

Bolstering AI by tapping human testers

September 25 2020, by Peter Grad



Credit: CC0 Public Domain

Advances in artificial intelligence depend on continual testing of massive amounts of data. This benchmark testing allows researchers to determine how "intelligent" AI is, spot weaknesses and then develop stronger, smarter models.



The process, however, is time-consuming. When an AI system tackles a series of computer-generated tasks and eventually reaches <u>peak</u> <u>performance</u>, researchers must go back to the drawing board and design newer, more complex projects to further bolster AI's performance.

Facebook announced this week it has found a better tool to undertake this task—people. In order to create better and more flexible AI, it built Dynabench, a platform that utilizes human and computer models to collect data and benchmark AI.

It relies on a procedure called dynamic adversarial data collection and, as a Facebook <u>white paper</u> posted Thursday explains, it "radically rethinks AI benchmarking."

By conversing with natural language processing models, humans attempt to trip up the program by using linguistically challenging questions. The program may trip up over challenging vocabulary or idioms, or it may misinterpret sarcasm. The more challenging the human questions, the more AI learns to navigate tricky terrain.

"It measures how easily AI systems are fooled by humans, which is a better indicator of a <u>model</u>'s quality than current static benchmarks provide," Facebook explains. "Ultimately, this metric will better reflect the performance of AI models in the circumstances that matter most: when interacting with people, who behave and react in complex, changing ways that can't be reflected in a fixed set of data points."

In fact, recent research has found that traditional benchmark tests are not reliable, finding that up to two-thirds of answers provided in natural language learning models were actually unwittingly embedded in the tests and allowed the models to merely memorize the answers.

Facebook researcher Douwe Kiela says reliance on faulty benchmarks



stunts AI growth.

"You end up with a system that is better at the test than humans are but not better at the overall task," Kiela says. "It's very deceiving, because it makes it look like we're much further than we actually are."

As the Facebook white paper points out, the Dynabench metric "will better reflect the performance of AI models in the circumstances that matter most: when interacting with people, who behave and react in complex, changing ways that can't be reflected in a fixed set of data points."

An AI researcher at the University of Washington emphasized that current <u>benchmark</u> tests of AI are distorted due to the ability of machine learning to masterfully detect dataset correlation imperceptible to humans: the machines correctly answer the question but don't have the requisite "understanding" of meaning.

Yejin Choi says, "We are seeing a Clever Hans situation." She was referring to the 1907 revelation that a horse could perform mathematical tasks. In fact, a psychologist discovered that the horse was responding to bodily cues from the trainer that tipped the animal off to the appropriate responses. Most interesting, the psychologist learned that the trainer, in fact, was unaware of his involuntary cues being read by the worse. The scenario has come to be known as the observer-expectancy effect, or the Clever Hans effect.

Likewise, Dynabench wants to ensure that AI is not merely responding to unintentional cues.

The public is invited to participate in the Dynabench project by conversing with its <u>natural language</u> processing models at <u>dynabench.org</u>.



"We want to convince the AI community that there's a better way to measure progress," Kiela says. "Hopefully, it will result in faster progress and a better understanding of why machine-learning models still fail."

More information: <u>ai.facebook.com/blog/dynabench ... king-ai-</u> <u>benchmarking</u>

dynabench.org/

© 2020 Science X Network

Citation: Bolstering AI by tapping human testers (2020, September 25) retrieved 3 May 2024 from <u>https://techxplore.com/news/2020-09-bolstering-ai-human-testers.html</u>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.