Data-Centric AI for trustworthy applications
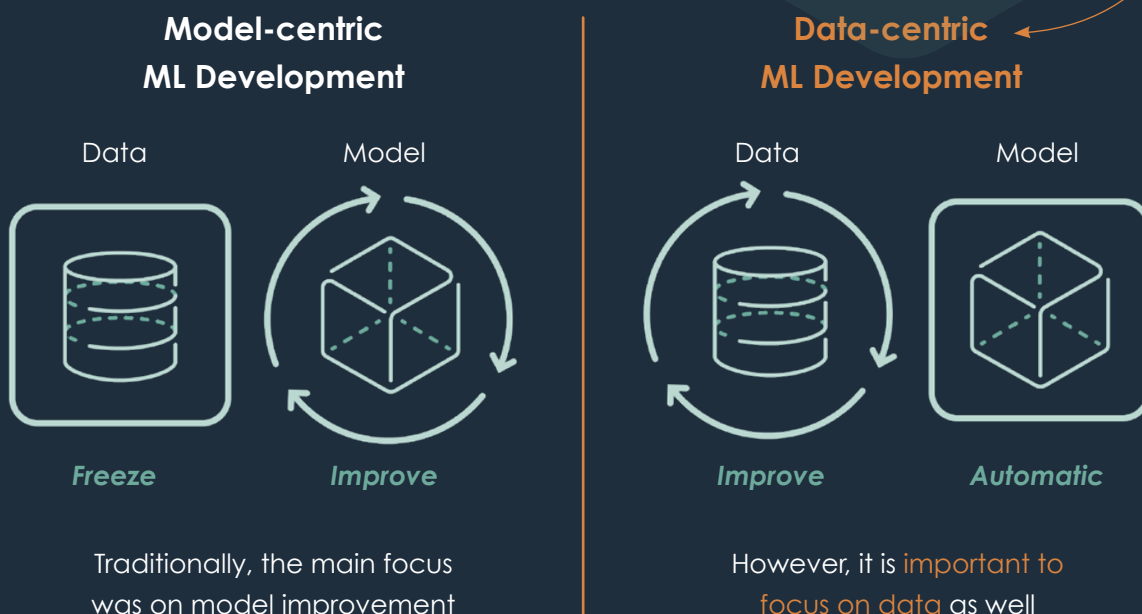
When conventionally creating ML models, data scientists focus their efforts on optimizing the model itself. They do this by playing with feature extractors, architectures, and hyperparameters to achieve the desired target measure (e.g., accuracy).

The problem with this approach is that dataset quality becomes a secondary concern and, as such, ML models will inevitably inherit dataset flaws. This has serious consequences for the trustworthiness of any results.

That's what we do!

Model-centric ML Development

Data   Model

Freeze   *Improve*

Traditionally, the main focus was on model improvement

Data-centric ML Development

Data   Model

Improve   *Automatic*

However, it is important to focus on data as well

Today, highly demanded data scientists spend a significant amount of their time doing what is considered the least enjoyable task of their discipline: cleaning big data. With little insight on the impact their actions have on resulting model performance, they can only proceed in a trial and error fashion.

modulos

Get a demo today at modulos.ai/get-started/

Modulos Data-Centric AI platform revolutionizes the way safe and trustworthy AI applications are built. System-guided recommendations assist business leaders in their AI journey to create reliable ML models in their own domain of expertise

Modulos takes a revolutionary approach

by centering the AI journey around data, providing actionable insights on how to address shortcomings in datasets—such as dirty labels, outliers, missing or incorrect labels, and missing and dirty feature values—to instead build ML models that result in desired performance outcomes.

This new approach is based on cutting-edge research done by leading AI researchers, including our co-founder at ETH Zurich, one of the world's top AI research institutions.

Human data scientists cannot predict what a change in a dataset implies for model performance. Will fixing a missing value have any real impact on the resulting model? What about noisy data?

Research indicates that the mission of data scientists, which is to create reliable and trustworthy models, would be carried out far more efficiently when augmented by machines like in the case of AutoML.

In this new data-centric approach to building ML models, data scientists delegate the model improvements to AutoML tools. They follow system recommendations to improve and correct flaws in the data, which leads to better models and per-

formance levels downstream. This changes the balance from focusing on purely technical aspects of model tuning to the real drivers of AI—use cases that must include domain expertise in conceptualizing a business challenge and translating it into an ML problem.

Given the latest regulatory focus on fairness, transparency, and robustness (the coming EU AI Act and AI legislation in many countries), new objectives are introduced when computing a model. A system-guided AI journey helps save time and resources in

fixing the portion of the dataset that matters most to model performance, fairness, and robustness. This will dramatically lower the cost of acquiring, labeling, and cleaning massive amounts of data.

With regards to AI trustworthiness, our innovative data-centric approach will facilitate the efforts of companies to comply with the requirements of regulators.

> ## "Human data scientists cannot predict what a change in a dataset implies for model performance. Will fixing a missing value have any real impact on the resulting model? What about noisy data?"

19%	Data loading
26%	Data cleansing
21%	Data visualization
11%	Model selection
12%	Model training & scoring
11%	Deploying models

Data Scientists' time spent on each task

Source: "2020 State of Data Science: Moving From Hype Toward Maturity," Anaconda.

Instead of starting with a given dataset and building a model, Modulos introduces a more data-centric approach where data quality benchmarks are goal-oriented (accuracy, fairness, and robustness). Utilizing different ML models, we give quantitative guidance on how these goals can be achieved most efficiently.

To summarize, our clients will understand whether their data leads to their desired outcome by creating different models for their varied business problems.

modulos

Get a demo today at modulos.ai/get-started/