Humanity 2.0, Not Golem 2.0

Stop pouring money down the rabbit hole, start believing in people

July 15, 2017

Kenneth B. Tingey, PhD, MBA, Master of Pacific International Affairs
Miroslaw Manicki, MD, MPH

Fluidity Finance
Logan, Utah/Piotrków Trybunalski, Poland

Contents

People, not golem

In Jewish lore, pre-modern Central European rabbis made human-like figures out of clay and then animated them. There is a long tradition of this – Adam, the first human, was said to have first been a golem, before he was enlightened, or energized.

A golem by legend could be activated by writing a 'shem' on his forehead or in his mouth and deactivating by removing at least a part of it. Other sources say once the golem had been physically made one needed to write the letters aleph, mem, tav, which is emet and means "truth," on the golem's forehead and the golem would come alive. Erase the aleph and you are left with mem and tav, which is met, meaning "death." One was made, according to legend, by Rabbi Elijah Ba'al Shem of Chelm, Poland. The rabbi created the golem to do hard work for him, which it did "for a long time".

FIGURE 1: A PRAGUE REPRESENTATION OF THE GOLEM. WIKIPEDIA/CREATIVE COMMONS

The golem grew large and the rabbi feared that it would "destroy the universe." He decided to deactivate the golem, which scratched the rabbi's face when the name was removed before returning to dust. Another golem was created in the late 16th century in Prague to defend the people. The golem was deactivated each week during the time of the Sabbath to keep it from desecrating the Sabbath and to let it rest. In some versions of the story, it fell in love and grew dangerous. Eventually, it was deactivated by the Rabbi Judah Loew ben Bezalel, who had created him. By legend, that golem is still in the attic of that synagogue, in inert form.

This is all different than the Gollum of J.R.R. Tolkien's *Lord of the Rings*. That is a very interesting, important story, but a different one.

'Artificial intelligence' golem a very bad idea, even if it were to work

Artificial intelligence, or AI, has been compared, and not favorably, to a golem. Benjamin Blech, in "Stephen Hawking's Worst Nightmare? Golem 2.0", characterizes Dr. Hawking's fear of AI with respect to mankind:

> Hawking is clearly someone undaunted by personal fears. Yet in a recent BBC interview, Hawking confided that he was deeply concerned for the future of humanity. The cause of his concern is artificial intelligence, or AI, the creation of intelligent machines able to "outthink" their creators. What began with IBM's Watson supercomputer, capable of handily beating chess grandmasters and the best players on "jeopardy!" may in the near future, Hawking warned, checkmate its designers to become the Earth's ruler.[1]

FIGURE 2: GOLEM OF PRAGUE. FINKORSWIM.COM: HTTPS://TINYURL.COM/Y7XC6DM3

We do not in any way think that AI is going to 'outthink humanity' in meaningful ways. People do not exist to serve some abstract, impersonalized purpose; all is available to meet our needs and preferences – wisdom and resource limits notwithstanding. Thinking and feeling are far too intimately connected to dissect them in such a manner. Machines are not a part of such activities, nor are the capable of sociality and group decisionmaking.

If granted more power and control over equipment and sensors, however, they can wreak havoc and create unwanted consequences of many kinds.

> Many scholars believe that it was the legend of the golem that inspired Mary Shelley to write her famous Frankenstein novel about an unorthodox scientific experiment that creates life, only to reap the horrifying results when the achievement goes terribly wrong.

> Creation without control is a formula for catastrophe. The history of scientific achievement bears ample testimony to the simple truth that progress detached from the restraints of moral and ethical considerations may grant us the knowledge to penetrate the secrets of nuclear fission, but at the cost of placing mankind in danger of universal annihilation.

> The story of the golem of Prague is a paradigm for the hazard of permitting what we create to go far beyond our intent. Artificial intelligence, as an extension of our intellectual ability, certainly

---

[1] Blech, B. 2015, January 4. *Stephen Hawking's worst nightmare? Golem 2.0* Forward. Available: https://tinyurl.com/pjskpdn.

has many advantages. Yet it cannot really "think," It has no moral sensitivity. It does not share the ethical limitations of its programmer. And it is not restricted by the values of those who brought it into being.

It cannot "feel" either, although it can interpret data from many kinds of sensors. Turning important decisions over to computer reasoning would create catastrophic messes where decisionmaking matches up with the realities of nature, choices, and human preferences. What about situations of preference, choice, degree, and group and social decisionmaking. What about persuasion? What about exceptions to the rule? What about religious and political issues? We are in the throes of new governance and social models of great import, with deep cultural and sociological roots. Are computers going to just 'figure them out'?

The question in such cases is, "how much damage will there be before the machines in question can be turned off, or at least diverted to other activity?"

The promise of AI is unusual. It is a solution to an unarticulated problem. The problem is easy to identify: It is a lack of confidence in humanity, in the capabilities of people. There is vague worry that mankind is not up to the task. Given the limitations of people – cognitively and socially – we need to turn to machines to make up for our shortcomings. This is utter folly. We need to turn this around, to operationalize hope. We need to reorient mankind in this manner – to achieve "Humanity 2.0", certainly in lieu of "Golem 2.0", which is what AI would become to us – very problematic even if it appeared to work at some point.

FIGURE 3: JOSEPH MCKENNA AND ROBERT PROSKY IN FISHELSON'S ADAPTATION OF THE GOLEM AT MANHATTAN ENSEMBLE THEATRE, 2002

In short, AI is a presumed phenomenon that will allow computers to think, to become sentient in the way that humans are sentient. The idea is that if you can load enough facts into computers, they will learn to line them up and make decisions – and eventually to 'fill in the blanks' and figure things out on their own. They will get better at this than we are, and will pass us up. Such a state is called a 'singularity' in the AI literature.[2] Since the 1956 "Dartmouth Workshop," w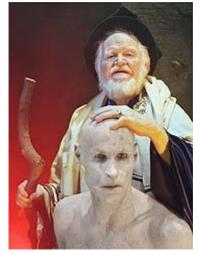hen the ideas were first mentioned,[3] forecasts for the singularity have continually been 'moved back',[4] but their implications have been expanded on.[5] As of now, they say that the singularity won't happen in the foreseeable future, but the implications have been fully fleshed out in the form of many Hollywood blockbuster movies. Of course, such works are universally frightening.

Do we need to have computers think? Isn't that what *we* do? The point with AI is that computers will think better than we do. We must ask: Is there a problem with the way people think? It is a fair question. Before we turn the task over to something other than people, aren't there many questions that need to be considered? What do people need? Are they getting it? If so, fine. If not, is there anything that people could do to achieve this – and is there something that used to occur that could well be brought back? Recall the philosophy of Plato between the 5th and 4th centuries BCE, as expounded beautifully in the Republic: We would do well under the "Rule of the

---

[2] Kurzweil, R. 2005. *The singularity is near: When humans transcend biology.* New York: Penguin Books.

[3] Haugeland, J. 1985. *Artificial intelligence: The very idea*. Cambridge, MA: A Bradford Book/The MIT Press, 176, .

[4] Haugeland, J. 1996. What is mind design? In Haugeland, J. (Ed.), 1997, *Mind design ii: Philosophy, psychology, artificial intelligence,* Rev. & enlarged. Cambridge, MA: A Bradford Book/The MIT Press, 1-28.

[5] Kurzweil, R. 2012. *How to create a mind: The secret of human thought revealed*. New York: Penguin Books.

Wise".[6] This involves, to be sure, working out of human issues by and for humans. ***Let us do that better: Let us introduce Humanity 2.0 in form of renewal.***

Having considered such matters, we now comment on such questions based on the seven categories of the Miller living systems model as seen in *Table: Thought and computational needs of Miller living systems levels*.

Table: Thought and Computational Needs of Miller Living Systems Levels		
Living system level	"Thinking" issue	AI Comment
Cell	No known abstract space	There is discussion as to beneficial use of AI in such environments; priority should be given to understanding and support of natural processes such as they exist, rather than replacing them altogether
Organ	No known abstract space	
Organism	Abstract reasoning is a defining aspect of being human; many choices	This is the AI target; no cohesive argument is given, except, perhaps, speed and complexity within a given set of rules
Group	Complex questions of collaboration and governance; many choices	Discussions among reasoning computers? There is little discussion of this; computers do a poor job of supporting group decisionmaking as it stands
Organization	Complex questions of collaboration and governance; many commitments	Computers do a poor job of supporting group decisionmaking; as it stands, they introduce many inane factors and arbitrary complexities that fundamentally compromise performance of organizations
Society	Very complex need to combine traditions; many choices with multilevel implications	Little consideration for social issues is given; AI in social media tools run roughshod over social needs and preferences; AI would do a very poor job of identifying and supporting social questions, including traditions and preferences
Supranational	Broad policies, but with far-reaching implications; many choices	AI is poorly positioned to assist at this level, although improved computing infrastructures would be helpful, a more beneficial use of time and resources

Computers should be significantly improved as to their ability to support human social needs generally before improved decisionmaking – particularly that of replacing human thought – is worthy of consideration. What is needed is improved support of human decisionmaking at all levels of living systems as per Miller.

It will not work

There is a second important example of AI folly, this time a literary one. Once you accept the AI proposition, you find yourself falling down Alice's rabbit hole. Nothing is real; nothing is sound. You are in a world turned upside down – or inside out, or whatever is your preferred metaphor for unmoored interactions.

The rabbit in the story being an erstwhile 'AI guy', running and talking, in a hurry and always in motion, lured Alice – the rest of society – into a morass of irresolution, to be sure, of money of untold quantities:

>*…suddenly a white Rabbit with pink eyes ran close by her.*

>*There was nothing so very remarkable in that; nor did Alice think it so very much out of the way to hear the Rabbit say to itself "Oh dear! Oh dear! I shall be too late!" (when she thought it over afterwards, it occurred to her that she ought to have wondered at this, but at the time it all seemed quite natural); but, when the Rabbit actually took a watch out of its waistcoat-pocket,*

---

[6] Stevenson, L. 1987. *Seven theories of human nature*, 2nd ed. Oxford, UK: Oxford University Press, 27-39.

and looked at it, and then hurried on, Alice started to her feed, for it flashed across her mind that she had never before seen a rabbit with either a waistcoat-pocket, or a watch to take out of it, and burning with curiosity, she ran across the field after it, and was just in time to see it pop down a large rabbit-hole under the hedge.

In another moment down went Alice after it, never once considering how in the world she was to get out again.

FIGURE 4: ALICE AND THE 'RABBIT IN A HURRY'. BBC LEARNING ENGLISH DRAMA: HTTPS://TINYURL.COM/Y7TZYP3C.

The rabbit-hole went straight on like a tunnel for some way, and then dipped suddenly down, so suddenly that Alice had not a moment to think about stopping herself before she found herself falling down what seemed to be a very deep well.[7]

…and all bets were off. Why was the Rabbit – representative of AI-proponents in our case – in such a hurry? In the case of AI, there is a powerful drive to make 'it' happen. Some of this is associated with market cycles. IBM, for example, has staked much of its future in such plans. They have outside investors to please. All major Silicon Valley technology companies refer to AI prospects – arguably not just 'singularity' prospects, but means of introducing high levels of logic and decisionmaking in their systems. Some of this is driven by a capitalistic motive – computers do not 'eat' and people do, making AI systems theoretically less expensive. Kenneth has worked with AI proponents in the past. There was – and is – an enormous '*fudge factor'* in what they do. Consider this: Think of the fundamental compromise to our institutions – from governments on down – from people exercising their personal prerogatives rather than the general good.[8] The Silicon Valley is all about this. Imagine the distraction when machines can exercise prerogative in such ways.

FIGURE 5: MONEY, TIME, EFFORT, AND HOPES DOWN THE GOLEM 2.0/AI HOLE

Our personal belief is that AI proponents are in such a hurry because they are concerned that people will learn to resolve their own issues – perhaps aided by computers, but not depending on them to fill in the blanks, cognitively-speaking. Stated in another way, they are in a hurry because of an instinctive sense that the public will 'wake up', that the trance regarding technology will end, and the era of outsized profits and unearned, prerogative-drive controls, will suddenly end. *In short, they are building Golem 2.0 for fear of Humanity 2.0*.

If you don't believe us, just ask them – not just the AI ones, but all of them:

---

[7] Carroll, L., and Gardner, M. (Ed.). 1865/1960/2015. *The annotated Alice: 150th ann. deluxe ed. Alice's adventures in wonderland & Alice through the looking-glass.* New York: W. W. Norton & Company, 11-12.

[8] Hamburger, P. 2014/2015. *Is administrative law unlawful?* Chicago: The University of Chicago Press

> *What about that glorious 'singularity' when people are empowered to make machines –*
> *computers in particular – do what they want them to do? What about that glorious time when*
> *computers become fully responsive to the desires, needs, and requirements of the people –*
> *individually, in groups and in organizations? What about that time, when people find and use the*
> *'silver bullet' that allows these things to happen, with a multitude of concurrent benefits?*

You may be surprised at the answer you will get, but we will not. Rest assured, they will only say that such would be a good thing only if they do not think you will find out a way to do it.

Benefits of Humanity 2.0

Humanity 2.0 is a very good idea. If there is a problem, it is that we have lost faith in ourselves. This can be characterized as an inability to operationalize our needs and desires, individually and in the collective. As to the individual, let us reflect on Plato's mantra of so many years ago…

> *From what point of view, then, and on what ground can we say that a man is profited by injustice*
> *or intemperance or other baseness, which will make him a worse man, even though he acquire*
> *money or power by his wickedness?*
>
> *From no point of view at all.*
>
> *What shall he profit, if his injustice be undetected and unpunished? He who is undetected only*
> *gets worse, whereas he who is detected and punished has the brutal part of his nature silenced*
> *and humanized; the gentler element in him is liberated, and his whole soul is perfected and*
> *ennobled by the acquirement of justice and temperance and wisdom, more than the body ever is*
> *by receiving gifts of beauty, strength and health, in proportion as the soul is more honorable than*
> *the body.*
>
> *Certainly, he said.*
>
> *To this nobler purpose the man of understanding will devote the energies of his life. And in the*
> *first place, he will honor studies which impress these qualities on his soul, and will disregard*
> *others?*
>
> *Clearly, he said. [9]*

Cannot we be true, if not worthy, of such a heritage?

When did he come up with such ideas? During the horrendous Peloponnesian War, when anything but the high-minded philosophies of Plato were in full play. [10]This calls to mind the work of Confucius and his contemporaries. Writing and teaching during the tumultuous times of the Warring States, their concepts were only applied – to 'multi-millennial effect' to support peace and prosperity in China and its surrounding region. [11]

As to the individual, so to the group. For this, we reach back about a hundred years. Polish-American engineer Alfred Korzybski made his point clear: Mankind must grow up! [12] This is all about language and information – the

---

[9] Plato, and Jowett, B (Tran.). 1952/1984. The republic ix, 426-427.

[10] Thucydides, and Strassler, R. B., Ed., and Crawley, R. (Tran.). 404BCE/1996. *The landmark Thucydides: A comprehensive guide to the Peloponnesian War*. New York: Free Press.

[11] Fairbank, J. K. 1992. China: *A new history.* Cambridge, MA: The Belknap Press of Harvard University Press, 51-71.

[12] Korzybski, A. 1921/1950. *Manhood of humanity*, 2nd ed. Englewood, NJ: Institute of General Semantics.

kind of conquest people need first, to be translatable to machines. Let us put our time and attention and resources to this end! This question of 'time-binding' he raised, it needs to be a foundation of achieving *Humanity 2.0*. Korzybski describes such an endeavor in an interesting way, with which we close:

> *The world affairs have seemingly come to an impasse and probably, without the help of scientists, mathematicians, and psychiatrists included, we shall not be able to solve our urgent problems soon enough to prevent a complete collapse. Now those who are professionally engaged in human affairs, economists, sociologists, politicians, bankers, priests of every kind, teachers, 'mental' hygiene workers, and psychiatrist included, do not even suspect that material and methods of great general semantic value can be found in mathematics and the exact sciences. The drawing of their attention to this fact, no matter how clumsily done at first, will stimulate further researches, produce better formulations and understanding, and ultimately create conditions where sanity will be possible.[13]*

---

[13] Korzybski, A. 1933/1995. *Science and sanity: An introduction to non-aristotelian systems and general semantics*, 5th ed. Englewood, NJ: Institute of General Semantics, 538.