



Dancing with AI

On a Three-Legged Stool

By Leopold B. Willner, PhD January 15, 2020

Abstract – *The most promising AI revolution now underway, and the pre-birth of a degree of general intelligence in machines, requires that we get it done right, really right. Herein, how to do so is analyzed using a three-legged model that focuses on balancing machine intelligence with human reasoning plus an appropriate dose of positive emotions to guide the way. In doing so, the roles of agency, realism about AI, context in action, causal reasoning and meaning are also addressed. The future of AI will evolve in a successful way if we act as good stewards of the process.*

1.0 On a Three-Legged Stool

Getting AI, and artificial general intelligence (AGI), to the point that it can work reliably and safely while also preserving what is best in human society is a major challenge. Ensuring that the rise of AI is not overly disruptive of markets, democracy, freedom, family life, and human institutions is at the heart of this challenge. The question is how to make further progress in the advancement of man-machine interplay without inadvertently undermining the present order of human affairs. And do so while preserving best practices and what already works effectively. The uncomfortable transition into AI/AGI underway will surely test the

patience, imagination and persistence of business and all of society. To do so successfully will require, in addition to ways of preserving political, social and economic institutions, carefully switching over to a world dominated by intelligent machines that cooperate with man.

This challenge suggests that future man-machine engagement, to be effective and safe, may need to function as a carefully choreographed ‘dance’ while standing on a three-legged stool, whose legs represent:

- ❖ The unique capabilities of intelligent machines and automata properly and fully utilized.
- ❖ The benefits of humans acting to bridge the gaps that may exist in machine thinking so as to compensate where machine intelligence may be lacking, flawed or limited.
- ❖ The essential value of using positive human emotions to empower and balance the man-machine model so it may reliably act in a useful manner.

This model is needed to maintain a balance between man and machines as the inevitable power of AI and AGI continues to grow. Over time machine power will continue to gain importance as an aid and a counterbalance to human perception, cognition, analysis and proposed action. It will need to do so in ways that conform with human society and human values. Here the full range of positive human emotions may come into play to help energize the joint cognitive man-machine process as it moves forward to join two cognitive streams – man and machine - to thoughtful responsible action. This three-legged stool ‘dance’ will be triumphant, but only if each of its legs is viable and coordinates with the other two in a well-balanced ensemble.

2.0 Dancing with Agency

Bringing human characteristics and attributes into play in this manner is a way of securing that AI/AGI and other automation cooperate with humans so they may succeed in practice.

In addition, as amply described by B. Berberian¹ in *Man-Machine Teaming: A Problem of Agency*, ensuring that people are able to discern when and to what degree they are truly in charge and responsible, instead of the machines they employ, is essential to success and safety.

Otherwise, as the Boeing 737 Max8 disasters have so terribly demonstrated, bad things can and will happen. After all, as with doubles in tennis, when the ball falls between the players, and neither is sure who is in charge or meant to act, bad results are likely. This underlies the need for effective 'agency' among man and machine, as a form of empowerment that is essential to success.

As explained with clarity in the Berberian paper, the role of agency in man-machine cooperation is central to any effective deployment of these technologies. This applies, whether the case in point is ordinary automation, as with a factory punch press, or AI automation as in the use of face recognition, or with a future AGI able to contemplate what-if-questions prior to advising its human master on what to do in a situation, task or challenge.

In human affairs the roles of leadership and empowerment (to act) play a key part in family life, on the job, in the arts and at play. As with a deer in the headlights, standing still in the face of eminent opportunity or danger, is not often

the best choice. Not when positive action may hold the promise of achieving a desirable result, or the deer surviving. The ability to engage requires agency to act along with the positive emotions needed to provide the intention and energy to do so. Here training and leadership along with agency play a key role, or else an opportunity may be lost or a danger not averted.

But what of job-sharing with advanced machines, is it always self-evident when the individual man or woman is in charge while working alongside a powerful intelligent machine? Much progress has been made in so called HCI and HMC, human-computer interface and human-machine cooperation in recent years. As a result, there is a good deal of carefully gathered statistical information on the subject of agency and its risk in HMC². This suggests, in part, that a person is likely to defer to a machine under a variety of circumstances – some of which may be dangerous. Even though the cyborg or robot to which the task is being deferred may be less able than a trained person on site to deal with the problem/opportunity at hand.

In his article on HMC, focused on dynamic situations in the flying business, Jean-Michel Hoc spells out the four basic types of risks of this sort in the cockpit of an airliner, in industrial process control, in anesthesiology and elsewhere that can lead to failure and at times disaster, as follows:

- ❖ Loss of Expertise - When the machine is in charge, it, not the humans, acts and controls while pilot skill, training and conditioning over time begins to fade.

¹ B. Berberian; *Man-Machine Teaming: A Problem of Agency*, IFAC (International Federation of Automatic Control), IFAC PapersOnLine 51-34 (2019) 118-123 sciencedirect.com.

² Jean-Michel Hoc; *From Human-Machine Interaction to Human-Machine Cooperation*, *Ergonomics*, 2000 43 (7), 833-843. Which contains an excellent references list.

- ❖ Complacency – Over time, the human participant who defers to the machine may have a tendency to pay too little attention. This is a most unfortunate circumstance should a problem occur that can overwhelm the intelligent machine’s guidance systems.
- ❖ Trust and Self-Confidence – Are two of the most important ‘positive human emotions’ necessary to balance the three-legged stool, or else the system may at times fail.
- ❖ Loss of Adaptability – Adaptive skill is a hallmark of human capability which even the most advanced machines do not share to the same high degree, today and for a long time to come.

3.0 A Case in Point

The crash, a few years ago, of a San Francisco bound airliner that fell short of the runway and into the waters of the Bay, is such a case in point. While the pilots deferred to the automatic landing system by accepted practice, the AI lost situational awareness and crashed the flight into the water of SF Bay before any pilot could override the automated system.

They, highly trained well-paid pilots, apparently felt they had little or no agency over the automated landing system as it engaged to land the plane, and it failed disastrously. All the while, the autopilot, an unaware unconscious automaton, knew not of its agency or indeed fully understood what it was about.

This type of failure has occurred with major automated systems in the past, as in the Three Mile Island and Chernobyl nuclear power plant disasters. Indeed, when agency is improperly

assigned such failure can be said to be designed into the system. No human-in-the-loop, with power to fix or override when necessary, implies a misguided faith that nothing will go wrong, until it does.

There is much else to appreciate about the workings of intelligent machines acting in cooperation and support of people, now and in the future. For example, humans are wired to behave in various tribal ways that go beyond simple reason, as with the herd instinct to defer to others, to a leader, or in this case a machine. This is a potentially dangerous situation not easily overcome with more training. For it requires overriding the primal emotional response of man to look around for a leader during a rapidly evolving critical situation. Such a self-caused pause or delay can prove to be very undesirable, especially when quick effective action by a human agent is required.

4.0 Realism in Dancing with AI/AGI

There is much hubris in the press about AI and its future, which some experts note is unwarranted as so-called intelligent machines are prone to unpredictability, error, bias, vulnerability to hacking and insufficient justification for results³. This may be fueled by some commercial interests promoting AI services. The result is that more than a few people, including investors, may be misled into unrealistic expectations, not to be met in the near-term decades. Even AI practitioners may at times be fooled into accepting mere conjectures about the future of AGI. All of which have a tendency to send misleading messages to the public at large.

As Brian Cantwell Smith notes in *The Promise of Artificial Intelligence*⁴- *Reckoning and Judgment*, there is at present and for the foreseeable future

³ Maryland Mitchell; *Artificial Intelligence*, Farrar, Strauss and Giroux press, 2019 pp. 120-121. For ‘meaning’ and ‘understanding’ see Part V.

⁴ Brian Cantwell Smith; *The Promise of Artificial Intelligence*, The MIT Press 2019

a wide gap between the cognitive skills of the most intelligent machines, yet built or even imagined, and the human ability to act as competent effective agents. Put simply, computational and reckoning skills alone do not and at times cannot readily lead to good judgment in real-world situations on the nature of opportunities, challenges and barriers.

To many of the leading lights in AI and AGI, including Stuart Russell of Cal Berkeley, Judah Pearl of UCLA and Brian Cantwell Smith of the U of Toronto, there is no clear path from AI to truly human class intelligence. This they note is due to the complexities of the human mind while artificial or synthetic intelligence may not naturally extend to understanding, causal reasoning, imagination, creativity, out of the box thinking, emotional behavior, human values, situational awareness, responsible agency and much else that humans do.

In the absence of scientific proof, the degree to which AGI may ultimately be limited will remain an unsettled matter for many years to come. Consequently, it would be unwise and even dangerous for people to ignore the value that human-in-the-loop approaches can contribute to man-machine cognitive systems. And may continue to do so until such a time as the limitations of machine intelligence are far better understood.

In the meantime, except for some abstract analysis, simple algorithms and such things as game playing, many activities in the real world, where time and much uncertainty are at play, will greatly benefit from human knowledge, instinct, skill and positive emotions – while dancing.

5.0 Contextual Action on the Dance Floor

⁵ Judah Pearl and Dana Mackenzie; The Book of Why, The New Science of Cause and Effect, Basic Books 2018

In the act of doing most anything, situational awareness and the context within which matters are to be understood is most critical, or else foolishness and error are likely to abound. For humans, beginning with childhood, the focus is now as it was then on being aware and in context; when listening to a conversation, driving in traffic or grasping an opportunity. Most always, when an action or a decision is to be taken, context sensitive execution of the matter at hand is a key to success. Such vital ‘consex’ is very difficult to train or program into an automaton due to the subtleties and nuances that so often underlie situations and affairs, and the need to understand consequences.

Context sensitive execution, as a guide to effective action is required to do things right, effectively and safely in a real world that abounds in risk, uncertainty and partial information. It is here again that the role of positive emotions is essential to man as the force behind movement and as the guide to success. Without this leg, the three-legged stool will not stand. Without it, the man-machine cooperative is always at risk of lack of direction and of failure. Without context the stool will fall.

5.1 Causal Reasoning a Key to Dancing

One of the most useful and powerful tools that the human mind is designed to readily handle is the ability to engage in causal analysis and speculation⁵. Without this capability, imagination is ineffective, as out-of-the-box queries cannot be readily answered, while prior statistical and factual data may not suffice to envision what may happen, what is possible, and what may arise out of creativity. Yet it is the ability to ask the right questions, and to imagine the result, which is at the heart of productive

thinking, as noted by none other than Albert Einstein.

Thus, speculating on what may happen in this and that situation, while pondering what-if-questions on alternatives, is of importance when deciding what to do and how to do it. To engage in this process a thinking machine must possess causal reasoning, a high bar indeed for a thinking machine, as this form of thinking goes well beyond what is attainable out of statistical analysis and big data learning, the pillars upon which the machine was raised.

It is not in the remit of Dancing with AI to delve into causal reasoning, except to note that any machine whose power rests on training and machine learning, say by way of deep convoluted neural networks, is unlikely to be equipped to engage causal reasoning on uncertain matters, issues and situations yet to come. All the while, it is this gift of the human mind, out of its historic need to survive in a dangerous and complex world, that is so very valuable to stabilizing the process of dancing on our three-legged stool; and make the rise of a better man-machine world more likely and effective.

6.0 The Dancing Lesson Summarized

In this article we suggest, as does the futurist Ray Kurzweil and many others, that human emotions must be brought into play to make the man-machine world of the future work successfully. To do so, machine intelligence and human intelligence working in conjunction must be supplemented, inspired and driven by positive human emotions. Nothing less will do.

In this endeavor, the role of agency, who is in charge between man and machine, must be kept command and control clear at all times. Realism about what AGI is likely to achieve in the next generation or two demands that human-in-the-loop methods not be overlooked for reason of expediency or simplicity; as the consequences of

such shortsightedness may undermine human confidence in AGI as it progresses and at times fails.

Getting things to work out right in man-machine play requires effective consex, or context sensitive execution of all tasks and actions, carefully and consistently executed.

Then, to look forward and plan ahead, computer scientists, engineers and social scientists should take a page out of the gospel of business leaders and learn not to rely too heavily on past data. Instead, they should learn to apply causal reasoning so they may ponder what if questions whose results let them spy on the future, with its opportunities and challenges.

Our Dancing with AI article offers an approach to adding a highly useful level of ‘meaning’ and ‘understanding’ to machine cognition and AI solutions during the upcoming years. As such, it provides a proposed way to begin to impart a degree of meaning and understanding to machine results. Such an addition, if successful, is understood by leading AI researchers to be helpful, even necessary, to the further empowerment and rise of true artificial general intelligence, AGI. For more on this subject, see the M. Mitchell footnote.

Additional references on request

Leopold B. Willner, PhD 1/15/2020
President, Dual Stream Technology, Inc.



From AI to Intelligent Behavior