Arya.ai

AI EXPLAINABILITY FRAMEWORK IN FINANCIAL SERVICES: THE TRUST IMPERATIVE

WHITEPAPER

INTRODUCTION

It is imperative that AI is going to be at the core for next generation apps. With almost all major tech players offering AI enabled solutions, we see it as a default feature. Though the early cycle started in early 2012 to 2017, the hype cycle started around 2017 - 2019 followed by reality correction between 2017 to 2020. So we saw more success in recent times. Adoption was only further fueled by global pandemic and remote workflow.

COVID is as big an influencer as any CIO/CDO towards driving AI adoption in enterprises

With packaged AI APIs in the market, more people are using AI than ever before without the constraint of compute, data or R&D. This provides an easy entry point to use AI and gets them hooked for more. This is not limited to small companies, a similar trend is emerging strongly in B2B markets. More enterprises are using APIs; something they would never have considered before.

Large tech companies have expanded their API offerings portfolio rapidly; which was previously the core business proposition for many early and growing startups.

While APIs address standard business requirements, adoption of AI in vertical specific use cases are largely driven by horizontal platforms coupled with implementation services. This has been the approach for enterprises to use AI since the 2014s. While enterprises appreciate the scalability offered by end-point solutions, control and flexibility are of priority for enterprises, particularly for regulated industries like Financial Services, Healthcare, Aviation etc. As most of the data sets in enterprises are within organization this approach has been the most adopted right now. The issue with such an approach is - it takes a long time to make necessary customizations, success is highly dependent on the experience & expertise of vendors and is costly!

A simplified picture of AI adoption can be represented as below:

Early wave: Hype, excitement and confusion

Early success creates more hype, with unrealistic expectations that corrects over time through learnings from experience or succcess stories

Intermediary phase: Services driven adoption and key use case discovery

Wave I: APIs-as-Service

Standard APIs gain attention and adoption as they provide the capability and scale without the need to build from ground-up

Intermediary phase: End-point solutions that are pre-customized; Horizontal ML platform with easy to use interfaces

Wave II: Verticalized AI PaaS

With more B2B demand, vertical specific PaaS gain more traction as the generic platform overwhelm with need of customizations

Intermediary phase: Vertical specific operating platforms, run-on-interactions

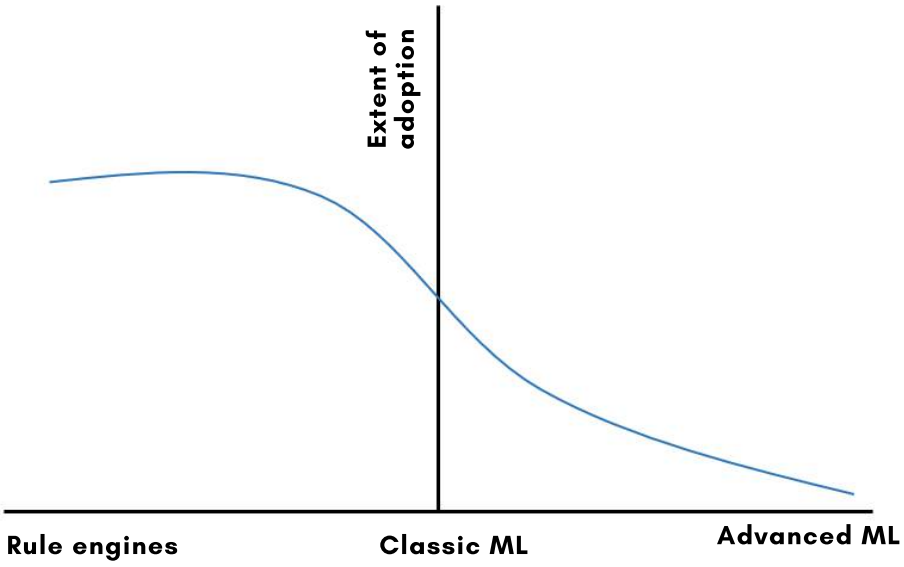
Wave III: Expert AI-as-Service

With aggregated knowledge, task specialists will evolve offering expert decisions through APIs

The other path being followed, especially by the larger enterprises is building and growing in-house data science teams using open source frameworks or horizontal platforms. Horizontal platforms and frameworks are matured and quick for general use cases! But as the use cases become more specific to an industry, the need for customizations on the horizontal platform becomes painfully expensive on both human and cost resources.

AI IN FINANCIAL SERVICES

Financial services industry has both sides of the story - very early adopters using advanced AI techniques and late entrants are very far from using any kind of AI. But the majority of financial institutions today use some kind of AI and have strong ambitions to deploy AI sooner.



Adoption of various types of AI engines in Financial Services

Many Financial Institutions (FIs) today use rules engines to automate decisions, raise triggers and monitor transactions. The primary reason is transparency, secondary is simplicity followed by ease of use. FIs have confidence in systems where they are clear on system functioning. When a case is processed by rules engines, it is very clear, based on the activated rules, as to why a case is declined or accepted ! And such rules can be written by any expert in a short time and added quickly. But, rules don't carry context. Not every application with 'age <30' is an ideal customer nor every customer 'age > 60' is a risky customer. And in many cases, even though all rules conclude that - the case is acceptable, an expert would understand why the case should not be accepted because of an additional logic not written in the rule engines. Such rules based AI engines are heavily used by FIs in use cases like - Underwriting, Claims processing, AML, Grid based pricing, Quote creation, Fraud monitoring, Audit etc. across products like Insurance, Banking and FS.

To control the risk of such a system, users tend to change the strictness of AI rules engines from time to time. Many times, these are eventually proven to be too stringent or too lenient! The impact of rules engines is largely efficient for standard cases in the transactions as rules

can be defined for simpler boundaries. Hence, it is observed that rules engines are applicable to limited volume of transactions and the rest are directed to manual experts.

On the contrary, self learning AI engines using advanced techniques like Deep Learning are much more sophisticated and proven to perform much better than rules engines or classic ML engines. Unlike rules based AI engines, self-learning algorithms using Deep Learning can learn more nuanced patterns and features from the data that can replicate 'high skill' human expertise. These systems are better at identifying and interpreting key markers for decisioning across multi-dimensional data inputs like an expert. They can include different types of data while processing and provide end-point outcomes like an expert.

If self learning systems are all so good, why are such systems not more widely adopted today?

Reasons start from - trust, time to roll out, cost and ease of controls.

TRUST & EXPLAINABILITY

If it is not explainable and a black box - FIs would be highly reticent/cautious to use such systems.

LONG TIME TO CUSTOMIZE

If it take very long to customize and requires long R&D - opportunity cost overrides value add

HIGH COST TO BUILD

If it needs special skills to control - makes it very tough to be accept by business users

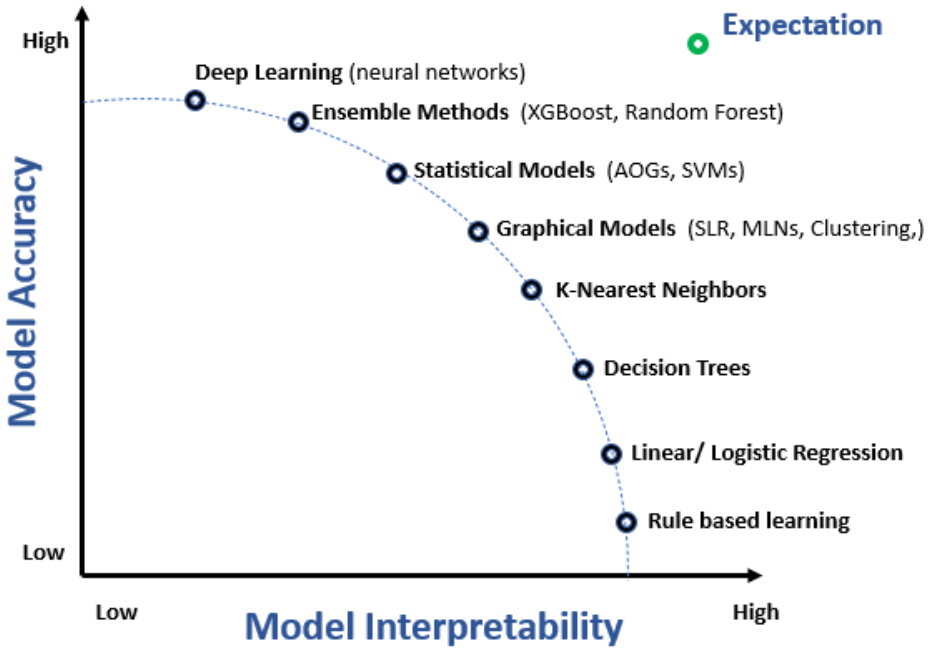
EASE OF CONTROLS

If all the efforts are limited to one use case and are not reusable - ROI becomes a challenge!

There are multiple horizontal tools aiming to solve these issues and assist AI experts in expedited roll outs of complex AI engines. However, many of them are focusing on the build side of the problem addressing points #2 and #3 above. While this has helped to diversify the use cases of complex AI engines from traditional use cases like - image classification, object detection, NLP etc. but there remains a large opportunity for FIs to leverage such complex systems. As these are getting there, solving point #1 and point #2 is super critical to fuel the success of AI in Financial Services.

The trade-off between Accuracy Vs. Interpretability

Considering the various models available today, there is usually a trade-off required between accuracy and interpretability - as the AI algorithm moves towards accuracy, interpretability starts fading. The graph below shows some of the machine learning techniques mapped along with their levels of explainability and accuracy.



Technical users will be more inclined towards accuracy as they can deliver better ROI, consistency in performance and have longer models life cycle. Whereas business users or compliance teams would expect transparency in understanding the outcomes. An ideal explainable model will be on the upper right quadrant, where accuracy and interpretability are the highest.

AI EXPLAINABILITY IN FINANCIAL SERVICES

Trusting the engine is really important for any user primarily if the user acts based on the recommendation from the engine. And providing explanations is a basic requirement of an AI engine. Given the diligence and attention required for financial transactions, if the user does not trust the model, they will not use it no matter how accurate it is! Such explanations can be categorized as - prediction related, model related, data related and controls.

- **Prediction related** - how did the engine arrive at the prediction
- **Model related** - how did the model analysed data and what has it learnt from it
- **Data related** - how is the data used to train the model
- **Influence and controls** - what can influence the system and thereby ways to controls

The user needs to trust the model functioning in a reasonable manner if deployed, has enough confidence on the model prediction and agrees that the data is used in an acceptable way. Once user trust is established, users need to understand 'when it can fail'! Such that necessary controls can be built.

And yes, regulatory frameworks are emerging! European Commission has proposed the first legal framework for AI aiming to develop 'human-centric, sustainable, secure, inclusive and trustworthy AI'. One of the many mandates in the proposal is to make AI systems adhere to transparency and traceability. Regulatory compliance and accountability on financial transactions is not new but as the decisioning authority here is a 'machine', it was unclear on the applicable guidelines. The EU legal framework for AI not only provides clarity but is an encouraging move for 'AI' engines to take up more responsibility and also the reward with it.

Moreover, the General Data Protection Regulation (GDPR) which came into force in May, 2018 also creates an obligation for companies to provide detailed explanations on how and when companies are making automated decisions about customers, and also the right to challenge these decisions. These additional requirements highlight the ever-increasing need for explainable AI.

With growing complexity of AI engines, it is becoming harder to explain the AI models. AI/ Deep Learning systems are hard to build. Once built, you need to achieve consistency and meet various metrics for being production ready. The systems

operations during training and inference is a highly resource intensive exercise. Even then, organizations steer clear of deploying these systems in the core process, since such solutions carry a somewhat justified reputation for being 'black boxes' characterized by poor transparency. To ensure that the solution enjoys confidence and trust, it is important to build additional transparency layers and controls for the users.

There is a lot of attention and buzz around explainable AI since 2016/17. There are various approaches proposed for self learning AI engines. Most prominent ones are SHAP, LIME, Integrated Gradients etc. Most of the techniques use a permute and predict approach to explain the model functioning. Such approaches do offer approximations of model functioning but are understandable only to an AI expert! Like we saw, a growing need for simplistic tools & frameworks for rapid adoption of AI, is similar to the need for a simpler framework for explainability as well. And at the same time, generalized frameworks may work for common use cases but as the use cases get specific to the users, such frameworks tend to be underwhelming! Heavily regulated industries like Financial services (FS) is one such vertical where generic explanations are not enough because of - regulatory risk, reputation risk and finally transactional cost! They need an acceptable framework that can cover the needs of various stakeholders in the process from the end customer to regulator!

Expectations from various stakeholders in the transaction:

BUSINESS OWNER:

- How is the system arriving at the outcome?
- What variables are considered and how?
- What are the acceptable and unacceptable boundary limits on this transaction?

AUDIT/ RISK TEAM:

- How can I defend the AI decision in front of a regulator or customer?
- What are the influencing factors in the decisioning and learning?
- How can I trust the model outcome?
- How can I ensure the consistency in the model in production?

DATA SCIENTIST:

- Is there a bias in the data?
- What has worked in the network and what isn't?
- How can I improve the model performance?
- How should I modify the network?
- How can I be informed about model deviation in production?

END CUSTOMER:

- How did you arrive at the decision?
- How is my data used for the decision making?

AI CONTROLS FOR FINANCIAL INSTITUTIONS

While there is due attention for 'explainability', there is less attention around 'Controls & continuous monitoring'. It is not possible to define necessary controls without understanding how the system works! In addition to explainability, providing necessary controls is also very important to Financial Institutions. There are ample examples when a software glitch caused serious financial and reputation damage to financial institutions. Depending on the volume and importance of the transaction, there needs to be necessary controls. EU guidelines highlighted the need for controls.

Continual Monitoring

Once a model is deployed, explainability can help with system monitoring. In production, over time, there are many reasons why a model's performance can drift and therefore needs to be watched.

"Financial Institutions needs a robust explainable framework and simple to use controls to increase large scale adoption of complex AI systems"

One reason is rare or outlier cases; in such cases understanding the parameters on which the model is making a decision can flag off anomalies and routing to manual processes or audits. To give an insight into the working; if in such cases the model's decision is based on comparatively too few attributes, then that could be a 'relook' trigger.

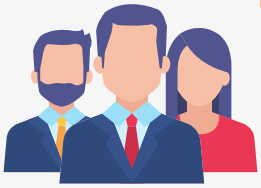
The second reason is data drift or changes in the 'environment' - Data drift is a very common scenario but can lead to serious harm if not handled properly. As AI engines learn from the data, the scope of their knowledge is also limited by the data provided. Transfer learning and progressive learning can expand the usage of the learning but applicability can not be very different from training data. So, it is important to track how the data is drifting in production. Data drift can be more than simple variations in a few parameters, it can be complex relational changes also. If tracked well, errors can be prevented before processed. This can prompt the system designers to recalibrate or retrain the system.

'AI Control framework' is matured in industries like transportation where autonomous vehicles are deployed today. While the AI is driving the vehicle, the driver always has necessary control to override or modify the actions whenever they want. Such controls

should be simple and easy to use. If the controls are too technical, then it needs specialized experts and increases chances of human errors! An autonomous vehicle is a very technical product, but the controls are super simple. Any driver can use it without too much technical knowledge.

A strong and scalable 'AI Control' framework can also enhance the applicability of the AI engines. eg: company has launched a new product aimed at new market segments (or geographies or demographics). The model can be able to weigh 'new' data points and continue to make decisions with larger attributions to 'familiar' data points and use the guidelines to improve its inferencing accuracy!

Expectations from users:



BUSINESS OWNER:

- Can I override the machine outcome?
- Can I modify the variables in the transaction?
- How can I ensure it is following all guidelines of the transaction?
- How can I control the system outcome?

RISK TEAM:

- Who are the participants in the transaction?
- How can I quantify and cap the risk?
- How can I define controls on the engine?
- What are the quality markers for the systems? And relevant back-up actions?

DATA SCIENTIST:

- How can I adjust the bias in the data?
- How to understand if the model drifted because of new learnings?
- How can I define the recursive actions for data drift scenarios?

CONCLUSION

Financial Institutions, carrying an inherent fiduciary responsibility towards their customers big and small, need to be able trust the systems they use and deploy and explainability is really at the core of trust. FIs needs a robust explainable framework and simple to use controls to increase large scale adoption of complex AI systems.

Daniel Dennett, a renowned philosopher and cognitive scientist who studies consciousness and the mind said[1] "I think by all means if we're going to use these things and rely on them, then let's get as firm a grip on how and why they're giving us the answers as possible." This was in 2017, when discussions on explainability had just begun. Today AI explainability has moved past the proof-of-concept phase, to scalable, deployable solutions.

REFERENCES

1. M. T. Ribeiro, S. Singh, and C. Guestrin. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In SIGKDD, 2016.
2. The Dark Secret at the Heart of AI (https://www.technologyreview.com)
3. Giles Hooker and Lucas Mentch. 2019. Please Stop Permuting Features: An Explanation and Alternatives.
4. Europe & The Dream For Ethical AI (https://analyticsindiamag.com/)
5. Explaining AI by Harry Shum (https://a16z.com/2020/01/16/biases-and-black-boxes-a-call-for-ai-transparency/)
6. Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) (https://digital-strategy.ec.europa.eu/)
7. The Dark Secret at the Heart of AI (https://www.technologyreview.com/2017/04/11/5113/the-dark-secret-at-the-heart-of-ai/)
8. Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador Garcia, Sergio Gil-Lopez, Daniel Molina, Richard Benjamins, Raja Chatila, Francisco Herrera. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. In Information Fusion, 2020.

Arya.ai offers a low-code 'AI Operating Platform' for financial institutions to deploy, scale and maintain responsible self-learning AI engines for core business functions like Underwriting, Risk Monitoring, Claims Processing etc. all on the same platform. The modular construct of the platform allows financial institutions to scale one function at a time and offers utmost flexibility to centralize the control on AI assets like - Models, Data pipelines, APIs, Security Guidelines etc.. This makes it the only AI platform required to achieve organization wide adoption of autonomous AI. Arya brings in the best of both worlds - products & platforms onto a single unified technology stack.

Arya.ai was one of the first startsups to deploy deep learning in Financial Institutions since 2013! Arya.ai founders named in Forbes Asia 30 under 30 under 'Technology' category. Arya.ai is also named in 'Top 61 AI Startups global' list by CB Insights.

Contact:
hello@arya.ai

Lithasa Technologies Private Limited,
1102, K.P.Aurum,
Marol, Mumbai,
India - 40007=59