COSMOS
Navigate the Compliance Universe

ethikos Volume 34, Number 1. January 01, 2020
Can artificial intelligence operate within compliance guidelines?

G. Elaine Wood, CCEP, and Alan Brill

G. Elaine Wood (elaine.wood@duffandphelps.com) is a Managing Director at Duff & Phelps,and a former federal prosecutor focusing on compliance and risk assessment. **Alan Brill** (alan.brill@duffandphelps.com) is a Senior Managing Director at Kroll, a Division of Duff & Phelps, and founder of the firm's cyber risk practice. He is also an Adjunct Professor at Texas A&M University School of Law. Elaine Wood and Alan Brill are both headquartered in New York City.

Traditionally, computer systems follow a set of rules defined by their programming and operate according to pre-established guidelines. Put another way, what they did yesterday, they will do today, and what they do today is what they will do tomorrow. This consistency is one of the basic tenets of compliance testing. But what happens if a system's programming is designed to evolve, so that the system itself can change the rules by which it operates? The same transaction processed today might have a different result than if it were processed yesterday, and yet another result if processed tomorrow.

That's the nature of deep learning artificial intelligence (AI) systems. An AI system examines characteristics of transactions and uses them to make changes in the way the system processes data. Consider the example of an AI system built for a bank that is designed to make decisions on applications for personal loans. The bank feels that having the decisions made by a single automated system will protect against claims that individual loan officers are acting in a discriminatory way or otherwise treating loan applicants unequally.

But is an AI system free from bias?

In a traditional system, a computer would look at a loan application, apply weights to the information provided, and spit out a decision. An AI version is different—the system is constantly reviewing outcomes of its prior decisions (here, whether loans approved by the system are being repaid) and modifying the weights given to different parameters to improve on the repayment rate. In fact, AI systems are often "trained" by providing them with large numbers of cases and outcomes, so this system might have been trained with 100,000 prior approved loans and each loan's repayment history.

Let's say that the bank's system determines through experience that there is a connection between the probability of successful loan repayment and the postal code of the applicant. Because of statistically significant differences based on zip codes, the AI system might well increase the weight given to the "applicant home zip code" field in the decision-making process going forward. As a result, two loan applications that are very similar in content, except for where the applicant lives, could have different outcomes. Wrong zip code, no loan.

Questions to consider in the design of the AI system are whether to impose limits that may be apparent to the compliance officer but not necessarily understood by the systems technologist designing the AI deep learning program.

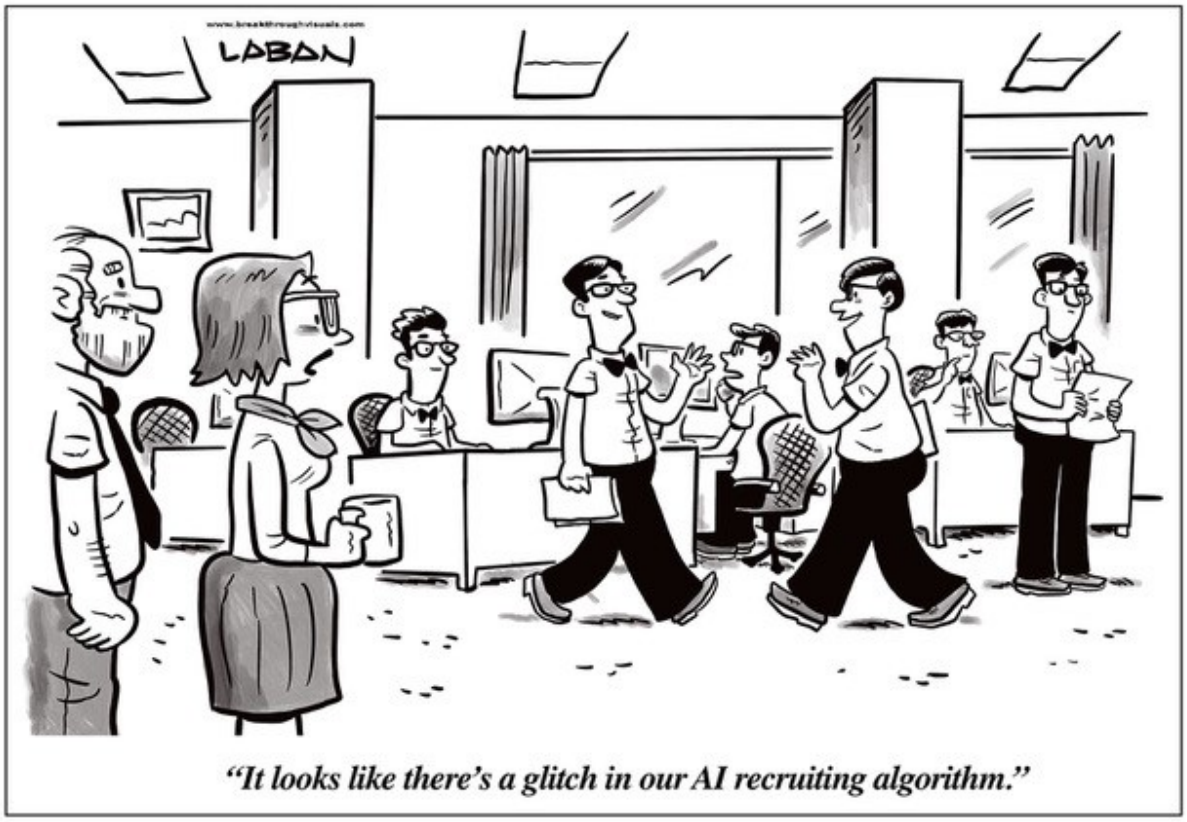
Changing the weight given to residential location would likely be recognized by both loan officers and compliance officers as a potential problem. "Redlining", according to a recent article in ThoughtCo, refers to when "banks and other institutions refuse to offer mortgages or offer worse rates to customers in certain neighborhoods based

on their racial and ethnic composition." This definition seems to require knowledge of the racial and ethnic composition of an area, but the result is key. If a bank says, "We don't care who lives there. We aren't going to make any loans there because of the high default rate," it seems naïve to assume that a regulator will accept that the institution didn't know (or reasonably should have known) that such a policy would have an adverse effect on certain groups. An argument that the bank's decision not to offer loans on an equal basis across neighborhoods is unbiased, because it was made by an AI system that is objective, is likely to fall flat.[1]

The reality is that all systems, even those employing AI and deep learning technologies, exist within a real-world set of legal and regulatory obligations. Discriminating against potential customers based on where they live is illegal.[2] To argue otherwise would be seen as supporting—overtly or otherwise—racial or ethnic bias.

So how do you prevent an AI deep learning system from veering off course and making decisions that a compliance officer would recognize as being illegal or outside the range of outcomes that would be acceptable to management? The answer, of course, is that a recognition of those limits would have to be built into the decision-making rules of the system. Without these "fences," the risk of an AI system making unacceptable decisions is real.

Amazon found out the hard way when it discovered in 2015 that its internal AI hiring tool (used to rank job candidates) had developed a bias against women. Its computer models were trained to compare applicants to patterns observed in resumes previously submitted to the company—mostly from men, because of the scarcity of women in the tech industry. In effect, the AI system taught itself that male candidates were preferable. The system had to be scrapped, as reported by Reuters.[3]



This cartoon initially appeared on ReWork, *a blog published by Cornerstone.*[4] *Used here with permission.*