# Towards an Artificially Intelligent System: Philosophical and Cognitive Presumptions of Hybrid Systems

Ondřej Vadinský
*Department of Information and Knowledge Engineering*
*University of Economics, Prague*
*náměstí Winstona Churchilla 4, Prague 3, Czech Republic*
*Email: ondrej.vadinsky@vse.cz*

***Abstract*—This contribution summarizes philosophical and cognitive presumptions of intelligence and looks at their realization in paradigms of artificial intelligence with emphasis on hybrid systems. It gives specifications of research concerning analysis of presumptions of intelligence in computer systems. Finally, the paper outlines preliminary research results, regarding the relationship of intentionality and representation and about viewing the strong artificial intelligence in the context of perception, action, learning and development. The paper suggests that there are aspects of intentionality which can be captured by hybrid representation.**

***Keywords*-strong artificial intelligence; hybrid systems; representation; intentionality; stratified model of perception.**

## I. Introduction

In this contribution, some ideas on intelligence in computer systems will be presented, as well as the preliminary results of the research concerning the analysis of the presumptions of intelligence of computer systems.

The main goal in the field of artificial intelligence (AI) is considered to be the creation of so-called *strong AI*, that is, artificial intelligence in all its aspects as powerful as human intelligence. For such an effort to succeed, a thorough analysis of the characteristics of intelligence has to be done to identify its essential aspects and properties [1], [2], [3].

Here, an opportunity arises to use philosophical reflection and knowledge of cognitive sciences about intelligence to make a better AI. This contribution focuses on hybrid systems and hybrid-system-based cognitive architectures. Hybrid systems are systems using both symbolic and connectionist representation of knowledge while cognitive architectures are domain-generic models of human cognition [3], [4], [5].

In Section II, presumptions of intelligence will be detailed, namely from the point of view of philosophy (II-A) and of the cognitive sciences (II-B). Section III will refer to current paradigms of AI, including logical symbolism (III-A), connectionism (III-B) and hybrid systems (III-C). Section IV will discuss how to combine philosophy, the cognitive sciences and artificial intelligence into one multidisciplinary endeavour. The research program will be specified (IV-A) and initial results outlined (IV-B and IV-C). This contribution will be concluded in Section V together with a discussion about the future work.

## II. Presumptions of Intelligence

If an intelligent machine is to be constructed, the following questions should first be considered: What is intelligence and what are its essential properties? How do we decide what is intelligent and what is not? How is intelligence related to other abilities of the mind?

Current scientific study of intelligence within the broader context of the human mind and cognition is underway in the cognitive sciences. The cognitive view on artificial intelligence was discussed in more detail in [6]. Therefore, only the main ideas will be presented in this contribution. Several areas of philosophy also discuss these issues from interesting perspectives, as explored thoroughly in [3]. Only a brief summary will be given here.

### *A. Philosophical Point of View*

In philosophy, there is a long debate about whether or not a machine can think. It can be traced back to Descartes, who raises two presumptions of intelligence: ability of rational speech and universality of thought [7].

This question was famously inverted by Turing. He asks whether a machine can mimic human beahviour and thinking in such a way, that an average human cannot tell who is the human and who is the machine while communicating with both. This is known as *the Turing test* and, as is the case with Descartes, it is strongly related to rational speech [1].

The issue was further addressed by Searle [2] in his attempt to look into the machine in what is known as *the Chinese room argument*. Searle states that for rational speech, *understanding* is needed. Humans have it due to *intentionality*, that is "[a] feature of certain mental states by which they are directed at or about objects and states of affairs in the world." Searle also notes that humans incline to ascribe *intentionality* to others based only on similarity in behaviour, which can lead to serious errors, as is the case with computers.

Pstružina [8] points out that the special settings of both *the Turing test* and *the Chinese room argument* disable many

 ISBN: 978-1-61208-273-8 

aspects of real-life communication and also neglect many aspects of thought that take part in it. He also gives a more elaborate *definition of intentionality* in which the meaning of mental states is given by the integration of intentional concept into the structure of other concepts and endocepts, and also by reflection of this emergence of the meaning.

### B. Cognitive Point of View

A central point of cognitivism is that information processing requires a system to have an internal *representation* of its environment. Such a model is called a *world view*, and as De Mey [9] shows, there are several of them dynamically combined to cognitive schemata. This way a structure combining and organizing knowledge about concepts is created.

Another important issue brought forward by cognitivism is the role of perception. Perception is always understood as an indirect process mediated by a certain *world view*. De Mey mentions the so-called *stratified model of perception*: A subject accesses a perceived object in several layers of granularity, which are partially independent of each other. A percept is, therefore, an amalgam of perceived shapes strengthened by relations of concepts within a subject's *world view*. This means that both the subject and the object contribute to the act of perception [9].

De Mey's cognitivism also stresses the importance of the *connection between knowledge and action*, in which he is inspired by Piaget. Acquiring knowledge is about realizing the interaction between the subject and the object. It goes through several phases, firstly building *implicit* knowledge which is later made *explicit*. As a suitable *representation* of knowledge, De Mey sees Minski's frames. *Frames* create a backbone of stereotypical knowledge in which specific knowledge can be inserted. The structure of *frames* can be recombined during the development of *world views* [9].

## III. Hybrid Systems and the Paradigms of AI

As the initial questions about intelligence have been considered, the ways of making it artificial should now be examined. This includes questions such as: What means can be used to create an AI system? How adequate those means are to the properties of intelligence? Are they limited by some of the properties? A more elaborate description of the means to create an AI system was given in [3]. Also, the question of adequacy of the paradigms was discussed there, having led to the focus on hybrid systems. For the sake of completeness, a brief summary of the paradigms of AI and their adequacy will be given in this section.

As the field of artificial intelligence developed, several paradigms emerged. Logical symbolism has foundations in logic and *explicit symbolic representation*. Trying to solve the shortcomings of symbolism, connectionism draws its inspiration from biological neural networks. Recently, there was a growing effort to combine symbolic and connectionist paradigms into so-called hybrid systems. Aside from these paradigms which are somewhat similar due to their focus on *computation* and *representation*, there is also another paradigm called enacted cognition or enactivism. It seems to be a promising point of view, but it will not be discussed in this contribution.

### A. Logical Symbolism

Logical symbolism operates with a term *physical symbol system* which is a structure of instances of symbols arranged in physical patterns. The meaning of a symbol in a system is given by its connections to other symbols [4].

The issue of meaning is tightly connected with the *grounding problem*: that is, the question of where the meaning of symbols comes from. Rapaport tries to solve this with his *syntactic semantics*. He shows a way how meaning can ensue from connections between symbols of two distinct symbolic systems, say of language and sensual perception. One symbol system can then be grounded in the other [10].

A similar approach as *syntactic semantics* is used in *semantic computing*. The meaning of the data is given by its connection to metadata in the form of ontology.

Symbolism creates *explicit* highly structured and abstract models. Despite several extensions of the original concept, it is problematic to deal with incomplete, vague or noisy data [4].

### B. Connectionism

The Churchlands [11] point out that the *massively parallel architecture* of the human brain is what makes it powerful in the tasks it performs. They suggest making *computers architecturally similar to the human brain*. This could be done by employing artificial neural networks and other connectionist methods.

Connectionism uses great numbers of simple computational units such as artificial neurons connected together into a network. Although a single unit can solve only simple tasks, the network as a whole succeeds in solving much more complex problems [4].

However, the complexity of the human brain is far beyond current connectionist models. There are both quantitative and qualitative aspects in which the models are lacking. An example of the quantitative aspect is the number of neurons and synapses. Some examples of the qualitative aspects include: the way in which the electric impulses arise and in which they are propagated in the network, various kinds of oscillations and continuity of learning. Currently, there are projects underway which try to find out and model the ways in which the human brain works such as the *Blue Brain Project* [12].

Connectionism creates *implicit* models capable of learning. Due to its parallelism it can cope with incomplete, vague or noisy data easily. Such models are, however, difficult to understand by the human observer [4].

 ISBN: 978-1-61208-273-8 

### C. Hybrid Systems

Since both symbolism and connectionism have their strong and weak points, and since a weak point of one is often a strong point of the other, it seems natural to combine them.

Hybrid systems in general use different models of knowledge *representation*. It is possible for the models to be a mere duplication of each other, to take part in different tasks, to cooperate or to compete. The models can be combined in different ways, loosely or tightly. All these architectural decisions influence greatly the properties and capabilities of such a hybrid system [4].

Further on, only such hybrid systems which combine symbolic and connectionist models will be considered. *Combining explicit and implicit representation creates synergy* which makes such systems robust and general-purpose. In such a system, there are processes of internal learning, when knowledge is transferred between different models: *explicit knowledge can gradually descend to the implicit model* while, on the other hand, *implicit knowledge can cause the explicit knowledge to emerge from the implicit model*. Due to its multiple *representations* of knowledge, hybrid systems keep the advantages of both connectionist and symbolic approaches [4].

An example of hybrid systems usage is Sun's *CLARION cognitive architecture*. It strives to create a *domain-generic* model of human cognition. Such a model should be cognitively realistic on social, psychological, componential and physiological levels [5].

## IV. Combining Philosophy, Cognitive Science and Artificial Intelligence

After the initial discussion of the presumptions of intelligence and paradigms of the AI, this paper tries to combine philosophical and cognitive knowledge to make better artificial intelligence. Furthermore, the two research ideas concerning the relation of *intentionality* and *representation* in hybrid systems and the context of perception and action in hybrid systems will be sketched.

### A. Research Specification

The long-term goal related to my research is to create such an artificially intelligent system that would be as powerful as human intelligence. Such artificial intelligence has been called *strong AI* by Searle. Recently, it has also been called *general artificial intelligence* in a new attempt to bring back this project from the beginnings of artificial intelligence.

The analysis of presumptions of intelligence of computer systems includes three goals:

- To examine different possible ways of creating a *strong AI*;
- To analyze architectures of hybrid systems suitable for creating a *strong AI*;
- To propose and verify an extension of a chosen architecture.

To reach these goals, a multidisciplinary approach is needed. Therefore, inspirations will be drawn from the fields of philosophy, cognitive sciences and artificial intelligence. Many specialists have examined related partial topics, but without much effort to integrate their results. However, such an integration and synthesis is crutial if we are to reach the long-term goal.

### B. Intentionality and Representation in Hybrid Systems

Let us review Pstružina's definition of *intentionality*. In [8], he describes three aspects of *intentionality*:

- Directedness at things or aboutness;
- Gaining the meaning through integration into the structure of other intentional concepts and endocepts;
- Realization of the previous two aspects.

Are these aspects of *intentionality* somehow connected to the *hybrid representation*?

Let us first consider the second aspect of *intentionality*. The hypothesis is that the structure of intentional concepts can be understood as the symbolic model of a hybrid system. Also, the structure of endocepts can be understood as the connectionist model of a hybrid system. The process of integration of an intentional concept into such structures can then be identified with the process of adding new symbols into the symbolic model or the process of learning of the connectionist model. Through its position in a symbolic system, concept gains a certain aspect of meaning. The meaning of a concept is not only about connections of a symbol to other symbols, although those connections among symbols certainly participate in the emergence of meaning. This meaning is further enriched by the connections between the symbolic and connectionist model, and by the connectionist model itself.

A brief look at the aboutness aspect of *intentionality* will now be presented. The preliminary hypothesis states that to represent and to be directed at are two sides of the relation between the representing and represented thing. Therefore, there is a kind of directedness and aboutness in a *representation*. The question is whether it is the same, or to what degree it is similar or different.

As for the realization of the previous two aspects of *intentionality*, a hypothesis has not yet been formulated. This aspect is connected to consciousness and requires further study.

### C. Model of Stratified Perception in Hybrid Systems

An important issue when considering *intentionality* is its origin. As has been mentioned by Searle in [2], it can come from the artificial system itself, or from the human observer. Searle's conclusion is the latter. But can the former be somehow achieved?

*Robotic answer to the Chinese room argument* shows a promising way. Having a computer in a robotic body connected to sensors and actuators enables it to interact with the world. This should ground its *representation* of the world [2].

The robotic answer can be further specified, if a hybrid system is considered as the architecture of the computer system. Then, an implementation of a *stratified model of perception* can be used to integrate the information into the hybrid system in relation to its own expectations. Finally, Piaget inspired learning of knowledge from the interaction with the world through concretization of *implicit* knowledge can be used. Such implementation of artificial intelligence respecting the context of perception, action, learning and development should give it a more solid *grounding* than just manipulation with *representation*.

## V. Conclusion and Future Work

In this contribution, philosophical presumptions of intelligence were summarized. The main focus of philosophical reflection of intelligence lies in the ability of rational speech and universality of thought. These are tightly connected with the concept of *intentionality*, as used by Searle and Pstružina. Other notable concepts include the *Turing test*, though in need of extension, and the *Chinese room*.

Furthermore, cognitive presumptions of intelligence as understood by De Mey were summarized. His position emphasizes the concept of *world views* and the importance of perception and its integration with the *world views* in a *stratified model of perception*. On top of that, it shows a special link among knowledge, perception and action, in which knowledge comes from realization of the interaction of the subject and object.

Symbolic and connectionist paradigms of artificial intelligence were described and it was shown how they can be combined in *hybrid systems*. These are useful for generic *cognitive architectures* due to synergy between symbolic and connectionist knowledge *representation*.

Since the focus is on hybrid systems together with *strong AI*, as it have been specified, two ideas about how to improve them using previously described presumptions of intelligence were pursued. First, the ideas about *intentionality* and *representation* were sketched. Then, Pstružina's definition of intentionality was used to show similarities between those concepts. Furthermore, the ideas about viewing the hybrid-system-based AI in the context of perception, action, learning and development were outlined.

Future work in several areas of the research is needed. First, the ideas mentioned in this contribution need some more development. Especially, the aspect of intentionality related to consciousness is in a need of thorough examination. Furthermore, a way of putting all three aspects together in a hybrid system should be devised. Also, the research of other presumptions of intelligence in philosophy and cognitive sciences should be done as part of the first, most theoretical, research goal. In the next phase the research should focus on analysis of existing architectures of hybrid systems. Theoretical findings of the first stage should be compared to the analysis, so that in the third phase concrete ways to apply them can be found.

The other area related to the research is to propose ways to validate the results. This may pose a serious chalenge, since opposing factors need to be reconciled. As mentioned in [5], testing cognitive architectures is a very demanding process due to their general-purpose nature. As the goal of the research is to provide an extension of a hybrid system to make it closer to *strong AI*, similar or even more adverse cicumstances as with testing cognitive architectures are to be met. Therefore, it may be possible to adapt some approaches and experiments which are used to validate cognitive architectures. However, a mere comparison of performances in several experiments may not be enough. Currently, hopes are placed in existing extensions of *the Turing test*.

## References


[1] A. M. Turing, "Computing machinery and intelligence," *Mind*, vol. 59, no. 236, pp. 433–460, 1950.

[2] J. R. Searle, "Minds, brains, and programs," *Behavioral and Brain Sciences*, vol. 3, no. 3, pp. 417–457, 1980.

[3] O. Vadinský, "Na cestě k uměle inteligentním systémům: Řešení vybraných filozofických a kognitivních předpokladů inteligence [Towards an artificially intelligent system: Addressing selected philosophical and cognitive presumptions of intelligence]," in *Kognice a umělý život XII*, J. Kelemen and P. Nahodil, Eds. Prague: ČVUT, 2012, pp. 238–242.

[4] R. Sun, *Artificial intelligence: Connectionist and symbolic approaches*. Oxford: Pergamon/Elsevier, 2001, pp. 783–789.

[5] ——, "The importance of cognitive architectures: An analysis based on clarion," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 19, no. 2, pp. 159–193, 2007.

[6] O. Vadinský, "Kognitivní pohled na umělou inteligenci [Cognitive view on the artificial intelligence]," *E-Logos*, no. 17, pp. 1–12, 2012, [retrieved: 2013-03-28]. [Online]. Available: http://nb.vse.cz/kfil/elogos/mind/vadinsky12.pdf

[7] R. Descartes, *A Discourse on Method*. Oxford: Oxford University Press, 1637.

[8] K. Pstružina, "Turing v Searleově čínské místnosti [Turing in Searle's Chinese room]," *Acta oeconomica Pragensia*, vol. 11, no. 8, pp. 9–16, 2003.

[9] M. de Mey, *The cognitive paradigm*. University of Chicago Press, 1992.

[10] W. J. Rapaport, *Syntactic semantics: Foundations of computational natural-language understanding*. Dordrecht: Kluwer Academic Publishers, 1988, pp. 81–131.

[11] P. M. Churchland and P. S. Churchland, “Could a machine think? Classical AI is unlikely to yield conscious machines; systems that mimic the brain might,” *Scientific American*, vol. 262, no. 1, pp. 32–37, 1990.

[12] H. Markram, “The blue brain project,” *Nature Reviews Neuroscience*, vol. 7, no. 2, pp. 153–160, 2006.

 ISBN: 978-1-61208-273-8