ARTIFICIAL INTELLIGENCE FOUNDATIONS

Learning from experience

Andrew Lowe and Steve Lawless
BCS, THE CHARTERED INSTITUTE FOR IT

BCS, The Chartered Institute for IT, is committed to making IT good for society. We use the power of our network to bring about positive, tangible change. We champion the global IT profession and the interests of individuals, engaged in that profession, for the benefit of all.

Exchanging IT expertise and knowledge
The Institute fosters links between experts from industry, academia and business to promote new thinking, education and knowledge sharing.

Supporting practitioners
Through continuing professional development and a series of respected IT qualifications, the Institute seeks to promote professional practice tuned to the demands of business. It provides practical support and information services to its members and volunteer communities around the world.

Setting standards and frameworks
The Institute collaborates with government, industry and relevant bodies to establish good working practices, codes of conduct, skills frameworks and common standards. It also offers a range of consultancy services to employers to help them adopt best practice.

Become a member
Over 70,000 people including students, teachers, professionals and practitioners enjoy the benefits of BCS membership. These include access to an international community, invitations to a roster of local and national events, career development tools and a quarterly thought-leadership magazine. Visit www.bcs.org/membership to find out more.

Further information
BCS, The Chartered Institute for IT,
3 Newbridge Square,
Swindon, SN1 1BY, United Kingdom.
T +44 (0) 1793 417 417
(Monday to Friday, 09:00 to 17:00 UK time)
www.bcs.org/contact
http://shop.bcs.org/
ARTIFICIAL INTELLIGENCE FOUNDATIONS
Learning from experience

Andrew Lowe and Steve Lawless
# CONTENTS

Figures and table vii  
Authors ix  
Acknowledgements x  
Abbreviations xi  
Glossary xiii  
Useful websites xix  
Preface xxi

1. **INTRODUCTION: ETHICAL AND SUSTAINABLE HUMAN AND ARTIFICIAL INTELLIGENCE** 1  
1.1 The general definition of human intelligence 1  
1.2 Definition of artificial intelligence 3  
1.3 A brief history of artificial intelligence 4  
1.4 Sustainable AI 23  
1.5 Machine learning – a significant contribution to the growth of artificial intelligence 24  
1.6 Summary 25

2. **ARTIFICIAL INTELLIGENCE AND ROBOTICS** 26  
2.1 Understanding the AI intelligent agent 26  
2.2 What is a robot? 31  
2.3 What is an intelligent robot? 36  
2.4 Summary 36

3. **APPLYING THE BENEFITS OF AI AND IDENTIFYING CHALLENGES AND RISKS** 37  
3.1 Sustainability – how our values will change humans, society and organisations 37  
3.2 The benefits of artificial intelligence 39  
3.3 The challenges of artificial intelligence 43  
3.4 Understanding of the risks AI brings to projects 44  
3.5 Opportunities for AI 46  
3.6 Funding AI projects – the NASA Technology Readiness Levels 47  
3.7 Summary 48

4. **STARTING AI: HOW TO BUILD A MACHINE LEARNING TOOLBOX** 49  
4.1 How we learn from data 49  
4.2 Neural networks 80  
4.3 Agents’ functionality 85
4.4 Using the cloud – cheap high-performance computing for AI 86
4.5 Summary 87

5. **ALGORITHMS** 88
5.1 What is an algorithm? 89
5.2 What is special about AI/ML algorithms? 90
5.3 What are self-learning algorithms? 90
5.4 What are the algorithms used in machine learning and artificial intelligence? 90
5.5 What is a deep learning algorithm? 93
5.6 The ethical use of algorithms 94
5.7 Summary 95

6. **THE MANAGEMENT, ROLES AND RESPONSIBILITIES OF HUMANS AND MACHINES** 96
6.1 Artificial intelligence will drive humans and machines to work together 96
6.2 Future directions of humans and machines working together 99
6.3 ‘Learning from experience’ Agile approach to projects 99
6.4 Summary 101

7. **AI IN USE IN INDUSTRY: REIMAGINING EVERYTHING IN THE FOURTH INDUSTRIAL REVOLUTION** 102
7.1 AI already in use 102
7.2 Summary 118

8. **AI CASE STUDIES** 119
8.1 Predicting the performance of a warehouse 119
8.2 The weather – concentration on data preparation 124
8.3 A Cambridge spin-out start-up 130
8.4 Summary 133

REFERENCES 134

FURTHER READING 140
Machine learning 140
Advanced theoretical text 140

Index 141
<table>
<thead>
<tr>
<th>FIGURE</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>Where AI sits compared to ML, NNs and deep learning</td>
</tr>
<tr>
<td>Figure 1.2</td>
<td>Dilts’ logical levels of change</td>
</tr>
<tr>
<td>Figure 2.1</td>
<td>The learning agent – based on the structure of agents in Russell and Norvig</td>
</tr>
<tr>
<td>Figure 2.2</td>
<td>The state of the world – the agent’s internal world that can make up for incomplete information or help to make decisions</td>
</tr>
<tr>
<td>Figure 2.3</td>
<td>Honda’s ASIMO</td>
</tr>
<tr>
<td>Figure 2.4</td>
<td>Hierarchical paradigm in robotic design</td>
</tr>
<tr>
<td>Figure 2.5</td>
<td>Reactive paradigm in robotic design</td>
</tr>
<tr>
<td>Figure 2.6</td>
<td>Hybrid deliberative/reactive paradigm in robotic design</td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>The three pillars of sustainability</td>
</tr>
<tr>
<td>Figure 4.1</td>
<td>A position vector, ( \mathbf{p} ), made up of three scalars, ( x, y ) and ( z )</td>
</tr>
<tr>
<td>Figure 4.2</td>
<td>Example problem for vector calculus</td>
</tr>
<tr>
<td>Figure 4.3</td>
<td>Complexity and randomness</td>
</tr>
<tr>
<td>Figure 4.4</td>
<td>Heuristics can simplify complexity and randomness</td>
</tr>
<tr>
<td>Figure 4.5</td>
<td>Venn diagrams for false (none of the space ( S ) is covered) and true (all of the space ( S ) is covered)</td>
</tr>
<tr>
<td>Figure 4.6</td>
<td>Venn diagrams for the probability of finding a letter A and the probability of finding a letter B; A and B are not mutually exclusive</td>
</tr>
<tr>
<td>Figure 4.7</td>
<td>Venn diagram where A and B are not mutually exclusive. Centre region is accounted for twice in the addition law of probability</td>
</tr>
<tr>
<td>Figure 4.8</td>
<td>Venn diagram defining the conditional probability of event A, given event B</td>
</tr>
<tr>
<td>Figure 4.9</td>
<td>Example of a probability mass function of rolling two dice</td>
</tr>
<tr>
<td>Figure 4.10</td>
<td>Example of a probability density function of continuous data</td>
</tr>
<tr>
<td>Figure 4.11</td>
<td>Uniform and Gaussian probability density functions of continuous data</td>
</tr>
<tr>
<td>Figure 4.12</td>
<td>An example of the central limit theorem using five dice and rolling set of scores</td>
</tr>
<tr>
<td>Figure 4.13</td>
<td>Supervised learning</td>
</tr>
<tr>
<td>Figure 4.14</td>
<td>Unsupervised learning</td>
</tr>
<tr>
<td>Figure 4.15</td>
<td>Semi-supervised learning</td>
</tr>
<tr>
<td>Figure 4.16</td>
<td>Reinforcement learning</td>
</tr>
<tr>
<td>Figure 4.17</td>
<td>A simple example of over-fitting</td>
</tr>
<tr>
<td>Figure 4.18</td>
<td>A simple example of under-fitting</td>
</tr>
</tbody>
</table>
Figures and Table

Figure 4.19  Algorithm assessment, the trade-off and finding the balance between under- and over-fitting 77
Figure 4.20  Visualisation is learning from experience at every stage of an ML project 77
Figure 4.21  Standard data visualisation examples 79
Figure 4.22  Open-source software to build simulated environments – Blender 80
Figure 4.23  AI in medicine – humans and machines working together 81
Figure 4.24  The schematic of a neural network neuron 83
Figure 4.25  The schematic of a neural network with an input, hidden and output layer 84
Figure 8.1  Store’s heat load: Vault One and Vault Two 123
Figure 8.2  Niagara Falls – the turbulence here dissipates enough energy to power an area the size of north Wales 126
Figure 8.3  The movement of the Earth’s oceans 127
Figure 8.4  The visualisation of a large data set from a simulation of turbulence that is found in the Earth’s atmosphere or a soap film 129
Figure 8.5  Optalysys’s first optical mathematical operation was the derivative; the first prototype was built on a 250kg optical bed and was over a metre in length 132
Figure 8.6  Optalysys’s first commercial optical AI processor – FT: X Optical co-processor system with PCIe drive electronics. It fits onto a standard PCI card about 100mm wide by 300mm long. The thumbnail, top left, is the next generation of AI processor and measures 32mm × 32mm 132

Table 5.1  Examples of the types of learning and algorithms 90
Andrew Lowe is an engineer specialising in using computers to solve the challenges engineers face. He started as a nuclear design apprentice when computers were replacing drawing boards. Learning from experience has been a fundamental part of his career, and artificial intelligence helps him in his day-to-day work – some of the problems he has worked on can only be solved with supercomputers. He obtained his PhD from Cambridge and has worked in academia and industry. He has also helped an AI start-up company gain traction. He is married with two daughters, and lives and works in the Lake District, a World Heritage Site. He volunteers with local organisations to give people better opportunities.

Steve Lawless has always had a keen interest in science and technology, and has worked in the computer industry for over 40 years. He has trained literally thousands of people to give them the skills they need to make computers work. He knows Andy from volunteering at the local ski club. He runs a successful training company who have a worldwide customer base. He has written more than 100 training courses on IT, and enjoys making technology easy to understand and accessible. He is married with three sons and two daughters, and also lives and works in the Lake District.
ACKNOWLEDGEMENTS

We would like to thank our families, who have been patient. It is amazing how much time it takes to pull together learning from experience in an easy-to-understand book.

The technical side of AI can be challenging, so the numerous teachers, lecturers, supervisors, tutors and mentors are too many to mention individually. Thank you; you know who you are. Especially to those who have made their lectures and notes available on the web so everyone can learn.

We would like to thank Dr Paul Mort, of Sellafield Sites Ltd, who has been a keen enthusiast and supporter of the BCS AI Essentials and AI Foundation courses. His challenging questions while reviewing the courses made us think very carefully. Paul is a roboticist, and we were highly aware that AI is a much more enriching and broader subject than machine learning. Indeed, we have all left the digital revolution behind in the third industrial revolution, and we are now well on our way to the fifth. The AI agent is a really important and fundamental concept that we can easily overlook if we think AI is a digital machine learning technology.

BCS have played a big role in bringing this book and the courses together. We would like to thank, in no particular order, Ann Winskill, Felicity Page, Ian Borthwick, Rebecca Youé, Chris Leadbeater, Becky Lemaire, Sharon Pillai, Blair Melsom, Helen Silverman, Natalie Rew, Jo Massiah, Pam Fegan and Suky Sunner. There have been many others involved, we are sure. Thank you.

Many thanks go to Debbie Patten and Nigel Crowdy, at Purple Griffon, who make course organisation look easy. People who have attended the course have given highly valued contributions to the course as we have developed it. Learning from experience is exactly what we do on the courses; we value your feedback, creative input and challenging questions.

We are sure you will find something useful in this book, and hope that above all you are sceptical, critical and rigorous in your study and learn from experience. We are human, we make mistakes, and would love to hear all feedback, good and bad. This is part of the scientific method and learning from experience.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABM</td>
<td>agent-based modelling</td>
</tr>
<tr>
<td>AGI</td>
<td>artificial general intelligence</td>
</tr>
<tr>
<td>AI</td>
<td>artificial intelligence</td>
</tr>
<tr>
<td>AloT</td>
<td>Artificial Intelligence of Things</td>
</tr>
<tr>
<td>ANN</td>
<td>artificial neural network</td>
</tr>
<tr>
<td>AR</td>
<td>augmented reality</td>
</tr>
<tr>
<td>ATEAC</td>
<td>Advanced Technology External Advisory Council Agency</td>
</tr>
<tr>
<td>CGI</td>
<td>computer-generated imagery</td>
</tr>
<tr>
<td>CNN</td>
<td>Convolutional Neural Network</td>
</tr>
<tr>
<td>CPU</td>
<td>central processing unit</td>
</tr>
<tr>
<td>DARPA</td>
<td>Defence Advanced Research Projects Agency</td>
</tr>
<tr>
<td>DBN</td>
<td>deep belief network</td>
</tr>
<tr>
<td>DNN</td>
<td>deep neural network</td>
</tr>
<tr>
<td>EU</td>
<td>European Union</td>
</tr>
<tr>
<td>GAN</td>
<td>generative adversarial network</td>
</tr>
<tr>
<td>GPU</td>
<td>graphics processing unit</td>
</tr>
<tr>
<td>IA</td>
<td>intelligent automation/augmentation</td>
</tr>
<tr>
<td>ICT</td>
<td>information and communications technology</td>
</tr>
<tr>
<td>IoT</td>
<td>Internet of Things</td>
</tr>
<tr>
<td>IT</td>
<td>information technology</td>
</tr>
<tr>
<td>k-NN</td>
<td>k-Nearest Neighbour</td>
</tr>
<tr>
<td>LSTM</td>
<td>long short-term memory</td>
</tr>
<tr>
<td>ML</td>
<td>machine learning</td>
</tr>
<tr>
<td>MLP NN</td>
<td>multilayer Perceptron neural network</td>
</tr>
<tr>
<td>NGO</td>
<td>non-governmental organisation</td>
</tr>
<tr>
<td>NI</td>
<td>natural intelligence</td>
</tr>
<tr>
<td>NLP</td>
<td>natural language processing</td>
</tr>
<tr>
<td>NN</td>
<td>neural network</td>
</tr>
<tr>
<td>NPC</td>
<td>non-player character</td>
</tr>
<tr>
<td>OCR</td>
<td>optical character recognition</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>OPU</td>
<td>optical processing unit</td>
</tr>
<tr>
<td>PDF</td>
<td>probability density function</td>
</tr>
<tr>
<td>PESTLE</td>
<td>Political, Economic, Sociological, Technological, Legal and Environmental</td>
</tr>
<tr>
<td>R&amp;D</td>
<td>research and development</td>
</tr>
<tr>
<td>RBM</td>
<td>restricted Boltzmann machine</td>
</tr>
<tr>
<td>RNN</td>
<td>recurrent neural network</td>
</tr>
<tr>
<td>RPA</td>
<td>robotic process automation</td>
</tr>
<tr>
<td>SDG</td>
<td>sustainable development goal</td>
</tr>
<tr>
<td>SFIA</td>
<td>Skills for the Information Age</td>
</tr>
<tr>
<td>SVM</td>
<td>support vector machine</td>
</tr>
<tr>
<td>TPU</td>
<td>tensor processing unit</td>
</tr>
<tr>
<td>TRL</td>
<td>Technology Readiness Level</td>
</tr>
<tr>
<td>UX</td>
<td>user experience</td>
</tr>
<tr>
<td>VR</td>
<td>virtual reality</td>
</tr>
</tbody>
</table>
Artificial intelligence (AI): Artificial intelligence refers to systems designed by humans that, given a goal, act in the physical world by perceiving their environment; to build intelligent entities.

Assistive robot: A robot designed to assist a person.

Autonomous system: A system that can make decisions autonomously.

Autonomy: The ability of an object to make its own decisions.

Axon terminals: Axon terminals are terminations of the telodendrions (branches) of an axon.


Bayesian network: A Bayesian network, or belief network, is a probabilistic graphical model that represents a set of variables and their conditional dependencies.

Bias: Bias is the deviation of a statistical estimate from the actual quantity value. It can also mean the conscious and unconscious bias demonstrated in humans’ behaviours.

Big Data: Big Data refers to data sets that are so big and complex that traditional data-processing application software are inadequate to deal with them.

Boosting: Boosting is an ensemble algorithm that aims to reduce bias, and to convert weak learners to strong ones.

Bootstrap aggregating – bagging: Bootstrap aggregating is an ensemble algorithm used in classification and regression.

Chatbot: A chatbot is an artificial intelligence program that conducts a conversation via auditory or textual methods.

Classification: Classification is the problem of identifying to which set of classes an object belongs.

Clustering: Clustering groups a set of objects in such a way that objects in the same group are similar to each other.
**Cognitive simulation:** Cognitive simulation uses computers that test how the human mind works.

**Combinatorial complexity:** Combinatorial complexity is the exponential growth in computer power required to solve a problem that has many ever-increasing complex combinations.

**Combinatorial explosion:** A combinatorial explosion is the rapid growth of the number of combinations a system has to deal with.

**Connectionist:** Cognitive science that hopes to explain intellectual abilities using artificial neural networks.

**Data analytics:** Meaningful patterns are found from data via discovery and interpretation.

**Data cleaning:** Data cleaning prepares data for analysis.

**Data mining:** The process of discovering patterns in large data sets.

**Data science:** Data science uses scientific methods, processes, algorithms and systems to understand data.

**Data scrubbing:** See data cleaning.

**Decision trees:** A decision tree uses a tree-like graph or model of decisions.

**Deep learning:** Deep learning is a multi-layered neural network.

**Dendrites:** Dendrites are branched extensions of a nerve cell that propagate the electrochemical stimulation.

**Edges:** Edges are the geometric machine learning name for the brain’s axons.

**Emotional intelligence or emotional quotient (EQ):** The understanding of our emotions and the emotions of others.

**Ensemble:** Ensemble methods use multiple or a combination of learning algorithms to obtain better learning outcomes.

**Ethical purpose:** Ethical purpose is used to indicate the development, deployment and use of AI that ensures compliance with fundamental rights and applicable regulation, as well as respecting core principles and values. This is one of the two core elements to achieve trustworthy AI.

**Expert system:** A computer system that emulates the decision-making ability of a human expert.

**Feedforward neural network:** A feedforward neural network is an artificial neural network that iterates to find the weights of the network from information passing from the input to the output.
**Fourth industrial revolution**: The fourth industrial revolution represents new integrated approaches where technology becomes embedded within objects and societies.

**Functionality**: The tasks that a computer software program is able to do.

**Genetic algorithms**: A genetic algorithm is an algorithm inspired by the process of natural selection.

**Hardware**: Hardware are the physical parts or components of a computer.

**Heuristic**: Heuristic is a strategy derived from previous experiences with similar problems.

**High-performance computing – supercomputing**: A computer with very high performance.

**Human-centric AI**: The human-centric approach to AI strives to ensure that human values are always the primary consideration. It forces us to keep in mind that the development and use of AI should not be seen as a means in itself. The goal of increasing citizens’ wellbeing is paramount.

**Hyper-parameters**: Hyper-parameters, set before the learning process begins, are parameters that tune the algorithms’ performance in learning.

**Inductive reasoning**: Inductive reasoning makes broad generalisations from specific observations.

**Intelligent quotient (IQ)**: A standard test of intelligence.

**Internet of Things (IoT)**: IoT is the network of devices. Devices have embedded technology and network connectivity.

**k-means**: k-means is a clustering algorithm that generates k clusters. An object belongs to a cluster with a nearest mean to a prototype.

**k-Nearest Neighbour (k-NN)**: The simplest clustering algorithm.

**Layers**: Neural networks are made up of multiple layers. Each layer is made up of an array of nodes.

**Linear algebra**: The branch of mathematics concerning linear equations and functions and their representations through matrices and vector spaces.

**Logistic regression**: Used in classification to predict the probability of an object belonging to a class.

**Machine learning (ML)**: Machine learning is a subset of artificial intelligence. This field of computer science explores a computer’s ability to learn from data.
Model optimisation: The improvement of the output of a machine learning algorithm (e.g. adjusting hyper-parameters).

Natural language processing (NLP): An area of artificial intelligence concerned with the interactions between computer and human (natural) languages.

Natural language understanding: A term used to describe machine reading comprehension.

Nearest neighbour algorithm: The nearest neighbour algorithm was one of the first algorithms used to find nearest neighbours between sets of data (e.g. the travelling salesman problem).

Neural network (NN): A machine learning algorithm based on a mathematical model of the biological brain.

Nodes: Nodes represent neurons (biological brain) and are interconnected to form a neural network.

One-hot encoding: Transforms object features into a numerical form for use in algorithms (e.g. false is given the number 0 and true is given the number 1).

Ontology: The philosophical study of the nature of objects’ being and the relationships of objects.

Optical character recognition (OCR): The conversion of images of typed, handwritten or printed text into a form a computer can use.

Over-fitting or over-training: Over-fitting is a machine learning model that is too complex, has high variance and low bias. It is the opposite of under-fitting or under-training.

Probabilistic inference: Probabilistic inference uses simple statistical data to build nets for simulation and models.

Probability: Probability is the measure of the likelihood that an event will occur.

Pruning: Pruning reduces the size of decision trees.

Python: A programming language popular in machine learning.

Random decision forests or random forests: Random decision forests are an ensemble learning method for classification, regression and other tasks.

Regression analysis: In machine learning, regression analysis is a simple supervised learning technique used to find a trend to describe data.

Reinforcement (machine) learning: Reinforcement learning uses software agents that take actions in an environment in order to maximise some notion of cumulative reward or minimise a loss.
Robotic process automation: More often known as RPA, robotic process automation is a business process automation technology based on the notion of software robots.

Robotics: Robotics deals with the design, construction, operation and use of robots.

Scripts: Scripts are programs written in an application’s run-time environment. They automate tasks, removing the need for human intervention.

Search: The use of machine learning in search problems (e.g. shortest path, adversarial games).

Semi-supervised machine learning: Machine learning that uses labelled and unlabelled data for training.

Sigmoid function: A Sigmoid function is a mathematical function. It is a S-shaped, or Sigmoid, curve and is continuous.

Software: A generic term for the instructions that make a machine deliver functionality.

Software robots: A software robot replaces a function that a human would otherwise do.

Strong AI or artificial general intelligence (AGI): Strong AI’s goal is the development of artificial intelligence with the full intellectual capability of a human and not just to simulate thinking.

Supervised machine learning: Supervised machine learning uses labelled data to map an input to an output.

Support vector machine: A support vector machine constructs a hyperplane or set of hyperplanes (a linear subspace or subspaces).

Swarm intelligence: Swarm intelligence is the collective behaviour of self-organised systems, natural or artificial.

Symbolic: Symbolic artificial intelligence is the term in artificial intelligence research based on high-level ‘symbolic’ (human-readable) representations of problems.

System: A regularly interacting or interdependent group of objects that form a whole.

Trustworthy AI: Trustworthy AI has two components: (1) its development, deployment and use should comply with fundamental rights and applicable regulation as well as respecting core principles and values and ensuring ethical purpose; (2) it should be technically robust and reliable.

Turing machine: A mathematical model of computation.

Under-fitting or under-training: Under-fitting is when the machine learning model has low variance and high bias. It is the opposite of over-fitting or over-training.
**Universal design:** Universal design (a close relation to inclusive design) refers to broad-spectrum ideas meant to produce buildings, products and environments that are inherently accessible to all people.

**Unsupervised machine learning:** Machine learning that learns a function from unlabelled data.

**Validation data:** A set of data used to test the output of a machine learning model that is not used to train the model.

**Variance:** The expectation of the squared deviation of a random variable from its mean.

**Visualisation:** Any technique for creating images, diagrams or animations to communicate a message.

**Weak AI or narrow AI:** Weak artificial intelligence (weak AI), also known as narrow AI, is artificial intelligence focused on one narrow task. It is the contrast of strong AI.

**Weights:** A weight is a mathematical object used to give an object or objects additional/diminished influence or weight.
USEFUL WEBSITES

ARTIFICIAL INTELLIGENCE

European Defence Agency, European Defence Matters journal

European Parliament Committees, Hearings on artificial intelligence

Open AI
https://openai.com/

ROBOTS

Skills for Care, Scoping study on the emerging use of AI and robotics in social care

ETHICS

Alan Turing Institute, Data ethics: how can data science and artificial intelligence be used for the good of society?
https://www.turing.ac.uk/research/data-ethics

European Union, Ethics guidelines for trustworthy AI

Future of Life Institute
https://futureoflife.org/
USEFUL WEBSITES

MACHINE LEARNING

Google for Education, Google's Python class
https://developers.google.com/edu/python

The Royal Society, What is machine learning?
https://royalsociety.org/topics-policy/projects/machine-learning/

SUSTAINABILITY

International Organization for Standardization
https://www.iso.org/home.html

SmartCitiesWorld
https://www.smartcitiesworld.net/

United Nations, Take action for the sustainable development goals
This book was written with the express purpose of supporting the BCS AI Essentials and the BCS AI Foundation training courses and other scheduled courses already under development in the BCS AI course pipeline at the time of writing.

Its aim is to document what artificial intelligence is and what it is not, separate fact from fiction and educate those with an interest in AI. We have also included a number of topics that introduce the basics of machine learning and ethics.

We believe that this book is unique in that it brings together information and concepts that until now have been spread across numerous other volumes. The book also aims to simplify (where possible) complex and confusing AI concepts, making the topics highly accessible to those without a high-level degree in the subjects covered.

Our aim here is to bring these concepts to life by balancing theory with practice. We want to make the human part of an AI project as important as the AI itself. After all, machines are here to take the heavy lifting away from us humans. Not only that, but to give us extra capabilities that we wouldn’t have by ourselves. Humans and machines have unique capabilities, and it is important to find the right balance between them.

People, society and governments are quite rightly concerned with AI and its potential. As such, we have adopted the EU guidelines (https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai) on the ethical use of AI. These guidelines ask us to build human-centric ethical purpose that gives us trustworthy and technically robust AI. This puts us humans at the goal setting lead in AI.

Stuart Russell’s\textsuperscript{1} recent book on human compatible AI puts the human into our consideration when undertaking an AI project. His take on this asks if we should think about AI as serving our needs, telling us how to be better humans. In doing so it gives us an alternative to controlling AI, by asking the AI what is best for us as humans.

As we move into the fifth industrial revolution, we have the opportunity to think about humans and machines. How do we complement each other? How can AI and machines leave humans to undertake more valued work and deal with ambiguous or contradictory situations, to become more human, to build better societies and for all to exploit their talents? What are the new roles for humans, humans and machines, and machines only? Simply considering these roles and focusing on opportunities paves the way for a richer environment as we progress. We focus on our needs and are less distracted by the notion of robots coming to take over our jobs! Can you, your current field or organisation benefit from learning from experience? If so, read on.
This chapter sets the scene for artificial intelligence (AI). We look at intuitive definitions of human and artificial intelligence. We also introduce the European Union’s (EU’s) ethical guidelines for AI and take a look back at the progress of AI over the past couple of centuries, examining how we as humans relate to this disruptive technology.

1.1 THE GENERAL DEFINITION OF HUMAN INTELLIGENCE

It is a vast understatement to say that human beings are one of the wonders of the universe. Human intelligence is the culmination of billions of years of evolution from single cell organisms to what we are today, which is ultimately marked by our ability to undertake complex mental feats and be self-aware. It also includes the ability to recognise our place in the universe and ask annoying philosophical questions such as ‘Why are we here?’ and ‘What is our purpose?’

There are many definitions of human intelligence. Our chosen definition is useful because it is intuitive and gives us a practical base that builds a strong foundation for AI. In fact, it needs to be a little more than intuitive: it also needs to guide us as to what AI is useful for in practice. When considering this definition, we must keep in the back of our minds that the need to find the right balance of theory and practice is paramount. We will also need to understand that AI and machine learning (ML) have significant limitations, and this will become apparent as we moved through the book.

Take five minutes to think about what it means to be human.

Leonardo da Vinci captured the essence of his science in art. René Descartes stated ‘cogito, ergo sum’ (I think therefore I am). Ada Lovelace wrote the first algorithm and notes on the role of humans and society with technology. Neil Armstrong was the first to put one foot on the moon, and changed our perspective of the world completely. Roger Bannister was the first to run a mile in under four minutes. Dr Karen Spärck Jones gave us the theoretical foundations of the search engine. Tu Youyou is a tenacious scientist who discovered a cure for malaria, which won her the Nobel Peace Prize in 2015. Tu’s intellectual talents are amazing, and, after perfecting her cure, she volunteered to be the first person for it to be tested on. We could ask ourselves if this was confidence or bravery.

We have set ourselves up here to introduce the concept of subjectivity. We have free will and all of us have our own unique subjective experience. We are conscious and
subjective, and conscious experience is something that we will need to be all too aware of as we develop AI.

We must always consider what effects artificial intelligence will have on humans and society.

Robert Sternburg gives us a useful definition of human intelligence, at least in so far as it relates to AI:

Human intelligence: mental quality that consists of the abilities to learn from experience, adapt to new situations, understand and handle abstract concepts, and use knowledge to manipulate one’s environment.

Here we can quickly recognise the desire to manipulate our environment in some way, and that we will use all our human talents to do so. It is general, and we still need to identify the type of learning from experience that humans do, or, to put it another way, what machines can help us with.

Sometimes called natural intelligence (NI), human intelligence is generally considered to be the intellectual accomplishment of humans and has been discussed by philosophers for thousands of years. Of course, other living things possess NI to some degree as well, but for now let’s just consider human intelligence. We are biased, but it’s fair to say that humans are the most intelligent living organisms on the planet. Just as we developed tools in the Stone and Iron Ages, we are now equipping ourselves with machines to help us intellectually.

We may phrase this as coming to the correct conclusions – hypothesising and testing – and understanding what is real, although we sometimes get it wrong. It’s also about how to understand complex problems like weather prediction or winning a game of chess; adapting what we have learned through things like abstraction, induction, simplification and creativity. It allows us to adapt and control our environment and interact socially, giving us an evolutionary advantage.

It may also make sense to consider human intelligence from a number of perspectives, such as:

- **Linguistic intelligence** – the ability to communicate complex ideas to another.
- **Mathematical intelligence** – the ability to solve complex problems.
- **Interpersonal intelligence** – the ability to see things from the perspective of others, or to understand people in the sense of having empathy.

So, how do we acquire these particular skills or traits? Through learning from experience.
1.1.1 Human learning

Human learning is the process of acquiring knowledge. It starts at a very early age, perhaps even before we are born. Our behaviour, skills, values and ethics are acquired and developed when we process information through our minds and learn from those experiences. Human learning may occur as part of education, personal development or any other informal/formal training, and is an ongoing process throughout our lives. Each person has a preference for different learning styles and techniques (e.g. visual, aural, kinaesthetic, etc.).

Machine learning can give us super-human capability; we can search every research paper using a search engine on our smartphone. This could take 'old school' academics a lifetime of hard work, travel and focused attention. Machine learning is changing our beliefs, behaviour and speed of progress.

Now we understand the basics of human intelligence and human learning, let’s see how that compares to artificial intelligence and machine learning. We will also explain further some of the jargon that is thrown around in AI circles and explain exactly what artificial intelligence is – and what artificial intelligence isn’t.

1.2 DEFINITION OF ARTIFICIAL INTELLIGENCE

In simple terms, AI is intelligence demonstrated by machines, in contrast to the NI displayed by humans and other animals.

Stuart Russell and Peter Norvig, the authors of the standard AI textbook *Artificial Intelligence: A Modern Approach*, explain that AI is a universal subject and helpful to us all. Learning from experience is AI’s signature. We will use this concept a lot; it applies to machines as it does, perhaps more so, to humans.

Einstein is often quoted as saying: ‘The only source of knowledge is the experience’.

Ask yourself the question: Can machines help us to learn from experience? If the answer to this question is yes, then AI can help you!

1.2.1 Artificial general intelligence (AGI)

AGI is the hypothetical intelligence of a machine that has the capacity to understand or learn any intellectual task that a human being can understand or learn. There is wide agreement among AI researchers that to do this AGI would need to perform a full range of human abilities, such as using strategy, reasoning, solving puzzles, making judgements under uncertainty, representing knowledge (including common-sense knowledge), planning, learning and communicating in natural language, and integrate all these skills towards common goals. AGI might never be achieved, and not everyone agrees whether it is possible or if we will ever get there.
A number of tests have been put forward to decide if and when AGI (human-like intelligence) has been achieved. One of the first tests was the ‘Turing’ test devised by the British scientist Alan Turing. The ‘Turing’ test goes along the lines of a machine and a human conversing while heard but unseen by a second human (the evaluator), who must evaluate which of the two is the machine and which is the human. The test is passed if they can fool the evaluator a significant fraction of the time. Turing did not, however, prescribe what should qualify as intelligence. Several other tests have since been defined, including visual and construction tests. In reality, a non-human agent would be expected to pass several of these tests.

Current AGI research is extremely diverse and often pioneering in nature, and estimates vary from 10 to 100 years before AGI is achieved. The consensus in the AGI research community seems to be that the timeline discussed by Ray Kurzweil in *The Singularity is Near* (i.e. between 2015 and 2045) is plausible. Kurzweil has based his estimate of 2045 on the exponential advances in four key areas of research: AI, robotics, genetic engineering and nanotechnologies.

There are other aspects of the physical human brain besides intelligence that are relevant to the concept of strong AGI. Strong AGI is AGI with consciousness. We discuss consciousness in more depth later in the book. If strong AI or some kind of conscious AI emerges in the future, then our ethical development of these technologies is paramount. Ethical use of AI is a fundamental requirement, and we need to build in our ethics from the start. It’s not too late.

### 1.3 A BRIEF HISTORY OF ARTIFICIAL INTELLIGENCE

Today we are mainly concerned with the current use and future applications of AI, and the benefits we would like to obtain from its use. But we should not ignore where and when AI first appeared, the history of AI and the challenges encountered along the way. To ignore our AI history would be to risk some of those challenges recurring.

Before we look at AI’s history, it’s worth noting that it is now generally recognised that John McCarthy coined the term ‘artificial intelligence’ in 1955. John McCarthy is one of the ‘founding fathers’ of AI, together with Alan Turing, Marvin Minsky, Allen Newell and Herbert A. Simon.

#### 1.3.1 Back in antiquity

Way back in antiquity there were legends, stories and myths of effigies and artificial mechanical bodies endowed with some form of intelligence, typically the creation of a wise man or master craftsman. We still have references to golems of biblical times, which were magical creatures made of mud or clay and brought to ‘life’ through some incantation or magic spell.

Aristotle (384–322 BC), the Greek polymath and father of later Western philosophy, was the first to write about objects and logic and laid the foundations of ontology and the scientific method. As a result, today we teach natural science, data science, computer science and social science.
Without the required technology, many centuries passed without any progress in the pursuit of AI.

### 1.3.2 The 18th and 19th centuries

In the 18th century we saw the mathematical development of statistics (Bayes theorem) and the first computer description and algorithm from Ada Lovelace.

During the 19th century, AI entered the world of science fiction literature with Mary Shelley’s *Frankenstein* in 1818 and Samuel Butler’s novel *Erewhon* in 1872, which drew on an earlier (1863) letter he had written to *The Press* newspaper in New Zealand, ‘Darwin among the Machines’. Butler was the first to write about the possibility that machines might develop consciousness by natural selection.

In 1920 Czech-born Karel Čapek introduced the word ‘robot’ to the world within his stage play, *R.U.R.* (Rossum’s Universal Robots). ‘Robot’ comes from the Slavic language word *robota*, meaning forced labourer.

AI has since become a recurrent theme in science fiction writing and films, whether utopian, emphasising the potential benefits, or dystopian, focusing on negative effects such as replacing humans as the dominant race with self-replicating intelligent machines. To some extent, we can say that what was yesterday’s science fiction is quickly becoming today’s science fact.

### 1.3.3 Technological advances during the ‘golden years’ (the second half of the 20th century)

In 1943 Warren McCulloch and Walter Pitts created a computational model for neural networks (NNs), which opened up the subject. The first was an electronic analogue NN built by Marvin Minksy. This sparked the long-lasting relationship between AI and engineering control theory.

Then in 1950 the English mathematician Alan Turing published a paper entitled ‘Computing Machinery and Intelligence’ in the journal *Mind*. This really opened the doors to the field that would be called Artificial Intelligence. It took a further six years, however, before the scientific community adopted the term ‘artificial intelligence’.

John McCarthy organised the first academic conference on the subject of AI at Dartmouth, New Hampshire, in the summer of 1956. The ‘summer school’ lasted eight weeks and brainstormed the area of artificial intelligence. It was originally planned to be attended by 10 people; in fact, people ‘dropped in’ for various sessions and the final list numbered nearly 50 participants. Russell and Norvig quoted McCarthy’s proposal for the summer school (p. 17):

We propose that a 2-month, 10-man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make
machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.

The mid-1950s into the early 1960s also saw the start of machines playing draughts, checkers and chess. This was the start of 'Game AI', which is still big business to this day and has developed into a multi-billion-dollar industry.

In 1959, Arthur Lee Samuel first used and popularised the term 'machine learning', although, today, Tom Mitchell’s definition is more widely quoted (see Section 1.5). Samuel’s checkers-playing program was among the world’s first successful self-learning programs, and was amazing given the limited technology available to him at the time.

The 1960s saw the development of logic-based computing and the development of programming languages such as Prolog. Previously it took an astronomical number of calculations to prove simple theorems using this logic method. This opened up a debate regarding how people and computers think.

Hubert Dreyfus, a professor of philosophy at the University of California Berkeley (1968–2016), challenged the field of AI in a book, explaining that human beings rarely used logic when they solved problems. McCarthy was critical of this argument and the association of what people do as being irrelevant to the field of AI. He argued that what was really needed were machines that could solve problems – not machines that think as people think.

In 1973 many funding resources were withdrawn from AI research, mainly as a result of Sir James Lighthill’s report into the state of AI, and ongoing pressure from US Congress and the US and British governments. As a result, funding was reduced for research into artificial intelligence, and the difficult years that followed would later be known as an 'AI winter', lasting from 1974 to 1980. Sir James highlighted that other sciences could solve typical AI problems; AI would hit combinatorial explosion limits and practical problems would resign AI to solving only trivial toy problems. In his view, general AI could not be achieved and there was no prospect of a general AI robot ever.

During the 1970s, however, a number of breakthroughs were made, most notably in 1974 by Ray Kurzweil’s company, Kurzweil Computer Products Inc, developing optical character recognition (OCR) and a text-to-speech synthesizer, thus enabling blind people to have a computer read text to them out loud. It was unveiled to the public in 1976 at the National Federation for the Blind, and became commercially available in 1978. Kurzweil subsequently sold the business to Xerox. It is widely considered to be the first AI product, although today we don’t associate OCR with AI or ML because it is now routine and we take it for granted. It was the precursor to reading handwritten text and the development of natural language processing (NLP).

Following on from the ‘AI winter’, the 1980s saw a boom in AI activity. In 1986 David Rumelhart and James McClelland developed ideas around parallel distributed processing and neural network models. Their book, Parallel Distributed Processing: Explorations in the Microstructure of Cognition, described their creation of computer
simulations of perception, giving computer scientists their first testable models of neural processing.14

The 1980s also saw the rise of the robots, with many researchers suggesting that AI must have a body if it is to be of use; it needs to perceive, move, survive and deal with the world. This led to developments in sensor-motor skills development. AI also began to be used for logistics, data mining, medical diagnosis and other areas.

During the 1990s a new paradigm called ‘intelligent agents’ became widely accepted in the AI community. An intelligent agent is a system that perceives its environment and takes actions that maximise its chances of success.

In 1997, IBM’s Deep Blue became the first computer chess-playing system to beat a reigning world chess champion, Garry Kasparov. It won by searching 200,000,000 moves per second. By comparison, 20 years on, Apple’s iPhone7 was 40 times faster than Deep Blue had been in 1997.

1.3.4 Bringing us up to date: the first 20 years of the 21st century

Over the last 20 years, developments in technology with cheaper and faster computers have finally caught up with our AI aspirations. We have started to gain the computing power to really put AI to work.

In 2002 iRobot released Roomba, which autonomously vacuums a floor while navigating and avoiding obstacles. It sold a million units by 2004, and over 8 million units by 2020. iRobot then went on to create a range of other commercial, environmental, military and medical robots.

In 2004 the Defence Advanced Research Projects Agency (DARPA), a prominent research organisation of the United States Department of Defense, introduced the DARPA Grand Challenge, offering prize money for competitors to produce vehicles capable of travelling autonomously over 150 miles. Then in 2007, DARPA launched the Urban Challenge for autonomous cars to obey traffic rules and operate in an urban environment, covering 60 miles within six hours.

Google entered the self-driving autonomous market in 2009 and built its first autonomous car, which sparked a commercial battle between Tesla, General Motors, Volkswagen and Ford, to name a few entrants into the same market.

From 2011 to 2014 a series of smartphone apps were released that use natural language to answer questions, make recommendations and perform actions: Apple’s Siri (2011), Google’s Google Now (2012) and Microsoft’s Cortana (2014).

SCHAFT Inc of Japan, a subsidiary of Google, built robot HRP-2, which defeated 15 teams to win DARPA’s Robotics Challenge Trials in 2013. The HRP-2 robot scored 27 out of 32 points over eight tasks needed in disaster response. The tasks were to drive a vehicle, walk over debris, climb a ladder, remove debris, walk through doors, cut through a wall, close valves and connect a hose.
In 2015, an open letter petitioning for the ban in development and use of autonomous weapons was signed by leading figures such as Stephen Hawking, Elon Musk, Steve Wozniak and over 3,000 researchers in AI and robotics.

In 2016 Google’s DeepMind AlphaGo supercomputer beat Lee Se-dol, a world Go champion, at a five-game match of Go (a strategic board game that originated in China more than 2,500 years ago and is considered one of the most complex strategy games in the world). It took just 30 hours of unsupervised learning for the supercomputer to teach itself to play Go. Lee Se-dol was a 9 dan professional Korean Go champion who won 27 major tournaments from 2002 to 2016. He announced his retirement from the game in 2019, declaring that AI has created an opponent that ‘cannot be defeated’. In 2017 AlphaGo Zero, an improved version of AlphaGo, beat the world’s best chess-playing computer program, StockFish 8, winning 28 of the 100 games and drawing 72 of them. What is astonishing is that AlphaGo Zero taught itself how to play chess in under four hours.

Also in 2017, the Asilomar Conference on Beneficial AI was held near Monterey, California. Thought leaders in economics, law, ethics and philosophy spent five days in discussions dedicated to beneficial AI. It discussed AI ethics and how to bring about beneficial AI while at the same time avoiding the existential risk from AGI.

1.3.5 The industrial revolutions

We are currently in the middle of the fourth industrial revolution and some would argue that we have already reached the fifth, although that has yet to be defined. In each industrial revolution, mankind has designed and developed technologies that have made a paradigm shift in human capabilities and exploited those technologies to drive through progress, although some would argue that not all of mankind benefited as a result of these advancements.

The first industrial revolution, sometimes referred to as the Industrial Revolution, occurred during the 18th and 19th centuries, primarily in Europe and the United States. It mainly grew between 1760 and 1840 (although the exact dates are open to debate) and resulted in the transition from hand production methods to machine production – initially led by the development of the steam engine powering large textile factories. It was a major turning point in history, led to worldwide trading and rapid population growth and resulted in large swaths of rural societies becoming urbanised and industrial.

The second industrial revolution occurred between 1870 and 1914, primarily across Europe, the United States and Japan. Sometimes known as the Technological Revolution, again it was a paradigm shift for mankind with the introduction of mass production and assembly lines. Increased use of electricity allowed for advancements in manufacturing and production, and resulted in technological advances such as the internal combustion engine, the telephone and the light bulb.

The third industrial revolution started in the 1950s and is often known as the Digital Revolution. It brought us space exploration, biotechnology, semiconductors, mainframe computing and information and communications technology (ICT), and embedded technology into society with personal computers, the internet and automated production of goods.
There is a bit of a blur between the third and fourth industrial revolutions. We believe that we are in the fourth today. The fourth exploits the gains made in the ‘digital revolution’ and is disruptive, driven by AI, robotics, Internet of Things (IoT), three-dimensional (3D) plastic printing, nanotechnology, bioengineering and so on.

I know what you’re thinking: ‘What happened between the industrial revolutions?’ We didn’t stop inventing or improving between the revolutions, and different parts of the globe experienced them at different times at different speeds, but these particular periods were paradigm shifts in thinking and invention.

The Diffusion of Innovation theory, developed by E. M. Rogers in 1962, explains how, over time, an idea or product gains momentum and spreads (diffuses) through a specific population, and he classifies adopters of innovations into five adopter categories: innovators, early adopters, early majority, late majority and laggards. This is based on the idea that certain individuals are inevitably more open to adoption and adaption than others. The adoption of AI technologies globally will be faster than the adoption of other technologies because we have seen a reduction in the time an industrial revolution lasts, from centuries to decades and now fractions of a decade, but we will still be led by innovators and there will still be laggards and even Luddites.

Human intelligence has led us through the various industrial revolutions. We are in the middle of the fourth, but what does this mean if we think about AI? It means increasingly that we will have more robots doing routine monotonous, laborious and dangerous tasks – doing the ‘heavy lifting’. This introduces the idea of humans and machines working together at what they are each good at.

1.3.6 AI as part of ‘universal design’

The concept of universal design was coined by the architect Ronald Mace, and is the design and composition of an environment so that it can be accessed, understood and used to the greatest extent possible by all people. For example, door handles, elevator controls and light switches should be designed for use by all people regardless of their age, size, ability or disability. A consideration, therefore, for any new AI service, system or product is that it is designed for all.

Incorporating the potential of AI in universal design can allow someone who is blind to ask a ‘home assistant’ what the weather is like, or for someone who is physically incapacitated to turn on the heating, or someone who is travelling home to turn on the heating while travelling.

A human working alongside an intelligent AI-enabled machine has the capability to do a lot more, whether that is in a work environment reducing physical risk or exertion, or on a personal basis educating us or translating our conversations. AI systems have the potential to make us more human.

Our efforts and endeavours developing AI-enabled products, systems and services should be focused on allowing us to be more human, improving us as humans (improving our physical and mental performance or by making us more active) or improving our ability to communicate or socialise.
The continual emergence of AI systems and products means that we as individuals and as part of wider society are going to have to reimagine every area of our lives to use AI in a positive way for all.

1.3.7 The concept of intelligent agents

A computer scientist may view AI as ‘intelligent agents’ perceiving their environment and taking actions to achieve a goal. Russell and Norvig describe intelligent agents in more detail.

As we mentioned previously, the scientific method has allowed humans to develop at an ever quicker pace, and we can, broadly speaking, think of these as the industrial revolutions. From the 1980s AI has also adopted the scientific method and in doing so has been absorbed into the revolutions. The basic intelligent agent can be very rudimentary and it’s often difficult to see why they would be considered intelligent. Learning from experience is a common phrase used in AI and the learning agent is an intelligent agent that learns from experience.

The learning agent, as proposed by Russell and Norvig, is an agent with an explicit ability to learn, to be autonomous. It is useful to always relate what we are doing in AI to the learning agent that can perceive its environment and take actions to achieve a goal. It also gives us an intuitive insight into what artificial intelligence and machine learning are – for instance to understand that ML is about learning from data – but that this alone is insufficient when we are designing products and services that will certainly operate in an environment.

We can define other types of agent, but these may not learn. We explore this further in Section 2.1, Understanding the AI intelligent agent.

Deep learning is a technique that gives learning agents the ability to learn from sensors and actuators and has been very successful in achieving very complex tasks.

Figure 1.1 shows how deep learning fits into the overall schematic of AI.

1.3.8 Consciousness – the unsolved problem

Human consciousness, sometimes referred to as sentience, is, in its simplest terms, having an awareness of an internal or external existence or having a mental state you are aware of being in. This can be compared to subconsciousness, which is that part of your mind that notices and remembers information when you are not actively trying to do so and can often influence your behaviour even though you do not realise it.

Some humans fear conscious machines, however unlikely they may be. Consciousness is a complex area at the cutting edge of AI research; our knowledge is growing but we might never understand what consciousness actually is. Interfacing the human
Machine consciousness or artificial consciousness is the domain of cognitive robotics and is primarily concerned with endowing an artificial entity with intelligent behaviour by providing it with a processing architecture. This will allow it to learn and reason about how to behave in response to complex goals in a complex world. The aim of the theory of artificial consciousness is to define what would have to be synthesised to develop artificial consciousness in an engineered artefact such as a robot.

Many organisations are pursuing developments in artificial consciousness, but comparing consciousness in functional machines to consciousness in functional
humans is more difficult than expected and the topic has raised a debate around the risks of developing conscious machines.

**1.3.9 Modelling subjective humans – neuro-linguistic programming**

By understanding how some humans achieve amazing things, can we then teach others to achieve amazing things? Can we teach a machine? What makes a good astronaut, for instance? How do charismatic people convince others to follow them? What makes a good marketing strategy? If we can’t describe – and by this we mean explicitly – or teach other humans how to do something, what chance do we have to design, build and teach a machine to do it? Our source of intellect is the human. The role models for any intelligent machine are currently humans, but we are subjective and objective beings. The human intellect is much more than scientific objectivity; neuro-linguistic programming gives readers another perspective of human intelligence.

AI learns from humans and we humans give the machines goals, environments and learning techniques. Humans are subjective beings and, if we are going to teach machines, we need an understanding of what it takes to understand subjective humans. Neuro-linguistic programming can help us with that.

Practically, we will also need to champion our AI, especially if AI is new to an organisation. This means dealing with the subjective nature of management, stakeholders and people in your team. We will need to capture hearts and minds in order to champion our AI projects. Neuro-linguistic programming is also a useful approach in this regard; it has been adopted by some of the world’s most successful organisations. This section is quite general, but vital because, in developing or championing your AI projects, you will need these skills.

Neuro-linguistic programming asks: ‘Are the processes that humans use to achieve amazing things the same as the processes that humans use to hold them back?’ Sometimes when humans are held back, this can be negative and destructive.

---

**Neuro-linguistic programming**

Neuro-linguistic programming comes from the work of John Grinder and Richard Bandler, presented in their book, *The Structure of Magic*. They diligently studied how skilled psychologists counselled humans, and realised that we humans do the following:

- **generalise**;
- **delete**;
- **distort**.

Neuro-linguistic programming is now a creative, exciting array of tools, techniques and approaches to understand how humans and organisations achieve amazing things. However, it is not possible to cover neuro-linguistic programming in depth in a book on AI; instead, we’ll give an overview of a couple of fundamental points taught...
INTRODUCTION: ETHICAL AND SUSTAINABLE HUMAN AND ARTIFICIAL INTELLIGENCE

to practitioners. The objective here is to understand our AI role models – humans – well and be able to handle the challenges of running an AI project and the subjective parts of the project specifically, such as motivating a team, dealing with the fears of individuals and organisations, creating a culture supporting AI, building business cases, communicating and influencing stakeholders. Neuro-linguistic programming could also play a role in how humans and machines interact in the future.

Our first neuro-linguistic programming example is building rapport, something that we need to be able to do well. One of the ways we might do this is understanding the type of person we are communicating with. Are they, for example, visual and express themselves in terms of visual words, such as ‘I see what you want to achieve’, ‘I saw red the moment I read that’? Are they auditory and express themselves in auditory words, such as, ‘I hear what you are saying’, ‘Your objective comes across loud and clear’? Or are they kinaesthetic and describe things in words we might associate with feelings, such as ‘I don’t feel happy about your objective’, ‘I’m all warm about the new plan’? If we mirror the body movements and language of the person or group we are communicating with it gives us better rapport with them. Rapport can be the difference between influencing a stakeholder or losing the sale. If you are working on a sensitive or emotive area of AI, such as healthcare, building rapport with key stakeholders will be fundamental to the success of the project.

1.3.9.1 Dilts’ logical levels
The second example is Dilts’ logical levels of thinking; we have picked a simple version here to illustrate. It is the fundamental structure in understanding change, demonstrated in Figure 1.2.

Figure 1.2 Dilts’ logical levels of change

<table>
<thead>
<tr>
<th>Culture or attitude</th>
<th>Spirituality</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Identity (my role)</td>
</tr>
<tr>
<td></td>
<td>Value and beliefs (what?)</td>
</tr>
<tr>
<td>Habits</td>
<td>Skills and capabilities (how?)</td>
</tr>
<tr>
<td></td>
<td>Behaviours (what?)</td>
</tr>
<tr>
<td>Environment (where, when, with whom?)</td>
<td>Ethics probably fits here</td>
</tr>
</tbody>
</table>

Human logical levels is part of neuro-linguistic programming. This has been a passion and interest of Robert Dilts for decades, and he has made neuro-linguistic programming into a successful international business, helping numerous organisations and individuals.
Our first exploration of the logical levels can be top down or bottom up. So, as an engineer, we might go top down. As an engineer, I value objectivity, science, safety and my values; the things that motivate me are integrity, professionalism, honesty and so on. I believe engineers can make a better sustainable society. The skills I need to do this are mathematics, craft skills, rapport skills and the ability to work in teams. I need the capability to organise my thoughts and draw clear engineering drawings and plans. I behave in a professional manner and work in a variety of places, including a design office and workshop.

This is a pretty basic description of an engineer. How do we make this more explicit? So detailed, in fact, that we could teach other engineers (or a machine) what we are doing so that they are competent and can describe explicitly what they are good at? What happens if we meet other engineers who don’t have our values? How does this change their engineering? Neuro-linguistic programming builds models of individuals and organisations so that we can understand them explicitly and are fully aware of what it is that makes us good at something.

At the higher levels of neuro-linguistic programming, spirituality, human identity and values and beliefs play a pivotal role in the success of a congruent outcome. Ethics plays a large role in occupations that involve humans heavily, such as engineering, care, health, politics. These are part of higher levels in Dilts’ logical levels. In AI, we need to explore ethics because we are now embracing humans and machines working together. It is possible that machines may have sufficient autonomy that they will influence and determine the outcome of human actions!

The European Union have put together guidelines on AI. They want AI to have an ethical purpose that is trustworthy and technically robust. This, at first inspection, seems easy, but humans are subjective, and measuring what we do is not a simple case of finding an objective measure. Neuro-linguistic programming gives us the tools to work with others and understand those who have different opinions. Simply surrounding ourselves with those we already have rapport with may not necessarily lead to a well-informed outcome. When we think about AI and the need for an ethical purpose, we need to work in the higher levels of Dilts’ logical levels. These levels can be emotionally more involved, and we will certainly need tools and skills to do this.

When we frame AI challenges as those that must have a human ethical purpose, the skills we need to make progress shift to human skills. These are far removed from learning computer skills, such as a new parallel programming language or how to use a cloud AI service. If we are not careful, a machine learning approach to AI will not fully embrace humans and machines working together to achieve goals that have this human ethical purpose.

1.3.10 Ethics and trustworthy AI

Ethics is a complex subject even before we start to include the topic of AI, but it has to be a consideration from the outset. We need to consider how ethics should be applied, who is responsible for our AI systems and products, what the law is, how human rights may be affected and even whether we will need to develop robotic rights.

If we ever develop AGI or strong AGI, then robotics becomes a major ethical area to understand.
This book was never intended to be a book on AI ethics, which is a constantly evolving space. This section references what we believe will give you an insight into the subject. From here you can do further in-depth research into the AI ethics that are most applicable to your organisation, your industry or your country.

Some areas of work include:

The Engineering and Physical Sciences Research Council

The Future of Life Institute
[https://futureoflife.org/](https://futureoflife.org/)

and

The International Electrotechnical Commission, First International Standards committee for entire AI ecosystem

The world of AI is constantly changing, and the rate of change is not linear; you may not have noticed, but it is actually growing exponentially, which basically means it is accelerating faster and faster on an annual basis. Some would argue that its growth is potentially out of control and we need to put the brakes on; others suggest that we just let it run and see where it goes.

Many people have fears that not all AI improvements are necessarily beneficial to humankind or in societies’ best interests. For example, there is the potential to weaponise AI, for it to intrude into our lives and privacy by listening and watching everything, and for individuals’ identification without consent with covert AI systems. There is also the potential to have control of our personal data taken away from us. So, how should AI be managed and controlled? Or, what happens if we try to do the right thing – whatever that means – and we get it wrong?

1.3.11 General definition of ethics

Ethics is a philosophical subject that manifests itself in individuals as what we term our moral principles; in fact, many people use the words interchangeably. Basically, both ethics and morals relate to ‘right’ and ‘wrong’ conduct.

It is worth noting that Ethics, the subject that academics and students study, is treated as singular, for example, ‘my elected subject this term is Ethics’. Our moral principles or ethical principles that motivate and guide us are treated as plural. This way we can differentiate between the subject, which is singular, and a person’s ethical principles, which are plural.

Further expansion on ethics can be found in *An Intelligent Person’s Guide to Ethics* by Lady Warnock. ¹⁸
Morals tend to refer to an individual’s own personal principles regarding what is right and wrong. Ethics generally refer to rules provided by an external source or community, for example codes of conduct in the workplace or an agreed code of ethics for a profession such as medicine or law.

Ethics as a field of study is centuries old and centres on questions such as:

- What is a ‘good’ action?
- What is ‘right’?
- What is a ‘good’ life?

1.3.11.1 The role of AI ethics

We all want beneficial AI systems that we can trust, so the achievement of trustworthy AI draws heavily on the field of ethics. AI ethics could be considered a subfield of applied ethics and technology, and focuses on the ethical issues raised by the design, development, implementation and use of AI technologies.

The goal of AI ethics is therefore to identify how AI can advance or raise concerns about the ‘good’ life of individuals, whether this be in terms of quality of life, mental autonomy or freedom to live in a democratic society. It concerns itself with issues of diversity and inclusion (with regard to training data and the ends to which AI serves) as well as issues of distributive justice (who will benefit from AI and who will not).

What comes out of consideration of AI ethics are some general themes that most agree on, and these are:

- **Transparency** – we might think of this as how we understand what went wrong when AI gets it wrong. Can we be explicit as to why an AI technology failed? For instance, a NN is generated by an algorithm that we can understand; however, for it to be transparent we would also need to understand explicitly why the NN came to a particular outcome. It gives us a basis on which to learn how the AI system performed and why.
- **Accountability** – we need to be careful or we could get inequality, bias, unemployment and so on.
- **Weaponisation** – creating lethal autonomous weapons is a red line that we must not cross; however, national security is a special subject that is beyond the general use of AI in society.
- **Harm** – the AI must do no harm.

1.3.11.2 What is already in place?

There is already a lot of legislation in place globally and locally, such as the Human Rights Act and data protection legislation that protect our basic human rights. However, often this is abused by the state and by private organisations alike, so why should we trust the creators of AI systems and services to operate responsibly?
1.3.12 Human-centric ethical purpose – fundamental rights, principles and values

A human-centric ethical purpose aims to enhance human capabilities and societal wellbeing. It builds on the Universal Declaration of Human Rights and the EU's Charter of Fundamental Rights of the European Commission. Typical measures of how this might be demonstrated are the United Nations sustainability goals.

There are several bodies and many organisations worldwide currently working on codes of ethics around AI.

1.3.12.1 The European Union

AI ethics guidelines are produced by the European Commission’s High-Level Expert Group on Artificial Intelligence. The guidelines are for trustworthy AI.

Trustworthy AI basically has two components:

1. It should respect fundamental rights, applicable regulation and core principles and values, ensuring an ‘ethical purpose’.
2. It should be technically robust and reliable, since, even with good intentions, a lack of technological mastery can cause unintentional harm.

It also considers:

- **Rights** – a collection of entitlements that a person may have and which are protected by government and the courts.
- **Values** – ethical ideals or beliefs for which a person has enduring preference and which determine our state of mind and act as a motivator.
- **Principles** – a fundamental well-settled rule of law or standard for good behaviour, or collectively our moral or ethical standards.

At the time of writing, the latest revision of the EU’s ethical guidance on AI was produced in March 2019 and we expect it to be revised on a regular basis.

1.3.12.2 Future of Life Institute (US founded)

The Future of Life Institute has developed the Asilomar Principles for AI (https://futureoflife.org/ai-principles/?cn-reloaded=1). There are 23 principles relating to:

- **Research** – goals, funding, policy, cultures, race avoidance (speed of progress).
- **Ethics and values** – safety, failure transparency, judicial transparency, responsibility, value alignment, human values, personal privacy, liberty and privacy, shared benefit, shared prosperity, human control, non-subversion, AI arms race.
- **Longer term issues** – capability caution, importance, risks, recursive self-improvement, common good.
The name Asilomar was inspired by the highly influential Asilomar Conference on Recombinant DNA in 1975. The biotechnology community are still influenced by the voluntary guidelines developed at this conference.

1.3.12.3 Individual organisations
Many organisations are in the process of developing their own ‘ethics for AI’.

Google is a good example of an organisation that decided to create an in-house AI ethics board, only to hit problems within a week of its launch. Google’s Advanced Technology External Advisory Council (ATEAC) was supposed to oversee its work on artificial intelligence and ensure it did not cross any lines, and could be dedicated to ‘the responsible development of AI’. It was, however, dissolved after more than 2,000 Google workers signed a petition criticising the company’s selection of an anti-LGBT advocate.

1.3.13 Trustworthy AI

What should you be doing? In the same way that we manage health and safety or compliance within our organisations, if we are working with AI or developing AI systems and services then we should adopt specific AI principles and values, for example:

- The principle of beneficence: Do Good.
- The principle of non-maleficence: Do No Harm.
- The principle of autonomy: Preserve Human Agency.
- The principle of justice: Be Fair.
- The principle of explicability: Operate Transparently.

These principles can be found in various disciplines such as medicine, engineering, accountancy and law; the five listed here are taken from HDSR.20 The principles and values need to be applied to our individual use cases, for example autonomous vehicles or personal care systems, or AI to decide insurance rates. Achieving trustworthy AI means that the general and abstract principles documented above need to be mapped into concrete requirements for AI systems and applications.

The 10 requirements documented below have been derived from the rights, principles and values detailed previously. While they are all equally important in different application domains and industries, the specific context needs to be considered for further handling. These requirements for trustworthy AI are given in alphabetical order to stress the equal importance of all requirements.

1. **Accountability** – means to be answerable for actions or decisions, and is essentially about ownership and initiative. Typically, one person should be accountable, although many may be responsible. To hold someone accountable means the person is being asked to explain why they did (or didn’t) do something. In our personal lives, we hold people accountable all the time. In AI product development, it means that an employee who is accountable will take responsibility of results and outcomes – they won’t presume that this is purely the concern of management.
2. **Data governance** – the process of managing the availability, usability, integrity, confidentiality and security of the data within enterprise systems, and also increasingly across international boundaries. Internally, it is based on data standards and policies that control data usage; internationally, it is based on regulations including Sarbanes–Oxley Act, Basel I, Basel II, Health Insurance Portability and Accountability Act and General Data Protection Regulation, many of which have been enshrined in law (e.g. Data Protection Act 2018 in the UK). Effective data governance ensures that data are consistent and trustworthy and don’t get misused; for example, the unauthorised use of personal medical data without the owner’s consent for a purpose that was not agreed at the time the data were collected. In an insurance setting, miscollected medical data may have a detrimental effect on a person gaining medical insurance.

3. **Design for all** – effectively universal design, that is, designing a system or service that all can use regardless of gender, age, disability or impairment, rather than the creation of individual systems suited to just part of society. It is effectively design for human diversity, social inclusion and equality.

4. **Governance of AI autonomy (human oversight)** – the framework of guidance and controls that need to be put in place to ensure that AI systems or products do no harm. For example, provision of advice that a suitably qualified person needs to be sitting in the driving seat of an autonomous vehicle while it is in operation; or that a medical diagnostic tool should be not left to decide treatment of a patient without the oversight of a suitably qualified medical practitioner.

5. **Non-discrimination** – the fair and unprejudiced treatment of different categories of people. Typically, it is managed through a non-discrimination policy that ensures equal employment opportunity without discrimination or harassment on the basis of race, colour, religion, sexual orientation, gender identity or expression, age, disability, marital status, citizenship, national origin, genetic information, or any other characteristic.

6. **Respect for (and enhancement of) human autonomy** – autonomy can be defined as the ability of the person to make his or her own decisions, and describes the ability to think, feel, make decisions and act on his or her own.21 Autonomy includes three facets, consisting of behavioural, emotional and cognitive self-government. In a medical AI context, it is the respect for a patient’s personal autonomy to be considered and built into an AI system of service. This is one of many fundamental ethical principles in medicine, and is usually associated with allowing or enabling patients to make their own decisions about which healthcare interventions they will or will not receive.

7. **Respect for privacy** – the ability of an individual, group or organisation to seclude themselves or information about themselves from others. Concerns around privacy often overlap with data security, and may be covered within controls established for that domain. In terms of the growing use of AI within the public environment, it may include the use of surveillance cameras within part of a city making use of facial recognition software without informing people that they are being monitored. Another example may be around mobile phone providers installing geolocation monitoring and data collection apps within software updates.

8. **Robustness** – requires that AI systems be developed with a preventative approach to risks and in a manner such that they reliably behave as intended while
minimising unintentional and unexpected harm and preventing unacceptable harm. An example would be for an AI system to operate as designed within agreed parameters and not learn and apply new behaviours outside those parameters that may harm humans either physically, socially or financially.

9. **Safety** – linked to robustness, safety must become a key element of AI to ensure human acceptance. Rigorous techniques must be established for building safe and trustworthy AI systems and establishing confidence in their behaviour and robustness. This will facilitate their successful adoption in wider society. Ethical design, transparency and testing are key to ensuring safe AI, especially advanced AI systems, to ensure that they are aligned with core human values.

10. **Transparency** – defined in AI terms as being able to explain what is happening within the AI system. How has the AI come to a particular conclusion or decision or action? Without transparency, how can we ensure fairness and remove racial and gender bias, for example? We require increased explicability from AI products and systems, without which trust cannot be established and AI products and systems will not become fully accepted by society.

In 2019, the European Commission’s High-Level Expert Group on Artificial Intelligence released *Ethics Guidelines for Trustworthy AI* and consolidated these 10 requirements down to seven:

1. **Human agency and oversight** – including fundamental rights, human agency and human oversight.
2. **Technical robustness and safety** – including resilience to attack and security, fall-back plan and general safety, accuracy, reliability and reproducibility.
3. **Privacy and data governance** – including respect for privacy, quality and integrity of data, and access to data.
4. **Transparency** – including traceability, explainability and communication.
5. **Diversity, non-discrimination and fairness** – including the avoidance of unfair bias, accessibility and universal design, and stakeholder participation.
6. **Societal and environmental wellbeing** – including sustainability and environmental friendliness, social impact, society and democracy.
7. **Accountability** – including auditability, minimisation and reporting of negative impact, trade-offs and redress.

Moving on from principles and requirements, we can now look at how these could be established and implemented. We can use both technical and non-technical methods to achieve trustworthy AI systems that support the principles and requirements mentioned.

Using technical methods, we have:

1. **Ethics and rule of law by design (X-by-design)** – there are numerous ethics boards around the world, some nationally based and some internationally, others corporately or societally based, all establishing various AI ethics and guidelines. There are laws around health and safety, privacy and data protection. These can be incorporated within the design of AI systems rather than bolted on as an afterthought.
For example, ‘Privacy by Design’ means that organisations need to consider privacy at the initial design stages and throughout the complete development process of new AI products or services that involve processing personal data.

2. **Architectures for trustworthy AI** – establishing trustworthy AI systems architecture, with clear boundaries or limitations of what an AI learning system is capable of, will ensure appropriate controls are introduced to prevent unexpected behaviours or actions that may be detrimental to the user of such a system.

3. **Testing and validating** – the process of evaluating AI software during or at the end of the development process to determine whether it satisfies specified business requirements. Validation and testing also ensure that the AI product actually meets the client’s needs.

4. **Traceability and auditability** – traceability is the ability to look at why something occurred and also the effect a request or action has had. Auditability is using traceability to ensure that the requirement or conformity has been met. It can also be used to identify deviations or non-conformance in an AI product or service.

5. **Explanation (XAI research)** – explainable AI is the ability of a human to understand and communicate the results or a solution produced by an AI system. This is compared to a ‘black box’ solution where even the designers may not be able to interpret why the results were produced.

6. **CE mark** – applicable within the European Economic Area (EEA), a CE mark is an administrative mark for AI and non-AI products which indicates that the manufacturer has checked conformity with health, safety and environmental protection standards. It allows for the free movement of products within the EEA. The CE marking is also found on products sold outside the EEA that have been manufactured to EEA standards.

Non-technical methods include:

1. **Regulation** – the act of controlling an AI product or system either through a law, rule or order. For example, stating that an AI product cannot be sold without a support contract, or an AI service must not be used by minors.

2. **Standardisation** – the AI product or service conforms to a documented standard. For example, all models are produced to the same standard to ensure quality conformance.

3. **Accountability governance** – in AI ethics this translates to answerability, blameworthiness, liability and the expectation of account-giving (i.e. being called to account for one’s actions). It is basically where the buck stops.

4. **Code of conduct** – an agreement on rules of behaviour as compared to a code of ethics (which is a set of principles distinguishing right from wrong). A code of conduct normally outlines appropriate actions and regulations for employers, employees or members, as well as what is not acceptable or expected and the legal consequences of breaking the rules.

5. **Education and awareness to foster an ethical mindset** – this plays an important role, both to ensure that knowledge of the potential impact of AI systems is widespread and to make people aware that they can participate in shaping the societal development of such systems.
6. **Stakeholder and social dialogue** – increasingly we are expected to understand the importance of stakeholder engagement and dialogue to help establish social and environmental reporting. We can address some of the key issues surrounding the use of AI and difficulties involved in the implementation and deployment of AI systems by establishing a stakeholder engagement and social dialogue process. For example, in gaining societal agreement to deploy facial recognition and tracking systems, which are contentious, whereas car number plate recognition systems are widely accepted.

7. **Diversity and inclusive design teams** – the creation of teams with diversity of experience, perspective, creativity and inclusion of race, ethnicity, gender, age, sexual identity, ability/disability and location help to create truly inclusive designs.

### 1.3.14 Trustworthy AI – continual assessment and monitoring

How do we check and confirm that we have hit our AI ethics principals? We do it through ongoing:

- accountability;
- data governance;
- non-discrimination;
- design for all;
- governance of AI autonomy;
- respect for privacy;
- respect for (and enhancement of) human autonomy;
- robustness;
- reliability and reproducibility;
- accuracy through data usage and control;
- fall-back plan;
- safety;
- transparency;
- purpose;
- traceability: method of building the algorithmic system; method of testing the algorithmic system.

So, in summary, here is a very simple mental checklist:

1. Adopt an assessment list for trustworthy AI.
2. Adapt an assessment list to your specific use case.
3. Remember that trustworthy AI is not about ticking boxes, it is about improved outcomes through the entire life cycle of the AI system.
1.4 SUSTAINABLE AI

As well as trustworthy and robust AI we need to ensure that what we are doing in AI is sustainable. In 2015, the United Nations produced its Agenda for Sustainable Development and identified 17 sustainable development goals (SDGs) and 169 targets. Amazingly, it was agreed by 193 countries, and had effectively produced ‘The framework for good’.

An open paper produced by a number of sustainability experts from around the world in 2019 identified that AI may support the achievement of 128 targets (76 per cent) across all SDGs, but AI may also negatively impact or inhibit 58 (34 per cent) of those targets.22 The paper further broke down the targets within the three pillars of sustainable development: society, economy and environment.

We should recognise that AI has a role to play in sustainability and that the current capabilities of AI, including automating routine and repetitive tasks, analysing Big Data and bringing intelligence and learning to various processes, have expanded and continue to expand our capacity to understand and solve complex, dynamic and interconnected global challenges such as the SDGs.

At the time of writing, with just 10 years remaining (up to 2030) to achieve the ambitions outlined in the United Nations SDGs, we should use AI to achieve those SDG targets we know it can contribute to, while at the same time identifying the impact and inhibitions that AI will have and mitigating the negative effects where possible. AI may also trigger inequalities that could act as inhibitors on some of the SDGs. For example, in helping to deliver the achievement of the ‘No Poverty’ SDG, AI can help to identify areas and pockets of poverty and their causes; however, it may also increase the gap between poor and rich countries and between the poor and rich within each country.

Guided by the United Nations SDGs, it is now up to all of us – businesses, governments, academia, multilateral institutions, non-governmental organisations (NGOs) and others – to work together to accelerate and scale the development and use of AI in the interest of achieving the SDG targets, while at the same time recognising possible negative effects and addressing them.

We also need to consider that not all stakeholders in the development and implementation of AI will be interested in sustainability, and may well have their own measures, such as return on investment.

Rogue actors may develop malicious AI. We must be on guard and develop intelligent cyber protection. This can be considered as the development of robust trustworthy AI.

1.4.1 Political, Economic, Sociological, Technological, Legal and Environmental (PESTLE)

Although the pursuit of SDGs may be seen as altruistic, that does not say that it should be left to others to pursue. When an organisation is designing, developing or implementing
an AI solution, they should consider at least undertaking a PESTLE analysis. This is a framework to analyse the key factors influencing an organisation from the outside.

PESTLE is a strategic and systematic way of learning about key factors that influence a range of stakeholders. It can be used by an organisation for tactical decision making and often is.

1.5 MACHINE LEARNING – A SIGNIFICANT CONTRIBUTION TO THE GROWTH OF ARTIFICIAL INTELLIGENCE

In this section, we need to make the distinction between machine learning and artificial intelligence. This is no easy task until we remind ourselves that AI is a universal subject that can help any pursuit of learning from experience to achieve a goal. ML is only part of this – the distinction starts with the definitions of an AI agent and ML. Tom Mitchell’s definition is the one usually quoted; it relates well to digital computing that is part of all our lives.

AI is about intelligent entities, or AI agents, interacting with an environment to achieve a goal. Not just one entity but multiple. AI is about humans and machines, how they interact, how they learn, what they experience. AI asks us to be explicit about humans achieving their goals. Can we build machines that we can work with to improve us as humans? AI asks hard questions of consciousness, philosophy, ethics, science. It deals with complex phenomena that we do not currently have the machines to explore. AI of the future is about how humans and machines will co-exist. In 2019 Stuart Russell released his latest book on human compatible AI, *Human Compatible: Artificial Intelligence and the Problem of Control*, which captures the true nature of AI and how, used wisely, we can benefit from a future of humans and machines.

As we are coming to realise the benefit of digital computing ML, this will unlock the potential of AI in other areas:

- engineering and building intelligent entities;
- medicine and improving health- and social care;
- business analytics, and others.

When we simply think about products, representing the world in a digital simulation, or what is (at the moment) ones and zeros, it is not ideal. Digital computation has its limitations, and we need better machines. We have run out of digital processing power. AI is much more than high-performance computing and programming of machines that deal with ones and zeros – digitally simulated pizza and weather don’t taste of anything and do not get us wet.

The AI machines of the future will incorporate digital computers, but, when we think about it, it’s actually hard to represent the mathematical operations we need. We are limited by the processing power, energy and accuracy of today’s technology. And a result, we can only concentrate on narrow ML, focused on specific well-defined tasks or goals. These tasks are defined by Tom Mitchell’s often quoted definition (p. 2).
A computer program is said to learn from experience, \( E \), with respect to some class of tasks, \( T \), and performance measure, \( P \), if its performance at tasks in, \( T \), as measured by, \( P \), improves with experience, \( E \).

The examples given of these types of tasks are playing games such as chess, checkers and draughts. Modern-day games include simulation games and very advanced strategic, well-engineered games like Go; these types of games can be explicitly defined on a digital computer. Practical examples in the real world include optimising where aircraft park at an airport or the logistics of delivering a parcel; again, a reasonably well-ordered and engineered environment in which ML can optimise something.

ML is focused on explicitly defining a problem that can be solved on a computer. These problems can be complicated, non-linear and statistical. In simple terms, if we are to use ML in our AI, we must be able to represent our problem mathematically and in such a way that it can be solved by a machine. Today, we typically use digital computers. However, quantum, analogue, optical and biological computers are on their way.

Digital ML has become very popular recently with the success of convolutional deep neural networks. These numerical techniques have given us an understanding of how the human mind solves problems. So, when we think about ML, we can think about an AI agent learning from data. These machines are now so good at playing games that they can beat the world champion at these types of games.

The AI agent is much more than a narrowly focused computer program. ML works on data in the computer. We must work really hard to think of this as an interface with actuators and sensors in an environment. It is sensible to think, here, that ML learns from data in a computational environment. This is a good starting point to opening up the world of AI. AI is about humans and machines working together to achieve goals. We might even go on to say that ML is an AI enabler setting the foundation for a future of humans and machines.

1.6 SUMMARY

This chapter has been a whistle-stop tour of human and AI intelligence. It has drawn out key concepts such as agents, ethics and machine learning and their historical context. It has highlighted that humans are subjective and objective conscious beings, something AI is nowhere near achieving. With our introductory knowledge of human intelligence and the progress of AI over the past few centuries, we can look in more detail at ethical AI for human good and how this may evolve into products and services in an age of humans and machines.
ARTIFICIAL INTELLIGENCE FOUNDATIONS
Learning from experience
Andrew Lowe, Steve Lawless

In alignment with BCS AI Foundation and Essentials certificates, this introductory guide provides the understanding you need to start building artificial intelligence (AI) capability into your organisation.

You will learn how AI is being utilised today to support products, services, science and engineering, and how it is likely to be used in the future to balance the talents of humans and machines. You will explore robotics and machine learning within the context of AI, and discover how the challenges AI presents are being addressed. You will delve into the theory behind AI and machine learning projects, examining techniques for learning from data, the use of neural networks and why algorithms are so important in the development of a new AI agent or system.

- Engaging introduction to artificial intelligence, including robotics and machine learning
- Explains how AI agents and systems are developed
- Contains real-life case studies of AI in use now and looks at how AI may be used in the future
- Supports BCS AI Foundation and Essentials certificates

ABOUT THE AUTHORS
Andrew Lowe is Director of iA42 Ltd. Steve Lawless is CEO of Purple Griffon, IT service management training and consultancy specialists. Together Andrew Lowe and Steve Lawless developed the BCS AI Foundations and AI Essentials certificates.

This is a great book for executives to understand how AI may fit into their company and the types of benefits that they may gain and some of the pitfalls they may fall into.

Dr Paul Edward Mort, B.Eng. (Hons), PhD, MBA, MIMechE CEng FNucl, Innovation Lead for Special Nuclear Materials, Sellafield Ltd

A 'must have' guide in preparation for the BCS AI Foundations exam, this book will not only assist with exam success, but will capture your imagination to the possibilities in front of us when reimagining the human vs machine relationship.

Richard Webber, Digital Transformation Programme Manager, Ravensbourne University London

Successfully brings the beginner up to speed and is written with such enthusiasm that the more complex topics become manageable.

Rosie Sheldon, Senior Trainer, TSG-training

You might also be interested in:

MANAGING DATA QUALITY
A practical guide

MASTERING DIGITAL BUSINESS
How powerful combinations of disruptive technologies are enabling the next wave of digital transformation

PROJECT-MANAGEMENT FOR IT-RELATED PROJECTS

Cover photo: iStock © Lovegtrv6

ISBN 978-1-78017-328-7
Paperback available