The Last Byte

Are You Sure You Love That Store?

Scott Davidson

■ **I'VE WORKED IN** test and reliability for more than 35 years, so when I look at a piece of hardware or software my first thought is how it can break. My model of the installed base of chips involves them failing left and right, though I know from the data and my own experience that they are very reliable.

This is due to the psychological principle of availability, which makes you think the probability of something happening is correlated to the amount of information you receive about it. That's why people used to (and maybe still do) think air travel was more dangerous than car travel. Airplane crashes got big headlines; car crashes were buried in the back of the newspaper if covered at all.

Until I read the abstracts of the articles in this issue of *Design&Test* though, I did not think about how machine learning systems could break. I knew that they might be incorrect in learning, and I knew that no one understands how they make decisions. Nevertheless, I never thought about them breaking. It seems that there are reliability issues and security issues too.

Now I'm even more nervous.

Sometime soon we'll have real personal digital assistants (PDAs). By this, I mean one which adds appointments when it reads an email about one or hears someone set one up. It will be location sensitive enough to tell you that you've made a wrong turn.

Such an assistant will incorporate machine learning. But what happens if the neural network is found to be unreliable?

We can't just throw it out and buy another, since it has accumulated knowledge of your habits and wants. Can we save enough of the learning to be able to back it up? If we can, how do we know the latency of the error? Is the advice of your assistant what you really want? Is that restaurant it recommends one you like? Did it really make that vital appointment? Is the gift it tells you to buy one your partner likes?

If reliability failures made the system fail as soon as they occur, this wouldn't be a problem. But how wrong can learning be under defects? Has this been studied?

We should be just as worried about security breaches. One thing a real PDA would learn is your preferences, such as brands and restaurants. What if an unscrupulous merchant hacked into the machine learning logic and modified it to say that their business was just wonderful. It might tell you that you found a certain restaurant the best, and you should eat there again.

Nonsense, you say. You'd remember the restaurant, or you would be convinced you never ate there. Are you sure? After we start to depend on our PDAs? How many times have you put faith in your GPS, following its directions though you have no idea where you are? Say a billboard owner hacked the GPS to lead you past his billboard? Would you ever catch on?

The articles in this issue don't deal with these specific issues, but they should open your eyes—as they opened mine—to the problem. Anyone who has lived through the past few years should not deny the inevitability of this concern. If you build it, they will hack it. ■

Digital Object Identifier 10.1109/MDAT.2020.2971651
Date of current version: 20 April 2020.

■ Direct questions and comments about this department to Scott Davidson; davidson.scott687@gmail.com.