An interactive platform that explains machine learning models to its users
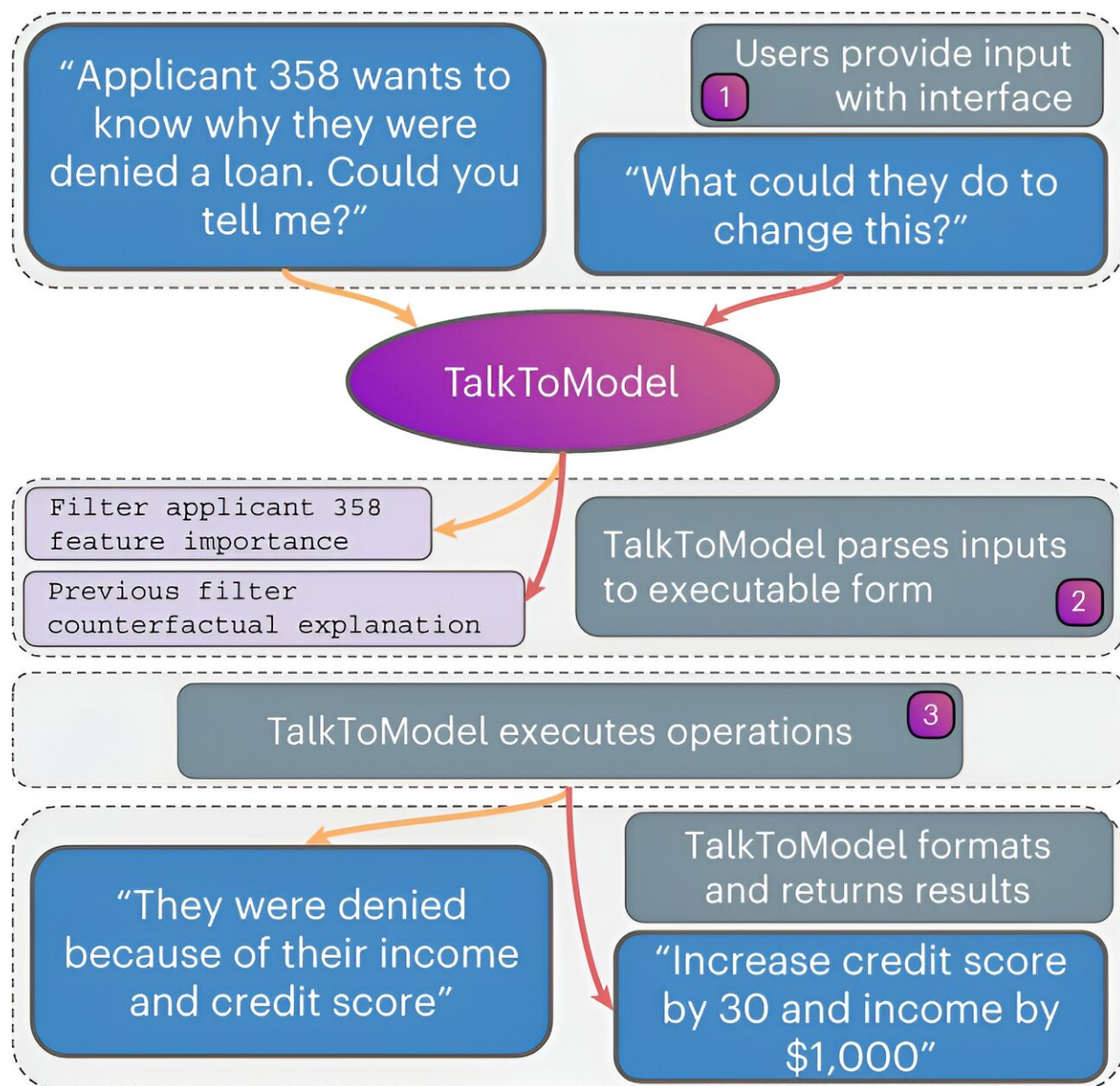
September 12 2023, by Ingrid Fadelli

Overview of TalkToModel. Credit: *Nature Machine Intelligence* (2023). DOI: 10.1038/s42256-023-00692-8

Machine learning models are now commonly used in various professional fields, while also underpinning the functioning of many smartphone applications, software packages and online services. While most people are exposed to these models and interact with them in some form or the other, very few fully understand their functioning and underlying processes.

Moreover, in recent years, machine learning algorithms have become increasingly sophisticated and complex, making the processes behind their predictions harder to explain even for experienced computer scientists. To increase people's trust in these highly advanced and promising computational tools, some research teams have been trying to create what is known as explainable artificial intelligence (XAI).

These are essentially machine learning models that can explain, at least in part, how they reached a given conclusion or what "features" of data they focused on when making a particular prediction. While XAI techniques could be more robust and reliable, most of them have not achieved particularly promising results, as their explanations often leave room for interpretation.

Researchers at University of California-Irvine and Harvard University recently developed TalkToModel, an interactive dialog system designed to explain machine learning models and their predictions both to engineers and non-expert users. Their platform, introduced in *Nature Machine Intelligence*, allows users to receive simple and relevant answers to their questions about AI models and their functioning.

"We were interested in finding ways to better enable interpretability of models," Dylan Slack, one of the researchers who carried out the study, told Tech Xplore. "However, practitioners often struggle to use interpretability tools. So, we thought it could be better if we let practitioners 'talk' to machine learning models directly."

The recent study by Slack and his colleagues builds on their earlier works focusing on XAI and human-AI interaction. Its key objective was to introduce a new platform that would explain AI to users in a simple and accessible way, similarly to how OpenAI's conversational platform ChatGPT answers questions.

Their system has three key components: an adaptive dialog engine, an execution unit and a conversational interfaced. The adaptive dialog engine was trained to interpret input texts in natural language and generate sensible responses to these texts.

The execution component essentially composes the "AI explanations" that are then translated into accessible words and sent to users. Finally, the conversational interface is essentially the software through which users can type their prompts and view answers.

"TalkToModel is a system for enabling open ended conversations with machine learning models," Slack explained. "You simply ask the system a question about why your model does something and get an answer. This makes it easy for anyone to understand models."

To determine whether users might find their system useful, the team asked different professionals and students to test it and share their feedback. Overall, most of their study participants found it somewhat useful and interesting, with 73% of participating health care workers stating that they would use it to better understand the predictions of an AI-based diagnostic tool, and 85% of machine learning developers

confirming that it was easier to use than other XAI tools.

In the future, this platform could be improved further and released to the public. This could contribute to ongoing efforts aimed at increasing people's understanding of AI and their overall trust in its predictions.

"The findings of our studies on humans, including graduate students, machine learning engineers, and health care workers were really interesting," Slack added. "They suggested that the system could be quite useful for anyone to understand models and how they worked. We now hope to keep exploring ways to use more advanced AI systems, such as ChatGPT style models, to improve experiences with this type of system."

More information: Slack, D et al, Explaining machine learning models with interactive natural language conversations using TalkToModel. *Nature Machine Intelligence* (2023). DOI: 10.1038/s42256-023-00692-8.