System Safety in Healthcare

by Dev Raheja

Application of AI on System Safety

The purpose of system safety is to prevent bad things from happening. Lately, systems have been using artificial intelligence (AI), and verification and validation is becoming a major challenge. First, we have to understand the four main types of AI [Ref. 1], in addition to much-less-used methods on the web. The main types of AI are:

- Reactive Machines
- Limited Memory
- Theory of Mind
- Self-aware

Reactive Machines

Reactive Machines perform basic operations. This level of AI is the simplest, reacting to an input with an output — there is no learning that occurs. This is the first stage of any AI system. A machine-learning system that takes a human face as input and outputs a box around the face to identify it as a face is a simple, reactive machine. The model stores no inputs, and it performs no learning.

Limited Memory

Limited Memory types refer to AI's ability to store previous data and/or predictions, using that data to make better predictions. With Limited Memory, machine-learning architecture becomes a little more complex. Every machine-learning model requires limited memory to be created, but the model can be deployed as a reactive machine type.

Theory of Mind

We have yet to reach Theory of Mind AI types. These are only in their beginning phases and can be seen in technologies like self-driving cars. In this type of AI, *it begins to interact with the thoughts and emotions of humans.*

Presently, machine-learning models help people achieve tasks. Current models have a one-way relationship with AI. Alexa and Siri bow to every command. If you angrily yell at Google Maps to take you in another direction, it does not offer emotional support and say, "This is the fastest direction. Who may I call and inform that you will be late?" Google Maps, instead, continues to return the same traffic reports and ETAs that it has already shown, with no concern for your distress.

A Theory of Mind AI will be a better companion.

Fields of study tackling this issue include Artificial Emotional Intelligence and developments in the theory of Decision-Making. Michael Jordan presented some of his decision-making research at the May 13, 2020 event, The Future of ML and AI with Michael Jordan and Ion Stoica, and more coverage was presented at the ICLR 2020 conference.

Self-Aware

Finally, in some distant future, perhaps AI achieves nirvana and becomes self-aware. For now, this kind of AI exists only in stories, and, as stories often do, instills immense amounts of both hope and fear into audiences. A self-aware intelligence beyond the human has an independent intelligence, and likely, people will have to negotiate terms with the entity they created. What happens, good or bad, is anyone's guess.

In his blog, Jonathan Johnson offers detailed information on these four types of AI [Ref. 2].

AI and Reliability

System reliability is an important part of system safety, writes Dr. Klaus M. Blache of the University of Tennessee Reliability & Maintainability Center (RMC). If there is no reliability, he opines, there is no system safety.

"Interest in Artificial Intelligence (AI) has been gaining speed due to big data, the cloud, increased

> “Finally, in some distant future, perhaps AI achieves nirvana and becomes self-aware. For now, this kind of AI exists only in stories, and, as stories often do, instills immense amounts of both hope and fear into audiences. A self-aware intelligence beyond the human has an independent intelligence, and likely, people will have to negotiate terms with the entity they created. What happens, good or bad, is anyone’s guess.”

computing power, greater connectivity, and advancements in sensors and signal processing,” he writes. “Machine perception, i.e., using cameras to recognize objects, has been around for years. Today’s machine-learning systems can do much more. Think of enhanced speech and facial-recognition technology, Tesla autonomous vehicles, IBM Watson, and reliability and predictive maintenance modeling efforts.” [Ref. 3]

Assessing the Reliability of AI Programs

“Reliability assessments of AI programs must consider not only possible program bugs which remain in the program due to insufficient testing and debugging, but also faults due to intrinsic characteristics of AI programs that cannot be removed even after the program is fully debugged,” write Bastani Farokh and Chen Ing-Ray in their paper “Assessment of the Reliability of AI Programs” [Ref. 4] They go on to describe the development of an analytical tool for assessing the reliability of AI programs, where possible intrinsic faults of AI programs are identified and modifications to existing software reliability models for conventional programs are suggested.

References

1. Simplilearn. “Artificial Intelligence Tutorial | AI Tutorial for Beginners | Artificial Intelligence | Simplilearn,” https://www.youtube.com/watch?v=FWOZmmIUqHg.
2. Johnson, Jonathan. “4 Types of Artificial Intelligence,” *Machine Learning & Big Data Blog,* https://www.bmc.com/blogs/artificial-intelligence-types.
3. Blache, Klaus M. “AI and Reliability: How Much, How Fast?” *Efficient Plant,* November 14, 2017, https://www.efficientplantmag.com/2017/11/ai-reliability-much-fast/.
4. Bastani Farokh, Chen Ing-Ray. “Assessment Of The Reliability Of AI Programs,” *International Journal on Artificial Intelligence Tools,* Vol. 2, No. 2, https://www.worldscientific.com/doi/epdf/10.1142/S0218213093000138.