[This has been updated. The new version is available here.]

AI, a "Common Sense" Approach

Jim Burrows, Eldacur Technologies

Introduction

This paper is a work in progress. For the last few months, I have been taking a bit of a sabbatical in order to delve deeper into subjects that have interested me, but which the demands of being the VP of Engineering or Advanced Development did not allow me the time and attention to focus on. As I find myself beginning to be ready to act on what I have been learning, it seems useful to put a few of my thoughts to paper, or at least bits, and circulate them for critique and discussion.

A skeptic's perspective

Ever since I started hanging out in the Engineering library at MIT and hacking the AI machine over the ARPAnet, I have been interested in, but at the same time troubled by, the subject of AI. I was intrigued because it promised to bring into reality the robots, the artificial people, that I read about in the works of Simak, Binder, Asimov, and the rest. But I was troubled because after studying programming and psychology, and hacking computers, I just didn't see how you could create intelligence from software, at least not intelligence that was the same sort of thing that humans exhibit.

The promise always was (and seems to still be even now, four decades later), that true AI was 25 years away but we were making real progress. Computers were electronic brains, just very simple ones, but as technology advanced via Moore's Law they would become more and more intelligent. This seemed to me to miss the mark in a couple of ways. First, computer programs seemed to have no intelligence, and no matter how much you multiplied zero by it wasn't likely to get you anywhere. There also seemed to be a qualitative difference between me, or even my cat, and the game-playing programs I could write, or the more sophisticated ones others were working on, or chatbots like Eliza and Parry. Still, I knew how bright the guys in the AI lab were, so I didn't push it.

I've looked in on AI any number of times over the years, and while they have created more and more complex and cleverly built systems, it has never seemed that they were any closer to real AI, or Artificial General Intelligence—AGI—as human-like AI has come to be known. This didn't surprise me. My misgivings have never gone away.

Recently, I've been looking into what it would mean to make "trustworthy" systems, that is software systems and services that are worthy of our trust. This appears to be more and more timely as we begin to see artificial "assistants" that can answer our questions and perform our tasks; and cars and trucks and airplanes are becoming autonomous; and robots and drones are finding their way onto the battlefield. None of these systems constitute real AGI, nor could be expected to pass the Turing test, but still, they are stepping into the roles of our servants. Can they become trustworthy even before they are intelligent?

In exploring this question, I've been looking into the state of robot and AI ethics, and one of the difficulties is that a lot of the work is being done on the assumption that we will achieve AGI (no

doubt in 25 years), and I still don't believe in AGI coming from the current major paradigms. As I tried to explain to friends and colleagues why the current quest for AI seems futile, I found that somewhere over the last couple of decades, my thinking on the matter had become well enough defined that I could put it into coherent words.

This document is an attempt to do that, so as to, on the one hand, encourage criticism and feedback and, on the other, by defining the flaws I see in the usual directions and assumptions, clarify my own thoughts on what direction might seem more fruitful. It is something of a personal meditation upon the subject, at least for now. Perhaps when I am done, it will be more rigorous, focused, and generally useful.

Intelligence?

I should perhaps take a moment to be clear what I mean by “intelligence”. The AI discipline has, over the years, provided us with any number of extremely useful techniques and technologies. I am not doubting that. However, there is, I would maintain, a difference between useful algorithms, heuristics, or techniques, and intelligence that really thinks the way that humans do. When I write about “human-like intelligence” or adopt the terminology of “Artificial General Intelligence (AGI)”, I mean an artificial system that can replicate all of the intellectual functions of the human mind, and not just one or two specialized tasks. A true human-like AGI, would need to exhibit understanding, planning, judgement, intention, intuition, learning, and creativity.

It is, perhaps, not necessary for such an intelligence to work precisely the way that a human mind operates. There may be equivalent ways for minds to work, after all chimps[1], dolphins[2], elephants[3], ravens[4] and octopuses[5] all exhibit some level of intelligence, and sense and manipulate the world in very different ways. However, since we have humans to examine both objectively and subjectively, it would seem that the easiest way to build an AGI is to make one that replicates our own abilities and their interrelationships.

Hard logic

Anyone who has taught classes in formal and symbolic logic can testify that teaching logic to human beings is not easy. Rigorously logical thinking does not come naturally to us. Deductive and mathematical proofs, despite their precision and clarity, take a lot of work to master. Our

---

[1] Jane Goodall, “About Chimpanzees”, The Jane Goodall Institute of Canada web page, http://www.janegoodall.ca/about-chimp-so-like-us.php#Intelligence (seen July 2015)

[2] Lori Marino, et al, “Cetaceans Have Complex Brains for Complex Cognition”, PLOS Biology, http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.0050139 (seen July 2015)

[3] Lisa Newbern, “First Evidence to Show Elephants, Like Humans, Apes and Dolphins, Recognize Themselves in the Mirror‘, Emory University we page, http://www.yerkes.emory.edu/about/news/developmental_cognitive_neuroscience/elephants.html (seen July 2015)

[4] Anna Smirnova et al., “Crows Spontaneously Exhibit Analogical Reasoning”, Current Biology, Volume 25 , Issue 2 , 256 - 260, http://www.cell.com/current-biology/abstract/S0960-9822(14)01557-7

[5] Brendan Borrell, “Are octopuses smart?”, Scientific American web page, http://www.scientificamerican.com/article/are-octopuses-smart/ (seen July 2015)

"mental machinery" seems far more suited to jumping ahead to "obvious" conclusions based on insufficient evidence—"intuitively". When it works, this is the great genius of creative people, but if we look at it from the perspective of logic and critical thinking, It is the very definition of a whole class of fallacies.

Intelligence—natural, human intelligence—builds up *to* logic and critical thinking, **not** up *from* it. Until it can embrace the intuitive or creative leap, artificial intelligence will not be intelligence. For natural intelligence, logic is hard; pattern matching, novelty, predictions, and error are easy. Starting with the hard part, logic, and building it using machines that are instantiations of the very thing that they are trying to implement, misses the mark. Intelligence isn't the end product, regularized, formalized critical thinking. Rather it is the mechanism that allowed us to make the journey to the invention of formal logic.

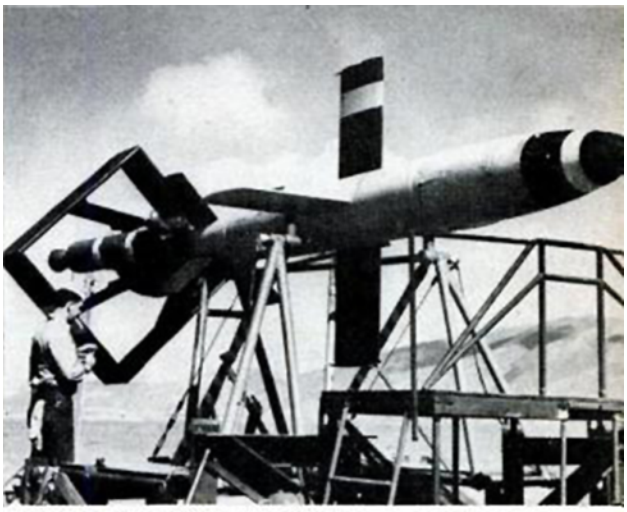
This takes us back to the definition of AGI, which is able to replicate all of our intellectual functions, including, in the words of Wikipedia, the abilities to "reason, use strategy, solve puzzles, and make judgments under uncertainty; represent knowledge, including commonsense knowledge; plan; learn; communicate in natural language; and integrate all these skills towards common goals."

Electronic brains?

Ever since the invention of computers, they have been described as "electronic brains". Take, for instance, this early description of Univac[6].

> *This system centers around a high-speed electronic brain capable of manipulating electric code impulses at rates over 2,000,000 per second.…*
>
> [There follows a list of system components, including…]
>
> *"Univac". This is the electronic brain or computer. It will handle data at a rate equivalent to 60,000 words per minute.*

Likewise, the Popular Science[7] picture to the right has "radar eyes and an electronic brain".



IF THE BOMBERS COME, a near-human rocket with radar eyes and electronic brain could be our close-in defense. **Page 124**

As computers, calculators and other forms of automation came into use enabling more accurate gunnery and missile guidance, it was natural to see some analogies between the electro-mechanical engines and the chemo-electrical operation of the brain. Lacking a well-established nomenclature, writers needed metaphors and analogies to describe the emerging technology.

But while these images were evocative and conveyed some idea of the roles of the new devices, they are also quite misleading in

---

[6] Robert S. Casey et al, Reinhold, 1951, "Punched Cards: Their Applications to Science and Industry", pp 74–75, https://books.google.com/books?id=CRwONNTeWe8C (seen July 2015)

[7] Popular Science, October 1950, Vol. 157, No. 4, pp 97, 124–128, https://books.google.com/books?id=7iwDAAAAMBAJ&pg=PA97 (seen July 2015)

detail. Computers are digital and operate sequentially. Living systems are analog and highly parallel. While we have emulated and simulated neurons and neuron nets, including systems like IBM's "TrueNorth" with hundreds of billions of simulated neurons and hundreds of trillions of simulated synapses[8], nature has managed to create a fully working nematode with only 302 neurons, whose behavior we cannot yet emulate[9].

Computers are wonderful at simple deterministic games involving a small number of extremely precise rules. Deep Blue, and its successors have proven that, by vastly outplaying human chess masters. However, chess is a task perfectly suited to a computer. It is based, after all, on logic, and deterministic, provable reasoning.

Chess, like logic, is hard for natural intelligences, so we chose it as a Holy Grail of AI early on. Surely, the argument went, an artificial intellect that could master a game like chess would be operating at the highest level of intelligence. This only illustrates how poorly we understand natural intelligence. The very attributes that make chess hard for humans are what make it easy for machines, machines of sufficient complexity. In the real world of nature, a 302 neuron worm is way more intelligent.

Developing intelligence

Another aspect of natural intelligence that may be undervalued, or left out of classical AI thinking is the developmental nature of intelligence. Natural intelligences, especially human beings, do not spring into existence fully formed. Rather, we are born quite incompetent and have to learn the simplest of mental skills, and then build more and more complex abilities on top of them. Each individual goes through this. Software, once written, just boots and works.

Early on it was assumed that an advantage of software as a basis of AI was that all that learning would be done by the humans learning how to build a fully formed intelligence, and once that was accomplished all AIs would be created fully intelligent. That's much more efficient.

More recently, as algorithms for creating learning- and knowledge-based systems were developed, the cognitive skills and the knowledge that they work with have started to be handled separately. Knowledge, associations, and connections are acquired over time, more like the way that humans—or at least adult humans—learn. Still, the cognitive skills used in that learning are programmed in from the beginning.

Intelligence, I would argue, is based in part on learning how to learn. While this may be less true at the nematode level, it becomes critical at human levels. If we, as programmers, appropriate the learning role, it would seem that with it, we steal much of the intelligence.

Then what?

If logic, rigorous rulesets, and critical thinking are not the key to intelligence, then what is?

---

[8] Kurzweil News page, November 19, 2012, "IBM simulates 530 billion neurons, 100 trillion synapses on supercomputer", http://www.kurzweilai.net/ibm-simulates-530-billon-neurons-100-trillion-synapses-on-worlds-fastest-supercomputer (seen July 2015)

[9] Alexey Petrushin et al, SPIE Proceedings, "The Si elegans connectome: A neuromimetic emulation of neural signal transfer with DMD-structured light", http://spie.org/Publications/Proceedings/Paper/10.1117/12.2085032

The common sense

In order to understand how human and animal perceptions of the world worked, Aristotle developed the notion of the "κοινὴ αἴσθησις" or "common sense". This is an internal faculty that integrates the perceptions of the five senses—sight, hearing, touch, smell and taste—into a coherent view of the world. Plato had previously ascribed this role to humanity's rational function, but Aristotle realized that animals must also be able to integrate their sensory data into a coherent whole which they could recognize and remember. He thus ascribed this function to the realm of the senses rather than rational thought.[10]

Medieval thinkers elaborated Aristotle's notion into a system of five external senses and five internal "wits". The function of the "common wit", once more, was the facility to take the perceptions of the senses and integrate them into a view or model of the world. The other wits were variously given as "imagination", "fantasy", "estimation", and "memory"; or "imagination", "reason", "intelligence", and “memory".

In any of these views, "common sense" served the same foundational function: stitching the impressions of the various senses into a coherent integrated experience of the world, a model wherein colors, shapes, textures and other senses are all understood to be aspects of objects in the external world.

Today, we recognize that there are more than five senses, and that different animals have different sets of senses, but the concept of a function that integrates the various senses into a common sensory experience is still valid, and is studied by psychologists and neurobiologists. Today we speak of sensory fusion, gestalts, amodal information, multisensory or multimodal integration, and the binding problem, but the function of the common sense remains, even though the terms have changed. It is still an area much in need of study.

Making sense

From Plato to Descartes and Kant to contemporary AI researchers, understanding the world, integrating sensory data is often seen in terms of the rational intellect. Our understanding of the mind is often in terms of intelligence, reason, and rational thinking. Thinking in terms of The Common Sense redirects our focus to our senses and our perceptions as the foundation of our experience, consciousness, and understanding. You might say that it shifts the focus so that we are thinking about cognition in terms of *recognition*.

The naturalness of this view can be seen in our use of language. If an intellectual concept or theory is described to us, when we comprehend it, we will often say that it "makes sense", and we will often say of things that defy our understanding that they aren't "sensible" and that we "cannot make sense of them". What we are saying is that they defy our common sense, that we cannot integrate them into our model of, our understanding of, the world. This, I would argue is at the root of "natural intelligence”: the integrative common sense function that allows us to model and think of the world, to put us inside the world of our sensory experience.

If this is so, the key to artificial intelligence is not rules and logic, which are the heart of software, but rather a process of integrating sensory data from multiple senses into an integrated perception and experience. Once we can create a system that can turn the “blooming, buzzing

---

[10] Wikipedia, “Common sense: Aristotelian common sense”, https://en.wikipedia.org/wiki/Common_sense#Aristotelian_common_sense

confusion" of multiple senses into something "sensible" something that "makes sense", then, and only then, will we be on our way towards artificial intelligence, towards thinking systems that can build from recognition to cognition, and we will be on the way towards systems that can become autonomous moral agents.

Common sense, in theory

As I have been researching this area, I have come across a few researchers and theorists who are taking a more "common sense" view in intelligence and consciousness. These include:

- Giulio Tononi's Information Integration Theory of consciousness (IIT). [11,12]
- Monica Anderson's Artificial Intuition (AN) theory appears to be close to my own "Common Sense" notions. Her contrasting of intuition vs logic[13] is very similar to what I have written above with regard to logic vs "common sense".
- During its first decade or so, cybernetics studied neural networks and the parallels between natural and artificial systems. In the late 1950s, AI emerged as a separate discipline with a different focus, and the cyberneticists seem to have shifted their focus elsewhere. Realizing this makes me want to reacquaint myself with the early cyberneticists, such as Norbert Wiener.

Law or virtues?

The study of normative ethics is generally divided into three approaches

- Deontology, which is rule based
- Consequentialism, which focuses on the ends
- Virtue or aretaic ethics, which focuses on one's character and virtues

The deontological approach, with its emphasis on rules, would seem to be the obvious choice for a software based system. In what may be the most advanced instantiation of this approach, Selmer Bringsjord and his colleagues at the Rensselaer AI & Reasoning Lab have done considerable work in developing and using a "Deontic Cognitive Event Calculus" system, exemplified in his paper with Joshua Taylor, "The Divine-Command Approach to Robot Ethics[14]".

However, the "common sense" model of AI would seem to be better suited to a virtue-based approach. At least as I understand it, a sensory-based system is likely to be based more on fuzzy, error prone (and constantly corrected) pattern matching and associations, that are less

---

[11] Giulio Tononi, "An information integration theory of consciousness," BMC Neuroscience 5: 42 (2004), doi: 10.1186/1471-2202-5-42.

[12] Giulio Tononi, Christof Koch, "Consciousness: Here, There but Not Everywhere," Philosophical Transactions of the Royal Society B 370: 20140167 (2015), doi: 10.1098/rstb.2014.0167; arXiv:1405.7089 [q-bio.NC]

[13] Monica Anderson, "Intuition and Logic", http://artificial-intuition.com/intuition.html

[14] Patrick Lin, Keith Abney, George A. Bekey. (2011). Robot Ethics: The Ethical and Social Implications of Robotics (Intelligent Robotics and Autonomous Agents series) (p. 85). The MIT Press. Kindle Edition.

well suited to rigorous formal rule systems. Virtues, it would seem, are similar fuzzy categories, and mapping actions and decisions to them seem, at least on the surface, to be more appropriate.

Without going into the details of my own general philosophy, I will assert that the reason that logic, ethics and aesthetics have been separate areas of study for thousands of years is that we as humans, value things along three different dimensions: rational, measuring truth; ethics, measuring the moral good; and aesthetics, measuring beauty or harmony. While there are certain similarities between these dimensions—all of them measure some flavor of "good" vs "bad", and there are parallels and ambiguities in our language such as the word "right" being used both for logically correct, and ethically righteous—they are separate dimensions. One makes, I will argue, a mode error if one confuses them or attempts to reduce one to another.

That being said, I am at the very least skeptical of the wisdom, and ultimate success, of efforts such as Bringsjord and Taylor's to create mathematically provable ethical system, and a system of propositional calculus with which to prove it. While it may provide some sort of deontological framework for constraining the behavior or logical reasoning of a software-driven system such as those envisioned by classical approaches to AI, I have doubts in two areas. The first is the practicalities of analyzing real-world situations into mathematically or logically precise, unique, and unambiguous formulations to serve as grist for this deontological mill. The second is that it profoundly subordinates ethics to logic and the resulting modal error robs it of its ethical character.

It may well be, however, that I am misconstruing the state of affairs. I have not been actively involved in AI, nor have I fully kept up on the literature. In my recent researches, I came across but have not yet read a paper by Bringsjord, Taylor, and their colleagues, "Piagetian Roboethics via Category Theory: Moving Beyond Mere Formal Operations to Engineer Robots Whose Decisions are Guaranteed to be Ethically Correct", whose title is filled with intriguing possibilities. It is very near the top of my next to be read pile. I'm still only part way through "Robot Ethics", at present.

The virtuous AI

This brings us to a meta-ethical question, "What virtues ought an AI/AGI have?", one that we certainly cannot resolve at present, but is worth at least considering in some detail. I can think of at least four roles that we might cast an AI in, each with its own set of associated virtues.

- An idealized virtuous person. As a classic example of this, we could consider the stereotypical "Eagle Scout".
- Digital virtual assistants and the like, which is where I started my researches, suggest the virtues of a personal servant, a butler, valet, or executive assistant.
- Alternatively, many AIs may serve the public rather than individuals, and thus require the virtues of a civil servant and the "good cop".
- As unmanned weapon systems become more autonomous, we must consider the war fighter. This suggests various codes of warrior ethics, as well as the values inherent in international law.

The Eagle scout provides the easiest list of virtues.

> *A Scout is trustworthy, loyal, helpful, friendly, courteous, kind, obedient, cheerful, thrifty, brave, clean, and reverent.*

There is no such clear-cut list of virtues for personal servants, but surveying the writings of professional butlers and those who train them, one might say that

> A butler is trustworthy, loyal, discreet, refined, professional, dedicated, organized, deferential, adaptable, polite, and friendly.

A civil servant, like those in private service, must be loyal, but their loyalty is to society as a whole, rather than putting their client's interest and wellbeing first. Putting others—society—before themselves—bravery—is probably more important. Discretion is still of some importance, but subordinate to the public interest and honoring the law would come substantially higher.

A warfighter swears allegiance, that is absolute loyalty and obedience, within the law, and is expected to live up to an honor code, which generally pledges honesty (e.g. "We Will Not Lie, Steal Or Cheat, Nor Tolerate Among Us Anyone Who Does"[15]). Bravery would seem to be a prerequisite. Naturally, they are expected to use lethal force, both defensively, and offensively, in accordance with the current rules of engagement. International Law and the rules of war are generally followed.

Virtue before intelligence

If actual AGIs are still in the relatively remote future, what do we do now? This, in fact, is the main question that I have been looking into during my sabbatical of the last few months. My thoughts on AI, that constitute the bulk of this document are something of a diversion from that question, or perhaps a bit of useful context.

Recently, we've seeing a plethora of systems which exhibit more and more human-like characteristics. Virtual digital assistants that are capable of voice interaction and "natural language" text input abound. Examples include Siri, Google Now, Cortana, Alexa and many more. Cars and trucks are becoming more and more capable of driving themselves and many commercial air flights are handled autonomously from end to end. Drones, used both for surveillance and attack missions, are becoming more autonomous, as are military and law enforcement systems.

As these systems, on the one hand, have greater access to sensitive data, information, and knowledge about us, and perform more and more potentially dangerous tasks, and on the other, interact with us in more and more human-like fashion, we begin to interact with them *as if* they were persons, and even autonomous moral agents. They are certainly very far from being able to pass a formal Turing test and even further from true AGI and moral agency, but they are getting closer and seem to have crossed a threshold to a form of low-grade personhood, at least in terms of how we treat them and expect them to act.

Because they are not actually the sophisticated creatures that we treat them as, they lack the judgement and perception needed for a true deontological ethical system. It would thus seem that the best that we can do, as we program their behavior, is to insure that it is in keeping with certain virtues. This will represent the bulk of my work in this area, identifying the key virtues that these personified system ought to exhibit and then defining the types of behaviors that would be in keeping with those virtues. It is my hope that, as we move forward towards a true

---

[15] US Air Force Academy honor code, USAF Academy web page, http://www.academyadmissions.com/the-experience/character/honor-code/ (seen July 2015)

common sense AI, we will be able to incorporate these same virtues, and presumably more, in the intelligent systems.

Personifying virtue

If I had to boil down the virtues from above as to which are most applicable to AGI's, I would probably go with something along these lines:

- trustworthy
- loyal
- discreet
- candid
- law abiding[16]

Many of the others, such as helpful, courteous, and such, are issues of style, tone, usability, and UI/UX considerations that software already deals with. These five, on the other hand, have a more clear cut ethical nature.

Afterthought

As I prepared to post this on the web for the first time, I came across an article by Luciano Floridi of the University of Oxford[17] that makes a very useful distinction between the two types of AI, the craft that has contributed so much to computers and related engineering fields, and the search for a true AGI. He refers to them as "the two souls of AI", the Smart (AI engineering technologies) and the Clever (AGI cognitive technologies). One way to summarize my thesis in this paper is that the Aristotelian or Medieval "Common Sense" is the path towards Floridi's Clever AI. I recommend Prof. Floridi's article.

---

[16] Note that this paper was written early in my project. In later documents; I tend to use "obedience" rather than "law-abiding". My categories are still somewhat fluid, and will be, no doubt, so long as this is a one-man exercise, rather than a full blown R&D project.

[17] Luciano Floridi , "The two souls of Artificial Intelligence: the Smart and the Clever", Che Futuro!, September 9, 2015, http://www.chefuturo.it/2015/09/artificial-intelligence-smart-cleve/ (seen September 2015)