AI in the 21^{st} Century – With Historical Reflections

Max Lungarella[1], Fumiya Iida[2], Josh C. Bongard[3], and Rolf Pfeifer[1]

[1] Dept. of Informatics, University of Zurich, Switzerland
[2] Computer Science and Artificial Intelligence Lab, MIT, USA
[3] Dept. of Computer Science, University of Vermont, USA
lunga@ifi.uzh.ch, iida@csail.mit.edu, josh.bongard@uvm.edu,
pfeifer@ifi.uzh.ch

Abstract. The discipline of Artificial Intelligence (AI) was born in the summer of 1956 at Dartmouth College in Hanover, New Hampshire. Half of a century has passed, and AI has turned into an important field whose influence on our daily lives can hardly be overestimated. The original view of intelligence as a computer program – a set of algorithms to process symbols – has led to many useful applications now found in internet search engines, voice recognition software, cars, home appliances, and consumer electronics, but it has not yet contributed significantly to our understanding of natural forms of intelligence. Since the 1980s, AI has expanded into a broader study of the interaction between the body, brain, and environment, and how intelligence emerges from such interaction. This advent of embodiment has provided an entirely new way of thinking that goes well beyond artificial intelligence proper, to include the study of intelligent action in agents other than organisms or robots. For example, it supplies powerful metaphors for viewing corporations, groups of agents, and networked embedded devices as intelligent and adaptive systems acting in highly uncertain and unpredictable environments. In addition to giving us a novel outlook on information technology in general, this broader view of AI also offers unexpected perspectives into how to think about ourselves and the world around us. In this chapter, we briefly review the turbulent history of AI research, point to some of its current trends, and to challenges that the AI of the 21st century will have to face.

1 Introduction

For a long time, humans have been romanced by the idea of understanding how thinking works, or how to construct intelligent machines and replicate the intelligent behavior displayed by many natural systems. Traditional Jewish mysticism, for instance, includes tales of the Golem, a thinking automaton made from the sticky clay of the bank of the river Moldau. In the 17th century, philosopher Gottfried Wilhelm von Leibniz outlined plans for a thinking machine by conceiving an artificial universal language composed of symbols, which could stand for objects or concepts, and logical rules for their manipulation. A little more than half a century ago (when Norbert Wiener was devising Cybernetics [1] and Gray Walter was building robotic tortoises [2], the English mathematician Alan Turing proposed a much-discussed imitation game used as a yardstick for assessing if a machine is intelligent or not, which since then has been known as the Turing Test for artificial intelligence [3, 4].

M. Lungarella et al. (Eds.): 50 Years of AI, Festschrift, LNAI 4850, pp. 1–8, 2007.

The advent of the general-purpose computer in the fifties of last century did substantially alter the dreams and ambitions of people and made AI an even more tangible possibility. So, in 1956, the fathers of “modern” AI, Marvin Minsky, John McCarthy, Allen Newell, Nathaniel Rochester, Claude Shannon, and Herbert Simon came together for a summer school at Dartmouth College (Hanover, New Hampshire) under the premise “that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it” [5]. That date can be considered the birth of AI because it is only thereafter that numerous research groups around the world began to engage in the construction of artificial systems with the professed goal of emulating, equaling, or even surpassing human mental and physical abilities (a different view on the history of AI can be found in [6]). The many attempts to synthesize intelligence or higher cognition (by formalizing knowledge and crystallizing cognitive principles obtained from the study of human beings) have resulted in many specialized AI systems that nowadays are at work in specific problem domains, such as knowledge discovery and data mining (for example, for the identification of customer profiles), fuzzy logic, probabilistic reasoning, and artificial neural networks (as used in intelligent control for robots, home appliances and consumer electronics), evolutionary optimization (for the design of antennae, circuits, and turbines), as well as statistical and machine learning (often employed in genomics and proteomics).

Although, as these examples show, AI has proven successful in many cases, it is clear that the original goals set out by the first generation of AI visionaries have not yet been reached. The holy grail of AI, the creation of general (human-like) machine intelligence (think of HAL from Stanley Kubrick’s *2001* or *Bladerunner*’s Roy Batty), has remained elusive. Natural intelligence is far from being understood and artificial forms of intelligence are still so much more primitive than natural ones. Seemingly simple tasks like object manipulation and recognition – tasks easily solved by a 3-year-old – have not yet been realized artificially. A look at the current research landscape reveals how little we know about how biological brains achieve their remarkable functionalities, how these functionalities develop in the child, or how they have arisen in the course of evolution. Also, we do not understand the cultural and social processes that have helped shape human intelligence. Because basic theories of natural intelligence are lacking and – despite impressive advances – the required technologies for building sophisticated artificial systems are yet not available, it is not surprising that the capabilities of current robots fall far short of even very simple animals. The inevitable conclusion then is that something important must have been missed or still needs to be discovered [7]. There is still no overall theory, no framework, explaining what thinking is and how to build intelligent machines.

Although we do not have any clear cut answer as to what aspects of natural intelligence are key for creating artificial intelligence, we will point at some avenues that might be worth pursuing in realizing the dreams of our ancestors.

2 Trends

The field of artificial intelligence has changed dramatically over the last 50 years. Back in the 1950s, the brain was conceptualized as some kind of powerful computer, and intelligence was thought to be the result of a symbol-crunching computer program located somewhere in our brain (this perspective is also known as the "brain as a computer metaphor" or the "information processing view"; [8, 9]). Since then, the by now "classical", computational, or cognitivistic approach has grown into a large discipline with many facets, and has achieved many successes in applied computer science and engineering. Think, for instance, of the victory of IBM's chess-playing supercomputer (Deep Blue) on reigning world chess champion, Garry Kasparov, on May 1997. While this early view of intelligence is probably adequate in formal or computational domains like chess playing or logical reasoning, in the past 20 years it has become increasingly evident that conceptualizing intelligence as a mere computational process cannot explain natural, adaptive forms of intelligence. The latter kind of intelligence requires a complete physical organism interacting with the real world: in other words, intelligence requires a body [10-15].

The new approach to understanding intelligence has led to a paradigm shift which emphasizes the physical and information-theoretical implications of embodied adaptive behavior, as opposed to the disembodied view of the computational framework [16-19]. The focus in this new paradigm is on systems acting in the real, physical and social world [20]. The implications of this change in perspective are far-reaching and can hardly be overestimated. With the fundamental paradigm shift from a computational to an embodied perspective, the kinds of research areas, theoretical and engineering issues, and the disciplines involved in AI have also changed substantially. The research effort in the field, for instance, has shifted towards understanding the lower level mechanisms and processes underlying intelligent behavior, as well as realizing higher forms of intelligence by first starting with simpler ones [21]. Cognition and action are viewed as the result of emergence and development rather than something that can be directly built (i.e. programmed) into a robot (such research draws heavily on insights from child development; e.g. [17, 22-24]. Automated design methods using ideas borrowed from biological evolution and ontogeny have also provided novel insights into the general nature of intelligence [25].

Physical agents in the real world, whether biological or artificial, are highly complex, and their investigation requires the cooperation of researchers from many different fields. In terms of research disciplines contributing to AI, we observe that in the classical approach computer science, psychology, philosophy, and linguistics played major roles, whereas in the embodied approach, computer science and philosophy [26] are now complemented by robotics, biomechanics [17, 27], material science and biology [28] and neuroscience [29]. The shift from high-level considerations (as raised in psychology and linguistics) to lower level sensory-motor processes – with the neurosciences now offering explanations at both the sensory-motor and cognitive levels of intelligence – is evident [30, 31]. There have also been changes concerning the notions used for describing the research area: a significant number of researchers dealing with embodiment no longer refer to themselves as working in artificial intelligence but rather in robotics, engineering of adaptive systems, artificial life, adaptive locomotion, bio-inspired systems, or neuroinformatics. But more than that, not only have researchers in artificial intelligence migrated into neighboring fields, but researchers in these neighboring fields have also started to contribute to artificial intelligence [29, 32]. This brings us to the issues of cross-disciplinarity and cross-fertilization.

3 Cross-Disciplinarity

Although most researchers in AI today are still working within the boundaries of the field as it was defined 50 years ago, the trend seems to be an expansion of AI into a broader study of the interaction between body, brain, and environment, and how intelligence emerges from such interaction [18, 33]. It has become evident that many fields (linguistics, cognitive sciences, robotics, material science, neuroscience, and morphogenesis) are highly relevant in order to advance the state of the art. The community is interested in locomotion, manipulation, and, in general, how an agent can act successfully in a dynamic and complex world. Concepts such as embodiment, autonomy, situatedness, interactive emergence, development, semiotic dynamics, and social interaction have come to the forefront and have spawned a spectrum of new and interdisciplinary approaches to the study of natural and artificial intelligence. There are important implications of the embodied view of intelligence for science, engineering, economics, and society at large. Here, we only mention but a few.

One of the important insights is that intelligence is not a “box” sitting inside the brain, but is distributed throughout the organism and requires the organism to interact with and explore its local environment. With respect to engineering, it has been demonstrated that systems which are not centrally controlled or do not feature a strict hierarchical structure tend to be more robust and adaptive against external perturbations [34]. It is important to realize that the underlying concepts might be directly mapped onto robot design (e.g. for de-mining operations, waste cleanup in hazardous environments, environmental exploration, service in hospitals and homes, and so on) and onto the design of embedded systems (i.e. systems equipped with sensors and motors which interact autonomously and continuously with their surrounding environment). The application of concepts from embodied AI in embedded systems could have profound economical and technological implications, because such systems are extremely widespread: they are used, for instance, in fuel injection devices, water purification plants, air conditioning systems, remote monitoring and control systems, as well as in many systems designed for human computer interaction. Finally, companies, societies and political bodies with an organizational structure based on local interactions and self-organization have been shown to react more robustly to unpredictable market forces than hierarchically organized ones (a good example of this is the financial difficulties currently experienced by many major airlines, compared to the proliferation and success of smaller, regional airlines).

4 Challenges and Outlook

Niels Bohr once famously quipped: “It’s hard to predict, especially the future”. Similarly, it is difficult to make predictions, especially about the future of artificial intelligence. However, after 50 years of explorations into artificially intelligent systems (and many false predictions), we have learned enough about the nature of intelligence to venture plausible guesses as to where AI could go as a field and which challenges it might face in the coming decades. In this section, we list some of these challenges and point to chapters in this volume that discuss them.

It seems to be generally accepted nowadays that the behavior of any system physically embedded within a particular environment is not the mere outcome of an internal control structure (e.g. the central nervous system) but is also affected by the system's morphology and the its material properties . Increasing evidence suggests that a direct link exists between embodiment and information: coupled sensory-motor activity and body morphology induce statistical regularities in sensory input and within the control architecture and therefore enhance internal information processing [35]. Mechanisms need to be created for the embodied and embedded systems to collect relevant learning material on their own [36] and for learning to take place in an "ecological context" (i.e. with respect to the environment). These and related issues are addressed in the chapters by Pfeifer et al. [18] and Polani et al. [19]. Potential avenues of how to exploit the interaction between control structure and physical properties (morphology and material characteristics) are described in the chapters by Iida et al. [16] and Behkam and Sitti [18].

Another challenge will be to devise a systematic theory of intelligence [14]. As argued by Nehaniv et al. [17] and Polani et al. [19], a promising approach may have to place special emphasis on information theory as a descriptive and predictive framework linking morphology, perception, action, and neural control. It will also be important to understand how raw (uninterpreted) information from unknown sensors can be used by a developing embodied agent with no prior knowledge of its motor capabilities to bootstrap cognition through a process of autonomous self-structuring in response to a history of self-motivated interaction with a rich environment [18, 19, 37]. This is closely connected to the issue of self-motivation and curiosity which are discussed in the chapters by Kaplan and Oudeyer [22] and Lipson [25]. Research on general capacities such as creativity, curiosity, motivations, action selection, and prediction (i.e. the ability to foresee consequence of actions) will be necessary because, ideally, no tasks should be pre-specified to a robot, but only an internal "abstract" reward function, some core knowledge, or a set of basic motivational (or emotional) "drives." Robots need to be endowed with the capacity to explore which are the activities that are maximally fitted to their current capabilities. Robots controlled by some core motives will hopefully be able to autonomously function in their environments not only to fulfill predefined tasks but also to search for situations where learning happens efficiently.

In the same vein, it will be necessary to address the issue of how robots (and embodied agents in general) can give meaning to symbols and construct meaningful language systems (semiotic systems). A fascinating new insight – explored under the label of "semiotic dynamics" – is that such semiotic systems and the associated information structure may not be static entities, but are continuously invented and negotiated by groups of people or agents which use them for communication and information organization [9]. This quite naturally leads to investigation of the collective dynamics of groups of socially intelligent robots and systems [20].

A further challenge will be how to integrate artificial intelligence and other fields. For instance, Potter [29] argues that neuroscience and artificial intelligence could profit from each other and invites researchers in both fields to venture across the divide (e.g. tools developed in the context of neuroinformatics could be used to design and manage neural databases). Similarly, Steels [9] predicts that many of the insights gained in fields such as nonlinear dynamical systems theory or evolutionary biology may have a direct bearing on the foundations of artificial intelligence as well as on how communication systems and other forms of interaction self-organize. A closer (symbiotic) interaction between artificial and natural systems will also benefit rehabilitation engineering as shown by recent work on human-machine interaction [38, 39]. As envisioned by Fattori et al. [31], robotics and neurophysiology could meet in a research field where bioelectrical signals obtained by single or multi-electrode recordings can be used to drive a robotic device [40]. As a last challenge, one could envision situations where AI could be used in the context of computer-assisted human-human interaction technologies [41].

5 Epilogue

The reconsideration of brain and body as a fundamental unit, physically and informationally, as well as the emergence of a new quantitative framework that links the natural and artificial domains, has begun to produce new insights into the nature of intelligent systems. While much additional work is surely needed to arrive at or even approach a general theory of intelligence, the beginnings of a new synthesis are on the horizon. Perhaps, finally, we will come closer to understanding and building human-like intelligence. The superior adaptability of embodied, distributed systems has been acknowledged for a long time, but there is now theoretical evidence from artificial intelligence research and corroboration from computer simulations supporting this point. In summary, the ideas emerging from the modern, embodied view of artificial intelligence provide novel ways of approaching technological, social, and economic problems in the rapidly changing world of the 21st century.

References

1. Wiener, N.: Cybernetics or Control and Communication in the Animal and the Machine. MIT Press, Cambridge, MA (1948)
2. Walter, G.W.: An imitation of life. Scientific American 182(5), 42–45 (1950)
3. Turing, A.M.: Computing machinery and intelligence. Mind 59(236), 433–460 (1950)
4. Legg, S., Hutter, M.: Tests of machine intelligence. LNCS, vol. 4850. Springer, Heidelberg (2007)
5. McCorduck, P.: Machines Who Think, 2nd edn. A.K. Peters Ltd., Natick, MA (2003)
6. Schmidhuber, J.: 2006: Celebrating 75 years of AI – history and outlook: the next 25 years. LNCS, vol. 4850. Springer, Heidelberg (2007)
7. Brooks, R.A.: The relationship between matter and life. Nature 409, 409–411 (2001)
8. Nilsson, N.: The physical symbol system hypothesis: status and prospects. LNCS, vol. 4850. Springer, Heidelberg (2007)
9. Steels, L.: Fifty years of AI: from symbols to embodiment – and back. LNCS, vol. 4850. Springer, Heidelberg (2007)
10. Brooks, R.A.: New approaches to robotics. Science 253, 1227–1232 (1991)
11. Chiel, H., Beer, R.: The brain has a body: adaptive behavior emerges from interactions of nervous system, body, and environment. Trends in Neurosciences 20, 553–557 (1997)
12. Clark, A.: Being There – Putting Brain, Body, and World Together Again. MIT Press, Cambridge, MA (1997)
13. Pfeifer, R., Scheier, C.: Understanding Intelligence. MIT Press, Cambridge, MA (1999)
14. Pfeifer, R., Bongard, J.C.: How the Body Shapes the Way we Think – A New View of Intelligence. MIT Press, Cambridge, MA (2007)

15. Varela, F.J., Thompson, E., Rosch, E.: The Embodied Mind. MIT Press, Cambridge, MA (1991)
16. Iida, F., Pfeifer, R., Seyfarth, A.: AI in locomotion: challenges and perspectives of underactuated robots. LNCS, vol. 4850. Springer, Heidelberg (2007)
17. Nehaniv, C.L., Mirza, N.A., Olsson, L.: Development via information self-structuring of sensorimotor experience and interaction. LNCS, vol. 4850. Springer, Heidelberg (2007)
18. Pfeifer, R., Lungarella, M., Sporns, O., Kuniyoshi, Y.: On the information theoretic implications of embodiment – principles and methods. LNCS, vol. 4850. Springer, Heidelberg (2007)
19. Polani, D., Sporns, O., Lungarella, M.: How information and embodiment shape intelligent information processing. LNCS, vol. 4850. Springer, Heidelberg (2007)
20. Dautenhahn, K.: A paradigm shift in artificial intelligence: why social intelligence matters in the design and development of robots with human-like intelligence. LNCS, vol. 4850. Springer, Heidelberg (2007)
21. Huelse, M., Wischmann, S., Manoonpong, P., von Twickel, A., Pasemann, F.: Dynamical systems in the sensorimotor loop: on the interrelation between internal and external mechanisms of evolved robot behavior. LNCS, vol. 4850. Springer, Heidelberg (2007)
22. Kaplan, F., Oudeyer, P.-Y.: Intrinsically motivated machines. LNCS, vol. 4850. Springer, Heidelberg (2007)
23. Lungarella, M., Metta, G., Pfeifer, R., Sandini, G.: Developmental robotics: a survey. Connection Science 15(4), 151–190 (2003)
24. Lungarella, M.: Developmental robotics, Scholarpedia, p. 173175 (2007)
25. Lipson, H.: Curious and creative machines. LNCS, vol. 4850. Springer, Heidelberg (2007)
26. Froese, T.: On the role of AI in the ongoing paradigm shift within the cognitive sciences. LNCS, vol. 4850. Springer, Heidelberg (2007)
27. Boblan, I., Bannasch, R., Schulz, A., Schwenk, H.: A human-like robot torso ZAR5 with fluidic muscles: toward a common platform for embodied AI. LNCS, vol. 4850. Springer, Heidelberg (2007)
28. Behkam, B., Sitti, M.: Bacteria integrated swimming robots. LNCS, vol. 4850. Springer, Heidelberg (2007)
29. Potter, S.: What can AI get from neuroscience? LNCS, vol. 4850. Springer, Heidelberg (2007)
30. Arbib, M., Metta, G., van der Smagt, P.: Neurorobotics: from vision to action. In: Siciliano, B., Khatib, O. (eds.) Springer Handbook of Robotics – ch. 63, Springer, Berlin
31. Fattori, P., Breveglieri, R., Nicoletta, M., Maniadakis, M., Galletti, C.: Brain area V6A: a cognitive model for an embodied artificial intelligence. LNCS, vol. 4850. Springer, Heidelberg (2007)
32. O'Regan, K.J.: How to build consciousness into a robot: the sensorimotor approach. LNCS, vol. 4850. Springer, Heidelberg (2007)
33. Vernon, D., Furlong, D.: Philosophical foundations of enactive AI. LNCS, vol. 4850. Springer, Heidelberg (2007)
34. Kitano, H.: Systems biology: a brief overview. Science 295(2002), 1662–1664 (2002)
35. Lungarella, M., Sporns, O.: PLoS Comp. Biol., e144 (2006)
36. Bongard, J.C., Zykov, V., Lipson, H.: Resilient machines through continuous self-modeling. Science 314, 1118–1121 (2006)
37. Bonsignorio, F.P.: Preliminary considerations for a quantitative theory of networked embodied intelligence. LNCS, vol. 4850. Springer, Heidelberg (2007)
38. Hernandez-Arieta, A., Kato, R., Yu, W., Yokoi, H.: The man-machine interaction: FMRI study for an EMG prosthetic hand with biofeedback. LNCS, vol. 4850. Springer, Heidelberg (2007)

39. Clark, A.: Re-inventing ourselves: the plasticity of embodiment, sensing, and mind. J. of Medicine and Philosophy 32, 263–282 (2007)
40. Lebedev, M.A., Nicolelis, M.A.L.: Brain-machine interfaces: past, present, and future. Trends in Neurosciences 29, 536–546 (2006)
41. Weibel, A., Bernardin, K., Woelfel, M.: Computer-supported human-human multilingual communication. LNCS, vol. 4850. Springer, Heidelberg (2007)