% K1/K2

3.1 Explain the use of mathematics in enabling a machine to solve numerical problems

It is important for learners to have a basic understanding of the mathematics used within machine learning, regardless of whether the software they go on to use handles this automatically. Bayes Theorem is a method which can be used to calculate probability where other probabilities are known.

Mathematics in ML

- Forms the foundation for algorithms and models
- Enables accurate problem-solving and predictions
- Core components: probability, statistics, linear algebra



Probability and statistics

- **Probability:**

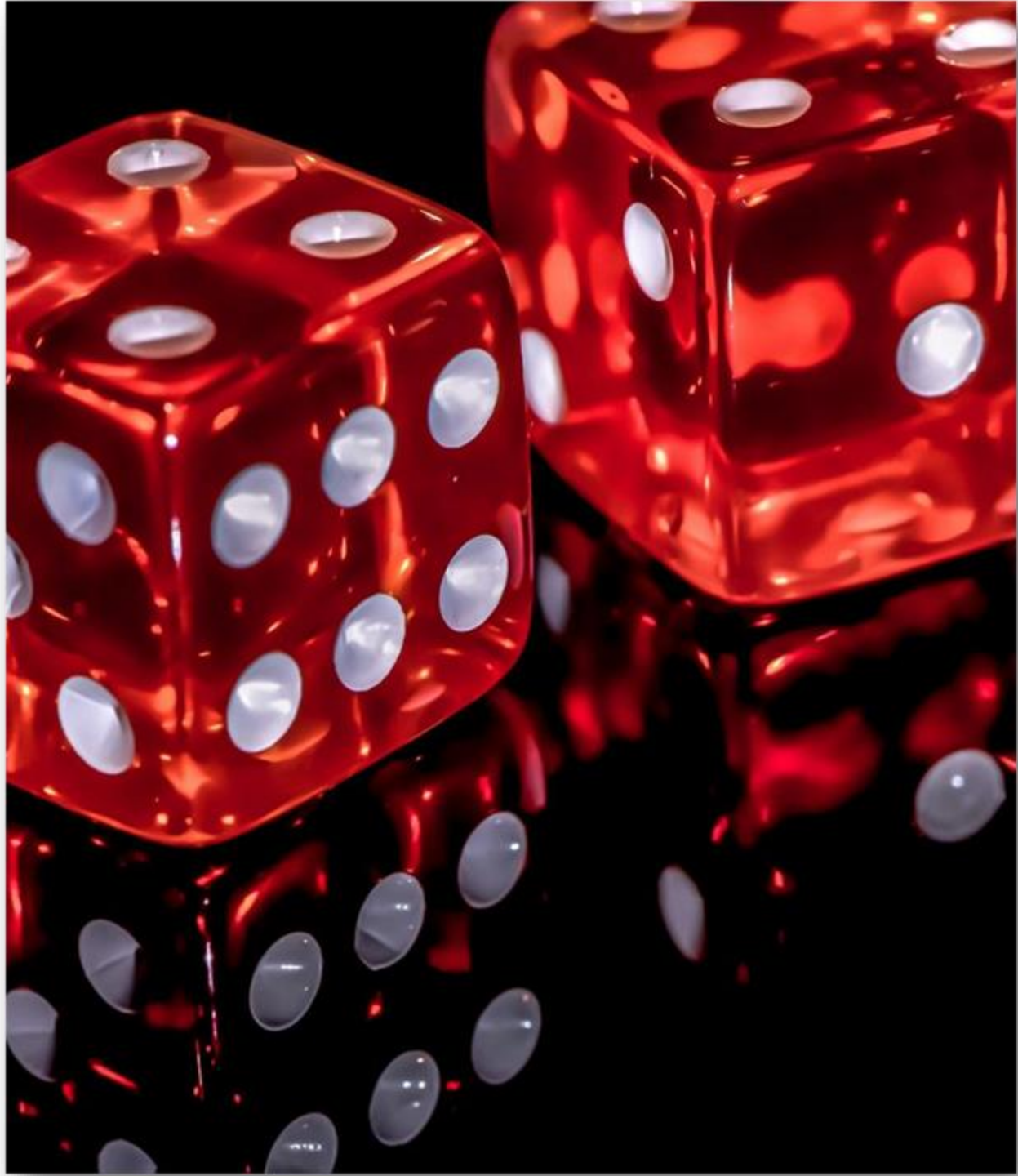
  Measures likelihood and uncertainty
- **Bayes Theorem:**

  Updates beliefs with new data
- **Descriptive statistics:**

  Summarises data

  (e.g. mean, median)
- **Inferential statistics:**

  Draws conclusions from samples



Linear algebra

- Vectors and matrices represent data
- Matrix operations underpin algorithms
- Concepts:
  - Multiplication
  - Eigenvalues
  - Vector spaces
- Enables efficient computations and model training

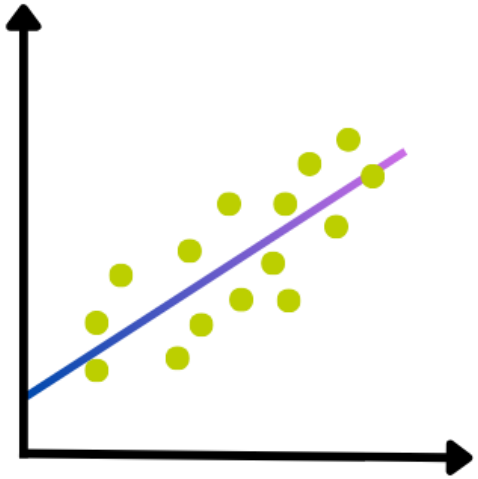
3.2 List and describe typical algorithms used in machine learning

Learners should have a basic understanding of some of the common algorithms used in machine learning and where they may be used in supervised or unsupervised learning. It is not essential at this level for them to understand the specific formulas used within each algorithm, however it is certainly advantageous to have a basic understanding of the maths involved in order to make it easier to program machine learning.

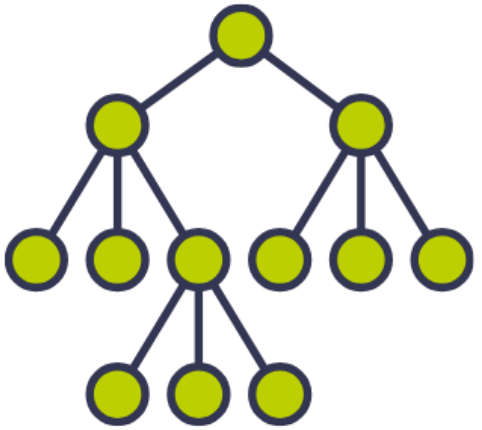
Algorithms – Forecasting & classification

A set of rules or instructions designed to enable a computer program to perform a specific task.

Common categories are forecasting, classification, and clustering*.

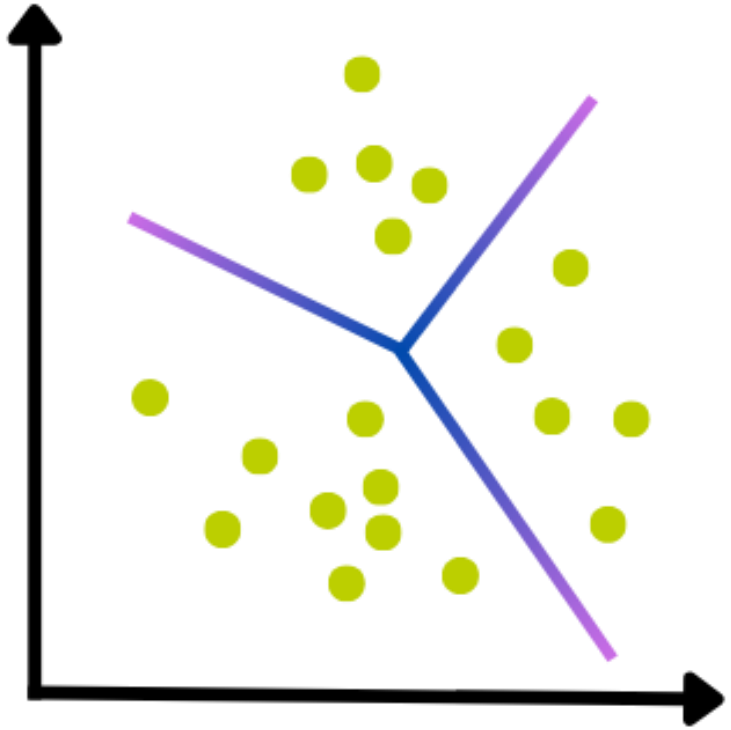


Forecasting, e.g. linear regression



Classification

Algorithms - Clustering



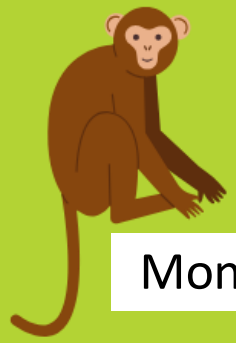
Clustering is used in unsupervised learning where there is unlabelled data.

3.3 Describe supervised, unsupervised and semi-supervised learning

It is useful for learners to have a basic understanding of the different types of approaches to machine learning to understand how it can be used to work with different types of data and where different algorithms are best used.

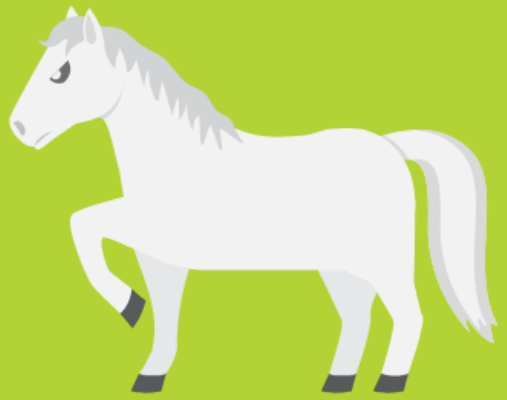
Labelled and unlabelled data

Labelled



Monkey

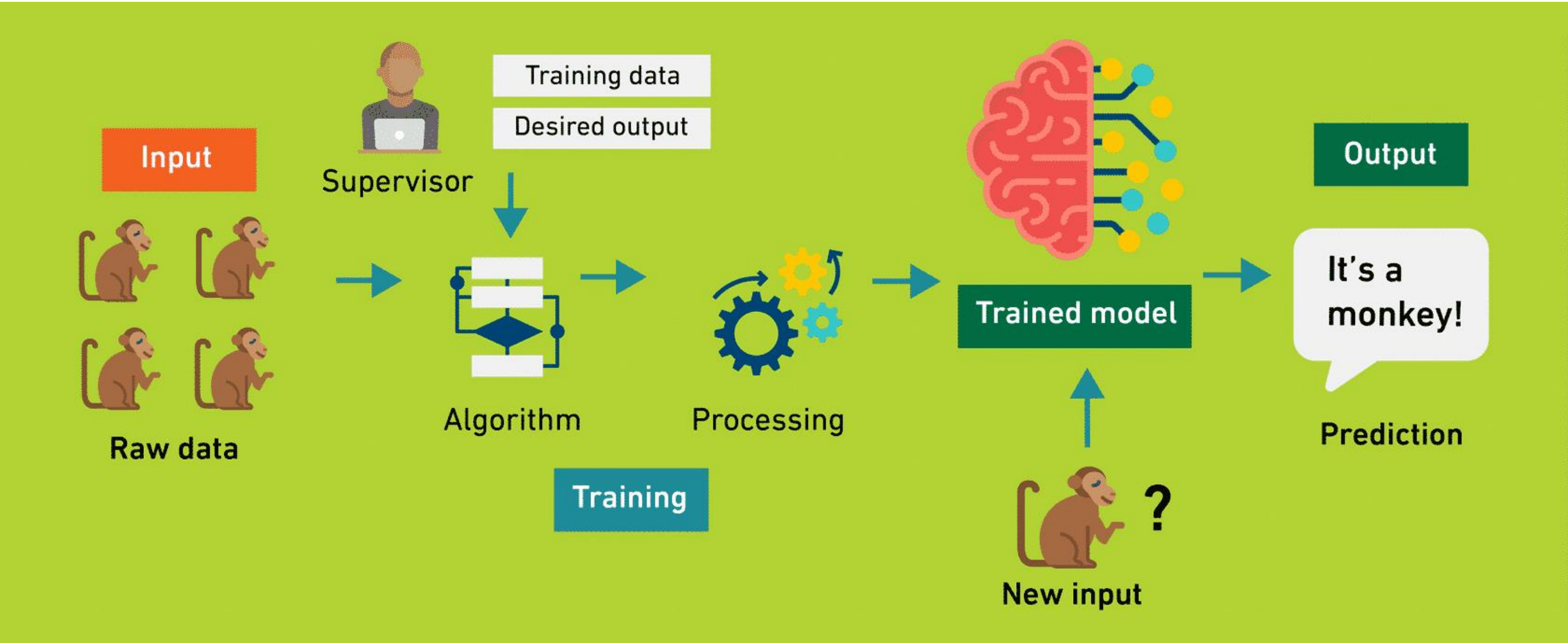
Cat



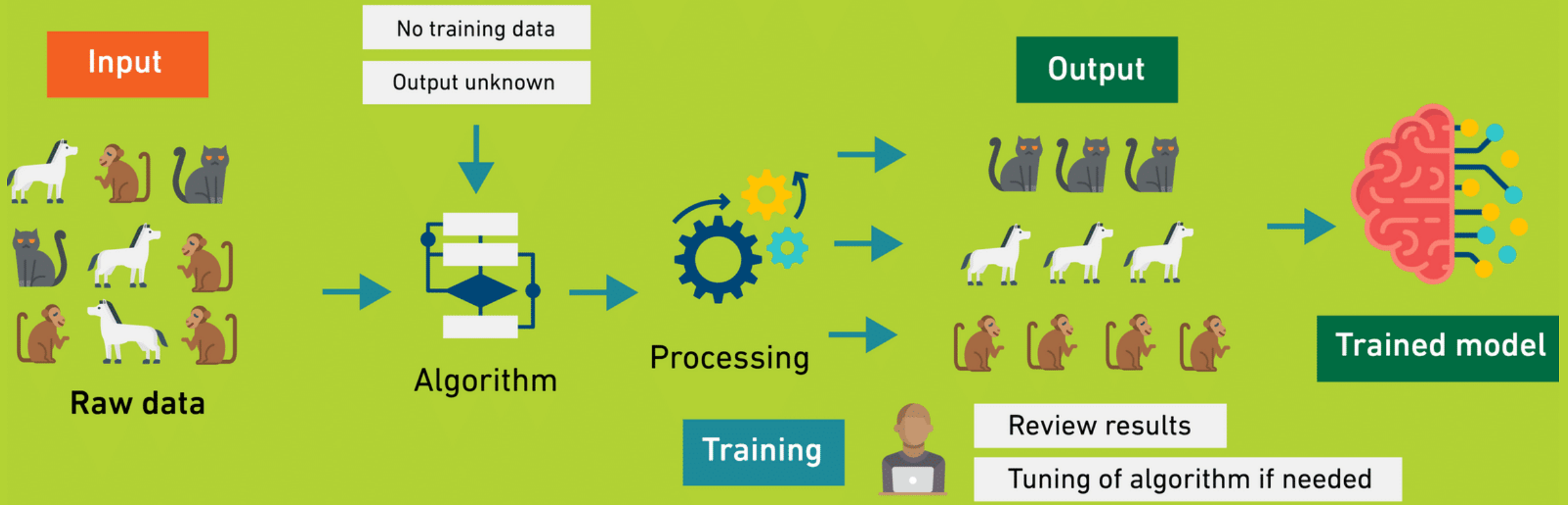
Horse

Unlabelled

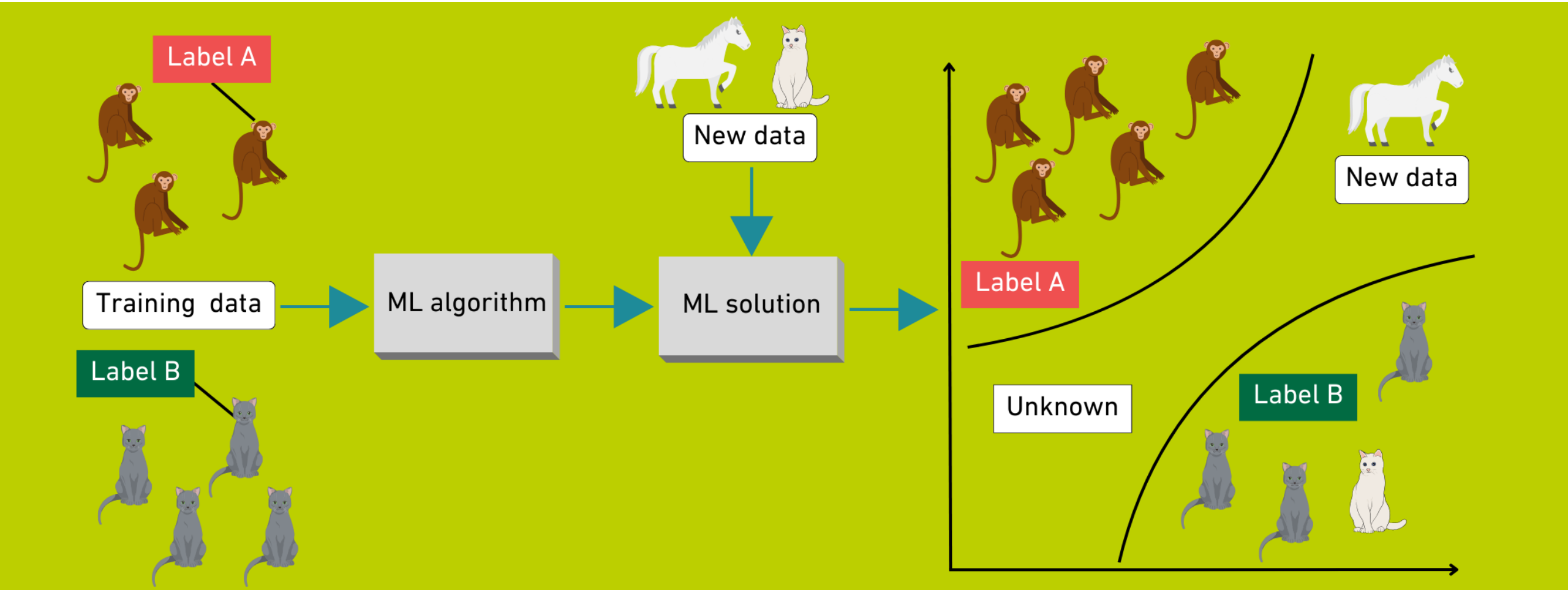
Supervised learning



Unsupervised learning



Semi-supervised learning



Adapted from: Lowe, A. and Lawless, S. (2024). *Artificial Intelligence and Machine Learning Foundations*, p.57. 2nd Edition. Swindon. BCS, The Chartered Institute for IT.

Machine learning in practice

Key topic 4

30% K1/K2

4.1 Describe a particular problem that can be addressed through the use of machine learning

Learners should be encouraged to identify a specific problem which could be solved through implementing machine learning.

Problem solving with ML

1. Identify the problem that needs solving, e.g.:

    - Slow diagnosis of medical conditions
    - Shoppers abandoning online carts
    - Unreliable system for equipment maintenance

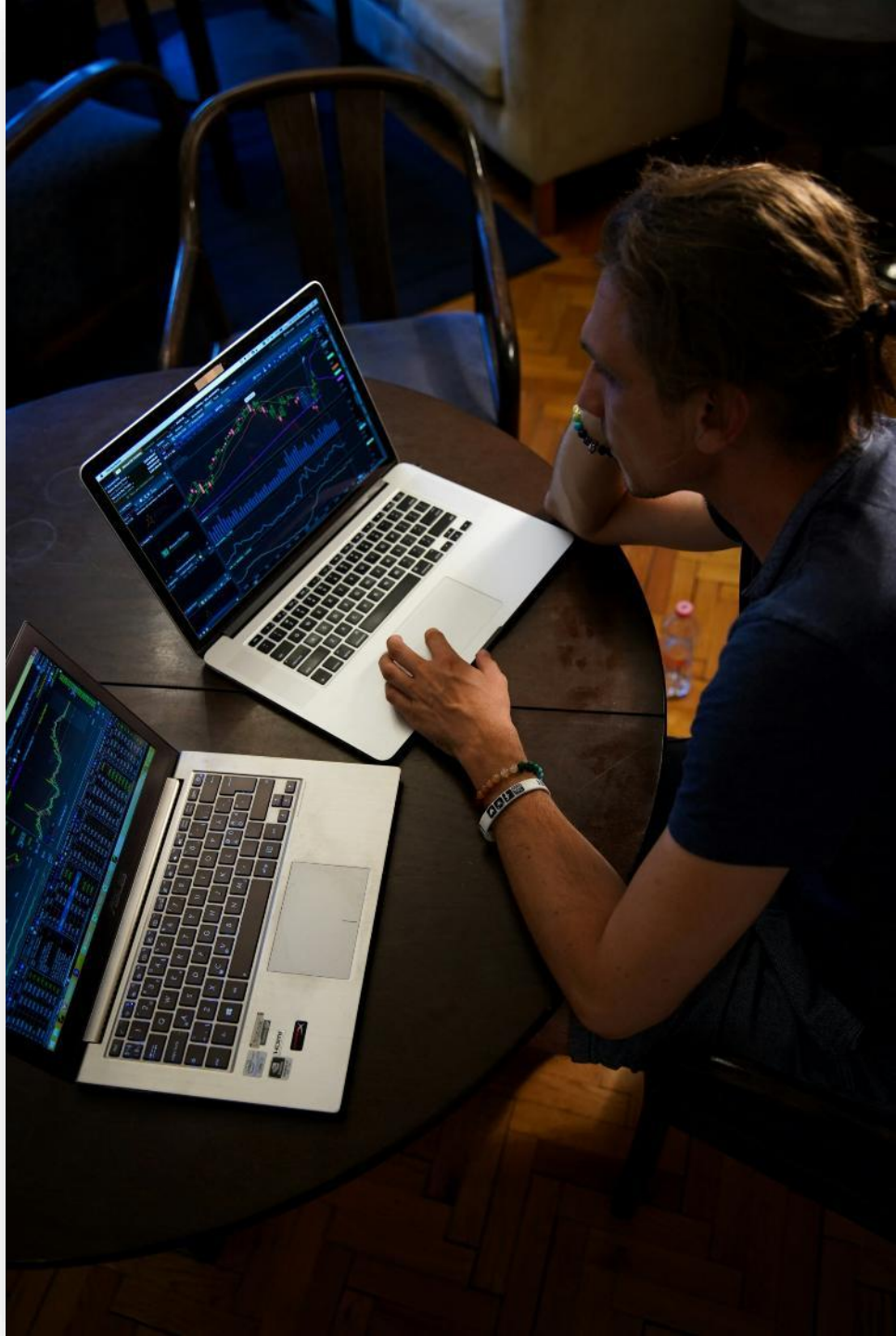
2. Determine data requirements and sources
3. Propose a suitable ML approach
4. Consider feasibility and constraints

4.2 Outline typical tasks required in the preparation of data for developing a particular application of machine learning

Learners should be able to outline the tasks they would need to undertake to prepare the data for use within an application of machine learning. This may include steps such as cleaning the data, data validation, and data transformation to ensure it is in a suitable format for using within a chosen software.

Preparing data for ML

1. Clean and handle missing values:
    - Identify outliers to either remove or correct them
    - Remove incomplete data fields or insert missing values by interpolation
2. Validate data quality and assess potential bias in data sources
3. Transform and scale features
4. Import/load data into chosen software



4.3 Explain the process of training a machine learning model

Learners should be able explain the process of training a particular algorithm using their prepared data.

Training an ML model

1. Ensure that the data meets training requirements (check quality, format, transformations)
2. Split data into three sets:
   - Training data
   - Validation data (used to check the model periodically during training, this is useful for avoiding overtraining)
   - Test data (comprising previously unseen examples, for checking the model)
3. Select an appropriate algorithm (match problem type and data structure)
4. Apply rules or constraints (e.g. regularisation, domain-specific conditions)
5. Choose supervised, unsupervised, or semi-supervised - based on available labels (see previous slides)

4.4 Explain the process of testing a machine learning model

Learners should be able to explain the process through which they tested a particular algorithm using their prepared data and how they identified whether it was performing well. They may use a number of methods to test their algorithm, and they may wish to test and compare multiple algorithms.

Testing an ML model

Rigorous testing during the development phase helps ensure that the AI models perform as expected and does not introduce unforeseen risks.

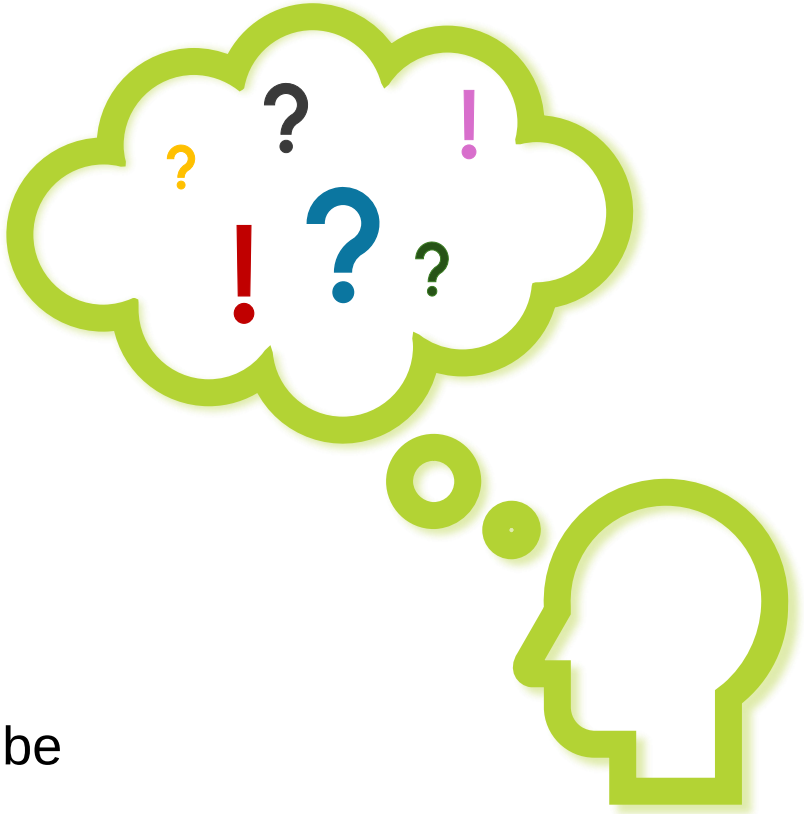
The testing process involves:

- Evaluating ML performance on unseen data
- Fine-tuning hyperparameters
- Combining multiple models (ensembles)
- Conducting statistical tests
- Reviewing and refining

4.5 Discuss how to evaluate the results of testing in order to identify the information to be shared with key stakeholders

Learners should be able to explain how they would go about identifying the key pieces of information to share with their stakeholders. They should also explain key considerations for sharing information with stakeholders e.g. type of information, presentation, language and use of technical terms, being prepared to answer questions.

Evaluating testing results



Check! Do the results align with your goals?

- Double-check the accuracy of the analysis
- Evaluate findings (interpret results, validate outcomes)
- Check for bias. ML algorithms start out neutral, but bias can be introduced unwittingly through the data and its use
- Identify stakeholder-relevant information (who needs to know what)
- Reflect on the problem (have we addressed it?)
- Draw conclusions (key takeaways, lessons learned)
- Plan next steps (future improvements, further investigation)

Who are our stakeholders? What information do they need?

Stakeholders

- Organisation leaders
- Developers
- End users
- Compliance teams
- Regulators

Information to be shared

- The AI model's impact on organisational operations and the wider environment, costs, return on investment
- Technical details for further refinement
- Transparency on how the model works, its limitations, implications of use, risks
- Ethical concerns, biases, how data is used

What should be communicated?

- **Performance metrics:**
  Accuracy, precision, recall, areas for improvement
- **Key insights and impact on the organisation:**
  Does the model meet the goals set? What are the risks and benefits?
- **Risk management strategies**
- **Challenges and limitations:**
  Bias, fairness, ethical concerns
- **Next steps:**
  Further training needed, adjustments/refinement, deployment plans

Considerations for effective communication

- Tailor presentation methods
- Use accessible language
- Highlight key insights
- Prepare for questions
- Document lessons learned


bcs
The
Chartered
Institute
for IT