EU privacy and data protection law applied to AI: unveiling the legal problems for individuals

Häuselmann, A.N.

2 Artificial Intelligence (AI)

This chapter aims to answer Subquestion 1, namely, what AI is and what AI disciplines exist.[50] It starts with existing definitions of AI (Section 2.1) and then provides an overview of the AI disciplines that seem to be the most problematic ones from a privacy and data protection perspective (Section 2.2). These disciplines include machine learning (Section 2.2.1), natural language processing (Section 2.2.2), computer vision (Section 2.2.3), affective computing (Section 2.2.4) and automated reasoning (Section 2.2.5). Section 2.3 answers Subquestion 1.

2.1 Definitions of AI

There is no officially agreed definition of Artificial Intelligence (AI). AI covers a wide range of concepts and terms, making it difficult to define. Available definitions often involve ambiguous terms such as 'thinking', 'learning' and 'intelligence'. In 1968, Minsky defined AI as 'the science of making machine do things that would require intelligence if done by men'.[51] Bellman defined AI in 1978 as 'the automation of activities that we associate with human thinking, activities such as decision-making, problem solving, learning, creating, game playing, and so on'.[52] Nilsson described AI as 'activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment.'[53] Russell and Norvig organised definitions of AI into four categories: a) thinking humanly, b) acting humanly, c) thinking rationally and d) acting rationally.[54] According to Munakata, AI involves abilities such as 'inference based on knowledge, reasoning with uncertain or incomplete information, various forms of perception and learning, and applications to problems such as control, prediction, classification, and optimization'.[55] More recent definitions are the ones adopted by the Organisation for Economic Co-operation and Development (OECD) and the National Institute of Standards and Technology (NIST). The OECD defines an AI system as a 'machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment'.[56] NIST defines AI as a 'branch of computer science devoted to developing data processing systems that performs functions normally associated with human intelligence, such as reasoning, learning, and self-improvement'.[57]

[50] A modified version of this chapter was published in Bart Custers, Eduard-Fosch Villaronga (eds) *Law and Artificial Intelligence* (Asser Press 2022). See Andreas Häuselmann, 'Disciplines of AI: An Overview of Approaches and Techniques' 43-70.
[51] Marvin Minsky, *Semantic Information Processing* (MIT Press 1968).
[52] Richard Bellman, *An Introduction to Artificial Intelligence: Can computers think*? (Boyd & Faser 1978) 3
[53] Nils J Nilsson, *The Quest for Artificial Intelligence: A History of Ideas and Achievements* (Cambridge University Press 2010).
[54] Stuart Russel, Peter Norvig, *Artificial Intelligence, A Modern Approach* (3rd edn, Pearson Education 2016) 2.
[55] Toshinori Munakata, *Fundamentals of the New Artificial Intelligence* (2nd edn, Springer 2008) xx.
[56] See < https://oecd.ai/en/wonk/ai-system-definition-update > accessed 8 February 2024.
[57] See < https://csrc.nist.gov/topics/technologies/artificial-intelligence > accessed 8 February 2024.

The field of AI may be divided into narrow and general AI. Narrow AI refers to systems that are able to solve a specific problem or performing a specific task. For an example on a narrow AI system, one can refer to IBM's 'Deep Blue' chess-playing computer. Deep Blue defeated the reigning world champion in chess, Garry Kasparov, in 1997.[58] This example indicates that computers can perform better than humans. However, this holds only true for a narrow domain, such as playing chess. General AI aims to build machines that generally perform on a human level and have a 'human-level' skillset. To achieve this goal, such a system must be able to mimic the functioning of the human brain in the most important aspects.[59] Unlike with narrow AI, general AI arguably has not been achieved yet despite rapid developments, for instance, ChatGPT. Although AI found its 'birth' at the Dartmouth Summer Research Project on AI in the summer of 1956 in New Hampshire,[60] there are many open challenges. According to Shi, AI research is still in its first stage since no breakthrough progress has been achieved for some key challenges such as common sense knowledge representation and uncertain reasoning.[61] Therefore, current AI systems must be considered examples of 'narrow' AI. However, computing power has become more affordable; the computers have become faster and contain larger memories. This led to the 'summer of AI' and it seems reasonable to expect major developments in the field of AI.

In his famous paper, called 'Computing Machinery and Intelligence',[62] Turing proposed the 'Imitation Game', which has later become known as the 'Turing test'.[63] Turing offered his test as a sufficient condition for the existence of AI.[64] This test involves three actors: (A) a machine, (B) a human and (C) another human called the interrogator (see Figure 1.1). In the Turing test, the human interrogator (C) stays in a room apart from the other two actors (A) and (B). The human interrogator knows the machine (A) and human (B) by labels (X) and (Y)[65] and therefore does not know which label is (A) or (B).[66] The object of the test is for the interrogator (C) to determine which of the other two actors is the human and which is the machine[67] by asking (X) and (Y) questions which they must answer.[68] In other words, the human interrogator engages in conversation with either a human or an AI natural language program which are both hidden from view. If the human interrogator cannot reliably

[58] https://www.livescience.com/59065-deep-blue-garry-kasparov-chess-match-anniversary.html, accessed 8 February 2024.
[59] Kevin Warwick, *Artificial Intelligence: The basics* (Routledge 2012) 65.
[60] Ronald R Kline, 'Cybernetics, Automata Studies, and the Dartmouth Conference on Artificial Intelligence' (2011) 4, EEE Computer Society, 5.
[61] Zhongzhi Shi, *Advanced Artificial Intelligence* (World Scientific 2011) 18.
[62] Alan Mathison Turing, 'Computing Machinery and Intelligence' (1950) Vol LIX Iss 236 Mind 433-460.
[63] Chris Bernhardt, *Turing's Vision: The Birth of Computer Science* (MIT Press 2016) 157.
[64] Stan Franklin, 'History, motivations, and core themes' in Frankish Keith and Ramsey William (eds) *The Cambridge Handbook of Artificial Intelligence* (2014) 17.
[65] Alan Mathison Turing, 'Computing Machinery and Intelligence' (1950) Vol LIX Iss 236 Mind 433-460.
[66] Chris Bernhardt, *Turing's Vision: The Birth of Computer Science* (MIT Press 2016) 157.
[67] Alan Mathison Turing, 'Computing Machinery and Intelligence' (1950) Vol LIX Iss 236 Mind 433-460.
[68] Chris Bernhardt, *Turing's Vision: The Birth of Computer Science* (MIT Press 2016) 157.

distinguish between the human and the program/machine, (artificial) intelligence is ascribed to the program.[69]

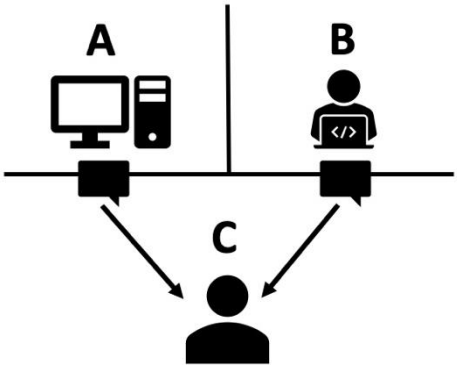

Figure 1.1 Illustration of the Turing test created by the author.

There are plenty of definitions for AI, which involve ambiguous terms such as those already mentioned. In this thesis, AI refers to adaptive machines that can autonomously execute activities and tasks that require capabilities usually associated with humans. 'Autonomously' in this sense means that the machine has the ability to make its *own* decisions and perform tasks on the designer's behalf.[70] 'Adaptive' refers to the machine's ability to learn from, and adapt to its environment in order to preserve its autonomy in dynamic environments.[71] Adaptivity is very important, since only a machine that *learns* will succeed in a vast variety of environments.[72] Learning in this context corresponds to 'adapt' the performance according to previously made experiences based on statistics and probability calculations.[73] This definition aligns well with the ones adopted by the OECD and NIST.[74]

2.2 AI disciplines

Since AI covers a broad range of concepts, this research will pay particular attention to AI disciplines which could be problematic in the light of the fundamental rights to privacy and data protection. These AI disciplines are coloured blue in Figure 1.2.[75] The remaining disciplines (white) will not be discussed in this thesis.

[69] Stan Franklin, 'History, motivations, and core themes' in Frankish Keith and Ramsey William M. (eds) *The Cambridge Handbook of Artificial Intelligence* (2014) 17, 18.
[70] Eduardo Alonso, 'Actions and agents' in Frankish Keith and Ramsey William M. (eds) *The Cambridge Handbook of Artificial Intelligence* (2014) 235, 236.
[71] Ibid 235.
[72] Stuart Russel, Peter Norvig, *Artificial Intelligence, A Modern Approach* (3rd edn, Pearson Education 2016) 39.
[73] Stefan Strauß, 'From Big Data to Deep Learning: A Leap Towards Strong AI or Intelligentia Obscura' (2018) 2 (3), Big Data and Cognitive Computing <https://www.mdpi.com/2504-2289/2/3/16> accessed 14 January 2019, 7.
[74] See < https://oecd.ai/en/wonk/ai-system-definition-update > and < https://csrc.nist.gov/topics/technologies/artificial-intelligence > respectively, accessed 8 February 2024.
[75] This figure shall not be considered as a complete overview of all AI disciplines, but serves as an illustrative overview for this thesis.

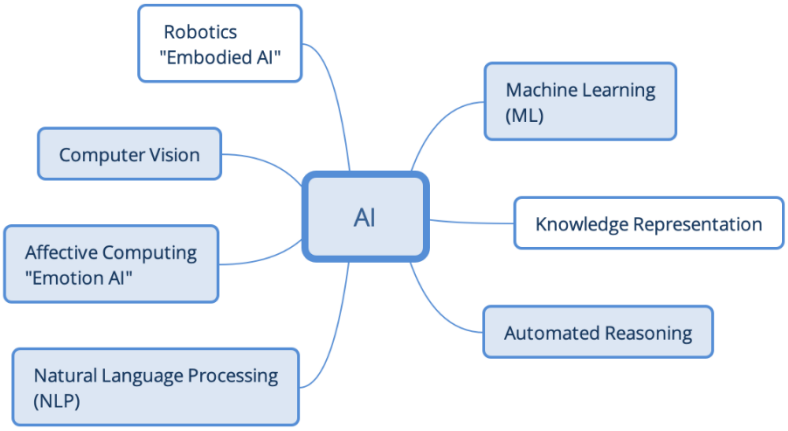

Figure 1.2 Graph created by the author outlining the AI disciplines inspired by Russel/Norvig[76] and slightly adjusted by adding the field of affective computing.

Methods of AI that combine AI with Robotics, i.e. 'Embodied Artificial Intelligence', are out of scope of this thesis. Applications of Embodied AI such as driverless vehicles, surgical robots and companions pose different questions such as liability issues or ethical issues in the context of robot-human interactions.[77] However, these questions are not in the scope of this research.

AI systems need to translate input into information or knowledge so that it can be processed to select output (action).[78] The discipline of AI research commonly referred to as knowledge representation focusses on the computers capabilities to store what it knows and hears.[79] Since research in this discipline of AI focusses on conceptual issues[80] not related to privacy and data protection, it will not be discussed here. However, the subfield of automated reasoning, which is a fundamental part of knowledge representation, will be discussed due to its implications on automated decision-making.

2.2.1 Machine learning (ML)

ML may be considered a discipline or one of the tools of AI.[81] I follow the former approach in this thesis and acknowledge that ML is often combined with other AI disciplines. Computer science has traditionally aimed to manually program computers. ML however aims to have computers program themselves based on experience.[82] In other words, the goal of ML is to adapt to new circumstances and to detect and extrapolate patterns.[83] Murphy defines ML as 'a set of methods that can

[76] Stuart Russel, Peter Norvig, *Artificial Intelligence, A Modern Approach* (3rd edn, Pearson Education 2016) 2, 3.
[77] Cándido García Molyneux, Rosa Oyarzabal, 'What Is a Robot (Under EU Law)?' (2018) Vol 1 RAIL: The Journal of Robotics, AI & Law 11, 12.
[78] Stan Franklin, 'History, motivations, and core themes' in Frankish Keith and Ramsey William M. (eds) *The Cambridge Handbook of Artificial Intelligence* (2014) 24.
[79] Stuart Russel, Peter Norvig, *Artificial Intelligence, A Modern Approach* (3rd edn, Pearson Education 2016) 2.
[80] E.g. the issue of whether or not to represent knowledge, Franklin Stan, 'History, motivations, and core themes' in Frankish Keith and Ramsey William M (eds) *The Cambridge Handbook of Artificial Intelligence* (2014) 24, 25.
[81] Vijay Kotu, Bala Deshpande, *Data Science* (2nd edn Elsevier 2019) 2.
[82] Tom M. Mitchell, 'The discipline of Machine Learning' (2006) 1 <http://www.cs.cmu.edu/~tom/pubs/MachineLearning.pdf> accessed 8 February 2024.
[83] Stuart Russel, Peter Norvig, *Artificial Intelligence, A Modern Approach* (3rd edn, Pearson Education 2016) 2.

automatically detect patterns in data, and then use the uncovered patterns to predict future data or to perform other kinds of decision making under uncertainty'.[84] ML can simply be described as the set of computational methods that use experience to improve its performance or to make accurate predictions.[85] This is achieved by using ML algorithms, algorithms that learn from experience.[86] Put simply, an algorithm is typically a numerical process that consists of a sequence of well-defined steps leading to the solution of a particular type of problem.[87] Experience refers to the data from the past available to the algorithm for analysis.[88] Learning in this context is about making computers modify or adapt their performance (actions) so that these actions become more *accurate*.[89] ML uses data-driven methods, combining fundamental concepts in computer science with approaches from statistics, probability and optimisation.[90] In fact, the probabilistic approach in ML is closely related to the field of statistics, but differs slightly in terms of its emphasis and terminology. The probabilistic approach is particularly helpful for handling ambiguous cases.[91] The main goal of ML is to generate accurate predictions for unseen data and to design efficient algorithms to produce these predictions.[92]

Before the specific *kind* of ML called deep learning (DL) will be discussed in Section 2.2.1.4, some of the most widely used ML *methods* will be elaborated on first in Sections 2.2.1.1 and 2.2.1.3. These methods are called supervised, unsupervised and reinforcement learning. In practice, the distinction between supervised and unsupervised learning is not always clear-cut. Therefore, semi-supervised learning creates a continuum between supervised and unsupervised learning: The algorithm is provided with a few labelled examples (supervised learning) but also has the task to uncover hidden patterns and structures in the data (unsupervised learning).[93] Another method deployed in ML is reinforcement learning (RL). RL is becoming increasingly relevant, in particular in natural language processing, a discipline of AI which aims to enable computers to process human language (see Section 2.2.2).

2.2.1.1 Supervised machine learning

Supervised ML aims to learn a mapping from input *x* to output *y,* given a labelled set of input-output pairs called the *training set* or training data. It can be used to make predictions on *new* input through generalisation.[94] Generalisation refers to the ability of the algorithm to categorise new examples that

[84] Kevin P Murphy, *Machine Learning: A Probabilistic Perspective* (MIT Press 2012) 1.
[85] Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar, *Foundations of Machine Learning* (MIT Press 2012) 1.
[86] Ian Goodfellow, Yoshua Bengio, Aaron Courville, *Deep Learning* (MIT Press 2016) 97 <www.deeplearningbook.org> accessed 8 February 2024.
[87] Yadolah Dodge, 'Algorithm' in: *The Concise Encyclopedia of Statistics* (Springer New York 2006) 1-2.
[88] Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar, *Foundations of Machine Learning* (MIT Press 2012) 1.
[89] Steven Marsland, *Machine Learning: An Algorithmic Perspective* (2nd edn Chapman & Hall 2015) ch 1.2.1.
[90] Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar, *Foundations of Machine Learning* (MIT Press 2012) 1.
[91] Kevin P Murphy, *Machine Learning: A Probabilistic Perspective* (MIT Press 2012) 1, 4.
[92] Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar, *Foundations of Machine Learning* (MIT Press 2012) 2.
[93] Stuart Russel, Peter Norvig, *Artificial Intelligence, A Modern Approach* (3rd edn, Pearson Education 2016) 695.
[94] Kevin P Murphy, *Machine Learning: A Probabilistic Perspective* (MIT Press 2012) 3.

differ from the ones used during the training phase.[95] In the supervised ML approach, the learning algorithm receives several examples, each *labelled* with the correct label (training data). Consider, for example, several labelled pictures with different animals (lions, horses, and cows). The goal is that the algorithm automatically recognises the correct label for the training data and *predicts* the value of unseen (unlabelled) inputs.[96] In other words, the aim is that the algorithm *generalises* accurately by producing a model that can classify input *not seen* during training.[97] The user who provides the correct labels to the algorithm is the teacher, knowing for each input the correct output. Therefore, this is called 'supervised' learning: the algorithm learns under the supervision and guidance of the teacher.[98] To measure the accuracy of the model generated by the algorithm, the teacher provides the algorithm with a set of examples that are *different* from the set of training.[99] Hence, the teacher feeds the algorithm with new pictures containing lions, horses and cows and evaluates the accuracy of the model, namely, whether the algorithm recognised the animals correctly. The algorithm learns by adjusting the relevant parameters so that the model makes the most accurate predictions on the data.[100]

There are basically two techniques used for supervised machine learning: classification and regression.[101] As indicated by its name, classification refers to situations where the predicted attribute is categorical, and regression applies to situations where the predicted attribute is numeric.[102] Classification orders data into exhaustive and exclusive groups or classes on the basis of their similarity. Consequently, all data can only be assigned to one class.[103] The example with the animal referred to the classification technique. Regression is suitable when the prediction to be made by the algorithm should be a numerical value. Regression could be described as a statistical approach that is used to identify the relationship between variables.[104] Therefore, the regression technique could be used to predict the number of people likely to click on an online advertisement based on the ad content and the user's previous surfing history. Other real-world examples using regression are predicting stock market prices given current market conditions or predicting the age of a viewer watching a given video on YouTube.[105]

[95] Christopher M Bishop, *Pattern Recognition and Machine Learning* (Springer 2006) 2.
[96] Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar, *Foundations of Machine Learning* (MIT Press 2012) 7.
[97] Ethem Alpaydin, *Machine Learning: The New AI* (3rd edn MIT Press 2016) 39.
[98] Toshinori Munakata, *Fundamentals of the New Artificial Intelligence* (2nd edn Springer 2008) 38.
[99] Stuart Russel, Peter Norvig, *Artificial Intelligence, A Modern Approach* (3rd edn, Pearson Education 2016) 695.
[100] Ethem Alpaydin, *Machine Learning: The New AI* (3rd edn MIT Press 2016) 39.
[101] Michele Usuelli, *R machine learning essentials* (Packt Publishing 2014) 155.
[102] Ibid 154.
[103] Toon Calders, Bart Custers, 'What is Data Mining and How Does it Work?' in Bart Custers et al. (eds) *Discrimination and Privacy in the Information Society* (Springer 2013) 32.
[104] However, note that decision tree regression would not be considered as traditional statistics.
[105] Kevin P Murphy, *Machine Learning: A Probabilistic Perspective* (MIT Press 2012) 9.

2.2.1.2 Unsupervised machine learning

Unlike supervised ML, the algorithm only receives *un*labelled training data.[106] That means that the algorithm is not told what the desired output is for each form of input and unsupervised ML does not require a human expert to manually label the data.[107] Due to the fact that there is no external comparison between actual and ideal output by the teacher, this approach is called unsupervised: There are no correct answers available.[108] Therefore, the algorithm tries to discover patterns in the input even though no explicit feedback is supplied.[109] The goal of unsupervised ML is to identify associations and patterns among a set of input data and categorise them accordingly.[110] It can be difficult to quantitatively evaluate the performance of the model, since there are no labelled examples available.[111] Two branches of techniques used for unsupervised learning are clustering and dimensionality reduction.[112]

Clustering in this context means dividing detected patterns into groups or clusters. Similar patterns are placed in the same group, while all others are put in different groups.[113] Simply put, clustering refers to the partition of unlabelled items into homogeneous regions.[114] Clusters may overlap, while classifications do not (see Section 2.2.1.1). Clustering is particularly performed to analyse very large data sets. A common example is to use clustering in the context of social network analysis, where the clustering algorithm tries to identify 'communities' within large groups of people.[115] The same applies to e-commerce, where users are clustered into groups based on their purchasing or online behaviour, which enables online shops to send customised targeted ads to each group.[116]

Dimensionality reduction aims to represent data with fewer dimensions[117] and is useful to project high-dimensional data to a lower dimensional subspace to capture the 'essence' of the data.[118] By reducing the dimensions, *hidden patterns* and *structures* in the data may be observed, and noninformative features are discarded. Dimensional representations often produce *better predictive accuracy* because they focus on the essence of the object and filter out non-essential features.[119] Dimensionality reduction is commonly used to pre-process digital images, in computer vision tasks[120] (see

[106] Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar, *Foundations of Machine Learning* (MIT Press 2012) 7.
[107] Kevin P Murphy, *Machine Learning: A Probabilistic Perspective* (MIT Press 2012) 9.
[108] Toshinori Munakata, *Fundamentals of the New Artificial Intelligence* (2nd edn Springer 2008) 38.
[109] Stuart Russel, Peter Norvig, *Artificial Intelligence, A Modern Approach* (3rd edn, Pearson Education 2016) 694.
[110] Hastie Trevor, Tibshirani Robert, Friedman Jerome, *The Elements of Statistical Learning* (2nd edn 2008) xi; Steven Marsland, *Machine Learning: An Algorithmic Perspective* (2nd edn Chapman & Hall 2015) ch 1.3.
[111] Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar, *Foundations of Machine Learning* (MIT Press 2012) 7.
[112] Michele Usuelli, *R machine learning essentials* (Packt Publishing 2014) 164.
[113] Toshinori Munakata, *Fundamentals of the New Artificial Intelligence* (2nd edn Springer 2008) 72.
[114] Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar, *Foundations of Machine Learning* (MIT Press 2012) 2.
[115] Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar, *Foundations of Machine Learning* (MIT Press 2012) 2.
[116] Kevin P Murphy, *Machine Learning: A Probabilistic Perspective* (MIT Press 2012) 11.
[117] Ethem Alpaydin, *Introduction to Machine Learning* (4th edn MIT Press 2020) 137, 138.
[118] Kevin P Murphy, *Machine Learning: A Probabilistic Perspective* (MIT Press 2012) 11.
[119] Ibid, 12.
[120] Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar, *Foundations of Machine Learning* (MIT Press 2012) 2.

Section 2.3) and applied in natural language processing (see Section 2.2.2), e.g., for acoustic signals.[121]

2.2.1.3 Reinforcement learning (RL)

Reinforcement learning (RL) is a distinct method in ML that differs from supervised and unsupervised ML approaches. In RL, the algorithm interacts with its environment and the method is inspired by behavioural psychology.[122] RL algorithms modify or acquire new behaviours incrementally and use *trial-and-error* experience without requiring complete knowledge or control of the environment.[123] Unlike supervised learning, RL learns with a 'critic' who does not instruct the algorithm what to do, but rather provides it with feedback in the form of a reward or punishment.[124] The reward depends on the correctness of the decision (the action by the agent).[125] In RL, the decision-maker is called the agent which interacts with everything outside the agent, called the environment. The agent and environment interact continuously: the agent selects actions, and the environment responds to these actions and presents new situations to the agent.[126] The agent has no prior knowledge of what action to take; it learns from interaction with the environment.[127] The object of the agent is to maximise its reward over a course of interactions with the environment.[128] Therefore, the agent uses the received feedback to update its knowledge so that it learns to perform actions that return the highest reward.[129]

An illustrative example is a machine (agent) that learns to play chess. The chessboard is the environment of the agent that must decide over a sequence of actions, namely, 'moves' on the chessboard (environment) to achieve a certain goal, namely, winning the game. In RL, the agent evolves and learns while analysing the consequences of its actions with the feedback received from the environment.[130] This is different from the unsupervised ML approach, where no feedback is distributed. RL also differs from supervised ML because the agent does not learn from the initially labelled training data, but from the interaction with the environment based on feedback in the form of a punishment or reward.[131] Combining it with deep learning techniques has made 'deep RL' increasingly successful in addressing challenging sequential decision-making problems such as mastering the game 'Go'[132] or

[121] Kevin P Murphy, *Machine Learning: A Probabilistic Perspective* (MIT Press 2012) 11.
[122] Ian Goodfellow, Yoshua Bengio, Aaron Courville, *Deep Learning* (MIT Press 2016) 104 <www.deeplearningbook.org> accessed 8 February 2024.
[123] Vincent François-Lavet et al, 'An Introduction to Deep Reinforcement Learning' (2018), Vol. 11, No. 3-4 Foundations and Trends in Machine Learning, 2, 15.
[124] Ethem Alpaydin, *Introduction to Machine Learning* (4th edn MIT Press 2020) 570.
[125] Andries P Engelbrecht, *Computional Intelligence – An Introduction* (2nd edn John Wiley & Sons 2007) 83.
[126] Zhongzhi Shi, *Advanced Artificial Intelligence* (World Scientific 2011) 365.
[127] Andries P. Engelbrecht, *Computational Intelligence – An Introduction* (2 edn John Wiley & Sons 2007) 83.
[128] Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar, *Foundations of Machine Learning* (MIT Press 2012) 8.
[129] Ethem Alpaydin, *Introduction to Machine Learning* (4th edn MIT Press 2020) 570.
[130] Zhongzhi Shi, *Advanced Artificial Intelligence* (World Scientific 2011) 362.
[131] Sumit Das et al., 'Applications of Artificial Intelligence in Machine Learning: Review and Prospect' (2015), Vol. 115, No. 9 International Journal of Computer Applications 31, 32.
[132] See <https://www.deepmind.com/research/highlighted-research/alphago> accessed 8 February 2024.

beating the world's top professionals in poker.[133] Its adaptive capabilities make RL very suitable for interactive applications. For example, deep RL is applied for dialogue systems and conversational agents, in particular for digital assistants and chatbots.[134] The most impressive and current example is ChatGPT provided by OpenAI. ChatGPT is a large language model trained to produce text. It was optimised by using reinforcement larning with human feedback.[135] Deep RL seems to possess promising potential for real-world applications such as robotics, self-driving cars, finance and smart grids.[136] Current ML applications based on the supervised method for natural language processing and speech recognition require vast amounts of labelled training data. This issue could be eliminated by applying deep RL methods.[137]

2.2.1.4 Artificial Neural Networks and deep learning

The human brain consists of a very large number of processing units called neurons.[138] These neurons have an output fibre called an axon and a terminal fibre called a synapse. The axons split up and connect to several dendrites, which are the input pathways of other neurons through the junction terminal synapse.[139] Because the neurons of the human brain are connected, it is called a neural network. Figure 1.3 shows a typical biological neuron.

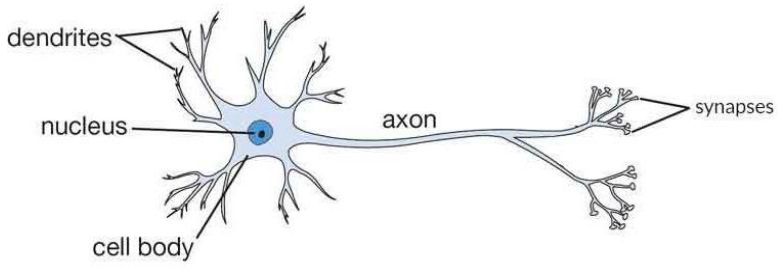

Figure 1.3 Biological neuron illustrated by Navdeep Singh.[140] Used with permission.

Although it is not entirely clear how the neural network of human brains actually works, it is considered to be the fundamental functional source of intelligence, which includes perception, learning and cognition.[141] The characteristic of a neural network is that the neurons operate in parallel and transfer

[133] See <https://www.nature.com/articles/d41586-019-02156-9> accessed 8 February 2024.
[134] Iulian Serban et al. 'A Deep Reinforcement Learning Chatbot' (2017) 1 <https://arxiv.org/pdf/1709.02349.pdf> accessed 8 February 2024.
[135] See FAQs about ChatGPT provided by OpenAI: < https://help.openai.com/en/articles/6783457-what-is-chatgpt > accessed 8 February 2024.
[136] Vincent François-Lavet et al., 'An Introduction to Deep Reinforcement Learning' (2018) Vol. 11 No. 3-4 Foundations and Trends in Machine Learning 3.
[137] Deng Li and Liu Yang, 'Epilogue: Frontiers of NLP in the Deep Learning Era' in Deng Li and Liu Yang (eds) *Deep learning in natural language processing* (Springer 2018) 316.
[138] Ethem Alpaydin, *Machine Learning: The New AI* (3rd edn MIT Press 2016) 86.
[139] Tommy Chow, Siu-Yeung Cho, *Neural Networks and Computing: Learning Algorithms and Applications* (Imperial College Press 2007) 2.
[140] Navdeep Singh Gill, 'Overview of Artificial Neural Networks and its application' <https://www.xenonstack.com/blog/artificial-neural-network-applications/> accessed 8 February 2024.
[141] Toshinori Munakata, *Fundamentals of the New Artificial Intelligence* (2nd edn, Springer 2008) 7.

information among themselves over the synapses so that the neurons are connected and influence each other.[142] The brain is believed to learn by examples and experience and to be highly capable of adapting to external changes.[143]

A single biological neuron would be too simple to make decisions like humans do. Similarly, a single artificial neuron would not be able to cope with challenging decision-making and prediction processes. Hence, to unleash the full potential of artificial neurons, they must operate in parallel and transfer information among themselves. That is why researchers such as Rumelhart and others in 1986 attempted to design artificial neural networks (ANN) with the aim to allow an arbitrarily connected neural network to develop an internal structure that is appropriate for a particular task.[144] ANNs can be simply described as an abstract model that is inspired by knowledge of the inner workings of the human brain that can be programmed on a computer. ANNs consist of artificial neurons and interconnections similar to the human brain. The network receives input, performs internal processes such as the activation of the neurons and finally yields output.[145] However, ANNs are generally not designed to be realistic models of the human brain. The neural perspective on deep learning is motivated by two main ideas: first, that the brain provides an example that intelligent behaviour is possible; and second, that it is possible to create machine learning models that shed light on the principles of the brain and human intelligence.[146] The pattern of connections between the artificial neurons is called the architecture or topology of the ANN and consists of distinct layers of neurons. The layers depend on the model used.[147] Each of the layers has a certain number of neurons which is usually determined by a specific application problem the model aims to solve. An example of a deep ANN is given in Figure 1.4.

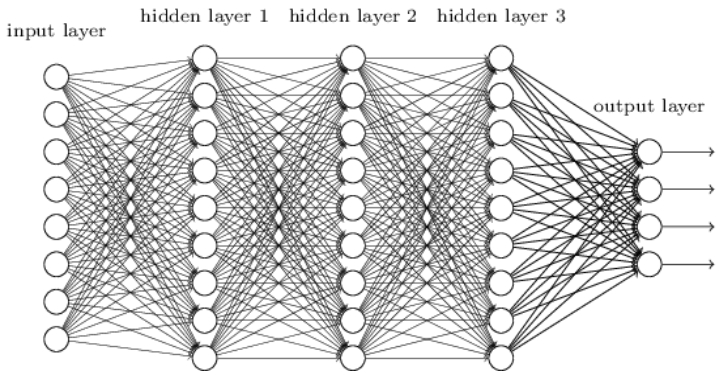

Figure 1.4 Example of a deep artificial neural network illustrated by Michael Nielsen.[148] Used with permission.

[142] Ethem Alpaydin, *Machine Learning: The New AI* (3rd edn MIT Press 2016) 86.
[143] Tommy Chow, Siu-Yeung Cho, *Neural Networks and Computing: Learning Algorithms and Applications* (Imperial College Press 2007) 2.
[144] David Rumelhart, Geoffrey Hinton, Ronald Williams 'Learning representations by backpropagating errors' (1986) Vol. 323 Nature 533.
[145] Toshinori Munakata, *Fundamentals of the New Artificial Intelligence* (2nd edn, Springer 2008) 3, 7.
[146] Ian Goodfellow, Yoshua Bengio, Aaron Courville, *Deep Learning* (MIT Press 2016) 13 <www.deeplearningbook.org> accessed 8 February 2024.
[147] Toshinori Munakata, *Fundamentals of the New Artificial Intelligence* (2nd edn, Springer 2008) 9.
[148] Michael Nielsen, 'Why are deep neural networks hard to train' in: *Neural Networks and Deep Learning* (Determination Press 2015) <http://neuralnetworksanddeeplearning.com/chap5.html> accessed 8 February 2024.

Generally, there are one input layer, one output layer and *any number* of hidden layers. Neurons of the input layer are connected to the neurons of the hidden layer through edges, and the neurons of the hidden layer(s) are connected to the output layer. A weight is associated to each edge. The input layer (see on the left side of Figure 1.4) consists of neurons that receive their input directly from the data, and its function is to merely send out input signals to the hidden layer neurons; it does not compute anything.[149] The hidden layer then applies computation methods to the inputs that depend on the model used for the neural network, transforming the received inputs to something the output layer can use. Hidden means that the values in these layers are not given in the data, but the model has the task of determining which concepts are useful for explaining the relationships in the observed data.[150] It then sends its output to the next layer, in the present case, to hidden layer 2, which sends it to hidden layer 3 and subsequently to the output layer (see the right side of Figure 1.4). Subsequently, the role of the output layer is to produce the output of the entire network. The output of ANNs can then be used to extract a prediction or a decision.

Deep learning (DL) is a particular kind of ML that represents the world as a nested hierarchy of concepts.[151] The human brain seems to execute many levels of processing with increasing levels of abstraction.[152] DL seems to resemble this by computing more abstract concepts in terms of less abstract ones.[153] Most of the models used for supervised and unsupervised ML have a simple two-layer architecture.[154] This is different with DL models, which use many different layers. Approaches in DL feed a large set of input data into the ANN that produces successive transformations of the input data, where each hidden layer combines the values in its preceding layer and learns more complicated functions of the input.[155] Then, the final transformation predicts the output.[156] The deep learning approach avoids the requirement that the human operator must specify all the knowledge which the computer requires. Deep learning solves this by enabling the computer to build complex concepts out of simpler concepts. When illustrating the approach in a graph by building the concepts on top of each other, that graph is deep, with many layers. Therefore, the approach is called deep learning (see Figure 1.4).[157] DL draws inspiration from many fields, especially from linear algebra and probabilistic statistics. Foundation models, namely models that are trained on broad data using self-supervision

[149] Toshinori Munakata, *Fundamentals of the New Artificial Intelligence* (2nd edn, Springer 2008) 10.
[150] Ian Goodfellow, Yoshua Bengio, Aaron Courville, *Deep Learning* (MIT Press 2016) 6 <www.deeplearningbook.org> accessed 8 February 2024.
[151] Ian Goodfellow, Yoshua Bengio, Aaron Courville, *Deep Learning* (MIT Press 2016) 8 <www.deeplearningbook.org> accessed 8 February 2024.
[152] Kevin P Murphy, *Machine Learning: A Probabilistic Perspective* (MIT Press 2012) 95.
[153] Ian Goodfellow, Yoshua Bengio, Aaron Courville, *Deep Learning* (MIT Press 2016) 8 <www.deeplearningbook.org> accessed 8 February 2024.
[154] Kevin P Murphy, *Machine Learning: A Probabilistic Perspective* (MIT Press 2012) 995.
[155] Ethem Alpaydin, *Machine Learning: The New AI* (3rd edn MIT Press 2016) 104.
[156] Yoav Goldberg, *Neural Network Methods in Natural Language Processing* (Morgan & Claypool Publishers 2017) 2.
[157] Ian Goodfellow, Yoshua Bengio, Aaron Courville, *Deep Learning* (MIT Press 2016) 1, 5 <www.deeplearningbook.org> accessed 8 February 2024.

and that can be adapted to a wide range of tasks, are based on deep neural networks.[158] Typical examples of foundation models include large language models (LLMs) as introduced in the AI discipline natural language processing (Section 2.2.2).

Interestingly, achievements in modern DL have been made with an astonishingly small number of neurons contained in the ANNs when compared with neural networks of the human brain. Although today's ANNs are considered quite large from a computational perspective, they are smaller than the neural networks of relatively primitive animals such as frogs. Goodfellow, Bengio and Courville, leading scholars in the field, predict that ANNs will not reach the same number of neurons as the human brain possesses before the 2050s unless new technologies enable faster scaling.[159]

However, most current DL models lack reasoning and explanatory capabilities, making them vulnerable to produce unexplainable outcomes. Despite the recent success of DL, DL methods based on ANN generally lack interpretability.[160] Foundation models and LLMs are no exception.[161] Interpretability remains a challenge due to the hierarchical and nonlinear structure of ANNs and the central concept in DL called connectionism. With deep learning models, each artificial neuron works *independently* by computing a relatively simple task, and therefore *partially* contributes to the output produced by the ANNs.[162] ANNs produce output based on the central concept in DL called *connectionism*, where the idea is that a large number of simple computational units (artificial neurons) achieve intelligent behaviour when networked together.[163] Consequently, combining the characteristic of artificial neurons to work independently with the concept of connectionism leads to a situation where thousands or hundreds of thousands of artificial neurons work in parallel in an ANN with hidden layers to jointly calculate certain output.[164] Hence, it seems neither possible to understand which artificial neuron contributed to a distinct part of the output nor to understand what happened in the intermediate (hidden) layers of the ANN.[165] In other words, it is not possible to extract any underlying rules that may be implied by the DL model.[166] This holds even true for DL algorithms using the supervised learning method, where the algorithm cannot learn without being given correct sample patterns. Therefore, even if an ANN has successfully been trained to achieve its goal, the many

[158] Rishi Bommasani et al, 'On the Opportunities and Risks of Foundation Models' (2022) Center for Research on Foundation Models Standford University 1, 3 < https://arxiv.org/pdf/2108.07258.pdf > accessed 8 February 2024.
[159] Ian Goodfellow, Yoshua Bengio, Aaron Courville, *Deep Learning* (MIT Press 2016) 21 <www.deeplearningbook.org> accessed 8 February 2024.
[160] Deng Li and Liu Yang, 'A Joint Introduction to Natural Language Processing and Deep Learning' in Deng Li and Liu Yang (eds) *Deep learning in natural language processing* (Springer 2018) 11, 12.
[161] Melanie Mitchell, David C Krakauer, 'The debate over understanding in AI's large language models' (2023) Vol 120 Iss 3 PNAS 1-5.
[162] Toshinori Munakata, *Fundamentals of the New Artificial Intelligence* (2nd edn, Springer 2008) 44.
[163] Ian Goodfellow, Yoshua Bengio, Aaron Courville, *Deep Learning* (MIT Press 2016) 16 <www.deeplearningbook.org> accessed 8 February 2024.
[164] Ethem Alpaydin, *Machine Learning: The New AI* (3rd edn MIT Press 2016) 155.
[165] Ethem Alpaydin, *Machine Learning: The New AI* (3rd edn MIT Press 2016) 155.
[166] Toshinori Munakata, *Fundamentals of the New Artificial Intelligence* (2nd edn, Springer 2008) 44.

numeric values of the weights produced by the model do not have a meaning to the supervisor.[167] Clearly, the model is parameterised by all these weights, but it remains unclear how these weights have been calculated and to what extent the various input variables contributed to the outcome. ANNs in use can be *updated dynamically* as new data are fed into the network.[168] Subsequently, this updates the weights produced by the model because they are learnt from experience. These updates contribute to further challenges regarding the interpretability of DL approaches.[169]

DL is well suited to deal with complex sensor data such as input from cameras and microphones that proved to be difficult to process when using conventional computational methods.[170] This applies in particular to cognitive tasks which include natural language processing and speech recognition or face recognition, which are discussed below.[171] Current research in DL attempts to decode speech directly from the human brain. Such approaches record the activity in the cortex to decode the characteristics of the produced speech.[172] State-of-the-art deep neural network models arguably contribute to an improved overall accuracy in speech reconstruction from neural recordings in the human auditory cortex.[173] The short-term goal of these research projects is to help individuals that are unable to communicate due to injuries or neurodegenerative disorders by creating a synthesised version of their voice that can be controlled by the activity of their brain speech centres.[174] However, the long-term goal of this could be much broader and very different. Facebook announced that it wants to 'build a non-invasive, wearable device that lets people type simply by imagining themselves talking.'[175]

2.2.2 Natural language processing (NLP)

Natural language processing (NLP), a subfield of AI, aims to give computers the ability to process human language. This interdisciplinary field comprises many concepts and methods such as speech and language processing, human language technology, natural language processing, computational

[167] Toshinori Munakata, *Fundamentals of the New Artificial Intelligence* (2nd edn, Springer 2008) 12, 25, 35.
[168] A production model has fixed weights after training. To continuously update weights is possible, but by no means necessary.
[169] Paul De Laat, 'Algorithmic Decision-Making based on Machine Learning from Big Data: Can Transparency restore Accountability' (2017) Vol. 31 Issue 4 Philosophy & Technology 14 < https://link.springer.com/article/10.1007%2Fs13347-017-0293-z> accessed 8 February 2024.
[170] Tommy Chow, Siu-Yeung Cho, *Neural Networks and Computing: Learning Algorithms and Applications* (Imperial College Press 2007), 1/2.; Mitchell Tom T., *Machine Learning* (Mc-Graw-Hill 1997) 95.
[171] Tommy Chow, Siu-Yeung Cho, *Neural Networks and Computing: Learning Algorithms and Applications* (Imperial College Press 2007) 2.
[172] David A. Moses et al, 'Real-time decoding of question-and-answer speech dialogue using human cortical activity' (2019) 10 Nature Communication <https://www.nature.com/articles/s41467-019-10994-4.pdf> accessed 8 February 2024.
[173] Minda Yang et al, 'Speech Reconstruction from Human Auditory Cortex with Deep Neural Networks' (Interspeech Conference, Dresden, September 2015) 1124 < https://dblp.org/db/conf/interspeech/interspeech2015.html > accessed 8 February 2024.
[174] Nicholas Weiler, 'Breakthrough device translates brain activity into speech' (University of California, 25 April 2019) < https://www.universityofcalifornia.edu/news/synthetic-speech-generated-brain-recordings> accessed 8 February 2024.
[175] 'Imagining a new interface: Hands-free communication without saying a word' (Tech@Facebook, 30 March 2020) <https://tech.fb.com/imagining-a-new-interface-hands-free-communication-without-saying-a-word/> accessed 8 February 2024.

linguistics, and speech recognition and synthesis.[176] NLP includes both the generation and understanding of natural language.[177] The advances in NLP have led to the development of large language models (LLMs). LLMs are advanced language models with massive parameter sizes (billions to trillions)[178] and strong learning capabilities.[179] These models can perform various NLP tasks, such as translation, text summarisation, and question-answering.[180] ChatGPT is the current prime example.

From an engineering perspective, NLP intends to develop novel practical applications to facilitate interactions between computers and human languages.[181] Current NLP systems require large amounts of labelled data.[182] Speech recognition is a typical application of NLP, and its aim is to *automatically transcribe* the sequence of spoken words. It may be defined as the process of converting a speech signal to a sequence of words by means of an algorithm implemented by a computer program.[183] In particular, speech recognition does not concern *understanding* but is simply responsible to *convert* language from spoken words to text form.[184] The observable 'physical' signal of natural language is called text in symbolic form, and its counterpart is the speech signal, that is, the continuous correspondence of spoken texts.[185] Speech recognition is based on the acoustic signal captured by a microphone as input. The classes are the words that can be uttered. A word is a sequence of phonemes that are the basic speech sounds.[186] Therefore, speech recognition converts phonemes (speech signal) into text. A specific challenge in speech recognition is that different people pronounce the same word differently due to factors related to age, gender or accent, which makes it more difficult to recognise the words.[187] Another challenge is that a common conversational utterance involves multiple queries with disfluencies such as pauses and hesitations. However, current NLP systems embedded in virtual assistants typically focus on 'unnatural' and one-sided interactions without hesitation or disfluency. For this reason, speech recognition involving conversiational speech is a challenging task.[188]

[176] Daniel Jurafsky, James H Martin, *Speech and Language Processing* (2 edn, Pearson Education Limited 2014) 1.
[177] Stan Franklin, 'History, motivations, and core themes' in Frankish Keith and Ramsey William M. (eds) *The Cambridge Handbook of Artificial Intelligence* (2014) 26.
[178] Melanie Mitchell, David C Krakauer, 'The debate over understanding in AI's large language models' Vol 120 Iss 3 PNAS 1-5; Rishi Bommasani et al, 'On the Opportunities and Risks of Foundation Models' (2022) Center for Research on Foundation Models Standford University 1, 3 < https://arxiv.org/pdf/2108.07258.pdf > accessed 8 February 2024.
[179] Yupeng Chang et al, 'A Survey on Evaluation of Large Language Models' (2023) 1, 4 < https://arxiv.org/pdf/2307.03109.pdf > accessed 8 February 2024.
[180] Yiheng Liu et al, 'Summary of ChatGPT-Related research and perspective towards the future of large language models' (2023) Vol 1 Meta-Radiology 1 – 14.
[181] Deng Li and Liu Yang, 'A Joint Introduction to Natural Language Processing and Deep Learning' in Deng Li and Liu Yang (eds) *Deep learning in natural language processing* (Springer 2018) 1.
[182] Deng Li and Liu Yang, 'Epilogue: Frontiers of NLP in the Deep Learning Era' in Deng Li and Liu Yang (eds) *Deep learning in natural language processing* (Springer 2018) 316.
[183] Abhang Priyanka, Gawali Bharti, Mehrotra Suresh, *Introduction to EEG- and speech-based emotion recognition* (Elsevier Inc 2016) 13.
[184] Gokhan Tur et al, 'Deep Learning in Conversational Language Understanding' in Deng Li and Liu Yang (eds) *Deep learning in natural language processing* (Springer 2018) 24.
[185] Ibid 24.
[186] Ethem Alpaydin, *Machine Learning: The New AI* (3rd edn MIT Press 2016) 67.
[187] Ibid.
[188] Shuo-ziin Chang et al, 'Turn-Taking Prediction for Natural Conversational Speech' (Interspeech Conference Incheon, September 2022) <https://arxiv.org/pdf/2208.13321.pdf> accessed 8 February 2024.

Speech signals cannot only reveal the intended message, but also the *identity of the speaker* because the ways in which prosodic characteristics are manifested in speech disclose important information regarding the identity of the speaker.[189] Prosody refers to the study of the intonational and rhythmic aspects of language.[190] Systems in the domain of speaker verification are capable of using the voice of an individual in order to identify an unknown person (speaker identification), verify the identity of a person (speaker verification) and classify specific characteristics like age or gender (speaker classification).[191] Text-based verification of an individual through voice analysis is technically possible with a very short text such as 'Ok Google', which takes approximately 0.6 seconds if uttered by an individual.[192] Hence, speaker identity is embedded in the speaker's voice and can be recognised using automatic speaker recognition systems, which apply DL approaches.[193]

Current research in speech recognition focusses on emotion recognition from speech signals, a major subject in human-computer interaction. This research focusses on how speech is modulated when a speaker's emotion changes from neutral to another emotional state. For example, it has been observed that speech in anger or happiness shows longer utterance duration and higher pitch and energy value with deep length. [194] Speech emotion recognition may be used for various areas, such as call centres, smart devices or self-driving cars.[195] A real-world application of affective computing (AC) that aims to derive emotional states from speech is Amazon's 'Halo' wearable, which analyses voice tones to detect user emotions.[196] The recent success in NLP and speech recognition has been powered by using the DL approach in ML, currently with supervised ML methods such as classification as described in Section 2.2.1.1. Therefore, the current bottleneck of these approaches is that they require large amounts of labelled data and lack reasoning abilities. However, it is tried to overcome this bottleneck by applying the unsupervised learning paradigm and particularly deep RL methods in NLP and speech recognition.[197] Deep learning has been successfully applied to real-world tasks in AI, in particular in

[189] Leena Mary, *Extraction of Prosody for Automatic Speaker, Language, Emotion and Speech Recognition* (2nd edn Springer 2019) 1, 8.
[190] Daniel Jurafsky, James H Martin, *Speech and Language Processing* (2 edn, Pearson Education Limited 2014) 238.
[191] Soufiane Hourri, Jamal Kharroubi, 'A deep learning approach for speaker recognition' (2020) Vol. 23 Iss. 1 International Journal of Speech and Technology 123.
[192] Gregor Heigold et al, 'End-to-End Text-Dependent Speaker Verification' (2015) < https://arxiv.org/pdf/1509.08062.pdf > accessed 8 February 2024.
[193] Leena Mary, *Extraction of Prosody for Automatic Speaker, Language, Emotion and Speech Recognition* (2nd edn Springer 2019) 7. See precedent references regarding DL approaches.
[194] Abhang Priyanka, Gawali Bharti, Mehrotra Suresh, *Introduction to EEG- and speech-based emotion recognition* (Elsevier Inc 2016) 14, 105.
[195] See services of the company audeering <https://www.audeering.com/> accessed 8 February 2024.
[196] Alex Hern, 'Amazon's Halo wristband: the fitness tracker that listens to your mood' *The Guardian* (London, 28 August 2020) <https://www.theguardian.com/technology/2020/aug/28/amazons-halo-wristband-the-fitness-tracker-that-listens-to-your-mood> accessed 8 February 2024; Austin Carr, 'Amazon's New Wearable Will Know If I'm Angry. Is That Weird?' *Bloomberg* (New York, 31 August 2020) < https://www.bloomberg.com/news/newsletters/2020-08-31/amazon-s-halo-wearable-can-read-emotions-is-that-too-weird > accessed 8 February 2024.
[197] Deng Li and Liu Yang, 'Epilogue: Frontiers of NLP in the Deep Learning Era' in Deng Li and Liu Yang (eds) *Deep learning in natural language processing* (Springer 2018) 316.

speech recognition as a part of the virtual personal assistants such as Google Assistant, Amazon Alexa, Microsoft Cortana or Apple Siri.[198]

2.2.3 Computer vision (CV)

Computer vision (CV) is a subfield of AI devoted to perceive objects, i.e. the automated understanding of visual images and comprises many fields of applications.[199] The goal of object detection is to detect all instances of objects from a known class, such as people, cars or faces in an image.[200] CV can also be described as the science and technology of machines that 'see', which refers to the ability of the machine to extract information from an image necessary to solve a task.[201] CV aims to infer properties from the observed visual data, which originate from a variety of sensors such as cameras, laser scans, etc.[202] CV algorithms reconstruct the properties of one or more images, such as such as shape, illumination and colour distributions. Researchers in computer vision develop mathematical techniques to recover the three-dimensional shape and appearance of objects in imagery. Real-world applications include optical character recognition (OCR) for automatic number plate recognitions (of vehicles), medical imaging for preoperative and intra-operative imagery, automotive safety to detect unexpected obstacles such as pedestrians on the street, surveillance to monitor intruders and fingerprint recognition for automatic access authentication.[203]

CV techniques are also currently used to identify individuals based on their gait. Biometric research implies that gait, i.e. the manner in which individuals walk, constitutes a unique identifier like a fingerprint or iris.[204] The biometrics necessary for gait identification may be captured in public places and from a distance in a rather ubiquitous manner. Methods used for identification are model-based approaches which consider the human body or its movements to acquire gait parameters (e.g., step dimensions, cadence, human skeleton, body dimensions) as well as model-free approaches that acquire gait parameters by that rely on gait dynamics and the measurement of geometric representations such as silhouettes.[205]

[198] Gokhan Tur et al, 'Deep Learning in Conversational Language Understanding' in Deng Li and Liu Yang (eds) *Deep learning in natural language processing* (Springer 2018) 23.
[199] Stuart Russel, Peter Norvig, *Artificial Intelligence, A Modern Approach* (3rd edn, Pearson Education 2016) 3, Stan Franklin, 'History, motivations, and core themes' in Frankish Keith and Ramsey William M. (eds) *The Cambridge Handbook of Artificial Intelligence* (2014) 26.
[200] Yali Amit, Pedro Felzenszwalb, 'Object Detection' in Katsushi Ikeuchi (ed) *Computer Vision – A Reference Guide* (Springer 2014) 537.
[201] Sota R. Yoshida, *Computer Vision* (Nova Science Publisher 2011) vii.
[202] Varun Jampani, 'Learning Inference Models for Computer Vision' (Dissertation, Universität Tübingen 2016) 1.
[203] Richard Szeliski, *Computer Vision: Algorithms and Applications* (Griets David, Schneider Fred Springer eds 2011) 3, 5.
[204] Ale Sokolova, Anton Konushin 'Methods of Gait Recognition in Video' (2019) Vol 45 No 4 Programming and Computer Software 213.
[205] Jure Kovač, Vitomir Štruc, Peter Peer 'Frame-based classification for cross-speed gait recognition' (2019) Vol 78 Multimedia Tools and Applications 5621, 5622.

Another real-world example is Amazon Go. Amazon Go is a checkout-free grocery store which is equipped with state-of-the-art cameras and sensors. Amazon Go is powered by computer vision, DL and sensor fusion[206] in order to track shoppers and their purchases. Sensor fusion exploits the best features of sensors (for example, cameras and small Bluetooth radio transmitters called 'beacons') installed in a given environment. It is particularly helpful in situations where the sensors themselves are not self-sufficient to achieve a certain goal, for example, comprehensive and precise tracking of shoppers.[207] In Amazon Go stores, shoppers enter by scanning an Amazon Go smartphone app and sensors track items that the shoppers take from the shelves. Once picked up, the items are automatically charged to the Amazon accounts of the shoppers when they leave the store. Where Amazon Go's inventory system cannot detect the object the user removed from the shelf, the system 'may consider past purchase history' of the user.[208]

Face recognition is one of the CV applications of particular relevance for this thesis. Section 2.2.3.1 introduces face recognition and Section 2.2.3.2 explains face recognition applications applying deep learning.

2.2.3.1 Face recognition

Face recognition refers to the technology capable of identifying or verifying the identity of subjects in images or videos based on biometric data.[209] It is one of the major biometric technologies and has become increasingly relevant due to the rapid advances in image capture devices and the availability of huge amounts of face images on the web.[210] Unlike other biometric identification methods, such as iris recognition (which requires individuals to get significantly close to a camera), face recognition can be used from a distance and in a covert manner.[211] Therefore, the range of potential applications for face recognition is wide because it can be easily deployed.[212]

[206] See https://www.amazon.com/b?ie=UTF8&node=16008589011 and Vasilios Mavroudis, Michael Veale 'Eavesdropping Whilst You're Shopping: Balancing Personalisation and Privacy in Connected Retail Spaces' (Living in the Internet of Things Conference, London, March 2018) 6 <https://ieeexplore.ieee.org/document/8379705> accessed 8 February 2024.
[207] For example, cameras offer a high level of precision, but might be too expensive to cover the whole shop. Beacons are not self-sufficient to provide tracking data for customer analysis, but can cover a wider operational range. Combined by means of sensor fusion, the sensors allow precise consumer path tracking. See Mirco Sturari et al, 'Robust and affordable retail customer profiling by vision and radio beacon sensor fusion' (2016) Vol. 81 Pattern Recognition Letters 30, 31, 40.
[208] Dilip Kumar et al. 'Detecting item interaction and movement' US Patent Number US 10268983 (Assignee: Amazon Technologies, Inc.) April 2019 at 9 <https://patentimages.storage.googleapis.com/01/0b/6e/de57009f5670ae/US20150019391A1.pdf > accessed 8 February 2024.
[209] Daniel Trigueros, Li Meng, Margaret Hartnett, 'Face recognition: From Traditional to Deep Learning Methods' (2018) 1 <https://arxiv.org/pdf/1811.00116.pdf > accessed 8 February 2024.
[210] Stan Li, Anil Jain, 'Introduction' in Li Stan, Jain Anil (eds) *Handbook of Face Recognition* (2nd edn, Springer 2011) 1.
[211] Stan Li, Anil Jain, 'Introduction' in Li Stan, Jain Anil (eds) *Handbook of Face Recognition* (2nd edn, Springer 2011) 1; Daniel Trigueros, Li Meng, Margaret Hartnett, 'Face recognition: From Traditional to Deep Learning Methods' (2018) 1 <https://arxiv.org/pdf/1811.00116.pdf > accessed 8 February 2024.
[212] Daniel Trigueros, Li Meng, Margaret Hartnett, 'Face recognition: From Traditional to Deep Learning Methods' (2018) 1 <https://arxiv.org/pdf/1811.00116.pdf > accessed 8 February 2024.

Face recognition systems operate in a face verification (authentication) and/or face identification (recognition) mode. The former involves a one-on-one match that compares a query face image of the person whose identity is claimed (e.g., for self-serviced immigration clearance using E-passports). The latter involves one-to-many matching, which compares a query face image against multiple face images in a database to associate the identity of the query face. Usually, finding the most similar face is not sufficient and a confidence threshold is specified. Therefore, only those faces whose similarity score is above the threshold are reported.[213] Face recognition systems are usually built on four building blocks:

1. Face detection, which finds the position of a face in an image
2. Face normalisation, which normalises the face geometrically and photometrically
3. Face feature extraction performed to extract salient information which is useful to distinguish faces such as reference points located at fixed locations in the face (e.g., position of eyes, nose, lips)
4. Face matching, where extracted features from the input face are matched against one or many of the enrolled faces in the database[214]

The facial features used for the third building block may be grouped into two classes of features: continuous and discrete. Continuous features are real valued numbers and are extracted using distances and angles between facial landmarks such as forehead height, eyebrow length, nose height, chin height, ears length, mouth length etc. Discrete features represent a finite number of categories, for example, the shape of the eyebrow or nose root width.[215] Figure 1.5 provides an example of such features.

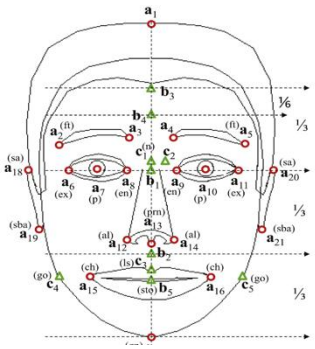

Figure 1.5 Face layout illustrated by Tome et al.[216] with examples of facial features extracted by using distances and angles between facial landmarks such as eyebrows, eyes, nose and lips. Used with permission.

[213] Stan Li, Anil Jain, 'Introduction' in Li Stan, Jain Anil (eds) *Handbook of Face Recognition* (2nd edn, Springer 2011) 3.
[214] Ibid, 4; Daniel Trigueros, Li Meng, Margaret Hartnett, 'Face recognition: From Traditional to Deep Learning Methods' (2018) 1 < https://arxiv.org/abs/1811.00116 > accessed 8 February 2024.
[215] Pedro Tome et al., 'Facial soft biometric features for forensic face recognition' (2015) Vol 257 Forensic Science International 271, 273.
[216] Ibid.

2.2.3.2 DL and face recognition

Current face recognition applications use DL methods based on convolutional neural networks (CNN) which are trained with very large datasets.[217] A CNN is a specific kind of neural network for processing data that has a known grid-like typology. For example, image data can be thought of as a 2D grid of pixels. As the name indicates, a CNN employs a mathematical operation called convolution, which is a specialised kind of linear operation.[218] Notably, the performance of a face recognition system largely depends on a variety of factors such as illumination, facial pose, expression, age span, hair and motion.[219] Whereas the building blocks of face recognition systems and the general architecture of the ANN are predetermined by the developer of the system, the ANN itself decides how to create the optimal score for determining similarity in the face matching building block mentioned in Section 2.2.3.1. Therefore, it remains often unclear how the similarity score is calculated by the ANN, even to the developer of the system.[220] Another issue is that face recognition systems perform poorly in recognising individuals of different ethnicities. For example, Hewlett Packard face recognition software could not recognise dark-coloured faces as faces.[221] A 'passport robot' in New Zealand rejected the passport picture of an Asian man because the 'subject's eyes are closed' although his eyes were open.[222]

However, face recognition systems are widely used in commercial applications and consumer products with built-in AI capabilities. Examples are cars with on-board cameras to deploy biometric identification and monitor driving behaviour[223] or connected retail spaces.[224] Furthermore, there is a trend to improve face recognition systems with the ability to monitor and analyse the emotions in real-time based on extracted biometric data and facial expressions. The gained knowledge is then used to build specific customer profiles.

[217] Daniel Trigueros, Li Meng, Margaret Hartnett, 'Face recognition: From Traditional to Deep Learning Methods' (2018) 1 <https://arxiv.org/abs/1811.00116> accessed 8 February 2024.
[218] Ian Goodfellow, Yoshua Bengio, Aaron Courville, *Deep Learning* (MIT Press 2016) 326 <www.deeplearningbook.org> accessed 8 February 2024.
[219] Stan Li, Anil Jain, 'Introduction' in Li Stan, Jain Anil (eds) *Handbook of Face Recognition* (2nd edn, Springer 2011) 3.
[220] Yana Welinder, Aeryn Palmer, 'Face Recognition, Real-Time Identification, and Beyond' in Selinger Evan, Polonetsky Jules, Tene Omer (eds) *The Cambridge Handbook of Consumer Privacy* (Cambridge University Press 2018) 104.
[221] Frederik Zuiderveen Borgesius, 'Discrimination, artificial intelligence, and algorithmic decision-making'(2019) Report for the Anti-discrimination department of the Council of Europe, 17 <https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73> accessed 8 February 2024.
[222] Regan James, 'New Zealand passport robot tells applicant of Asian descent to open eyes' (2016) *Reuters* <https://www.reuters.com/article/us-newzealand-passport-error/new-zealand-passport-robot-tells-applicant-of-asian-descent-to-open-eyes-idUSKBN13W0RL> accessed 8 February 2024.
[223] See <https://visagetechnologies.com/application-fields/driver-monitoring/> accessed 8 February 2024
[224] See <https://www.einfochips.com/blog/facial-recognition-in-retail-enhance-in-store-customer-experience-and-improve-retailer-operations/> accessed 8 February 2024.

2.2.4 Affective computing (AC)

Affective computing (AC), sometimes called 'emotion AI', is computing that relates to, arises from or influences emotion.[225] AC is a scientific and engineering endeavour inspired by psychology, neuroscience, linguistics and related areas.[226] Affective states are considered to be experiential phenomena such as emotions and moods.[227] Emotions form an important part of human intelligence and daily live, be it for decision-making, social interaction, perception or learning. In other words, emotions play a pivotal role in functions considered essential to intelligence.[228] Picard, the pioneer in the field of AC, therefore, concludes that if computers are to be genuinely intelligent, they too should have emotional capabilities.[229] In this thesis, the focus lies on affect detection from facial expressions and speech, since they may be easily deployed compared to more invasive approaches that include measurement of physiological factors such as cardiac activity (heart rate) or skin conductance (sweat).

The following sections elaborate on affect detection from facial expressions (Section 2.2.4.1), speech (Section 2.2.4.2) and discuss multimodal approaches in which different methods of AC are combined to detect emotions (Section 2.2.4.3).

2.2.4.1 Facial expressions

Facial expressions are probably the most natural way humans express their emotions.[230] According to Darwin's evolutionary theory of emotions, emotion expressions help in regulating the social interaction and increase the likelihood of survival.[231] Due to the developments in technology, it is possible to detect facial information automatically in real-time, for example, with the use of a simple video camera. However, automatic detection of emotions derived from facial expressions and their interpretation is not simple and context-driven.[232] Physically, a facial expression is a change in the face due to movements of several muscles demonstrating an emotional state. An emotional state is an individual's transient reaction to specific encounters with the environment, one that occurs and disappears depending on particular conditions. For example, someone is feeling or reacting with anger at a particular time and place.[233] A facial expression is communicated by a transient flexing of facial

[225] Rosalind W Picard, 'Affective Computing' (1995) MIT Media Laboratory Perceptual Computing Section Technical Report No 321 at 1 <https://hd.media.mit.edu/tech-reports/TR-321.pdf> accessed 8 February 2024.
[226] Rafael Calvo et al, 'Introduction to Affective Computing' in Rafael Calvo et al (eds), *The Oxford Handbook of Affective Computing* (OUP 2015) 2.
[227] Steffen Steinert, Orsolya Friedrich, 'Wired Emotions: Ethical Issues of Affective Brain–Computer Interfaces' (2020) Vol 26 Science and Engineering Ethics 351, 352.
[228] Rosalind W Picard, *Affective Computing* (MIT Press 1997) 47.
[229] Ibid preface x.
[230] Rafael Calvo et al, 'Introduction to Affective Computing' in Rafael Calvo et al (eds), *The Oxford Handbook of Affective Computing* (OUP 2015) 4.
[231] Avinash Awasthi, Manas K. Mandal, 'Facial Expressions of Emotions: Research Perspectives' in Manas K. Mandal, Avinash Awasthi (eds) *Understanding Facial Expressions in Communication* (Springer 2015) 3.
[232] Catherine Marechal et al, 'Survey on AI-Based Multimodal Methods for Emotion Detection' in Joanna Kołodziej, Horacio González-Vélez (eds) *High-Performance Modelling and Simulation for Big Data Applications* (Springer 2019) 314, 315.
[233] Richard S Lazarus, *Emotion and Adaption* (OUP 1991) 46, 47.

futures such as mouth, eyes and eyebrows due to the contraction of the muscles that make up the face.[234] These muscle contractions are controlled by two different areas of the brain, one controlling voluntary movements and the other involuntary reactions.[235] Facial expressions can easily be used for emotion detection because it only requires a simple video camera to register facial information automatically and in real-time.[236] Two approaches to measuring facial expressions will be discussed here: message-based and sign-based approaches.

Based on the assumption that the face provides a direct 'readout' of emotion, the message-based approach makes inferences about emotion or the affective state by assigning facial expression and movements to 'basic emotions' according to Ekman.[237] Facial movements and the 'basic emotions' hypothesised are illustrated in Figure 1.6.

Figure 1.6 Facial movements and hypothesised 'basic' emotion categories illustrated by Barret et al.[238] Used with permission.

It should be noted that this approach is problematic since the meaning of an expression depends on the context. For example, smiles accompanied by cheek raising express enjoyment, the same smile combined with head lowering and turning to the side convey embarrassment. Additionally, facial expressions can be posed or faked.[239]

The sign-based approach measures anatomic facial signs and then uses experimental or observational methods to discover the relation between these signs and emotion.[240] In 1978, the psychologists Ekman and Friesen proposed a model for measuring facial muscle contractions involved in facial expression called 'Facial Action Coding System' (FACS).[241] FACS is now a common standard used to

[234] Alice Caplier, 'Visual Emotion Recognition: Status and Key Issues' in Catherine Pelachaud (ed) Emotion-oriented Systems (Wiley-ISTE 2012) 107, 109.
[235] Hyisung C. Hwand, David Matsumoto, 'Emotional Expression' in Catharine Abell, Joel Smith (eds) *The Expression of Emotion* (CUP 2016) 139, 140.
[236] Catherine Marechal et al, 'Survey on AI-Based Multimodal Methods for Emotion Detection' in Joanna Kolodziej, Horacio Gonzalez-Vélez (eds) *High-Performance Modelling and Simulation for Big Data Applications* (Springer 2019) 314.
[237] Jeffrey F. Cohn, Fernando De La Torre, 'Automated Face Analysis for Affective Computing' in Rafael Calvo et al (eds), *The Oxford Handbook of Affective Computing* (OUP 2015) 132, 133.
[238] Lisa Feldman Barrett et al. 'Emotional Expressions Reconsidered' (2019) Vol 20 (1) Psychological Science in the Public Interest 1, 19.
[239] Jeffrey F. Cohn, Fernando De La Torre, 'Automated Face Analysis for Affective Computing' in Rafael Calvo et al (eds), *The Oxford Handbook of Affective Computing* (OUP 2015) 132.
[240] Ibid 133.
[241] Paul Ekman, Wallace V. Friesen, 'Facial Action Coding System: A Technique for the Measurement of Facial Movements' (1978) Consulting Psychologists Press.

systematically describe and quantify visible human facial movement[242] and describes facial activity in terms of anatomically-based action units (AU).[243] FACS defines 46 AUs to describe each independent movement of the face, including head and eye movements. FACS is used to verify the physiological presence of emotion. Due to its comprehensiveness, it also allows the discovery of new patterns related to emotional states.[244] FACS-coded facial events (AUs) such as 'Inner Brow Raiser', 'Chin Raiser', 'Lip Corner Puller' are classified into emotion categories by matching facial events with emotional events coded from previous empirical studies.[245] Figure 1.7 provides some examples of AUs.

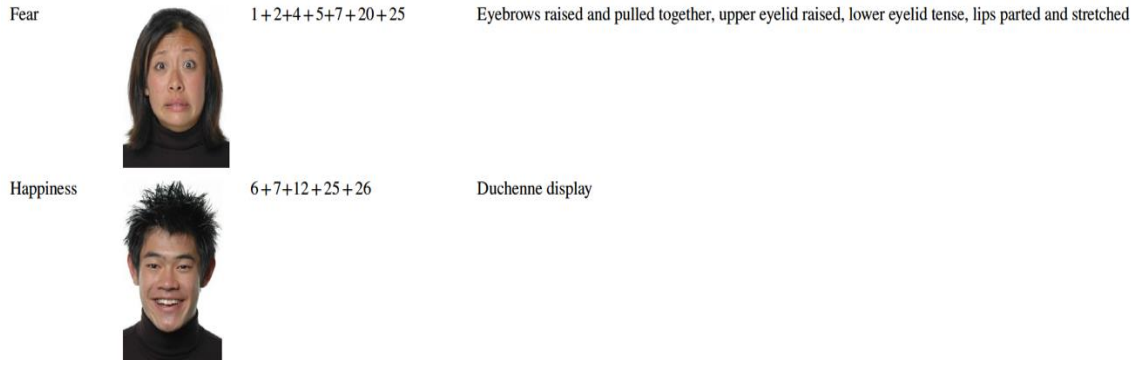

Figure 1.7 Facial expression examples for basic emotions 'fear' and 'happiness', the corresponding FACS action units and physical descriptions for each expression, illustrated by Keltner et al.[246] Used with permission.

Manual application of the FACS to videotaped behaviour is very time consuming. It takes approximately 100 hours to train a person to make judgements reliably and typically takes more than two hours to complete a one-minute video. [247]

Unsurprisingly, computer scientists started to use computer vision and graphics to automatically analyse and synthesise facial expression in automated face analysis (AFA) systems. Recently developed AFA systems claim to detect pain, frustration, emotion intensity, depression and psychological distress.[248] For example, a study aimed to predict depression, anxiety and stress levels from videos using the FACS approach built on ANN-based architecture.[249] Automated face analysis (AFA) systems seek to detect emotions using message-based and sign-based approaches. Such systems typically follow

[242] Lisa Feldman Barrett et al. 'Emotional Expressions Reconsidered' (2019) Vol 20 (1) Psychological Science in the Public Interest 1, 52.
[243] Jeffrey F. Cohn, Fernando De La Torre, 'Automated Face Analysis for Affective Computing' in Rafael Calvo et al (eds), *The Oxford Handbook of Affective Computing* (OUP 2015) 132.
[244] Marian Stewart Bartlett et al, 'Toward Automatic Recognition of Spontaneous Facial Actions' in Paul Ekman, Erika L. Rosenberg, *What the Face Reveals* (2nd edn OUP 2005) 393, 394.
[245] Erika L. Rosenberg, 'Introduction: The Study of Spontaneous Facial Expressions in Psychology' in Paul Ekman, Erika L. Rosenberg, *What the Face Reveals* (2nd edn OUP 2005) 14, 16.
[246] Dacher Keltner et al. 'Emotional Expression: Advances in Basic Emotion Theory' (2019) Vol 43 Iss 2 Journal of Nonverbal Behaviour 133, 142.
[247] Marian Stewart Bartlett et al, 'Toward Automatic Recognition of Spontaneous Facial Actions' in Paul Ekman, Erika L. Rosenberg, *What the Face Reveals* (2nd edn OUP 2005) 394.
[248] Jeffrey F. Cohn, Fernando De La Torre, 'Automated Face Analysis for Affective Computing' in Rafael Calvo et al (eds), *The Oxford Handbook of Affective Computing* (OUP 2015) 132, 133.
[249] Mihai Gavrilescu, Nicolae Vizireanu (2019) 'Predicting Depression, Anxiety, and Stress Levels from Videos Using the Facial Action Coding System' (2019) Vol 19 No 17, 1.

four steps: face detection, face registration, feature extraction and classification. In the first step, the face will be recognised using approaches from object detection in computer vision. During the face registration step, the face is rotated to an upright and frontal facing position to remove geometric differences.[250] In the feature extraction step, the algorithm extracts the main features of the face (e.g. mouth, eyebrows) and analyses movement, shape and texture composition of these regions to identify AUs.[251] After feature extraction, a machine learning (ML) component has the task to learn the relationship between the feature representation and the target facial expressions. Most of the current approaches use supervised learning[252], with a tendency to also make use of deep learning and ANN methods.[253] Fully automatic FACS coding systems use state-of-the-art ML techniques that can recognise any facial action.[254]

2.2.4.2 Speech in affective computing

Emotions of a person may be measured and quantified by observing speech signals from this person. This is exactly what speech-based emotion recognition systems aim at. Such systems are based on insight gained from research that investigates the mechanisms of emotional speech production.[255] Research in emotion recognition has shown that emotions in speech are related to prosody features such as pitch and energy.[256] Prosody refers to the study of the intonational and rhythmic aspects of language. Research has demonstrated specific associations between emotions such as fear, anger, sadness, joy and measures of pitch, voice level and speech rate.[257] Pitch is a perceptual property of a signal. The pitch of a sound is the mental sensation of fundamental frequency. In case a sound has a higher frequency, it is generally perceived as having a higher pitch.[258] The pitch of speech associated with emotions such as anger or happiness is higher than the pitch of speech associated with emotions such as sadness or disappointment.[259] In terms of speech rate, it has been shown that if the person who speaks is in an emotional state of anger or fear, the speech is usually faster. In case the person is bored or sad, then the speech is typically slower. Hence, effects of emotion tend to be present in features

[250] Michael Valstar, 'Automatic Facial Expression Analysis' in Manas K. Mandal, Avinash Awasthi (eds) *Understanding Facial Expressions in Communication* (Springer 2015) 144-150.
[251] Catherine Marechal et al, 'Survey on AI-Based Multimodal Methods for Emotion Detection' in Joanna Kołodziej, Horacio González-Vélez (eds) *High-Performance Modelling and Simulation for Big Data Applications* (Springer 2019) 315.
[252] Jeffrey F. Cohn, Fernando De La Torre, 'Automated Face Analysis for Affective Computing' in Rafael Calvo et al (eds), *The Oxford Handbook of Affective Computing* (OUP 2015) 137.
[253] Panagiotis Tzirakis et al, 'End-to-End Multimodal Emotion Recognition using Deep Neural Networks' (2015) Vol. 14 No. 8 Journal of Latex Class Files, 1.
[254] Marian Stewart Bartlett et al, 'Toward Automatic Recognition of Spontaneous Facial Actions' in Paul Ekman, Erika L. Rosenberg, *What the Face Reveals* (2nd edn OUP 2005) 395.
[255] Chi-Chun Lee et al, 'Speech in Affective Computing' in Rafael Calvo et al (eds), *The Oxford Handbook of Affective Computing* (OUP 2015) 171.
[256] Ricardo A. Calix, Leili Javadpour, Gerald M. Knapp, 'Detection of Affective States From Text and Speech For Real-Time Human-Computer Interaction' (2012) Vol 54 No 4 Human Factors and Ergonomics Society 530, 531.
[257] Christina Sobn and Murray Alpert, 'Emotion in Speech: The Acoustic Attributes of Fear, Anger, Sandess, and Joy' (1999) Vol 28 No 4 Journal of Psycholinguistic Research, 347.
[258] Daniel Jurafsky, James H Martin, *Speech and Language Processing* (2 edn, Pearson Education Limited 2014) 238.
[259] Ze-Jing Chuang, Chung-Hsien Wu, 'Multi-Modal Emotion Recognition from Speech and Text' (2004) Vol. 9 No. 2 Computional Linguistics and Chinese Language Processing, 45-62.

such as average pitch, pitch range and pitch changes, speech rate, voice quality and articulation.[260] Approaches in affective computing extract these acoustic signal features that characterise emotional speech. Machine learning algorithms map the automatically derived acoustic features described before to the desired emotion representations.[261] Research in the field aims to extract features from the voice to detect depressive people[262] or candidate stress levels during human resources interviews using ML and ANN.[263] Real-world applications of AC that aim to derive emotional states from speech are Amazon's 'Halo' wearable, which analyses voice tones to detect user emotions,[264] or Spotify's patented voice assistant,[265] which, based on commands or other utterances (e.g., 'ugh'), recognises when a user sounds sad and then offers encouragement by 'cheering' the user.[266] Methods applied to speech emotion recognition increasingly involve deep learning approaches.[267]

2.2.4.3 Multimodal approaches

Methods used in AC may be combined in multimodal approaches. For example, research in psychology aims to develop multimodal frameworks comprising audio-video fusion (facial expressions and emotions in speech) for the diagnosis, of depression to distinguish between people who suffer from depression and people who do not.[268]

Multimodal approaches may also include the detection from physiological factors such as cardiac activity (heart rate and heart rate variability). Research has shown that the variability of heart rate provides a novel marker to recognise emotions in humans.[269] Both heart rate and heart rate variability have been reported as indicators of fear, panic, anger and appreciation and are therefore used for affective computing.[270] Methods in AC can be integrated in commercial applications in order to track

[260] Rosalind W Picard, *Affective Computing* (MIT Press 1997) 179, 180.
[261] Chi-Chun Lee et al, 'Speech in Affective Computing' in Rafael Calvo et al (eds), *The Oxford Handbook of Affective Computing* (OUP 2015) 173, 177.
[262] Marius Dan Zbancioc, Silvia Monica Feraru, 'A study about the automatic recognition of the anxiety emotional state using Emo-DB' (E-Health and Bioengineering Conference, Iasi, 2015) 1.
[263] Kevin Tomba et al, 'Stress Detection Through Speech Analysis' (2018) Vol 1 ICETE 2018, 560.
[264] Alex Hern, 'Amazon's Halo wristband: the fitness tracker that listens to your mood' *The Guardian* (London, 28 August 2020) <https://www.theguardian.com/technology/2020/aug/28/amazons-halo-wristband-the-fitness-tracker-that-listens-to-your-mood> accessed 8 February 2024; Austin Carr, 'Amazon's New Wearable Will Know If I'm Angry. Is That Weird?' *Bloomberg* (New York, 31 August 2020) <https://www.bloomberg.com/news/newsletters/2020-08-31/amazon-s-halo-wearable-can-read-emotions-is-that-too-weird > accessed 8 February 2024.
[265] Daniel Bromand et al, 'Systems and Methods for Enhancing Responsiveness to Utterances Having Detectable Emotion' US Patent Number US 10566010 B2 (Assignee: Spotify AB) February 2020 11 < https://patentimages.storage.googleapis.com/2a/9d/2d/926b58a2bd956f/US10566010.pdf >, accessed 8 February 2024.
[266] Josh Mandell, 'Spotify Patents A Voice Assistant That Can Read Your Emotions' *Forbes* (New York, 12 March 2020) <https://www.forbes.com/sites/joshmandell/2020/03/12/spotify-patents-a-voice-assistant--that-can-read-your-emotions/> accessed 8 February 2024.
[267] Haytham M Fayek, Margaret Lech, Lawrence Cavedon, 'Evaluating deep learning architectures for Speech Emotion Recognition' (2017) Vol 92 Neural Networks 60.
[268] Jyoti Joshi et al, 'Multimodal assistive technologies for depression diagnosis and monitoring' (2013) Vol 7 Journal on Multimodal User Interfaces, 217.
[269] Quintana Daniel et al. 'Heart rate variability is associated with emotion recognition: Direct evidence for a relationship between the automatic nervous system and social cognition' (2012) Vol 86 No 2 International Journal of Psychophysiology 168.
[270] Jennifer Healey, 'Physiological Sensing of Emotion' in Rafael Calvo et al (eds), *The Oxford Handbook of Affective Computing* (OUP 2015) 206.

and analyse customer behaviour in retail stores, so-called behaviour inference systems. Behavioural inference systems apply deep learning (DL) and affective computing in order to monitor and analyse the shopper's behaviour based on extracted physiological factors (heart rate) and facial expressions.[271] An example for such a system comprises six modules: a speech recognition module, a biofeedback model, a facial expression and emotion recognition module, a gaze detection module, an age and gender recognition module and an identification module.[272]

2.2.5 Automated reasoning (AR)

Automated reasoning (AR) aims to develop computers that can use stored information to answer questions and draw new conclusions.[273] It may be described as the science of developing methods that intend to replace human reasoning with procedures that perform individual reasoning automatically.[274] Automated reasoning is devoted to answering questions from diverse data without human intervention and includes decision-making. As a form of reasoning, decision-making focusses on an autonomous agent trying to perform a task for a human.[275] Reasoning problems are of practical significance, they arise naturally in many applications that interact with the world, for example, reasoning about knowledge in the sciences or natural language processing. Furthermore, reasoning algorithms form the foundation for theoretical investigations into general AI (human-level AI).[276] Reasoning is the process of obtaining new knowledge from a given knowledge, where certain transformation rules are applied that depend only on knowledge and can be done exclusively in the brain without involving senses.[277] Research in automated reasoning focusses on logical reasoning, probabilistic reasoning and common sense reasoning.[278] Logical reasoning attempts to avoid any unjustified assumptions and confines itself to inferences that are infallible and beyond reasonable dispute.[279] Probabilistic reasoning deals with uncertainty about knowledge and belief. Uncertainty may be approached by applying tools from probability theory and statistics. Research in probabilistic reasoning focusses on the representation of different types of uncertainty and uncertain knowledge, reasoning with these types of knowledge, and learning them. It facilitates the development of applied systems of practical importance, such as machine vision, medical diagnosis and natural language processing. Probabilistic reasoning models are close to ML and serve as a medium between ML and AR.[280]

[271] Andrea Generosi, Silvia Ceccacci, Maura Mengoni, 'A deep learning-based system to track and analyse customer behaviour in retail store' (IEEE 8th International Conference on Consumer Electronics, Berlin 2018) 36.
[272] Ibid 37.
[273] Stuart Russel, Peter Norvig, *Artificial Intelligence, A Modern Approach* (3rd edn, Pearson Education 2016) 2.
[274] Tudor Jebelean et al, 'Automated Reasoning' in Buchberger Bruno et al (eds) *Hagenberg Research* (Springer 2009) 63.
[275] Amir Eyal, 'Reasoning and decision making' in Frankish Keith and Ramsey William M. (eds) *The Cambridge Handbook of Artificial Intelligence* (2014) 191.
[276] Ibid 191.
[277] Tudor Jebelean et al, 'Automated Reasoning' in Buchberger Bruno et al (eds) *Hagenberg Research* (Springer 2009) 63.
[278] Amir Eyal, 'Reasoning and decision making' in Frankish Keith and Ramsey William M. (eds) *The Cambridge Handbook of Artificial Intelligence* (2014) 193.
[279] John Harrison, *Handbook of Practical Logic and Automated Reasoning* (Cambridge University Press 2009) 1.
[280] Amir Eyal, 'Reasoning and decision making' in Frankish Keith and Ramsey William M. (eds) *The Cambridge Handbook of Artificial Intelligence* (2014) 201.

For a very long time, scientists and philosophers have tried to understand and formalise how humans reason and whether reasoning methods may be automatised.[281] Achieving common sense reasoning capabilities in computational systems has been one of the goals of AI since its beginning in the 1960s.[282] Common sense reasoning constitutes a central part of human behaviour and is a precondition for human intelligence. Unsurprisingly, the creation of systems that exhibit common sense reasoning is a central goal towards achieving general AI. History in AI has proven that it is more difficult to develop systems with common sense reasoning capabilities compared to systems that solve explicit reasoning problems, such as chess-playing programs or expert systems that assist in clinical diagnosis. Part of this difficulty is due to the all-encompassing aspect of common sense reasoning: It requires many different kinds of knowledge. Furthermore, most common sense knowledge is implicit and therefore difficult to explain and compute, unlike expert-knowledge which is usually explicit. Therefore, implicit common sense knowledge must be made explicit in order to develop common sense reasoning systems.[283]

Other problems that impede the development of automated common sense reasoning are the lack of a precise meaning of 'common sense reasoning', how to take into account of polysemy, ambiguity and vagueness of natural language and the difficulty in modelling the role of various forms of implicit knowledge such as context, background knowledge and tacit knowledge.[284] Therefore, common sense reasoning capabilities are still a challenge in AI applications.[285] According to Oren Etzioni, who oversees the Allen Institute for Artificial Intelligence, AI 'is devoid of common sense'.[286] Hence, to acquire common sense from massive amounts of data and implementing it in intelligent systems appears to be the next frontier in AI.[287] The lack of progress in providing general automated common sense reasoning capabilities underscores that this is a very difficult problem in the field of AI.[288] Common sense reasoning is not just the hardest problem for AI, it is also considered to be the most important problem.[289]

[281] Marco Gavanelli, Toni Mancini, 'Automated Reasoning' (2013) Vol. 7 No. 2 Intelligenza Artificiale 113.
[282] Brandon Bennet, Anthony G Cohn, 'Automated Common-sense Spatial Reasoning: Still a Hughe Challenge' in Stephen Muggleton, Nicholas Chater (eds) Human-Like Machine Intelligence (Oxford University Press 2021) 405.
[283] Ernest Davis, Leora Morgenstern, 'Introduction: Progress in formal common sense reasoning' (2004) Vol 153 Artificial Intelligence 1.
[284] Brandon Bennet, Anthony G Cohn, 'Automated Common-sense Spatial Reasoning: Still a Huge Challenge' in Stephen Muggleton, Nicholas Chater (eds) Human-Like Machine Intelligence (Oxford University Press 2021) 406.
[285] Shoham Yoav et al, 'The AI Index 2018 Annual Report' (AI Index Steering Committee Stanford University 2018) 64 <https://hai.stanford.edu/sites/default/files/2020-10/AI_Index_2018_Annual_Report.pdf > accessed 8 February 2024.
[286] Cade Metz, 'Paul Allen Wants to Teach Machines Common Sense' *The New York Times* (New York, 28 February 2018) <https://www.nytimes.com/2018/02/28/technology/paul-allen-ai-common-sense.html> accessed 8 February 2024.
[287] Niket Tandon, Aparna S. Varde, Gerard de Melo, 'Commonsense Knowledge in Machine Intelligence' (2017) Vol 46 No 4 SIGMOD Record 49.
[288] Brandon Bennet, Anthony G Cohn, 'Automated Common-sense Spatial Reasoning: Still a Hughe Challenge' in Stephen Muggleton, Nicholas Chater (eds) Human-Like Machine Intelligence (Oxford University Press 2021) 405.
[289] Gary Marcus, Ernest Davis, *Rebooting AI: Building Artificial Intelligence we can trust* (Pantheon Books 2019).

2.3 Conclusions

This chapter answered Subquestion 1, namely, what AI is and what disciplines exist therein. AI is an exciting, challenging and complex technology which accelerates at a tremendous pace. AI covers a broad range of approaches and techniques and at least five disciplines. These five disciplines are machine learning, natural language processing, computer vision, affective computing and automated reasoning.

As a major discipline of AI, *machine learning* (ML) is focussed on computers that program themselves based on experience. ML can be applied by means of several methods, ranging from supervised to unsupervised to reinforcement learning. *Deep learning* (DL) is a very powerful kind of machine learning considering that the achievements in the field have been reached with *artificial neural networks* (ANNs) comprising an astonishingly small number of neurons when compared with neural networks of the human brain. By means of *natural language processing* (NLP), machines can process human language. It includes both the generation and understanding of natural language. NLP significantly contributes to improved interactions between machines and humans. *Computer vision* (CV) facilitates the automated processing of visual images and thus enables machines to see. Face recognition, which is one of the applications of computer vision, empowers machines to identify or verify the identity of humans in images or videos based on biometric data. Because emotions form an important factor of human intelligence and daily life, *affective computing* (AC) aims to equip machines with emotional capabilities. Approaches in AC which derive emotions from facial expressions and speech may be easily deployed and widely used. Efforts in the discipline of *automated reasoning* (AR) seek to perform individual reasoning automatically.