

On a Scientific Discipline (Once) Named AI

Wolfgang Bibel

Darmstadt University of Technology

Bibel@gmx.net

Abstract

The paper envisions a scientific discipline of fundamental importance comparable to Physics or Biology, reminding that a discipline of such a contour was originally intended by the founders of Artificial Intelligence (AI). AI today, however, is far from such an encompassing discipline sharing the respective research interests with at least half a dozen of other disciplines. After the analysis of this situation and its background we discuss the consequences of this splintering by means of selected challenges. We deliberate thereby what could be done to alleviate the disadvantages resulting from the current state of affairs and to leverage AI's current prominence in the public attention to re-engage in the field's broader mission.

1 Introduction

The contours of a scientific discipline are formed by two different sources. One source consists in the discipline's subjects and the methods used for their explorations. In a natural science such as Physics this source remains rather stable over time, thus resulting in a rather stable contour.

The other source forming a discipline consists of the people working in the discipline, ie. its community, and their interests. These may change considerably over time, to a major extent also by the successes achieved in certain parts or by the actual challenges within the discipline. In the case of an established discipline like Physics these dynamics exert a relatively minor effect on the discipline's contour.

Not so in a field like Artificial Intelligence (AI). While AI's pioneers seemed to have a clear and rationally founded vision about the field as a scientific discipline with well-defined subjects and research methods, this source has lost some of its influence on AI's contours. So AI's contours have to a larger degree been influenced by the successes achieved in applying the accumulated scientific knowledge to applications with relevance for people's use as well as for industry, business and administration. This is true particularly during hypes like the present AI hype worldwide. In consequence, the contours of AI have shifted considerably during its short history of somewhat more than six decades.

At the surface one might regard these shifts as the fruitful search for the right direction in AI research. Unfortunately, this lack of a stabilizing influence from the source characterized by a common understanding of the subjects and methods has disadvantageous consequences which, in the case of AI, can be observed in various aspects pointed out in the sections below. In short, this is because the scientific discipline envisioned by AI's pioneers today has become spread over about half a dozen of other disciplines who compete with AI for harvesting the honors (and profits) going along with the subjects' successful explorations. This distribution was not induced by internal structures of the subject under investigation in any natural way but was rather driven by the influence and interests of historically grown communities. The resulting lack of exchanging insights among those competitors retards the progress and leads to plenty of redundancy in the research efforts, disadvantages which in a more coherent disciplinary constellation could easily be avoided.

Whether or not this trend towards a rationally unjustified splintering could yet be reversed is hard to judge. Even if it could be reversed in principle it is difficult to say at this point in time, how this could be achieved. The present article is therefore meant to provide just an analysis of the present state of affairs rather than pointing out a concrete plan for changing directions except for some indications in view of the given possibilities. It thereby continues the discussion started already with the articles [Bibel, 2008] and [Bibel, 2014]. In detail, we proceed as follows.

In the subsequent section we characterize the discipline as it was envisioned by the pioneers and as many of us still see it after six decades of remarkable progress in our research. We do so by identifying the subject of inquiry which in this case is by its very nature elusive and hardly tangible. In terms of the methods used for exploration we roughly distinguish two related levels of inquiry consisting of the lower process level and the higher functional level. In Sect. 3 we very briefly sketch the history of the emergence of our discipline. Sect. 4 is meant to convince the readers that, because of the splintering of the research efforts, the present situation is less than optimal, both for AI as well as for scientific progress in general. However, AI could leverage its current prominence in the public attention to re-engage in its broader mission. To this behoof the section lists five selected challenge areas which definitely require AI expertise, but seem

to have slipped away to other disciplines at least to a large extent or suffer from lack of collaboration within AI. In the final Conclusions section we indicate possible choices how to proceed into the future.

2 Vision and Subject of an Emerging Discipline

Before discussing the current situation of AI we would like the reader to share the view of a fundamental scientific discipline that has emerged along with AI in the last century. Therefore, in this section, we first characterize this discipline within a longterm perspective and then roughly distinguish two levels of explorations in it.

2.1 The Evolution of Science

During thousands of years of experience humans have learned that the chances for survival and well-being are the better, the deeper an understanding of the world has been attained. This fundamental wisdom is true today, as it was for thousands of years.

Only since the last half millenium have we found out that an understanding of the world can reliably be attained especially by way of scientific methods including experiments for testing theories preferably formalized in mathematical terms. This methodological discovery by Galilei, Newton and others has led to an accelerating accumulation in our understanding of the world.

Roughly, we may distinguish three succeeding steps in this scientific evolution during the last centuries. Natural science first acquired scientific knowledge mainly about Physics, i.e. about matter, its structure, its motion and behavior through space and time, and about related entities of energy and force. While there are still fundamental open problems even after about 500 years of research in classical and modern Physics, our understanding of this aspect of the world has become truly overwhelming.

During the last two centuries Natural Science then additionally focussed on the study of life and living organisms, i.e. on biological phenomena and mechanisms such as cells, DNA, reproduction, metabolism, evolution, and so forth. Again the accumulated biological knowledge in the meantime is enormous while some fundamental questions like that for the origin of life are still not settled.

Physics and Biology, as representative natural sciences in the rather high-level view taken here, are covering all of the natural phenomena – with one fundamental exception. This then should be the domain of a third representative discipline which as yet has not been named by any comprehensive term. This exception comprises phenomena like signal processing and evaluation, attitudes, intentions, sensing, action control, communication, consciousness, emotion, thinking, goal setting, psyche, mind(s), intellect(s), spirit(s), and so forth. Let us summarily refer to these phenomena as to informational, psychological and intellectual stuff, or to *IPsI-stuff* for short. A scientific study of IPsI-stuff, including the phenomena related to it, has started during the last 100 years under various aspects and in various fields including Neuroscience, (Cogni-

tive) Psychology, Cognitive Science (CogSci), Computer Science (CS) resp. Informatics, and Artificial Intelligence (AI).

2.2 Two Related Levels of Inquiry

IPsI-stuff is elusive and lacks tangibility for which reason it has proved so hard, and took so long, to be studied in a precise scientific way. How, for instance, is it possible that a bee during its life-time fulfils its duty of collecting nectar and transporting it back to its beehive? Or, how can a mathematician find out a proof for a conjectured theorem? Questions like those for thousands of years, if posed at all, found only speculative rather than scientific answers and these questions are definitely outside of the contours of Physics or Biology.

This situation has changed radically through discoveries like universal computability by Turing and others, which in the last century led to the development of the modern universal computer. Only within such a new mind frame became it possible to start research about IPsI-stuff in a truly scientific way (rather than the way philosophers pondered it throughout millennia), namely by validating theories through experiments.

We may distinguish two different levels at which such research can be carried out and acknowledge that this amounts to a very coarse-grained simplification just for the purpose of this article. At the lower level one may study the physical, chemical, or physiological processes that can be observed in natural systems featuring IPsI-stuff such as the nervous system in animals or humans. Physiology and Neuroscience are among the fields in which this kind of research has been performed in an observational and experimental way resulting in astonishing insights.

Unfortunately, the insights gained at the lower level just outlined, with regard to the very nature of IPsI-stuff, remains very limited. Neither the bee-question nor the mathematician-question above could ever be answered convincingly on the basis of just the observations and experiments at this level. Therefore, we need additional insights from the higher level characterized by the functional meaning (or semantics) of the processes occurring at the lower level and associate the two correspondingly. Psychology is studying this higher level featuring functional behaviors of, say, mathematicians and, if taken broadly, also of bees, although thereby ignoring the correspondence with what is going alongside at the lower level.

Enters the field of Artificial Intelligence (AI). Its pioneers recognized right away that information processing and computation is at work in IPsI-stuff throughout and that universal computers thus could lead us to its deeper understanding in the following way. The processes in a computer during the execution of a program could be set in analogy with, for instance, the processes in the nervous system. This analogy has suggested that the latter processes might be regarded as the executions of higher-level “programs” of a kind. So, one could write computer programs whose executions simulated the functionality of the natural “programs”; let us refer to this as a *higher-level simulation*.

Given the enormous power of modern computers the pertinent processes observed at the lower level may as well be simulated computationally by modelling them in fine-grained detail rather than in a higher-level, functional way; let us refer

to this as a *lower-level simulation*. Extracting the functionality from such a lower-level simulation generates a higher-level one whose execution processes, in turn, can be seen as a lower-level simulation, thus establishing a revealing correspondence between the two levels.

These kinds of simulations and this correspondence open all kinds of possibilities for experimentation and thus for scientific exploration of IPsi-stuff, of its nature and role in the world. This fundamental idea was initiated by the AI-pioneers for the first time in the history of science and has stayed undisputed during the last six decades, leading to fascinating insights in the meantime.

3 Emergence of a Scientific Discipline

In the previous section we have identified three major steps in the evolution of science as well as IPsi-stuff as the subject of study in the third one. This third step has a long prehistory described, for instance, in [Nilsson, 2009] (with a few additions in [Bibel, 2014]). It is generally agreed that the emergence of the science studying IPsi-stuff started in earnest only with the Dartmouth Conference in 1956 [McCarthy *et al.*, 1955]. It is the purpose of the present section to give a very coarse-grained outline of the historical development of this science in the last sixty years.

Let us first note that the initiators of this legendary conference did not intend to propose a name for the new discipline they had in mind. For instance, John McCarthy, the driving force in realizing the conference, for some time thought about naming it “Cognology” (see [Bibel, 1992] for more details). But the rapidly growing community then simply picked the “Artificial Intelligence” in the title of the conference, there naming the object of study, as a convenient term for also naming the discipline as such without further deliberations which then stayed stuck to this discipline. Later alternative proposals like the “Intellectics” [Bibel, 1992] of the present author could no more displace the term firmly established by then. Since in the meantime AI has changed its meaning and contours considerably and has ceded much of the terrain characterized by IPsi-stuff to other disciplines, I will nevertheless, in lack of another name, use here Intellectics (or IPsi-science, if you prefer) as the name for the discipline studying IPsi-stuff in the originally broad sense.

It should also be noted that the initiators explicitly envisaged both aspects outlined in terms of the higher and lower levels in the previous section. For instance, they talked of “*every aspect of learning or any other feature of intelligence*” which corresponds to the higher level, but indicated also the lower level and its correspondence to the higher one with passages like “*How can a set of (hypothetical) neurons be arranged so as to form concepts*” [McCarthy *et al.*, 1955]. AI has become so successful exactly because it has been characterized as a science in the right way from the outset.

Strategically, the term AI turned out to be an excellent choice, at least under the following aspects. It immediately gave some idea about the topic to everybody, helped to focus the efforts of all those involved and kept raising public interest in the discipline. This along with initial successes of the early systems eventually led to the first hype about AI in the

eighties of the last century, eventually resulting in disappointment for the most part.

AI like all sciences has a strong potential for technological applications. It is these that raise the interest of the general public and make politicians spend money for AI research. But in any science fundamental research and applications both need to be supported in a balanced way. During the first AI hype this balance was badly neglected with the sad consequence that AI narrowed its research focus mainly to applications and, in consequence, to a rather limited extent of IPsi-stuff. For frustration, AI researchers with a focus on fundamental research about cognitive phenomena (rather than on quickly achieved applicational results) founded Cognitive Science (CogSci) as a separate discipline, formerly being part of the encompassing AI. Both, AI and CogSci study IPsi-stuff in the same methodological way; hence the separation is not justified by scientific arguments.

Independently, other historically grown disciplines such as Biology, Medicine, Psychology and Philosophy, observing AI’s progresses, began to recognize the power of the computational paradigm excelling within AI and founded new subdisciplines such as Neuroscience, Cognitive Psychology, Cognitive Philosophy and Neuroinformatics, to mention just four of them. In terms of IPsi-stuff their independent existence can hardly be justified. As a consequence, the broad scientific discipline IPsi-science concerned with IPsi-stuff, formally represented exclusively by AI, has been splintered among more than half a dozen of different disciplines. While the interest of so many disciplines in what originated from AI is encouraging, the splintering of IPsi-science into separate disciplines causes many disadvantageous effects, two of which are the following ones.

An obvious disadvantage is the multiplication of independent scientific efforts striving for the same solutions which results in an enormous redundancy and hinders synergy. Avoiding this redundancy could otherwise lead to a more focussed progress. Another disadvantage arises from the lack, or ignorance, of pertinent know-how required in one of the different disciplines while often readily available in a neighbouring one. Note thereby that the contours of those splintered disciplines have not been formed by structural IPsi-stuff features in a rational way but by the existence of – from the viewpoint of IPsi-stuff – accidental, historically grown communities. The situation is thus very different from the subjects- and methods-oriented and thus rational division of other disciplines like the well-justified separation of Chemistry from Physics. Let us illustrate this particular disadvantage with a rather telling example.

The huge 1 billion euros Human Brain Project (HBP)¹ aims at replicating the brain and its workings on a computer. Thus the goals of HBP are exactly what in the previous section we called lower-level simulations. In the twelve subprojects of this European flagship project AI is not even mentioned nor are leading AI researchers involved although AI techniques such as learning are a crucial part of it. Due to this lack of AI

¹See the websites <https://www.humanbrainproject.eu/>, https://cordis.europa.eu/project/rcn/99188_en.html, and https://cordis.europa.eu/project/rcn/109805_en.html.

participation the AI techniques available for establishing the correspondence with the higher-level simulation as described in Sect. 2.2 cannot at all be taken into account in a state-of-the-art way, thus limiting the progress from the outset.

This absence of AI from the HBP reflects the prevailing public view of AI as a subject quite different from the goals of the HBP, a truly unfortunate misunderstanding, which demonstrates that the lack of an appropriately named and defined discipline in the sense of IPsi-science leads to rather negative consequences, not only for AI but also for science as a whole. The challenge of a deep understanding of IPsi-stuff will not be achieved unless research focusses on its subject at both levels described in the previous section in a coherent and disciplined effort, a point we will resume in view of the HBP in the next section.

Despite this somewhat less than optimal situation, AI, after two decades sometimes tabbed as the “AI winter”, is currently back in its second and much more substantially founded hype. While this is good for our field, the hype again carries with it problems to be concerned about. In particular, AI’s current prominence in the public attention again focusses on rather narrow aspects of technology originating from our field, especially on mining big data, learning, neural nets and to a certain degree robotics. In the public view of today, AI often is identified with just these techniques.² In consequence, fundamental and basic Intellectics research is not pursued as broadly and uniformly as would be necessary for substantial progress in the big issues of our discipline.

4 Challenges for Future Research

In all parts of the world AI is in the headlines these days as mentioned in the last section. Governments as well as companies are investing huge amounts of money in order to win the ongoing technological race. In some cases this might open the possibilities for AI researchers to influence the directions of research and leverage their current prominence in the public attention to re-engage in the broader mission of our discipline. In other words, the AI community, in addition to and beyond striking successes in applications, should seize the moment to demonstrate our insights on the fundamental IPsi-science (or Intellectics) as well as on broader societal questions around AI phenomena.

It is the purpose of the present section to outline five selected exemplary and challenging research directions which seem desirable and promising in view of such a more encompassing and ambitious conception of AI as a scientific discipline and to the benefit of mankind. These – along with many other – directions crucially require AI methods and could be attacked now by taking advantage of the current public focus on AI.

²Unfortunately, even outstanding colleagues involved in the research of IPsi-stuff propagate such a narrow view of AI which, for instance, happened in the open letter published under <https://ellis-open-letter.eu/> on 23.04.2018, leading to numerous reactions in the press thereafter.

4.1 Functionality of Brain Processes

In Sect. 2.2 we introduced the lower and higher levels of inquiry of IPsi-stuff. In Sect. 3 we pointed to the Human Brain Project (HBP) as an initiative for simulating the working of the brain at the lower level. In fact, HBP already maintains a platform for researchers worldwide to access the obtained simulation data.

As we pointed out in Sect. 2.2 the simulation of the working of the brain during the execution of a particular task is certainly of great scientific interest, but by itself does not really provide a satisfactory explanation and a full understanding of the underlying brain process. Only if we could correlate such a particular simulation with what that process means in terms of a functional interpretation at the higher level would this lead to an exciting insight into the working of the brain. In Sect. 2.2 I spoke of extracting the functionality from the simulation in this context.

In order to understand the importance of both levels for a deeper understanding, by analogy think of the processes occurring during the execution of a program on a computer. Imagine observing those processes in all their details exclusively at the lower process level where bits are shifted around and without any information about the executed program. The resulting data would definitely not reveal an understanding of what is going on until we would have a look into the program at the higher level, or even better at the program’s declarative specification. It is possible to extract the program from the process data at the lower level and that is what we refer to as functionality extraction.

Who if not AI researchers (or computer scientists for that matter) could succeed in working out the solution to this challenging extraction problem in the case of brain research! Imagine for a moment what we would learn from such an insight if the brain’s task happened to consist in the search for a proof of a theorem (one of our examples from Sect. 2.2)! That, of course, seems like a thought whose realization still lies in the far distant future. But the analogous experiment for the bee fulfilling its daily duty – our second example – seems already realistic enough according to the present state of affairs. So I suggest here to extend the HBP by relating some of its results to the higher level through functionality extraction (which might as well be carried out independently of the framework of the HBP per se).³

What I suggest here was similarly envisioned by AI’s pioneers when they asked the research question cited already above: “*How can a set of (hypothetical) neurons be arranged so as to form concepts*” [McCarthy *et al.*, 1955]; what they had in mind is obviously similar to what I term functionality extraction. Unless AI researchers take up on this sixty years old challenge now, HBP researchers from other disciplines might solve it eventually (although taking much longer than with AI involvement) and this way further the progressive shrinking of the contour of original AI. Including AI into the

³One referee commented this suggestion with the following statement: “*My sense is that if some group of AI scientists could (for example) show how to extract programs from brain simulations, that would be appreciated as a breakthrough and central to initiatives like the Human Brain Project.*”

bunch of disciplines forming a commonly understood IPsI-science would be a more effective solution, of course.

4.2 Hybrid Two-Level Techniques

One of the reasons for the recent prominence of AI is the spectacular success of deep learning combined with reinforcement learning techniques in systems like Alpha-Zero [Silver *et al.*, 2017]. The fact that this system succeeds with tabula rasa reinforcement learning from self-play games within 700.000 steps has triggered speculations that such a tabula rasa approach via deep learning might succeed similarly in many other domains. In terms of the levels discussed in Sect. 2.2 this would mean that intelligent behavior could be generated by simulation just at the lower level (thus leaving the higher level obsolete).

While this approach may indeed successfully work for many more domains beyond game playing, it seems very unlikely that it also applies to domains characterized by the involvement of structured knowledge. Take for instance mathematical theorem proving. Given that mathematicians take years for acquiring specialized knowledge until they become creative in theorem discovering and proving, it seems that the structured knowledge involved does play an essential role. In other words, in domains like theorem proving, or generally in reasoning about represented knowledge for that matter, the higher level seems to be indispensable. Due to this general feeling this kind of AI research up to this point has been done mostly at the higher level. It remains important not only for a deeper understanding of IPsI-phenomena, as illustrated with the HBP in the last subsection, but also for success of AI in numerous relevant domains.

Nevertheless, I strongly believe that a hybrid approach of the following kind would promise enhanced success in such knowledge-based domains. In Sect. 4 of [Bibel, 2017] I have outlined a way how to engage deep learning techniques in the search for (connection) proofs. Informally, that part of the work of a mathematician which consists of the structured proof steps seems to require the higher-level techniques of Automated Theorem Proving (ATP) also in the future. But the part, where unstructured information in form of the gut feelings of humans get involved in the search for a proof, might be more suitably modelled at the lower level with techniques like deep learning. As an illustrating example, if mathematicians look out for applying one out of a variety of possible lemmas in conceiving a proof then this selection process might be supported with deep learning techniques simulating the expert's lower-level gut feeling. Such a hybrid approach seems promising in many knowledge-based areas other than ATP, opening a vast new research area. For initial work in that direction see the references in [Bibel, 2017].

In this particular challenge of combining higher-level formalisms with lower-level learning techniques the experienced difficulties in advancing this subarea are due to the lack of collaboration (or sometimes even frictions) among current AI subfields rather than between AI and other disciplines (like neuroscience as in the previous subsection). In the present example we may simply observe that the collaborations between experts in connectionism (the so-called subsymbolic approach) and those in reasoning (the symbolic approach) are

the rare exception rather than the rule. It is just hoped that this situation will be changing in the near future.

4.3 Societal Problems

“We humans are sentenced to spend our lives trapped in our own heads ... [and] will inevitably encounter an unbridgeable gap between self and other. ... [This] empathy gap is responsible for most interpersonal conflicts, from prosaic quibbles over who should wash the dishes to violent disputes over sacred land.” states the Psychologist Molly Crockett in [Crockett, 2015]. She goes on therein to speculate about intelligent machines that could bridge the empathy gap.

With this statement she points to one of the many problems in our societies and a truly important one at that. Its (partial) solution could improve the quality of our personal lives as well as that of the politics worldwide. But where in the world is a group of scientists who could set about its solution? In which discipline were such a group located? I fail to have an answer to these questions ready at hand although the problem obviously is one concerning IPsI-stuff.

If AI would still follow the directions envisioned by its pioneers like Alan Turing or the Dartmouth initiators the answer would clearly be: some group in AI. After the separation of AI into AI and CogSci one might expect that such a topic would now be at home in CogSci. But again I fail to know of any cognitive scientist who is aiming into this or a similar research direction. The general public would expect that such a task lies in the domain of Psychology (or of Sociology). But where is the psychologist who has acquired the technical and scientific expertise in building a challenging AI system of this kind?

There are many problems in our societies like – or related to – the empathy gap which could be alleviated or solved with AI technology. For instance, many personal or political disputes arise due to the limitedness and uncertainty of the knowledge available to the agents involved. So, each participating agent tends to jump to a different solution of the faced problem based on the individual resources in its knowledge base. An AI system might clarify the differences in the various assumptions and preferences. Already with such a rational clarification could the mutual understanding among the disputants substantially be improved if engaged suitably enough. The area of law and legislation claims the responsibility for resolving such disputes, but badly lacks the precision offered by AI technology for the same purposes [Bibel, 2005].

In lack of an all-embracing discipline like the envisaged IPsI-science (or Intellectics), the area of Law – despite the conference series “AI and Law” and the small community backing it – is generally not considered a subject for AI technology. Nor is Psychology or Sociology. In the meantime inter- (or cross-) disciplinary initiatives or projects might be an intermediate substitute (although experiences with some of those are not overly encouraging). The important message here is that such pressing problems harassing our people and societies should finally be attacked by scientific methods including AI ones. *“Particularly in the current climate, there is a receptiveness to technical ideas and proposals from AI scientists,”* as one of the paper's referees has put it.

4.4 Consciousness

Consciousness remains one of the greatest mysteries around IPsi-stuff. The literature on this subject is huge. The topic is discussed still mainly in philosophical circles, although in the past few decades it has begun to interest cognitive scientists and neuroscientists as well. I just mention two recent books by (cognitive) neuroscientists Baars and Gage [Baars and Gage, 2010] and Damasio [Damasio, 2010] which contain extensive lists of pertinent references. I also mention the Association for the Scientific Study of Consciousness (ASSC) along with its journal and conference series as well as the Society for Mind-Brain Sciences.

In terms of the levels introduced in Sect. 2.2 Damasio's and Baars&Gage's books study consciousness more or less exclusively at the lower level, similarly as the HBP discussed above. So what was generally said in Sect. 4.1 about extracting functionality applies here as well, except that there we had in mind particular tasks studied on both levels along with their correlations while here consciousness refers to a complex phenomenon of a particular kind.

At the higher level consciousness could perhaps be understood in formal terms which might more easily be explored in a first approach. In Sect. 2.5 of [Bibel, 2003] the present author outlined such an approach based on the logic of knowing [Fagin *et al.*, 1995], here to be applied to a self which knows about its knowing. With such a functional model of consciousness one might then find an access to an understanding of what is going on in the brain leading to consciousness and the conscious self.

While this is just one out of several ideas how to approach the mystery of consciousness at the higher level in a top-down manner – see eg. [Perlis, 1997] for an interesting different one –, it seems at least clear that there is great potential in engaging methods from AI in the challenging task of solving the mysteries of consciousness. Not surprisingly, nothing of that kind can be found in the work of neuroscientists like the ones mentioned above. The fundamental aspect of computation plays no noticeable role in their handling of the topic. Working in completely separated disciplines they apparently lack the expertise available in AI. Also, they do not even think of the possibility to collaborate with AI experts in work of that kind; and vice versa. Would neuroscientists and AI researchers understand their respective work because of sharing the same encompassing discipline, such collaborations would perhaps be more likely if not even self-understanding.

4.5 Morality, Ethics and Fairness

AI systems influence societies to an ever growing degree. Like a species of foreign agents they have by now mingled among human agents. This raises issues of morality, ethics and fairness for artificial agents similar to those for human agents. It also changes the traditional market structures familiar in Economics thus requiring new rules for a fair distribution of the realized profits. Governments have therefore started to work on establishing norms for autonomous systems, to begin with.

Typically, the approach taken thereby is to set up a committee, usually consisting of politicians, theologians, philosophers, etc. Rarely do they include AI experts who would be

knowledgeable enough to foresee the forthcoming possibilities of technologies. If we want to change these attitudes, AI within its own community would first have to take more serious these kinds of issues. We should carry out research of this nature with the goals to elaborate ethical rules and to equip autonomous AI systems with ethical guidelines for their behavior. We also should help the public to understand the dangers which might be involved in the use of AI systems and thus should rather be eliminated as far as possible right away.

Fortunately, there is already promising, if rare, work into this direction. For instance, the article [Scheutz, 2017] made a strong case for explicitly building ethical agents. And Stuart Russell [Russell, 2017] and others have started an international campaign in order to sensitize the public for the fact that AI technology like any other technology can be used in ways that might cause enormous harm to people and the societies worldwide. For instance, a worldwide ban on lethal autonomous weapons systems (LAWS) should definitely be agreed upon and enforced.

So the inclusion of these particular aspects within this section is meant to strongly support these initiatives and to argue for an increased awareness and concern of the AI community about these issues. We need to achieve a common understanding that it is highly desirable that the issues of morality, ethics and fairness receive much more attention within our field which in consequence might lead to a more human-compatible and human-centered AI with many benefits for mankind. It might then also lead to more influence of AI experts in the efforts of governments towards a competent and thus adequate regulation of AI technology.

5 Conclusions

In this paper we have envisioned a scientific discipline of a similarly fundamental importance as Physics or Biology. To have a name for talking about it, we used IPsi-science (or Intellectics suggested by the author more than three decades ago). While Physics is concerned with matter and Biology with life, IPsi-science deals with IPsi-stuff (ie. informational, psychological and intellectual stuff).

When AI was founded more than six decades ago, AI denoted the same envisioned discipline as IPsi-science does here. Today, however, AI has ceded much of that terrain to other disciplines and is mostly used as a name for a far more restricted scientific area with a particular focus on the technology evolving from it. Since the subjects and methods of IPsi-science as a discipline are independent of what the contours of current AI actually are, we now experience an unfortunate discrepancy between IPsi-science and AI. IPsi-science by now is distributed among more than half a dozen disciplines including AI, a fact demonstrated in the paper.

In order to get a better feel for the strangeness of this situation, imagine that the study of protons, neurons and electrons, an undisputed terrain of Physics, would all of a sudden be pursued also in Chemistry, because these particles are occurring in chemical substances, in Biology for the analogous reason, similarly in Geology and so forth. The thought seems ridiculous for good reasons. But that is exactly what currently happens with IPsi-stuff in an analogous way as it is indepen-

dently studied in a variety of disciplines.

It would, of course, be naïve to expect that the wheel could be turned back and AI's contour could again become that of IPsI-science (or of original AI for that matter). The competing disciplines would simply not allow for that. But given the disadvantages resulting from that splintering, which have been discussed in the previous sections, one might perhaps try to convince all participating disciplines to put their IPsI-science-related subareas under a combined umbrella to the benefit of a bright future of our discipline. Any success resulting thereof would increase the standing of each participating subfield.

There are many ideas of detail how such an alignment of scientific forces might be achieved. Small steps into such a direction have been made in the past such as inviting speakers from different disciplines at major conferences, eg. a cognitive psychologist speaking at an AI conference. The initiators of this Special Track on the Evolution of the Contours of AI within IJCAI-ECAI-18 have done a laudable step in this sense. But more courageous steps would be needed, like, for instance, an all-embracing international conference, pertinent courses of studies at universities, etc. In any case, AI could leverage its current prominence in the public attention to re-engage in the broader mission in some way or another.

It is impossible to enforce such a development. So the only thing an author can do is to make people in all those disciplines more aware of the existence of this particular problem and of the disadvantages for science resulting from it and argue for an even more successful course. The present paper was meant to contribute to exactly that purpose.

Acknowledgments

I thank Ulrich Furbach for helpful comments on a first draft of the paper and two anonymous referees for very instructive suggestions towards improving the paper.

References

- [Baars and Gage, 2010] Bernard J. Baars and Nicole M. Gage. *Cognition, Brain, and Consciousness – Introduction to Cognitive Neuroscience*. Elsevier, Burlington, 2nd edition, 2010.
- [Bibel, 1992] Wolfgang Bibel. Intellectics. In S. C. Shapiro, editor, *Encyclopedia of Artificial Intelligence*, pages 705–706. John Wiley, New York, 1992.
- [Bibel, 2003] Wolfgang Bibel. *Lehren vom Leben – Essays über Mensch und Gesellschaft*. Deutscher Universitäts-Verlag, Wiesbaden, 2003.
- [Bibel, 2005] Wolfgang Bibel. AI and the conquest of complexity in law. *Artificial Intelligence and Law Journal*, 12:159–180, 2005.
- [Bibel, 2008] Wolfgang Bibel. Über ein Fach ohne Namen und Struktur. *KI Künstliche Intelligenz*, (2):66–67, 2008.
- [Bibel, 2014] Wolfgang Bibel. Artificial Intelligence in a historical perspective. *AI Communications*, 27(1):87–102, 2014.

- [Bibel, 2017] Wolfgang Bibel. A vision for Automated Deduction rooted in the connection method. In R. Schmidt and C. Nalon, editors, *The 26th International Conference on Automated Reasoning with Analytic Tableaux and Related Methods (TABLEAUX 2017)*, volume 10501 of *LNAI*, pages 3–21. Springer, 2017.
- [Crockett, 2015] Molly Crockett. Could thinking machines bridge the empathy gap? <https://www.edge.org/annual-question/what-do-you-think-about-machines-that-think>, 2015.
- [Damasio, 2010] Antonio R. Damasio. *Self Comes to Mind: Constructing the Conscious Brain*. Pantheon Books, New York NY, 2010.
- [Fagin et al., 1995] Ronald Fagin, Joseph Y. Halpern, Yoram Moses, and Moshe Y. Yardi. *Reasoning About Knowledge*. MIT Press, Cambridge MA, 1995.
- [McCarthy et al., 1955] John McCarthy, Marvin L. Minsky, Nathaniel Rochester, and Claude E. Shannon. A proposal for the Dartmouth summer research project on artificial intelligence. August 31 1955.
- [Nilsson, 2009] Nils J. Nilsson. *The Quest for Artificial Intelligence: A History of Ideas and Achievements*. Cambridge University Press, 2009.
- [Perlis, 1997] Donald Perlis. Consciousness as self-function. *Journal of Consciousness Studies*, 4:509–525, 1997.
- [Russell, 2017] Stuart Russell. Provably beneficial Artificial Intelligence. The Next Step: Exponential Life, BBVA OpenMind, 2017.
- [Scheutz, 2017] Matthias Scheutz. The case for explicit ethical agents. *Almagazine*, 38(4):57–64, 2017.
- [Silver et al., 2017] David Silver, Thomas Hubert, and Julian Schrittwieser et al. Mastering chess and shogi by self-play with a general reinforcement learning algorithm. *arXiv:1712.01815 [cs.AI]*, pages 1–19, 2017.