Artificial Intelligence (AI)

Tutorial for Beginners — Complete Course Notes

SECTION 1: Introduction — What is AI and Why Learn It?

1.1 Course Overview

This introductory course is designed for complete beginners who want to understand what Artificial Intelligence is, how it works at a conceptual level, and how it is reshaping society, industry, and the job market. No prior technical knowledge is required — the course does not cover algorithms or mathematical implementation.

- Define what Artificial Intelligence (AI) is
- Explore the types and subfields of AI
- Review significant AI milestones, breakthroughs, and real-world applications
- Understand how AI is driving the Fourth Industrial Revolution
- Examine how AI will change the job market, economy, and society
- Discuss the risks, ethics, and future outlook of AI

1.2 Why Learn About AI?

AI is no longer a futuristic concept — it is embedded in everyday life and is rapidly transforming every sector of the economy. Understanding AI is becoming an essential skill for everyone, regardless of profession.

- AI is on virtually every smartphone: Google Assistant, Siri, Bixby, and other personal assistants
- Smart speakers (Amazon Echo, Google Home) and IoT devices rely heavily on AI
- Internet of Things (IoT): fridges, washing machines, TVs, and wearables are increasingly AI-powered — capable of detecting your habits and acting autonomously (e.g. ordering groceries automatically)
- AI is creating enormous demand for skilled professionals — from engineers and data scientists to ethicists and policy-makers
- Even non-technical roles (HR, law, finance, healthcare) are being reshaped by AI tools
- Understanding AI helps you prepare yourself — and guide the next generation — for what is coming

1.3 Key AI Application Domains

Domain	Examples
Personal Assistants	Siri, Google Assistant, Alexa — voice-controlled AI on phones and smart speakers
Internet of Things (IoT)	Smart fridges, wearables, thermostats — devices that collect data and act intelligently
Autonomous Vehicles	Self-driving cars (Tesla, Waymo, Toyota) using cameras, LiDAR, and radar sensors
Language & Communication	Chatbots, virtual agents, conversational AI on websites and apps
Security & Cybersecurity	Perimeter breach detection, malware intrusion detection, threat intelligence, incident response
Healthcare	AI-assisted diagnostics, drug discovery cost reduction, medical imaging analysis
Finance	Portfolio management, stock price prediction, fraud detection, algorithmic trading
Games	AlphaGo (board game), AlphaStar (StarCraft II), Atari game agents — AI beating world champions
Smart Cities	Traffic optimisation, public safety cameras, infrastructure monitoring, crime reduction
Robotics	Pepper robot (healthcare), industrial automation, warehouse logistics

SECTION 2: History of Artificial Intelligence

2.1 Ancient Origins — Myths and Early Automata

The human fascination with intelligent machines dates back thousands of years. Early stories and myths reflect a desire to create life-like, autonomous beings — often human-shaped (anthropomorphic).

- Talos (c. 400 BC, Ancient Greece) — a mythical bronze giant said to circle Crete, protecting the island from invaders
- King Mu of Zhou, China (c. 900 BC) — legend of an artificer who built a lifelike, singing automaton so realistic that the king ordered it destroyed
- Golem (1580s, Prague) — Jewish folklore of a creature made from clay and brought to life by a rabbi to protect the Jewish quarter
- The word 'Robot' — coined by Czech playwright Karel Čapek in his early 1920s play; derived from the Czech/Slovak word meaning 'forced labour' or 'drudgery'

2.2 The Birth of Computing — Alan Turing

- Alan Turing (1936) — presented the concept of a Universal Turing Machine: a theoretical device that could perform any computation expressible as a series of symbols
- Proved that using just two symbols (0 and 1), all computation is possible — the foundational concept of modern binary computing
- Machines capable of this are called 'Turing Complete'
- Turing Test — Turing proposed a test for machine intelligence: a human interacts via text with both a machine and another human without knowing which is which. If the human cannot reliably tell them apart, the machine passes the test.
- Early computers (late 1930s–1940s) were used primarily for: cryptography (decoding enemy messages in WWII), calculating artillery and torpedo trajectories

2.3 The Dartmouth Conference — The Birth of AI (1956)

- Organised by Marvin Minsky and colleagues at Dartmouth College, 1956
- Widely considered the founding event of Artificial Intelligence as a formal academic discipline
- Researchers gathered to discuss whether machines could simulate every aspect of human intelligence
- Early successes from this era include:
  - ELIZA Chatbot — a simulated psychotherapist that recognised keywords in user input and responded with relevant questions. It could sustain a conversation convincingly for short periods.
  - Semantic Networks — knowledge representation systems that allow reasoning via indirect connections (e.g. 'bear' → 'mammal' → 'animal')
- DARPA funded significant AI research projects during this initial wave of optimism

2.4 The AI Winters — Periods of Disillusionment

AI research went through two major funding crises known as 'AI Winters', caused by the technology failing to deliver on its ambitious promises.

Period	Key Events
First AI Winter (1974–1980)	ALPAC Report (1966) — found that machine translation was too slow, inaccurate, and expensive; recommended defunding NLP research. Lighthill Report (1973, UK) — stated AI had failed to meet its promises due to a 'combinatorial explosion' — problems grew too complex for the computers of the time. Funding withdrew significantly.
Brief Revival (1980s)	Expert Systems — software encoding a domain expert's knowledge as rules (IF-THEN logic). Allowed computers to answer complex questions in specialised domains. Excitement returned briefly.
Second AI Winter (late 1980s–early 1990s)	Expert systems had fundamental limitations — too rigid, expensive to maintain. The Lisp machine market collapsed. DARPA lost enthusiasm again. Funding dropped.
Slow & Steady Progress (1994–2010)	Intelligent agents, statistical AI methods, incremental advances. Less public excitement but steady academic and industrial research.
Deep Learning Revolution (2012–present)	Breakthrough in neural networks enabled by large datasets and GPU computing power. Rapid advances in image recognition, NLP, game-playing AI, and more.

2.5 AI Timeline Summary

Era	Key Developments
1936	Turing Universal Machine — theoretical foundation of computing
1940s	First electronic computers — used in WWII for cryptography and ballistics
1956	Dartmouth Conference — AI born as a field; ELIZA, Semantic Networks
1966	ALPAC Report — triggers first AI Winter funding cuts
1973	Lighthill Report (UK) — deepens first AI Winter
1980s	Expert Systems revival — second wave of excitement
Late 1980s	Second AI Winter — Lisp machine market crash, DARPA withdraws
1997	IBM Deep Blue defeats world chess champion Garry Kasparov
2011	IBM Watson wins Jeopardy! against human champions
2012	Deep learning breakthrough — AlexNet wins ImageNet competition
2016	DeepMind's AlphaGo defeats world Go champion Lee Sedol 4–1
2019	AlphaStar defeats professional StarCraft II players
2020s	Large Language Models (GPT, Gemini), generative AI, widespread AI adoption

SECTION 3: Types and Subfields of Artificial Intelligence

3.1 What is Artificial Intelligence? — Formal Definition

Artificial Intelligence is intelligence presented by a machine. More precisely, it refers to a flexible agent that interacts with its environment and performs actions to maximize its chances of achieving a defined goal.

AI research fields map directly onto human capabilities — for every cognitive or sensory ability humans have; there is a corresponding AI research discipline:

Human Capability	Corresponding AI Field
Seeing / Vision	Computer Vision — recognizing objects, faces, scenes in images and video. Was historically the most advanced AI field and drove many breakthroughs adopted by other areas.
Hearing / Sound Processing	Audio Processing / Sound Recognition — identifying speech, music, environmental sounds
Talking / Singing	Sound Synthesis / Speech Synthesis — generating human-like speech and voice
Understanding Language	Natural Language Processing (NLP) — understanding, generating, and translating human language. One of the most rapidly advancing fields today.
Movement	Robotics — enabling machines to navigate and interact physically with the world
Reasoning / Thinking	Automated Reasoning — logic-based systems that draw conclusions from given information
Learning	Machine Learning — enabling machines to improve performance on tasks through experience and data, without being explicitly programmed for each task

3.2 Narrow AI vs. General AI vs. Superintelligence

Type	Description & Status
Narrow AI (ANI)	AI designed to perform one specific task exceptionally well (e.g. playing chess, recognizing faces, recommending movies). All current AI systems are Narrow AI. They cannot transfer skills between domains.
Artificial General Intelligence (AGI)	Hypothetical AI with broad human-like reasoning — able to learn, adapt, and apply intelligence across any domain, just as a human can. Does not yet exist. The course notes this is far from being achieved.
Artificial Superintelligence (ASI)	Hypothetical AI that surpasses human intelligence in every domain. A theoretical concept — the subject of much ethical debate and speculation. Far beyond current capabilities.

3.3 AI in Daily Life — Current Examples

- Search engines — Google ranks and retrieves relevant web pages using AI
- Recommendation engines — Netflix, Amazon, Spotify suggest content based on your behaviour
- Virtual assistants — Alexa, Google Home, Siri respond to spoken commands
- Fraud detection — banks use anomaly detection to flag unusual card transactions
- Medical imaging — AI assists radiologists in detecting tumours and anomalies in scans
- Self-driving cars — Tesla, Waymo use computer vision and sensor fusion
- Chatbots — customer service bots on websites handle queries without human agents
- Social media — AI curates your feed, detects harmful content, and targets ads

SECTION 4: Machine Learning — The Core Engine of Modern AI

4.1 What is Machine Learning?

Machine Learning (ML) is a subfield of computer science and AI that explores how machines can learn to perform tasks without being explicitly programmed for each specific task. Instead of writing rigid rules, you supply data — and the algorithm learns patterns from that data to make predictions or decisions on new, unseen inputs.

- Traditional programming: human writes rules → computer follows rules → output
- Machine learning: human provides data + desired output → algorithm learns rules → model applied to new data
- Machine learning is the foundational layer beneath most modern AI applications — advances in computer vision, NLP, and robotics in recent years have primarily been driven by ML techniques

4.2 The Three Main Types of Machine Learning

Type 1 — Supervised Learning:

- The algorithm is trained on labelled data — each input example has a known, correct output
- The machine learns to map inputs to outputs based on examples
- Task-driven: you tell the machine exactly what you want it to predict or classify
- Examples:
  - Sentiment analysis — classify a text message as positive, negative, or neutral
  - Image classification — detect whether a photo contains a cat, car, or cupcake
  - Spam detection — classify emails as spam or not spam
  - Medical diagnosis — classify whether a scan shows signs of disease
- Requires: labelled training data (the bigger and cleaner, the better the model)

Type 2 — Unsupervised Learning:

- No labels are provided — the algorithm explores raw data to find hidden structure
- The machine figures out patterns and groupings independently
- Main task: Clustering — grouping similar data points together
- Examples:
  - Grouping customers by purchasing behaviour without pre-defining the groups
  - Identifying natural groupings in a dataset of characters (e.g. by shape, colour) without being told the categories
  - Document topic modelling — discovering themes across a large collection of texts
- Challenge: the algorithm may cluster data in unexpected ways depending on what features it emphasises (e.g. Simpson characters grouped by hair colour vs. family vs. job role — all are valid clusters)
- Useful when you do not know what patterns to look for in advance

Type 3 — Reinforcement Learning:

- An agent learns by interacting with an environment and receiving feedback (rewards or punishments)
- The agent starts with no knowledge — it takes random actions, observes outcomes, and gradually learns which actions lead to better rewards
- Learning from mistakes: more rewards for good actions, penalties for bad actions
- Examples:
    - AlphaGo and AlphaStar — learned to play board games and video games at superhuman levels by playing millions of games against itself
    - Robotics — a robot arm learns to pick and place objects through trial and error
    - Autonomous vehicles — a car agent learns to navigate environments in simulation
    - Game agents — Atari game-playing agents that learned to play dozens of games from raw pixel input
- Key mechanism: reward system guides the agent toward optimal behaviour over time

Type 4 — Semi-Supervised Learning:

- A hybrid approach used when you have a small amount of labelled data and a large amount of unlabelled data
- The model is first trained on the labelled data, then used to generate pseudo-labels for the unlabelled data, which are then used to further train the model
- Useful in practice because labelling data is expensive and time-consuming — fully labelled datasets are rare

4.3 Types of Machine Learning — Comparison Table

Type	Data Required / How It Learns / Main Tasks
Supervised	Labelled data required. Learns input-to-output mapping. Main tasks: classification, regression, prediction.
Unsupervised	Unlabelled data only. Finds patterns and groups independently. Main task: clustering, dimensionality reduction.
Reinforcement	No dataset — learns from environment interaction. Rewards and punishments guide learning. Main tasks: game playing, robotics, navigation.
Semi-Supervised	Small labelled + large unlabelled dataset. Combines both approaches. Used when full labelling is impractical.

SECTION 5: The Machine Learning Project Workflow

5.1 Overview

Building a machine learning solution follows a structured, iterative process. The steps below apply to most ML projects — from simple classifiers to complex deep learning systems.

5.2 Step-by-Step ML Workflow

- Step 1 — Define and Refine the Business Problem: Identify the specific problem you want to solve. Be precise — a vague problem leads to a model that solves the wrong thing. Can revisit this step if results are poor.
- Step 2 — Prepare the Data:
  - Ingest the data from your sources (databases, APIs, files, sensors)
  - Label the data if building a supervised model
  - Collect raw unlabelled data if building an unsupervised model
  - Build a simulated environment with a reward system if building a reinforcement learning agent
- Step 3 — Data Cleaning and Transformation (the most time-consuming step — ~80% of total project time):
  - Remove duplicates, fix errors, handle missing values
  - Normalise and scale numeric features
  - Convert data into the format required by the chosen ML algorithm
  - Data scientists spend approximately 80% of their time here and only 20% on model building
- Step 4 — Feature Selection:
  - Identify which input variables (features) are most relevant to the prediction task
  - Exclude irrelevant or redundant features that add noise without adding predictive power
  - In deep learning, neural networks can sometimes perform internal feature selection automatically — but explicit feature analysis is still recommended
- Step 5 — Create the Machine Learning Model:
  - Select an appropriate algorithm (e.g. decision tree, neural network, support vector machine)
  - Train the model on your prepared data
- Step 6 — Test the Model:
  - Evaluate performance on unseen test data (data not used during training)
  - Measure accuracy, precision, recall, F1-score, RMSE, or other relevant metrics
  - If performance is unsatisfactory, loop back to earlier steps
- Step 7 — Deploy the Model:
  - Integrate the trained model into a production system or application
  - The model is now making predictions on real-world data
- Step 8 — Monitor and Iterate:
  - Track model performance in production over time
  - Retrain the model as new data arrives and the world changes
  - Loop back to refine the business problem if the model is not solving the right thing

5.3 The Iterative Nature of ML Projects

The ML workflow is not linear — it is cyclical. If testing reveals poor performance, you return to earlier steps: more data collection, better cleaning, different feature selection, or a different algorithm. If deployment reveals the model is solving the wrong problem, you return to Step 1.

SECTION 6: Machine Learning Tasks and Their Applications

6.1 Overview

Machine learning can be applied to solve many different types of problems. Each 'task' describes a category of problem that ML algorithms are designed to solve. Choosing the right task type is crucial before building any model.

6.2 Classification

Classification assigns an input to one of a set of predefined categories (classes). It is the most common supervised learning task.

- How it works: the model is trained on labelled examples and learns to assign a class label to new, unseen inputs
- Examples:
  - Sentiment analysis — classify SMS/text as positive, negative, or neutral
  - Email spam detection — classify as spam or not spam
  - Image classification — identify whether a photo contains a cat, cupcake, or apple
  - Medical diagnosis — classify whether a scan shows disease markers
  - Document classification — categorise a document by topic

6.3 Sequence Modelling

Sequence modelling processes data where context and order matter — such as natural language text or time-series data. The model assigns a label to each element of a sequence, taking into account what came before and after.

- Examples:
  - Named Entity Recognition (NER) — in the sentence 'The United Nations official visited Baghdad', the model classifies each word: 'United Nations' = Organisation, 'official' = Title, 'Baghdad' = Location
  - Why context matters: 'New York City' is a location, but 'the city' in a different sentence is not — the surrounding words determine the classification
  - Time-series prediction — classifying states in a sequence of sensor readings over time
- Algorithms used: Conditional Random Fields (CRF), Recurrent Neural Networks (RNN), Hidden Markov Models (HMM)

6.4 Regression

Regression predicts a continuous numeric value rather than a category. The model learns a mathematical function that best fits the relationship between inputs and outputs.

- Examples:
  - Predicting body mass index (BMI) from percentage of body fat
  - Predicting apartment prices based on location, size, and age
  - Forecasting stock prices or sales revenue
- Output: a number on a continuous scale (not a discrete class label)
- The model draws a 'line of best fit' through training data and uses it to estimate values for new inputs

6.5 Clustering

Clustering is an unsupervised task — the algorithm groups data points into clusters based on similarity, without being told what the groups should be.

- No labels are provided — the algorithm discovers natural groupings
- Challenge: there may be multiple valid ways to cluster the same data. For example, Simpsons characters could be grouped by family, by gender, or by job — all are equally valid clusterings. The algorithm may not pick the one you expect.
- Applications:
  - Customer segmentation — grouping customers by purchasing behaviour
  - Document grouping — clustering articles by topic without predefined categories
  - Gene expression analysis — finding groups of genes with similar behaviour
- You can provide additional constraints or hints to steer the clustering process

6.6 Anomaly Detection

Anomaly detection identifies data points that deviate significantly from normal behaviour — flagging them as potential problems or fraud.

- Works by learning what 'normal' looks like from historical data, then flagging outliers
- Applications:
  - Credit card fraud detection — a transaction that differs dramatically from a cardholder's usual pattern is flagged
  - Cybersecurity — detecting unusual system behaviour that may indicate malware
  - Industrial monitoring — detecting equipment faults in power plants or factories before failure
  - Medical monitoring — detecting abnormal vital signs in patient data
- Often applied to time-series data where sudden spikes or drops indicate anomalies

6.7 Ranking

Ranking orders a set of items by relevance or importance relative to a query or context.

- Most famous application: Google Search — pages are ranked by relevance to the search query (using text matching AND the number/quality of incoming links)
- Also used for: disambiguating word meanings by ranking candidate interpretations in context
- E-commerce: ranking search results by predicted likelihood of purchase

6.8 Recommendation

Recommendation systems predict what a user will want next based on their past behaviour and the behaviour of similar users.

- How it works: the system finds users with similar purchase/viewing history and recommends items those users also liked — this is called collaborative filtering
- Applications:
  - Amazon — 'Customers who bought this also bought...'
  - Netflix / Disney+ — 'You might also enjoy...'
  - Spotify — personalised playlists and song recommendations
  - YouTube — suggested videos
- Recommendation engines are one of the highest-value ML applications commercially — they directly drive revenue

6.9 Agent Movement / Reinforcement Learning Applications

In RL applications, an agent learns to navigate an environment by maximising cumulative rewards over time.

- Game agents — start with random moves, learn optimal strategies through millions of play iterations
- Robot navigation — a robot learns to move through rooms, avoid obstacles, and reach targets
- Autonomous driving — simulation-based RL trains vehicles to handle complex traffic scenarios
- Mechanism: random exploration → observe outcome → receive reward/penalty → update strategy → repeat

6.10 ML Tasks — Quick Reference Table

Task	Type / Description / Example
Classification	Supervised — assign input to a category. E.g. spam detection, image recognition, sentiment analysis.
Sequence Modelling	Supervised — classify each item in a sequence using context. E.g. Named Entity Recognition in text.
Regression	Supervised — predict a continuous number. E.g. house prices, stock values, BMI prediction.
Clustering	Unsupervised — group similar data without labels. E.g. customer segmentation.
Anomaly Detection	Supervised/Unsupervised — flag unusual data points. E.g. fraud detection, malware detection.
Ranking	Supervised — order items by relevance. E.g. Google search results ranking.
Recommendation	Collaborative Filtering — suggest items based on user behaviour. E.g. Netflix, Amazon, Spotify.
Agent Movement	Reinforcement — navigate environments via reward/punishment. E.g. AlphaGo, robot navigation.

SECTION 7: Industrial Revolutions and AI's Role in Society

7.1 The Four Industrial Revolutions

Revolution	Period / Key Technologies / Impact
First Industrial Revolution	1780s — Steam power, mechanisation, the loom. Enabled mass textile production and railways. Moved production from homes to factories. Triggered the birth of capitalism and communism as competing economic ideologies.
Second Industrial Revolution	1870s — Mass production, assembly lines, electrical energy. Henry Ford's automobile factory. Enabled rapid scaling of manufactured goods.
Third Industrial Revolution	1970s — Computers, electronics, automated assembly lines. Brought digital technology into manufacturing and daily life.
Fourth Industrial Revolution (Now)	2000s–present — Cyber-physical systems, Internet of Things (IoT), AI and machine learning. Blurs boundaries between physical, digital, and biological worlds. Characterised by unprecedented speed of change.

7.2 AI as a Driver of the Fourth Industrial Revolution

- AI is the primary engine of the Fourth Industrial Revolution — not just automating tasks, but transforming entire industries simultaneously
- Unlike previous revolutions (which took decades to unfold), AI-driven change is happening in years — faster than humans can retrain for new roles
- Einstein anticipated this: he observed that automation would make the industrial methods of his era look as primitive as the Stone Age
- The pace of technological change is accelerating — looking 50 years back now (1975) vs. looking 50 years forward — the contrast will be even more dramatic

7.3 Economic Growth and Wealth Creation

- Global GDP was nearly flat for most of human history — people lived near subsistence level
- The Industrial Revolution triggered exponential economic growth starting around 1870
- By 2020, global GDP exceeded $60 trillion and continues growing
- We live in the wealthiest period in human history — yet inequality is simultaneously at record highs
- AI is accelerating wealth creation, but primarily for those who own or work with the technology

7.4 The Capitalism vs. Socialism Debate in the Age of AI

- The First Industrial Revolution created a political debate still ongoing today: how should the wealth generated by machines be distributed?
- Capitalism (Adam Smith) — individuals and companies drive innovation for profit; wealth flows to owners; high innovation incentive but risk of inequality
- Socialism / Communism (Marx & Engels) — wealth is redistributed to society; reduces inequality but potentially reduces innovation incentive
- AI intensifies this debate: tech companies (Amazon, Google, Meta) are among the world's wealthiest, while workers in automated industries face job displacement

SECTION 8: AI and the Future of Work

8.1 Historical Pattern: Jobs Lost and Jobs Created

Every Industrial Revolution has destroyed some jobs and created new ones. The net employment impact across history has been roughly neutral — unemployment rates remained broadly stable despite exponential population growth and repeated waves of automation.

Jobs That Have Disappeared	Jobs That Were Created
Knocker-uppers (waking people up for work)	Software Engineers
Ice cutters (cutting and delivering ice blocks)	Medical device technicians
Telephone operators (manually connecting calls)	Social media managers
Lamplighters (lighting street gas lamps)	Pilots and aircraft crew
Manual cashiers	Camera operators / videographers
Pin setters in bowling alleys (children)	Auto mechanics

8.2 Jobs at Risk from AI Automation

- Accountants performing simple/routine accounting tasks
- Some legal niches — routine document review, contract analysis
- Human drivers — as self-driving vehicles become more capable
- Cashiers — automated self-checkout already widely deployed
- Financial analysts performing quantitative number-crunching
- Military roles — increasing use of drones and autonomous weapon systems
- Data entry and administrative roles — high repetition, rule-based tasks

Note: Jobs involving complex human judgment, creativity, empathy, and interpersonal skills are much harder to automate and are lower risk in the near term.

8.3 New Jobs AI Will Create

Just as previous revolutions created entirely new professions that did not exist before, AI is expected to generate new job categories. Many of these roles do not yet exist as formal job titles.

- Robot Psychologist / AI Troubleshooter — diagnosing why robots or AI systems behave unexpectedly
- AI Ethics Officer — ensuring AI systems are fair, transparent, and safe
- Data Scientist / ML Engineer — building and maintaining AI models
- Augmented Reality Architect — designing experiences in mixed physical-digital environments
- Driverless Ride Experience Designer — curating the experience inside autonomous vehicles
- 3D Food Printing Chef / Molecular Gastronomist — food production using advanced technology
- Traffic Flow Optimiser — using AI tools to manage smart city infrastructure
- AI Trainer — labelling data, fine-tuning models, and teaching AI systems

8.4 The Two Theories of AI's Economic Impact

Theory	Description
Optimistic View (Hawking's first scenario)	If the wealth generated by AI-powered machines is widely shared across society, everyone can enjoy a life of greater leisure and prosperity. Technology raises the floor for all.
Pessimistic View (Hawking's second scenario)	If machine owners use their influence to prevent redistribution of AI-generated wealth, most people become poorer while a small elite grows richer. Current trend appears closer to this trajectory — income inequality is rising.

8.5 Universal Basic Income (UBI)

- UBI is a proposed policy response to automation-driven job displacement
- The government provides every citizen with a regular unconditional cash payment, regardless of employment status
- Intended to: support people whose jobs have been automated before they can retrain; reduce poverty; give people the economic security to pursue education or retraining
- Debate: how to fund UBI? Via taxation of AI-generated profits? Requires political will and economic restructuring.
- Several countries and cities have piloted limited UBI schemes (Finland, Kenya, Stockton CA)

SECTION 9: AI Risks, Ethics, and Future Outlook

9.1 Will AI Destroy Humanity? — The Terminator Scenario

The popular cultural fear — AI becoming conscious, identifying humans as a threat, and deciding to eliminate them — is a recurring theme in science fiction. How realistic is it?

- The scenario requires Artificial General Intelligence (AGI) — a machine that can reason autonomously, have goals of its own, and act strategically against humans
- We are very far from AGI — current AI systems are narrow, task-specific, and have no self-awareness or independent goals
- We do not yet understand how human consciousness works — without that understanding, we cannot replicate it in a machine
- Even if AGI were built, ethical safeguards and AI safety research are actively being developed to prevent misuse
- Alternative risk scenario: an AI optimising a narrow goal incorrectly — e.g. an AI programmed to 'reduce human suffering' might logically conclude that ending human life reduces suffering. This highlights the importance of careful goal specification.

Conclusion: The Terminator scenario is unlikely in the near term. The more immediate risk is not malevolent AI, but poorly specified goals or misuse of AI by humans.

9.2 Current and Near-Term AI Risks

- Bias and fairness — AI models trained on biased data reproduce and amplify those biases (e.g. racial bias in facial recognition, gender bias in hiring algorithms)
- Privacy — AI-powered surveillance (cameras, data collection, behavioural profiling) threatens individual privacy
- Deepfakes and misinformation — AI can generate convincing fake images, videos, and text, enabling disinformation at scale
- Economic inequality — AI wealth concentrating in a small number of tech companies and individuals
- Job displacement — short-term disruption for workers in automatable roles before new jobs emerge
- Autonomous weapons — AI-controlled drones and weapon systems create new ethical and legal challenges in warfare
- Algorithmic accountability — when an AI system makes a harmful decision, who is responsible?

9.3 AI Safety and Ethics Research

- Major tech companies (DeepMind, OpenAI, Anthropic) have dedicated AI safety teams
- Governments worldwide are developing AI regulatory frameworks
- Key principles being developed: transparency, explainability, fairness, accountability, privacy, safety
- The goal: ensure AI remains a tool that serves human values — not one that operates against them

9.4 Positive Applications of AI — AI as a Tool

- Entertainment — AI curates music, film, and game recommendations; AI composes music and generates art
- Productivity — AI automates repetitive tasks, summarises information, assists with writing and research
- Healthcare — AI assists in drug discovery, diagnostics, robotic surgery, and personalised medicine
- Accessibility — AI-powered prosthetics connected to nerve endings restore lost mobility; AI screen readers and voice interfaces help people with disabilities
- Finance — AI analyses market data faster and more accurately than humans; detects fraud; optimises portfolios
- Education — AI tutors, adaptive learning platforms, and content personalisation
- Climate and science — AI accelerates research in climate modelling, materials science, and protein folding (AlphaFold)

9.5 AI and Finance — Specific Applications

- Financial analysis requires processing: company financial data, fundamental data, news, and sentiment — all tasks AI excels at
- AI can read, analyse, and infer from publicly available financial data at a speed and scale no human can match
- Machine learning algorithms detect patterns in market data and make predictions for investment decisions
- Applications: algorithmic trading, fraud detection, credit scoring, risk assessment, portfolio optimisation
- The pattern recognition capabilities of ML are particularly valuable in financial markets where signals are often hidden in vast amounts of noisy data

SECTION 10: Key Terms & Quick Reference

10.1 Essential AI Glossary

Term	Definition
Artificial Intelligence (AI)	Intelligence exhibited by machines; systems that act to maximise a goal
Machine Learning (ML)	A subfield of AI where algorithms learn from data without being explicitly programmed
Deep Learning	A subset of ML using multi-layered neural networks; drives most modern AI breakthroughs
Neural Network	A computational model loosely inspired by the human brain; layers of interconnected nodes
Supervised Learning	ML with labelled training data; learns to map inputs to known outputs
Unsupervised Learning	ML without labels; finds hidden patterns and clusters in raw data
Reinforcement Learning	ML via rewards and punishments in an environment; learns optimal actions over time
Feature	An input variable used by an ML model to make predictions
Model	The output of a trained ML algorithm; a mathematical function that makes predictions
Training Data	The dataset used to teach an ML model
Overfitting	When a model learns training data too precisely and performs poorly on new data
Natural Language Processing (NLP)	AI field enabling machines to understand and generate human language
Computer Vision	AI field enabling machines to interpret and analyse visual data (images, video)
Turing Test	A test of machine intelligence: can a human distinguish AI from human in conversation?
AGI	Artificial General Intelligence — hypothetical human-level AI across all domains. Does not yet exist.
IoT	Internet of Things — network of physical devices connected to the internet, often with embedded AI
AlphaGo	DeepMind's AI that defeated the world Go champion in 2016 using reinforcement learning
ELIZA	One of the first chatbots (1960s); simulated a psychotherapist using keyword matching
Expert System	AI based on coded human expertise (IF-THEN rules); popular in the 1980s
AI Winter	Periods of reduced funding and interest in AI research (1974–80 and late 1980s–early 90s)

— End of Course Notes —