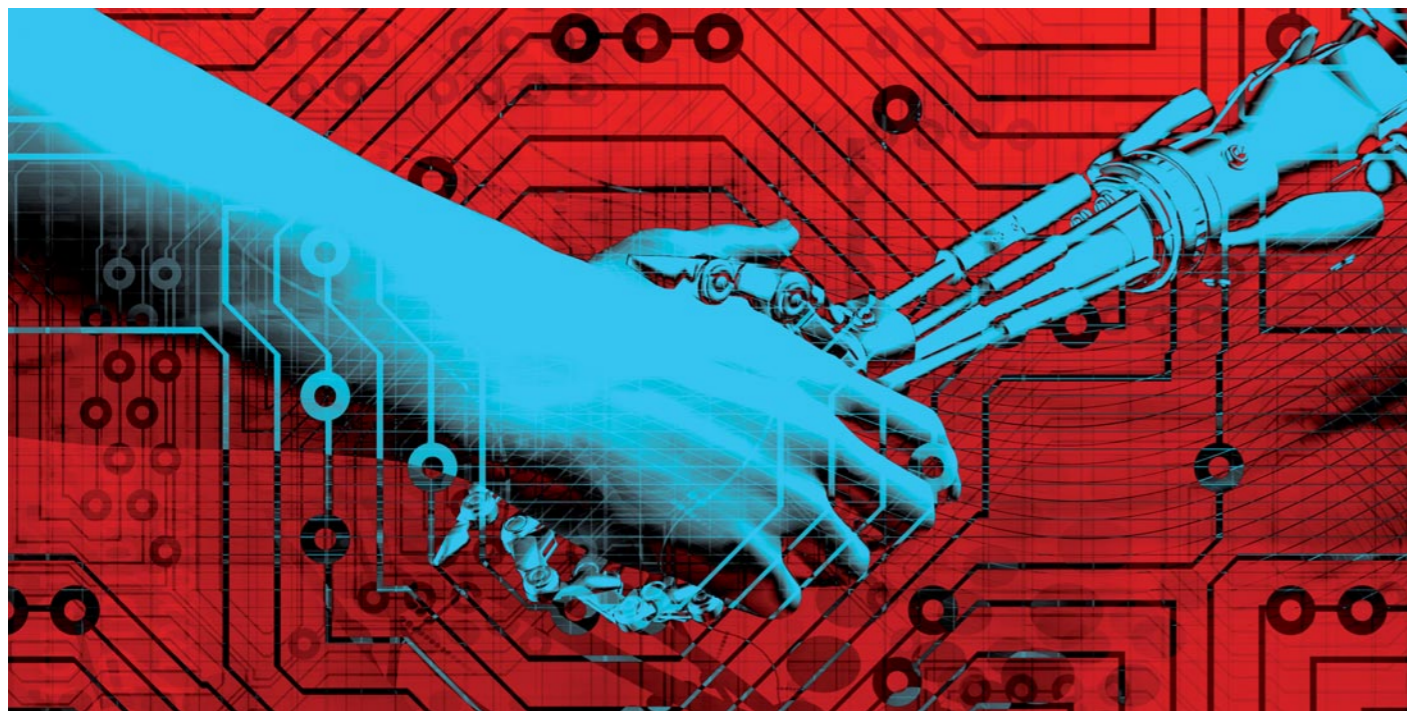BY FATIMA ABAT
AND JOOST VAN GIJN

AI in an operational environment, how human and machine work together

HUMANS AND AI NEED EACH OTHER

This was the unexpected conclusion of several years of data science work within the insurance industry, that led to the development of our AI deployment strategy via a digital assistant that helps operational teams to improve. This article explores some of the reasons why people and AI models complement each other well and some of the challenges that come with implementing AI in an operational environment.

F. Babat MSc. (left) and
J. van Gijn MSc are both Data
Scientist at Experience Data.

MODELLING IN AN OPERATIONAL ENVIRONMENT

Like all models, AI models represent only part of the real world. And while models can be trained on ready-made, cleaned and fully understood datasets, performing in an operational environment brings its own challenges. Missing data, untimely data or sometimes even faulty data can easily cause a model to underperform or misinterpret data. And even with a robust data pipeline, quality monitoring and logging, outliers still occur. Commonly it is then up to the data scientist to come up with a solution to prevent risky predictions or classifications from reaching the operational team.

But in practice, people within the operation will quickly spot flukes, or outliers that are unlikely and promptly incite further investigation. It may therefore be much easier (and less costly) to have a person second guess model output, rather than spend months optimizing a model and trying to circumvent every possible problem that the model may run into. It is then the feedback of all the people using the model predictions that gives direction and significance to the data scientist on how to make the pipeline and model more robust for better overall performance.

INFORMATION OVERLOAD

People can take in only so much information to base their decisions on. More likely the constant feed of information leads to a general intuition in the decision making, which we call experience. In itself, gaining experience is a great mechanism for learning, but in an operational setting with multiple people doing similar tasks this has its flaws. Besides any personal biases that a person may have, not everyone has the same level of experience. And those experiences are likely built upon different subsets of cases, creating a unique view on case handling.

This is of course where models shine. The ability to ingest millions of rows of information within minutes and translate that information into a model, is something that humans cannot compete with in terms of speed and completeness. In that sense, a model is perhaps the most senior person within the team, having seen all previous examples and learned from them all.

CONTEXT, REAL CONTEXT

Models are only as good as the data they are trained on, is a common phrase used in data science. And for most models this is certainly true. Interestingly enough, this goes for people as well. People experience the world with their many senses every day, which results in a very rich information stream. Besides the technically stored information a model would use, people ingest real context information. Tone of voice, formulation and word choice in a letter, other worldly impactful situations (like Covid recently) affecting a situation, are all examples of information that may be relevant to a decision, but are rarely provided to a model during training. In that sense (no pun intended), a person's experience is built upon much more than just bits of data fed into a model. This contextual experience can be important input into the decision-making process, be it for good or bad. By keeping the human as part of the decision making process, we use this information, together with the power of AI.

OPERATIONS ARE A MOVIE, REGARDLESS OF WHICH PICTURE WE CHOOSE TO TRAIN OUR MODELS ON

When analyzing data, in preparation of modelling a process, data scientists prefer to use a set dataset. A single data dump or subset query of cases, that have run its course and thus, contains all the elements of the full process. This is done to reduce complexity and gain a faster understanding of the process. For most models (not you, reinforcement models), they then take the same set and use it for training and validation of the model. During implementation it is then discovered the hard way that an operational production environment is not as forgiving. Missing critical datapoints can reduce predictive quality to rubbish. When this occurs, the model immediately hits its limits, but thankfully there is someone who can help. In such cases the digital assistant asks the person using it for help in finding out the values for the critically missing datapoints and to enter those into the system. In itself, the request for information isn't that interesting, but it does make the codependency between human and machine visible on an operational level.

BUT WHAT ABOUT ETHICS?

A machine learning model is susceptible to bias. When the training data used is incomplete or prejudicial the resulting model is affected by this and the bias becomes a systemic implementation. Especially in cases where the predicted output is based on previous human-made decisions. Having a human in the loop of the training and the monitoring of the model improves the chances of a model's biases not going unnoticed. Besides the training data, the machine learning technique used to model the process has a big influence on the transparency of the choices being made by AI. While complex models using thousands or millions of parameters in non-linear networks can result in great performance, their black box characteristics make their decisioning impossible for human interpretation. In conclusion, for ethical machine learning modelling, it is imperative that the data

scientist fully understands both the technical and the functional (business impact and influence) sides of the model.

Perhaps the most important ethical step gets taken before a data scientist starts modelling. How, when and where to use a model in an operational setting decides largely how big of an impact the model is allowed to have. For some cases, like routing emails to the proper recipient, may be allowed to act without case by case human supervision. In other cases, decisions that directly affect clients may require case by case human supervision. In that case there is a clear distinction whether a model acts as the deciding or an advisory role. So before asking if a AI model could support you, think about the how and in what role you want the model to act.

INTRODUCING DATA SCIENCE TO AN OPERATIONAL TEAM AND VICE VERSA

A good data scientist is aware of how his model performs and knows how to measure this performance and how to test it well. You could say that at some point a data scientist trusts his model. This trust does not come naturally for someone working in an operational team, who probably only interacted with models unknowingly (who doesn't use a search engine, the weather forecast or a phone). This is especially the case when such a model is supposed to advise him or her on something she does every day. At the same time, the model or data scientist needs experience and to learn from the actual process by getting constant feedback from the person using it. There are two key ingredients to make this successful. The first is a practical and user-friendly interface to facilitate the feedback loop and smooth interaction between human and machine. A digital assistant provides a more familiar feel to a model and triggers a more natural response than interacting with a 'model' or model predictions. The second important thing is learning by doing. Starting small and with a dedicated group of people the interactions can be shaped and improved so that there is a better fit between the model, its interactions and the persons using it. This also builds trust. These early adopters can be great ambassadors later on and are vital to solid adoption. After all, data scientists can't do everything.

CONCLUSION

Humans and AI need each other when operationalizing AI. True collaboration, making use of both human and AI strengths makes a huge difference in the adoption off models in the operational environment. In the end it is not the best model, but the best used model that wins.

"Learning by doing" ∎