

TEXAS
ACG



19TH ANNUAL

CAPITAL CONNECTION

MARCH 7-8, 2023 | MARRIOTT MARQUIS HOUSTON

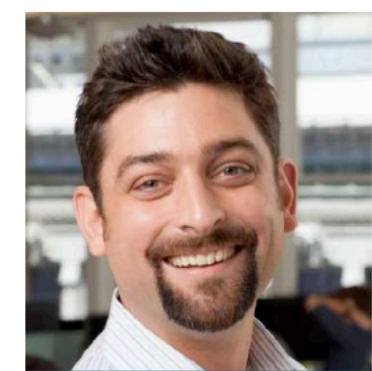
REGISTER HERE



TEXAS
BUSINESS JOURNALS

WORRYING ABOUT AI RUN AMOK? IT NEEDS A HUMAN IN THE LOOP

Jonathan Dodson, Software Engineering,
Data Science & Operations, SourceScrub



Jonathan Dodson,
Software Engineering,
Data Science & Operations,
SourceScrub

Worrying about AI Run Amok? It Needs a Human in the Loop

Scientists already are fretting about the prospect that out-of-control artificial intelligence (AI) systems could end up wiping out humankind.

Let's pause for a reality check. AI today has problems, but they're much less existential than the kind of future catastrophe that these scholars are brooding about. The immediate problem is that AI models that fail to incorporate reasoning and logic from human beings simply aren't that smart.

When it comes to building a software program that includes exactly the kind of data that private equity investors and other dealmakers yearn for as they develop their acquisition target lists, the ideal outcome is the result of the best that both AI and humans bring to the table.

That's why we have combined human logic with artificial intelligence and machine learning processes in building SourceScrub's software. Our goal is to provide our clients with a comprehensive view of the universe of founder-owned, investable companies, and deliver more data and insights on those businesses. The machine learning process, as you can see illustrated below, is able to create a kind of virtuous cycle. The process starts by defining a business problem (in the case of SourceScrub, the need to obtain and properly classify data that provides insight into these bootstrapped companies) and describes what successfully solving it might look like. Machine learning helps to refine the model that's created as we pursue those goals – and constantly refines that model as it 'learns' from what was right or wrong about earlier versions, until it delivers the best possible outcome. Feedback loops permit us to constantly improve and fine-tune the model.

But the bootstrapped, founder-led companies that make

up SourceScrub's database don't exist in a static universe. They're subject to economic and business trends, innovation, world events and many more unpredictable occurrences. If models aren't retrained to 'learn' from current, real-world data, unprecedented or anomalous events can cause predictions to become flawed.

Deployment is Only a Milestone

The big risk of having a ML-based artificial intelligence model running without human in loop supervision, is that the original model prediction will begin to drift, becoming biased and outdated. For instance, a model that only knows what happened in the past won't be able to update itself in a timely manner, and without getting the right new training data, won't update itself to reflect a new reality. A case in point: the COVID-19 pandemic dramatically reshaped the mergers and acquisitions landscape in ways that even the best model based solely on AI/ML trained prior to the event couldn't have imagined.

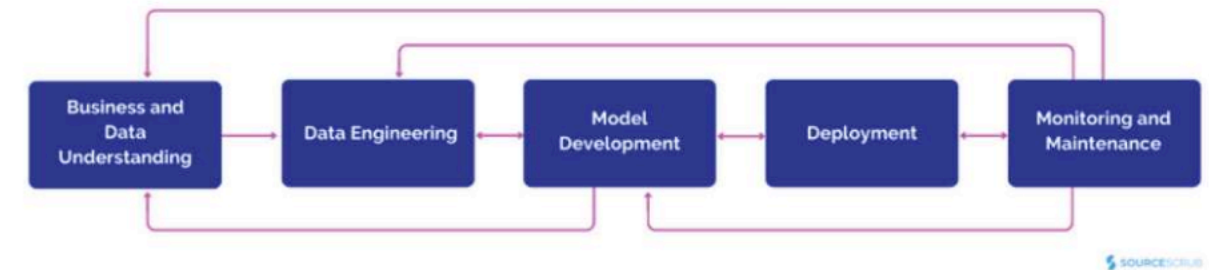
Since any AI-based machine-learning system is only as good as the data we train it with, over time the model will become distorted and degraded. The results can become irrelevant or even misleading.

Adding a Human to the Loop

Blending the best that both machine learning and human logic and intellect bring to the party addresses this problem. At SourceScrub, we embrace a semi-supervised machine learning strategy to constantly re-train our AI-based models with the input of trained experts at specific points in the loop.

Humans enter the loop once a model has been created and deployed to prevent that model from degrading in the

Machine Learning Lifecycle



first place. That happens when human beings update the dataset that the machine-learning process draws on as it retrains itself and refines the model. Our skilled team members also can spot any degradation that does occur, and take action to figure out what might have changed that isn't captured by the data on which the model relies. We're constantly monitoring a model's output and trying to identify ways to help it do even better.

But our approach goes further than that. "Human in the Loop" allows us to look at specific data sets in which the model has only a low confidence level, and assess and label that as needed. And our team also can validate the model's predictions and monitor how the model reacts to new labels and different data sets.

If we asked human beings to undertake the kind of challenges involved in building and maintaining the kind of vast database that SourceScrub now offers its clients without the assistance of AI, even a workforce 10 times the size our current team would struggle to keep up. But if we left it all up to the machines and didn't have our skilled team supporting the loop, results down the road would get messy.

Relying on AI tools alone is a trap. Only by adding humans to the loop, with their ability to provide continuous feedback about a model's strengths and weaknesses, is it possible to deliver a truly effective ML-based model that delivers exactly what dealmakers need.