Human-Level Artificial Intelligence Must Be a Science

Tarek R. Besold

Institute of Cognitive Science, University of Osnabrück, tbesold@uos.de

Abstract. Human-level artificial intelligence (HAI) surely is a special research endeavor in more than one way: The very nature of intelligence is in the first place not entirely clear, there are no criteria commonly agreed upon necessary or sufficient for the ascription of intelligence other than similarity to human performance, there is a lack of clarity concerning how to properly investigate artificial intelligence and how to proceed after the very first steps of implementing an artificially intelligent system, etc. These and similar observations have led some researchers to claim that HAI might not be a science in the normal sense and would require a different approach. Taking a recently published paper by Cassimatis as starting point, I oppose this view, giving arguments why HAI should (and even has to) conform to normal scientific standards and methods, using the approach of psychometric artificial intelligence as one of the main foundations of my position.

The dream of successfully (re-)creating intelligence has fascinated men for centuries. For instance legendary creatures from the Middle Ages like the golem or the homunculus, van Kempelen's 18th century Mechanical Turk, or the robots described in science fiction stories from the early 20th century to different degrees bear witness of the never-ceasing attraction of this idea. In the second half of the 20th century, arguably starting out with Turing's well-known paper "*Computing Machinery and Intelligence*" [1], this dream also was given its proper manifestation in science: Research in what today has become an established field of scientific enquiry known as Artificial Intelligence (AI) started. Still, from a certain perspective AI seems to stand out between the modern sciences for more than one reason: Neither is there agreement upon what shall be AI's overall objective (i.e., whether the purpose of AI is the implementation of technical systems supporting humans in their everyday tasks and facilitating human intellectual activity, or if the purpose of AI is the creation of a computer system exhibiting general intelligence — in doing so possibly outperforming humans in tasks requiring reasoning and thought —, or something in between these two extremes), nor is there a commonly accepted methodology for conducting research in AI, nor is there consensus concerning the valuation of previous developments and the actual status quo in AI as a story of success or perpetual failure.

These and related observations repeatedly caused philosophers of science and even some researchers from within AI to wonder about AI being a special type of science or to even question (and occasionally finally deny) the status of AI as a science. In what follows, I want to share my view and arguments concerning these discussions specifically focussing on the subbranch of AI dealing with Artificial General Intelligence (AGI) or Human-Level AI (HAI), i.e., the endeavor to create computer systems that

exhibit intelligence at a level similar to humans.[1] Recently, Cassimatis — himself a researcher in AI with the set goal of creating machines with human-level intelligence — in [2] advocated that, when dealing with HAI research, normal scientific standards and methods are often incidental and even antithetical to achieving human-level intelligence in an artificial system and that a different approach would thus be required. I oppose this view, giving an account of Cassimatis' position and criticism (as both, paradigmatic example and synecdoche for a more general line of argumentation prevailing in quite some academic discourse) and contrasting them with arguments showing why AI can be, should be, and even has to be a science and adhere to scientific principles.[2]

Cassimatis' argument for why HAI is not a normal science starts out with proclaiming a significant qualitative difference in the specificity and the level of ambition of the respective goals between research in HAI and other sciences. From his point of view, the objectives of HAI are "*more concrete and much more ambitious*" [2] than the counterparts from other fields. Also, he claims that (H)AI historically had not been conducted as a normal science, using Winograd's SHRDLU as representative of the systems of these days which allegedly did not witness experimental evaluation or formal proofs that are obligatory part of science or engineering research reports. As the evaluation of the first claim to me seems to be a question of personal judgement only — try to decide which field has the more ambitious objectives, the one trying to (re-)create intelligence at a human level (i.e., HAI) or the one aiming at obtaining "*knowledge which relates to the order of nature, or (...) to the regular succession of events*" [3] (i.e., physics) —, the latter claim should at least be considered debatable: Already in his original dissertation [4], Winograd himself compared his SHRDLU system with other parsers and programs (even offering rudimentary quantitative performance comparissons). As additional evidence to the contrary e.g. consider the numerous introspection reports that Newell and Simon collected and methodologically analyzed as basis for the development of their General Problem Solver [5], in doing so implementing systematic observation, data collection, analysis and subsequent model building as classical methodological pattern from the natural sciences. Therefore, different from Cassimatis, I do not see a sign of "science envy" in the observation that (almost) all of AI in its approach by now has adopted the standards of science, but gleefully approve the diagnosis from [6] that "*[i]n terms of methodology, AI has finally come firmly under the scientific method*".

But the line of reasoning given in [2] does not stop with stating that HAI is not a science in the normal sense, but continues with a from my point of view much stronger and (if correct) even more serious claim. In Cassimatis' eyes, the application of scientific norms and methodological standards is not favorable for achieving HAI and often even stands against swift progress towards this goal. He gives several arguments for why he thinks that normal scientific methods fail HAI, drawing on comparisons with

[1] Although I am well aware that this entails a particular interpretation of the term Artificial General Intelligence, in this paper I will treat AGI and HAI as synonymous.

[2] Cassimatis in [2] also gives a set of principles which he deems meaningful as basis for an approach to HAI. In this note, I will only deal with his inquiry into the scientific status of AI, for the time being leaving the second part of his considerations aside.

other scientific fields or general practices which he considers as important sources of (potentially harmful) influence on HAI. In what follows, I sketch what I consider the four main issues and rebut them.

Cassimatis' first and second main quarrels lie with formal or empirical demonstrations of correctness or optimality, and the connected computational requirements in terms of processing power and speed. In his eyes, the believe that importance should be assigned to showing the formal correctness of a theorem, or to empirically demonstrating a method's optimality with respect to a certain normative standard, goes against the nature of human intelligence. By making reference to Simon's notion of bounded rationality [7], Cassimatis tries to demonstrate that human rationality falls far from optimality or formal correctness in all but a few cases, and thus also HAI should not use normative notions of correctness or optimality in judging and evaluating results. Still, I do think that Cassimatis here has an overly narrow view and that the problem he tries to point at does not reside with optimality considerations or formal proofs itself, but rather with the chosen normative standards. It goes without doubt that human rationality is not optimal in any of the classical senses (i.e., with respect to the standard of classical logic, Bayesian probability theory, game theory, etc.). Also, it should be clear that human cognition most likely does not solve difficulties arising from a problem's exorbitant computational complexity or intractability (i.e., to the best of my knowledge, there is no convincing evidence suggesting that humans can be understood as super-Turing computers). But this does not mean that no quantitative assessment is possible per se. I am aware of two possible remedies for the diagnosed problem: There is an ongoing interdisciplinary discourse trying to define new frameworks of rationality that better encompass actual human performance than their predecessors did (cf., e.g., [8]). But more importantly, going along a similar path, there are approaches within AI itself that successfully apply quantitative measures to problems in HAI research, namely research going by the name of Psychometric Artificial Intelligence (PAI) [9, 10]. PAI applies the full battery of techniques from psychometrics to an HAI context, setting its internal standard by declaring that "*some agent is intelligent if and only if it excels at all established, validated tests of intelligence*" [10]. This clearly makes PAI a very quantitatively focused field of research with clear normative principles, but still avoids the pitfalls Cassimatis (who himself has close ties to PAI-style research endeavors) meant to diagnose. The progress of an HAI system towards achieving the overall goal of (re-)creating human intelligence is measured against actual human performance on what commonly is agreed as standard means of assessing the relevant human mental capacities. By doing so, optimality is not demanded anymore with respect to a hypothetical idealized standard but with respect to achievable and reproducible testing scores of human subjects.

The third and the fourth main point Cassimatis raises relate to the fields of empirical (cognitive) psychology and neuroscience. Both fields adhere to normal scientific standards in their methodology, relying on experimental studies of (mostly) isolated individual capacities and functions, according to [2] by this inevitably not being aligned with the standards needed for HAI. Firstly, the phenomena studied very often may not be directly crucial for progress towards the goal of solving the overall intelligence puzzle. Secondly, the approach of averaging over (a possibly high number of) subjects in

an experimental evaluation abstracts away from the individual, thus yielding results which provide general, decontextualized average predictions of the behavior of a cognitive system instead of individual-specific, context-sensitive ones — thus seemingly also not contributing to solving the intelligence puzzle on an individual level. Whilst these observations by themselves seem correct to me, I do not agree on Cassimatis' assessment and interpretation. Clearly, whilst general cognitive psychology is significantly different from differential psychology in its methods, results and ambitions, the insights gained in cognitive capacities on a general level are as valid to HAI as are insights on an individual level. This is not only the case because each average trivially also can be interpreted as one possible case within a population (namely the one coinciding with the average), but also can an average prediction be bootstrapped into an individual prediction by contextualizing it with the respective initial conditions and accompanying factors and context. And staying under the umbrella established by scientific standards does bring along important advantages: Quantitative and comparative methods allow for measuring and judging progress (e.g., in the case of PAI , against human performance, which seems to be reasonable in the case of HAI), and in doing so also enable us to make goals and milestones specific (thus alleviating yet another problem within the HAI context).

I strongly advocate the necessity of viewing HAI as a normal science — thus also demanding that research in HAI has to be conducted within the framing constraints of "normal" scientific standards —, relying on quantitative and comparative experimental and assessment methods whenever possible, and trying to adhere to overall principles and laws underlying cognition and intelligence that have been identified within neighboring fields. The lesson is clear: Human-Level Artificial Intelligence must be a science.

References

1. Turing, A.: Computing Machinery and Intelligence. Mind **LIX**(236) (October 1950) 433–460
2. Cassimatis, N.I.: Human-Level Artificial Intelligence Must Be an Extraordinary Science. Advances in Cognitive Systems **1** (2012) 37–45
3. Maxwell, J.C.: Matter and Motion. Courier Dover Publications (1920)
4. Winograd, T.: MIT AI Technical Report 235: Procedures as a Representation for Data in a Computer Program for Understanding Natural Language. (1971)
5. Newell, A., Simon, H.: Human problem solving. Prentice-Hall (1972)
6. Russell, S.J., Norvig, P.: Artificial Intelligence: A Modern Approach. Prentice Hall (2009)
7. Simon, H.: Models of man: Social and rational. Wiley (1957)
8. Besold, T.R., Kühnberger, K.U.: E Pluribus Multa In Unum: The Rationality Multiverse. In Miyake, N., Peebles, D., Cooper, R.P., eds.: Proc. of the 34th Annual Meeting of the Cognitive Science Society, Austin, TX, Cognitive Science Society (2012)
9. Bringsjord, S., Schimanski, B.: What is Artificial Intelligence? Psychometric AI as an Answer. In: Proc. of the 18th International Joint Conference on Artificial Intelligence (IJCAI'03), Morgan Kaufmann (2003)
10. Bringsjord, S.: Psychometric artificial intelligence. Journal of Experimental & Theoretical Artificial Intelligence **23**(3) (2011) 271–277