EazyML
XAI Platform

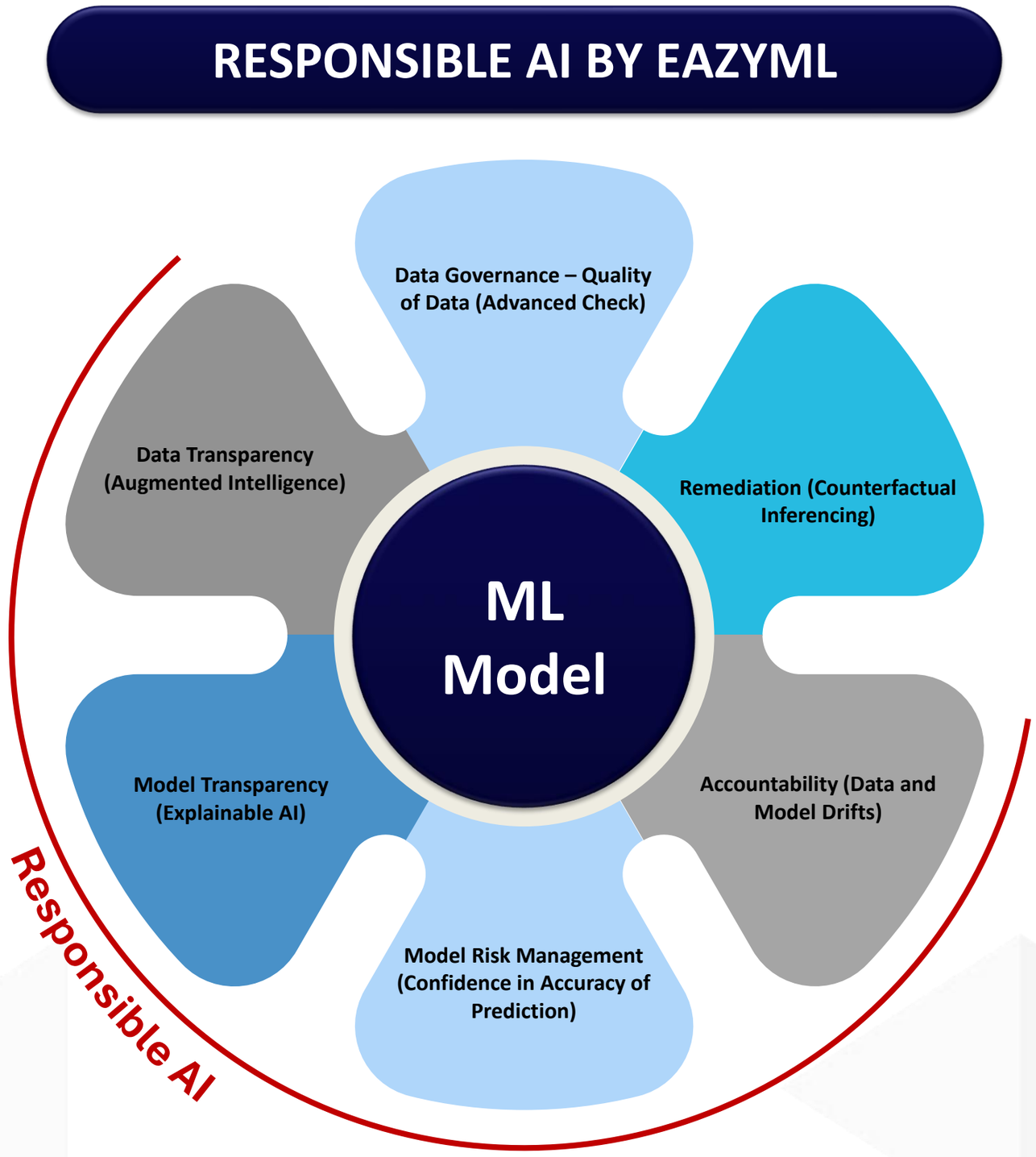
EazyML's

Responsible AI Alleviates Risk

Introduction to **Responsible AI**

EazyML is acknowledged as a leader for its Responsible AI by customers (large enterprises for regulated sectors like finance and pharmaceutical, for instance) and analysts (for instance, Gartner and AIM Research) alike. Its Responsible AI helps with regulatory compliance for regulated sectors and earns trust of the experts for sustainable automation and true ROI.

Below are the various facets of what constitutes EazyML's Responsible AI – exemplified by the diagram for Classical AI.



ML Model – surrounded by Responsible AI functions – delivers trusted Decision AI for an enterprise

Each Responsible AI function for EazyML is packaged as a Python executable (resident in EazyML SaaS or customer cloud), easily accessible by a JSON API

Briefly, we explain how each facet alleviates risk in the enterprise:

Data Governance

Data is your key asset for AI. Is your data afflicted with bias? It typically is. How can you trust Decision AI if it trains on biased data? EazyML Data Quality Assessment Bias Detector identifies bias – so that you can surgically fix for accurate ML models.

Data Quality | EazyML

Data Transparency

Data holds the secrets about your business. Can you trust the insights you mine from data? Is there an insight score that informs you which insights can be trusted, which not (so as to not mislead)? EazyML Augmented Intelligence mines insight, expresses it in simple form, accompanies it with a score to identify those that can be trusted as valid.

Augmented Intelligence | EazyML

Data Transparency

Your ML model makes a prediction, but does it explain the reasons behind it? How can experts trust the model if they don't understand how it works? EazyML's Explainable AI explains the reasons behind your model's prediction by an API call, alleviating risk. Its patented explainability score advises which reasons can you trust.

Explainable AI | EazyML

Model Risk Management

Your ML model many times predicts correctly, but some times incorrectly. How do you know if a prediction is correct or not? You don't; you cannot automate downstream workflow. EazyML will accompany each prediction with a confidence score – the ones above a high threshold are always correct, enabling automation of these records. You've managed the risk of Decision AI's automation for true ROI.

Accountability

Your model was trained by data that's no longer as relevant for a dynamic business. There's risk of model predicting incorrectly to mislead, adversely impacting your business as this error goes unchecked for quite some time. EazyML's Data and Model Drifts continuously monitor the model to alert degraded performance to prevent this. It helps refreshand retrain the model.

Remediation

Your model predicts an unfavorable outcome (for instance, a fraud alert). How can we remedy it – transform it into a favorable outcome (for instance, arrest fraud)? EazyML's Counterfactual Inferencing will inform you which predictors to change, and by how much, to optimally transition from an unfavorable outcome to favorable, making AI actionable.

EazyML Counterfactual Inferencing

> Do you want to use AI to enhance productivity, cut costs, enhance customer service, and increase revenues, but don't know how to proceed?

Professional Services can be packaged with EazyML to help you deliver your business objectives.

EazyML Responsible AI Delivers for **Generative AI**

Of the various facets of Responsible AI, EazyML delivers two core features for Generative AI: Accountability and Transparency. These are vital for the success of Generative AI projects. The sequel discusses them.

Accountability

Generative AI's Large Language Models (LLMs) learn from a wide spectrum of training data, much of it external sources, publicly available, whose accuracy can't be verified. How do we then hold the LLM's accountable? They shouldn't mislead us with incorrect predictions – the response it generates. It's difficult to enforce accuracy because much of the training data from which LLM learns is suspect. This behavior of LLM to mislead is called hallucination – a prime source of risk for an enterprise, especially for regulated sectors – like finance and healthcare, where the risk can have significant adverse financial impacts. (In fact, this is a big reason why Small Language Models are becoming popular).

How do we fix hallucination to alleviate risk? EazyML's confidence score does exactly that. We discuss briefly how.

Consider LLM's Information Retrieval from textual sources. We'll frame in the context of a popular technique called Retrieval Augmented Generation (RAG). RAG retrieves information most relevant for answering the query from a compendium of internal documents of an enterprise; the LLM is instructed (through prompt engineering) to answer the query based on the information contained in the RAG. While the LLM attempts to do that, invariably some background information it has learnt from unreliable public sources creeps in the response, leading to pockets of hallucination.

To remedy hallucination, EazyML computes the confidence score for the accuracy of the LLM response to a query. It is called the **Response Score**, computed using Bayesian logic:

⇨ **P(R) = P(R|RAG) x P(RAG) + P(R|not(RAG)) x P(not(RAG))**

where the events are defined as:

R = correct response by LLM to a query

RAG = RAG is correct as it contains information to generate R

P(R|not(RAG)) can be safely assumed to be zero; we've the following approximate P(R):

⇨ **P(R) = P(R|RAG) x P(RAG)**

P(R|RAG) can be experimentally derived by checking for entailment of each sentence in response with the RAG, and

P(RAG) can be computed using probability theory for the likelihood that the number of chunks assembled as RAG contain the relevant information.

P(R) is hence computed as the confidence score for correct LLM response – what we call Response Score.

An important observation is made here. It is P(R|RAG) that differentiates between the accurate internal documentation and the suspect external information – in the response. The more the external information in the response that can't be verified for accuracy, the less the entailment, and the lower the Response Score.

This logic applies to Agentic RAG framework as well. Each of the multi-agents performs a sub-task; the sub-task's accuracy of response is measured by the formula above. The accuracy of the final response requires that each agent performs its task correctly:

P(Response) = P(Response by Agent 1) x P(Response by Agent 2) x … x P(Response by Agent N)

for N agents.

A simple logic for automation of LLM response follows – one that makes it accountable for risk-averse implementation:

If (Response Score > Threshold)

/ % of workload that's automated for significant ROI */*

send response downstream for automated workflow

else

/ % of workload that leads to small ROI due to quicker manual inspection */*

response manually inspected by the expert before flowing downstream

The threshold for Response Score is set based on tolerance for risk; as an example, a FinTech use case for credit risk set it at 0.87.

EazyML holds LLM's accountable for accuracy to avoid hallucination.

Transparency

Let's say that the Response Score is low, requiring manual inspection. Or even if the score is high, the workflow is in a critical sector that demands the expert certify the response before it can be used. How would an expert mine through voluminous internal documentation to determine if the LLM generated correct response? The LLM must be transparent to explain how it predicted the response.

EazyML implements Transparency comprehensively for Generative AI in two ways – by answering the following questions:

- What were the relevant information that were considered in deriving the LLM response?
- What parts of the relevant information were most influential in determining the response?

Let's briefly explain how EazyML answers the two questions. The first one is straightforward: RAG is a compilation of relevant information; in fact, the LLM is instructed to use it to generate the response by prompt engineering. EazyML displays the RAG as evidence of correctness of the response. Regarding the second question, EazyML uses diversity of LLMs to find a common ground of what were the passages in the RAG that were most influential.

Conclusion

Enterpises are risk-averse. Hallucination is a major challenge for them. How much of the response has been afflicted by the unvetted suspect external information used to train an LLM? Without a clear answer, the benefits of Generative AI is limited. This has handicapped Generative AI implementation. Worry not – EazyML helps!

Its Responsible AI alleviates risk for big-ticket projects - rather than small projects that nibble on the periphery of the business, especially true for Generative AI projects, go after the big-ticket projects, core to the business, for significant savings, but do so in a risk-averse way to get the approvals necessary to tinker with a high visibility project – convinces executives of AI's value-add.

And the best part: EazyML is easy to trial as you download the specific packages from pypi.org into your ML environment, within the confines of your firewall to enforce data privacy, then use EazyML's standard JSON API to trial it, stitching your own data processing pipeline, and integrating in your workflow for automation.

Trial EazyML today as a complement to your existing tools for Classical and Generative AI to deliver business objectives.

For more information contact: info@EazyML.com