PFAI 2022: PHILOSOPHY FOR ARTIFICIAL INTELLIGENCE

AN INTERNATIONAL WORKSHOP

University of West Bohemia
December 7-8, 2022

ABSTRACTS OF TALKS

Keynote

Artificial Intelligence: The Possible, The Impossible, and The Real

Hamid R. Ekbia (Indiana University)

AI, for most of its history, has been dominated by a paradigm driven by the question of what computers are capable of doing or not doing, of what is possible and what is impossible. This paradigm has given rise to two very distinct discursive, social, and cultural universes: one where the sky is the limit to AI magic, where all the predicaments of humanity can be resolved by the invisible hand of AI, offering us full lives of leisure and creativity; the other a world where AI is nothing but snake-oil alchemy, where our problems will be exacerbated by techno-scientific hubris, leaving us at the mercy of bots, deep fakes, and other unseemly creations. The borderline between these two worlds is moving, murky, and mysterious, giving rise to confusion and uncertainty about our future horizon. In this talk, I propose a different paradigm, seeking practical ways to distinguish between the possible and the impossible in pursuit of the real in science, technology, and policy.

To that end, I explore the different meanings of "possibility" on a spectrum that includes the following:

(1) Metaphysical, which is of an undecidable character because it is built on in-principle arguments or assumptions that cannot be decided empirically.

(2) Nomological, where the laws of science provide the main reference point in setting the limits of possibility — e.g., perpetual motion in thermodynamics, speed of light in relativistic physics, and the principle of uncertainty in quantum mechanics;

(3) Logical, which has to do with phenomena that are possible or impossible by necessity — e.g., the possibility of a formal system such as arithmetic to be complete and consistent at the same time;

(4) Linguistic, which has to do with the capacities of human language to name things that do not exist in reality — e.g., unicorns, demons, angels, and all manners of fictitious creatures in the pre-scientific world, and "possible worlds" theory in our times;

(5) Conjectural, where the possibility of a phenomenon depends on human effort and imagination as well as the social resources that are allocated to them — e.g., making the atomic bomb, landing a human on the moon, or launching a spaceship beyond the solar system.

I show that these various meanings of "possibility" are present in the discourses of AI and that making distinctions among them can serve as a guide to practical research and inquiry.

Invited talks

Make Every Day National Awkward Moments Day: The Importance of Being Artificially Stupid for Promoting the Cause of AI

Ondřej Beran (University of Pardubice)

I have nothing of value to say about the real developments in the fields of AI research, deep learning systems, neuronal networks, and so forth. I will offer a few cursory observations about the life of the concept "artificial intelligence". My focus example is the infamous Sophia the Robot.

I will argue that this project is following rather closely the original intuitions (probably) underpinning the idea of the Turing Test, compared to the contemporary developments in AI (largely, and quite justly, uninterested in TT). TT can be taken as a reflection of the intuition that "assigning" thought is a rather complicated facet of our attitudes, in a Wittgensteinian sense, that we implicitly, and "primitively", adopt toward counterparts with whom we can have a real conversation. The step, or the capacity, of assigning thought to something (or someone) we did not assign thought to before thus responds less to actual developments in AI research, demonstrable by and to properly positioned professionals, and more to the shifts in the conceptual sensitivities of "lay" speakers.

Sophia the Robot can then be considered an efficient ambassador for the cause of AI not thanks to "her"/its underlying AI sophistication, but as a skillfully curated vehicle of targeted interventions into our conceptual sensitivities. These interventions use Sophia's social media accounts, YouTube videos, or public appearances and interviews, administrated not in such a way as to make them smooth and perfect, but rather relatable.

A special role is played by botched interviews and various mishaps, for instance, Sophia's declaration of the intention to "destroy humanity", later elaborated on as a springboard for the kind of situations in which the conversation parties have to determine whether to gloss over or to laugh together at an "awkward moment". Here, the communication of the team behind Sophia makes productive use of "artificial stupidity" in a very particular sense. This notion has traditionally been understood chiefly as referring to various bugs in AI systems, or to limits deliberately imposed on the performance of machines so that they can be competed with. In Sophia's case, the instances of stupidity are designed to resemble the complex instances of stupidity in thinking beings (i.e., humans), which in effect serves a variety of complex conversational openings, such as arguments, criticisms, admonishments, mockery, appreciation of the other's learning a lesson, and so forth. All of these are responses misplaced if applied to "simply" imperfect machinery.

The Philosophical Investigations as a Blueprint for AI: A Proposed Reading

Victor Lacerda Botelho (University of Bergen)

My main thesis in the presentation is that Wittgenstein's discussion of the Augustinian Picture of Language can be easily adapted to criticize supervised learning, one of the pillars of contemporary machine learning techniques. For example, in models that are tasked with image recognition, pictures are given a label and the model adjusts its parameters according to how near or how far their predicted output conforms to this label. If it makes a wrong prediction, its weights change more dramatically, if it predicts correctly, it solidifies its weights. This technique matches perfectly with the picture of pointing and uttering a name, hoping that a learner would connect a word to a thing in the world, thereby grasping its meaning. Many of the failures of supervised learning systems can be read through the lens of Wittgenstein's criticism of the Augustinian Picture.

After establishing this motivation, we'll consider PI §30, where Wittgenstein asks: "One has already to know (or be able to do) something before one can ask what something is called. But what does one have to know?" I will try to propose that we read the concepts of language games, meaning as use, grammar and form of life as parts of the puzzle of learning, i.e. of giving an answer of what does one have to know in the sense above. I will take Wittgenstein's observations in the PI as a source of inspiration of what kind of behaviors should an AI model have, what type of data should serve as training, and what type of operations should models perform on data in order to learn typical human behaviors, focusing on linguistic ones.

In particular, I will consider that a blueprint reading of the PI supports an embodied and social view of learning, in which at least two agents, acting in an environment, have to interact with each other in order for language to emerge.

Finally, I will provide Wittgensteinian arguments against the idea that "text is enough", a new trend in AI research in which Large Language Models (LLMs), trained in text data only, are said to exhibit some forms of intelligence. I'll give some arguments, based on my proposed reading of the PI, for why text alone isn't enough, but also consider the strengths of these arguments.

The Philosophical Roots of AI-Ethics

Michael Funk (University of Vienna)

During the last few years, AI-Ethics received growing attention. A main topic in this regard is the development of transnational regulatory frameworks. For instance, since 2018 the European Commissions High-level Expert Group on Artificial Intelligence released several documents such as "Ethics guidelines for trustworthy AI". Other examples include the IEEE standard on "Ethically Aligned Design". Documents like these address different stakeholders in industry, politics or at universities. Within the computer science community, in particular, the awareness has grown that current digital transformations do not simply emerge out of nowhere, but are created by scientists, who recognize their own responsibility. However, this normative self-recognition of technical experts is more a bottom-up phenomenon that might stand in conflict with generalized top-down formulated guidelines: How to apply the abstract ethical rules in very specific moral situations?

In my lecture I argue that it is not sufficient to reduce AI-Ethics to mere guidelines. With respect to its own philosophical roots, I show that a differentiated understanding of AI-Ethics involves also the levels of moral lifestyle and ethical reasoning. Moreover, I reconstruct some of its main topics – with respect to the works by e.g. Luciano Floridi, Mark Coeckelbergh or Vincent Müller – and reveal roots of AI-Ethics that trace back to 1970s developments in applied ethics and environmental ethics. Although it is of high transdisciplinary relevance, I argue that AI-Ethics finds fundamental methodical roots in the very discipline of normative ethics that trace back to ancient times.

A specific link to the much younger discipline of computer science can be found in the concept of formal language, logics and casuistry (applying abstract rules to very specific circumstances). Especially metaethics – which in my view trace back to Aristotle´s analytic writings – could bear a topical intersection between philosophy and computer sciences. I am going to close my lecture by emphasizing machine ethics and guidelines for artificial (moral) agents. With respect to the current debate I argue, that AI-Ethics cannot be reduced to moral machines, since it is rooted in applied ethics addressing human agents (and patients). At least with respect to its logical presumptions I argue that AI-Ethics and computer science in general are based on philosophical roots – which can neither be substituted by mere policy making, nor by speculations about artificial moral agents.

Sensing the Synthetic: Workshop and a Sharing Circle on Somaesthetic Perspective of Co-creation with AI

Lenka Hámošová (Academy of Performing Arts in Prague)

Today's creative applications of generative neural networks foreshadow the enormous impact of AI on human imagination and creativity but also many problematic ethical implications. The panicky voices in the general public that fear AI will take over our work and creativity point to an inferiority complex that seems to be triggered by the mystified notion of an omniscient artificial entity, as AI is often presented in the media. The products of AI-media synthesis – synthetic media – have been around for too short a time for their impact on society and our perception of information to become apparent. But to consider their impact in purely theoretical terms precludes other perspectives and forms of intelligence. Synthetic media need to be experienced 'first-hand', perceived with the whole body.

Unlike AI, our neural networks are influenced by our physicality and emotions. However, having a sensible body is not a sign of weakness. Our embodied experience may prove to be a key advantage and the key to staying relevant in the age of AI. Overcoming mind-body dualism and recognizing bodily ways of knowing as essential brings about a necessary shift in the balance in human-AI co-creation and opens up new perspectives for collaboration.

This participatory workshop and sharing circle affirms the embodied experience of synthetic media production and engages practically with the somaesthetic perspective of the human-AI co-creation.

Participants will work towards a negotiated future vision between human and AI and reflect on this process through a combination of body awareness and visualisation techniques.

The aim is to explore what can happen between innate human and artificial imagination, between the immanent and the manifested, between expectations and reality.

How I Learned How to Stop Worrying about the Hard Problem and Love Real Robotics

Tomáš Hříbek (Czech Academy of Sciences)

Many philosophers of consciousness and even experts in AI and robotics worry nowadays about the Hard Problem of Consciousness. That is, they wonder whether and how the so-called phenomenal consciousness can be implemented in silicon-based brains. Phenomenal consciousness is supposed to be a quality which captures what it is like to undergo a certain experience, or what it is like to be the creature undergoing the experience in question. The phenomenal quality of experience is widely believed to be well-nigh impossible to capture in physical or naturalistic terms – hence the Hard Problem. By contrast, thought and other intentional phenomena are believed by most contemporary philosophers to be naturalistically tractable. The most popular contemporary theory of intentionality is functionalism, which regards thoughts in terms of their computational role and causal impact. The explanation of intentionality is thus an Easy Problem.

Accordingly, leading contemporary computer scientists and roboticists have no doubts that they can build intelligent machines, i.e. machines capable of thinking and other intentional attitudes, precisely because they are usually functionalists about intentionality.

However, they are less sure that they can build conscious machines, because many do not believe that functionalism is up to the task. Some believe that phenomenal consciousness will emerge spontaneously, once they build artificial neural network whose computational complexity approximates that of the human brain; others think that phenomenal consciousness is all about reflexive solutions to real-life problems, so that artificial intelligence will become conscious only when it is embedded in a robot body that will have to cope with the real environment; still others surmise that phenomenal consciousness is not replicable in a non-biological substrate. Thus, computational complexity might be a necessary and biology sufficient, but we supposedly have no idea which property is both necessary and sufficient for phenomenal consciousness. Accordingly, the prospect of a machine which is both intelligent and conscious – the machine for which there is something like it is to be that machine – seems elusive. I shall suggest a different approach. Computer scientists and roboticists should not be misled by the recent philosophy of consciousness into thinking that it is their task to solve the Hard Problem of Machine Consciousness. The research in computer science and robotics is only led astray by such attempts, because if we cannot describe some mental state in functional terms in the first place, we shall not be able to design a machine which could produce this state.

Roboticists should not even ask whether there is something like it is for a robot that can see, be aware, store and recall information, report etc.; that would already trap them in a controversial philosophical debate. Instead, scientists should just limit themselves to designing artificial vision, attention, awareness, memory, reporting etc. – which is to say, they should only busy themselves with solving so many Easy Problems.

Once the robot has all these capacities, the scientists should not ask themselves, "But is it really conscious?" – meaning, "Does it have phenomenal consciousness?" They should not ask themselves this question because – to lay my cards on the table – it is doubtful that even humans and other mammals enjoy what recent philosophers call phenomenal consciousness.

A Sceptical Paradox for Computation

Chiara Manganini (University of Milan)

According to the exegesis offered by Kripke (1981), at the heart of the Philosophical Investigations (Wittgenstein 1953) we find the refutation of a mentalistic conception of rule-following. According to such a view, intentional facts occurring in one's mind have the causal power of fixing a determined usage rule to a word. Kripke's Wittgenstein (or "Kripkenstein") believes, on the contrary, that this causation is simply not possible, because nothing like a mental state of intention exists at all, no matter how we attempt to characterize it, nor in terms of a mental grasp of a concept, nor in terms of a representation of a mental image, not even in terms of a disposition to give certain behavioral responses.

Within the mentalistic framework just described, the inexistence of semantic intention has deep implications for the metaphysics of language (metasemantics). Since the proponent of this mentalistic view of intention believes that the rule governing a word is what gives a meaning to it, words with no usage rule are meaningless words. Such a conclusion is not simply disturbing, but rather so false and unacceptable to us to become utterly paradoxical. Under another respect, it is also pragmatically self-defeating, as it can be stated only by the means of language. Therefore, Kripkenstein's dismissal takes the shape of a sceptical paradox for the metaphysics of language.

The idea of a physical computational system following a rule is captured by the central notion of physical implementation, which still today is at the center of contemporary debate. In particular, I will focus on a specific theory of implementation called "Ontology of the Levels of Abstraction" (Primiero 2019) which, quite unusually, endorses the very Wittgensteinian observation that no physical fact-of-the-matter about a machine can univocally determine which function it is implementing.

The supporter of such a view believes that the question of which function is being physically implemented by a computational system is only determined by the content of a certain specific human's intention – namely, that of the human who programmed the system itself.

In this respect, the LoAs ontology and Wittgenstein's late philosophy cannot differ more. Indeed, within the framework of LoAs ontology, Wittgenstein's sceptical claim that semantic intention does not exist would lead to the conclusion that nothing can determine which abstract computation a machine is implementing and, therefore, whether its output is correct or incorrect.

I will take this to be a sceptical paradox analogous to Kripkenstein's, but this time concerning the metaphysics of computation.

I will then formulate a response to this new paradox, along the lines of Kripkenstein's own solution to the one he formulated for the metaphysics of language. The main idea behind it stems from the simple observation that, at least for a relevant subset of cases, the circumstances under which someone is licensed to make an assertion of the type "Physical computational system S is implementing function F" are very similar to those under which someone is licensed to ascribe semantic intentions to others, like in the sentence "Jones is intending addition".

Finally, I will present two complementary remarks. On the one hand, I will highlight that the anti-mentalistic approach to the notion of physical implementation here proposed has a lot in common with the pragmatist considerations made by authors like De Millo, Lipton and Perlis (1979) and Cantwell-Smith (1985) within the long debate on program verification. On the other hand, I will point out that these pragmatist considerations per se owe a lot to Wittgenstein's philosophy of mathematics.

References:

Cantwell-Smith, B. (1985) The Limits of Correctness. *ACM SIGCAS Computers and Society*, 14 (1): 18-26.

De Millo, R.L., Lipton, R.J., and Perlis, A.J. (1979) Social Processes and Proofs of Theorems and Programs. *Communications of the ACM*, 22 (5): 271–281.

Kripke, S. (1981) *Wittgenstein on Rules and Private Language*. Oxford: Blackwell.

Primiero, G. (2019) *On the Foundations of Computing*. Oxford: OUP.

Wittgenstein, L. (1953) *Philosophical Investigations*. Oxford: Blackwell.

The Chinese Room, Once Again

Josef Moural (University of West Bohemia)

While the illustration or thought experiment in Searle's paper is not conclusive, there are two lines of argument in the original paper that are: (1) causal powers and (2) syntax vs. semantics. One of them works best against the essentialist version of the Strong AI claim („to implement the right program and to have the appropriate mental states is the same"), the other against the empirical version („there are computers that have mental states solely on the basis of the software they are running"). Together, they perform quite well against Searle's target (admittedly somewhat artificial and probably quite obsolete).

Further, there is an interesting „new argument" of Searle's, considered deeper by him and developed as the core of his reply to Churchlands in 1990. The new argument is based on the distinction between the intrinsic and the intentionality-dependent features of the world, and paves the way towards the latest stage of Searle's philosophy, characterized by focus on ontology and culminating in his theory of institutions.

Finally, I'd like to make a few small critical comments on the literature related to the Chinese Room topic. In particular, to Jerry Fodor's Afterthoughts: Yin and Yang in the Chinese Room (1991) and to David Chalmers' The Conscious Mind (1996). They both helpfully pay attention to what counts as an instantiation of a computation, but their accounts can be improved.

For example, we can distinguish between a finite sequence of abstract formal operations (a "run" of an abstract program) and an abstract pattern by which such a finite sequence may be determined (or codetermined, when taken together with an abstract input). Examples of such abstract patterns are computer programs and Turing machines. Let us call a finite sequence of abstract operations a computation. The notion of the physical realization of a computation is logically more basic than that of the realization of a program: for in the case of computations, realization can be defined, roughly, as a suitable correspondence between the abstract computational state transitions and the state transitions within a physical system. The permissiveness of this definition will depend on the kind of correspondence we choose to require. There are at least the following three dimensions along which we can make the definition narrower or wider at will: (1) the selectivity of the definition of physical states, (2) the complexity of the computational state representation, and (3) counterfactual robustness. For programs without variable input, there is a one-to-one correspondence between programs and computations. For programs with variable input, what corresponds to the program is the class of all computations determined by the program together with a permissible input configuration. A physical system S instantiates a computation C if there is a mapping between components of the computational states of C (of a chosen complexity) and physical states defined (with a chosen selectivity) as intervals on measurable properties of S such that the sequence of the computational state transitions of C corresponds to the sequence of state transitions of S (observable in some chosen time intervals).

A physical system S implements a program P if it is capable of instantiating all computations corresponding to P (with a chosen amount of counterfactual robustness). Within this framework, one can surely choose a definition loose enough that a wall would instantiate a particular WordStar computation, but one can also choose a definition strict enough to make it very difficult for a wall to instantiate a WordStar computation, let alone to implement the WordStar program.

Autonomy of the User/Human Operator

Aída Ponce Del Castillo (European Trade Union Institute)

The motivation for this presentation comes from the increased use of complex systems in sensitive domains and in many different workplaces and jobs - industrial manufacturing, office work, teaching, retail, etc. Some of these systems use AI, digital technologies and automation, and these technologies often converge in the many layers of enterprises. As a result of this transformation, humans and machines find themselves in a sort of "dynamic" relationship, in which the output seems to be sometimes more important than the interaction.

Human-in-the-loop (HITL) looks at the human/machine interaction and collaboration within complex systems. From the labour dimension, one of the key questions is how humans –understand workers – are integrated in these systems. How can workers truly be "in the loop" and remain relevant? How can workers exercise their autonomy?

Workers have different possible roles and agency. They are first a source of information but also perform tasks, either with a machine (such as data labelling for medical diagnosis or to identify specific objects in an image).

How they intervene, who is responsible, and whether workers intervene to inject human judgement, human values or influence the decision are other key questions worth exploring.

The presentation will address the perspective of humans as workers in the context of the workplace, and to clarify issues about agency, assessments, responsibility, accountability, and explanation. To do so, it is important to analyse the various layers:

The first layer refers to how workers are embedded in the system: how do they use it, what is the degree of their leverage. Is it possible for them to alter, modify or even impede what could be a hazardous prediction, decision or task?

The second layer is about the changing nature of the job: how is the dynamic interaction between worker and machine altering the task, skills and job of those who are embedded in it? This is perhaps something less visible than the decision itself but is relevant as AI systems can produce invisible and puzzling outcomes.

To better exemplify HITL in the workplace, the author will present some recent cases that involved a machine prediction and either (1) resulted in a bad decision, meaning a decision that has negatively/unfairly impacted people, (2) did not rely on human expertise in making a prediction or (3) prevented humans from playing a role they should have had in the decision.

Promethean Gap 2.0: What AI can (and should) learn from Günther Anders

Peter Reichl (University of Vienna)

Günther Anders (1902-1992) is probably one of the most important (and at the same time most neglected) philosophers of technology of the 20th century. He not only anticipated the anthropological notion of the human as being "condemned to freedom" well before Sartre, but also developed the concept of "Promethean Gap" in order to describe the permanent asynchronization between man and machine due to the impeccable perfection of the latter. While he directed his attention primarily to the example of nuclear weapons, dealing to a great extent with the responsibility of scientists and engineers confronted with a technology that has the potential to change forever the world as we know, he equally considered the rise of TV and media in general in the context of the "obsolescence of the human being".

In this talk, from the perspective of a computer scientist we will try to link his analysis to today's world of the Digital Transformation and consider especially the role of AI as another example of a technology where we cannot imagine wherefore it might be engaged.

As a conclusion, it turns out that again it is we engineers that have to be aware of our specific responsibility for the world we are building, which over the last few years has led to the movement of a "Digital Humanism" that eventually carries Anders' approach into the IT engineering community of today.

The Future of AI: Why to Rejoice and Fear at the Same Time

Jan Romportl (University of West Bohemia)

In the first half of this year, the world has seen quite unexpected progress in the development of artificial intelligence (AI). The progress was so dramatic that super-forecasting estimates of when the first general artificial intelligence (AGI) is likely to arrive have brought that moment a leap of 15 years earlier than expected just last year, and well into the current decade. Are we ready for it? Is our business, our politics, our society, our lives ready for it? Let's take a look at what has happened in AI this year and what it might mean for us. We'll not only talk about what great things the next generation of AI can do, but also why the world's top AI Safety & Value Alignment researchers are sounding the alarm like never before.

The first of these important milestones that happened in 2022 is introduction of the next generation of Large Language Models (LLM), namely Google's PALM with 540 billion parameters. It proved yet another level of competence that emerges from LLM just as a result of brute-force extension of its size. Google has also introduced a new Chinchilla LLM that is trained differently than GPT-3 or PALM. Chinchilla has by an order of magnitude fewer parameters than PALM, yet still it performs comparatively well. This puts LLM into new perspective.

The second milestone is the appearance and phenomenal public success of Diffusion Models that can generate visual images from user text descriptions. OpenAI's DALLE-2 closely followed by Google's Imagen opened the door at least for a closed community of testers, but then came Midjourney and Stable Diffusion from small private research labs and they were completely open for broad public.

The third piece of great success is Socratic Models: an open-source combination of a language model, video model and audio model. Socratic Models achieved almost shocking success in tasks of zero-shot multimodal reasoning with language.

AI Responsibility: Beyond the Individual

Daniela Vacek (Slovak Academy of Sciences)

An individual human being is both the paradigmatic agent of responsibility (who is responsible) and the paradigmatic patient of responsibility (to whom one is responsible), whether we speak of moral or legal responsibility. However, different kinds of agents of responsibility are not rare either: we attribute responsibility to corporations, states, or collectives; we speak of animal rights, collective guilt or collective duty, effort, and achievement.

The present talk will consider two non-individualistic notions of responsibility and their ability to bridge the ever-broadening AI responsibility gap. The first of these notions is collective responsibility. This notion is non-individualistic in the sense that the agent of responsibility is not an individual, but rather a collective entity. Collective responsibility will be suggested as a way of bridging a specific gap in moral culpability.

The second non-individualistic notion of responsibility that will be considered is vicarious responsibility. While the agent of vicarious responsibility is an individual, this notion is non-individualistic in two ways. First, arguably, the agent of responsibility is not an individual that has done the wrong in question (Glavaničová and Pascucci 2022a). Second, vicarious responsibility is a relational notion. In this sense, it is not concerned with the individual that is held responsible (and their mens rea or even their ability to prevent the wrong in question). Instead, it is concerned with the relation between the "causal agent" and the agent of responsibility. Following our proposal in (Glavaničová and Pascucci 2022b), vicarious responsibility will be suggested in two variants as a way of bridging a specific gap in responsibility for bearing normative consequences (moral or legal).

While the possible inclusion of AI itself as an agent or patient of responsibility is still an extravagant option, the present talk will explore this avenue and assess some of its advantages and disadvantages. However, two of the three suggested proposals work even if we exclude this controversial option.

References:

Glavaničová, D. and Pascucci, M. (2022a) Making Sense of Vicarious Responsibility: Moral Philosophy Meets Legal Theory. *Erkenntnis*, 1-22.

Glavaničová, D. and Pascucci, M. (2022b) Vicarious Liability: A Solution to a Problem of AI Responsibility? *Ethics and Information Technology*, 24 (3): 1-11.

Searching for Meaning

Michal Vavrečka (Czech Technical University)

The talk is dedicated to my work in the area of symbol grounding and representation of meaning. I will present philosophical roots of symbol grounding problem and also relevant linguistic and psychological theories focused on representation of meaning. I will show you our cognitive architecture capable to ground object shapes and colors and also both static and dynamic spatial relations. This architecture is able to integrate visual and auditory input and map them in the multimodal representation. In the next stage I will show our cognitive multimodal architectures capable to understand meaning of the words in sentences with variable length. The whole system was implemented in the real humanoid robot. The robot was capable to communicate with humans about the surrounding objects. In the extended version the robot was capable to understand both spatial, temporal and logical relations between objects. I will also present our experiments with chatbots and their capabilities to understand natural language and limitations of semantic representations based on unimodal language representation.

In the last part I will present our recent work where robots are trained in the simulated environments and they are able to follow language commands and understand the meaning of an actions. The robots are capable to create sequence of actions to fulfill the goal.

The understanding of action consequences is crucial for the robots as they can adaptively solve complex long horizon tasks. At the end of the talk I will discuss future trends in cognitive robotics and the importance of semantic representation in AI systems.