

Quo vadis, Intelligent Machine?

Rosemarie Velik

Biorobotics and Neuro-Engineering Department, Fatronik – Tecnalia,
Paseo Mikeletegi, 7 - Parque Tecnológico E-20009 Donostia - San Sebastián, Spain
Institute of Computer Technology, Vienna University of Technology,
Karlsplatz 13, 1040, Vienna, Austria
rvelik@fatronik.com

Abstract

Artificial Intelligence (AI) is a branch of computer science concerned with making computers behave like humans. At least this was the original idea. However, it turned out that this was no easy task to solve. This article aims to give a comprehensible review on the last 60 years of artificial intelligence taking a philosophical viewpoint. It is outlined what happened so far in AI, what is currently going on in this research area, and what can be expected in the future. The goal is to mediate an understanding for the developments and changes in thinking in the course of time about how to achieve machine intelligence. The clear message is that AI has to join forces with neuroscience and other brain disciplines in order to make a step towards the development of truly intelligent machines.

Keywords: Artificial intelligence, brain modeling, cognitive computation, brain-inspired automation, neuro-symbolic networks, affective computation, neuroscience, psychoanalysis.

1. Introduction

Over the last 60 years, basically two disciplines have evolved in the area of artificial intelligence (AI), having quite disparate points of view, approaches to solutions, and (often implicit) definitions of the same terms. This article aims to give a review on the development of this research field from the beginning to the present and an outlook to its future. Considerations are not intended to focus on mathematical definitions or particular AI techniques, but on the mediation of an awareness of what changes this research area went through in order to better understand what we can expect in the future.

When talking about artificial intelligence, there first has to be answered the question what artificial intelligence actually is – or at least what it should be. The definitions that can be found about artificial intelligence are various. To specify this area in one sentence, a definition as given by [1] seems to be appropriate: “*Artificial intelligence is a branch of computer science concerned with making computers behave like humans.*” Taking a look on the web or on non-scientific press, the concept that non-engineers and non-scientists often have about artificial intelligence is heavily influenced by film industry. One classical example is the android Data in Star Trek: An artificial, intelligent being, possessing similar or superior intelligence as humans have and maybe lacking a bit of emotions and feelings. Their expectation is that research is not too far away from this aim. What comes up with this expectation are concerns and discussions about ethics. On the one hand, if machines get superior intelligence as we have, there arises the fear that machines will rule the world, kill us, or keep us as slaves (as we see in movies like Terminator or Matrix). On the other hand, if we assume that machines will not possess superior intelligence but feelings and a soul, questions of ethics will come up concerning the rights of machines, as we might keep them as slaves for doing our work. Scientists talking to public press are often confronted with this question of ethics. One answer given in such an interview at the Engineering-Neuropsychology-Forum (ENF) in Vienna 2007 by Gerhard Zucker, one of the keynote speakers, is the following: “*Society will find a way to handle these issues like society found solutions before for slavery and animal testing*” [2]. Actually, this is a wise answer for a scientist in terms of publicity. It implies that our research is already in a very advanced stage and it is time to face these questions. However, reality

is still very different. Today, science is still far away from the goal of creating something that is at least close in intelligence to humans and it is not yet sure whether we will ever reach this goal. To understand where we currently are and what we can expect in the future, first, one must take a look at the history of artificial intelligence.

2. A Brief History of AI

The wish to create artificial intelligent life might be as old as mankind. Already in the ancient Greece, we can find myths about the Greek god of technology Hephaestus who was lame and therefore constructed two golden robots to help him move about. Another well-known narrative is the one of the British scientist Frankenstein who designed a human being from scavenged body parts. Such literature clearly shows that the intention to create human-like intelligent beings was already there for at least some thousand years, only the required means were missing. This situation seemed to change after 1950. Round 1950, it was the time of the advent of the first computers. With their processing power, they offered completely new possibilities. For the first time, the dream of designing an electronic brain seemed to be realizable. The research field of artificial intelligence started to emerge [3].

2.1 The Golden Years (1950-1975)

The first years of artificial intelligence were marked by great successes. It was the era of discovery and sprinting across new ground. The programs developed during this time were, to most people, simply "astonishing". Computers were playing chess, solving algebra problems, proving theorems in geometry, and learning to speak English. Few people at that time would have believed that such "intelligent" behavior of machines was possible at all. Machines were seemingly easily executing "cognitive" tasks that were difficult even for humans. A lot of public money was invested into this promising area, and researchers were very optimistic that a fully intelligent machine would soon be built. The following well-known statements may best catch the spirit of this time [4]:

- 1965, H. A. Simon: *"Machines will be capable, within twenty years, of doing any work a man can do."*
- 1967, M. Minsky: *"Within a generation ... the problem of creating 'artificial intelligence' will substantially be solved."*
- 1970, M. Minsky: *"In from three to eight years we will have a machine with the general intelligence of an average human being."*

The question that logically arose at the same time with the attempt of building intelligent machines was the question of how to prove intelligence of machines. There was the need for a certain evaluation mechanism. The most prominent and best accepted evaluation mechanism suggested for this purpose at that time was the so-called Turing test designed by Alan Turing in 1950 [5]. The original version of the Turing test is illustrated in figure 1a. There is one person, C, that can be regarded as judge. Furthermore, there exist two persons, A and B (one female and one male), that cannot be seen by person C, because they are in different rooms. Person C communicates with person A and B (according to certain rules) via a computer and afterwards has to decide from the communication which person is the woman and which the man. To evaluate machine intelligence, this Turing test was adapted for computers (see figure 1b). Again, person C is the judge. This person now communicates with a computer and another person, and has to decide which one is the human and which the computer. In later times, the Turing test has been criticized a lot [6]. However, at that time, it was the standard measure for evaluating machine intelligence and shall therefore be considered as such for the moment. The interesting question that rises with this is if any computer passed the Turing test? Starting with 1966, this question had to be answered positively.

From 1964 to 1966, John Weizenbaum worked on a computer program called ELIZA [7, 8], which actually passed the Turing test. ELIZA is a program that mimics a Rogerian psychotherapist. This means it confronts its opponent with slight modulations of its own words formulated as questions to achieve the impression that it speaks freely.

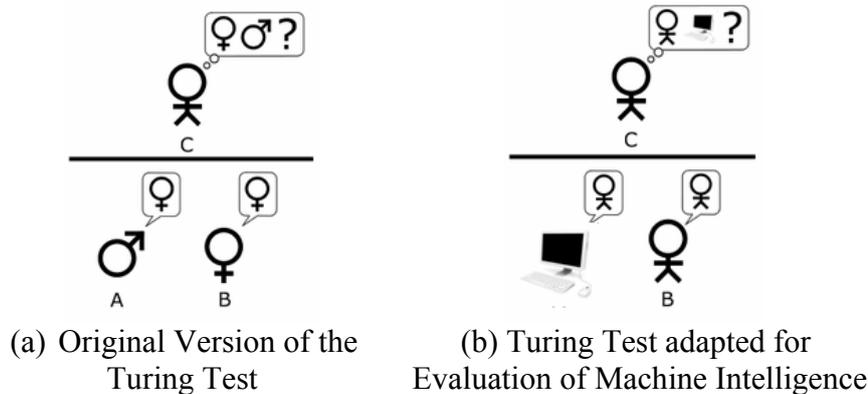


Figure 1. The Turing Test Principle

As the facts so far outlined show, round 1970, researchers were very optimistic that machines would soon (latest in one generation) reach human intelligence level. There already existed the first programs passing the Turing test, which was the official test for proving computer intelligence. Looking at statements of that time today and recognizing that approximately 40 years have passed since then, the logical question that arises is why there are still no intelligent machines among us.

2.2. The Years of Reconsideration (1975-2000)

In 1996, round 30 years after the statement of M. Minsky that “*within a generation ... the problem of creating 'artificial intelligence' will substantially be solved*”, a young scientist called Push Singh, who happened to work under M. Minsky, published an article with the title: “*Why AI failed*”. This fact clearly illustrates that after the first years of enthusiasm, AI went through a change. AI began to get stuck. Researchers had to admit that making computers actually think – even on a childlike level – was far more complicated than they expected [4]. One explanation for this could be that in the first years, scientists focused on problems that were difficult for humans (like playing chess, solving algebra problems) and therefore seemed to be particularly challenging for intelligent machine design. Researchers generally considered constrained problems and problem domains. They had the illusionary hope that when accumulating all single efforts together, soon an intelligent machine would emerge. They did so far not put emphasis on problems that were easy for humans like e.g., perceiving their surroundings, evaluating complex situations (what is currently important?), and taking decisions in real world environments. When starting to consider these issues, it turned out that they were very difficult to implement into a computer.

3. The Current State of AI

Today, artificial intelligence has split into two branches. These branches are probably best referred to as method-based artificial intelligence and brain-inspired artificial intelligence. Nowadays, many researches are still not aware of the existence of these two sub-disciplines which are in fact very disparate in their basic dogmas. The following section outlines the main principles and differences of both sub-disciplines.

3.1 Method-based AI

During the history of artificial intelligence, it had to be admitted that creating human-like intelligence is far from trivial. When recognizing that creating truly intelligent machines seemed to be almost infeasible, researchers started to focus on simpler and more constraint problems. The goal was no longer to achieve a machine with a human intelligence level for all circumstances but to develop particular solutions for particular problems. Examples of classical methods of method-based AI are symbolic systems, expert systems, genetic algorithms, artificial neural networks, etc. This classical AI domain is a mature research field. Hundreds of textbooks can be found about these methods. Problems that are solved with these methods are pattern recognition problems (image processing, language processing, etc.), prognoses, path planning, etc. These problems are solved by certain mathematical models and algorithms but have hardly anything to do with how the brain works and solves these problems or with human intelligence. One fact that might be surprising is that also tools like artificial neural networks are assigned to this category as they are inspired from the function of biological neurons in the brain. Artificial neural networks however do not emulate a neural brain network but just quite simplified neurons. The way in which artificial neurons are interconnected has not much to do with how interconnection takes place in the brain, which rather seems to be the secret of the complexity of the brain than the function of single neurons alone [9].

3.2. Brain-inspired AI

So far, there is no technical system that can even nearly compete with the capacity and the capabilities of the human mind. Within the last years, it had to be admitted that the reduced approaches often focused on in classical method-based AI can never lead to technical systems with skills and capabilities comparable to humans' mental abilities. Therefore, as stated in [9], "*like at the beginning of artificial intelligence research, again, findings about how natural intelligence works have to be the basis for developing concepts for technical approaches trying to achieve intelligence*". This is the basic dogma of the new generation of brain-inspired AI approaches. Here, archetypes for model development of intelligent systems are the structure, functional systems, and information processing strategies of the brain. Approaches followed in this area are various and based on different disciplines of brain research like neuroscience, psychology, pedagogy, psychoanalysis, etc. The following section is aimed at mediating a basic understanding of what research efforts are currently going on in this research community based on some concrete projects realized within the last 10 years by an about 25-headed interdisciplinary research team of the Vienna University of Technology.

3.3. Showcase Projects of Brain-inspired AI

"*Current technology will not be able to handle future demands of automation systems*" [10]. With this statement, in the year 2000, Dietmar Dietrich, the head of the Institute of Computer Technology of the Vienna University of Technology, formed an interdisciplinary research team consisting of electrical engineers, computer scientists, physicists, neuroscientists, psychoanalysts, and psychologists with the aim of developing next generation intelligent automation systems. As the human brain can be regarded as the most efficient, effective and flexible known automation control system, the basis for development should be insights about the brain gained from neuroscience, psychology, neuro-psychology, psychoanalysis, and neuro-psychoanalysis. To do so, the team had to enter an unknown territory and establish new ways of thinking and interdisciplinary collaborations. The following section will briefly present three of the models that were the outcome of this interdisciplinary effort. One of the application areas for the models is building automation in order to create more "intelligent" situation-aware buildings that can automatically adapt and (re-)act according to what is currently going on in their environment [11, 12, 13]. This domain is today also referred to as "intelligent buildings", "digital home", "ubiquitous computing" or "domotics". The

basic idea is to equip buildings with sensors and actuators of different types (similar to the way in which our body is equipped with different sensory receptors for perception and limbs for actuation) and recognize what is going on in the building in order to decide how to react on these situations and carry out selected actions. This is of interest for safety and security critical applications, to save energy, to improve the user's comfort, and for user entertainment [14, 15]. A second application field is autonomous agents and robots that have to navigate independently in their environment and manipulate objects. Also for them, an effective and efficient perception-decision-making-action-execution system is necessary [16, 17, 18].

3.3.1. Autonomous Decision Making based on Emotions, Drives, and Memory

The first model that is briefly introduced is an approach inspired by research findings in neuro-psychology and psychoanalysis (see figure 2). The model considers the environment, the body, and the brain/mind of an artificial being. The environment and the body are perceived via sensors. (Re-)actions can be carried out via actuators. The model focuses on the decision-making process of how to (re-)act according to situations currently perceived. In this decision-making process, there are involved concepts like emotions, desires, drives, and different types of memory as well as concepts like the Ego-Id-Superego model of Sigmund Freud. There are considered fast reactions in form of reflexes and slower reactions that need reflection, thinking, and planning. For a detailed model description, see [14, 17, 18].

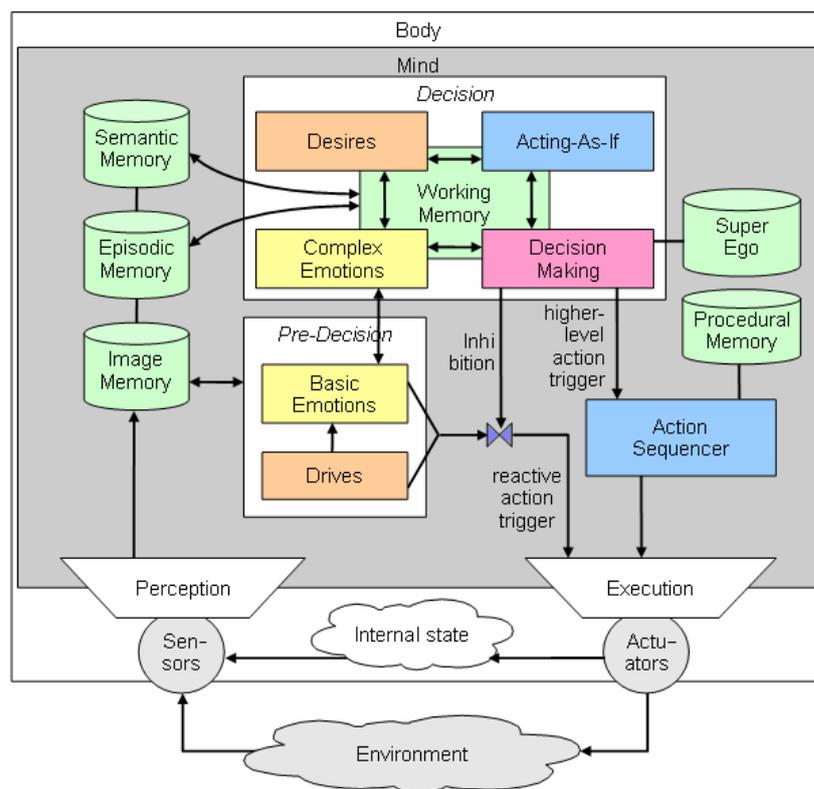


Figure 2. Decision Making Model based on Emotions, Drives, and Different Memory Types

3.3.2. Machine Perception based on Neural and Symbolic Information Processing Strategies

The second model developed is a model for human-like machine perception based on research findings in neuroscience and neuro-psychology. The principal idea of the model is to use so-called neuro-symbols as basic processing units. This concept is inspired by the fact that the brain

is made up of neurons but we think in term of symbols. In analogy to the brain, starting from sensor values, the sensory information is combined and condensed in a modular hierarchical manner to more and more complex neuro-symbolic information until this results in a complete, unitary, multimodal perception of the environment (see figure 3). For a detailed explanation of the model, see [9, 19, 20, 21].

3.3.3 Neuro-Psychoanalysis for Modeling the Mind

The third approach is based on research findings in psychoanalysis and the Ego-Id-Superego model of Sigmund Freud [22, 23]. The particularity of this model is that it is based on top-down design strategies. The basic idea is to start from the function of the whole brain and then divide the brain functions like in a top-down approach into different modules starting from the Id, Superego, and Ego (see figure 4).

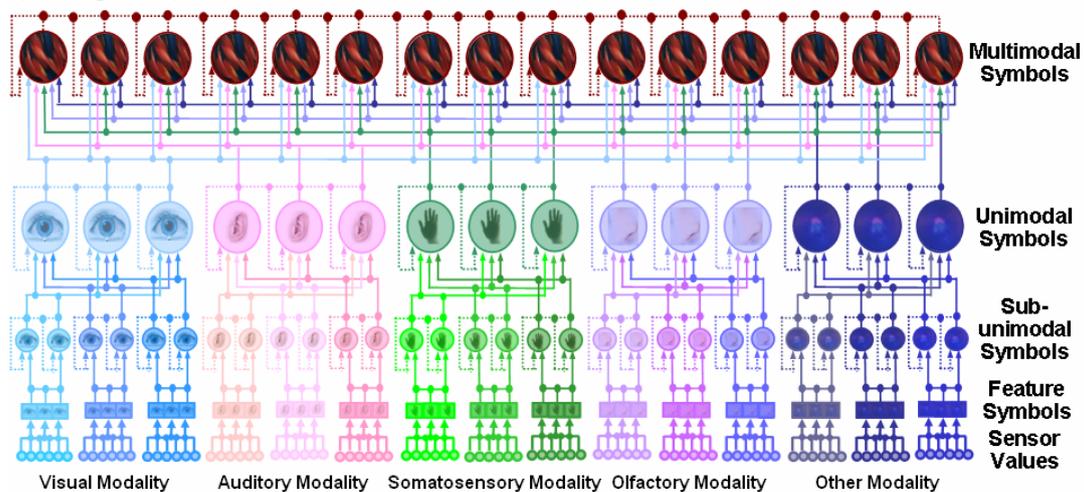


Figure 3. Combination of Neural and Symbolic Information Processing Strategies

The Id is correlated with internal drives of the body like hunger, fear, or sexuality. The Superego contains social rules like “I must not kill.” or “I have to wear clothes in society.” The Id and Superego are always in conflict tending to different behavior. The task of the Ego is to mediate between them and decide what to do. According to the top-down design strategy, these three blocks are further subdivided into other psychic functions until ideally the neural level is reached. Details about this model can be found in [2, 24].

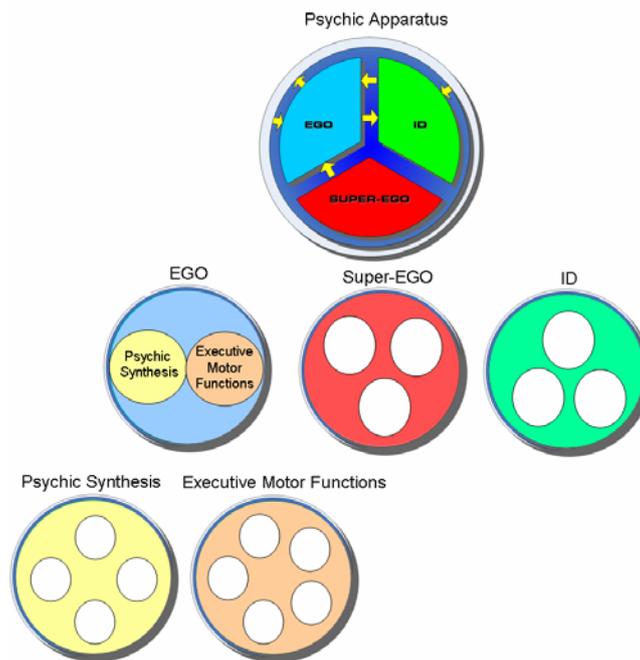


Figure 4. Applying Neuro-Psychoanalysis for building a Technical Model of the Brain

4. Future Perspectives of AI

Having sketched some examples of what is currently going on in the research area, we shall now attempt to give a short outlook on what we can expect in the future and what are the great challenges to face.

Concerning the vision on how the research area of artificial intelligence will develop, there exist two different opinions – a pessimistic and an optimistic one. According to the pessimistic view, we will never be able to build machines similarly intelligent to humans. The main reason for this is that we will not be capable of understanding how the brain works, or even that there is more about the brain and the mind than just a huge bunch of neurons interacting with each other. The computer pioneer Prof. Heinz Zemanek, who also formed part of the early artificial intelligence community, used to say: “If one light switch is not intelligent, why 1000 should be” and he generally added: “I built a computer. I can tell you that there is nothing intelligent in it... and if you call the computer intelligent, then I am not, then I am something else.” [9]

On the other hand, there also exist more optimistic views considering the task of emulating the human brain as feasible, maybe not with today's computer technology but with technologies using the structural organization and information processing principles of the human brain as archetype. According to Etienne Barnard, there exist two different possibilities for how this research field will progress [25]. One possibility is that as until now, small but continuous progress will be made. The other possibility is that the next Albert Einstein – the Einstein of Artificial Intelligence – will appear and a big leap forward will be made.

Assuming that the optimistic views hold true, an outlook on a number of challenges that will have to be faced in the research area of artificial intelligence in the future should now be given. For sure, the most challenging goal of research in this area is to achieve consciousness of machines. Consciousness of an individual is its subjective experience – to know what it is like to be oneself [26]. Metaphorically speaking, the aim is to create a machine that one day opens its eyes and asks us “Who am I?” and maybe adds “And who are you?” It does not ask, because we have programmed it to do so, but because it is aware of itself as an individual living being. Today, there are many discussions going on in different research communities concerning conscious machines. The fact is

that we are still far away from this goal. Modeling consciousness is last but not least a problem because it is subjective and not measurable. According to [26], not even a human being can be sure about the consciousness of other human beings. Consciousness cannot be technically modeled by adding a further function block (the consciousness block) to a brain model. It rather emerges from the sum of all other physiological and mental functions [27]. Therefore, to model consciousness, the focus of research has first to be directed towards issues that are a prerequisite for it. We shall further mention some of these issues.

One issue that we might have to integrate into machines is constituted by emotions. So far, the role of emotions in thinking and intelligence has widely been ignored. New research results however show that they have a major influence on our thinking and our decision-making processes [20, 28]. In the first model introduced in the last section, as well as in models of some other research groups, the concept of emotions was already integrated. Nevertheless, emotions are still a topic needing further thorough investigation from both the perspectives of brain research and engineering.

A second point is the embodiment of machines. According to the theory of embodied intelligence, we can never be intelligent and conscious without having a living body that has needs, that has sensors to perceive its environment and its body, and actuators to act on the body and the environment [18].

A third point that might have to be considered when creating artificial intelligent life is survival and reproduction. It is an uncontestable fact that the brain becomes useless as soon as the organism dies. According to the neuroscientist and psychoanalyst Mark Solms, who tries to give the meaning of life on a scientific basis, the purpose of life is "survival for reproduction" [26]. He further outlines what is necessary to achieve this goal: We as human beings with our body live in an environment – the world. To survive and reproduce, we need to get from the environment food and a partner of the opposite sex for reproduction. For this purpose, we have to be able to perceive the environment and to act on it. All these tasks of perceiving, of evaluating what was perceived, and reacting accordingly are controlled by the brain. The task of the brain is to mediate between the internal needs (I am hungry, I want social interaction) of our body and the environment in which our needs can be satisfied. The brain perceives the environment and the internal needs of our body, evaluates these perceptions, decides what to do, and prepares signals for (re-)acting on the environment. As the whole organism and the brain seem to be designed to achieve the basic goal of survival for reproduction, it might not make sense or even not be feasible to design true artificial intelligence without considering this issue.

Having mentioned reproduction, a related topic for investigation is evolution. Our brain did not evolve from one day to the other but is the result of millions of years of optimization processes through variation and selection. Therefore, it might not be the best way to rely on "intelligent design" only but rather to make available mechanisms for self-optimization as it is the case in evolution.

In conclusion, it can be said that the research domain of biologically and brain-inspired artificial intelligence is by far not saturated. It is an area where still astonishing discoveries can be made, secrets can be unveiled, and new grounds can be broken. To achieve this, engineers will have to join forces with brain scientists and life scientists and carry out research in a tight collaboration.

5. References

- [1] Wall, B. (2009), Artificial Intelligence and Chess. <http://www.geocities.com/SiliconValley/Lab/7378/ai.htm>.
- [2] Dietrich, D., Sauter, T. (2000), Evolution Potentials for Fieldbus Systems. *Proc. 3rd IEEE International Workshop on Factory Communication System (WFCS'00)*, 343-350
- [3] McCarthy, J., Minsky, M., Rochester, N., Shannon, C. (1955), A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence

- [4] Crevier, D. (1993), AI: The Tumultuous Search for Artificial Intelligence, *Basic Books*, New York.
- [5] Turing, A. M. (1950), Computing Machinery and Intelligence, *Mind*, 59, 433-460.
- [6] French, R. M. (1990), Subcognition and the Limits of the Turing Test, *Mind*, 99 (393), 53-66.
- [7] Weizenbaum, J. (1966), ELIZA – A Computer Program For the Study of Natural Language Communication Between Man And Machine, *Communications of the ACM* 9 (1), 36–45, January.
- [8] Bruckner, D. (2007), Probabilistic Models in Building Automation: Recognizing Scenarios with Statistical Methods, *PhD Thesis*, Vienna University of Technology.
- [9] Velik, R. (2008), A Bionic Model for Human-like Machine Perception, *PhD Thesis*, Vienna University of Technology.
- [10] Dietrich, D., Sauter, T. (2000), Evolution Potentials for Fieldbus Systems, *Proceedings of 3rd IEEE International Workshop on Factory Communication Systems*, 343-350.
- [11] Pratl, G. (2006), Processing and Symbolization of Ambient Sensor Data, *PhD Thesis*, Vienna University of Technology.
- [12] Burgstaller, W. (2007), Interpretation of Situations in Buildings, *PhD Thesis*, Vienna University of Technology.
- [13] Velik, R. (2008), A Bionic Model for Human-like Machine Perception, *VHS Verlag*.
- [14] Velik, R., Zucker, G. (2010), Autonomous Perception and Decision Making in Buildings, *IEEE Transactions on Industrial Electronics*, article in press.
- [15] Velik, R., Boley, H. (2010), Neuro-symbolic Alerting Rules, *IEEE Transactions on Industrial Electronics*, article in press.
- [16] Deutsch, T., Lang, R., Pratl, G., Brainin, E., Teicher, S. (2006), Applying Psychoanalytic and Neuro-scientific Models to Automation, *Proc. of the 2nd IET International Conference on Intelligent Environments, Vol. 1*, 111–118.
- [17] Roesener, C. (2007), Adaptive Behavior Arbitration for Mobile Service Robots in Building Automation, *PhD Thesis*, Vienna University of Technology.
- [18] Palensky, B. (2008), From Neuro-Psychoanalysis to Cognitive and Affective Automation Systems, *PhD Thesis*, Vienna University of Technology.
- [19] Velik, R., Bruckner, D. (2009), A Bionic Approach to Dynamic, Multimodal Scene Perception and Interpretation in Buildings, *International Journal of Intelligent Systems and Technologies, Vol. 4*, 1.
- [20] Velik, R. (2010), Why Machines Cannot Feel, *Minds and Machines, Volume 20*, Issue 1 1-18.
- [21] Velik, R. (2010), Towards Human-like Machine Perception 2.0, *International Review on Computers and Software*, article in press.
- [22] Freud, S. (1915), Triebe und Triebchicksale, In: *Studienausgabe Band 3: Psychologie des Unbewussten*, Fischer Taschenbuch Verlag.
- [23] Freud, S. (1923), Das Ich und das Es, In: *Studienausgabe Band 3: Psychologie des Unbewussten*, Fischer Taschenbuch Verlag.
- [24] Dietrich, D., Fodor, G., Zucker, G., Bruckner, D. (2008), Simulating the Mind: A Technical Neuropsychological Approach, Springer Verlag.
- [25] Lang, R. (2010), A Decision Unit for Autonomous Agents Based on the Theory of Psychoanalysis, *PhD Thesis*, Vienna University of Technology, Institute of Computer Technology.
- [26] Barnard, E., Palensky, B., Palensky, P., Bruckner, D. (2008), Towards Learning 2.0, *Proceedings of IT Revolutions, Venice*.
- [27] Solms, M., Turnbull, O. (2002), The Brain and the Inner World – An Introduction to the Neuroscience of Subjective Experience, *Other Press*, New York.

- [28] Velik, R. (2010), From Single Neuron-firing to Consciousness – Towards the True Solution of the Binding Problem, *Neuroscience and Behavioral Review*, Volume 34, Issue 7 993-1001.
- [29] Damasio, A. R. (1995), Descarte's Error: Emotion, Reason, and the Human Brain, *Harper Perennial*.