Core AI Concepts	
Artificial Intelligence (AI)	AI is an umbrella term that encompasses various technologies, such as machine learning, deep learning, and natural language processing (NLP). It includes systems that perform tasks normally requiring human intelligence for example, reasoning, learning, decision-making.
Machine Learning (ML)	Machine Learning (ML) is a field of study within artificial intelligence (AI) that focuses on developing and analysing statistical algorithms that learn from data patterns rather than explicit programming. These algorithms learn from data and generalize their knowledge to perform tasks without explicit instructions.
Deep Learning	A type of ML using neutral networks with many layers to process complex data like images or speech.
Model	The trained system/algorithm used to make predictions or decisions based on input data.
Algorithm	A set of rules or instructions a computer follows to solve a problem
Data & Training	
Training Data	Data used to build and teach an AI model how to perform a task
Dataset	A structured collection of data used for training, testing, or validation
Supervised Learning	Training a model using labelled data (input + correct answer)
Unsupervised Learning	Finding patterns in data without labelled outcomes
Reinforcement Learning	Learning through trial and error using rewards and penalties
Feature	An individual measurable property used as input (e.g., age, location)
Label	The correct output associated with input data (used in supervised learning)
Generative AI & Language	
Generative AI (GenAI)	GenAI, also known as generative AI or GAI, is a subset of AI that can generate new content (such as text, images, videos, or other data) in response to prompts. Generative models learn patterns and structures from training data and then create new data with similar characteristics.
Large Language Model (LLM)	AI trained on vast amounts of text to understand and generate human-like language.
Prompt	The input or instruction given to an AI system.
Prompt Engineering	Crafting prompts to get better results from AI model.

Hallucinations (AI)	A phenomenon where an AI model generates or perceives patterns, objects, or meaning in data that do not actually exist. This has been observed across various AI models and continues to be prevalent today.
Model Development & Performance	
Training	Training – The process of teaching a model using data.
Inference	Using a trained model to make predictions.
Accuracy	How often the model is correct.
Precision & Recall	Metrics used to evaluate model performance, especially is classification tasks.
Overfitting	When a model performs well on training data but poorly on new data.
Underfitting	When a model is too simple to capture patterns in the data
Risks, Ethics & Governance	
AI bias	Refers to the tendency of an AI system to produce results that are systematically prejudiced due to erroneous assumptions in the machine learning process, biased training or design.
Fairness	Ensuring AI outcomes are equitable and non-discriminatory, especially in public service delivery
Transparency	Obligation to explain how AI is used in decisions affecting citizens
Explainability (XAI)	Ability to explain AI outputs in understandable terms to users, regulators, or affected individuals
Accountability	Requirement that organisations can justify and take responsibility for AI-driven decisions aligned with UK GDPR
Model drift	When an AI system's performance declines over time due to changing data.
Security & Misuse	
Adversarial Attack	Attempts to trick AI models using manipulated inputs.
Data Poisoning	Introducing malicious data into training to compromise a model
Deepfake	AI-generated media that convincingly imitates real people.
Implementation & Usage	
API (Application Programming Interface)	A way for systems to interact with AI services.
Automation	Using AI to perform tasks without human intervention.

Human-in-the-Loop (HITL)	A requirement that humans review or intervene in AI decisions, especially for high-risk or legal impacts.
Use Case	A use case is a methodology used in system analysis to identify, clarify and organise system requirements. The method creates a document that describes all the steps taken by a user to complete an activity and is often used in fraud detection, chatbots etc.
Other Associated Terms	
Algorithmic Transparency	Refers to the ability to understand and explain how an AI system makes decisions. This includes being able to audit the processes and decisions of AI models to ensure they align with ethical standards and regulatory requirements. In health and social care, transparency is crucial to maintain trust and ensure that AI-driven decisions can be justified.
AI Ethics	AI ethics encompasses the moral implications of AI technologies, including fairness, transparency, accountability, and non-discrimination. Ethical AI ensures that AI technologies are designed and implemented in ways that protect human rights, promote fairness, and avoid harm, especially in sensitive sectors like social care.
AI Lifecycle	The AI lifecycle refers to the stages an AI system goes through from conception to deployment and ongoing monitoring. It includes data collection, model development, training, deployment, performance evaluation, and continuous improvement.
Bias Mitigation	Bias mitigation refers to the processes and techniques used to reduce bias in AI models, including strategies such as data pre-processing, algorithmic adjustments, and post-processing methods that aim to minimise unfairness in AI outcomes.
Automated Decision Making	Automated decision making refers to decisions made by AI systems without human intervention. These decisions are typically based on data inputs and pre-set algorithms. In healthcare, automated decision making can assist in diagnostics, treatment recommendations, or administrative tasks, but must be overseen to ensure compliance with legal and ethical standards.
Data Governance	Data governance refers to the policies and processes in place to manage the availability, usability, integrity, and security of data within an organisation. Effective data governance ensures that data used for AI

	systems is accurate, secure, and used responsibly in compliance with data protection such as GDPR.
Clinical Risk Management	Clinical risk management involves identifying, assessing, and mitigating potential risks to service user safety that could arise from the use of AI systems. In the context of AI, this means ensuring that the deployment of AI tools does not introduce errors or safety hazards into clinical workflows or decision-making.
AI governance	Frameworks, policies, and processes used to ensure AI is used lawfully, ethically, and safely.
Data Protection Impact Assessment (DPIA)	A mandatory risk assessment when AI processing is likely to result in high risk to individuals (i.e., automated decisions)
UK GDPR (General Data Protection Regulation)	The primary UK law governing personal data use, which applies to AI systems handling personal data
Information Commissioner's Office (ICO)	The UK regulator providing guidance on AI, data protection and compliance
Automated Decision-Making (ADM)	Decisions made without meaningful human involvement, subject to strict rules under UK GDPR
Lawful Basis	The legal justification for processing personal data
Algorithmic Bias	When AI produces unfair or discriminatory outcomes (i.e., against protected groups under the Equality Act 2010)
Equality Impact Assessment (EqIA)	Assessment used by public sector bodies to ensure AI systems do not discriminate
Personal Data	Any information relating to identifiable individual (central to AI compliance in the UK)
Special Category Data	Sensitive data (i.e., health, ethnicity) requiring additional protections under UK GDPR
Information Asset Register (IAR)	A record of information assets, including datasets or information used in AI systems
Data minimisation	Principle requiring only the minimum necessary data to be used
Data Quality	Ensuring data used in AI is accurate, relevant and up to date
AI Assurance	Activities that ensure AI systems are safe, compliant, and trustworthy (i.e., audits, testing)
Data Protection Impact Assessments (DPIA)	Is a process to help identify and minimise the data protection risks of a project. DPIAs must be considered and, where necessary, completed in full for any new data processing activities.
Digital Technology Assessment Criteria (DTAC)	Is a framework designed to evaluate digital health technologies in health and social care settings. The DTAC provides a standardised approach assessing

	the safety, effectiveness, and usability of digital health products
Natural Language Processing (NLP)	Is a branch of computer science/AI that gives computers the ability to understand, interpret, and generate human language in both written and spoken forms.
Processing	In relation to information or data means; obtaining, recording or holding the information or data or carrying out any operation or set of operations on the information or data, which may include adaptation or alteration of the information; retrieval, or use of the information or data; disclosure of the information or data by transmission, combination, blocking, erasure or destruction of the information or data. In summary anything you do with data is “processing”.
Robotic Process Automation (RPA)	Is the form of business process automation technology that uses software robots to automate tasks typically performed by humans. These software robots, also known as Bots, follow predefined workflows to perform repetitive office tasks such as extracting data, filling in forms, and moving files. RPA combines Application Programming Interface (APIs) and user interface (UI) interactions to integrate and execute tasks across different enterprise and productivity applications.
Shadow AI	Refers to the use of AI systems and tools within an organisation without explicit approval or oversight from the IT or data governance teams. This often involves colleagues using generative AI tools like ChatGPT or other AI applications to perform tasks such as drafting documents, creating images, or creating code, without the knowledge of consent of your IT department (if you have one).