Build trustworthy and data-centric AI applications

The growth of artificial intelligence (AI) has been unmistakable in recent years and is expected to continue. According to Fortune Business Insights, the global AI market should enjoy a compound annual growth rate (CAGR) of 33.6 percent between 2021 and 2028.

It's clear that, no matter what industry you are working in, AI will likely play an increasing role moving forward. For many, this reality has already arrived.

The Modulos platform is a powerful tool to better harness the transformative power of AI that is sweeping the globe.

But AI is not a simple plug-and-play technology. Its power is profound, stemming from the complexity and the raw speed at which decisions can be made at scale. Although coined—in terms of computing power—ages ago in the 1950s, GIGO (garbage in, garbage out) still holds. No matter how powerful a processing system is, if it is working with flawed data, then flawed results are going to be returned.

To use a sports analogy from the world of skiing, it's easy to "get out over your skis" when using AI. A series of misjudgments based on faulty data can rapidly accumulate and, suddenly, one is in a position of danger.

Poor data management and governance are **some of the chief reasons** that AI projects do not meet expectations. And this seems to be the norm, with **Gartner Inc. predicting** that 85 percent of all AI projects will fail in 2022.

Due to this dynamic, the field of AutoML is quickly developing. Machine learning (ML) applications can be crucial to efficiently prepare raw data for entry into AI systems. But our ML platform, Modulos, can provide additional vital tools. For example, our platform is unique in being able to examine data and judge how the AI model will perform in advance, which is a remarkable way to prevent costly systemic problems down the road.

And, in some cases, Modulos can judiciously repair faulty data to improve such anticipated poor AI model performance and achieve, instead, successful business outcomes. This capability means the significant cost of re-canvassing data can be avoided by carefully intervening in the existing dataset to make corrections before it is fed into the AI solution.

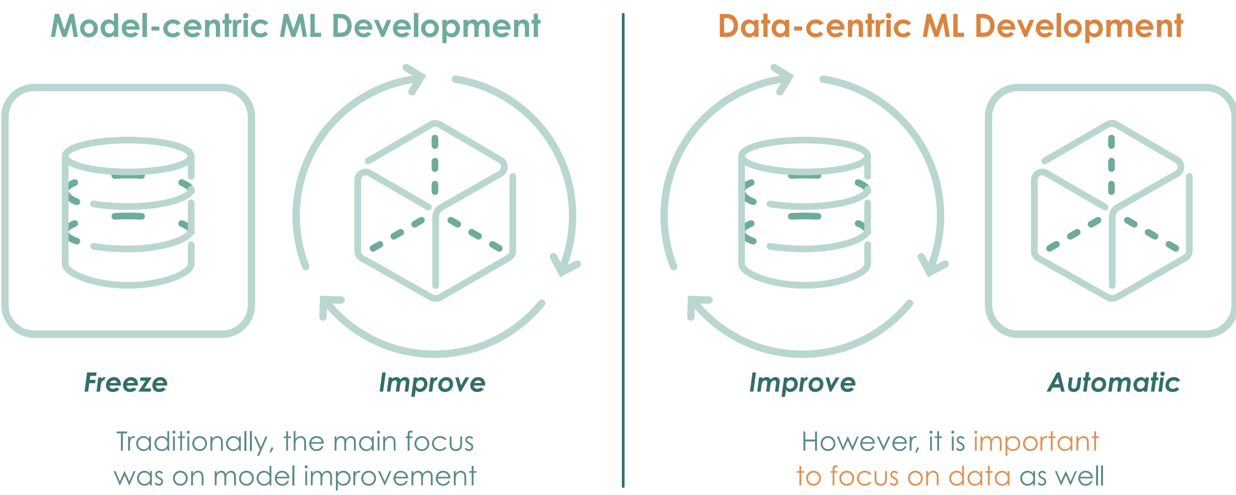
Modulos is a robust tool that can automate much-needed quality control of data. It allows companies to move forward with confidence that their AI initiatives will return useful findings that will result in superior decision-making.

This white paper will now walk you through what Modulos can offer, though first, we'll take a deeper dive into the world of ML modeling.

Data Analysis That Drives Modeling

When conventionally creating ML models, data scientists start with a dataset and 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, largely separated from the model's performance objectives and, as such, ML models will inevitably inherit dataset flaws.



This has serious consequences for the trustworthiness of any results. AI research recently started to focus on data debugging and finding new ways to improve model performance. This is done by correcting the data against a specific objective among accuracy, fairness, and robustness.

Modulos takes a revolutionary approach by centering the AI journey on data, rather than models. The search for good models can now be automated (with AutoML), while the factor that limits model performance is now the data itself.

We solve this by 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 data-centric 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? In fact, blindly cleaning data and filling it with interpolated values can *worsen* model performance.

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, as is the case with AutoML.

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

"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? In fact, blindly cleaning data and filling it with interpolated values can worsen model performance."

Given the latest regulatory focus on how AI solutions are developed and deployed (the 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 by 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.

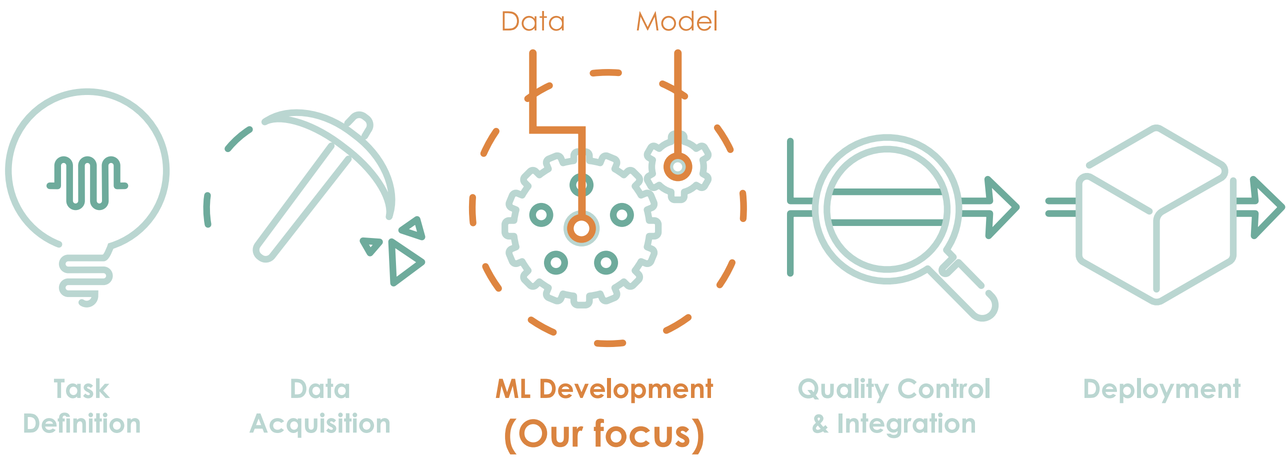
With regards to AI trustworthiness, our innovative data-centric approach will facilitate efforts by companies to comply with regulatory requirements.

To summarize, instead of starting with a given dataset and building a model, Modulos introduces a more data-centric approach in which 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. This empowers our clients to understand whether their data will lead to desired outcomes by creating different models for their varied business problems.

Modulos Data-Centric AI Platform

Developing trustworthy AI applications is a lengthy process and, for most companies, it is too complex as well.

What we offer is a low-code platform that simplifies using AI by reducing dependence on scarce human resources, especially high-demand professionals like data scientists. With Modulos, you can ensure superior model performance while delegating responsibilities to a system that can decide the most suitable models to deploy and which datasets are most impactful.



When dealing with data in a specific industry, business experts are required to make educated decisions based on their understanding of the context in which they operate. Unfortunately, it's uncommon for business domain experts to be competent in building trustworthy AI applications. They either require a strong in-house data science team or a proven automated platform that can, throughout the journey, provide guidance.

Modulos AG, Technoparkstr. 1, 8005 Zurich, Switzerland UID: CHE-233.391.866 | **contact@modulos.ai** | **+41 76 566 05 48**

Taking AI Regulations Into Account

We have also developed a unique approachto tackle future regulatory challenges, allowing companies to build trustworthy AI applications. Modulos will empower companies by providing tools that will anticipate such challenges and simplify efforts to comply with regulations.

On 21 April 2021, the European Commission presented the details of the **Artificial Intelligence Act**, which is expected to be approved by the EU Parliament in the near future. It will set the de-facto international standard for AI regulation in the same way that the EU's General Data Protection Regulation (GDPR) set the global standard for online privacy. Non-EU companies wishing to do business in Europe will have to comply, and other countries will introduce similar regulatory frameworks as citizens demand that the AI systems they interact with are regulated. There is a powerful extraterritorial aspect to the act, which attributes risk categories to AI applications. The penalties for violations are significant, including up to 6 percent of global annual revenue or, for smaller businesses, EUR 30 million.

Topics like accuracy, fairness, and robustness of AI models will have to be at the top of the agenda for any company offering services in the European market, based on the EU's recognition that AI modeling will have a critical impact on society.

Data-centric AI will help companies stay in compliance with these new regulations. Humans will be deeply involved in automated processes and empowered to address data and model aspects such as accuracy, fairness, and robustness.

Let's take the example of a portfolio manager

Often, investment decisions made by humans may exhibit biases like loss aversion (displayed preference to avoiding losses versus realizing gains of the same size) or confirmation bias (interpretation of new events as confirmation of prior beliefs). These biases can drive poor portfolio performance.

Aware of this, portfolio managers want to avoid such biases, which can be far better avoided by ML models through the analysis of historical portfolio performance. However, lacking the technical expertise to build AI solutions, portfolio managers will most often fail in their effort to improve performance by using speedy and effective data-driven decision-making. This is where data-centric AI can come to the rescue, allowing managers to benefit from trustworthy AI solutions.

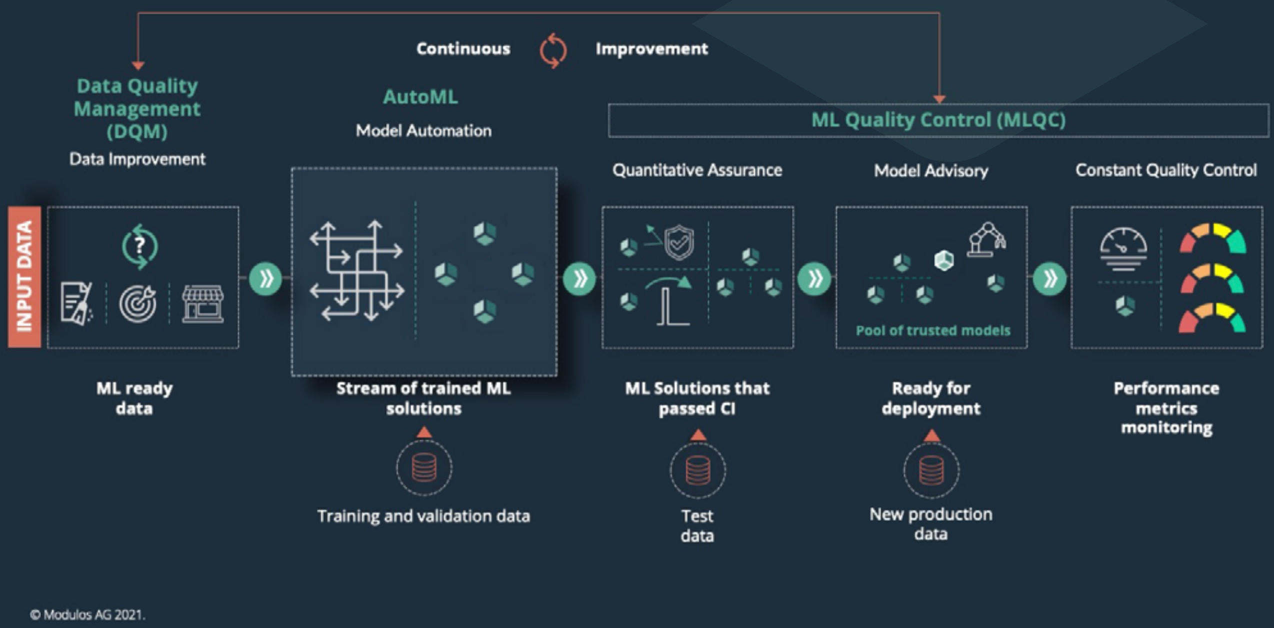
The mission of Modulos is to democratize AI by hiding all the complexity with an abstraction layer that is more understandable by nonexperts, though at the same time ensuring the reliability of AI applications that are built in-house.

Modulos AG, Technoparkstr. 1, 8005 Zurich, Switzerland UID: CHE-233.391.866 | **contact@modulos.ai** | **+41 76 566 05 48**

A Glimpse into MLOps and End-to-End Data-Centric Architecture

Everything related to the life cycle management of ML models is commonly referred to as MLOps. The performance of ML models is directly tied to data quality.

In traditional software development, the developer explicitly defines algorithms. They are complex but transparent, in that for every specific input the output is clear. But with ML models, rules indirectly follow the data to solve the task at hand. This is the genius of the ML model. But because of this, the quality of data used to train models must be carefully considered, as well as the data that is used for inference.



If the relationship between input and output data changes over time, the model's performance will degrade. This concept shift can appear gradually, abruptly, or even be recurring. For example, many ML models trained on normal human behavior had to be recalibrated when the world was hit with COVID-19.

Modulos AG, Technoparkstr. 1, 8005 Zurich, Switzerland UID: CHE-233.391.866 | **contact@modulos.ai** | **+41 76 566 05 48**

MLOps revisited with Modulos - ML Quality Control

Some emerging tools address aspects of the data-dependence dynamic but very few provide an end-to-end solution. MLOps often require significant human supervision by experienced specialists, which is time-consuming and expensive. At Modulos, we are developing reliable, cost-effective, and efficient end-to-end solutions for ML model life cycle management.

As outlined, a model's performance can degrade over time because real-world data is variable. It is important to always use the most recent training dataset and to constantly monitor the model's accuracy in production. And the life cycle of a dataset's value is limited not only because it becomes outdated in relation to the real world; its statistical power also decreases over time simply because reusing the same dataset to assess a model's performance will overfit this dataset. The assessment of the model's accuracy will be overly optimistic and, in production, the model will not generalize well to real-world data. This is especially important when using a test dataset to determine if a model is production-ready. To ensure reliability, statistically rigorous dataset management is central to our MLOps system quality control model.

For the MLOps process, fresh test datasets are always needed. But the cost of collecting and processing data has to be considered. Depending on the use case, obtaining labels (i.e., the ground truth) might involve manual processing that can be very expensive. Therefore, there is a dynamic of wanting test datasets to be large enough to be representative while also wanting smaller datasets for affordability.

The Modulos system leverages the fact that not all samples are of equal statistical importance, optimizing for required dataset size and providing recommendations for which samples should be labeled. This makes the MLOps process more cost-effective.

"Our system provides regular recommendations to guide users and make procedures transparent."

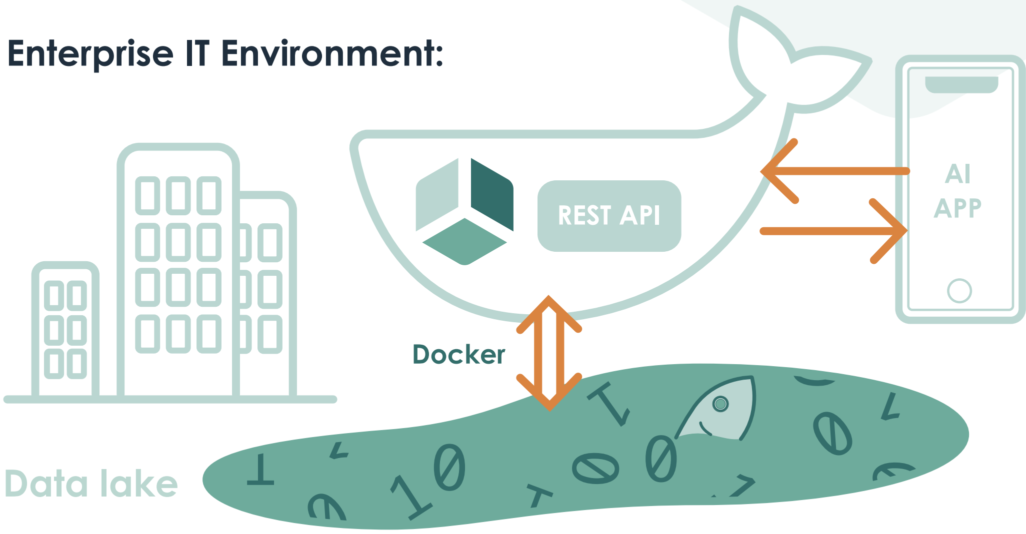
In addition to data scientists, a diverse group of experts is usually involved in ML model management—data engineers, software engineers, business analysts, DevOps engineers, and domain experts. If many models are deployed in production, the task can be very time-consuming. At Modulos, we aim to turn the traditional MLOps into accessible and easy Machine Learning quality control, allowing users to manage the process more independently. Our system provides regular recommendations to guide users and make procedures systematic and transparent. Strong default settings provide security for novice users, while still guaranteeing full control for experts. Overall, this will make MLOps more efficient, reliable, and cost-effective.

Deployment Options, Scalability and Ownership

One of the main barriers to making AI transformational for businesses is overcoming the complexities around transmitting model deployment to production.

Other equally important challenges for making ML models an integral part of companies' infrastructure are ownership issues, access to source code, and flexible pricing. Modulos chose a bold strategy to remove burdens that would limit or slow full internal adoption. When the Modulos data-centric AI platform converges to a best ML model that is ready for production, it is packaged in a Docker container. This is a technology that facilitates the installation of the model at scale into any IT infrastructure close to the data lakes. Additionally, it provides REST API so no additional coding is required around the model itself, which is now ready to start predicting (the inference phase) and differentiating the company's business models with a maximized ROI.

Enterprise IT Environment:



It is common for large corporations to be concerned about not having access to the source code or, in some cases, even legally owning the outcome. At Modulos, we provide the source code of the model, the legal right to use it, and full ownership. Conversely, our pricing strategy does not depend on the inference phase, meaning unlimited predictions can be made. The Modulos platform is based on a fully comprehensive annual SaaS fee. And platform security is inherited by a companies' IT infrastructure, since Modulos can be deployed on-premise—an Internet connection is not even required—making it extremely safe.

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 AG, Technoparkstr. 1, 8005 Zurich, Switzerland UID: CHE-233.391.866 | **contact@modulos.ai** | **+41 76 566 05 48**

Conclusion

As has been emphasized in this white paper, AI is here to stay. It is an innovation to operations comparable to the influx of the IBM System/360 model line in the 1960s, which had a profound impact on business decision-making.

The data-centric AI that the Modulos platform offers allows businesses of all types to harness the full potential of AI by making data ML readiness affordable and consistent. Our product is like a sports trainer who, by enhancing the quality of preparation, allows an athlete to compete at a higher level (and hopefully prevents them from getting out over their skis).

A full range of operational tools are embedded into the Modulos system. With drag-and-drop datasets that include entrenched performance models that are constantly tested and monitored for performance, operational degradation is avoided. We provide system-guided recommendations that do not require in-house data science expertise.

The result is AI-processed data that you can use. Modulos can help you automate your AI journey and, in turn, harness the power implicit in leading-edge data analysis using artificial intelligence.

References

"Artificial Intelligence (AI) Market to Hit USD 360.36 Billion by 2028; Surging Innovation in Artificial Internet of Things (AIoT) to Augment Growth: Fortune Business Insights™."
Fortune Business Insights. September 16, 2021.

Dilmegani, Cem. "Reasons for Artificial Intelligence (AI) Project Failure."
AI Multiple. August 16, 2021.

"The ease.ml Project."
DS3Lab: Data Sciences, Data Systems, & Data Services.

"Gartner Says Nearly Half of CIOs Are Planning to Deploy Artificial Intelligence."
Gartner. February 13, 2018.

Hürtgen, Holger, Sebastian Kerkhoff, Jan Lubatschowski, and Manuel Möller.
"Rethinking AI Talent Strategy as Automated Machine Learning Comes of Age."
McKinsey & Company. August 14, 2020.

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

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 AG, Technoparkstr. 1, 8005 Zurich, Switzerland UID: CHE-233.391.866 | **contact@modulos.ai** | **+41 76 566 05 48**