

Jan 25, 2021, 12:48pm EST | 8 views

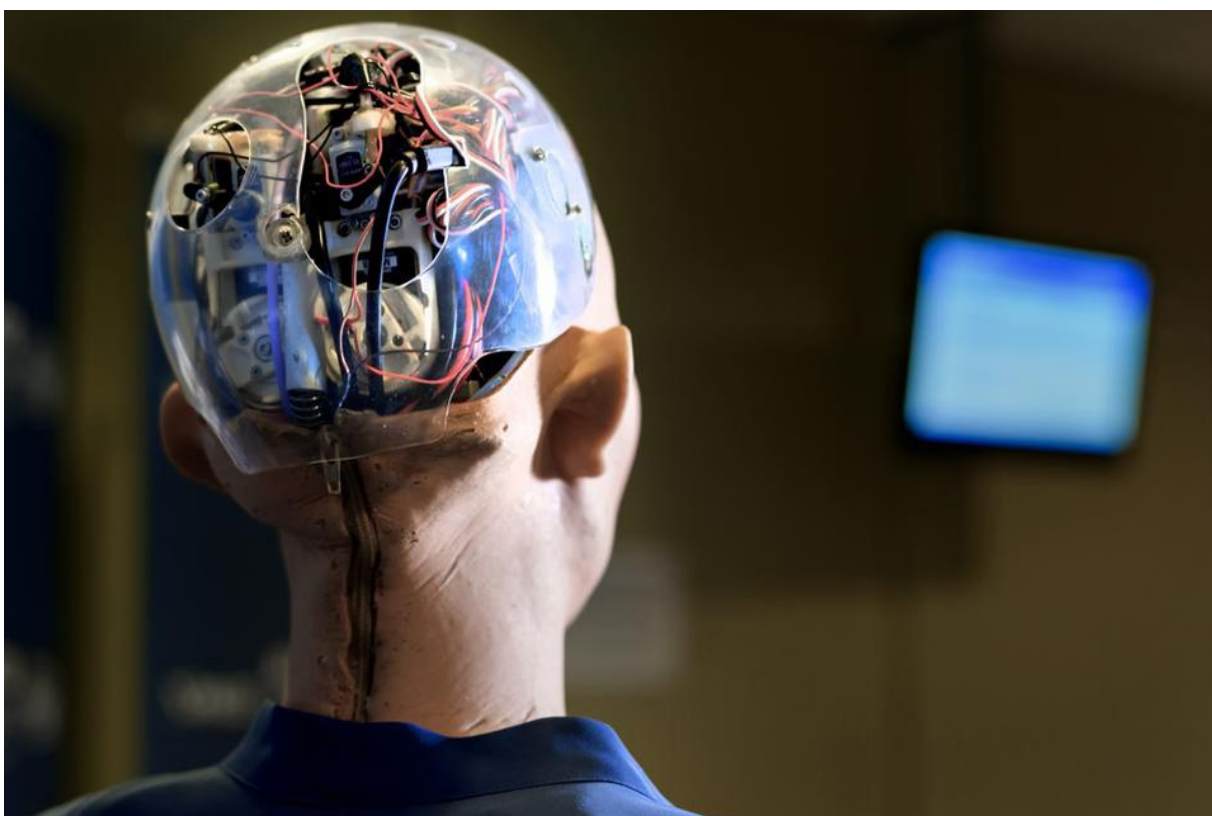
Can We Trust AI? When AI Asks For Human Help (Part One)



Charles Towers-Clark Contributor ⓘ ⊕

[AI](#)

I write about digital transformation, AI, self-management & data in the digital age



Making AI more 'humble' could not only help improve AI decision making, but could also help inspire ...

[+] AFP VIA GETTY IMAGES

AI is notoriously difficult to explain, and some deep learning algorithms can be too complex for even their creators to understand their reasoning. This makes it hard to trust what AI is doing, and even harder to find mistakes before it's too late. Having an algorithm stop partway through its reasoning to check with a human-in-the-loop could inspire more trust in AI, and open the door for the technology to be used in more sensitive and mission-critical

applications. Injecting some ‘humility’ into AI in this way could not only make AI more trustworthy and change how companies think about AI, but it could also help to demystify AI and reveal it as the logical and reliable technology that it is.

How to trust the process

According to enterprise AI company DataRobot, making AI more ‘humble’ is key to inspiring confidence in the technology, and can help to show how AI works in real time. “Think of [Humble AI] like a guard rail,” says Ted Kwartler, VP, Trusted AI at DataRobot, “if the algorithm is unsure of a prediction or is coming up with answers that are wildly outside of expectations, it can get a human involved to apply that contextual knowledge.” This kind of check and balance stops an AI from ‘wildly’ following its own reasoning, and makes sure that bias and faulty logic can be corrected before getting too far along in the process.

There are, of course, instances when AI is perfectly fine to make a decision on its own - after all, that is the main purpose of a properly functioning AI. Different levels of decision-making require different levels of trust in what an algorithm is doing. These differences can be thought of in terms of ‘fast’ decisions, “like serving a banner ad” and ‘slow’ decisions, “such as in a medical context,” explains Kwartler: “If an algorithm serves you the wrong banner ad, it’s not the end of the world... but if an algorithm says that the tumor on that scan is cancer, we want a human in the loop to check that rather than just accepting that decision, because it’s really about quality of life in that scenario.” Applying this ‘contextual knowledge’ as Kwartler puts it is essential in ‘slow’ decisions because they have direct consequences on a human’s life or livelihood. The true test for humble AI is for it to be trusted in these situations because it will always alert a human rather than continuing a potentially faulty line of reasoning.

Learning to accept limitations

This humility layer is important because of the different ways that algorithms learn and make decisions as they progress. An algorithm first learns through inductive learning (“learning patterns through groups of data”), then makes predictions through deductive learning (“given these inputs it will always be this output”), and uses abductive learning when in production. “Abductive learning follows the logic that all cats have fur, and that animal has fur, therefore that animal has to be a cat,” says Kwartler, and it is this kind of learning that can cause problems if left unchecked. “That kind of generalization is what we are trying to protect against [with humble AI],” says Kwartler. If an algorithm can be shown to avoid these logical leaps then it can also be trusted to avoid drastic consequences by checking that its reasoning makes sense in the real world. This also acts as the final filter against bias, because an algorithm is stopped from bringing biased or faulty reasoning into the final stages of decision-making, although Kwartler points out that it is better to check for bias at the start of the process “as the algorithm is being built.”

Assessing the situation

Being conscious of the flaws of AI can help to avoid dangerous mistakes that come with having blind faith in AI’s abilities. This kind of blind faith on a large scale also causes distrust in those using AI. Furthermore, a bit more humility in the boardroom could well improve trust in AI and technology overall. “Several of the world’s largest companies have all had missteps with their AI... and that is partly due to the lack of an AI impact assessment. When you build a factory you need an environmental impact assessment, so why not [have one] for automation?” argues Kwartler. An impact assessment for AI could be as simple as questioning whether the aim of an algorithm is to limit harm or to actively do good, or it could explore all possible outcomes of an algorithm’s decisions. “That [would] build trust at an organizational level,” says Kwartler, “it shows that [a company has] proactively thought about how their algorithm impacts the world.”

MORE FOR YOU

Is Contact Tracing A Public Or Private Concern?

Do Algorithms Dream Of Electric Speech? What Virtual Philip K. Dick Can Teach Us About AI

Is Facebook Doing Enough To Stop Racial Bias In AI?

Trust goes hand in hand with transparency, especially on an organizational level, but Kwartler is careful to point out the reasoning behind being transparent to make sure that it is not simply a case of opening the black box of AI. “Being handed the source code [of an algorithm] is not going to be useful or give the transparency you want,” says Kwartler, “it’s more important from a societal standpoint to be able to rationalize that decision.”

Humility from the top down

As AI gains more influence over the lives of ordinary people, humility in AI and those developing AI is needed to show that the technology will indeed be kept in check as it continues to advance: if there will always be that ‘guard rail’ to make sure that the consequences of an AI’s decision make sense in the real world, then the idea that we are in control of AI’s decisions will be easier to comprehend for non AI experts.

Trust is not an easy thing to inspire in people, especially on a wide scale and with such a transformative technology as AI. Blind faith in AI’s ability to do good has already led to serious mistakes, and a more logical, measured, and humble approach to such a powerful technology can only help AI to do more good in the world, and help humanity further as a result.

Part two of this article will look at how making AI more trustworthy needs to come from the top, and explore the wider impact of AI and trust in our society moving forward.

Follow me on [Twitter](#) or [LinkedIn](#). Check out my [website](#) or some of my other work [here](#).



Charles Towers-Clark

I have been working in the M2M, IoT, and data space since founding Pod Group (a provider of IoT connectivity & billing software) in 1999, and have become greatly...

Read More

Reprints & Permissions

ADVERTISEMENT
