




The Future of AI: Neat or Scruffy?

Bernardo Gonçalves^(✉)  and Fabio Gagliardi Cozman 

Escola Politécnica, Universidade de São Paulo, São Paulo, Brazil
{begoncalves,fgcozman}@usp.br

Abstract. The “neat” and “scruffy” portraits have long been painted to describe viewpoints, styles of reasoning and methodologies in AI research. Essentially, the neats defend techniques based on first principles and grounded in mathematical rigor, while the scruffies advocate diversity within cognitive architectures, sometimes meant to be models of parts of the brain, sometimes just kludges or ad-hoc pieces of engineered code. The recent success of deep learning has revived the debate between these two approaches to AI; in this context, some natural questions arise. How can we characterize, and how can we classify, these positions given the history of AI? More importantly, what is the relevance of these positions for the future of AI? How should AI research be pursued from now on, neatly or scruffily? These are the questions we address in this paper, resorting to historical analysis and to recent research trends to articulate possible ways to allocate energy so as to take the field to maximal fruition.

Keywords: Neat vs. Scruffy · History of AI · Future of AI · Scientific method · Styles of scientific reasoning

1 Introduction

Which architecture should be implemented in a machine to make it best reproduce human intelligence in relevant intellectual tasks? In the history of AI, two answers have often been given to this question by different researchers, who have been labeled either as “neats” or “scruffies.” Capturing viewpoints, styles of reasoning and methodologies, these labels have often been used as caricatures of the AI researcher—the analytical, sensible, dry *vs.* the empirical, messy, creative. But is this a fair account of what is at stake?

The debate about which methodologies are best for AI and what AI actually is or should be is an old one. But the advent of deep learning brought it back to the front stage of the discussion about the present and the future of AI. With considerable success in image, speech and natural language processing, AI has increased its social impact. Arguably the question about which architecture is most promising for AI underwent a twist, and it now comes with an attached social concern about the future.

Which future is best for AI? Is it one developed around a (super)intelligence that emerges through learning and that makes its judgements with little oversight? Or is it one based on a useful yet more predictable AI that is managed

more closely? Possible answers to this question are strongly tied to different research directions that have been proposed by the neats, on one side, and by the scruffies, on the other side. Recently hybrid or the so-called neurosymbolic AI has also been proposed as an attempt to leverage on the strengths of both sides. This approach to AI needs to be tested by time and deserves further study. In any case, it shall also benefit from a reflection upon its neat *vs.* scruffy origins.

In this paper we contribute to the discussion with:

1. an in-depth examination of the past through a review of notable neat and scruffy positions in the history of AI up to the present;
2. the identification of three types of attitudes towards AI research from neats and scruffies;
3. the formulation of their implications for the future of AI from the point of view of both research and its social impact.

In Sects. 2, 3 we shall study positions and uses of the “neat” and the “scruffy” terminology through the history of AI in order to arrive in Sect. 4 at three types of attitude towards AI research. Our contributed scheme is aimed at clarifying which notions of the terms have been available. In Sect. 5 we study implications of the three approaches for the future of AI. We conclude the paper in Sect. 6.

2 The “neats” Vs. “scruffies” Debate in the History of AI

We start with a review that partly extends a recent survey paper [3]. Our exposition is not chronological.

2.1 Marvin Minsky (1985–1995)

Minsky did not explicitly refer to the “neat” vs. “scruffy” dichotomy yet he posited an influential position in the discussion. In his 1985 book *The Society of Mind* and later, Minsky appealed to the complexity of the human mind and/or brain to justify that his explanations “rarely go in neat, straight lines from start to end.” He associated it with the very nature of the mind:

Perhaps the fault is actually mine, for failing to find a tidy base of neatly ordered principles. But I’m inclined to lay the blame upon the nature of the mind: much of its power seems to stem from just the messy ways its agents cross-connect. If so, that complication can’t be helped; it’s only what we must expect from evolution’s countless tricks. [18, p. 18]

A few years later Minsky developed the point in connection with the brain:

The brain’s functions simply aren’t based on any small set of principles. Instead, they’re based on hundreds or perhaps even thousands of them. In other words, I’m saying that each part of the brain is what engineers call a kludge — that is, a jury-rigged solution to a problem, accomplished

by adding bits of machinery wherever needed, without any general, overall plan: the result is that the human mind — which is what the brain does — should be regarded as a collection of kludges. The evidence for this is perfectly clear: If you look at the index of any large textbook of neuroscience, you'll see that a human brain has many hundreds of parts — that is, subcomputers — that do different things. Why do our brains need so many parts? Surely, if our minds were based on only a few basic principles, we wouldn't need so much complexity [19].

So according to Minsky, one is led to think, AI systems shall be untidy like the brain. But did Minsky suggest that AI as a discipline shall be messy as well? This question will be revisited in a discussion with Yann Lecun in Subsect. 2.6.

2.2 Nils Nilsson (2009)

Nilsson addressed Minsky's position that the brain is a kludge so an AI system should be likewise. In his 2009 book *The Quest for Artificial Intelligence*, Nilsson acknowledged the wide disagreement in the field about what AI research should be like. He observed:

Of course, just because the brain is a kludge does not mean that computer intelligences have to be. Nevertheless, some AI researchers favored systems consisting of collections of experimentally derived, ad hoc routines designed to solve specific problems. These people called themselves “scruffies” to distinguish themselves from the “neats” who favored programs based on theoretically based principles. (These terms were apparently first used by Roger Schank in the 1970s to contrast his approach to building natural language processing systems with the more theoretically based work of McCarthy and others.) In his keynote address at the 1981 annual meeting of the Cognitive Science Society, Robert Abelson compared the two camps by saying “The primary concern of the neat is that things should be orderly and predictable while the scruffy seeks the rough-and-tumble of life as it comes [...]” [20, p. 417]

Nilsson thus resumed to put his own opinion over the issue:

I believe that both neats and scruffies are needed in a field as immature as AI is. Scruffies are better at exploring frontiers outside the boundaries of well-established theory. Neats help codify newly gained knowledge so that it can be taught, written about, and thus remembered. (*Ibid.*)

So it seems that Nilsson supported the coexistence of both approaches in AI research, each one in its own camp. He suggested that this is particularly important while AI is still a young endeavor.

2.3 Herbert Simon (1972)

Simon had received in 1978 the Nobel Prize in Economics for identifying the limited and messy nature of decision making as opposed, say, to idealized metrics of utility. Pamela McCorduck interviewed Simon extensively. She reported:

Simon also took up the problem that had divided the AI (and cognitive science) community from the beginning: Is thinking best viewed as a process of reasoning from premises (the Neats?) or as a process of selective search through a maze (the Scruffies)? He had no final answer to that when he died in 2001, though we can assume he leaned toward the Scruffy point of view, given his book, *Human Problem Solving*, that he'd published with Allen Newell in 1972. [15, p. 452-3]

In his theory of human problem solving with Allen Newell, Simon emphasized the role of heuristics. He thus described their famous Logic Theorist (LT) program:

There were important differences between LT's processes and those used by human subjects to solve similar problems. Nevertheless, in one fundamental respect that has guided all the simulations that have followed LT, the program did indeed capture the central process in human problem solving: LT used heuristic methods to carry out highly selective searches, hence to cut down enormous problem spaces to sizes that a slow, serial processor could handle. Selectivity of search, not speed, was taken as the key organizing principle [...]. Heuristic methods that make this selectivity possible have turned out to be the central magic in all human problem solving that has been studied to date. [29, p. 147]

From a mathematical point of view, heuristics may be considered ad-hoc techniques as it is hard to assign to them accurate theoretical guarantees. This is probably related to what McCorduck meant when classing Simon (and his LT co-designed with Newell) as an example of scruffiness in AI. Now, if the event in the northern summer of 1956 in Dartmouth marks the birth of AI as a discipline, then the scruffy LT can be considered the first AI program.

2.4 John McCarthy (1958)

McCarthy was the primary organizer of the Dartmouth workshop [15]. But it was Newell and Simon who seem to have taken most of the attention at Dartmouth, as they were the ones that had something real to show—the LT. McCarthy's reaction would come soon. In 1958, he compared the LT with his own program:

The *advice taker* is a proposed program for solving problems by manipulating sentences in formal languages. The main difference between it and other programs or proposed programs for manipulating formal languages (the *Logic Theory Machine* of Newell, Simon and Shaw and the Geometry Program of Gelernter) is that in the previous programs the formal system

was the subject matter but the heuristics were all embodied in the program. In this program the procedures will be described as much as possible in the language itself and, in particular, the heuristics are all so described. [14, no emphasis added]

McCarthy’s reservations with respect to Newell and Simon’s LT had to do with the organization of the knowledge manipulated by the program. For McCarthy, this knowledge—about the world and about the problems the program is expected to solve—should be held transparent in an expressive language. So the language should be “most likely a part of the predicate calculus” (Ibid.).

McCarthy’s concern with the knowledge representation language to be used in an AI program seems to have been the core of his research agenda. Further on in his 1958 text he outlined five features that, in his opinion, a system “which is to evolve intelligence of human order” should have at least. He then summarized: “We base ourselves on the idea that: *In order for a program to be capable of learning something it must first be capable of being told it*” (no emphasis added).

McCarthy had also expressed back then his hope to collaborate with Minsky:

The design of this system [so-called “the advice taker”] will be a joint project with Marvin Minsky, but Minsky is not to be held responsible for the views expressed here. [14]

Later on, as of 1996, McCarthy updated his text with a note “[t]his was wishful thinking,” for “Minsky’s approach to AI was quite different.” In fact, McCarthy is seen as perhaps the main advocate of neatness in AI. His divergence with Minsky seems strongly related with our main subject in this paper. And yet it is still hard to pinpoint what exactly their differences were. Before we conclude this review and proceed to address the issue, we shall refer to another AI textbook and to the advent of deep learning as the most recent success in AI.

2.5 Russell and Norvig (1995–2020)

In the 1995 (first) edition of their textbook [24], Russell and Norvig reported that AI research had just seen a “sea change in both [its] content and [its] methodology” (p. 25). They referred to “the field of speech recognition” as a paradigmatic example, as it went through a shift from the use of “ad-hoc and fragile architectures and approaches” to eventually find in Hidden Markov Models (HMM’s) an approach that is “based on a rigorous mathematical theory” and whose models “are generated by a process of training on a large corpus of real speech data.” This, the authors observed, allowed “speech researchers to build on several decades of mathematical results developed in other fields,” and ensured “that the performance is robust.” They further remarked:

Some have characterized this change as a victory of the *neats* — those who think that AI theories should be grounded in mathematical rigor — over the *scruffies* — those who would rather try lots of ideas, write some

programs, and then assess what seems to be working. Both approaches are important. A shift towards increased neatness implies that the field has reached a level of stability and maturity. (Whether that stability will be disrupted by a new scruffy idea is another question.) [24, p. 25, note 17, no emphasis added]

The last phrase, as appearing in this 1995 edition of the book, has been replaced in the 2020 (fourth and latest) edition by this note: “The present emphasis on deep learning may represent a resurgence of the scruffies” [25, p. 24, note 14]. It seems that, for these authors, the advent of deep learning is a result of scruffiness while that of HMM and Bayesian Networks are results of neatness. But to what extent is deep learning not based on mathematical theory?

2.6 Yann LeCun (2018)

With the success and also the hype around deep learning, there has been a criticism that machine learning techniques are pseudo-science and that their empirical results are not explainable nor reproducible. A report in *Science* magazine [9] covered the NIPS 2017 “Test-of-Time Award” keynote address by AI researchers Ali Rahimi and Benjamin Recht [23], who made the case that various aspects of machine learning algorithms are so badly understood, even by engineers, that they amount to alchemy. This has triggered a polemic with co-recipient of the 2018 Turing Award and Facebook’s VP and chief AI scientist Yann LeCun, who reacted to their talk on the Internet:

In the history of science and technology, the engineering artifacts have almost always preceded the theoretical understanding: the lens and the telescope preceded optics theory, the steam engine preceded thermodynamics, the airplane preceded flight aerodynamics, radio and data communication preceded information theory, the computer preceded computer science. Why? Because theorists will spontaneously study “simple” phenomena, and will not be enticed to study a complex one until there [is] a practical importance to it. [11]

Along the same lines, LeCun is reported by the *Science* magazine reporter to have said that “shifting too much effort away from bleeding-edge techniques toward core understanding could slow innovation and discourage AI’s real-world adoption.” He concluded that “It’s not alchemy, it’s engineering;” and added: “Engineering is messy.”

Now, is engineering messy? It seems that it is only in the eyes of the scruffies that it is. We can think of researchers and practitioners who contribute to the field with standards and specification, design theory and so on, and altogether with the tidy application of principles in engineering. In any case, LeCun suggested that deep learning is engineering. And yet there has been no question that deep learning is AI.

We shall then make a detour to inquire about the nature of AI.

3 Is AI a Science of Intelligence or a Branch of Engineering?

There is something odd going on here—for Minsky and Simon seem to have thought of AI more as a *science* of intelligence and less as a field of engineering.

Minsky had a point about the possibility of intelligence itself and the brain being messy as products of evolution by natural selection. For him, neuroscience—and AI —, as scientific disciplines, would be accordingly untidy as well. Now, is this reasonable? Let us consider an analogy. Just like the brain, the human organism as a whole is a (messy) biological product of evolution. Yet that did not stop the discovery of the molecular structure of DNA as a unifying principle of all life. So Minsky’s point may be seen rather as an *a priori* assumption.

Simon also seems to have thought that brain processes were to some extent messy. But he found in heuristics sort of a unifying principle. Given the initial success achieved by Newell and Simon’s heuristics, Minsky himself felt compelled to refer to them in his 1959 paper [16] and in his follow-up survey [17]. Moreover, Simon presented a view of the design of artificial systems as an empirical science, and strove to manage the complexity of large systems [28]. Later in 1995 he even wrote a dedicated piece to show that AI is an empirical science [27]. And yet it is unlikely that he would endorse Minsky’s view that AI should be messy. For Simon, AI is a lawful empirical science just as physics is. He wrote:

The natural laws that determine the structure and behavior of an object, natural or artificial, are its internal constraints. An artificial system, like a natural one, produces empirical phenomena that can be studied by the methods of observation and experiment common to all science. [27, p. 99]

In fact, Simon accepted no significant distinction between natural and artificial objects from the point of view of empirical studies.

But LeCun’s point looks really different. He seems to be talking about AI as an engineering discipline, say, such as aerospace engineering. And this connects to an older point about AI but also about computer science more generally. In a 1984 talk “The threats to computer science” [5], the physicist and computer programming pioneer Edsger Dijkstra set out one of his ingenious metaphors:

The Fathers of the field had been pretty confusing: John von Neumann speculated about computers and the human brain in analogies sufficiently wild to be worthy of a medieval thinker and Alan M. Turing thought about criteria to settle the question of whether Machines Can Think, a question of which we now know that it is about as relevant as the question of whether Submarines Can Swim [5].

Ten years later a variant of the same metaphor was caught by Noam Chomsky and appeared (with no source given) in the May 1994 lectures that further composed his 1995 *Mind* paper [2]. Many of these debates “over such alleged

questions as whether machines can think,” Chomsky referred (p. 9), “trace back to the classic paper by Alan Turing.” They fail to take note, he objected, that Turing himself declared to believe that the question “can machines think?” was “too meaningless to deserve discussion.” Chomsky thus concluded:

It is not a question of fact, but a matter of decision as to whether to adopt a certain metaphorical usage, as when we say (in English) that airplanes fly but comets do not [...] Similarly, submarines set sail but do not swim. There can be no sensible debate about such topics; or about machine intelligence, with the many familiar variants. [2, p. 9]

Overall, Chomsky denied Turing’s question to have a seat within the empirical sciences. Given Minsky’s position as we have quoted above, it is unlikely that he would agree with Dijkstra and Chomsky. But perhaps LeCun and others would do it, and then the discussion must go back to Turing’s 1950 vision of AI [31].

In an influential keynote lecture at IJCAI 1995 [8], Patrick Hayes and Kenneth Ford indirectly associated themselves with Chomsky’s variant of Dijkstra’s metaphor to argue against “the Turing test vision” of AI (p. 974–5). They complained that AI systems back then (e.g., expert systems) were sufficiently successful as task-specific cognitive artifacts and yet were seen as a failure because of “Turing’s ghost” (p. 976). They then urged:

[I]f we abandon the Turing Test vision, the goal naturally shifts from making artificial superhumans which can replace us, to making superhumanly intelligent artifacts which we can use to amplify and support our own cognitive abilities, just as people use hydraulic power to amplify their muscular abilities. [8, p. 974]

According to the vision laid out by Hayes and Ford, the primary concern of AI should be the engineering of intelligent systems.

Now, it is interesting to note that Patrick Hayes hardly fits in the scruffy portrait. In fact, Hayes’ intellectual project in AI had significant overlapping with McCarthy’s and can be best seen as a neat one. This shows that the discussion about the nature of AI as a discipline is present among both scruffies and neats. Among neats, however, the science *v.* engineering question seems much less pronounced. This is perhaps because neats tend to agree on the value of understanding—the logic of intelligent systems must be well-understood. In fact, lack of principles and understanding is the core of most critiques of machine learning today, e.g., Rahimi and Recht’s at NIPS 2017 (cf. Subsect. 2.6).

In 2012, Chomsky claimed that AI departed from the tradition of modern science as it gave up the understanding of cognitive phenomena and their rendering in artificial systems. A reporter mentioned AI’s recent shift from the so-called “Good Old Fashioned AI” to the use of “probabilistic and statistical models.” He was trying to get Chomsky’s opinion on what could explain that shift and whether or not it was a step in the right direction. Chomsky answered:

Chomsky: [An] approach, which I think is the right approach, is to try to see if you can understand what the fundamental principles are that deal

with the core properties, and recognize that in the actual usage, there's going to be a thousand other variables intervening — kind of like what's happening outside the window, and you'll sort of tack those on later on if you want better approximations [...That] is what science has been since Galileo, that's modern science. The approximating unanalyzed data kind is sort of a new approach, not totally, there's things like it in the past. It's basically a new approach that has been accelerated by the existence of massive memories, very rapid processing, which enables you to do things like this that you couldn't have done by hand. But I think, myself, that it is leading subjects like computational cognitive science into a direction of maybe some practical applicability...

Interviewer: "... in engineering?"

Chomsky: ... But away from understanding. [10].

Chomsky's opinion is worth quoting also because Google's director of research Peter Norvig bothered to reply it. Thus wrote Norvig:

I agree that engineering success is not the goal or the measure of science. But I observe that science and engineering develop together, and that engineering success shows that something is working right, and so is evidence (but not proof) of a scientifically successful model. Science is a combination of gathering facts and making theories; neither can progress on its own. I think Chomsky is wrong to push the needle so far towards theory over facts; in the history of science, the laborious accumulation of facts is the dominant mode, not a novelty. The science of understanding language is no different than other sciences in this respect [21].

Norvig's view seems strongly related with LeCun's.

We shall now be well positioned to characterize types of past and present attitudes towards AI in view of its future as a science and engineering discipline.

4 Types of Neat and Scruffy's Attitudes in AI

Considering neat and scruffy approaches to AI and the further heterogeneity that is found within each camp, we shall distinguish three types of attitudes. Although there is some overlapping of concerns in between them, the primary commitment in each type is different and looks clear, as we elaborate next.

4.1 Scruffy Type I: The Empirical Scientists

For the scruffies type I, AI should be above all an empirical science of intelligence and of the mind and/or the brain. In short, their primary goal is:

Goal: to understand and reproduce the phenomenon of intelligence.

This attitude was quite strong among influential figures in the early phase of the discipline. We identify it in Minsky and Simon but first, very early on in Turing himself. After World War II, Turing was recruited by the National Physical Laboratory (NPL) near London to build a digital computer that would match US-based initiatives such as the construction of the ENIAC. Turing’s first technical report *Proposed Electronic Calculator* in late 1945 is an outcome of his task of specifying a computer architecture. He was engaged in the job of building a computer system (“a machine”). But the historical sources show that, unlike some of his colleagues at the time (e.g., the physicist and computer pioneer Douglas Hartree [7]), Turing’s goal was not to enable scientific computing applications but rather to try to imitate the human brain [6, p. 237]. In fact, when he joined the NPL after the northern summer of 1945 he said that he was going to “build a brain” [6, p. 233]. It is possible to identify Turing’s shift from neatness to scruffiness—from his seminal *On computable numbers* paper in 1936 and his 1938 doctoral thesis in mathematical logics to codebreaking and empirical problem solving ever since his wartime service in 1939 on [6, p. 230-1]. This illustrates the ambition of the scruffy type I, who will resort to systems building as a means to empirically study natural intelligence and the brain.

The same inclination can be found in Minsky and Simon’s intellectual projects. Minsky opened his 1985 *Society of Mind* by positing his goal of explaining “how minds work” straightforwardly and asking “[h]ow can intelligence emerge from non-intelligence?” To answer that, he added, “we’ll show that you can build a mind from many little parts, each mindless by itself.” [18, p. 17]. As known, Minsky’s experimental approach to pursue that project was based on computer techniques and systems. This is described, for example, in his well-known 1961 survey of AI techniques [17].

Simon’s take is another example of the scruffy type I. Together with Allen Newell, he developed plenty of computer techniques and software systems in order to test his theories of intelligence. The Logic Theorist as we have seen was the first of the line, and was followed by the General Problem Solver and others.

To mention an active AI researcher that we identify with this type, Rodney Brooks has recently called upon AI to get back at taking the human brain as reference [1].

In short, among scruffies, the *empirical scientist* seems to aim at discovering whatever cognitive architecture and kludges that can be shown to emulate human intelligence. Emphasis is given on the best knowledge then available about the human brain as a product of evolution by natural selection over billions of years.

But in fact, as we have seen (Sect. 3), this is not the only kind of attitude among scruffies. We shall discuss a second position next.

4.2 Scruffy Type II: The System Builders

For this other class of scruffies, AI should not be primarily concerned with the deeper mysteries that surround human intelligence.

The *system builders* seem to aim at inventing whatever techniques to achieve or surpass human-level intelligence. If some contribution to the understanding

of human intelligence is made in the process, the better. Emphasis is given on leveraging the best computer resources then available. In short, we can write their goal as:

Goal: to build intelligent systems that are as autonomous as possible.

We identify this view in LeCun and Norvig as quoted above. Both posed their commentaries while being in charge of two of the biggest companies that invest in AI today. Their wish to reply to criticisms can also be understood as a defense of the research strategy and methodology underlying the AI systems that have been deployed in society by the companies they represent.

Another AI researcher that seems to express this view is Richard Sutton, Distinguished Research Scientist of the company DeepMind. In a 2019 web commentary [30], he wrote that “[t]he biggest lesson that can be read from 70 years of AI research is that general methods that leverage computation are ultimately the most effective.” Sutton called this “the bitter lesson.” He was implicitly referring to the recent success of deep learning and suggested that “[t]he ultimate reason for this is Moore’s law.” He completed: “researchers seek to leverage their human knowledge of the domain, but the only thing that matters in the long run is the leveraging of computation.” Now, there is no doubt that the human brain has huge computation power. Yet Sutton’s line of thought seems to depart from a concern with discovering the true nature of human intelligence and with imitating it closely. It characterizes most clearly the goal of resorting to whatever techniques are available in order to achieve performance that is compatible with the output of human intelligence.

4.3 Neats: The Computer Epistemologists

We have examined the varieties of discourse among neats and found no significant differences. This class may even include logicians and analytic philosophers, as it closely resembles the field of epistemology (or the theory of knowledge) in modern philosophy. But it also includes builders of knowledge bases (KB’s), as long as there seems to be a concern with the soundness of both the knowledge that gets into the KB and the knowledge that is derived out of it.

The *computer epistemologists* seem to aim at understanding and developing knowledge management and inference techniques to imitate the rational aspects of human-level intelligence. Emphasis is given on the soundness and effectiveness of reasoning. In short, we can write their goal as:

Goal: to understand and reproduce rational aspects of intelligence.

As mentioned, McCarthy advocated the need to address commonsense knowledge representation and management in view of human-level intelligence early

on in the history of AI. He was perhaps the main champion of the neats in AI research. But there is a lot more diversity in this class.

The related tradition of expert systems is also a notable example of the neats at work. Douglas Lenat and Ed Feigenbaum thus tried to summarize their view:

We articulate the three major findings and hypotheses of AI to date:

- (1) The Knowledge Principle: If a program is to perform a complex task well, it must know a great deal about the world in which it operates. In the absence of knowledge, all you have left is search and reasoning, and that isn't enough.
- (2) The Breadth Hypothesis: To behave intelligently in unexpected situations, an agent must be capable of falling back on increasingly general knowledge and analogizing to specific but superficially far-flung knowledge. (This is an extension of the preceding principle.)
- (3) AI as Empirical Inquiry: Premature mathematization, or focusing on toy problems, washes out details from reality that later turn out to be significant. Thus, we must test our ideas experimentally, *falsifiably*, on large problems. [12, p. 185, no emphasis added]

Considering all three “findings and hypotheses,” one may note Lenat and Feigenbaum’s concern with justified reasoning through sound knowledge in real-world scenarios. Also, they address the caricature of the neat profile which is often criticized for tackling toy problems.

Lenat’s initiative to develop the large-scale commonsense KB he called Cyc was based on an expressive knowledge representation language “involving first-order predicate calculus plus ZF set theory, meta-level assertions, contexts, and modal operators” [13, p. 38]. It did not pay much attention to uncertainty management. But several other approaches to knowledge representation and reasoning have been developed by the neats based on, say, multi-valued logics and probability theory. We identify as neats thinkers such as Isaac Levi and Henry Kyburg, who developed theories of uncertain and approximate reasoning; but also thinkers such as David Lewis who studied causality and counterfactuals. Other neats are Ronald Fagin and Joseph Halpern, who theorized on reasoning about beliefs; but also Judea Pearl and Adnan Darwiche who contributed to reasoning through graphical models and probability distributions.

Darwiche posited that the results achieved by “function-based” approaches to AI such as deep learning are closer to animal-like abilities” than to “human-level intelligence.” He pointed out that the latter requires a “model-based approach” [4]. For Pearl, true AI can only come when a machine is able to test cause-and-effect statements, which would allow it to explain events [22].

5 Implications for the Future of AI

Related to the goals of scruffy types I and II and the neats are implications for the future of AI research in society. The discussion between researchers from different categories is often heated as if there was no room for the fellow’s approach in AI. In dialogue with the scruffy type II, e.g., Ali Rahimi said at the 2017 of NIPS:

We are building systems that govern healthcare and mediate our civic dialogue. We would influence elections. I would like to live in a society whose systems are built on top of verifiable, rigorous, thorough knowledge, and not on alchemy [23].

But also, in dialogue with the neats Simon had written in 1995:

Artificial objects, including computer programs, are what they are because they were designed to be that way. This fact has led some to claim that there can be no science of artificial objects, but only an engineering technology. Those who hold the most extreme form of this view look to the discovery and proof of mathematical theorems about intelligent systems as the only genuine route to a science of AI, and denigrate the role of system building and experiment as “only engineering.” [27, p. 98-9]

Now, in acknowledgement of the values underlying each of the three attitudes let us pack their core values and implications for the future of AI in society.

- **Scruffy type I.** The empirical scientists hold the promise to contribute to an in-depth understanding of the phenomenon of intelligence, perhaps in cross-fertilization with neuroscience, cognitive psychology and the behavioral sciences, psychiatry and human development. This attitude towards AI research was present early on from the beginning with Turing, Minsky and Simon. There is hardly any ethical concern to be brought about in connection with it, and this is particularly true if we consider its software-based (abstract and non-invasive) methodology.¹ As intelligence is distinctive of the human, this approach to AI can in principle deliver an improved understanding of our own nature.
- **Scruffy type II.** The system builders are primarily committed to deliver AI systems that can learn for themselves from experience and change their environment, be it physical or virtual. This form of AI needs essentially two skills: perception or the ability to recognize things, and control or the ability to do things in its environment. So the recent success in image, speech and natural language processing, that is, in the semantic interpretation of opaque data, is for sure a step forward towards it. This attitude towards AI research is more recent and arguably flourished with the emergence of large AI projects in the big tech companies. It holds promise to deliver value by pushing the amount of automation in industry to the next level. It is key for applications such as tumor detection, face and speech recognition, language translation and self-driving cars. It can also replace workers in dangerous jobs. However, hand in hand with this great value there is increasing social concern that it shall lead to significant job losses, privacy risks and concentration of power.

¹ And yet, as mentioned, these types of attitude towards AI do have some overlapping. It is worth noting that Simon’s RAND-corporation collaboration with DARPA during the Cold War has something of the scruffy type II as well (e.g., cf. [26]).

- **Neats.** The computer epistemologists are primarily concerned with the study of reasoning and the delivery of AI systems that can truly interact with us humans in our own language. This form of AI needs essentially one skill: knowing or the ability to understand and form new ideas and judgements, perhaps even structured theories, which must be communicated and explained in a conversation. This approach to AI is key for achieving non-shallow chatbots and personal assistants, question answering and domain-specific reasoning and decision making in, say, law, science and engineering, healthcare and so on. It is also important, of course, in teaching AI. It has cross-fertilization with analytic philosophy and logics.

We hope that this description of types and their implications for the future of AI can be helpful for the AI community in its reflection about methodologies.

6 Conclusion

In the history of AI as a discipline, researchers tended to adopt either a neat or a scruffy (of whatever type) approach and seldom both. It is important to recognize that the forms of AI derived from each of them are in fact essentially different and hard to integrate. In spite of that, recently an approach to AI called hybrid or neurosymbolic AI has been proposed. This purportedly hybrid approach to AI needs to be tested by time and deserves a dedicated study.

In this paper we have striven to describe the state of affairs within AI research. Hopefully this descriptive effort can be of value for the AI community to make a step forward in its own reflection towards the future.

References

1. Brooks, R.: Is the brain a good model for machine intelligence? *Nature* **482**, 462–3 (2012). <https://doi.org/10.1038/482462a>
2. Chomsky, N.: Language and nature. *Mind* **104**(413), 1–61 (1995). <https://doi.org/10.1093/mind/104.413.1>
3. Cozman, F.: No canal da Inteligência Artificial: nova temporada dos desgrenhados e empertigados. *Estudos Avançados* **35**(101), 7–20 (2021). <https://doi.org/10.1590/s0103-4014.2021.35101.002>
4. Darwiche, A.: Human-level intelligence or animal-like abilities? *Commun. ACM* **61**(10), 56–67 (2018)
5. Dijkstra, E.: The threats to computing science. In: Talk delivered at the ACM 1984 South Central Regional Conference, November 16–18, Austin, Texas (November 1984). <http://www.cs.utexas.edu/users/EWD/transcriptions/EWD08xx/EWD898.html>. Accessed 10 Jun 2021
6. Gonçalves, B.: Machines will think: structure and interpretation of Alan Turing’s imitation game. Ph.D. thesis, Faculty of Philosophy, Languages and Human Sciences, University of São Paulo, São Paulo (March 2021). <http://dx.doi.org/10.11606/T.8.2021.tde-10062021-173217>
7. Hartree, D.: *Calculating Instruments and Machines*. University of Illinois Press, Champaign (1949)

8. Hayes, P., Ford, K.: Turing test considered harmful. In: Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI 1995), pp. 972–7 (1995)
9. Hudson, M.: IA researchers allege that machine learning is alchemy. *Science* (3 May 2018). <http://dx.doi.org/10.1126/science.aau0577>
10. Katz, Y.: Noam Chomsky on where artificial intelligence went wrong: An extended conversation with the legendary linguist. *The Atlantic* (1 Nov 2012) (2012). <http://www.theatlantic.com/technology/archive/2012/11/noam-chomsky-on-where-artificial-intelligence-went-wrong/261637/>
11. LeCun, Y.: My take on ali rahimi’s ”test of time” award talk at nips. <https://www.facebook.com/yann.lecun/posts/10154938130592143>. Accessed 3 June 2021
12. Lenat, D., Feigenbaum, E.: On the thresholds of knowledge. *Artif. Intell.* **47**(1–3), 185–250 (1991). [https://doi.org/10.1016/0004-3702\(91\)90055-O](https://doi.org/10.1016/0004-3702(91)90055-O)
13. Lenat, D.: Cyc: a large-scale investment in knowledge infrastructure. *Commun. ACM* **38**, 33–8 (1995)
14. McCarthy, J.: Programs with common sense. In: Proceedings of the Teddington Conference on the Mechanization of Thought Processes, Her Majesty’s Stationery Office, London (December 1958). <http://www-formal.stanford.edu/jmc/mcc59.pdf>. Accessed 3 June 2021
15. McCorduck, P.: *Machines Who think: a Personal Inquiry into the History and Prospects of Artificial Intelligence*. A. K. Peters, second edn. CRC Press, Boca Raton (2004 [1979])
16. Minsky, M.: Some methods of heuristic programming and artificial intelligence. In: Blake, D.V., Uttley, A.M. (eds.) *Proceedings of the Symposium on Mechanisation of Thought Processes*, vol. 2, H. M. Stationery Office, London (1959)
17. Minsky, M.: Steps toward artificial intelligence. *Proc. IRE* **49**, 8–30 (1961). <https://doi.org/10.1109/JRPROC.1961.287775>
18. Minsky, M.: *The Society of Mind*. Simon & Schuster, New York (1985)
19. Minsky, M.: *Smart Machines*. In: Brockman, J. (ed.) *The Third Culture: Beyond the Scientific Revolution*, chap. 8. Simon & Schuster, New York (1995)
20. Nilsson, N.: *The Quest for Artificial Intelligence*. Cambridge University Press, Cambridge (2009)
21. Norvig, P.: On chomsky and the two cultures of statistical learning (2012). <http://norvig.com/chomsky.html>. Accessed 3 June 2021
22. Pearl, J.: *The Book of Why*. Basic Books, New York (2019)
23. Rahimi, A., Recht, B.: NIPS ”test-of-time award” keynote address (2017). <http://www.youtube.com/watch?v=Qi1Yry33TQE>. Accessed 3 June 2021
24. Russell, S., Norvig, P.: *Artificial Intelligence: a Modern Approach*. 1st edn. Prentice Hall, Hoboken (1995), ISBN 0-13-103805-2
25. Russell, S., Norvig, P.: *Artificial Intelligence: a Modern Approach*. Pearson Series in Artificial Intelligence, Pearson, 4th edn. (2020). ISBN 9781292401133
26. Sent, E.M.: Herbert A. Simon as a cyborg scientist. *Perspect. Sci.* **8**(4), 380–406 (2000). <https://doi.org/10.1162/106361400753373759>
27. Simon, H.: Artificial intelligence: an empirical science. *Artif. Intell.* **77**(1), 95–127 (1995). [https://doi.org/10.1016/0004-3702\(95\)00039-H](https://doi.org/10.1016/0004-3702(95)00039-H)
28. Simon, H.: *The Sciences of the Artificial*. 3rd edn. MIT Press, Cambridge (1996 [1969])
29. Simon, H., Newell, A.: Human problem solving: the state of the theory in 1970. *Am. Psychol.* **26**(2), 141–59 (1971). <https://doi.org/10.1037/h0030806>

30. Sutton, R.: The bitter lesson. <http://incompleteideas.net/IncIdeas/BitterLesson.html>. Accessed 15 June 2021
31. Turing, A.M.: Computing machinery and intelligence. *Mind* LIX (236), 433–60 (1950). <https://doi.org/10.1093/mind/LIX.236.433>