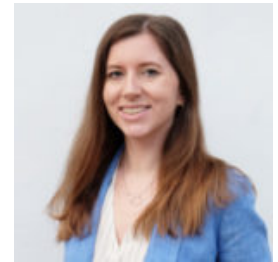# Case study: Evidently AI
State of Open: The UK in 2023
Phase Two "Show us the Money"
Part 2: AI Openness

EVIDENTLY AI

Elena Samuylova,
Co-Founder

Evidently AI is a company that is developing Open Source Software to help monitor machine learning models in production. They don't develop AI but help others who work with AI and machine learning. Their tool helps ensure that other businesses' models work and deliver the value they expect. AI and Machine Learning technologies sit on top of Open Source Software by design and this has been the case for a while now. What is new is large pre-trained models that can also be in Open Source.

Elena is CEO and co-founder of Evidently AI, a company developing Open Source Software to help monitor machine learning models in production. Her background includes working at the Russian search engine, Yandex and five years focused on applying machine learning in different industries, from telecom and retail to more traditional internet applications. Overall she's spent 10 years on the applied side of machine learning. Both she and her co-founder are based in London with a dispersed team working from Argentina to the Netherlands. The organisation doesn't have office space.

Evidently's MLOps tool

Users of Evidently's Open Source Software product come from a variety of companies – both startups and enterprises, a lot of ecommerce, FinTech, banking, retail, even manufacturing, so it's very broad. The company doesn't develop AI but helps others who work with AI and machine learning. Companies run different types of models for example, demand forecasting models used to optimise logistics, or a marketing personalisation model and maybe a text classification model. All of these power the business' back end, some of them are user facing, some of them internal. When these models are in production, Evidently's tool helps ensure that they actually work and deliver the value they expect. This is part of what is known as MLOps – machine learning operations. It's a technology stack on the back end of all things that are happening throughout the ML model lifecycle, from model creation to model deployment. Evidently exists in after Model Deployment and is a Python library that is used to enable visibility into production model performance.

AI and Open Source

AI and Machine Learning technologies sit on top of Open Source Software by design. Most tooling used to develop these systems is Open Source. It's been this way for the last decade – it's not something new happening now. It's almost expected by any developer, data scientist, machine learning engineer that works with this technology, that most of the tools will be available Open Source. Typically, you expect your data to be proprietary, but in creating these models, you use Open Source Software.

The new part that's happened is that we are talking about large pre-trained models that can be also in Open Source. This is the new development and huge discussions relate to the LLMs, Chat-GPT and everything that's happening today. Elena's strong belief is that for AI and machine learning it's impossible to create AI without Open Source. In the last six months Chat-GPT 4 created a very easy interface to interact with these sorts of systems which popularised the idea.

She notes, "There was no new technological breakthrough, it's more that this technology has become known and available to so many companies and people that are now experimenting and thinking how to develop it. It's more about a marketing spike that has happened, providing the tool in a very accessible form."

She sees automation as having the biggest impact – automating some specific decisions or processes powered by data and machine learning to help you achieve a lot of different operational efficiencies. These operational efficiencies and productivity are the key to the benefits being created. Chat-GPT and other LLMs bring actual use cases across different industries – that's the most exciting part, the application.

She feels very strongly about disclosure of datasets, not just from the legal standpoint but because disclosure gives a lot of insight to potential users on model capabilities and biases. This is very important for pragmatic use of the model to be successful for anyone. Understanding what the model might know means understanding what type of data sources and curation process it has experienced. There is a lot of interesting work around ML model cards, Data Set cards and data sheets. These are basically documents that describe what the data the model was trained on – even when talking about large language models.

It would be interesting to talk not just about Open Source AI, but about Open Source data sets and Open Source Models, because these are the two most interesting things in this space and they might exist separately.

First published by OpenUK in 2024 as part of
State of Open: The UK in 2024
Phase Two "The Open Manifesto"
© OpenUK 2024

openuk.uk   @openuk_uk   openukopentechnology