Chapter 11
Artificial Intelligence in Human-Robot Interaction

Edirlei Soares de Lima and Bruno Feijó

Abstract Human-Robot Interaction challenges the field of research on Artificial Intelligence in many ways, especially regarding the complexity of the physical world. While physical interactions require Artificial Intelligence techniques to handle dynamic, nondeterministic, and partially unknown environments, the communication with humans requires socially acceptable responses and common-sense knowledge to handle a broad variety of situations with complex semantics to interpret and understand. In the context of emotional design, different Artificial Intelligence techniques are necessary to allow robots to express, understand, and induce emotions as part of the interaction process. This chapter explores Human-Robot Interaction from the Artificial Intelligence point of view, presenting the main challenges, techniques, and our particular vision for future developments in this research area.

Keywords Artificial intelligence · Robotics · Human-robot interaction · Machine learning

11.1 Introduction

In the emotional design of Human-Robot Interaction, we should consider the connections that can form between humans and robots, and the emotions that can arise from them. In this context, the most central technological question is the intelligence of the machines. What is Artificial Intelligence and what are its limitations?

Over the last decades, Artificial Intelligence (AI) has emerged into the public view as an important frontier of technological innovation with potential influences in many areas. The first use of the term Artificial Intelligence is attributed to John McCarthy, who created the term in his 1955 proposal for the 1956 Dartmouth Conference (Russell and Norvig 2009), which is considered the seminal event for Artificial Intelligence as a field. Today, applications of Artificial Intelligence are all around us

E. S. de Lima (✉)
School of Technology, Arts and Communication, Universidade Europeia, Lisbon, Portugal
e-mail: edirlei.lima@universidadeeuropeia.pt

B. Feijó
Department of Informatics, Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro, Brazil

H. Ayanoğlu and E. Duarte (eds.), *Emotional Design in Human-Robot Interaction*,
Human–Computer Interaction Series, https://doi.org/10.1007/978-3-319-96722-6_11

(virtual assistants, recommendation systems, robotics arms in assembly lines) and there are more to come in the near future (autonomous vehicles, autonomous drone delivery services, robot assistants, etc.).

The term Artificial Intelligence can have different definitions depending on the context and the intended application. The English Oxford Dictionary defines it as "the theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages." (Stevenson 2010). A more general definition is given by the Merriam-Webster dictionary: "a branch of computer science dealing with the simulation of intelligent behavior in computers." (Merriam-Webster 2016). Similarly, The Encyclopedia Britannica states that AI is "the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings." (Encyclopedia Britannica 2003). However, all these definitions do not consider the fact that intelligence itself is not very well defined or understood.

In general, Artificial Intelligence can still be considered a young discipline, and its structure, concerns, and methods are less clearly defined than those of more mature sciences (such as physics).

Many different disciplines contributed to the development and establishment of the field of research on Artificial Intelligence, including Philosophy, Mathematics, Psychology, Neuroscience, Linguistics, and Computer Engineering. The general goal of the research on Artificial Intelligence is to create the technology necessary for computers to work in an intelligent manner. Over the past years, several techniques have been proposed and successfully applied for a variety of different tasks, such as market analysis, medical diagnosis, speech recognition, simulation, training, weather forecasting, emotion analysis, facial recognition, and robotics.

There are many approaches and methods to creating intelligent systems, including search and optimization, logic and planning, probabilistic reasoning, and machine learning. Some approaches are becoming synonyms of AI, such as machine learning, which gives computers the ability to learn without being explicitly programmed to solve a specific problem (Russell and Norvig 2009). Machine learning techniques are employed in a vast range of computing tasks, where designing and programming explicit algorithms with good performance is difficult or infeasible (Mitchell 1997).

Human-Robot Interaction represents a challenge for the field of research on AI (Lemaignan et al. 2017). Most classical AI techniques were not designed to handle the dynamic, nondeterministic, and partially unknown environments of the physical world. Over the last decades, the most successful applications of robots were limited to simple tasks that involve predictable situations (e.g., packaging, welding, and spray painting). Robot automation obtained a huge commercial success because it is usually applied to highly repetitive processes that hardly vary and require little dexterity, such as those done in industrial plants. However, physical interactions and communication with humans require socially acceptable responses and common-sense knowledge to handle a broad variety of situations with complex semantics to interpret and understand.

Robots that interact with humans are very different than those used in assembly lines: they require more intelligent behavior than simply following a set of instruc-

tions to complete repetitive tasks. As a result, robotics is moving into areas where sensor input becomes increasingly important and the AI must be robust enough to anticipate and handle a range of different situations (Thrun et al. 2005). Robotics, thus, is increasingly becoming a software science, where the goal is to develop robust software that enables robots to handle the challenges that arise when dealing with complex and dynamic environments.

Human-Robot Interaction requires intelligent robots able to recognize, understand, and participate in communication situations, both explicit (e.g., the human addresses verbally the robot) and implicit (e.g., the human points to an object). In addition, an intelligent robot must be able to take part in joint actions, both proactively (by planning and proposing resulting plans to the human) and reactively (by following the human instructions) (Lemaignan et al. 2017). All these actions must be complemented with the robot's ability to move and act in a safe, efficient, and legible way, considering all social rules relevant to the situation. This kind of behavior requires more than just AI algorithms; it requires support from Psychology, Philosophy, Interaction studies, and General computer engineering.

11.2 Artificial Intelligence in Robotics

From the Artificial Intelligence point of view, robots are physical agents that perform tasks by manipulating the physical world (Russell and Norvig 2009). They are equipped with effectors (e.g., legs, wheels, joints, and grippers) and sensors (e.g., cameras, lasers gyroscopes, and accelerometers). While the sensors allow the robot to perceive the environment, effectors are used to asserting physical forces on the environment.

In the physical world, robots must interact with environments that are partially observable, nondeterministic, dynamic, continuous, and multiagent. Partial observability, continuously, and non-determinism are the result of dealing with a large and complex world. Most robot sensors cannot see around corners, and motion commands are subject to uncertainty due to gear slipping and friction (i.e., it is not possible to guarantee that all planned actions will have the desired results). The physical world is also dynamic, so it can change while the robot is planning or performing an action, which requires real-time responses from the robot. In addition, some robots can interact with other robots or with humans, which adds the multiagent characteristic to the environment.

Figure 11.1 illustrates a general robot interaction process. First, the robot sensors capture raw signals from the environment (e.g., visual signals, audio signals, and tactile signals). Then, feature extraction methods are used to obtain meaningful data from the raw signals. Based on the extracted features, Artificial Intelligence and Computer Vision techniques are performed to procedure a semantic understanding of the current situation (e.g., object recognition, object tracking, and emotion recognition). With this information, the robot can plan and perform the most appropriated actions, such as movements, gestures, and speech.

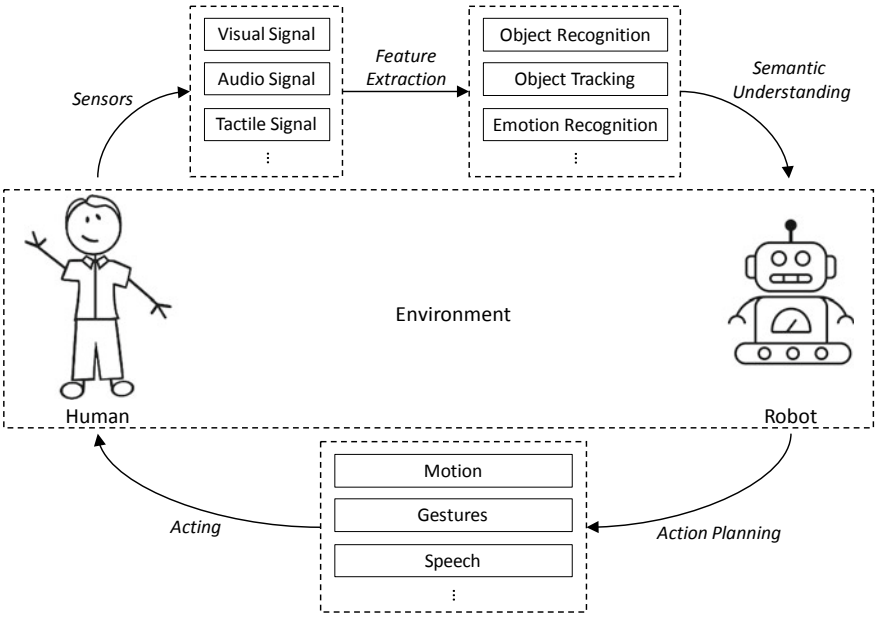

Fig. 11.1 General robot interaction process

Perception is one of the key elements when designing robots that interact with complex environments (Russell and Norvig 2009). It consists of the process of mapping sensor measurements into internal representations of the environment. The perception process can be divided into two steps: (1) feature extraction, which has the objective of converting the raw signals from sensors to feature descriptors for subsequent understanding tasks; and (2) semantic understanding, which aims at inferring the objects or human behaviors from the extracted features. Typical semantic understanding tasks include object detection and recognition, human tracking and identification, speech recognition, emotion recognition, and touching detection and recognition.

One of the most basic perceptions a robot requires is localization, which is used to determine where things are in the environment (including the robot itself). This kind of knowledge is the key element of any successful physical interaction with the environment (Thrun et al. 2005). For example, robot manipulators must know the location of objects they seek to manipulate, navigating robots must know where they are to find their way around, and assistant robots must know where their human subjects are.

Localization has received a lot of research attention in the past decades and, as a result, significant advances have been made on this front. The most common method used to determine the position of the robot in the environment is called Monte Carlo localization (MCL) (Dellaert et al. 1999). The MCL algorithm estimates the position and orientation of a robot as it moves and senses the environment (Thrun

et al. 2005). The algorithm uses a particle filter to represent the distribution of likely states (each particle represents a possible state, that is a hypothesis of where the robot is). Initially, the particles are uniformly distributed based on prior knowledge. Whenever the robot moves and senses something new, the particles are resampled using a recursive Bayesian estimation (Berger 1985). This process repeats until all the particles converge toward the actual position of the robot. In some situations, no map of the environment is available. In these cases, the robot must acquire a map while navigating through the environment. This problem is known as simultaneous localization and mapping (SLAM) (Thrun and Leonard 2008). Usually, this problem is solved using probabilistic techniques, including the extended Kalman filter (Jetto et al. 1999).

Not all robot perceptions are about localization. Social robots also need to recognize objects, identify humans, recognize gestures, track subjects, recognize emotions, and so on. Motivated by the fact that most information received by human beings are visual signals (Castleman 1996), most robot systems use visual signals to simulate human-like perceptions (Yan et al. 2014). Most of these visual signals are usually obtained using traditional or stereographic camera sensors. Then, Computer Vision techniques are used to extract meaningful information from the captured images.

Different tasks usually require different features and specific techniques to extract them. The field of research on Computer Vision provides a vast repertory of techniques for feature extraction, including color, texture, shape, and motion. Color can be used to detect objects with distinct color components (Khan et al. 2012). It can also be used to efficiently detect human skin and identify the presence of human subjects (Darrell et al. 2000; Wang et al. 2008). However, color can be easily affected by illumination conditions, which requires special treatment. Visual texture is another important property for object and face detection. The Local Binary Pattern (LBP) (Wang and He 1990; Ojala et al. 1996) and the Scale Invariant Feature Transform (SIFT) (Lowe 1999) are both popular texture descriptors for feature representation that have been widely used in object recognition, robotic navigation, video tracking, and image matching. The shape is also a useful feature for visual signal representation, especially for facial image analysis and human detection. Popular shape descriptors include the snake model (Kass et al. 1988), which can capture features like lines and edges; and the Hu descriptors (Hu 1962), which are based on non-orthogonalized central moments that are invariant to image rotation, translation, and scale. Another important visual feature is motion, which is widely used for object tracking. Optical flow is a typical motion feature that represents the distribution of velocities of brightness patterns' movement in an image (Horn and Schunck 1981; Brox et al. 2004; Bab-Hadiashar and Suter 1998).

Motivated by the fact that speech is an important communication channel for human beings, audio signals are also an important source of information for robots that interact with human subjects. By analyzing the collected audio signals, robots can acquire more information related to their interaction subjects, such as their positions, commands, and emotional states (Yan et al. 2014). In addition, audio signals are essential to establish a communication channel between humans and robots through speech recognition and speech synthesis.

With all relevant features extracted from the sensors' signals, semantic understanding tasks must be performed in order to generate semantic knowledge to be used by the robot to plan future actions. For these tasks, Artificial Intelligence techniques—especially machine learning methods—are essential for general solutions. While some simple tasks can be solved only with Computer Vision techniques (such as human tracking), most of the semantic understanding tasks require machine learning algorithms (e.g., object recognition, emotion recognition, human identification).

11.3 Machine Learning

Machine learning tasks are typically classified into three categories, depending on the type of data available to the system: supervised learning, unsupervised learning, and reinforcement learning (Russell and Norvig 2009). In supervised learning tasks, the system learns a function that maps an input (features describing an instance of a problem) to an output (the correct answer for the instance of the problem) based on examples of input–output pairs. There are several algorithms for supervised machine learning, including artificial neural networks (Priddy and Keller 2005), decision trees (Rokach and Maimon 2014), support vector machines (Steinwart and Christmann 2008), k-nearest neighbors (Altman 2012), etc. In contrast, unsupervised learning tasks require from the system the ability to learn patterns in the input even though no explicit output is supplied. Algorithms for unsupervised learning include clustering methods (Aggarwal and Reddy 2013), artificial neural networks, and latent variable models (Loehlin 1998). In reinforcement learning tasks, the system learns from a series of reinforcements (rewards or punishments). The system does not know which actions to take, but instead, it must discover which actions yield the highest rewards by trying them. Algorithms for reinforcement learning include Q-learning (Watkins and Dayan 1992) and State-Action-Reward-State-Action (Szepesvári 2010).

Artificial Neural Network is a very popular machine learning algorithm used in robotics for a variety of tasks. Inspired by the biological neural networks that constitute biological brains (Russell and Norvig 2009), artificial neural networks comprise several artificial neurons interconnected with each other to form a network with input, hidden, and output layers (Fig. 11.2). Neural networks "learn" by example and can be trained to extract patterns and detect trends.[1]

Over the last years, several works explored the use of neural networks to solve robotic tasks. Seemann et al. (2004) presented a method for estimating a person's head pose using neural networks trained with grayscale and disparity images from a stereo camera. Ge et al. (2011) described a neural network to estimate human motion intention based on the desired trajectory in human limb model. Yin and Xie

[1] In a nutshell, given a set of training points (x_i, y_i), a system that learns by example tries to find a function f that maps a given x to its corresponding y (within a certain error tolerance). In a neural network, this function is represented by numerical weights associated with each node. During training, these numbers are continually adjusted until training data with the same labels consistently yielding similar outputs.

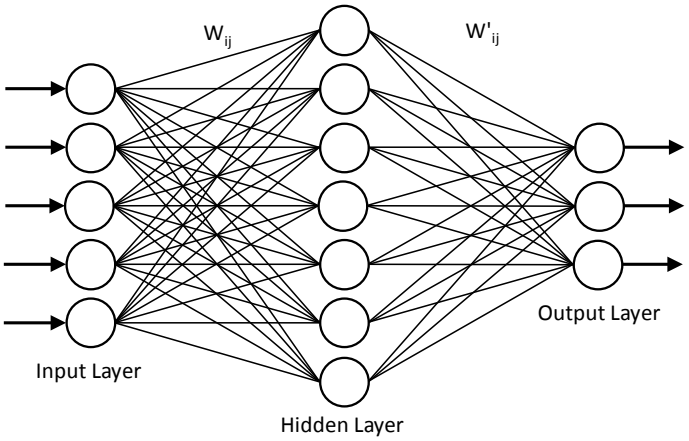

Fig. 11.2 Structure of an artificial neural network. Each neuron is connected to all neurons in the next layer and each connection has a numeric weight (W_{ij} and W'_{ij}) that determines the strength and sign of the connection

(2007) proposed a hand posture recognition system for humanoid robots that uses neural networks trained with topological features extracted from the silhouette of the segmented hand. Ito et al. (2006) used a dynamic neural network model to allow a small humanoid robot to learn object handling behaviors. Other authors proposed solutions for general problems that also exist in robotics using neural networks. Maturana and Scherer (2015) proposed a real-time object recognition method that uses a supervised 3D convolutional neural network capable of recognizing hundreds of objects per second. Bhatti et al. (2004) presented a language-independent emotion recognition system for the identification of human affective state in the speech signal using a neural network. Lawrence et al. (1997) presented a hybrid neural network for human face recognition that combines local image sampling, a self-organizing map neural network, and a convolutional neural network.

11.4 Future Developments

Future developments in Artificial Intelligence in HRI should be mainly driven by breakthroughs in deep learning. However, a realistic assessment and a solid understanding of these new possibilities require a review of the core of what Artificial Intelligence means and what are its most challenging problems.

If we take the claimed goal of AI seriously—i.e., the production of AI—then we shall specify the most important features of intelligence. There is no consensus in the AI community about the most adequate theory of intelligence, mainly because this depends on the context. For the purposes of the present writing, we can assume that intelligent behavior arises from the ability to learn, to adapt behavior to new and

challenging environments, and to be creative.[2] Alternatively, we can suppose that intelligent behavior arises from the balance of the following abilities: (i) the ability to evaluate, analyze, and compare information; (ii) the ability to generate invention and discovery; (iii) the ability to apply what have been learned in the appropriate situation. This is exactly the Triarchic Theory of Successful Intelligence proposed by Cianciolo and Sternberg (2008). However, no matter which intelligence theory we adopt, the challenge of producing Artificial Intelligence is enormous.

The story of AI consists of successes and failures, ups and downs, abundance and scarcity of investments.[3] From a theoretical viewpoint, this story is marked by the rivalry between two lines of thought: *logic-based AI* (also known as *logicism* or *symbolic AI*) and *machine learning* (mainly, artificial neural networks). In the first line, cognition involves operations on symbols. In contrast, neural networks exhibit intelligent behavior without processing on symbolic expressions.

Logicism was the predominant theory within the AI community until the mid-1980s, when neural networks, genetic algorithms, and other machine learning paradigms started producing impressive results. Currently, in the late 2010s, after almost 20 years of experiencing slow development, AI innovation has exploded and deep learning (mainly in the form of neural networks) has been dominating the AI scenario.

A deep learning algorithm (also known as hierarchical learning) attempts to learn in multiple levels, corresponding to different levels of abstraction. Deep learning requires extremely large datasets (called big data), which are complex sets to process, manage, and maintain. Furthermore, current deep learning models also require training datasets that are labeled (i.e., data that have been classified or categorized by humans).

Deep learning typically uses artificial neural networks leading to the so-called deep neural networks (DNNs). In a DNN, there are multiple layers to process features, and generally, each layer extracts some piece of information. In this architecture, higher level features are defined in terms of lower level ones. Deep neural networks can have hundreds of millions of parameters (LeCun et al. 2015), allowing them to model complex functions such as nonlinear dynamics. They form compact representations of states from raw, high-dimensional, multimodal sensor data commonly found in robotic systems.

The current burst of AI advances occurred due to the emergence of powerful GPUs (Graphics Processing Units) [4] being used for complex computations of deep learning models, and the availability of big data. Deep learning has achieved astonishing performance in many complex tasks like language translation (Wu et al. 2016),

[2] Apart from being "creative", this is totally aligned with early psychological theories, such as the one by Edward Thorndike in the very ending of the nineteenth century (Thorndike 1911), which are one of the first references on learning mentioned by researchers of artificial neural networks (also known as connectionists) (Knight 2017; Ertugrul and Tagluk 2017).

[3] A fun but complete and accurate history of AI until the early 1990s can be found in Crevier (1993).

[4] GPUs are designed for rendering graphics by having a large number of simple process units for massively parallel calculation. However, we can use GPUs to perform any sort of computation (e.g., deep learning computation). We can use multiple GPUs to increase processing power.

strategic games playing (Silver et al. 2016), and self-driving cars (Bojarski et al. 2016). Although deep learning has been successful in perception and classification problems, it is far from solving real reasoning problems. The next AI revolution is when *deep reasoning* becomes effective. Cognitive robots[5] rely not only on deep learning but also on deep reasoning. Deep reasoning is required for cognitive tasks, such as common sense, dealing with changing situations, planning, remembering, and making complex decisions. Robots and other Artificial Intelligent systems are still far from real deep reasoning. Yet there are even more complex issues that current technology cannot deal with, such as *consciousness* (especially the ability to obtain and process information about ourselves) and *ethics*. An interesting approach to machine consciousness can be found in Dehaene et al. (2017). The present authors believe that logicism and machine learning should cooperate with each other to deal with these challenging questions and the situation of AI safety in general. A deep discussion about AI safety is presented in Amodei et al. (2016).

While disruptive solutions for deep reasoning are yet to come, important improvements in deep learning can be pursued: (i) to train systems on less data ("small data"); (ii) to use unlabeled training datasets (unsupervised learning); and (iii) to open the black box of deep learning systems—the interpretability problem (Lipton 2017). These subjects are somewhat intertwined, and we may envisage a future system that can attain rapid learning from small unlabeled data and, in case of an accident, can track down the cause.

The use of small data is necessary not only because developing AI systems using big data is a costly and time-consuming task (or because extremely large sets of data are not available in many domains) but also because an AI system must quickly adapt to single unexpected observations. Many situations require rapid inference from small quantities of data. As pointed by some researchers of Google DeepMind (Santoro et al. 2016): "in the limit of 'one-shot learning', single observations should result in abrupt shifts in behavior." The use of statistical models (e.g., Gaussian process and Bayesian optimization) to deal with the problems of small data and interpretability have been reported by the media (Metz 2017). In a different approach to interpretability, a recent work by Google Brain explains how a deep neural network can make decisions by combining feature visualization (*what is a neuron looking for?*) with attribution (*how does it affect the output?*) (Olah et al. 2018).

The second research direction mentioned above efers to the use of raw, unlabeled data to train AI systems with little or no human intervention (known as unsupervised learning). The use of massive labeled dataset training presents many drawbacks: it is costly, consumes time, and introduces human bias into the systems (either unintentionally or caused by malicious attackers). One of the first experiments with large-scale unsupervised learning was presented by Google and Stanford University (Le et al. 2012). We can reduce the use of labeled data if we use *transfer learning*, a technique in which the first layers of a network are a copy of the first layers of another network (Yosinski et al. 2014).

[5]Cognitive robots, different from industrial robots, are robots that reason, remember, learn, anticipate, plan, and communicate with humans and with each other.

The next AI breakthrough can be driven by new hardware currently under development. The two most promising new hardware paradigms are *neuromorphic chips* and *quantum computing*, according to current media reports (Knight 2018).

A review of deep learning in robotics can be found in Pierson and Gashler (2017). These authors argue that large training data, long training times and unsupervised learning for critical robotic systems[6] are the main barriers to the adoption of deep learning in robotics. They also claim that a promising perspective is crowdsourcing training data via *cloud robotics* (Pratt 2015). However, as we have mentioned in the present section, advances in small data, unsupervised learning, and interpretability can also lower the barriers to adoption of deep learning in robotics.

As far as HRI is concerned, deep learning is of ultimate importance for communication and assistance. For example, deep neural networks can recognize spontaneous emotional expressions (Barros et al. 2015), which is essential for Human-Robot Interaction. Assistance has been considered a premium goal in AI systems, in the sense that AI should be used to augment human intelligence. In this case, we should create user interfaces that let us work with the representations inside machine learning models (Carter and Nielsen 2017). As a general prognosis, future developments in AI systems must maintain a strong adherence to the concept of creating an interactive and intelligent conversation between a human and a machine.

References

Aggarwal C, Reddy C (2013) Data clustering: algorithms and applications. Chapman and Hall/CRC, London

Altman NS (2012) An introduction to kernel and nearest-neighbor nonparametric regression. Am Stat 46(3):175–185. https://doi.org/10.2307/2685209

Amodei D, Olah C, Steinhardt J, Christiano P, Schulman J, Mané D (2016) Concrete problems in AI safety. arXiv preprint: 1606.06565

Bab-Hadiashar A, Suter D (1998) Robust optic flow computation. Int J Comput Vis 29(1):59–77. https://doi.org/10.1023/A:1008090730467

Barros P, Weber C, Wermter S (2015). Emotional expression recognition with a cross-channel convolutional neural network for human-robot interaction. In: Proceedings of the 2015 IEEE-RAS 15th international conference on humanoid robots (humanoids). https://doi.org/10.1109/HUMANOIDS.2015.7363421

Berger J (1985) Statistical decision theory and Bayesian analysis. Springer, New York

Bhatti M, Wang Y, Guan L (2004) A neural network approach for human emotion recognition in speech. In: 2004 IEEE international symposium on circuits and systems. https://doi.org/10.1109/ISCAS.2004.1329238

Bojarski M et al (2016) End to end learning for self-driving cars. arXiv preprint: 1604.07316v1

Britannica Encyclopedia (2003) Encyclopedia Britannica, 15th edn. Encyclopedia Britannica Inc., Chicago

Brox T, Bruhn A, Papenberg N, Weickert J (2004) High accuracy optical flow using a theory for warping. In: Pajdla T, Matas J (eds) Computer vision—ECCV 2004, vol 3024, pp 25–36. https://doi.org/10.1007/978-3-540-24673-2_3

[6]Such as aerial vehicles, where a single failure is catastrophic.

Carter S, Nielsen M (2017) Using artificial intelligence to augment human intelligence. Distill J. https://doi.org/10.23915/distill.00009

Castleman K (1996) Digital image processing. Prentice Hall, New York

Cianciolo AT, Sternberg RJ (2008) Intelligence: a brief history. Wiley-Blackwell Publishing, Malden

Crevier D (1993) AI: the tumultuous history of the search for artificial intelligence. Basic Books, New York

Darrell T, Gordon G, Harville M, Woodfill J (2000) Integrated person tracking using stereo, color, and pattern detection. Int J Comput Vis 37(2):175–185. https://doi.org/10.1023/A:1008103604354

Dehaene S, Lau H, Kouider S (2017) What is consciousness, and could machines have it? Science 358:486–492. https://doi.org/10.1126/science.aan8871

Dellaert F, Fox D, Burgard W, Thrun S (1999) Monte Carlo localization for mobile robots. In: Proceedings of the 1999 IEEE international conference on robotics and automation. https://doi.org/10.1109/ROBOT.1999.772544

Ertugrul O, Tagluk ME (2017) A novel machine learning method based on generalized behavioral learning theory. Neural Comput Appl 28(12):3921–3939. https://doi.org/10.1007/s00521-016-2314-8

Ge S, Li Y, He H (2011) Neural-network-based human intention estimation for physical human-robot interaction. In: Proceedings of the 8th international conference on ubiquitous robots and ambient intelligence (URAI). https://doi.org/10.1109/URAI.2011.6145849

Horn B, Schunck B (1981) Determining optical flow. Artif Intell 17(1–3):185–203. https://doi.org/10.1016/0004-3702(81)90024-2

Hu M (1962) Visual problem recognition by moment invariants. IRE Trans Inf Theory 8:179–187

Ito M, Noda K, Hoshino Y, Tani J (2006) Dynamic and interactive generation of object handling behaviors by a small humanoid robot using a dynamic neural network model. Neural Netw 19(3):323–337. https://doi.org/10.1016/j.neunet.2006.02.007

Jetto L, Longhi S, Venturini G (1999) Development and experimental validation of an adaptive extended Kalman filter for the localization of mobile robots. IEEE Trans Robot Autom 15(2):219–229. https://doi.org/10.1109/70.760343

Kass M, Witkin A, Terzopoulos D (1988) Snakes: active contour models. Int J Comput Vis 1(4):321–331. https://doi.org/10.1007/BF00133570

Khan F, Anwer R, Weijer J, Bagdanov A, Vanrell M, Lopez A (2012) Color attributes for object detection. In: Proceedings of the 2012 IEEE conference on computer vision and pattern recognition (CVPR), pp 3306–3313. https://doi.org/10.1109/CVPR.2012.6248068

Knight W (2017) Reinforcement learning. MIT Technology Review. https://www.technologyreview.com/s/603501/10-breakthrough-technologies-2017-reinforcement-learning/. Accessed 12 June 2018

Knight W (2018) Intel's new chips are more brain-like than ever. MIT Technology Review. https://www.technologyreview.com/s/609909/intels-new-chips-are-more-brain-like-than-ever/. Accessed 12 June 2018

Lawrence S, Giles C, Tsoi A, Back C (1997) Face recognition: a convolutional neural-network approach. IEEE Trans Neural Netw 8(1):98–113. https://doi.org/10.1109/72.554195

Le QV, Ranzato M, Monga R, Devin M, Chen K, Corrado GS, Dean J, Ng AY (2012) Building high-level features using large scale unsupervised learning. arXiv preprint: 1112.6209v5 [cs.LG]

LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444. https://doi.org/10.1038/nature14539

Lemaignan S, Warnier M, Sisbot E, Clodic A, Alami R (2017) Artificial cognition for social human–robot interaction: an implementation. Artif Intell 247:45–69. https://doi.org/10.1016/j.artint.2016.07.002

Lipton ZC (2017) The mythos of model interpretability. arXiv preprint: 1606.03490v3

Loehlin JC (1998) Latent variable models: an introduction to factor, path, and structural analysis, 3rd edn. Lawrence Erlbaum Associates Publishers, Mahwah

Lowe D (1999) Object recognition from local scale-invariant features. In: Proceedings of the Seventh IEEE International Conference on Computer Vision, pp 1150–1157. https://doi.org/10.1109/ICCV.1999.790410

Maturana D, Scherer S (2015) VoxNet: a 3D convolutional neural network for real-time object recognition. In: Proceedings of the 2015 IEEE/RSJ international conference on intelligent robots and systems (IROS). https://doi.org/10.1109/IROS.2015.7353481

Merriam-Webster (2016). The Merriam-Webster dictionary, New Edition. Merriam-Webster, Massachusetts

Metz C (2017) AI is about to learn more like humans—with a little uncertainty. Wired, Business. https://www.wired.com/2017/02/ai-learn-like-humans-little-uncertainty/. Accessed 12 June 2018

Mitchell T (1997) Machine learning. McGraw-Hill, New York

Ojala T, Pietikäinen M, Harwood D (1996) A comparative study of texture measures with classification based on featured distributions. Pattern Recognit 29(1):51–59. https://doi.org/10.1016/0031-3203(95)00067-4

Olah C, Satyanarayan A, Johnson I, Carter S, Schubert L, Ye K, Mordvintsev A (2018) The building blocks of interpretability. Distill J. https://doi.org/10.23915/distill.00010

Pierson H, Gashler M (2017) Deep learning in robotics: a review of recent research. Adv Robot 31(16):821–835. https://doi.org/10.1080/01691864.2017.1365009

Pratt GA (2015) Is a cambrian explosion coming for robotics? J Econ Perspect 29(3):51–60. https://doi.org/10.1257/jep.29.3.51

Priddy K, Keller P (2005) Artificial neural networks: an introduction. SPIE Publications, Washington

Rokach L, Maimon O (2014) Data mining with decision trees: theory and applications, 2nd edn. World Scientific Publishing Company, Singapore

Russell S, Norvig P (2009) Artificial intelligence: a modern approach, 3rd edn. Pearson, London

Santoro A, Bartunov S, Botvinick M, Wierstra D, Lillicrap T (2016) One-shot learning with memory-augmented neural networks. arXiv preprint: 1605.06065v1 [cs.LG]

Seemann E, Nickel K, Stiefelhagen R (2004) Head pose estimation using stereo vision for human-robot interaction. In: Proceedings of the sixth IEEE international conference on automatic face and gesture recognition. https://doi.org/10.1109/AFGR.2004.1301603

Silver D et al (2016) Mastering the game of Go with deep neural networks and tree search. Nature 529:484–489. https://doi.org/10.1038/nature16961

Steinwart I, Christmann A (2008) Support vector machines. Springer, Berlin

Stevenson A (2010) Oxford dictionary of English, 3rd edn. Oxford University Press, Oxford

Szepesvári C (2010) Algorithms for reinforcement learning. Morgan and Claypool Publishers, California

Thorndike EL (1911) Animal intelligence. Macmillan, New York

Thrun S, Leonard JJ (2008) Simultaneous localization and mapping. In: Siciliano B, Khatib O (eds) Springer handbook of robotics. Springer, Berlin, Heidelberg

Thrun S, Burgard W, Fox D (2005) Probabilistic robotics. MIT Press, Massachusetts

Wang L, He D (1990) Texture classification using texture spectrum. Pattern Recognit 23(8):905–910. https://doi.org/10.1016/0031-3203(90)90135-8

Wang X, Xu H, Wang H, Li H (2008). Robust real-time face detection with skin color detection and the modified census transform. In: IEEE international conference on information and automation, pp 590–595. https://doi.org/10.1109/ICINFA.2008.4608068

Watkins C, Dayan P (1992) Q-learning. Mach Learn 8(3–4):279–292. https://doi.org/10.1007/BF00992698

Wu Y et al (2016) Google's neural machine translation system: bridging the gap between human and machine translation. arXiv preprint: 1609.08144v2

Yan H, Ang MH, Poo A (2014) A survey on perception methods for human-robot interaction in social robots. Int J Soc Robot 6(1):85–119. https://doi.org/10.1007/s12369-013-0199-6

Yin X, Xie M (2007) Finger identification and hand posture recognition for human–robot interaction. Image Vis Comput 25(8):1291–1300. https://doi.org/10.1016/j.imavis.2006.08.003

Yosinski J, Clune J, Bengio Y, Lipson H (2014) How transferable are features in deep neural networks? In: Proceedings of the 27th international conference on neural information processing systems (NIPS'14), pp 3320–3328

Edirlei Soares de Lima is Assistant Professor at Universidade Europeia, Portugal. He holds a Ph.D. degree in Informatics from Pontifical Catholic University of Rio de Janeiro (PUC-Rio), an M.Sc. degree in Computer Science from Federal University of Santa Maria (UFSM), and a bachelor's degree in Computer Science from Contestado University (UnC). His areas of research include Artificial Intelligence, Computer Graphics, and Games. During his career, his work and research received several awards and honors, including best paper awards at SBGames 2016 (XV Brazilian Symposium on Computer Games and Digital Entertainment), ICEC 2015 (14th IFIP International Conference on Entertainment Computing), and WebMedia 2014 (20th Brazilian Symposium on Multimedia and the Web). In 2011, his research on video-based interactive storytelling received an honorable mention on "Innovation" from the International Telecommunication Union (ITU) and, in 2012, he received another honorable mention from the ITU, at this time on "Interactivity".

Bruno Feijó is Associate Professor of the Department of Informatics at Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Brazil, and CNPq researcher Level 1. He is founder and coordinator of ICAD/VisionLab (Research Lab on Visualization, Digital TV/Cinema, and Games) of PUC-Rio and co-founder of the Special Commission of Games and Digital Entertainment of the Brazilian Computer Society. He holds a Ph.D. from the University of London/Imperial College and an engineering degree from Technological Institute of Aeronautics (ITA). His areas of research include Computer Graphics, Artificial Intelligence, and Games.