WHITE PAPER

ROADMAP TO AI ADOPTION IN FINANCIAL SERVICES

Infosys®
Navigate your next

Introduction

Organizations that successfully adopt emerging technologies are most likely to survive in the ever-competitive business environments. For majority part, embracing new technologies has been a linear evolution, mostly entailing migration from existing technology stack (labelled "legacy") to the emerging technology stack (labelled "modern"). But ever so often, a paradigm shift comes along that necessitates a deeper and lateral reorganization of structures and processes. Artificial intelligence is one such paradigm.

With its emphasis on "data first" (as opposed to "rules first"), AI presents a completely new way of applying technology to solve business problems. Enterprises need a carefully thought-out plan if they expect to be successful at adopting AI. The problems are compounded for financial services organizations, as most of their data loads continue to remain on legacy platform, thereby necessitating an overhaul of their complete data infrastructure to modern technology stacks. Moreover, FS organizations are highly regulated and traditionally have been wary of adopting a culture of innovation due to the inherent risk of non-compliance of its carries. Both these elements are essential for an enterprise-wide AI adoption strategy.

Any new technology goes through a rapid phase of evolution before it is deemed as "having arrived". During this phase, it goes through challenges such as a lack of standards leading to portability issues, or parallelly emerging tool stacks that result in interoperability issues. Such issues are temporary as the natural process of evolution will automatically address them. AI brings in challenges that the IT industry had not yet encountered, such as

- A shift from requirements-first to data-first approach towards problem solving
- Rapid co-evolution of technology with scientific research
- Imperative expectations of on-demand availability of high-quality data

In this whitepaper, we will attempt to distil out distinct patterns of AI adoption, understand why these patterns will need to continually coexist, how modular components can be assembled together to compose business solutions, with particular focus on Financial Services, and what skills would be needed to create these components.

External Document © 2021 Infosys Limited

Enterprise AI Adoption Models

We describe four patterns of AI adoption across organizations. This somewhat simplistic view is useful to understand the essential flow of activities and their accompanying concerns as AI adoption gains maturity in an enterprise. Do note, that these patterns should not be viewed as successive evolutionary stages. Rather, while most organizations will move from one pattern to the next as indicated in figure 1, we believe these patterns will continually coexist, as each provides set of functions complimentary to the others.

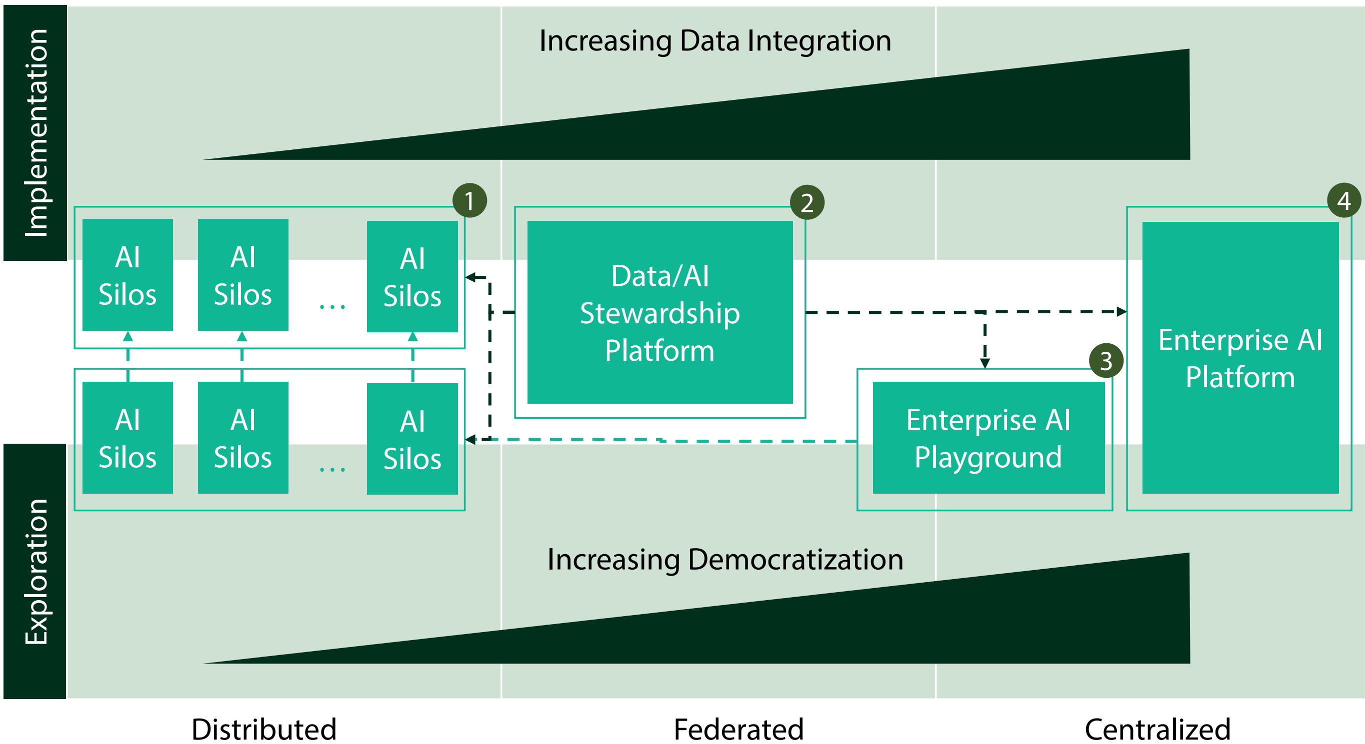


Figure 1: Patterns of Enterprise AI adoption

AI Silos – It begins with some enterprising employee taking a csv dump from a transactional system and to train a machine learning model on their own device (or a personal cloud account). If a trend picks up, it soon mushrooms out into a community of experimenters. It is vital to harness this momentum and stifle all attempts to suppress it. Early initiatives should be to establish an innovation framework and a culture of learning. This includes providing access to incubation labs with the right infrastructure, proving upskilling opportunities through certifications and establishing a ringfenced center of excellence that interfaces with project teams to work on business relevant use cases.

Data Stewardship Platform – Machine learning models are best trained on real data (synthetic generation which possibly introduces a risk of performance accuracy). This means that to train model, data moves out from the source-of-truth repositories and go through a pipeline of transformations over multiple systems to finally arrive at the predictive models. Data quality issues get amplified over this process resulting in less-than-acceptable accuracy. It is imperative to set up a formal Data Stewardship program that seeks to eliminate data quality issues at source. It is equally important for the program to address issues related to data privacy and ethical concerns arising out of hidden biases.

Enterprise AI Playground – Increasing desire of democratization leads to creation of AI playgrounds. Doing AI requires deep data science and machine learning skills, which in turn, ask for familiarity with algorithms and statistical concepts. Platforms have sprung up that claim to make it easy for anyone to train models through easy-to-use user interfaces. Such platforms are excellent for rapid experimentation. However, they do not work as well for production implementations as real-world problems are not solved just by training a model, rather a pipeline composed of multiple steps, each of which could be probabilistic or deterministic, must be built.

Enterprise AI Platform – An enterprise wide platform is highly desirable from the separation-of-concerns design principle. A centralized platform that services all AI needs across the organization will bring a strong governance through standardization and benefits of shared usage. On the other hand, it will necessitate movement of data from all source systems to the centralized platform, almost mandating a modernization of the entire data infrastructure. Even leaders in AI adoption will practice a degree of federation, as the complexities and costs associated with data integration outweigh the benefits of complete centralization.

AI Roadmap and Use Cases

Vast amount of literature is available in public domain on typical AI uses cases being explored by financial services organizations. However, there is little that can act as a coherent guide for an enterprise AI adoption strategy. We attempt to remit that gap with our findings.

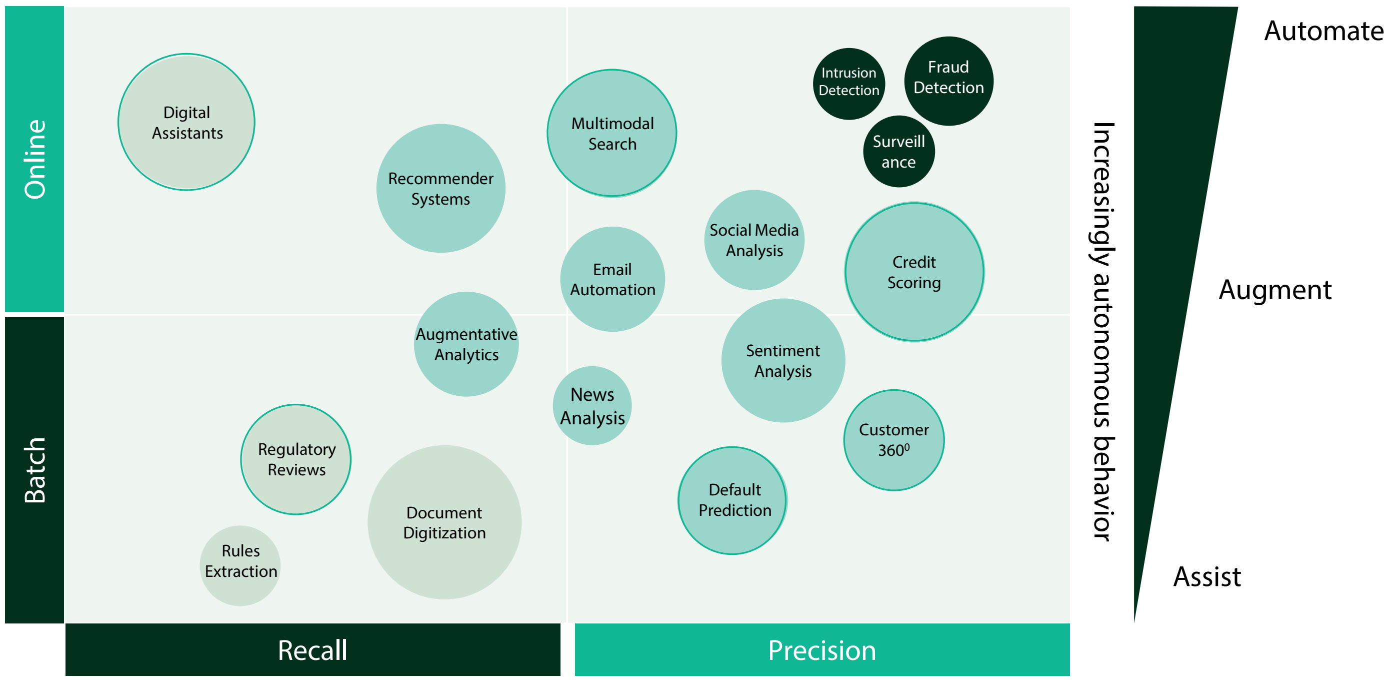


Figure 2: AI Building Blocks

Use Cases – Organizations need solutions for business relevant problems and they are turning more-and-more towards technology for fulfilling this need. For financial services organizations this could mean detecting fraudulent transactions in real-time or for assessing risks associated with credit approvals. We categorize them as business use cases (henceforth simply called use cases) as they are end objectives in themselves. However, use cases can also be about building technology components that can go on to service one or more business uses cases. To differentiate as well as to highlight their essential nature, we henceforth refer to them as building blocks.

AI Roadmap – It's best to attempt low-hanging fruit first and leave the most complex problems for the last. A good roadmap is one that starts with a pragmatic approach to assessing this complexity of problems. Our recommendation is to use "degree of autonomous behavior" of a bot as the proxy for its complexity. Start by embedding intelligence in bots that act to assist humans and then progress to building bots that augment human capabilities. Note that in both cases "human-in-the-loop" is a common feature. The difference lies in the degree of importance given to the recommendations of the intelligent bot. Only when augmented intelligence is proven to be performing successfully in production should the full automation blocks be taken up for building. Two factors help in arriving at this segmentation of the degree of autonomous behavior. First, consider whether a bot must infer in real time (or worse still, learn and infer in real time, called "online learning"). Real time bots cannot afford a human intervention, and hence are completely autonomous. Second, evaluate the relative importance of precision vs. recall (they are inversely related) expected from a bot's performance. Typically, high precision (low false positives/negatives) is hard to achieve and such cases tend to hit a ceiling at the level of augmentation. Regulatory use cases such as KYC and AML belong to this category.

In figure 2, use cases are depicted with boundaries whereas building blocks are boundaryless. Circles towards bottom-left belong to assist category and should be attempted first whereas those clustered around top-right tend toward autonomous behavior and should be taken up only after a solid AI foundation has been laid.

Use Cases and Building Blocks

Use cases are the end objectives, the problem statements that are looking out for solutions. Building blocks is one level lower architectural layer that is used to compose these solutions. Real world use cases translate into problem statements that, in most cases , require two of more building blocks to be bundled within custom-built use case specific codebase. Use cases inherit the complexities of their underlying building blocks. Go with the highest degree of autonomous behavior to determine the priority of a use case.

Businesses face problems aplenty, which means the use case catalog quickly burgeons into a big backlog of desired capabilities. Tackling individual business use cases is not a scalable approach. Rather, we have observed that a limited set of core AI components can be used as building blocks for solving majority of business problems. These building blocks are best crafted as capabilities deployed as REST services that can be consumed to compose all kinds of solutions across FS sub-domains. When aspiring to build an enterprise AI platform, target building these core building blocks and leave solving business problems to units.

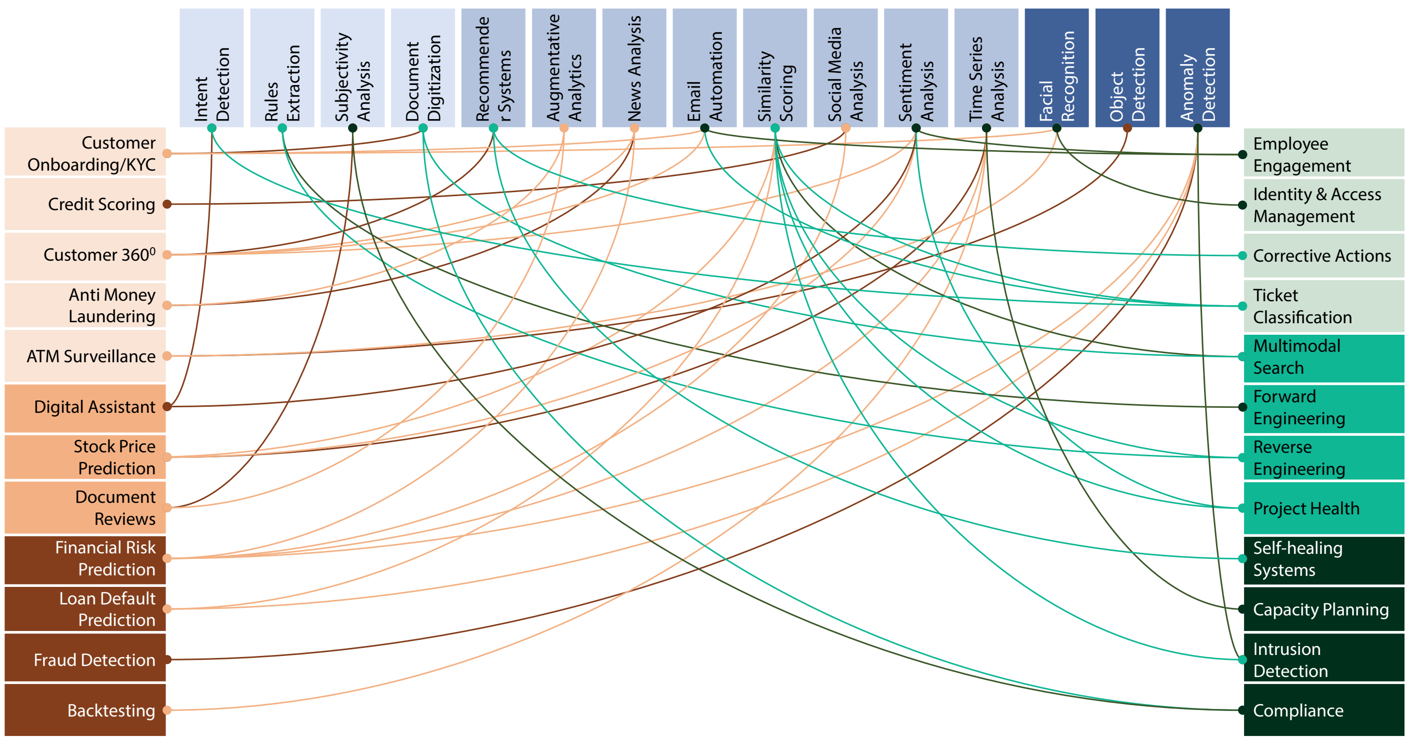


Figure 3: AI Use Cases

Figure 3 links building blocks to a range of use cases. Vertical and horizontal use cases are depicted as columns. Vertical use cases are specific to financial services. Horizontal use cases, in the context of financial services institutions, will automate business functions such as HR, legal, procurement, sales, marketing or other enabling functions. The stacks are color coded to segregate front-, middle-, and back-office use cases. Connectors in bold represent a deeper dependence on those building blocks.

A mapping of this kind serves as an instrument for exploring possibilities that are not readily evident. As an example, when planning on creating a digital assistant, teams are apt to focus on intent detection to enable the bot to respond to user queries. But the same channel can include detection of sentiments over the course of a chat conversation, thereby making the digital assistant an invaluable source of customer predilections. The thought can be extended further by building a recommendation engine that suggests "next-best-actions" to relationship managers.

External Document © 2021 Infosys Limited

Building Blocks and Capability Tracks

AI is a multidisciplinary field and even at a practitioner level requires building capabilities in fields of machine learning, natural language processing, computer vision and many others. All of these are distinct from each other. However, for a successful AI adoption strategy, an organization needs to go way beyond just AI disciplines and invest heavily on building complimentary skills without which AI initiatives will fail to meet business expectations.

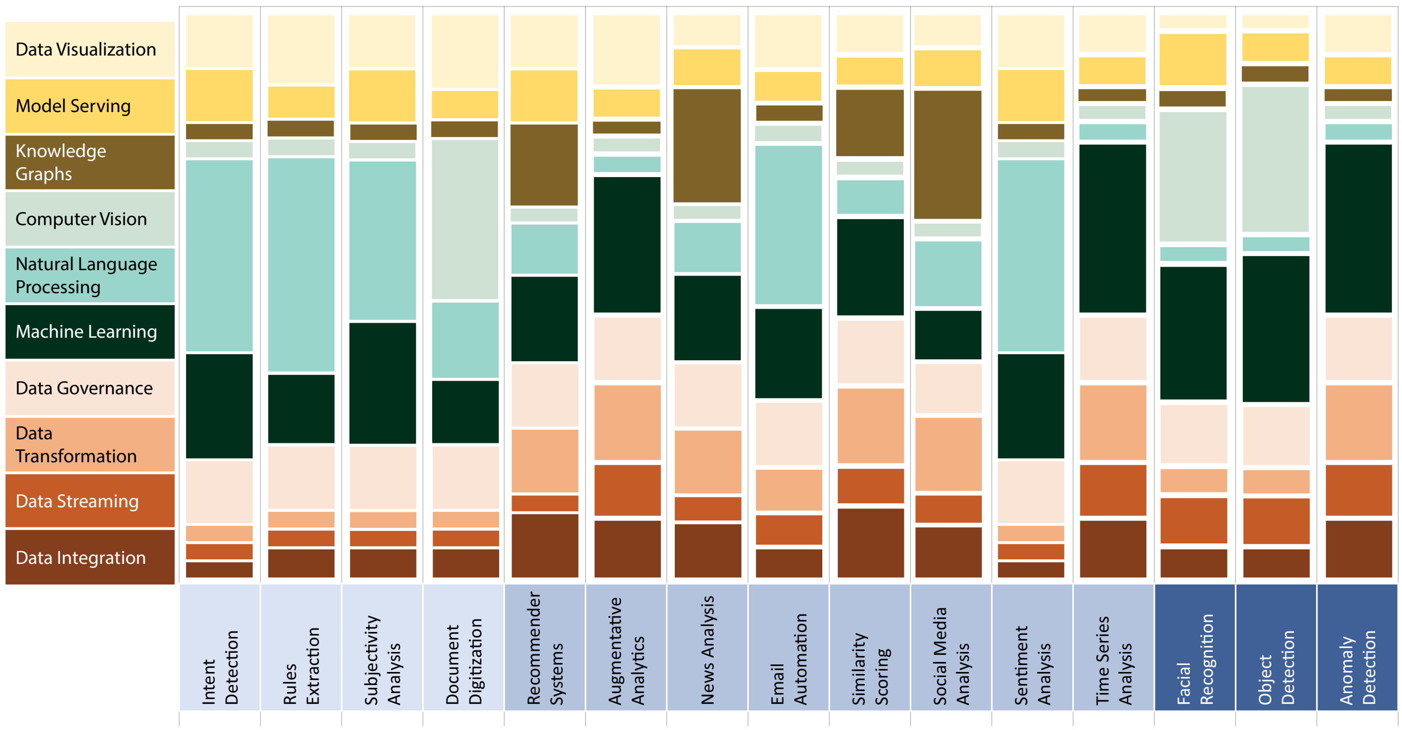


Figure 4: AI Capabilities

As depicted in figure 4, a modern data infrastructure forms the bedrock on which an edifice of AI will stand. Data is a vast domain and would require a complete book to be written just to introduce its major concerns. We have aggregated the most relevant ones into four data capability tracks.

1) **Data Integration** establishes a robust infrastructure for movement of data from source systems to data lakes.

2) **Data Streaming** aims to achieve the same in real-time.

3) **Data Transformation,** especially in the context of machine learning, is converting raw data into features for training models.

4) **Data Governance** is the overarching construct that regulates the discovery and movement of data across platforms.

The core AI capability tracks of ML, NLP and CV sit atop this data foundation. We have added 3 capability tracks that complete the picture.

1) **Knowledge Graphs** are repositories of information stored in form of relationships between entities. This is a powerful representation that augments AI capabilities for several building blocks.

2) **Model Serving** brings the ability to expose inferencing as a service. Several elements are involved here, ranging from containerized microservices or cloud-based deployments to automated model deployments (called MLOps).

3) **Data Visualization** targets creating an enhanced user experience, whether for analytics or for taking actions.

These capability tracks as best set up as technology COEs serving innovation pods. Innovation pods themselves can be set up as agile teams delivering POCs over 4 to 6 sprint cycles.

Determining ROI of AI use cases

Quick notes on steps to be followed for determining ROI:

Investments –	Returns –
1. Determine constituent building blocks (as per figure 3) 2. Determine capability tracks that serve each constituent building block (as per figure 4) 3. Seek costs (effort, infrastructure) from each capability track 4. Aggregate	1. Determine technology metrics to measure model performance (e.g. prediction accuracy or false positives) 2. Establish baseline performance for these metrics 3. Establish a mechanism to translate incremental improvement in performance to business terms (e.g. cost savings, increase in sales) 4. Measure model performance 5. Determine returns in business terms

For existing use cases, baseline is the performance of the existing rule-based implementations. For new use cases it is tempting to go with the state-of-the-art models. However, SOTA performances are achieved in labs working on curated data and will not be replicated in real-world scenarios. It is a lot more prudent to go with a heuristics-based model based on reasonable expectations.

External Document © 2021 Infosys Limited

Capability Tracks and Technology Stacks

Organizations can build their technology solutions by

1) Building them from scratch
2) Buying products that offer those capabilities, or
3) Subscribing to cloud-based services.

For the first option, industry is opting for open source softwares, where experimentation is encouraged on community editions, but production deployments are on enterprise editions. Buying products that offer out-of-the-box functionality is generally a declining trend, principally because such products have failed to deliver on their promises, and clients are vary of getting locked into a vendor. This has resulted in most product vendors adopting cloud-based subscription model. However, traditional vendor products are not fully cloud-native and neither fully benefit from auto-scaling nor would necessarily provide integrated digital experiences. FinTechs had beaten them hands down, but do not yet have the experience for support enterprise-scale implementations. Traditionally organizations faced the "buy vs. build" dilemma. Cloud complicates this further, as organizations are faced with the task of double decisions – what workloads to move to cloud and whether to build cloud-native solutions or simply subscribe to them.

	Open Source	Cloud Service Providers				FinTechs/Partners
Data Visualization	Angular, React, D3, Kibana, Grafana, Streamlit, Plotly, Tableau Public	QuickSight, Athena	Data Explorer	Data Studio	NA	Whatfix
Model Serving	Flask, TFServing, Docker, Kubernetes, MLFlow, KubeFlow	SageMaker	Azure ML	Cloud AutoML	OpenShift	Altair
Knowledge Graphs	Neo4j, ArangoDB, GraphDB	Neptune	Microsoft Graph	Knowledge Graph	NA	StarDog, OntoText
Computer Vision	OpenCV, Tesseract, Detectron	Textract, Rekognition	Computer Vision	VisionAI, DocAI	OpenShift AI/ML	Gieom, Signzy
Natural Language Processing	NLTK, Spacy, Flair, Rasa, BERT, Stanford CoreNLP,	Comprehend, Lex	Synapse Analytics, Bot	Natural Language	OpenShift AI/ML	Clinc, Acitve.AI, Ailleron, Glia
Machine Learning	Scikit-learn, Tensorflow, Keras, Pytorch, Spark Mllib	Amazon SageMaker	Azure ML Studio	Cloud AutoML	OpenShift AI/ML	Cognino, Fiddler
Data Governance	Apache Atlas, TrueDat, XAI	DIDL	Purview	Data Catalog	DataOps	DataRobot
Data Transformation	Python, Pandas, Numpy, Apache Spark, Talend OpenStudio	Glue	Data Factory	OpenRefine	Fuse	Elastic
Data Streaming	Apache Kafka, Storm, Flink	Kinesis	Stream Analytics	Stream Analytics	AMQ Streams	NA
Data Integration	Apache Airflow, Camel, Nifi, LogStash	Data Pipeline	SSIS, Data Factory	Data Fusion	Data Integration	Couchbase, MongoDB

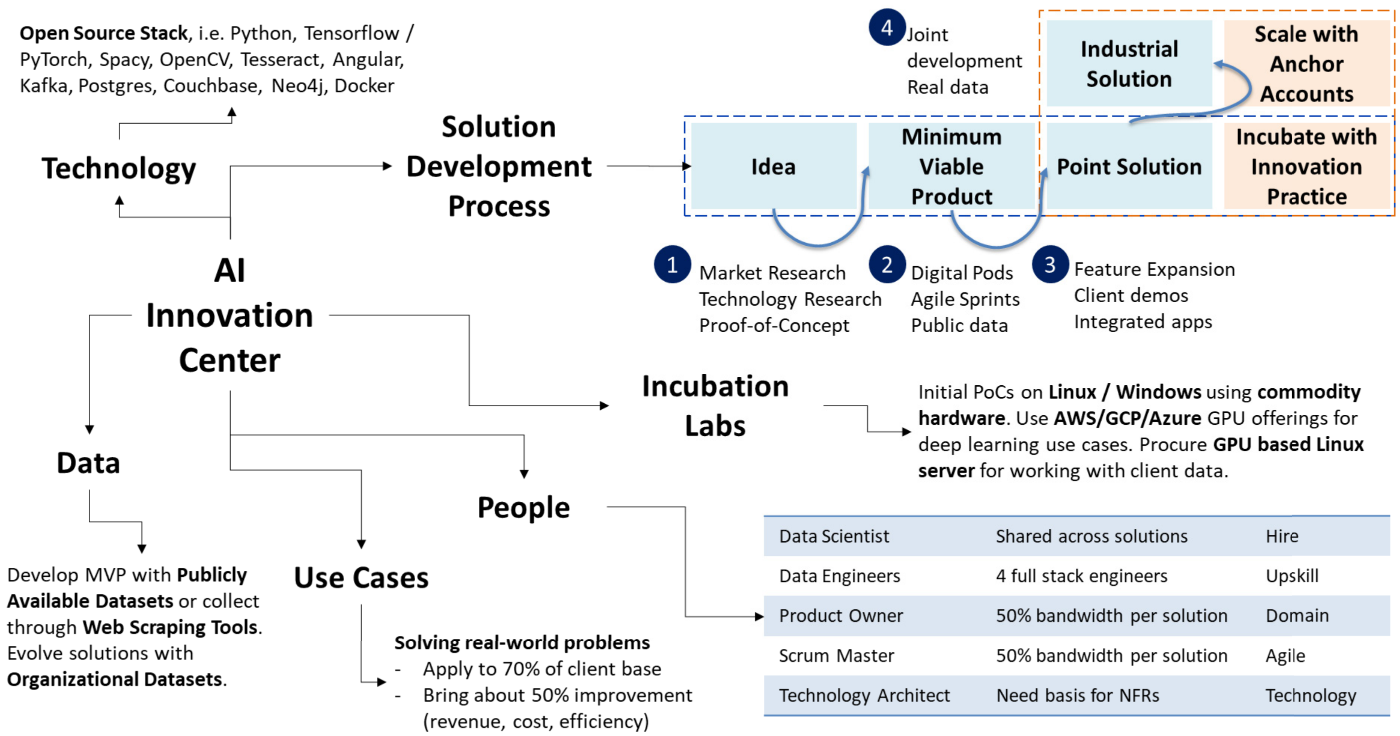
Figure 5: AI Skills

Figure 5 lists some of the industry leading frameworks, libraries, services and products for each of the capability tracks. From this view, it becomes evident that the complete AI stack spans a vast spectrum of technologies. While this appears daunting, segregating these technologies into capability tracks helps in clustering related technologies together. Once the team (or even a team member) has mastered one of the skills within a track, picking up the adjacent skills becomes a relatively simpler exercise.

External Document © 2021 Infosys Limited

Innovation model for AI POCs

A formal framework that industrializes innovation is essential to embrace emerging technologies. It is best to experiment with the expectation that most POCs will never graduate from incubation labs. These must be executed in the "fail-fast" spirit.



Data Scientist	Shared across solutions	Hire
Data Engineers	4 full stack engineers	Upskill
Product Owner	50% bandwidth per solution	Domain
Scrum Master	50% bandwidth per solution	Agile
Technology Architect	Need basis for NFRs	Technology

Figure 6: AI Incubation Framework

Design-thinking led ideation brings in customer empathy, the most vital ingredient, to the understanding of problem statements. A product-backlog based approach translates these problem statements into actionable user stories. Build happens in isolated incubation labs that allow free access to the internet for downloading open source softwares and publicly available datasets. Teams are grouped into pods that execute timeboxed POCs. Team members are encouraged to scour research papers and apply the learnings to solve given problems.

External Document © 2021 Infosys Limited

Productionizing AI Projects

Models are binary objects that get deployed in production environment, very much like compiled source code. But that's where the similarity ends. Models learn patterns from a training dataset and depend on a set of additional parameters that cannot be learnt from the dataset (called hyperparameters). Consequently, every model links back to the training dataset and the set of hyperparameters. Over time, multiple variants of the same model get built by varying either the training dataset or hyperparameters. Moreover, model performance tends to degrade over time as datasets evolve (called model drift). Ensuring that always the best performing model runs in production, through a champion-challenger model, requires a robust model management framework.

External Document © 2021 Infosys Limited

Conclusion

To summarize, it is imperative that financial services organizations need to develop a well-thought out enterprise AI adoption strategy. Our key recommendations include:

- Establishing the four patterns of coexisting AI adoption models
- Prioritizing AI use cases on the basis of "Degree of Autonomous Behavior"
- Creating a foundational layer of AI "Building Blocks" that can be used to compose solutions for one or more business use cases
- Setting up pods as "Capability Tracks" that work on AI POCs
- Rolling out a structured and measurable culture of innovation

We believe that Financial Services firm will continue to bring AI to bear fruit unless they adopt a robust strategy as outlined above. This paper presents a generic view that applies to all AI projects. NLP is of specific importance for FS as vast volumes of data is present in the form of unstructured text. In a subsequent paper, we plan to address concerns specific to NLP adoption in FS organizations.

External Document © 2021 Infosys Limited

About the Author

Amitabh Manu
Delivery Manager, Infosys

He is leading Innovation for Financial Services.

For more information, contact askus@infosys.com

Infosys®
Navigate your next

© 2021 Infosys Limited, Bengaluru, India. All Rights Reserved. Infosys believes the information in this document is accurate as of its publication date; such information is subject to change without notice. Infosys acknowledges the proprietary rights of other companies to the trademarks, product names and such other intellectual property rights mentioned in this document. Except as expressly permitted, neither this documentation nor any part of it may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, printing, photocopying, recording or otherwise, without the prior permission of Infosys Limited and/ or any named intellectual property rights holders under this document.