Powered by
IEEE SA

IEEE Finance Playbook Version 1.0

Trusted Data and Artificial Intelligence Systems (AIS) for Financial Services

IEEE

NOTES ABOUT THIS VERSION

Public comments are invited on the **IEEE Finance Playbook Version 1.0: Trusted Data and Artificial Intelligence Systems (AIS) for Financial Services:** An industry-specific implementation playbook that encourages technologists in the financial services to prioritize human well-being and ethical considerations in the applications Data and Autonomous and Intelligent Systems (AIS).

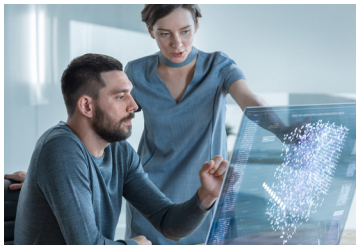
To our knowledge, this is the first global industry-specific Ethical AI standards and certification initiative that brings together financial institutions, academia, legal and compliance, technology and services providers, and fintech companies to accelerate the adoption of IEEE's Ethically Aligned Design, P7000 standards, and The IEEE Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS) certifications.

This document has been created by committees of the IEEE Trusted Data and Artificial Intelligence Systems (AIS) Playbook for Finance Initiative composed of 50+ industry participants from Canada, the US, and the UK, who are thought leaders from banks, credit unions, pension funds, law firms, academia, and technology services organizations in the related risk management, legal and compliance, data governance, data and analytics, systems development, and sustainable business development disciplines to identify and find consensus on timely issues. For Version 1.0, all six Canadian major banks (i.e., RBC, TD, CIBC, BMO, Scotiabank, and National Bank) and representatives from credit unions (i.e., PPJV), pension funds (i.e., OMERS), and fintechs have participated in this initiative.

The document's purpose is to:

1. Curate, summarize, and contextualize trusted data and AIS implementation high level requirements and best practices for financial services Line of Business (LoB) executives; Digital, Innovation, and Transformation Program executives; Chief Risk Officers (CRO); Chief Technology Officers (CTO); Chief Data and Analytics Officers (CAO/CDO/CDAO); Chief Information Officers (CIO); and other executive sponsors with relevant mandates.

2. Provide a trusted data and AIS sandbox/community to help financial services organizations learn more about the *Ethically Aligned Design* (EAD) methodology, standards (e.g., IEEE P7000), and certification programs (e.g., ECPAIS) for high value AIS use cases, digital applications, and intelligent workflows.



3. Organize participation in quarterly benchmarking surveys, a trusted data and AIS sandbox, best practices community calls, and the latest version of the IEEE Finance Playbook.

How to Get Involved

The IEEE Trusted Data and Artificial Intelligence Systems (AIS) Playbook for Finance Initiative invites comments for Version 1.0, the IEEE Finance Playbook provides the opportunity to bring together diverse voices from the related finance and data science communities, to generate a broad consensus on pressing ethical and societal issues as well as welcome recommendations on the development and implementation of AIS technologies.

Input on the IEEE Finance Playbook Version 1.0 should be sent by email no later than 1 May 2021 and 1 June 2021. These collected responses will be made publicly available at the website no later than 1 June 2021.

Email: financeplaybook@ieee.org
Website: https://standards.ieee.org/industry-connections/ais-finance-playbook.html

Details on how to submit comments are available via our Submission Guidelines, provided in Appendix A. Publicly available comments in response to this request for input will be considered by committees of The IEEE Trusted Data and Artificial Intelligence Systems (AIS) Playbook for Finance Initiative for potential inclusion in the final version.

Companion Survey

Another way to participate in the development of the playbook is to participate in our Trusted Data and AIS Readiness Survey. Take the 20-minute anonymous online survey here to receive an evaluation of your organization's Trusted Data and AIS Readiness and guidance on how best to use the playbook. An anonymous summary of results will be included in future versions.

If you are a journalist and would like to know more about the IEEE Finance Playbook, please contact the IEEE SA PR team.

Support Statements from the Executive Steering Committee and Key Contributors

"Ethical deployment of Data and Technology is at the heart of Open Banking and ESG Transformation for Financial Services. In this IEEE Finance Playbook global initiative, we will continue to curate, survey, summarize, and contextualize trusted data and AI implementation high level requirements and best practices for financial services. Please join our journey to develop a global consensus-based standards and certifications program for high value AI applications and intelligent workflows."

Pavel Abdur-Rahman
Head of Trusted Data and AI, IBM Canada

"We are in the business of trust. A primary goal of financial services organizations is to use client/member data to generate new products and services that deliver value. Best in class guidance assembled from industry experts in IEEE's Finance Playbook addresses emerging risks such as bias, fairness, explainability, and privacy in our data and algorithms to inform smarter business decisions and uphold that trust."

Sami Ahmed
SVP Data and Advanced Analytics (DNA), OMERS

"Data ethics aren't a cost of doing business, they are an investment in good business. We have a responsibility to set ethical standards that ensure transparency, minimize biases, and encourage accountability."

Terry Hickey
SVP and CIO Enterprise Data, CIBC

"We are at a critical junction of industrial scale of AI adoption and acceleration. This IEEE Finance Playbook is a milestone achievement and provided a much needed practical roadmap for organizations globally to develop their trusted data and ethical AI systems."

Amy Shi-Nash
Global Head of Analytics & Data Science, HSBC

"Upholding customer privacy and trust is paramount. The acceleration of digital and AI technologies makes it increasingly important and complex. It is critical to install practical and scalable approaches across the AI development lifecycle to align with corporate values, customer expectations, and regulatory requirements. The right balance between risk and innovation will accelerate AI adoption in a responsible way."

Dr. Ren Zhang
Chief Data Scientist and Head of AI Center of Excellence, BMO

"Like all disruptive technology, Artificial Intelligence offers both great promise and the risk of great peril; and the increasing sophistication of AI has rightly sparked a vibrant debate on how best to balance the two. As AI systems become more ubiquitous, trust in these systems will be critical and organizations will need to demonstrate that they have thoughtfully considered the unique risks of AI and responded in a manner that is consistent with their corporate values and aligned with the expectations of clients, employees, and society. This Playbook will help organizations and practitioners understand the rapidly evolving landscapes of Trusted Data and Ethical AI and offers a pragmatic approach for ensuring ethical principles are built into AI systems from the start."

William Stewart
Head of Data Use and Product Management, Data and Analytics (DNA), Royal Bank of Canada (RBC)

"The IEEE Playbook, and it's wealth of knowledge from a consortium of thought leaders, is a must-read for companies and practitioners alike wanting to use AI to make better, more trusted decisions."

Mark Wagner
VP of Advanced Analytics and AI, Scotiabank

"In financial services, deployment of AI technologies is both an opportunity and a responsibility. It can positively transform client and employee experience, and augment core capabilities of organizations. We also have a key role to play in deploying AI technologies responsibly and ethically, in order to maintain trust of clients and other stakeholders.

IEEE has successfully brought the ecosystem together to tackle important questions, an essential component in helping the financial industry continue its journey to set exemplary standards for responsible and ethical AI."

Mathieu Avon
VP, Integrated Risk Management, National Bank of Canada

"Our world is being shaped by some increasingly interconnected trends around data and technology availability, open ecosystems and privacy. They are accelerating individually and amplifying each other at the same time and creating uncharted waters which are difficult to navigate. The IEEE Finance Playbook provides a framework and guidelines built on top of strong ethical foundations and it will help and enable organizations to become more confident and good data actors driving economic and social value. I am delighted to be able to support this agenda."

Vilmos Lorincz
Managing Director, Data and Digital Products, Commercial Bank, Lloyds Banking Group

"The cornerstone of innovation within banks and financial institutions will be rooted in the ethical use of data and artificial intelligence, to foster and maintain a trusted relationship with customers. Putting ethics and customer protection at the centre of how we develop and use technology is an enabler of innovation, not a barrier or a box to check off. Leveraging consortium guides such as IEEE's Playbook for Financial Services is pertinent for our business to stay abreast of ethical standards required to consistently protect our stakeholders and the communities we serve."

Elizabeth Chacko
VP Data and AI Risk Scotiabank

"Financial organizations need real, concrete, and practical advice tailored to their AI system, informed by the full knowledge of the legal and ethical risk they face. IEEE's Finance Playbook responds to this need and provides timely guidance on how the financial sector can continue innovating responsibly."

Carole Piovesan
Partner, INQ Law

"As Artificial Intelligent Systems are more and more widely adopted throughout organizations, technologists in those organizations need to be keenly aware of the bias, fairness, privacy, and other ethical issues at play. The IEEE Finance Playbook is the definitive guide for financial organizations to identify and resolve ethical issues in their AI Systems."

Dr. Stephen Thomas
Executive Director Analytics and AI Ecosystem, Smith School of Business, Queen's University

"Transparency generates trust, which is a prerequisite for Artificial Intelligent Systems contribution to economic and social development. This needed playbook reflects on all ethical aspects of trusted data and AI. It guides business leaders to execute an all ethical principles learning from real-life case studies in banking and financial markets. This collaborative effort is an inspiring approach for financial services internationally."

Paolo Sironi
Global Research Leader, Banking and Financial Markets, IBM Institute for Business Value

"Ethical AI practices that remove bias from platforms are critical to ensuring that everyone participates in this fast-growing innovation economy. An inclusive future is a prosperous future."

Armughan Ahmad
Managing Partner and President, KPMG

"This playbook is the pragmatic result from a deep collaboration across competitors and their ecosystem partners to drive from the principles of AI ethics to their practice...and then to put those practices into action. It was a multi-disciplinary, multi-stakeholder effort with a stated direction toward concrete execution and outcomes-- reflecting a broader desire to shift the AI ethics dialogue from outside conference and academic environs into the boardrooms, executive suites, IT labs, and even front-line operations of those companies implementing AI today-- that can serve as beacon for similar companies in other regions and even other industries."

Brian C. Goehring
Associated Partner, AI/Cognitive and Analytics Lead, IBM Institute for Business Value

"The IEEE Trusted Data and AI Playbook for Financial Services provides an accessible and usable directory of resources that supports AI practitioners as they encounter any broad range of challenges and contexts. By incorporating the playbook into our advisory and delivery methodologies, GFT is able to tap into the experiences and lessons learned by a diverse and innovative community, extending our capabilities to offer ethically aligned high-quality business AI solutions for financial services sector clients. We look forward to the further development of the playbook as the community of contributors diversifies and widens further, and the development and adoption of AI grows."

Dr. Simon Thompson
Head of Data Science, GFT Technology, UK

"With AI you can disrupt your industry, but if you don't mitigate unwanted bias and ensure fairness, your FinTech company runs the risk of being part of cancel culture."

Beth Rudden
IBM Distinguished Engineer and Principal Data Scientist,
Cognitive and AI, Trustworthy AI Center of Excellence

"As global commerce races toward a future of pervasive AI, business leaders must contend with the fact that trustworthy AI requires a trustworthy organization – more humane organization. To keep our AI capabilities in lockstep with our ethical frameworks we must gain insights into human capabilities that are still lacking in AI such as adaptability, abstraction, and common sense. We'd like to imagine that when the AI 'Black Box' will be pried open it will shine a light, more likely, it will shine a mirror. The IEEE Finance Playbook is a multi-disciplinary guidebook on how to start this fundamental conversation."

Tomer Meldung
Solution Executive, Data and AI, IBM Canada

"As revolutionary as Automated Intelligent Systems is poised to be, the risks associated with its development and adoption need to be addressed in equal measure. What an honour to collaborate with global thought leaders in the AI Ethics space delivering the Playbook for Financial Services. I'm confident in the value-add this playbook offers to organizations looking to propel their trusted data and AIS business practices for now and future generations."

Cindy Pham
Senior Manager, AI Risk, Scotiabank

"As AI adoption accelerates in the industry, Financial Services businesses find themselves at the precipice of a yawning well of innovation and opportunity. The industry is in a unique position to drive its own AI-led transformation as it undertakes its next wave of digitalization, faces algorithm-driven market disruptors, and proliferates goldmines of data—the fuel of the AI furnace! Yet, optimizing operations, generating insights, and intelligently automating processes is not enough. Organizations must marry these new powers with heightened responsibility, risk mitigation, and humanitarian values. Infusing ethics and trust into AI, IEEE can help drive Financial Services toward this vision."

Devan Leibowitz
Senior Management Consultant, Monitor Deloitte US

"We are now entering an era of operational ML models which provide real tangible yield. With this comes an increasing complexity of production grade models at scale needing risk mitigation.

The financial industry is specially positioned to de-risk negative expectancy by setting the gold standard for operational models. Cross domain experts provide significant value-add by leveraging their knowledge of compliance, ethics, KPI metrics, and regulatory oversight. The intersectionality of MLOPs and financial services is a perfect match to enter this critical phase. "

Wish Bakashi
Applied Machine Learning and AI Computer Scientist, Capital Power

"This playbook represents the diverse knowledge and collaborative approach needed for the complexity of developing trusted Artificial Intelligent Systems. By offering a practical approach to building these systems, we move toward real-world outcomes that promote human well-being and ethical considerations. The IEEE Trusted Data and AIS Playbook provides a unique opportunity for the financial industry to foster a responsible and inclusive future."

Tania De Gasperis
Responsible IoT/XR Researcher for ACE Lab at OCAD University, previously AI Ethics Researcher for Montreal AI Ethics Institute

"It is remarkable how rapidly advancements are being achieved in the realm of AI and Data Science. They have raised fundamental questions about what we should do with systems, what systems should accomplish, and what risks and externalities they raise in the short and long term. The IEEE Finance Playbook comes at a much-needed time to provide guidance and an approach to complex decisions that are meant to have lasting impacts."

Wendi Zhou
Manager, Strategic Research and Analytics, Borden Ladner Gervais LLP (BLG)

"The promise of a prosperous, fair, and uplifting future for banking cannot be realistically conceived without an unwavering commitment to ethics. As new technologies embolden us with possibilities, our responsibility to understand and govern these tools for the benefit and well-being of humanity becomes a shared priority. The IEEE Trusted Data and AIS Playbook is an honest, grounded, and rigorous attempt to initiate and guide practitioners into the complex world of ethics for intelligent and autonomous systems, so that together we may uphold and operationalize the virtues of trust and justice within and beyond the Financial Services industry."

Daniel Gomez Seidel
Senior Manager, Design Strategy, Capital One US

"As an ethical Islamic Fintech, it is imperative that our clients and partners are assured that we are committed to halal and ethical business practices. This includes and is not limited to the technologies we incorporate into our platform that collect and store data, as well as how that data is used to provide an exceptional customer service experience. AI Ethics will play a key role in attracting underserved and underbanked communities and will be an important piece in future open banking initiative frameworks."

Mohamad Sawwaf
Co-Founder and CEO, Manzil, Halal Financing and Investments

"At RateCo, we use our customers data in an ethical manner to better predict their future mortgage needs in order to offer them better financial solutions today. In addition to our regulatory compliance requirements, we feel this is the only way to bring transparency to the private mortgages marketplace."

Ajay Jain
CEO Rateco.ca

CONTENTS

Section 1: The Roadmap to Trusted Data and AIS

Section 2: Key Resources for the Roadmap

SECTION 1:
The Roadmap to Trusted Data and AIS

Introduction

Defining and Drafting a Treatise for Trusted Data and AIS for Financial Services

As the use and impact of autonomous and intelligent systems become pervasive, we need to establish societal and policy guidelines in order for such systems to remain human-centric, serving humanity's values and ethical principles. These are socio-technical systems and must be developed and should operate in a way that is explicitly beneficial to people and the environment, in addition to reaching functional goals and addressing technical problems. This approach will foster the heightened level of trust between people and technology that is needed for its fruitful use in our daily lives.

From the introduction to *Ethically Aligned Design*, First Edition

COVID-19 has fundamentally altered the way humans trust information and the institutions that track it. Human data, the fundamental building block for AIS[1], is currently being accessed and shared in multiple ways in every country of the world without unified policy, technology, or cultural guidelines.

Yet we know that our data directly reflects our identity. And it is our identity combined with how our data is shared that determines our worth in the algorithmic age. More than our credit or the specie reflecting our wealth, it is the ability to access and share our data that is the seminal innovation required in the algorithmic age to keep people at the center of how they perceive and curate their truest selves.

This is the vision of this first draft of the IEEE Finance Playbook—to holistically define and draft a treatise for Trusted Data and AIS for Financial Services. This is also why The IEEE Trusted Data and Artificial Intelligence Systems (AIS) Playbook for Finance Initiative was created to harness the collective power of the global AIS Finance industry to offer this actionable document with a survey and invitation to make it as specific, relevant, and actionable as possible.

[1] *Artificial Intelligence Systems (AIS) is the preferred term from the IEEE Standards Association for "Artificial Intelligence."*

We know that the financial sector has been a pioneer in sovereign data exchange tools and mindsets, leveraging blockchain and smart contract methodologies to begin the inevitable transition to participatory environments directly embracing the customer. Countries like Estonia and Singapore have likewise been forerunners in these new practices, demonstrating cross-pollination between empowered citizens with the financial and government entities responsible for their fiscal well-being.

In combination with these practices, countries like Canada have been among the first to prioritize human-centric, values-driven ethical methodologies for the application of AIS, for example, with their creation alongside France of the International Panel on Artificial Intelligence. Building from a basis of sovereign data, multiple financial institutions are recognizing that it is only by eliminating discriminatory bias or by increasing transparency that genuine trust with customers can flourish in the algorithmic age. This action-based trust is the central tool for cementing self-determination, thus also democracy in the digital era with tools and practices that will sustain and heal us both in our present crisis and beyond.

From Catalyst to Consensus

Version 1.0 of The Finance Playbook has been designed to harness the cutting edge thinking of dozens of financial experts in AIS working in collaboration with the thought leaders at IEEE (the world's largest technical professional organization with members in more than 160 countries) who created the globally recognized document, *Ethically Aligned Design*.

Utilized by more than a dozen global policy organizations (including the OECD, the UN High Level Experts Group, and UNICEF) and countless businesses, EAD provides a pragmatic and comprehensive guide for moving from principles to practice regarding AIS design. Rather than a focus on morals or fear, EAD provides implementable frameworks to instantiate responsible AIS at the outset of design.

Now these formative practices have been coupled with the sector specific knowledge of leading financial minds to bring you this playbook with an invitation to provide your insights via our survey (take the 20-minute anonymous online survey here) or via direct feedback to this work (see submission guidelines in Appendix A) for inclusion in future versions. We have provided the expert and informed catalyst leading toward trusted data and AI for AIS Finance—**now we need you** to move toward consensus.

Specifically, this Playbook will discuss how organizations can implement the *Ethically Aligned Design* ideology necessary to create value with data, advanced analytics, and AIS technologies. To do so, the Playbook highlights three types of critical business blocks we have observed in leading trusted data and AIS financial services firms: **technology, process, and people.** Within each of these categories there are several capabilities that organizations should focus on developing to deliver high value use cases, manage risk, and deliver value with trusted data and AIS.

The playbook will provide readers with **recommended key resources** to help develop these trusted data and AIS critical building blocks and deliver business value through high value AIS use cases, supported by robust AIS governance and reporting, and executive sponsorship and leadership. The Roadmap methodology, use cases, and resources throughout the book address a diverse set of topics of potential interest to AIS executives, designers, data engineers and scientists, developers, end users, risk and compliance teams, and governance stakeholders, who need to work collectively to embrace and advance as a community.

The leadership of IEEE Standards Association, the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, and the IEEE Finance Playbook Initiative wish to thank the Executive Steering Committee and expert volunteers for their dedication and insights in creating the **IEEE Finance Playbook**. By curating, summarizing, and contextualizing best practices within the industry, these senior thought leaders have provided key insights that will greatly expedite the efforts of their colleagues in every department of the financial organizations—most defining the nature of value for human and fiscal data.

Konstantinos Karachalios
Managing Director, IEEE Standards Association

Raja Chatlia
Chair, IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems

John C. Havens
Executive Director, IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems

Pavel Abdur-Rahman
Chair, IEEE Finance Playbook

How the Playbook Began

Our journey began in early 2019 when a collaboration with IEEE, Smith School of Business at Queen's University, Scotiabank, and IBM developed a certification program on AIS ethics principles, and the use of the technology in business decision making, applications and processes.

Daniel Moore, Group Head and Chief Risk Officer at Scotiabank summarized the inspiration behind the program: "AI will be a transformative force in business and in life, and the Bank has a duty to our customers to safeguard their trust and use AI responsibly. At Scotiabank, our investment in AI goes beyond the smart implementation of new tools and technologies, with a commitment to being leaders in the development of principles, guidelines, and training for the ethical application of this powerful technology. We are thrilled to be the first financial institution in Canada to offer a certification program specifically designed to help our leaders and senior executives better understand the framework for building trust in AI systems."

The two-day certification program delivered by the collaboration group provided a comprehensive overview of ethical AI principles for senior executives and business leaders with a focus on:

- Principles of AI and ethics in design
- Decision-making with analytics
- Dynamics of enterprise data and AI management
- Key Canadian information and privacy regulations
- Latest research and technology developments in AI
- Impacts on the future

The IEEE Finance Playbook, Version 1.0 is a continuation of this collaboration, with the hopes of broadening the availability of knowledge on Trusted Data and AIS for the financial services industry. "People are progressively more aware of what companies can do with their data and demanding transparency," said Dr. Yuri Levin, Professor, and Stephen J.R. Smith Chair of Analytics, Smith School of Business. "We hope that this IEEE Finance Playbook will improve employee awareness of the challenges and key questions they need to be asking whenever they deal with something related to AI."

Trusted Data and AIS Readiness Roadmap: How to Use This Playbook

This section features our Playbook designed to proactively and pragmatically build trusted data and AIS in your organization. It is one thing for an organization to agree to ethical principles in theory, but another thing entirely to build, execute, and scale trusted data and AIS applications. To guide this evolution and as a key feature of the Playbook, we (members of The IEEE Trusted Data and Artificial Intelligence Systems (AIS) Playbook for Finance Initiative) have developed a Trusted Data and AIS Readiness Roadmap (see Figure 1) to help guide organizations in the development of their trusted data and AIS in the near-term (1-3 years).

The Roadmap was developed based on extensive academic research, industry interviews, and our combined consulting experience. Our academic research team performed fifty in-depth interviews with trusted data and AIS ethics experts across the globe to identify the critical building blocks that allow leading organizations to build, execute, and scale trusted data and AIS applications. We spoke to a diverse set of experts at more than twenty financial services organizations, ten countries, and four continents, including data science practitioners, lawyers, compliance experts, risk experts, and analytics and data executives at deposit taking institutions, credit card companies, fintechs, as well as several regulators.

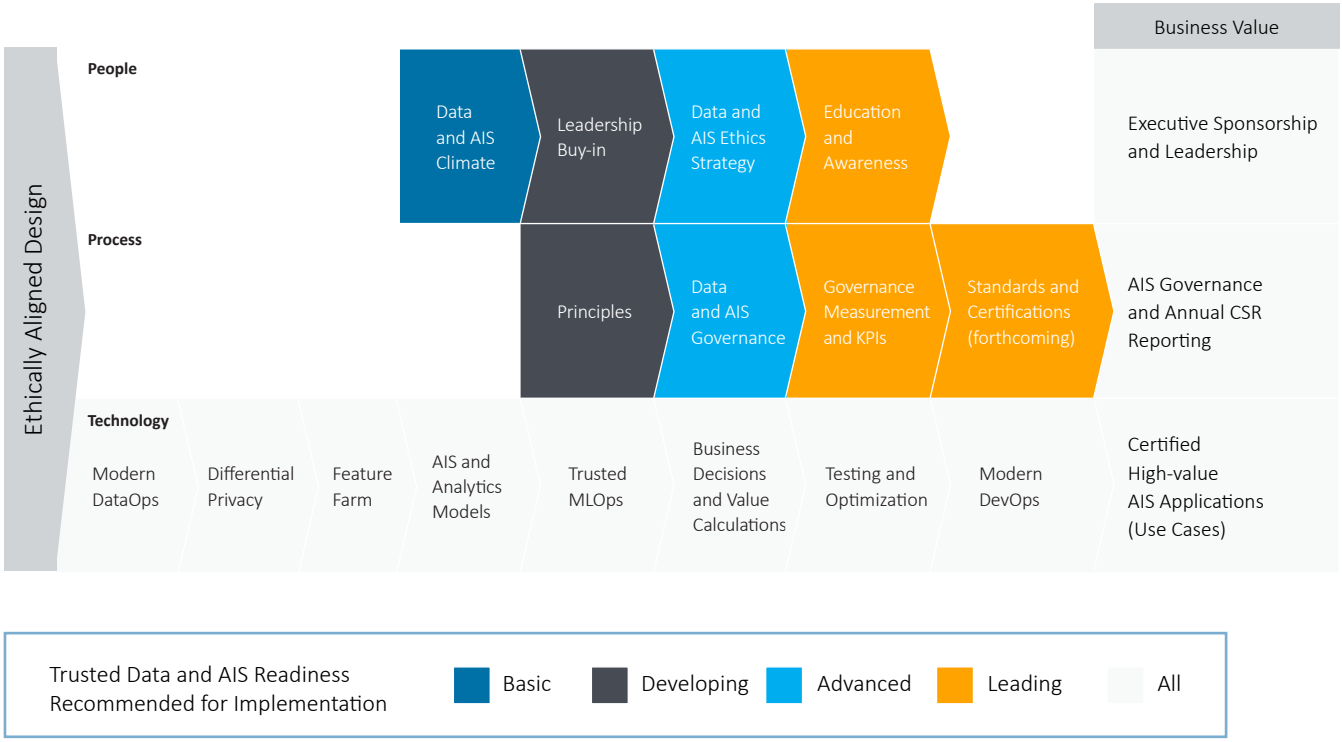
Through this work we have observed that leading trusted data and AIS organizations have in place three types of critical building blocks for Trusted Data and AIS covering people, process, and technology (see Figure 1). These critical building blocks allow financial services organizations to take the principles of ***Ethically Aligned Design*** and translate them into business value. Based on our research, in the context of Trusted Data and AIS business value means three things in the near-term (1-3 years):

1) Certified high-value AIS applications (also referred to as use cases),

2) Robust AIS governance and annual CSR reporting, and

3) Executive sponsorship and leadership.

Figure 1: Trusted Data and AIS Readiness Roadmap[2]



Take the 20-minute online survey here to determine how your organization ranks across the 7 Factors of Trusted Data and AI Ethics Readiness.

The survey results will provide you with guidance on which critical building blocks your organization should focus on in the Roadmap, and how to use the playbook resources to develop them.

[2] Several of these factors map directly to the more general AI Ethics Readiness Framework presented in *A Call to Action for Businesses Using AI*, from the Ethically Aligned Design for Business series, available here: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead/ead-for-business.pdf

Trusted Data and AIS Critical Building Blocks

As mentioned, there are three sets of critical building blocks that we have observed allow organizations to develop, execute, and scale Trusted Data and AIS; they are people, process, and technology.

People

Operational efforts focused on technology and process by themselves are not enough to ensure trusted data and AIS; leading organizations also must focus on developing ***people*** critical building blocks: a strong data and AIS climate, leadership buy-in, a clear data and AIS strategy, and education and awareness across the organization on the trusted data and AIS initiatives. These three levels of Trusted Data and AIS best practices are observed across all leading financial institutions, regardless of the level of data and AIS regulation in the home country. Speaking to firms at various levels of readiness allowed us to determine a common cadence of evolution, which is illustrated in the Trusted Data and AIS Readiness Roadmap in Figure 1.

Process

In addition to technical critical building blocks, leading organizations have in place trusted data and AIS ***processes***; specifically, a series of data and AIS ethics governance tools including principles, data, and AIS impact assessments, as well as measurements and KPIs. Standards and certifications are also starting to appear (i.e., IEEE std 7010™, ECPAIS), which will allow organizations to certify both their technical and governance initiatives for trusted data and AIS.

Technology

Institutions must have the appropriate ***technology*** critical building blocks to scaling trusted data and AIS projects. Through our industry and academic work, we have observed there to be a consistent scaling process that leads to the successful implementation of value driving AIS projects. This technical process is platform agnostic and should be applied to every single AIS project initiated in an organization to ensure it is being conducted ethically with trusted data and AIS in mind. The process includes modern DataOps, differential privacy, a feature farm, AIS and analytics models, trusted MLOps, business decisions and value calculations, testing and optimization, and modern DevOps. Within each of these capabilities some leading organizations are even working to develop consistent workflows to standardized operations across projects. Technical scaling capabilities by themselves however are not enough to ensure trusted data and AIS.

Trusted Data and AIS Readiness Levels

Given a firm's development across each of the three sets of critical building blocks (people, process, technology), it will fall into one of four trusted data and AIS readiness categories: basic, developing, advanced, or leading.

Basic organizations may not yet have started or will have just embarked on the evolution of their business architecture using AIS. Those who have started the journey will have an emerging data and AIS climate, and perhaps some core technical scaling capabilities, but will otherwise be underdeveloped across the three critical building blocks of people, process, and technology.

Developing organizations will have started the trusted data and AIS journey and will likely have some core technology scaling capabilities, along with a strong data and AIS climate. They are likely in the process of gaining leadership buy-in and/or developing principles but have yet to operationalize those into more formal governance or large-scale implementations.

Advanced organizations have taken several steps to evolve their business architecture using AIS, and most notably have strong leadership buy-in, principles, in addition to a strong climate and developing technical scaling capabilities. These organizations are likely refining their data and AIS strategy and governance procedures but are still lacking widespread education and/or awareness on trusted data and AIS and are not measuring their governance initiatives.

Leading organizations are well on their way to transitioning to a new business architecture powered by AIS, with trusted data and AI ethics initiatives in place across each of the three critical building blocks, most notably they are measuring and using KPIs for their trusted data and AIS governance, are educating and creating awareness for the initiatives across the organization, have well developed technical scaling capabilities (i.e., they may have implemented standardized workflows for each capability), and are working with certifications and standards bodies.

Note that the Trusted Data and AIS Readiness Roadmap (Figure 1) was developed given the initiatives in place across leading firms in 2020, so as the use of AIS and data evolves, so too will the roadmap and accompanying survey tool.

High Value AIS Use Cases for Financial Services

Recently we observed through our research and industry work a massive increase in the development, use, and implementation of AIS applications. The use of AIS has helped organizations deliver incremental value through several avenues including engaging customers, increasing share of wallet, customer acquisition, reduced credit loss, and improved profitability. Here we discuss the top twenty highest value delivering use cases for AIS observed throughout the global financial services industry.

The high value use cases are classified into one of four business categories:

Product and Customer

Risk

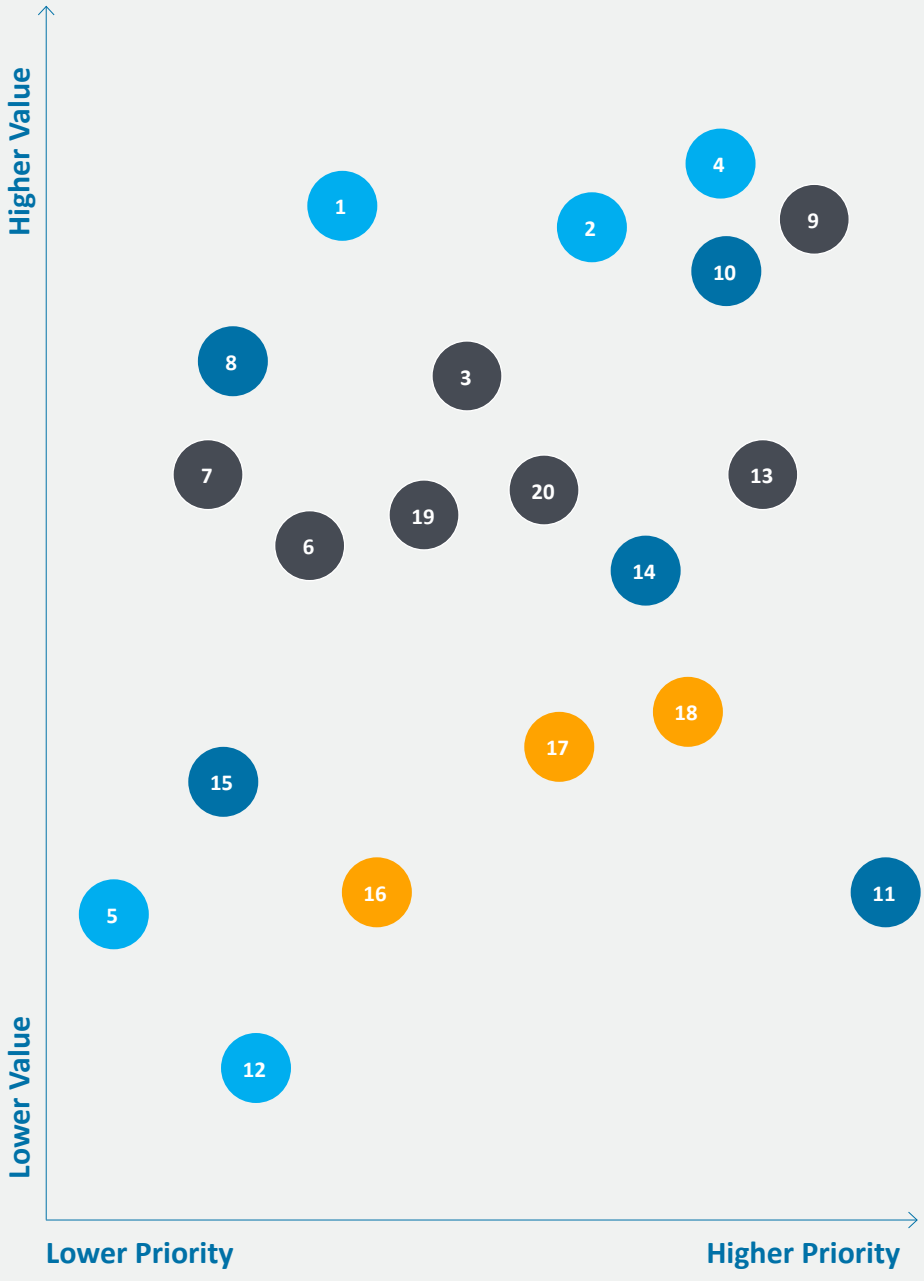
Operations

Corporate

We have observed through our industry research studies and consulting activities that organizations need to have evolved to the appropriate level of Trusted Data and AIS Readiness for effective and ethical implementation for particular AIS use cases. We present the high value use cases below in a value-priority map, accompanied by the level of readiness recommended for their implementation in Figure 2.

Value-Priority and Readiness Summary

Figure 2: Value-Priority Mapping of Top 20 High Value AIS Use Cases



		Use Cases
Product and Customer	1	Personalized Marketing Offers
	2	Next Best Action
	3	Loan and Deposit Pricing
	4	Credit Adjudication
	5	Customer Sentiment Tracking
	6	Customer Lifetime Value
	7	Customer Segmentation
Risk	8	High-frequency Trading/Robo-Advisors
	9	Cybersecurity
	10	Fraud Detection
	11	AML
	12	Model Validation and Bias Detection
Operations	13	Robotic Process Automation
	14	Operational Efficiencies
	15	Expense Management
Corporate	16	Talent Acquisition
	17	Talent Retention
	18	Audit
	19	Collections
	20	Customer Service

Trusted Data and AIS Readiness Recommended for Implementation — Basic, Developing, Advanced, Leading

Take the 20-minute online survey here to determine how your organization ranks across the 7 Factors of Trusted Data and AI Ethics Readiness.

The survey results will provide you with guidance on which critical building blocks your organization should focus on in the Roadmap, and how to use the playbook resources to develop them.

Next, we highlight the major ethical concerns that organizations need to consider in their implementation of these use cases, categorized by the EU's Assessment List for Trustworthy AI (ALTAI) principles in Figure 3. Additional details on the ALTAI principles can be found in Section 2: Key Resources for the Roadmap (Ethical Principles Used Throughout the Playbook, and the Global and Canadian Regulatory Landscape). Detailed descriptions of each use case, including the suggested level of readiness required for implementation, and execution concerns are also provided in Appendix C.

Figure 3: Key Ethical Concerns of the Top 20 High Value AIS Use Cases

	Use Cases (ALTAI Principles)	Human Agency and Oversight	Technical Robustness and Safety	Privacy and Data Governance	Transparency	Diversity, Non-discrimination and Fairness	Societal and Environmental Well-being	Accountability
Product and Customer	Personalized Marketing Offers			✓	✓	✓	✓	
	Next Best Action			✓	✓	✓	✓	
	Loan and Deposit Pricing				✓	✓	✓	
	Credit Adjudication				✓	✓	✓	
	Customer Sentiment Tracking	✓		✓			✓	✓
	Customer Lifetime Value			✓		✓	✓	
	Customer Segmentation			✓		✓	✓	
Risk	High-frequency Trading / Robo-Advisors	✓					✓	
	Cybersecurity		✓	✓			✓	
	Fraud Detection		✓	✓		✓	✓	
	AML		✓				✓	✓
	Model Validation and Bias Detection		✓		✓	✓		✓
Operations	Robotic Process Automation	✓	✓					
	Operational Efficiencies	✓		✓				
	Expense Management	✓				✓		✓
	Talent Acquisition			✓	✓	✓	✓	
Corporate	Talent Retention			✓	✓	✓	✓	
	Audit		✓	✓				
	Collections	✓		✓		✓	✓	
	Customer Service			✓		✓	✓	

Trusted Data and AIS Readiness Recommended for Implementation — Basic | Developing | Advanced | Leading

For more information on the implementation of these use cases, we recommend reviewing the key resources provided in Section 2, particularly those in the section titled "Developing the Critical Building Blocks: Key Resources". Each resource is tagged with the same ALTAI principles and the level of trusted data and AIS readiness we recommend for their implementation.

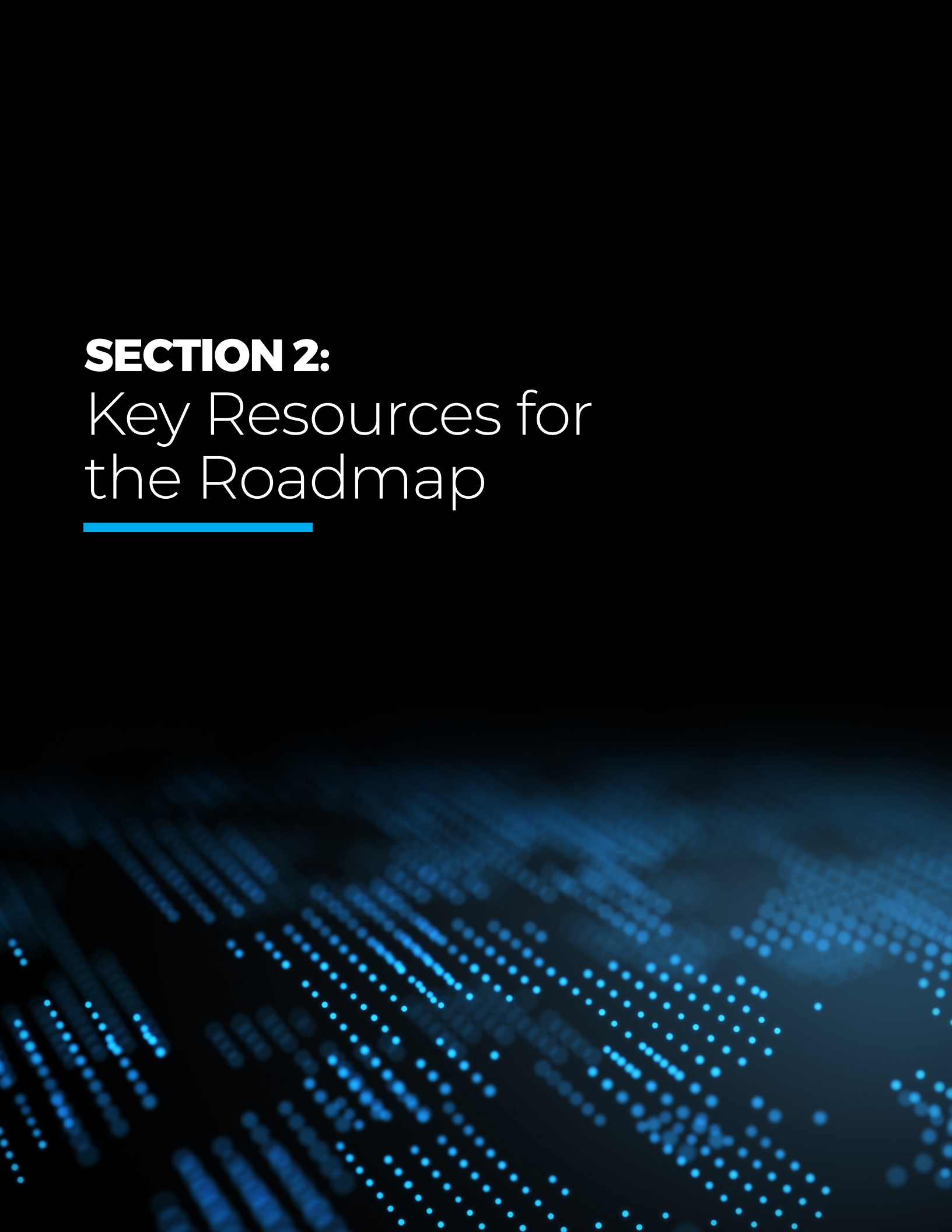
Roadmap Conclusion

Featured as part of the larger Playbook, this Roadmap is designed to provide a key tool to help you identify where you are in the process of adopting responsible, ethically aligned design at the outset of any AIS development in the financial sector.

While the landscape may seem challenging, we used the analogy of a map to help you feel empowered at what may be the beginning or middle of your journey. The entire AIS/Ethics sector is so nascent that the legislation and best practices are being discovered and applied by the cutting edge thought leaders in the space.

People like you. By simply identifying where you and your organization are in terms of readiness, taking our survey, and providing any and all feedback to this Roadmap section or any of the Playbook, you're in the driver's seat with us, heading toward the same destination in a consensus building, innovative way.

SECTION 2:
Key Resources for the Roadmap



Ethically Aligned Design: The Ideology Behind Trusted Data and AIS

At a time when data is abundant, the rise of data and privacy concerns is merely a reflection of the ongoing debate around ownership and governance of personal data. Furthermore, allowing intelligent and autonomous machines to make considerable decisions without putting humans in the center has further challenged the overall justice and suitability of such systems. All of them are hindering the advancement and adoption of AI and autonomous systems.

Ethically aligned design is a concept that emerged as part of a three-year process with more than 700 global experts who created a document by the same name: *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*. The original version came out in 2016, received more than 500 pages of feedback, and was released again in 2017(v2). This version of *EAD* was utilized by the OECD to create their AI principles, as well as FLI, the EU High Level Experts Group, and multiple companies like IBM.

Ethically aligned design as a methodology leverages the ethical and values-based design principles into design, development, and implementation of autonomous and intelligent systems. In the IEEE's published *Ethically Aligned Design* (the first edition), there are several general principles including: human rights, well-being, data agency, effectiveness, transparency, accountability, awareness of misuses, and competence, to guide us through our analytics and AI development journey.

The theoretical framework is based on three primary pillars:

1) Universal values
2) Political self-determination data agency
3) Technical dependency

Each pillar represents an important aspect of using data and AIS in business and technology decision making.

Together, the three pillars establish a solid foundation to understand the impacts of our decisions in creating and using such intelligent and autonomous systems.

Pillar 1: Universal human values

The pillar of universal human values advocates that advances in AI systems are not meant to serve the interest of a small group, a nation, or corporation, but really should be in service for all people. It is these values that should be used when developing policy, but also in the conception, development, and deployment stage for engineers, designers, and developers.

Pillar 2: Political self-determination data agency

The pillar of political self-determination and data agency reflects that people have the right to access, share, and benefit from their data and the resulting insight created and that data must be properly protected. By cultivating data agency over individual's digital identity and their data, we can nurture trust, develop accountability, and protect our private sphere.

Pillar 3: Technical dependability

The pillar of technical dependability emphasizes the importance of trust in AI systems. Here, we include reliability, safety, and active accomplishment of the system's intended objectives while ensuring the human values and considerations of the first pillar are reflected. To fulfill this goal, technologies should be monitored to ensure ethical practices and human values are respected, such as codified rights. Aspects of explainability, validation, and verification should be included to facilitate audits and certification of AI systems.

EU's ALTAI Ethical Principles: Principles Suggested for Practice

For consistency, throughout the playbook, we reference the ethical principles from the EU's Assessment List for Trustworthy Artificial Intelligence (ALTAI) in our high value use case and key resources assessments. The ALTAI includes seven ethical principles for Trustworthy AIS, which organizations can use for self-assessment: human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination, and fairness; societal and environmental well-being; and accountability.

A summary of the principles is provided in Table 1. Note that there are many commonalities between the Ethically Aligned Design principles and those presented in the ALTAI. We provide a more detailed discussion of the ALTAI and other global and Canadian data and AI regulation in *The Global and Canadian Regulatory Landscape section*. We also explore how each of the ALTAI principles maps to definitions in Ethically Aligned Design, and those used commonly in financial services, in Appendix B.

Table 1: The EU's Assessment List for Trustworthy AI (ALTAI)

#1 Human Agency and Oversight	#2 Technical Robustness and Safety	#3 Privacy and Data Governance	#4 Transparency	#5 Diversity, Non-discrimination and Fairness	#6 Societal and Environmental Well-being	#7 Accountability
• Human Agency and Autonomy • Human Oversight	• Resilence to Attack and Security • General Safety • Accuracy • Reliability, Fall-back Plans and Reproducibility	• Privacy • Data Governance	• Traceability • Explainability • Communication	• Avoidance of Unfair Bias • Accessibility and Universal Design • Stakeholder Participation	• Environmental Well-being • Impact on Work and Skills • Impact on Society at Large or Democracy	• Auditability Risk • Management

Developing the Critical Building Blocks: Key Resources

The increased adoption and use of autonomous and intelligent systems has been accompanied by an ever-increasing number of trusted data and AIS ethics resources made available by concerned stakeholders. These resources, which include initiatives such as principles, codes of conduct, policy observatories, frameworks, guidelines, white papers, and research papers, have been contributed by stakeholders such as academics, consulting firms, technology firms, regulators, governments, and intergovernmental organizations. Through our research interviews and consulting work with industry experts, we have heard that the wealth of resources can be overwhelming, and many resources may not be applicable for a financial services organization.

To address these concerns and aid in the implementation of the high-value AIS use cases, we consulted with several trusted data and AIS ethics experts in the financial services industry to determine which resources would be most impactful for a financial services organization looking to increase their trusted data and AI ethics readiness. The key resources are categorized into one of the three critical building blocks for Trusted Data and AIS: people, process, technology, and are tagged with the Trusted Data and AIS Readiness level we recommend for their execution; the details of which are provided in Figure 4. Note that not every critical building block has a key resource associated with it; this reflects where the focus has been to date and where there remain opportunities for additional guidance from various stakeholders.

If you haven't already done so, take the 20-minute anonymous online survey here to receive an evaluation of your organization's Trusted Data and AIS Readiness.

The survey results will provide you with guidance on which critical building blocks your organization should focus on in the Roadmap, and how to use the playbook resources to develop them.

We then review the 20 key resources on the following page, providing a summary of each that details what the resource is, how and by whom it can be used in a financial services organization, the key ethical themes covered, key findings, and actionable recommendations on how to immediately start using the resource.

Figure 4: Key Resources to Develop the three Critical Building Blocks for Trusted Data and AIS

	Key Resources	Critical Building Blocks	Human Agency and Oversight	Technical Robustness and Safety	Privacy and Data Governance	Transparency	Diversity, Non-discrimination and Fairness	Societal and Environmental Well-being	Accountability
People	Navigating Uncharted Waters: A roadmap to responsible innovation with AI in financial services	Data and AIS Climate	✓			✓	✓		
	The Future Computed: Artificial Intelligence and its role in society	Data and AIS Climate	✓	✓	✓	✓	✓	✓	✓
	Empowering AI Leadership: An Oversight Toolkit for Board of Directors	Leadership Buy-in	✓	✓	✓	✓	✓	✓	✓
	OECD.AI Policy Observatory	Data and AIS Strategy	✓	✓	✓	✓	✓		✓
	Responsible Innovation in the Algorithmic Era	Education and Awareness		✓	✓				✓
	Values by Design in the Algorithmic Era	Education and Awareness	✓						
	The Economic Advantage of Ethically Aligned Design	Education and Awareness	✓				✓	✓	✓
Process	Principles to Promote FEAT in the Use of AI and Data Analytics in Singapore's Financial Sector	Principles	✓	✓	✓	✓	✓		✓
	A Code of Conduct for the Ethical Use of AI in Canadian Financial Services	Principles	✓		✓	✓	✓	✓	✓
	Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI	Principles	✓	✓	✓	✓	✓		✓
	Responsible Bots: 10 Guidelines for Developers of Conversational AI	Principles		✓	✓	✓	✓		✓
	White Paper on Artificial Intelligence	Data and AIS Governance	✓	✓	✓	✓	✓	✓	✓
	AI Governance: A Holistic Approach to Implement Ethics into AI	Data and AIS Governance	✓	✓	✓	✓	✓		✓
	IIF Machine Learning Recommendations for Policymakers	Data and AIS Governance	✓			✓			✓
Technology	Modern Data Ops	Modern DataOps		✓		✓	✓		✓
	WhiteNoise: A Platform for Differential Privacy	Differential Privacy			✓				
	Trusted MLOps: Methodologies for Production-Ready AI	Trusted MLOps		✓		✓	✓		✓
	Annotation and Benchmarking on Understanding and Transparency of Machine Learning Lifecycles (ABOUT ML)	Testing and Optimization				✓			
	Adversarial Robustness 360 Toolbox (ART)	Testing and Optimization		✓					
	AI Explainability 360 Toolkit	Testing and Optimization				✓			✓
	AI Fairness 360 Open Source Toolkit	Testing and Optimization	✓			✓	✓		✓

Trusted Data and AIS Readiness Recommended for Implementation — Basic, Developing, Advanced, Leading, All

Developing the People Critical Building Block: Key Resources

Data and AIS Ethics Climate

Navigating Uncharted Waters: A Roadmap to Responsible Innovation with AI in Financial Services

Actionable Recommendation on How to Understand and Use this Resource

Strategic decision makers or policymakers can benefit from the proposed frameworks to mitigate risks posed by the wide scale adoption of AI in financial services.

Authors	World Economic Forum, Deloitte
What is it?	The report includes key findings and an executive summary of the concerns of the use of AI in financial services including: AI explainability, systemic risk and AI, bias and fairness, the algorithmic fiduciary, and algorithmic collusion.
What is its purpose?	The report seeks to understand the risks of AI to the financial system and proposes strategies for the mitigation of these key risks. The report aims to: • Propose decision-making frameworks to address key concerns surrounding the use of AI in financial services • Explore strategic upside in responsible and trust-first AI business models • Highlight areas of regulatory uncertainty
Who within a financial services organization would most benefit?	Top management team, strategic decision makers, regulators, and policymakers.
What could this be used for?	To guide executives and policymakers on the paradigms of governance and supervision of the use of AI in financial services.
Key Ethical Themes	• Human Agency and Oversight • Transparency • Diversity, Non-discrimination and Fairness
Key Findings	• Early adopters face risks such as: - The risk of customer backlash from AI failures that cause reputational damage. - The risk of regulatory scrutiny, censure, or depletion of goodwill. - The risk of alienating employees. • To safely unlock the power of AI, financial institutions, policymakers, and regulators should: - Responsibly deploy AI systems through new models of governance to challenge the foreignness of AI. - Responsibly scale AI that will drive cross-industry examination of competition policy, data-rights, and operational resilience. - Create the trust-first AI model of the future. • AI explainability is a growing concern and the black box effect of hidden layers between data inputs model outputs. There is a middle ground of informed trust that balances out unfair AI outcomes with the opportunity to benefit from the power of AI's full potential. • Systemic risk and AI raise new questions on financial stability and the management of system risk. New sources of system risk include the herding to move markets, the unpredictability of machines causing human panic, and the impact of optimizing algorithms locked in competition (e.g., rate setting). • Erosion of financial systems' defenses. - The loss of the effective challenge from the loss of skills in humans as they are removed from the process. - Undermining of regulatory control mechanisms. • Bias and fairness where AI can exacerbate unfair bias in financial decision making. The use of AI as the power to improve accuracy but can perpetuate unfair biases (human, data, model, second-hand). • The expansion of AI systems' responsibility raises questions about the ability of AI systems to effectively meet various standards of fiduciary duty. • Self-learning and the ability of AI systems to collude with each other.

Find this resource at: http://www3.weforum.org/docs/WEF_Navigating_Uncharted_Waters_Report.pdf

The Future Computed: Artificial Intelligence and its Role in Society

Actionable Recommendation on How to Understand and Use this Resource

Reading this report provides the user with an outline how to support the development of responsible AI systems. Organizations will be able to extract and leverage universal values and address the full range of societal issues that AI will raise.

Authors	Microsoft
What is it?	A report written to bring awareness to the growing use of AI in society and the challenges that it presents. The report also outlines implications to society that must be considered by researchers, policymakers, and leaders from government, business, and civil society.
What is its purpose?	The need to come together to develop a shared ethical framework for AI and help foster responsible development of AI systems that will engender trust.
Who within a financial services organization would most benefit?	Top management team, strategic decision makers, regulators, and policymakers.
What could this be used for?	A guidepost for the impact of AI on jobs and work and the principles that contribute to the development of responsible AI.
Key Ethical Themes	• Human Agency and Oversight • Technical Robustness and Safety • Privacy and Data Governance • Transparency • Diversity, Non-discrimination and Fairness • Societal and Environmental Well-being • Accountability
Key Findings	• Steps that promote the safety and reliability of AI systems: - Systematic evaluation of the quality and suitability of the data and models used to train and operate AI-based products and services, and systematic sharing of information about potential inadequacies in training data. - Processes for documenting and auditing operations of AI systems to aid in understanding ongoing performance monitoring. - When AI systems are used to make consequential decisions about people, a requirement to provide adequate explanations of overall system operation, including information about the training data and algorithms, training failures that have occurred, and the inferences and significant predictions generated. - Involvement of domain experts in the design process and operation of AI systems used to make consequential decisions about people. - Evaluation of when and how an AI system should seek human input during critical situations, and how a system controlled by AI should transfer control to a human in a manner that is meaningful and intelligible. A robust feedback mechanism so that users can easily report performance issues they encounter. • Privacy and Security—AI systems should be secure and respect privacy. • Inclusiveness—AI systems should empower everyone and engage people. • Transparency—AI systems should be understandable. • Accountability—The people who design and deploy AI systems must be accountable for how their systems operate. • Internal Oversight and Guidance—Microsoft's AI and Ethics in Engineering and Research (AETHER) Committee. • The Importance of Data—To help reduce the risk of privacy intrusions, governments should support and promote the development of techniques that enable systems to use personal data without accessing or knowing the identities of individuals. • Promoting Responsible and Effective Uses of AI—Governments have an important role to play in promoting responsible and effective uses of AI itself. • Liability—Governments must also balance support for innovation with the need to ensure consumer safety by holding the makers of AI systems responsible for harm caused by unreasonable practices.

Find this resource at: https://blogs.microsoft.com/wp-content/uploads/2018/02/The-Future-Computed_2.8.18.pdf

Leadership Buy-in

Empowering AI Leadership: An Oversight Toolkit for Boards of Directors

Actionable Recommendation on How to Understand and Use this Resource

Senior leaders and members of the board of directors should use this guide first to better understand, and then to set policy around business strategy as it relates to AI. The frameworks presented here support the creation of ethics policy that can be applied across all AI applications.

Authors	World Economic Forum (WEF)
What is it?	A guide for boards of directors consisting of 12 modules. Each module is aligned with typical board committees and provides a description of the topic, the relevant board responsibilities, oversight tools, and discussion resources. Modules include: brand, competition, customers, cybersecurity, operating model, people and culture, technology, audit, ethics, governance, risk, and responsibility. There is also a glossary of AI terms.
What is its purpose?	This resource is principally geared toward helping boards of directors learn, understand their responsibility for, and have meaningful discussions about AI. The paper aims to educate and empower board members with the knowledge they need to guide an organization through the changing world of AI.
Who within a financial services organization would most benefit?	Board of directors, top management team.
What could this be used for?	Educating executives, priming discussions at the board/senior executive level, developing/refining organizational policies around the adoption and fair use of AI.
Key Ethical Themes	• Human Agency and Oversight • Technical Robustness and Safety • Privacy and Data Governance • Transparency • Diversity, Non-discrimination and Fairness • Societal and Environmental Well-being • Accountability
Key Findings	• Ethics and responsible decision making must be at the heart of everything the board does, especially as it guides an organization through an AI transformation. • The board must consider the new normal and guide the organization through this transformation. Specifically, disruption with new business models, raising and serving customer expectations, and matching up with technically advanced ecosystem partners. • To successfully implement fair policies, the board needs an ethical framework that it can run all decision making through. This should be updated regularly. • One useful framework is ordered as follows, though the paper recommends tailoring this by adding other frameworks important to your organization: - Required: By law, by human rights, by contractual obligation. - Recommended: Promotes organizational values, promotes stakeholder trust, follows professional code. - Above and beyond: Shares wealth, benefits society, sustains the environment. • The board will have to consider monitoring and enforcement protocols to keep businesses aligned and on track, as well as keep the ethics guidelines updated regularly.

Find this resource at: https://spark.adobe.com/page/RsXNkZANwMLEf/

Data and AIS Ethics Strategy

OECD.AI Policy Observatory

Actionable Recommendation on How to Understand and Use this Resource

Use the Business Stakeholder Initiative page in the Countries and Initiatives section to view more than 20 organizational AI ethics codes of conduct, both from financial services and several other industries. Use the Countries and Initiatives hub to better understand the landscape for AI ethics in the various countries in which your organization operates.

Authors	The Organization for Economic Co-operation and Development (OECD)
What is it?	A database of global AI policies and AI policy research curated by the OECD.
What is its purpose?	To monitor and encourage the responsible development of trustworthy AI.
Who within a financial services organization would most benefit?	Head of AI/analytics, head of data, compliance, risk, regulators, and policymakers, training/education coordinators.
What could this be used for?	Developing/refining organizational AI ethics principles, understanding the global landscape for AI ethics principles and initiatives.
Key Ethical Themes	• Human Agency and Oversight • Technical Robustness and Safety • Privacy and Data Governance • Transparency • Diversity, Non-discrimination and Fairness • Accountability

Key Findings	• The online repository is broken into four pillars.
	• AI principles: An overview of the OECD Values-Based Principles on AI
	  - Inclusive growth, sustainable development well-being.
	  - Human-centered values and fairness.
	  - Transparency and explainability.
	  - Robustness, security, and safety.
	  - Accountability.
	• Policy areas: The most extensive collection of AI policies and policy research (at the time of publication), organized by field (e.g., finance and insurance, education, environment, agriculture). Each page aggregates and monitors in real time related publication, news, research reports for a given field.
	  - The finance and insurance page is of particular relevance.
	• Trends and data: Houses a series of reports and data from the OECD on AI development across the globe, complete with several data visualizations.
	• Countries and initiatives: Split into two primary sections for national AI policies and stakeholder initiatives. Each country or territory has its own page detailing a description of the initiative, objective, background, budget, and several other facts, along with a URL to the initiative itself. A similar page exists for each of the six stakeholder groups: businesses, academia, technical communities, civil society, intergovernmental organizations, and trade unions.
	  - The Business stakeholder page, houses more than 20 organizational AI ethics codes of conduct from financial services and several other industries; the largest gathering of organizational AI ethics principles (at the time of publication).
	  - Principles from Google, IBM, and Microsoft are all present, but the observatory team has gathered documents from Baidu, Telefonica, SONY, Deutsche Telekom, and Sage along with several other organizations.

Find this resource at: https://oecd.ai/

Education and Awareness

Responsible Innovation in the Age of Artificial Intelligence

Actionable Recommendation on How to Understand and Use this Resource

Assign the online educational video to business practitioners/data science leaders/executives to introduce them to AI ethics from a non-technical position. Use the course by itself as a general introduction to the topic, or with the other IEEE continuing education videos in the Artificial Intelligence and Ethics in Design program for a more robust education on AI ethics.

Authors	IEEE Standards Association and IEEE Continuing Education, presented by Rumman Chowdhury, John P. Sullins, and Virginia Dignum
What is it?	A 1-hour online educational video on AI ethics challenges, part of the IEEE Artificial Intelligence and Design Program.
What is its purpose?	To guide participants in understanding the ethical challenges faced by practitioners who are developing AI applications, with particular focus on responsible AI; law, compliance and ethics in AI; and practical applied AI ethics. It is targeted toward non-technical business practitioners.
Who within a financial services organization would most benefit?	Communications Directors, training/education coordinators, data scientists.
What could this be used for?	Developing AI ethics principles, guidelines, or other ethical initiatives. Educating business practitioners, data scientists, and others on AI ethics.
Key Ethical Themes	• Technical Robustness and Safety • Privacy and Data Governance • Accountability

Key Findings	• The course delivers on three key learning objectives to help participants.
	• Understand the principles of responsible innovation in AI.
	  - Define responsible AI and identify its value to AI innovation.
	  - Establish a preliminary framework to implement responsible AI in the workplace.
	• Discover the unique challenges AI is introducing for existing organizational ethics frameworks.
	• Determine how practical applied ethics concepts can be used in the design of AI systems.
	It is divided into three parts, each covering a series of important questions on AI ethics.
	• Responsible AI
	  - What question does responsible AI answer?
	  - What do business leaders (e.g., BOD, CEO, CRO/CISO) want to know about responsible AI?
	  - How can a responsibility framework enable AI to flourish and establish trust?
	  - How can I get started with responsible AI?
	• Law, compliance, and ethics in AI
	  - What is the difference between law, compliance, and ethics?
	  - Is it too early to worry about AI ethics?
	  - When is the right time to do something about the ethical impacts of AI?
	  - Should we be worrying about the ethical decisions made by AI?
	• Practical applied ethics for use in the responsible design of AI systems
	  - Why look at ethics?
	  - Which ethics?
	  - Whose ethics?
	  - How to implement ethics.

Find this resource at: https://ieeexplore.ieee.org/courses/details/EDP496

Values by Design in the Algorithmic Era

Actionable Recommendation on How to Understand and Use this Resource

Assign the online educational video to business practitioners/data science leaders/executives to introduce them to the concept of values by design and research ethics for AI from a non-technical position. Use the course with the other IEEE continuing education videos in the Artificial Intelligence and Ethics in Design program for a more robust education on AI ethics.

Authors	IEEE Standards Association and IEEE Continuing Education, presented by Sarah Spiekermann and Sara Jordan
What is it?	A 1-hour online educational video on research ethics and values-based design, the foundational design methodology for the IEEE Trusted Data and AI Standards and Certifications (i.e., IEEE P7010™, ECPAIS, and Ethically Aligned Design).
What is its purpose?	To educate participants on values-based system design and research ethics and the role these two concepts play in the development of trustworthy AI.
Who within a financial services organization would most benefit?	Communication, training/education coordinators, data scientists.
What could this be used for?	Developing AI ethics principles, guidelines, or other ethical initiatives. Educating business practitioners, data scientists, and others on AI ethics.
Key Ethical Themes	• Human Agency and Oversight
Key Findings/ Contributions	The course delivers on six key learning objectives to help participants: • Gain an understanding of what ethics means • Examine the most important streams of ethical thinking and reasoning. • Discuss what values are. • Understand the importance of stakeholder processes to derive the values for system design. • Examine the system engineering flow for value-based design. • Review responsible conduct of research including concepts, regulations, and applications. It is divided into two parts: • Values by design in the algorithmic era - What is value-based system design and how does it relate to ethics? - What are values? - What are the four dimensions of the value definition? - How do you derive values for your organization? - How can you integrate values into AI systems? - What is the role of stakeholders in values-based system design? • Ethics in Artificial Intelligence research and development. - What are: - Responsible conduct of research, - Detrimental research practices, - Research misconduct, - Research ethics, core values in research? - How can research ethics help your organization for responsible AI development?

Find this resource at: https://ieeexplore.ieee.org/courses/details/EDP498

The Economic Advantage of Ethical Design for Business

Actionable Recommendation on How to Understand and Use this Resource

Assign the online educational video to business practitioners/data science leaders/executives to introduce them to the economic benefits of AI ethics from a non-technical position. Use the course with the other IEEE continuing education videos in the Artificial Intelligence and Ethics in Design program for a more robust education on AI ethics.

Authors	IEEE Standards Association and IEEE Continuing Education, presented by Virginia Dignum and Matthew Scherer
What is it?	A 1-hour online educational video on AI ethics challenges, part of the IEEE Artificial Intelligence and Design Program.
What is its purpose?	To educate participants in the economic benefits and importance of responsible AI development, and diversity and inclusion in the development of AI.
Who within a financial services organization would most benefit?	Communication, training/education coordinators, data scientists.
What could this be used for?	Developing AI ethics principles, guidelines, or other ethical initiatives. Educating business practitioners, data scientists, and others on AI ethics.
Key Ethical Themes	• Human Agency and Oversight • Diversity, Non-discrimination and Fairness • Societal and Environmental Well-being • Accountability
Key Findings/ Contributions	The course delivers on five key learning objectives to help participants: • Identify and explain what responsible research and innovation look like. • Understand why AI success measurement must go beyond simple metrics and focus on human well-being • Understand the importance of diversity and inclusion in AI development. • Identify the stages of AI development in which diversity and inclusion matter most. • Understand the relationship between the personal, cultural, and systems dimensions of diversity. It is divided into two parts: • Prioritizing human well-being in the age of AI - Responsible research and innovation (RRI) and its role for AI development. - Diversity and inclusion. - Openness and transparency. - Anticipation and reflexivity. - Responsiveness and adaptation. - The importance of measuring human well-being for AI success. - Leveraging openness and transparency as a source of growth. - Discussion of ethical use cases. • Business advantage of diversity and inclusion. - A review of diversity and inclusion. - Three stages where it matters: lab, data, end user - Three dimensions: human, cultural, systems. • Why AI ethics is more than legal compliance or social responsibility. • Diversity, inclusion, and discrimination law.

Find this resource at: https://ieeexplore.ieee.org/courses/details/EDP497

Developing the Process Critical Building Block: Key Resources

Principles

Principles to Promote FEAT in the Use of AI and Data Analytics in Singapore's Financial Sector

Actionable Recommendation on How to Understand and Use this Resource

Use these principles as a starting point for developing or refining your organization's AI ethics principles.

Authors	Monetary Authority of Singapore
What is it?	A set of high-level principles developed by the financial regulator in Singapore, in consultation with financial services organizations, to guide the ethical use of AI and data analytics (AIDA) in financial products and services.
What is its purpose?	To provide foundational guidance on ethics principles, governance, and to aid in the development of public trust in the use of analytics and AI in Singapore's financial services sector.
Who within a financial services organization would most benefit?	Head of AI/analytics, head of data, compliance, risk, regulators, and policymakers.
What could this be used for?	Developing/refining organizational AI ethics principles.
Key Ethical Themes	• Human Agency and Oversight • Technical Robustness and Safety • Privacy and Data Governance • Transparency • Diversity, Non-discrimination and Fairness • Accountability
Key Findings	• The document lists 14 summary principles under the four sections of fairness, ethics, accountability, and transparency. It provides a brief explanation of each principle along with several illustrative examples for each, specific to financial services. • The fairness section covers the justifiability of the decision-making process including the use of personal attributes, and accuracy and bias, calling for regularly scheduled reviews and validation. • Ethics suggests that the use of AI and analytics should be aligned to other organizational ethics standards and code, and calls for analytics and AI to be held to the same ethical standards as human decisions. • Accountability addresses both internal and external accountability concerns. Internally, all uses of analytics and AI should be approved by an internal authority, whether the applications are developed internally or externally, and all use cases should be communicated to management and the board of directors. Externally, there should be a feedback loop for data subjects to enquire about and/or appeal any decision made by analytics or AI, and firms must verify and use any relevant data provided by the subjects in their decision-making process. • Transparency proposes several measures to ensure public trust: proactively disclosing the use of analytics and AI to data subjects, providing clear explanation upon request of the data used and the consequences of decisions made by analytics and AI applications.

Find this resource at: https://www.mas.gov.sg/~/media/MAS/News%20and%20Publications/Monographs%20and%20Information%20Papers/FEAT%20Principles%20Final.pdf

A Code of Conduct for the Ethical Use of AI in Canadian Financial Services

Actionable Recommendation on How to Understand and Use this Resource

Use these principles as a starting point for developing or refining your organization's AI ethics principles.

Authors	Stephanie Kelley, Yuri Levin, and David Saunders at the Smith School of Business at Queen's University
What is it?	A set of principles created in consultation with several Canadian financial services organizations to guide the ethical use of AI in financial institutions.
What is its purpose?	To provide practical guidance for financial services organizations to prevent ethical implications in their day-to-day use of AI. To close the gap between AI technology and the existing ethical guidelines and regulations currently in place which were designed to govern other technologies.
Who within a financial services organization would most benefit?	Head of AI/analytics, head of data, compliance, risk, regulators, and policymakers.
What could this be used for?	Developing/refining organizational AI ethics principles.
Key Ethical Themes	• Human Agency and Oversight • Privacy and Data Governance • Transparency • Diversity, Non-discrimination and Fairness • Societal and Environmental Well-being • Accountability
Key Findings	• The document is a first step in a practical, industry specific set of guiding ethical principles. It includes a series of definitions, including a practical definition of AI, along with 17 summary principles and practical examples. The principles are grouped into three topics: fairness, accountability, and transparency. • Principles of Fairness: Bias and discrimination guidelines denounce the intentional or unintentional disadvantage of individuals and suggest that an organization should conduct ethics reviews for material AI applications to ensure unfair discrimination, intentional and unintentional disadvantage are avoided. Justifiability is also covered; and the document suggests that in addition to being lawful, the aggregate input data and its outcome should be understood for material AI applications. • Principles of Accountability: Specific individuals in the organization must be responsible and accountable for the output of the organization's AI applications, regardless of where they are developed. These individual(s) are responsible for the level of autonomy assigned to an AI application and should be held to the same standard as human decisions, especially if those decisions impact the workforce. All applications must be approved by the accountable and responsible parties prior to production; this approval should ensure a consistent model development process has been followed and ensure at minimum each model is validated, approved, and monitoring ongoing, in line with its materiality. • Principles of Transparency: Explainability, upon request, is expected of models that make material decisions that impact individuals. In addition to being lawful, any data used in the AI application should adhere to the highest standards of data privacy and informed consent, regardless of whether the application or data is external or internal. To ensure trust, an organization should proactively announce the use of AI to individuals when there is direct interaction with the application (i.e., chat bot), or a material decision is made by an AI application (i.e., credit lending).

Find this resource at: https://www.stephaniekelleyresearch.com/a-code-of-conduct-for-ethical-ai

Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI

Actionable Recommendation on How to Understand and Use this Resource

Use the data visualization and accompanying report as a starting point for developing or refining your organization's AI ethics principles as it offers a comprehensive analysis and aggregation of 36 prominent AI ethics documents, with an online real-time data visualization that recaps the AI ethics documents.

Authors	Jessica Fjeld and a team of researchers at The Berkman Klein Center for Internet and Society at Harvard University
What is it?	A report that provides a qualitative comparison of 36 prominent ethical AI principle documents, with a companion online visual mapping of the documents.
What is its purpose?	To assess, identify, and summarize trends in the 36 ethical AI principle documents.
Who within a financial services organization would most benefit?	Head of AI/analytics, head of data, compliance, risk, regulators, and policymakers.
What could this be used for?	Developing/refining organizational AI ethics principles, understanding the global landscape for AI ethics principles.
Key Ethical Themes	• Human Agency and Oversight • Technical Robustness and Safety • Privacy and Data Governance • Transparency • Diversity, Non-discrimination and Fairness • Accountability
Key Findings	• There are eight key themes covered across the 36 prominent ethical AI documents. They are: privacy, accountability, safety and security, transparency and explainability, fairness and non-discrimination, human control of technology, professional responsibility, and promotion of human values. • The more recent documents tend to cover all of the eight themes, suggesting a convergence of the ethical AI principle conversation, and a potential normative core of a principles-based approach to ethical AI. • Principles are context-specific and should be interpreted in their "cultural, linguistic, geographic, and organizational context." • Although existing privacy regulations are often tied to the AI ethics conversation, it is human rights laws that are the most interconnected and referenced in the AI ethics principles documents. • Alone, ethical AI principles are not enough to prevent the unethical use of AI; embeddedness in the larger AI ethics ecosystem (including laws, regulation, national strategies, professional practices, and daily routines) is necessary for impact.

Find this resource at: https://cyber.harvard.edu/publication/2020/principled-ai

Responsible Bots: 10 Guidelines for Developers of Conversational AI

Actionable Recommendation on How to Understand and Use this Resource

These guidelines are most relevant to building bots that may affect people in consequential ways—such as helping people to navigate information relating to employment, finances, and such.

Authors	Microsoft
What is it?	Ten guidelines for developers for building bots responsibly.
What is its purpose?	These guidelines are aimed at helping developers to design a bot that builds trust in the company and service that the bot represents by following six key principles, including ethics, privacy, security, safety, inclusion, and transparency and accountability.
Who within a financial services organization would most benefit?	Head of AI/analytics, data scientists, compliance, risk, regulators, and policymakers.
What could this be used for?	Building bots in a responsible way.
Key Ethical Themes	• Technical Robustness and Safety • Privacy and Data Governance • Transparency • Diversity, Non-discrimination and Fairness • Accountability
Key Findings	• Articulate the purpose of your bot and take special care if your bot will support consequential use cases. Be sure to pause to research, learn, and deliberate on the impact of the bot on people's lives. Develop metrics to assess user satisfaction. • Be transparent about the fact that you use bots as part of your product or service. • Ensure a seamless hand-off to a human where the human-bot exchange leads to interactions that exceed the bot's competence. • Design your bot so that it respects relevant cultural norms and guards against misuse. Since bots may have human-like personas, it is especially important that they interact respectfully and safely with users and have built-in safeguards and protocols to handle misuse and abuse. • Ensure your bot is reliable and be transparent about bot reliability. • Ensure your bot treats people fairly. Systematically assess the data used to train your bot. Strive for diversity amongst your development team. • Ensure your bot respects user privacy. Inform users up front about the data that is collected and how it is used and obtain their consent beforehand, but collect no more personal data than you need. • Ensure your bot handles data securely. • Ensure your bot is accessible and have people with disabilities test your bots. • Accept responsibility. Developers are accountable for the bots they build.

Find this resource at: https://www.microsoft.com/en-us/research/uploads/prod/2018/11/Bot_Guidelines_Nov_2018.pdf

Data and AIS Governance

White Paper on Artificial Intelligence

Actionable Recommendation on How to Understand and Use this Resource

Use this to guide the development of internal risk and compliance policies on AI ethics in advance of formal regulation.

Authors	European Commission
What is it?	A white paper outlining the direction of the forthcoming EU AI regulation, rooted in the ethics principles released by the High-Level Expert Group on Artificial Intelligence in 2019. It also provides an overview of the European Commission's two-pronged approach to build both an ecosystem of excellence and trust for AI in Europe.
What is its purpose?	To assess, identify, and summarize trends in the 36 ethical AI principle documents.
Who within a financial services organization would most benefit?	Head of analytics/AI, legal, risk, compliance.
What could this be used for?	Guiding the development of internal risk and compliance policies on AI ethics, planning for future regulation.
Key Ethical Themes	• Human Agency and Oversight • Technical Robustness and Safety • Privacy and Data Governance • Transparency • Diversity, Non-discrimination and Fairness • Societal and Environmental Well-being • Accountability
Key Findings	The paper makes several broad claims about the EU's objectives on AI including hopes for upward regulatory convergence, access to key resources including data, and a vision to create a level playing field for AI for its members. It states the EU hopes to use AI to capitalize on existing strengths in industrial and professional markets, and lead the next data wave expected with edge and quantum computing. • Forthcoming AI regulation will apply to all AI-enabled products and services in the EU, regardless of whether they are developed in the EU or not (similar to GDPR) • It is undecided whether current regulation will be enforced as is, adapted, or new regulation developed, but the commission has noted several unique properties of AI that may induce legislative reform including lack of transparency, the changing functionality of AI, accountability of AI systems, cybersecurity risk, and product safety challenges (which are discussed in the accompanying Report on Safety and Liability Implications of AI). • Applications will be divided into high risk and other, with stronger controls on training data, data governance and explainability, reporting, robustness and accuracy, and human oversight for the high-risk applications. • At this time, it is unclear whether all or part of the financial services industry will be deemed high-risk by the forthcoming regulation. Examples of high-risk applications discussed in the white paper include those from healthcare, transport, energy, and parts of the public sector including border controls and migration; however, it states that some applications from these industries would be exempt from the stronger controls if they are not used in a manner that could generate risk (i.e., an appointment managing algorithm in a hospital setting may not be high risk, but a triaging algorithm likely would be). • Additional regulation will likely be proposed on remote biometric identification (i.e., identification of people at a distance using biometric identifiers such as iris, facial image, vascular patterns, etc.). • The stronger regulation will likely be enforced through inspection or certification; organizations may be held to higher standards and restricted in their use of inscrutable algorithms if using high-risk applications. • A voluntary labeling scheme has been proposed for applications not deemed high risk that would allow labelling of applications as trustworthy if they adhere to the stronger regulations. While the labelling itself would be voluntary, the same enforcement mechanisms would be used to ensure compliance. • A definition of AI is provided, an important first step in regulatory reform (found on page 16 of the white paper in the footnotes).

Find this resource at: https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

AI Governance: A Holistic Approach to Implement Ethics into AI

Actionable Recommendation on How to Understand and Use this Resource

Use this to guide the development of internal risk and compliance policies on AI ethics in advance of formal regulation.

Authors	World Economic Forum (WEF)
What is it?	This white paper aims to enrich the ongoing debate about implementing ethical considerations into AI by looking at possible means and mechanisms to apply ethical values and principles in AI-driven technology and machines in order to contribute to building a human-centric AI society.
What is its purpose?	The goal is to outline approaches to determine an AI governance regime that fosters the benefits of AI while considering the relevant risks that arise from the use of AI and autonomous systems.
Who within a financial services organization would most benefit?	Head of analytics/AI, legal, risk, compliance.
What could this be used for?	Developing an internal compliance program for AI ethics and trusted data, guiding regulators on the design and appropriate regulation of AI.
Key Ethical Themes	• Human Agency and Oversight • Technical Robustness and Safety • Privacy and Data Governance • Transparency • Accountability
Key Findings	There are several means to implementing ethics in AI applications: • To implement ethical decision-making criteria technically, both a bottom-up approach and a top-down approach are possible. • Casuistic approach: Machines would be programmed to react specifically in each situation where they may have to make an ethical decision. • Dogmatic approach: The systems could be programmed in line with a specific ethical school of thought. • Technical meta-level approach: An AI-driven monitoring system that controls a machine's compliance with a predetermined set of laws and ethical rules on a meta-level (guardian AI) could be developed. Such guardian AI could technically interfere in the basic AI's system and directly correct unlawful or unethical decisions. • The insufficiency of technical means and mechanisms: To make sure that AI systems behave according to ethical principles, it is necessary to adopt a variety of agile governance mechanisms including, for example, binding legal requirements or the creation of economic incentives to promote ethically aligned AI system design. There are two practical approaches to implement ethics in AI systems. • The IEEE Global Initiative. • The World Economic Forum project on Artificial Intelligence and Machine Learning.

Find this resource at: https://spark.adobe.com/page/RsXNkZANwMLEf/

IIF Machine Learning Recommendation for Policymakers

Actionable Recommendation on How to Understand and Use this Resource

Use this to guide the development of internal risk and compliance policies on AI ethics in advance of formal regulation.

Authors	Institute of International Finance (IIF)
What is it?	A set of recommendations for policymakers looking to regulate machine learning (ML), presented by IIF. The recommendations are based on the aggregated summary of challenges and opportunities faced by 87 financial institutions.
What is its purpose?	To provide policymakers with a set of recommendations to guide the development of future ML regulation to ensure a level playing field, support adoption, and avoid overregulation that could stifle innovation.
Who within a financial services organization would most benefit?	Head of analytics/AI, legal, risk, compliance.
What could this be used for?	Guiding the development of internal risk and compliance policies on AI ethics, planning for future regulation.
Key Ethical Themes	• Human Agency and Oversight • Transparency • Diversity, Non-discrimination and Fairness • Accountability
Key Findings	• Although supervisors are closely watching the adoption of ML, the Monetary Authority of Singapore has been the only one to release guidance, in the form of high-level principles.[3] In lieu of guidance from regulators, many FIs are developing their own internal principles to fill the gap, but the IIF warns that these documents should be coordinated across institutions and jurisdictions where possible to promote a consistent approach. Principles should be brought to life using current governance and risk management frameworks which are already well developed. • Recommendations to avoid overregulation: - Existing regulations, particularly those in privacy and data protection, should be assessed to identify gaps in their ability to govern AI, and where possible expanded before creating new regulation. - Regulatory initiatives should take a risk-based approach and consider varying levels of controls given the materiality of specific use cases. - Materiality could include factors such as: the extent of ML application, the complexity, the degree of automation, the severity and probability of stakeholder impact, the monetary and financial impact, the level of human oversight, the impact on regulatory compliance, the cybersecurity risk, and the potential reputational impact. - Any regulatory initiatives should be technology-agnostic and not overly prescriptive to encompass the rapid advances in ML methodologies and application. • Recommendations to support adoption: - All institutions should adhere to a robust governance structure to support end-to-end oversight of all applications of ML. This should be supported by a consistent definition of ML and AI to ensure all relevant applications are maintained in an inventory and governed accordingly. - Regulation should require firms to have clear guardrails to constrain their ML models and ensure firms have in place internal guidelines to mitigate additional risks, particularly for highly complex models. - Existing model governance practices should be leveraged, but where appropriate, additional principles-based guidelines can be used by institutions to outline unique risks and relevant examples related to the use of ML. These principles-based guidelines should be updated at minimum annually to ensure they remain relevant. • Recommendations to ensure a level playing field: - Regulatory initiatives should consider the various types of financial institutions (i.e., fintechs, bigtechs, shadow banks) to ensure all financial activities are regulated consistently, and regulatory arbitrage is avoided. - Globally harmonized regulatory initiatives should remain the focus; however, the IIF will remain open to non-harmonized initiatives where a one-size-fits-all approach is not warranted. • The appendices to the recommendations include a framework to adapt existing modeling frameworks to handle the bias and ethical implications of ML, a report on interpretability (explainability) techniques.

Find this resource at: https://www.iif.com/Portals/0/Files/content/Innovation/09252019_iif_ml_policymakers.pdf

[3] In September 2019, the Monetary Authority of Singapore was the only financial supervisor to have such principles, but the Hong Kong Monetary Authority followed with the release of additional principles in November 2019.

Developing the Technology Critical Building Block: Key Resources

Modern DataOps

Actionable Recommendation on How to Use this Module

This resource can be used by data science leaders, enterprise architects, and chief data officers to consider the key principles of creating a reliable data production framework for AI to enable faster real-world insights and return on investment.

What is it?	A summary of leading DataOps principles; a methodology which borrows concepts from development and operations (DevOps) to ensure the required people, processes, and technology are in place to support the data use for machine learning.
What is its purpose?	To provide a consistent methodology to guide data decisions for AI deployment frameworks. Readers may already be familiar with data governance and management principles for traditional analytics and reporting. The following guide outlines the DataOps principles to keep in mind when evaluating options and designing AI pipelines for production.
Who within a financial services organization would most benefit?	Head of AI/analytics/data, enterprise/AI architects, data scientists, data engineers, compliance, risk.
What could this be used for?	Educating data science leaders and enterprise architects on best practices when designing and evaluating data solutions.
Key Ethical Themes	• Technical Robustness and Safety • Transparency • Diversity, Non-discrimination and Fairness • Accountability

Key Principles to Operationalize AI Solutions (Common to DataOps and MLOps)	• Ensure the AI solution is ethical and follows necessary regulatory compliance up front.
	  - Do not spend time building and operationalizing a solution that is not ethical.
	  - Assemble all stakeholders early to review the intended purpose.
	• Clearly identify success criteria and a methodology for ongoing monitoring.
	  - Define a feedback loop that captures the agreed upon success criteria.
	• Keep reusability in mind: consider platform agnostic frameworks.
	  - Solution architecture might differ depending on your model, available tools, and infrastructure.
	• Start building the end-to-end pipeline in parallel to your AI model development.
	  - Avoid leaving the technical implementation details to the end.
	  - Engage your data engineering team as early as possible in your project. Data scientists and data engineers will have interdependent goals and should work closely.
	• Have an agile mindset when it comes to delivering AI solutions.
	  - Deliver value in small, iterative increments when possible.
	  - Start with a simple model to test the end-to-end pipeline.
	• Adopt best practices and tools to make data scientists comfortable with continuous integration (CI) and continuous deployment (CD). This is crucial!
	  - GitOps principles ensure repositories become the single source of truth and pull requests are used to version control all changes.
	  - Automation tools such as Jenkins® are used to build, test, and deploy code in a controlled and predictable manner.
	  - CI/CD pipelines create consistent environments for running AI solutions by provisioning necessary compute resources (e.g., virtual machines, Docker® containers, or cloud servers), checking out code from a version-controlled repository, compiling the code, running tests, packaging binaries, and deploying binaries without manual configuration or intervention.
	• Consider separating your data pipeline (DataOps) from your ML pipeline (MLOps).

Jenkins is a registered trademark of Software in the Public Interest, Inc

Docker is a registered trademark of Docker, Inc..

DataOps Key Principles

DataOps blends the principles of data management, data governance, and development operations (DevOps) together in order to oversee this invaluable resource. Data is the lifeblood to AI solutions and should be treated with the same control that application development receives today.

- Version control your data pipelines.
  - Changes to data pipelines will have downstream impacts. Use version control to monitor these changes and notify dependent models.
- Version control your data.
  - Data should be immutable and versioned, meaning any changes to the underlying data is saved as a new version. This is the only way to ensure features and model results are reproducible and can be rerun using historical snapshots of data.
  - Data versioning allows for proper auditing and issue investigation.
  - Any changes to the underlying data will result in dependent models needing to be rerun. By tracking these changes, models can be rerun after data issues are resolved.
- Monitor for data quality issues before data reaches AI models.
  - Data validation is especially important when the AI is empowered to make autonomous decisions.
  - Establish agreed upon metrics to sufficiently monitor for data quality issues. Decide if these metrics should be generic or specific (specific will require additional work but may produce more favorable results).
- Consider saving the output of your data pipelines to a feature store.
  - A feature store is a centralized data source of reusable, quality features (transformed data) available across teams.
  - This provides consistent features when moving from model training to model deployment (AI relies heavily on consistent data).
- Apply DataOps principles to model training as well as model deployment.
  - Encourage using features from the feature store to train AI models.
  - If features are not yet in the feature store, teams should follow the established DataOps principles to appropriately contribute their work into the feature store.
- Pair DataOps with appropriate data governance and management tools.
  - Governance includes metadata collection and specifying rules of use and access.
  - Can be done on an as-needed basis for specific use cases instead of trying to take on everything at once.
  - Having proper governance and management in place will avoid wasted effort on tasks that are not suitable for production either legally, ethically, or practically (refer to Principle #1)!

DataOps Resources

The following links provide additional information and technical guidance for DataOps.

- Dowling, J. (5 March 2020). MLOps with a Feature Store. Retrieved 3 July 2020, from https://towarddatascience.com/mlops-with-a-feature-store-816cfa5966e9
- Eckerson Group Report. DataOps: Industrializing Data and Analytics Strategies for Streamlining the Delivery of Insights. (2018). Retrieved 3 July 2020, from https://www.qlik.com/us/resource-library/dataops-industrializing-data-and-analytics
- The DataOps Manifesto. (2019). Retrieved 3 July 2020, from https://www.dataopsmanifesto.org/dataops-manifesto.html
- Special thanks to Joe DosSantos, CDO of Qlik (conversation on 10 July 2020).

Terms Used in DataOps

- AI, ML, predictive models, models: Used interchangeably in this context to represent the code that produces predictive insights.
- Pipeline: A series of events scheduled together.
- Training: Developing a model on initial training data.
- Testing: Testing the performance of the model on a test set that it has not seen before.
- Serving, deploying, scoring: Moving the AI model into production, using the model to generate new insights (inferences) in the real world.

Find this resource at: https://blogs.microsoft.com/wp-content/uploads/2018/02/The-Future-Computed_2.8.18.pdf

Differential Privacy

WhiteNoise: A Platform for Differential Privacy

Actionable Recommendation on How to Understand and Use this Resource

The heads of privacy, analytics/AI, legal, risk, and compliance can review the white paper and accompanying software repository to understand the potential impact of differential privacy. The resource could inspire change in internal privacy governance frameworks to allow for privacy-protected analysis of personally identifiable information. Regulators and policymakers can also use the resource to guide the next evolution of privacy regulation.

Authors	Microsoft and Harvard University's Open Differential Privacy (OpenDP) Project
What is it?	An open source software repository and accompanying white paper that details how to implement differential privacy, a technique based on a strong foundation of theoretical math and inspired by cryptography that allows the user to analyze personally identifiable information in a way that allows for the privacy-protected analysis of personally identifiable information and ensures individual records cannot be reverse-engineered. The software is available in several common programming languages (through bindings).
What is its purpose?	The white noise software itself uses differential privacy techniques. The software repository's aim is to aid in the practical adoption of differential privacy techniques by any interested data analytics party, including industry, government, data archivists, and human-subject research communities.
Who within a financial services organization would most benefit?	Head of privacy, head of analytics/AI/data, data scientists, compliance, risk, legal, regulators, and policymakers.
What could this be used for?	Inspiring change in internal privacy governance frameworks to allow for privacy-protected analysis of personally identifiable information. It could also be used by regulators and policymakers to guide the next evolution of privacy regulation.
Key Ethical Themes	• Privacy and Data Governance
Key Findings	• Differential privacy is a set of systems and practices to keep personally identifiable information and other sensitive data safe and private while allowing for its analysis. • The privacy system, which is referred to as a "trusted curator," sits between the user and the private dataset. For each query a user makes, the trusted curator: - Checks the user's access credentials - Checks to determine if the query fits within the predefined privacy budget for that dataset (the budget is defined by an administration based on the re-identification risk of the dataset) - If both checks pass, the credentials are used to access the private dataset - A privacy mechanism then adds noise to the private data - The differentially-private data and accompanying differential privacy metrics, which are close to, but not exactly equal to, the real results, are then returned to the user • Differential privacy does have a data usability and reliability tradeoff; however, it has already been deployed at large-scale at the US Census Bureau, Google, Apple, Uber, and Microsoft and clearly offers organizations much potential. • Differential privacy techniques have the potential to completely change both the privacy governance frameworks and privacy regulations for financial institutions, allowing for the analysis of personally identifiable information, unlocking new knowledge, all while preserving the privacy of the consumer.

Find this resource at: https://projects.iq.harvard.edu/files/opendp/files/opendp_white_paper_11may2020.pdf and https://github.com/opendifferentialprivacy/whitenoise-core

In addition to this resource, IBM has also released a similar code library which can be found at: https://github.com/IBM/differential-privacy-library

Trusted MLOps

Methodologies for Production-Ready AI

Actionable Recommendation on How to Use this Module

This resource can be used by data science leaders, enterprise architects, and chief data officers to consider the key principles of creating a reliable data production framework for AI to enable faster real-world insights and return on investment.

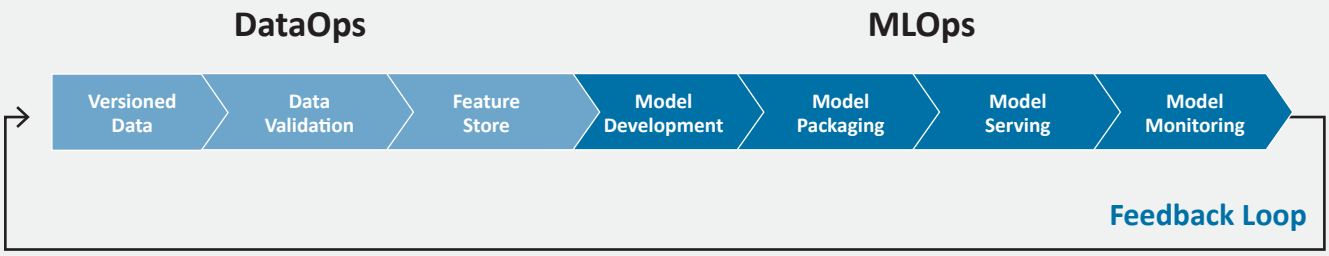
What is it?	A summary of leading MLOps principles; a methodology which borrows concepts from development and operations (DevOps) to ensure the required people, processes, and technology are in place to support the end-to-end development of ML solutions. Benefits include faster time-to-market, decreased technology spend, and better code governance.
What is its purpose?	To provide a consistent methodology to guide design decisions for AI deployment frameworks. Readers may already be familiar with data governance and management principles for traditional analytics and reporting. This resource explains the importance of applying the same rigor to AI models which are integrated into business processes and impact consumer experience. Therefore, these models need to be governed and monitored accordingly. The following guidelines outlines the key MLOps and DataOps principles to keep in mind when evaluating options and designing AI pipelines for production.
Who within a financial services organization would most benefit?	Head of AI/analytics/data, enterprise/AI architects, data scientists, data engineers, compliance, risk.
What could this be used for?	Educating data science leaders and enterprise architects on best practices when designing and evaluating ML solutions. Enable those responsible for AI to have confidence in model management and trust in reproducible AI results.
Key Ethical Themes	• Technical Robustness and Safety • Transparency • Diversity, Non-discrimination and Fairness • Accountability

Key Principles to Operationalize AI Solutions (Common to DataOps and MLOps)	• Ensure the AI solution is ethical and follows necessary regulatory compliance up front.
	  - Do not spend time building and operationalizing a solution that is not ethical.
	  - Assemble all stakeholders early to review the intended purpose.
	• Clearly identify success criteria and a methodology for ongoing monitoring.
	  - Define a feedback loop that captures the agreed upon success criteria.
	• Keep reusability in mind: consider platform agnostic frameworks.
	  - Solution architecture might differ depending on your model, available tools, and infrastructure.
	• Start building the end-to-end pipeline in parallel to your AI model development.
	  - Avoid leaving the technical implementation details to the end.
	  - Engage your data engineering team as early as possible in your project. Data scientists and data engineers will have interdependent goals and should work closely.
	• Have an agile mindset when it comes to delivering AI solutions.
	  - Deliver value in small, iterative increments when possible.
	  - Start with a simple model to test the end-to-end pipeline.
	• Adopt best practices and tools to make data scientists comfortable with continuous integration (CI) and continuous deployment (CD). This is crucial!
	  - GitOps principles ensure repositories become the single source of truth and pull requests are used to version control all changes.
	  - Automation tools such as Jenkins® are used to build, test, and deploy code in a controlled and predictable manner.
	  - CI/CD pipelines create consistent environments for running AI solutions by provisioning necessary compute resources (e.g., virtual machines, Docker® containers, or cloud servers), checking out code from a version-controlled repository, compiling the code, running tests, packaging binaries, and deploying binaries without manual configuration or intervention.
	• Consider separating your data pipeline (DataOps) from your ML pipeline (MLOps).

Jenkins is a registered trademark of Software in the Public Interest, Inc.

Docker is a registered trademark of Docker, Inc.

The DataOps and MLOps Process



Additional MLOps Key Principles

MLOps introduces principles of model management and model governance and borrows from development operations (DevOps) to provide controlled AI pipelines that enable smoother production deployments. The main steps include development, packaging, serving, and monitoring models.

- Version control your model code.
  - Apply to model training as well as model deployment.
  - Version control during model training provides hyperparameter tracking.
  - Software tools are available to help with this (e.g., DVC).
- Seek out a solution that easily packages and deploys models, preferably with interactive endpoints.
  - Having a machine-learning-as-a-service (MLaaS) platform is extremely beneficial. Maximizing the use of reusable pipelines or deployment software will greatly reduce time to market for AI models.
  - Consider supported languages/tools (e.g., Python™, R, Spark™), deployment infrastructure (e.g., Cloud, Docker containers, etc.), and required deployment patterns (e.g., canary, shadow, blue-green, etc.) when evaluating options.
  - Many options exist: AWS SageMaker, Azure Machine Learning, GCP, DataRobot, Domino Data Lab, seldon-core, mlflow, Kubeflow, TensorFlow, etc.
  - For example: TD's in-house Spinal Machine Learning eXecution (MLX) framework built in 2019 provides a CI/CD pipeline that leverages Docker to package model code, Jenkins to automate the build process, Bitbucket to version model code, seldon-core to expose RESTful APIs, and an in-house API to authenticate end-users.
- Consider performance requirements (scaling, recovery).
- Consider optimization requirements (throughput, latency).
- Provide a staging environment for quality assurance between the model development and production environments.
- Once deployed, monitor model performance on an ongoing basis.
  - Establish agreed upon metrics to sufficiently monitor for model quality issues. Decide if these metrics should be generic or specific (specific will require additional work but may produce more favorable results).
  - Consider re-training your model in case of a concept drift, identified using the metrics.
- Ensure appropriate model access management.
  - For example, if the model is exposed as an API endpoint, ensure there is appropriate authentication to guard against unauthorized access.
- Pair MLOps with appropriate model governance and model management principles.
  - Collect and store metadata to manage model access and use.
  - This information essentially creates a model store—models that can be reused by others and tracked/validated the same way as feature stores.

MLOps Resources

MLOps Resources

The following links provide additional information and technical guidance for MLOps.

- MLOps for Python models using Azure Machine Learning. Azure Reference Architectures. (2019). Retrieved 21 July 2020, from https://docs.microsoft.com/en-us/azure/architecture/reference-architectures/ai/mlops-python
- How to Deploy AI Solutions to Production. (2018). Retrieved 21 July 2020, from http://www.datarobot.com/blog/data-professional-persona-how-to-deploy-ai-solutions-to-production/
- Boykis, V. (9 June 2020). Machine learning is hard. Retrieved 8 July 2020, from http://veekaybee.github.io/2020/06/09/ml-in-prod/

Terms Used in DataOps and MLOps

- AI, ML, predictive models, models: Used interchangeably in this context to represent the code that produces predictive insights.
- Pipeline: A series of events scheduled together.
- Training: Developing a model on initial training data.
- Testing: Testing the performance of the model on a test set that it has not seen before.
- Serving, deploying, scoring: Moving the AI model into production, using the model to generate new insights (inferences) in the real world.

Testing and Optimization

Annotation and Benchmarking on Understanding and Transparency of Machine Learning Lifecycles (ABOUT ML)

Actionable Recommendation on How to Understand and Use this Resource

Develop a standardized and consistent documentation process across your organization. It could also be used by regulators as a basis for industry-wide documentation processes in future regulation.

Authors	Partnership on AI (PAI)
What is it?	Annotation and Benchmarking on Understanding and Transparency of Machine Learning Lifecycles (ABOUT ML) is a project of the Partnership on AI (PAI) working toward establishing new norms on transparency via documentation by evaluating best practices throughout the ML system lifecycle from design to deployment.
What is its purpose?	To offer a summary of recommendations and practices that is mindful of the variance in transparency expectations and outcomes. To provide an adaptive resource to highlight common themes about transparency, rather than a rigid list of requirements. To guide teams to identify and address context-specific challenges.
Who within a financial services organization would most benefit?	Head of AI/analytics, data scientists, compliance, risk, regulators, and policymakers.
What could this be used for?	Establish standardized processes on documentation in ML life cycles to improve transparency in ML systems.
Key Ethical Themes	• Transparency
Key Findings	Documentation for ML life cycles is not simply about disclosing a list of characteristics about the data sets and mathematical models within an ML system, but rather an entire process that an organization needs to incorporate throughout the design, development, and deployment of the ML system being considered. This documentation process begins in the ML system design and setup stage, including system framing and high-level objective design. At each step of this workflow, transparency and documentation need to be an explicit part of the discussion. The steps are: system feedback, system design and setup, system development, system deployment, and system maintenance. The paper provides suggested documentation sections for data sets and for models. It also points out current challenges of implementing documentation.
Key Ideas Include:	• Documentation is valuable both as a process and an artifact. • Internal documentation (for other teams inside the same organization, more detailed) and external documentation (for broader consumption, fewer sensitive details) are both valuable and should be undertaken together as they provide complementary incentives and benefits. • Avoiding misuse and harm from ML systems is a focus of current research and practice. Adhering to a documentation process that demands intentional reflection about how a system might be used and misused, in which contexts, and impacting whom is one first step toward potentially reducing harm. Incorporating feedback from diverse perspectives early, often, and throughout every stage in the ML lifecycle is another risk mitigation strategy. The Diverse Voices process from the Tech Policy Lab at the University of Washington is one formalized methodology for incorporating this type of feedback.

Find this resource at: https://www.partnershiponai.org/wp-content/uploads/2019/07/ABOUT-ML-v0-Draft-Final.pdf

Adversarial Robustness 360 Toolbox (ART)

Actionable Recommendation on How to Understand and Use this Resource

Leverage the 360 Toolbox open source algorithms and adversarial examples to defend deep neural networks (DNN) used in your organization. To get started with the Adversarial Robustness Toolbox (ART), visit https://developer.ibm.com/open/projects/adversarial-robustness-toolbox/.

Authors	IBM Research
What is it?	The Adversarial Robustness 360 Toolbox (ART) is an open source software library that supports both researchers and developers in defending algorithms against adversarial attacks, making AI systems more secure. The toolkit provides implementations of dozens of published attacks, with associated references to published papers. The attacks fall into three categories. Evasion attacks attempt to defeat a classifier causing it to produce an invalid result, such as misclassifying spam or malware. Extraction attacks probe a model in order to learn enough to reconstruct it. Poisoning attacks work against systems that continually retrain by injecting examples intended to compromise the learning process. The code library contains sample implementations of each of these types of attacks.
What is its purpose?	To educate researchers and practitioners about the threat of adversarial attacks on ML algorithms and to show how those algorithms can be hardened to mitigate such attacks.
Who within a financial services organization would most benefit?	Head of AI/analytics, data scientists, compliance, risk, regulators, and policymakers.
What could this be used for?	Researchers can use the toolbox to develop and benchmark novel defenses against state-of-the-art attacks. For developers, the library provides interfaces which support composition of comprehensive defense systems. The ART toolbox is developed with the goal of helping developers better understand and measure model robustness, model hardening, and improve runtime detection.
Key Ethical Themes	• Technical Robustness and Safety
Key Findings	• Adversarial attacks pose a real threat to the deployment of AI systems in security critical applications. • Virtually undetectable alterations of images, video, speech, and other data have been crafted to confuse AI systems. Such alterations can be crafted with or without access to training sets. • Adversarial attacks can be launched in the physical world: Instead of manipulating the pixels of a digital image, adversaries could evade face recognition systems. • Criminal access: Attackers can avoid facial recognition systems for critical access points. • Fraud: Attackers can: - Disrupt digital check deposit services. - Disrupt digital onboarding and Know Your Customer (KYC) services. - Propagate identity theft and fraud. • Skew training sets to cause misclassification or fraud. • Approach for defending DNNs is three-fold: - Measuring model robustness. First, the robustness of a given DNN can be assessed. A straightforward way for doing this is to record the loss of accuracy on adversarially altered inputs. Other approaches measure how much the internal representations and the output of a DNN vary when small changes are applied to its inputs. - Model hardening. Secondly, a given DNN can be "hardened" to make it more robust against adversarial inputs. Common approaches are to preprocess the inputs of a DNN, to augment the training data with adversarial examples, or to change the DNN architecture to prevent adversarial signals from propagating through the internal representation layers. - Runtime detection. Finally, runtime detection methods can be applied to flag any inputs that an adversary might have tampered with. Those methods typically try to exploit abnormal activations in the internal representation layers of a DNN caused by the adversarial inputs.

Find this resource at: https://github.com/IBM/adversarial-robustness-toolbox

AI Explainability 360 Toolkit

Actionable Recommendation on How to Understand and Use this Resource

This resource can be used by data scientists, model validators, and compliance teams to better understand the range of available AI explainability techniques. An accompanying paper gives an overview of the explainability toolkit. Together with links to published algorithms, code implementations, and tutorials, this site provides a trove of resources. In addition, a taxonomy and decision tree help guide the reader to select the appropriate explainability method for different applications and end users.

Authors	IBM Research
What is it?	An extensive open source Python® toolkit that supports interpretability and explainability of data and ML models. The toolkit includes a comprehensive set of algorithms that covers different dimensions of explanations along with proxy explainability metrics, with links to demos, videos, papers, and tutorials on explainable AI algorithms.
What is its purpose?	To educate the readers on AI explainability, provide examples and code, and guide the choice of the most appropriate explainable methods for different situations
Who within a financial services organization would most benefit?	Head of AI/analytics, data scientists, compliance, risk, regulators, and policymakers
What could this be used for?	Educating data scientists and regulators on current best practices in explainable AI. Developing trust and confidence in internal employees to determine how a decision was made. Developing trust and confidence in customers who want to understand an AI-based decision.
Key Ethical Themes	• Transparency • Accountability

Key Findings

- It provides an introduction and overview of the topic of explainability in AI. This topic has gained increasing importance as automated decision systems gain prevalence in high stakes applications that involve people in the decision making loop. In order to gain acceptance by all stakeholders, including the designers of such systems, users, executives, and the subjects of those decisions, there should be trust and insight into how those decisions are made. However, what constitutes an acceptable explanation is difficult to answer. It is an active area of research, and many different methods are being explored.
- A survey of recently published techniques covers a number of different approaches to explainability and highlights in what situations they might be applicable, depending on the level of understanding of the stakeholder and whether an explanation for particular decisions or the overall model is required.
- It provides a taxonomy of explainability methods and a decision tree to guide the user to the best algorithm for a given situation. The taxonomy classifies algorithms along two main properties: global vs. local and directly interpretable or post hoc explanation. A global model attempts to explain the overall working of a model, while a local explanation seeks to explain a particular decision and the factors that influenced it. Directly interpretable methods constrain the complexity of a model in such a way that the mechanics of the model's decisions can be inspected and understood. Post hoc explanations do not attempt to peer into the workings of the model, but provide an alternate means of understanding the model's workings (e.g., a simpler surrogate model). A further useful distinction for local explanations is whether they explain through whole samples or analyze the features of the particular sample in question. For example, a counterfactual explanation shows a similar sample that would have yielded a different decision. Saliency approaches highlight which features of the model inputs were important to making a decision and which were less relevant. Algorithms may be classed in the four quadrants of the taxonomy: global directly interpretable, global post hoc, local directly interpretable, and local post hoc.
- The key contribution of this resource is an open source library of Python code with implementations of 19 different explainability algorithms, some developed by IBM research and others taken from the collection of published papers cited on the website.
  - Algorithms: Eight algorithms are provided, which are Boolean Decision Rules via Column Generation (BRCG), Generalized Linear Rule Models (GLRM), ProtoDash, ProfWeight, Teaching Explanations for Decisions (TED), Contrastive Explanations Method (CEM), Contrastive Explanations Method with Monotonic Attribute Functions (CEM-MAF), Disentangled Inferred Prior Variational Autoencoder (DIP-VAE). The toolkit also includes two metrics from the explainability literature: faithfulness and monotonicity.
- The last important contribution of this resource is that it provides tutorials of decision making in several important contexts such as financial services, healthcare, and human resources. These tutorials include links to publicly available data sets and demos of how the algorithms provided in the open source toolkit are relevant to each situation.

In addition to this IBM resource, which appears to be the most comprehensive at present, our data science community also recommends several other explainability resources: Google Explainable AI, Microsoft Interpret, H2O Machine Learning Interpretability, and Oracle Skater.

Find this resource at: https://aix360.mybluemix.net/

Registered trademark of Python Software Foundation

AI Fairness 360 Open Source Toolkit

Actionable Recommendation on How to Understand and Use this Resource

It should be used to understand the state-of-the-art algorithms in this area and how they may be deployed to measure and mitigate systematic bias in automated decision systems. The open source code library may be directly used by data scientists and validators to measure bias or verify that models do not contain bias. However, as the guidance states, this is a complex area and the code and this material should only be taken as a starting point to broader discussion with multiple stakeholders.

Authors	IBM Research
What is it?	A web page with links to demos, videos, papers, tutorials, and a code library of implementations of algorithms for fairness in AI. It provides an overview of research on the topic of fairness in AI and an open source library containing 20 fairness metrics and 10 state-of-the-art algorithms for mitigating discrimination and bias in ML.
What is its purpose?	To educate the readers on AI fairness, provide examples and code, and guide the choice of the most appropriate methods for different situations.
Who within a financial services organization would most benefit?	Head of AI/analytics, data scientists, compliance, risk, regulators, and policymakers
What could this be used for?	To educate stakeholders such as model developers, model validators, executives, and regulators on the state-of-the-art algorithms in this area and how they may be deployed to measure and mitigate systematic bias in automated decision systems. The open source code library may be directly used by developers and validators to measure bias or validate that models do not contain bias.
Key Ethical Themes	• Accountability • Transparency • Bias and Fairness • Responsibility • Promotion of Human Values
Key Findings/ Contributions	• It provides an overview of the topic of fairness in AI, including a glossary of terms and a list of references to the important and most recent published papers in the field. • The resource includes a guide to published metrics of fairness and mitigation measures and where each might be appropriate. It gives a general taxonomy of the different measures of fairness. Fairness is a context-dependent social construct and as such, there is not a single agreed upon measure of fairness. There are many possible ways of measuring depending on the objectives. The taxonomy divides fairness measure into two basic groups: individual vs. group. Individual fairness is the objective that similar individuals should be treated similarly by a model. Group similarity requires statistical measures of a selected metric to produce similar results between groups, as defined by a protected or sensitive attribute (e.g., gender, ethnicity). For measures of group fairness, there is the additional consideration that the bias may exist in the data or the model. This means that the data used to train an AI model may itself be biased and should be assessed before building any algorithm. The toolkit provides implementations of many fairness metrics which may be applicable to individuals or groups, models, and data. • The open source code library includes implementations of dozens of metrics for measuring fairness in a data set or model output. The library also includes implementations of 10 different algorithms for mitigating bias. They are divided into three classes: preprocessing, in-processing, and postprocessing. The preprocessing algorithms act to de-bias data before fitting a model, either by changing features related to sensitive attributes, obfuscating sensitive attributes, or reweighting the data. In-processing algorithms influence the model during training, and postprocessing algorithms act to reweight the outputs of a trained model. • Two detailed tutorials that include code and data and explain how to work through a problem of measuring and mitigating bias in the context of credit decisions and medical expenditure. In the first use case, the sensitive attribute is age. They define the privileged group as those aged 25 or over and measure the difference in mean outcomes between the privileged and unprivileged sets. They apply a reweighting algorithm to adjust the outcomes in the data set. Re-running the same fairness metric shows this intervention on the data eliminates the measured bias in the data. The transformed data set may be used to fit a classifier, with further work to show whether the classifier trained on reweighted data more fair than one trained on the unmodified data set. In the second use case, a model predicts healthcare utilization of individuals in order to allocate resources. The sensitive attribute here is ethnicity. Two models are trained on the raw data and their between-group difference is measured using six different fairness metrics. The models are retrained using both data preprocessing and in-processing mitigation algorithms, showing a reduction in the disparity between the groups on the six metrics.

Find this resource at: https://aif360.mybluemix.net/

Coronavirus Impact on the Financial Services Industry

The pandemic and related containment measures have deeply affected Canadians and the Canadian economy. Data collection, data sharing, and digitization of behaviors have all accelerated, but the impacts have been different across sectors and regions, and the economic recovery will also look different across the country as local economies start to re-open.

How do we ensure the health and safety of our employees and customers?

How do we support the dire economic circumstances of our communities and country?

How do we respond to the appropriate demand for social justice?

And in the midst of all of these, how do we ensure our financial services organizations are meeting their regulatory requirements?

Each of us has questions about the current state and the evolution of the regulatory environment, given the pandemic's impact. Moreover, how are regulators likely to respond to the banks' financial and operational challenges resulting from the coronavirus' effect on the economy?

A CEO and board executive summit in June 2020 on the coronavirus's impact on the financial services industry hosted by a panel of former senior regulators highlighted **three short term (12 to 24 months)** major impacts:

1. **Compliance:** Despite regulators being more flexible in some areas, compliance is not going away—on the contrary, social media will ensure it is top of mind for institutions, the public, and regulators. Institutions should keep compliance, including lending discrimination, not just in mind but integral to policies, procedures, and actions. Further, it is impossible to keep up with a bank's compliance obligations and do so fairly and consistently without data and robotic and advanced technologies. The data and technology needed for compliance will affect all controls, including prudential controls and management.

2. **Operational Resiliency:** Banks are experiencing increased operational risk due to employees working from home, absenteeism, increased call volumes, cyber activity, and fraud schemes. Regulators will be looking at banks' ability to manage elevated operational risk successfully.

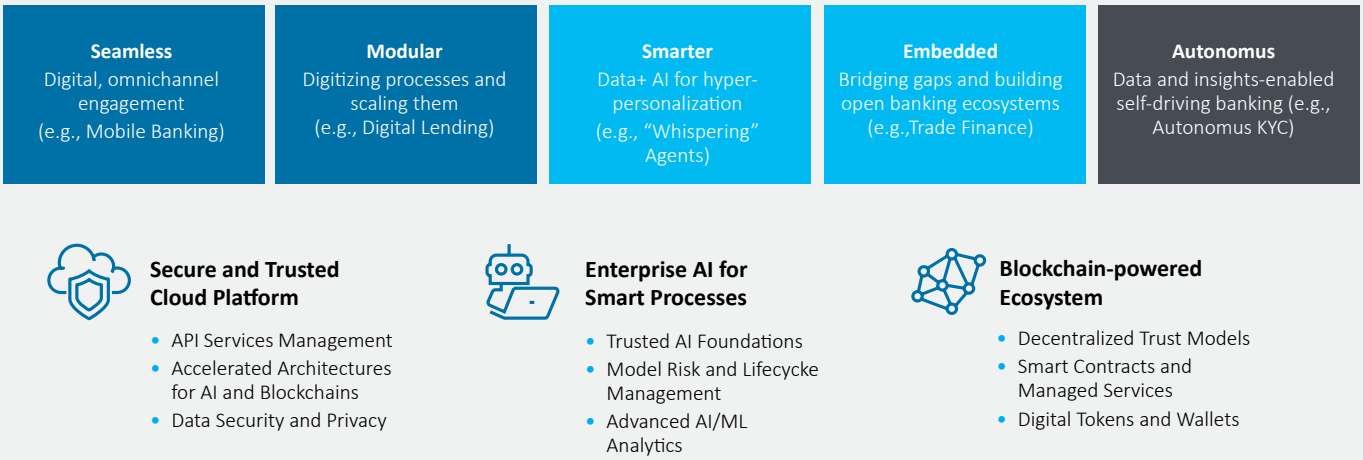
3. **Consolidation:** There will likely be additional consolidation because of the pandemic and other reasons, but it is not likely to result in a handful of very large banks. Consolidation is more likely to occur with smaller banks that have difficulty navigating through this crisis from a profitability or capital perspective or have not made the requisite investment in technology.

In the **mid term (1 to 5 years)**, the senior regulatory experts expect there to be cost pressures, technology disruptions, and regulatory changes that will drive the following ten unique characteristics for the financial services industry:

1. Fungibility of products and services,
2. Heterogeneity of customers or clients,
3. High intensity of service interactions,
4. Long-term customer relationships,
5. Direct impact on customer's sense of well-being and safety,
6. Global distribution channel with intermediaries,
7. Significant regulatory oversight from multiple agencies,
8. Convergence of sales, marketing, and operations,
9. Financial services influence and integration with other industry ecosystems, and
10. Most targeted industry for cybercrimes.

Given these characteristics, innovation powered by transformative technologies will drive seamless, modular, smarter, embedded, and autonomous operations in the financial services:

Figure 5: Post-pandemic financial services innovation, powered by transformative technologies



Characteristics of Post-Pandemic Financial Services Innovation

1. **Seamless:** Financial services firms are using mobile and social media technologies in addition to traditional channels to engage customers and provide a seamless omni-channel digital experience.(CIO)
2. **Modular:** Financial institutions are increasingly digitizing their internal processes and exposing them through APIs and "as-a-service" platforms. In some cases, this is driven by regulations (e.g., PSD2) as well as by the desire to partner with fintech firms and others.
3. **Embedded:** Once the financial processes are modularized and exposed as APIs, how can they be best embedded in other industry value chains? For instance, trade finance services reduce the financial friction in global supply chains by providing the working capital necessary to finance the export and import of goods.
4. **Smarter:** Financial services acquire data from myriad sources that help to deeply profile customers as well as personalize their services. Leveraging this data will require big-data analytics including AI/ML. Being a regulated industry, such AI/ML techniques will have to be explainable for them to be used; financial institutions may not be allowed to deploy "black box" methods. In some jurisdictions, it is already a consumer right to consent and appeal to any automated decisions.
5. **Autonomous:** Financial institutions are exploring all-digital end-to-end automated services such as digital banking channels with autonomous on-boarding and service management, thereby dramatically reducing the marginal cost of adding new services and disrupting the market. Such capabilities will require advanced security and scalability and decision intelligence and are actively being pursued through digital subsidiaries to lower the business risks.

Canadian Response to the Crisis to Date

Decisive policy actions by the Bank of Canada, governments, and other authorities have supported household incomes and helped businesses stay afloat through the lockdown period (Figure 6). These actions laid the necessary foundation for eventual recovery. Financial market conditions have improved, and credit is flowing to households and businesses when they need it most. These actions will ensure a well-functioning financial system to support the emerging recovery and help achieve Canada's 2% inflation target.

Figure 6: Potential impact of COVID-19 pandemic in Canada[4]

Slower recovery

- Spread of virus and related control measures
- Tightening financial conditions
- Uncertainity and confidence effects
- Long-lasting structural economic damage

Faster recovery

- Fiscal measures, including support to households and business
- Credit and liquidity support
- Monetary policy

[4] https://www.bankofcanada.ca/2020/06/our-covid-19-response-navigating-diverse-economic-impacts/

Canada's banks have launched comprehensive programs to make a positive difference for personal customers, small business customers, employees, and communities. According to Canadian Bankers Association (CBA), here are some examples:

1. Reducing credit card interest rates, deferring payments, and instituting low minimum payments on credit cards and lines of credit. Just two months after the pandemic was declared (24 June 2020), more than 450,000 credit card deferral requests were in progress or had been completed by eight banks.
2. Canada's banks are offering mortgage payment relief to customers by way of deferred mortgage payments. As of 24 June 2020, thirteen CBA member banks had provided help through mortgage deferrals or "skip a payment" to more than 743,000 Canadians, representing approximately 15% of the mortgages in bank portfolios.
3. Banks are delivering fast access to the Canada Emergency Business Account (CEBA) program, providing small and medium-sized companies interest-free loans of up to $40 thousand, including facilitating the program's recent Phase 3 expansion. As of 3 July 2020, 688,000 CEBA loans had been approved by financial institutions, including banks, representing roughly $27.4 billion in interest-free credit for eligible businesses.
4. Banks have implemented intensive cleaning programs to ensure that workplaces, including branches, remain as safe as possible for everyone. They pay employees bonuses and provide extra paid days off to customer-service employees who work in branches and call centers. Banks are also implementing broad-based work from home options for any roles that can be performed remotely to support public health efforts and employees' well-being.
5. For communities, seven banks have donated a combined $10.8 million to support frontline healthcare workers and community services urgently needed for vulnerable individuals affected by the public health, social, and economic consequences of COVID-19. Gifts to Canadian charities such as United Way Centraide Canada, Food Banks Canada, and Breakfast Club Canada provide essential services, community services, and senior citizens' services.

Envisioning the New Normal with Trusted Data and AIS

Long after the medical threat has passed, the pandemic will bestow lasting consequences on business and society. A dramatically different "new normal" calls for financial institutions to play a crucial role in their business operations and their clients' lives.

They will need to step up and guide customers through economic and financial instability, and they will need to help those customers navigate and even thrive in an uncertain world. Considered together, the challenges of this next environment point to the need for new business architecture. Forward-looking banks and financial services organizations can respond to these unique circumstances by accelerating their migration to the new architecture; many financial institution employees have already demonstrated admirable adaptability. This new architecture can accelerate workforce reskilling, agile ways of working, and strengthening both client and employee relationships, whether face-to-face or virtual.

Until now, client-centric operations have been anchored to products typical of "output economies," in which customers are those who buy. The next normal will likely accelerate the transformation of human-centric, service-based platforms—ones that place relationships front and center. These platforms are based upon value-generating interactions typical of "outcome economies," in which customers achieve their goals through seamless experiences. Consider interactions that create transparent banking relationships, directly or digitally augmented, where trust is generated through a value exchange between banks and their most precious assets, their clients. These human-centric, service-based platforms could dramatically change existing banking architectures and their corresponding business models.

A bank's purpose could evolve from credit institutions, which provide relevant accessory solutions (payments, investment, insurance), into platform-driven centers of competence (CoC). These CoCs would integrate lending operations into advisory relationships for families and businesses. The emphasis would shift from distribution channels of lower-margin products to relationship-based services built on client engagement and experience. The bank of the future will redesign customer proximity not only by using data to personalize their offers (output economy) but also by infusing AIS into interactions; the "data-driven bank" will be based on the "data-enabled customer." This model, driven by close digital relationships between banks and customers, will be powered by AIS and will be able to generate new value even during a crisis such as a pandemic lockdown. Trusted digital relationships are, therefore, not only a real asset of financial institutions facing a different normal but a necessary mechanism to help communities weather the storm and emerge more robust and ready for the future.

The Trusted Data and AIS Ethics Landscape

This section provides a view of trusted data and AIS in contexts: by country of issuer, private sector, and then a focused look at financial services, as of October 2020. From the perspective of strategic public communications and awareness, the sub-section illustrates where the financial services industry compares in terms of contribution to narrative and discourse regarding ethics and AIS in practice. Considering media engagement, events, and public awareness initiatives, it begins with a brief analysis of private sector responses to national AIS strategies and guidelines which are led predominately by Big Tech (e.g., Google, Microsoft, IBM, Facebook, etc.). Finally, it aims to demonstrate where the financial services industry ranks in the global conversation and suggests that as the discourse moves more toward ethics implementation, a considerable amount of work is required to solidify reputational management plans.

This sub-section is useful for individuals and teams responsible for strategic and global communications planning, public relations, marketing, or for those who are interested in knowing more about public engagement, education and awareness, and the mainstream ethical discourse on AI.

Landscape of AI Ethics in Action in the Private Sector

As demonstrated in the previous section, both governments and international organizations have responded to social fears of AI by appointing ad hoc expert committees often commissioned with drafting policy recommendations. The private sector's response has been swift addressing anticipated regulation or, as some advocates might argue, attempting to advocate for a "soft policy" approach. As a result, the industry embarked on numerous activities to manage the public perception of trusted AI.

A meta study of global guidelines and guiding principles done by researchers in 2019 identified 84 documents containing ethical principles or guidelines for AI, 88% of which were released after 2016. Most of these documents were produced by private companies (22.6%). A separate study done by PwC and based on 59 AI Principle documents highlights the increase in activity in the tech sector in 2018.

Ethical Guidelines for AI by Country of Issuer

Name of Document (Linked)	Issuer	Country of Issuer
Tieto's AI Ethics Guidelines	Tieto	Finland
AI Guidelines	Deutsche Telekom	Germany
Sony Group AI Ethics Guidelines	SONY	Japan
AI Principles of Telefonica	Telefonica	Spain
The Ethics of Code: Developing AI for Business with Five Core Principles	Sage	UK
DeepMind Ethics and Society Principles	DeepMind Ethics and Society	UK
The Responsible AI Framework	PriceWaterhouseCoopers UK	UK
Responsible AI and Robotics. An Ethical Framework	Accenture UK	UK
AI—Our Approach	Microsoft	USA
Artificial Intelligence. The Public Policy Opportunity	Intel Corporation	USA
IBM's Principles for Trust and Transparency	IBM	USA
Our Principles	Google	USA
Everyday Ethics for Artificial Intelligence. A Practical Guide for Designers and Developers	IBM	USA
Intel's AI Privacy Policy White Paper. Protecting Individuals' Privacy and Data in the Artificial Intelligence World	Intel Corporation	USA
Introducing Unity's Guiding Principles for Ethical AI—Unity Blog	Unity Technologies	USA
Responsible Bots: 10 Guidelines for Developers of Conversational AI	Microsoft	USA
SAP's Guiding Principles for Artificial Intelligence	SAP	Germany
Position on Robotics and Artificial Intelligence	The Greens (Green Working Group Robots)	EU
Commitments and Principles	OP Group	Finland

Beginning in 2018, there has been a significant increase in AIS-related legislation in congressional records, committee reports, and legislative transcripts around the world. This suggests that the mainstream is picking up on the language of harm prevention in AIS development opportunities. In more than 3,600 global news articles on ethics and AIS identified between mid-2018 and mid-2019, the dominant topics are framework and guidelines (32% of all articles analyzed) followed by Big Tech Advisory on Tech Ethics as indicated in the curated recap of key private sector actions, below.

Private Sector (Non-financial Services)

	Category	Company or Organization	Date
Project Proposal Collaboration	PR/Research	US National Science Foundation and Amazon	May 2020
Announcement of AI Ethics Board	PR/Ethics Advisory	Google	March 2019
Facebook invests US $7.5M in Centre on Ethics and AI at the Technical University of Munich, Germany	PR/Education Investment	Facebook	January 2019
Axon Announces Three New Members to AI Ethics Board	PR/Ethics Advisory	Axon	March 2020
Rome Call for AI Ethics	PR	Vatican Academy for Life, IBM, Microsoft, FAO, Italian Government	February 2020
Partnership on AI	Partnership	Amazon as founding member	January 2017
OpenAI	Partnership	Elon Musk founded	2015
HSBC's Principles for the Ethical Use of Big Data and AI	PR/Corporate Guidelines	HSBC	February 2020
DeepMind Ethics Council (London)	PR/Ethics Advisory	DeepMind	January 2016
Scholarships to open the field of AI	PR/Education Investment	DeepMind	n/a
Microsoft FATE	PR/Research	Microsoft Research	Ongoing
Aspen Institute Roundtable on AI	PR/Conference	Multiple	2019
Microsoft joins Veritas	PR/Collaboration	Microsoft and MAS	May 2020

Recent Media Coverage of AIS Ethics

Shifting to the financial services sector, at the top tier, central banks have significantly increased interest in AIS as reflected in the mentions of AI in official speeches and publications. This more intensive communication reflects greater efforts to understand AI and regulatory environments as they relate to the macroeconomic environment and financial services. The Bank of England, the Bank of Japan, and the Federal Reserve have mentioned AI the most in their communication.

In the corporate domain, compared to all other sectors, finance had the largest number of AI mentions in earnings calls from 2018 to Q1 of 2019, followed by the electronic, technology, and services sectors. However, with the tech sector leading in trusted data and AI public relations activities, this suggests that the financial sector can drive the discussion in trusted data and AI. Generating good stories could be in the form of initiatives to preserve the trust of customers as well as transparency around AI use.

Financial Services

	Category	Company or Organization	Date
Veritas	PR/Industry Consortium	Monetary Authority of Singapore (MAS)	2019
Code ethics in AI	PR/Strategy/Partnership	Standard Chartered	January 2020
Ethics and AI Conference	PR/Conference	Scotiabank	November 2019
Trusted Data and AI Training	PR/Education/Training	Scotiabank	November 2019
Ethics and AI Guidelines	PR	Scotiabank	November 2018
Corporate Social Responsibility	CSR	Scotiabank	2019
Philanthropic donation to launch Ethical AI Initiative University of Ottawa	PR/Media	Scotiabank	January 2020
Philanthropic donation to launch AI Research at University of Alberta	PR/Media	Scotiabank	November 2019
Op-ed: AI Revolution Needs a Rulebook, Here's a Beginning	PR/Media	Scotiabank	December 2018
Borealis AI	Research Lab	RBC	March 2018
Donation to CIFAR	PR/Philanthropy	RBC	October 2018
RBC backs university program on ethics in AI and data analytics	Philanthropy	RBC	November 2019
The new NextAI initiative	Donation	RBC	January 2017
University of Toronto AI Research Program	Investment	BMO	October 2019
Responsible AI in Financial Services	Industry Report	TD	September 2019

The financial industry is not far behind Big Tech both in AIS adoption and trusted data and AIS discussions. Most activity is driven by top-down edicts or an attempt at "soft law" and internal self-regulations. For trusted data and AIS to be realized, considerable effort should be invested into planning and monitoring key risk drivers and culture management.

References

Data can be accessed at https://drive.google.com/drive/folders/1Tl2HyuXHTGufDTsF-h0cb0InlMD3gvSQ

Raymond Perrault, Yoav Shoham, Erik Brynjolfsson, Jack Clark, John Etchemendy, Barbara Grosz, Terah Lyons, James Manyika, Saurabh Mishra, and Juan Carlos Niebles, "*The AI Index 2019 Annual Report*," AI Index Steering Committee, Human-Centered AI Institute, Stanford University, Stanford, CA, December 2019.

Anna Jobin, Marcello Ienca, Effy Vayena, "*Artificial Intelligence: The Global Landscape of Ethics Guidelines,*" A Health Ethics and Policy Lab, ETH Zurich, 8092 Zurich, Switzerland.

The Global and Canadian Regulatory Landscape

With respect to activities of financial institutions, two types of regulators have taken active interest or concrete steps to regulating AI globally: privacy regulators, who regulate consumer privacy matters, and prudential and securities regulators, who regulate banking and financial market matters.

Canadian and Global Overview

Privacy Regulatory Oversight

Global trends. The General Data Protection Regulation (GDPR) in the European Union (EU) has been the leading regulation in terms of privacy including AI-related issues. It was first introduced in 2016 (and later imposed in 2018) to address personal data protection issues within the European Economic Area (EEA). GDPR has served as a model and influenced other lawmakers to draft similar legislation to protect personal data. For instance, the California Consumer Privacy Act (CCPA), a state statute that is intended to enhance privacy rights and consumer protection in California, follows a similar chain of thought and forces some of the world's largest tech giants to comply. In Singapore, the privacy regulator—the Personal Data Protection Commission—issued two editions of its Model AI Governance Framework, which is a sector-, technology-, and algorithm-agnostic framework. The model framework recommends approaches to govern data analytics—including AIS—in a responsible/ethical way.

Canadian landscape. In Canada, analytics and AIS regulation are closer than one may think. In Quebec, a GDPR-like bill (Bill 64) was introduced in June 2020, which reflects most GDPR requirements pertaining to AIS. At the federal level, entities are governed by the Personal Information Protection and Electronic Documents Act (PIPEDA). The act applies to private-sector organizations across Canada that collect, use, or disclose personal information during their commercial activities. This act further restricts the use of personal data and actualizes the right of individuals to access information held by organizations and challenge its accuracy. A new Digital Charter Implementation Act (DCIA, or Bill C-11) was tabled in November 2020, which if passed would replace PIPEDA with the new Consumer Privacy Protection Act (CPPA), introducing significant updates related to AIS. It includes several GDPR-like requirements with respect to AI including the right to explainability and algorithmic transparency for all automated decision systems, the right to data mobility, the right to data deletion, changes to consent for data collection, and generally stricter enforcement powers for the Office of the Privacy Commissioner, including a tribunal who would have the authority to fine organizations in breach of the CPPA.

Innovation, Science and Economic Development Canada, the department of the Government of Canada with a mandate of fostering a growing, competitive, and knowledge-based Canadian economy, has also publicly discussed the creation of a new framework to further regulate AI and open banking.

Despite continuous efforts in planning for these regulations, how to effectively adopt these requirements in a systematic manner across the organization is still not entirely clear. The cost/resources/time to implement these regulatory trends are not to be underestimated. Financial institutions must be proactive and plan for these regulatory trends.

Prudential Regulatory Oversight Overview

Global trends. In terms of prudential and securities regulatory oversight, the Monetary Authority of Singapore (MAS) has been the international trendsetter. In 2018, MAS published its now renowned FEAT principles (Fairness, Ethics, Accountability, and Transparency), providing for a first high-level framework for FIs when instating AIS governance measures.

Canadian landscape. In Canada, the Office of the Superintendent of Financial Institutions (OSFI) circulated a survey in 2019 among several Canadian financial institutions inquiring about technology-related risks including strategic orientations, business uses, recruitment efforts, and AIS ethics and governance measures. The survey was followed up by the release of a discussion paper in September 2020 titled *Developing Financial Sector Resilience in a Digital World: Selected Themes in Technology and Related Risks.* The paper addresses several pressing technology (non-financial) risks, which includes AIS/ML (part of the broader category that OSFI refers to as advanced analytics). OSFI explains that AI may present some challenges when it comes to model risk management (i.e., the heart of guideline E-23), namely with respect to "continuously evolving models and the use of AI in validation." The discussion paper highlights three core ethical principles that institutions need to consider when using advanced analytics: 1) soundness (which includes auditability and fairness), 2) explainability, and 3) accountability. The paper also presents principles for two other technology related risks in addition to advanced analytics: cybersecurity (confidentiality, integrity, and availability), and third-party ecosystem (transparency, reliability, and substitutability). The continued emphasis on principles over prescription falls in line with existing OSFI regulations, and the discussion paper presents a promising first step toward updated regulatory and supervisory frameworks. From September through December 2020 OSFI accepted public comments on the paper and performed several consultations with academic and industry stakeholders to ensure relevancy and appropriateness of the principles.

Top Commitments to Consumers

In terms of concrete regulatory obligations for which FIs should prepare for, the following are the most important ones. The ones stemming from privacy legislation and regulators are clearer, as they are the most advanced in terms of legislative amendments and regulatory requirements.

Transparency. Transparency refers to the commitment of organizations that are honest and transparent while disclosing relevant information to their constituents. Bill 64 (and likely the PIPEDA reform package mentioned above) imposes a duty for data controllers including FSIs to disclose to data subjects when they are the subject of a decision based exclusively on automation using personal data. Automated decision making has to be monitored by human employees, with submission of human observations. Furthermore, with the OPC consultation, the regulator suggests a right of consumers to opt-out of decisions made solely by machines and requests for human decision making. In addition, Bill 64 outlines the right for the data subjects to "be informed when interacting with AI applications," which means FIs must disclose the existence of AI and automation to their constituents.

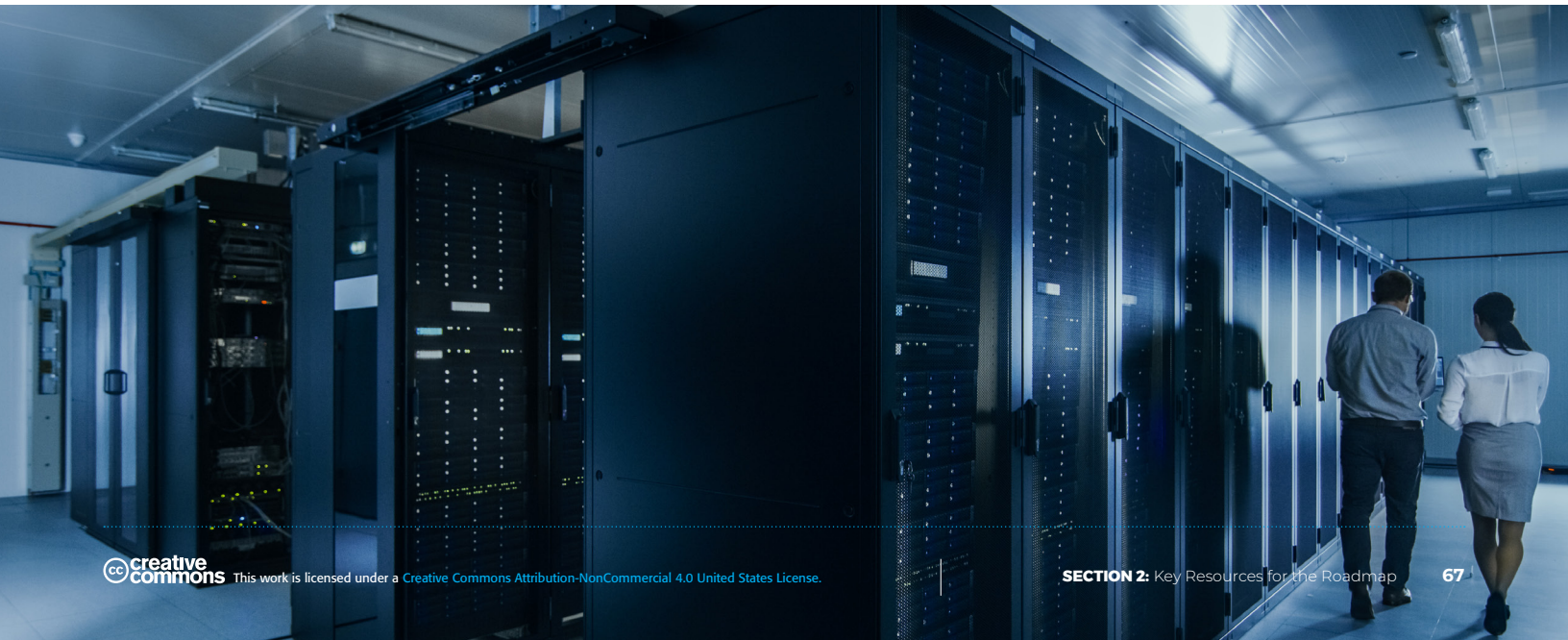
Explainability. Broadly speaking, explainability consists of the right to request an explanation for who and how the decision is made. Similarly, in both Bill 64 and likely the PIPEDA reform package, obligations regarding explainability are outlined for any decisions that are made exclusively automatically. In a nutshell, the bill requires an organization to inform data subjects of personal data that was used in automated decisions and all the factors that led to a decision. Some of these requirements are triggered only when the decision rendered is exclusively based on automated processing. The interpretation of "exclusively" is not further explained in Bill 64 and has not been analyzed by a tribunal.

The concept of "exclusively" based on automated processing is also present in the GDPR (see quote below); thus the analyses of what it means in the EU can be useful. The Information Commission's Office gives useful insight as to what may be considered like such a decision. Although not applicable in Canada, the GDPR can still guide Canadian FIs:

> *Solely means a decision-making process that is totally automated and excludes any human influence on the outcome. A process might still be considered solely automated if a human inputs the data to be processed, and then the decision-making is carried out by an automated system.*
>
> *A process won't be considered solely automated if someone weighs up and interprets the result of an automated decision before applying it to the individual.*
>
> *Many decisions that are commonly regarded as automated actually involve human intervention. However, the human involvement has to be active and not just a token gesture. The question is whether a human reviews the decision before it is applied and has discretion to alter it, or whether they are simply applying the decision taken by the automated system.*

Fairness. In Canada, legally speaking, fairness is largely governed by anti-discrimination legislation, which is rooted in protecting equality rights. These legal obligations can also inform ethical ones that a financial institute may want to undertake. For instance, they may want to consider the following from PIPEDA and the GDPR:

> *In terms of AI, the OPC consultation suggests addition of anti-discrimination measures to PIPEDA. Namely, it suggests a Human rights-by-design requirement. The proposals do not address how such provisions would interact with other anti-discrimination legislation in Canada.*
>
> *In the EU, the GDPR addresses equality rights matters in a way that is similar to what is being suggested by the OPC. The GDPR prohibits the processing of "personal data revealing facial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation" without explicit consent (and other applicable exceptions).*



Top Commitment to Regulators

Financial institutions can expect new legal requirements to provide transparency and auditability to consumers and regulators that will help build trust among the institutions' constituents. Regulators will also benefit from the increased transparency. Currently, most new requirements stem from privacy legislation; however, legislation is also expected in the two following areas:

Transparency, Auditability, and Explainability

After the OSFI's survey in 2019 among several Canadian FSIs, regulators announced a technology risk discussion paper to be published, which includes AI and machine learning (ML). OSFI explains that AI may present some challenges when it comes to model risk management (i.e., the heart of guideline E-23), namely with respect to "continuously evolving models and the use of AI in validation." For instance, E-23 contains requirements akin to those stated in responsible AI corpus as transparency and explainability.

OSFI will subsequently develop regulatory expectations around the use of AI. OSFI states that it is engaging with key stakeholders to ensure relevancy and appropriateness of these supervisory expectations. The discussion paper, titled "Developing Financial Sector Resilience in a Digital World: Selected Themes in Technology and Related Risks" was released in September 2020.

Accountability

The Digital Charter Implementation Act (DCIA, or Bill C-11) suggests several new governance requirements that might also overlap with OSFI requirements. Beside transparency, the principle that regulators strive to emphasize is accountability, instructing organizations to document model development and monitor the models afterward. To ensure the end-to-end accountability, data traceability is deconstructed into data lineage and other data protection practices. Execution of this accountability plan include: information such as where data is originally from, how it is collected and curated, where it moves in the organization from its source to its destination, how the data gets transformed, where it interacts with other data, how it is prepared and labeled, having audit trails recording the correlations and inferences made algorithmically in the prediction process, how data accuracy is maintained over time. All of these are based on the principle of accountability. Bill C-11 also codifies accountability principles, namely by a more prescriptive privacy management program to document privacy practices.

On a larger scale, the concepts like privacy-by-design and fairness-by-design have triggered the creation of tools like Risk Impact Assessment Tool and Privacy Impact Assessment to assess and mitigate privacy and fairness concerns. In short, regulators do not just ask the corporations to protect their constituents' data and ensure fairness in all decision making, they also build the fairness into the core of applied analytics applications, protecting all.

EU's Assessment List for Trustworthy Artificial Intelligence (ALTAI)

In 2019, High-Level Expert Group on Artificial Intelligence (AI HLEG), set up by the European Commission, published the *Ethics Guidelines for Trustworthy Artificial Intelligence.* In July 2020, the final Assessment List for Trustworthy AI (ALTAI) was presented by the AI HLEG. The ALTAI is intended for self-evaluation purposes. It provides an initial approach for the evaluation of trustworthy AI. It builds on the one outlined in the *Ethics Guidelines for Trustworthy AI* and was developed over a period of two years, from June 2018 to June 2020. In that period this ALTAI also benefited from a piloting phase (second half of 2019). Through that piloting phase, the AI HLEG received valuable feedback through fifty in-depth interviews with selected companies; input through an open work stream on the AI Alliance to provide best practices and via two publicly accessible questionnaires for technical and non-technical stakeholders.

The ALTAI is firmly grounded in the protection of people's fundamental rights, which is the term used in the European Union to refer to human rights enshrined in the EU Treaties, the Charter of Fundamental Rights (the Charter), and international human rights law. This ALTAI is intended for flexible use: organizations can draw on elements relevant to the particular AI system from the ALTAI or add elements to it as they see fit, taking into consideration the sector in which they operate. It helps organizations understand what trustworthy AI is, in particular what risks an AI system might generate, and how to minimize those risks while maximizing the benefit of AI. It is intended to help organizations identify how proposed AI systems might generate risks, and to identify whether and what kind of active measures may need to be taken to avoid and minimize those risks.

Organizations will derive the most value from this ALTAI by active engagement with the questions it raises, which are aimed at encouraging thoughtful reflection to provoke appropriate action and nurture an organizational culture committed to developing and maintaining trustworthy AI systems. It raises awareness of the potential impact of AI on society, the environment, consumers, workers, and citizens (in particular children and people belonging to marginalized groups). It encourages the involvement of all relevant stakeholders. It helps to gain insight on whether meaningful and appropriate solutions or processes to accomplish adherence to the seven requirements (as outlined above) are already in place or need to be put in place. This could be achieved through internal guidelines, governance processes, etc. A trustworthy approach is key to enabling responsible competitiveness by providing the foundation upon which all those using or affected by AI systems can trust that their design, development, and use are lawful, ethical, and robust. This ALTAI helps foster responsible and sustainable AI innovation in Europe. It seeks to make ethics a core pillar for developing a unique approach to AI, one that aims to benefit, empower, and protect both individual human flourishing and the common good of society. We believe that this will enable Europe and European organizations to position themselves as global leaders in cutting-edge AI worthy of our individual and collective trust.

To summarize the major Canadian and US AIS regulatory guidelines, we have included a cross-referenced table of the ALTAI ethical principles discussed in several Canadian and global regulations in Figure 7.

Figure 7: Cross-referenced Summary of ALTAI Principles and Canadian/US AIS Regulation

	Major Canada/US Jurisdictions								
EU's July 2020 Requirements Assessment List for Trustworthy AI (ALTAI)	**Canadian Federal Gov**	**Ontario**	**Quebec**	**Alberta**	**Manitoba**	**Saskatchewan**	**British Columbia**	**US Federal Gov**	**California**
#1 Human Agency and Oversight Human Agency and Autonomy Human Oversight	✓	✓	✓	✓	✓	✓	✓	✓	✓
#2 Technical Robustness and Safety Resilence to Attack and Security General Safety Accuracy Reliability, Fall-back Plans and Reproducibility	✓	✓	✓	✓	✓		✓	✓	✓
#3 Privacy and Data Governance Privacy Data Governance	✓	✓							✓
#4 Transparency Traceability Explainability Communication	✓		✓	✓	✓	✓			✓
#5 Diversity, Non-discrimination and Fairness Avoidance of Unfair Bias Accessibility and Universal Design Stakeholder Participation	✓			✓					✓
#6 Accountability Environmental Well-being Impact on Work and Skills Impact on Society at Large or Democracy	✓								✓
#7 Accountability Auditability Risk Management	✓	✓			✓			✓	✓

What are the Applicable Standards and Certifications that Support Ethically Aligned Design?

In IEEE, the leading project that is in progress to support the framework within ethically aligned design is IEEE P7000™. This initiative establishes a set of processes by which organizations can include consideration of human ethical values throughout the stages of concept exploration and development. This initiative supports management and engineering in transparent communication with selected stakeholders for values elicitation and prioritization.

Although IEEE P7000 is not being developed exclusively for FSIs, the standard project can be used in ensuring the early compliance of the leading standards for financial services providers in developing data and analytics solutions. The IEEE P7000 standards projects are a response to the pressing need of integrating ethically aligned design framework in analytics solution exploration, development, implementation, and distribution. IEEE P7000 provides traceability of ethical values throughout the end-to end-lifecycle, from developing an operational concept and creating a value proposition, to embedding value dispositions in the system design.

IEEE Std 7010™-2020, IEEE Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems on Human Well-Being, was released in April 2020 and speaks directly to issues of critical importance to FSIs focusing on what metrics are used to measure wealth, growth, or indicators beyond single bottom line measures to determine what makes a good society.

There is understandable confusion around the term well-being, where people tend to think the word is synonymous with happiness. But as IEEE Std 7010 points out, the term is more focused on long-term flourishing, which involves a comprehensive view on what brings a person physical and mental health. As with the triple bottom line approach, for IEEE Std 7010, well-being is defined to include aspects of the environment. This literally means the ecological surroundings as well as access to education or ability to feel safe. IEEE Std 7010 also includes and focuses on treatment of how well-being is measured. And what you measure matters—the most common metric or Key Performance Indicator (KPI) for societal success is Gross Domestic Product (GDP), but this was not built to measure the environment or aspects of human mental and physical health. Incorporating human well-being and environmental metrics into the AI systems we build increases innovation for FSIs and companies at large as it provides new ways to define, measure, and substantiate value beyond exponential growth alone. Where customers expect banks and financial institutions that care about things like mental health and the environment, utilizing tools and methodologies like those outlined in IEEE Std 7010 helps provide FSIs a roadmap to determine holistic value in the algorithmic era.

Ethics Certification Program for Autonomous and Intelligent System (ECPAIS) is an IEEE-approved certification program developed to promote responsible innovation in the algorithmic age. There are three expert focus groups focused on ethical transparency, accountability, and algorithmic bias for ECPAIS. The ECPAIS creates specifications for certification and processes that generate transparency, provide accountability, and reduce algorithmic bias in Autonomous and Intelligent Systems (AIS).

The IEEE Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS)

Autonomous decision making by algorithmic learning machines in financial products, services, or systems poses uncertainties and societal concerns over their ethicality and trustworthiness with respect to fairness, explicability and rationality of the automated decisions. This brings about a formidable challenge to the uptake and innovation in the deployment of AIS based solutions.

ECPAIS certification criteria consist of a suite of detailed specifications for the evaluation, assessment and certification of ethical properties of AIS products and services.

The following facets of ethical responsibility are applied:

- Transparency criteria relate to values embedded in a system design, and the openness and disclosure of choices made for development and operation.
- Accountability criteria recognize that the system/service autonomy and learning capacities are the result of algorithms and computational processes designed by humans who remain responsible for their outcomes.
- Algorithmic bias criteria relate to systematic errors and repeatable undesirable behaviors that create unfair outcomes.
- Privacy criteria respect the private sphere of life and public identity of an individual, group, or community, upholding dignity. (Coming Soon)

In the absence of or as a complement to broadly accepted and enforced regulations for AIS, ECPAIS certification of products, systems and services enhance confidence in the public and private enterprises that wish to realize the benefits of AIS, while mitigating risks, liabilities and adverse impacts on their reputation and market share. Whatever an enterprise's role in the development and delivery of AIS products, services, or systems-- as a developer, system integrator, vendor, operator or maintainer—it stands to gain from independent ECPAIS evaluation, assessment and certification.

An organization's key stakeholders benefit from engaging with a committed organization with their needs in mind. Implementation against ECPAIS criteria demonstrates this commitment towards a greater trustworthy experience with AI systems.

To learn more about how ECPAIS may benefit your interests, complement the certification needs of your clients' products/services, or license the ECPAIS suite, please contact us at: ecpais@ieee.org.

Playbook Contributors

EXECUTIVE STEERING COMMITTEE

IEEE Finance Playbook: Trusted Data and AIS for Financial Services (Version 1.0)

Terry Hickey
SVP and CIO Corporate Centre Technology, CIBC

Sami Ahmed
SVP of Data and Analytics
OMERS

Dr. Konstantinos Karachalios
Managing Director, IEEE Standards Association Member of the IEEE Management Council

William Stewart
Head, Data Use and Product Management, Data and Analytics, RBC

Matt Fowler
Head of ML, Enterprise Data and Analytics,
TD Bank

Dr. Raja Chatila
EU Commission- High-Level Expert Group on AIChair, IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems

Mark Wagner
VP of Advanced Analytics and AI, Scotiabank

Mathieu Avon
VP, Integrated Risk Management, National Bank of Canada

Dr. Francesca Rossi
EU Commission, IEEE and PAI
IBM Fellow and AI Ethics Global Leader

Dr. Ren Zhang
Chief Data Scientist and Head of AI CoE, BMO

Vilmos Lorincz
Managing Director, Data and Digital Products, Commercial Bank,
Lloyds Banking Group

Dr. Yuri Levin
Dean, Moscow School of Management Skolkovo, Russia

Pavel Abdur-Rahman
Chair, IEEE Finance Playbook Head of Trusted Data and AI,
IBM Canada

John C. Havens
Executive Director, IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems

Dr. John Macintyre
Editor-in-Chief, AI and Ethics Journal, SpringerPro Vice Chancellor, University of Sunderland

Elizabeth Chacko
VP, Data and AI Risk,
Scotiabank

EDITORIAL TEAM

IEEE Finance Playbook: Trusted Data and AIS for Financial Services (Version 1.0)

Pavel Abdur-Rahman
Chair, IEEE Finance PlaybookHead of Trusted Data and AI,
IBM Canada

Stephanie Kelley
PhD Candidate in Management Analytics and AI Ethics, Queen's University

Cindy Pham
Senior Manager, AI Risk,
Scotiabank

Daniel Gomez Seidel
Senior Manager, Design Strategy,
Capital One US

John C. Havens
Executive Director, IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems

Alexander Scott
Business Development,
Borealis AI

Michelle Liu
Managing Consultant, Mastercard, President,
Analytics by Design

KEY CONTRIBUTORS

IEEE Finance Playbook: Trusted Data and AIS for Financial Services (Version 1.0)

Ozge Yeloglu
VP, Enterprise Advanced Analytics,
CIBC

Charbel Safadi
Senior Partner, Financial Services,
IBM

Andrew Brown
Senior Director, Data Science and AI Research,
CIBC

Noel Corriveau
Legal Counsel, INQ Data Law
Ex Senior Advisor, AI Policy and Implementation, Treasury Board of Canada

Omolade Saliu, PhD
Executive Data Scientist,
AI and Analytics,
IBM

Lucy Liu
Director, Data Science,
RBC

Aaron Zhang
Product Strategy and Analytics,
NEO Financial

Amy Shi-Nash, PhD
Global Head of Analytics and Data Science,
HSBC

Tim Gordon
Product Manager,
Enterprise Data and AI,
BMO

Carole Piovesan
Partner and Co-Founder,
INQ Data Law

Joy Hopper
Manager, Advanced Analytics and MLOps,
TD Bank

Teresa Papaleo
Director, Data Ethics and Use,
Scotiabank

Bharat Bhushan
CTO, Banking and Financial Markets,
IBM Europe

Dominique Payette
Lawyer, Legal Affairs,
National Bank of Canada

KEY CONTRIBUTORS

IEEE Finance Playbook: Trusted Data and AIS for Financial Services (Version 1.0)

Manav Gupta
Director and Distinguished Engineer,
IBM Americas

Teuta Bitici Mercado
Director, Responsible AI Program,
Capital One US

Paolo Sironi
Global Research Leader,
Financial Markets,
IBM Institute for Business Value

Max Howarth
Director of Artificial Intelligence,
Ideon Technologies

Wendi Zhou
Manager, Strategic Research
and Analytics,
Borden Ladner Gervais LLP (BLG)

Tania De Gasperis
Responsible IoT/XR Researcher for ACE Lab
at OCAD University, previously AI Ethics
Researcher for Montreal AI Ethics Institute

Andrew Morgan
Senior Manager, Risk and Insurtech,
Deloitte UK

Joseph Kim
Associate Partner, Data Platform,
IBM Canada

Dr. Ali Hessami
Chair and Editor of IEEE P7000,
IEEE Standards Association

Devan Leibowitz
Senior Management Consultant,
Monitor Deloitte US

Kaoru Kajimachi
Business Lead, Risk Management and AI,
National Bank of Canada

Dan Wigglesworth
Senior Consultant,
Technology and Analytics

Simon Thompson
Head of Data Science,
GFT Group UK

Ally Karmali
Associate Partner, Risk and Compliance,
IBM Canada

Sarah Hossain
Associate Director, Capital Markets,
RBC

KEY CONTRIBUTORS

IEEE Finance Playbook: Trusted Data and AIS for Financial Services (Version 1.0)

Jonathan Briggs
Chief Investment Officer at Delphia
(turn data into investment capital)

Nathan Good
Principal at Good Research
(proactive, holistic, and user-centric approach to Privacy and Security)

Monica Holboke
CEO/Co-Founder at CryptoNumerics
(privacy protection for the data-driven enterprise)

Ajay Jain
CEO at Rateco.ca
(turn data into better mortgage)

Ilana Golbin
Director, Emerging Tech and Responsible AI,
PwC US

Andrew Chau
Co-founder @ Neo Financial
(Canada's new Challenger Bank)

Brian Goehring
Global Research Leader,
AI and Cognitive,
IBM Institute for Business Value

Gigi Dawe
Director, Corporate Oversight and Governance,
CPA Canada

Artur Kluz
Managing Partner at Kluz VenturesFounder,
Centre for Technology and Global Affairs at Oxford University

Matthew Carroll
CEO at Immuta
(legal and ethical use of data)

Olga Harris
Director, FS Advisory and Analytics Technology,
PwC US

David Lauer
Independent Director,
NEO Exchange
(Canada's Next Gen Stock Exchange)

Mohamad Sawwaf
CEO at Manjil
(Canada's 1st Islamic NeoBank)

Linda Briceno
Senior Data Privacy Manager,
Lloyds Banking Group, UK

Fouad Habib
Engagement Manager, Data and AI,
IBM Canada

Appendices

Appendix A: Submission Guidelines

Thank you for submitting your comments and feedback for the first version of *The IEEE Finance Playbook*. We appreciate your taking the time to read the document and provide your insights.

Please review Guidelines below as you prepare your responses. Once completed, submissions should be sent to: financeplaybook@ieee.org for consideration. Submissions sent to any other email address will not be considered.

We will be posting all submissions received in a public document available from the website of The IEEE Trusted Data and Artificial Intelligence Systems (AIS) Playbook for Finance Initiative no later than 1 June 2021.

Detailed Submission Guidelines:

- All submissions must be received by 15 April 2021.
- Submissions should include the name of the individual submitting, the organization they represent, and the page number(s) or Sections of ***The IEEE Finance Playbook, V1.0*** being referenced. When submitting potential Issues or Candidate Recommendations, background research or resources supporting comments should also be included.
- Please ensure comments provide actionable critique (e.g., "On page X, I recommend adding the following resource") versus opinion without clear recommendations, (e.g., "I didn't like the Issue on page X.")
- We will post messages exactly as they are received. Please make sure to list your affiliations exactly as you would like them to appear, with embedded hyperlinks.
- Please do not send attachments. If you would like to cite other works, please link to them with embedded hyperlinks only.
- We encourage brevity. Specifically, submissions less than 1-2 pages in length are desired. If you feel you would like to send more, please send all feedback in one document.
- Do not send language or content protected via Intellectual Property or Patent considerations of any kind. IEEE, The IEEE Global Initiative and any/all of its subsidiaries will not be liable for any messaging received in this regard.

How Comments will be Reviewed:

- A committee within The IEEE Trusted Data and Artificial Intelligence Systems (AIS) Playbook for Finance Initiative will review all comments.
- Based on the context and type of comments received, the committee cannot guarantee all feedback will be implemented in later versions of the Playbook. Therefore, we are posting all feedback received publicly as a thank you for your contributions and to ensure everyone can see the insights you have provided.

As a thank you for your time and efforts we will do the following:

- Issue a press release and/or blog post by 1 June 2021 that will link to the document listing all comments sent regarding ***The IEEE Finance Playbook, V1.0.***
- List contributors and their affiliations as they were sent in the "Our Appreciation" section of the next version(s) of ***The IEEE Finance Playbook***.

Warm regards and thank you,
Pavel Rahman

Chair, The IEEE Trusted Data and Artificial Intelligence Systems (AIS) Playbook for Finance Initiative

Appendix B: Glossary of Key Ethical Terms

	IEEE EAD Definitions	References	Proposed Definition for Financial Services
Privacy	• Privacy safeguards mean strict standards for the collection and use of personal information. • Compliance with the highest requirements for data privacy and law.	IEEE P7002™, IEEE Standards Project for Data Privacy Process Inspired by The Personal Data and Individual Agency Control Committee, and supported by IEEE Computer Society standards.ieee.org/project/7002.html. IEEE P7012™, IEEE Standards Project for Machine Readable Personal Privacy Terms Supported by IEEE Society on Social Implications of Technology standards.ieee.org/project/7012.html.	**Privacy** means protection of personal information. Personal information is information about an identifiable individual, whether alone or in combination with other data points.
Accountability	• Principle on accountability: "Accountability: AIS shall be created and operated to provide unambiguous rationale for all decisions made." • Closely linked to transparency. • It is about clarifying roles and responsibilities, culpability, and liability (accountability structures). • Oblige states, "As duty bearers to behave responsibly to seek to represent the greater public interest, and to be open to public scrutiny of the AIS policies." (200) • "Accountability: AIS should be adopted only if all those engaged in their design, development, procurement, deployment, operation maintain a clear and transparent lines of responsibility for their outcomes and are open to inquiries as appropriate." (221) • "The combination of governing model of accountability and an openness to meaningful audit will allow the maintenance of accountability." (240) • Some legal case examples. (242) • "Transparency is essential in determining accountability, but transparency serves purposes beyond accountability while accountability seeks to answer questions not addressed directly by transparency." (249)	*EAD* pp.11, 29, 31, 200, 221, 240, 242.	**Accountability** generally means the processes to provide unambiguous rationale for decisions made, including clear roles and responsibilities for implementing and enforcing those decision-making processes.

	IEEE EAD Definitions	References	Proposed Definition for Financial Services
Transparency	• "Transparency: The basis of a particular AIS decision should always be discoverable." (4) • "Transparency in the context of AIS also addresses the concepts of traceability, explainability, and interpretability." (27) • Lack of transparency increases the difficulty of ensuring accountability. • Lack of transparency increases the risk and magnitude of harm. • "Develop standards that describe measurable, testable levels of transparency so that systems can be objectively assessed and levels of compliance determined." (28) • Four ways to be transparent: traceability; verifiability, honest design, and intelligibility. • Need to have sufficient transparency to allow evaluation by third parties. (189) • Transparency: Measurement methods and results must be open to scrutiny by experts and the general public. • No trust without transparency. • IP cannot be used unduly as a shield to prevent transparency. (251) Protect what you must, disclose what you can. • Does not necessarily mean disclosing all of code. There are alternatives. (277)	*EAD* pp. 4, 27, 28, 189, 251, 277. IEEE P7001™, Draft Standard for Transparency of Autonomous Systems is one such standard, developed in response to this recommendation. A testing framework for validating adherence to well-being metrics and ethical principles such as IEEE Std 7010™, IEEE Standard for Well-being Metric for Autonomous and Intelligent Systems.	**Transparency** is the principle of ensuring that AIS decisions are always discoverable. Transparency in the context of AIS also addresses the concepts of traceability, verifiability, honest design, and intelligibility.
Explainability	• The discussions of explainability in EAD are always in the context of transparency. No specific considerations are provided. • Explainability is often tied to better auditability.	Also see ICO ExplAIn Project in association with the Alan Turing Institute.	**Explainability** can be divided into two subcategories: • Process-based explanations that give information on the governance of AIS across its design and deployment, and • Outcome-based explanations that tell an end user what happened in the case of a particular decision.
Fairness and bias	Fairness (as well as bias) can be defined in more than one way. In the EAD a commitment is made not to provide one definition—and indeed, it may not be either desirable or feasible to arrive at a single definition that would be applied in all circumstances. (267) • To address the risk of bias, AIS must be underpinned by ethical and legal norms. These should be instantiated through values-based research and design methods. (124) • The evaluation process should integrate members of potentially disadvantaged groups in efforts to diagnose and correct bias. This includes the planning and evaluation of AIS. (188) • Transparency can help identify potential bias. • Bias can be introduced in a number of ways: via the features taken into consideration by the algorithm, via the nature and composition of the training data, via the design of the validation protocol, and so on.	*EAD* pp. 124, 188, 267. IEEE P7003™, IEEE Standards Project for Algorithmic Bias Considerations Supported by IEEE Computer Society standards.ieee.org/project/7003.html.	**Fairness and bias** can be defined in more than one way. The EAD states that it may not be desirable nor feasible to arrive at a single definition that would be applied in all circumstances. The concern with algorithmic bias has much to do with addressing and eliminating issues of negative bias in the creation of algorithms, where negative bias infers the usage of overly subjective or uniformed data sets or information known to be inconsistent with legislation concerning certain protected characteristics (such as race, gender, sexuality, etc.); or with instances of bias against groups not necessarily protected explicitly by legislation, but otherwise diminishing stakeholder or user well-being and for which there are good reasons to be considered inappropriate. In addition, fairness can be a substantive or a procedural dimension. • The substantive dimension is related to the concept of bias. • The procedural dimension is related to the concepts of explainability, transparency, and accountability.

	IEEE EAD Definitions	References	Proposed Definition for Financial Services
Autonomy	• Machines do not, in terms of classical autonomy, comprehend the moral or legal rules they follow. They move according to their programming, following rules that are designed by humans to be moral. (40) • Autonomy in machines, when critically defined, designates how machines act and operate independently in certain contexts through a consideration of implemented order generated by laws and rules. • When addressing the nature of autonomy in autonomous systems, it is recommended that the discussion first consider free will, civil liberty, and society from a Millian perspective in order to better grasp definitions of autonomy and to address general assumptions of anthropomorphism in AIS. (42) • A two-step process is recommended to maintain human autonomy in AIS. The creation of an ethics-by-design methodology is the first step to addressing human autonomy in AIS, where a critically applied ethical design of autonomous systems preemptively considers how and where autonomous systems may or may not dissolve human autonomy. The second step is the creation of a pointed and widely applied education curriculum that spans grade school through university, one based on a classical ethics foundation that focuses on providing choice and accountability toward digital being as a priority in information and knowledge societies. (60) • Ethically aligned design should support, not hinder, human autonomy or its expression. (102)	*EAD* pp. 40-45.	**Autonomy** in AIS, when critically defined, designates how machines act and operate independently in certain contexts through a consideration of implemented order generated by laws and rules. AIS do not, in terms of classical autonomy, comprehend the moral or legal rules they follow. They move according to their programming, following rules that are designed by humans to be moral.
Responsibility	• The responsibility for the behavior of algorithms remains with the designer, the user, and a set of well-designed guidelines. • Achieving a distributed responsibility for ethics requires that all people involved in product design are encouraged to notice and respond to ethical concerns. Organizations should consider how they can best encourage and facilitate deliberations among peers. • All those engaged in their design, development, procurement, deployment, operation, and validation of effectiveness maintain clear and transparent lines of responsibility for their outcomes and are open to inquiries as may be appropriate. • Apportioning responsibility. An essential component of informed trust in a technological system is confidence that it is possible, if the need arises, to apportion responsibility among the human agents engaged along the path of its creation and application: from design through to development, procurement, deployment, operation, and, finally, validation of effectiveness. (236) • The goal of clarifying lines of responsibility in the operation of AIS is to implement a governing model that specifies who is responsible for what, and who has recourse to which corrective actions, i.e., a trustworthy model that ensures that it will admit actionable answers should questions of accountability arise. (239)	*EAD* pp. 236, 239. "Human Responsibility for Autonomous Agents," IEEE Intelligent Systems 22, no. 2, pp. 60–61, 2007.	**Responsibility** for the actions of AIS remains with the designer, the user, and a set of well-designed guidelines. AIS can sometimes destroy data, compromise privacy, and consume resources, such as bandwidth or server capacity. What is more troubling is that automated systems embedded in vital systems can cause financial losses, destruction of property, and loss of life. All those engaged in AIS design, development, procurement, deployment, operation, and validation of effectiveness should maintain clear and transparent lines of responsibility for their outcomes and open to inquiries as may be appropriate. The goal of clarifying lines of responsibility in the operation of AIS is to implement a governing model that specifies who is responsible for what, and who has recourse to which corrective actions, i.e., a trustworthy model that ensures that it will admit actionable answers should questions of accountability arise.

Appendix C: Product and Customer Use Cases

Personalized Marketing Offers

Advanced Maturity

Description	Personalized marketing allows institutions to target customized promotional material to individual customers. This transition from bulk mailers to personalized material has been found to increase both customer uptake and reduce marketing fatigue. AI models are trained on customer marketing behavior gathered from past marketing interactions and used to adapt the marketing materials. The models allow for multiple versions of copy, graphics, and offers to be combined into any possible permutation. Customers then only receive the permutation most likely to change their behavior. Better customer targeting means the institution is able to identify which customers are most likely, least likely, and on the fence about purchasing a product. Using this information, the team can better allocate marketing budget; for example, by providing higher offers to customers on the fence, gentle and inexpensive reminders to customers with high propensity to buy, and allocating no spend for customers with little chance of take-up. Once the right permutation of materials and customer propensity are determined, AI can also be used to determine the most appropriate channel to distribute the materials.
Trusted Data and AI Implications	• Privacy and Data Governance • Diversity, Non-discrimination and Fairness • Transparency • Societal and Environmental Well-being
Prioritization, Execution, and Governance Best Practices	If your institution is not working on personalized marketing offers, it is a quick win with high value both in terms of increased revenue and reduced cost. The AI behind targeted marketing is widely available, broadly understood, and easily executed. From an execution and governance perspective, there are a few key considerations: • Customer exclusion lists (risk policy, credit, fraud) • Customer contact policies and procedures • Access to detailed customer data and historical marketing campaigns

Next Best Action

Advanced Maturity

Description	Next Best Action (NBA) models optimize every customer interaction. Institutions use customer data to train an AI system to predict the ideal next interaction with any given customer across promotional offers, marketing, and servicing, as well as the ideal channel through which to make that next interaction. NBA models combine a customer's current interests and needs with the marketing and sales needs of an organization. Typically, this is executed by creating a centralized AI-based decisioning hub that takes into account data on each customer's individual propensities to purchase and how they would like to be interacted with (information typically collected through customer surveys or prior interaction records). That is paired with possible treatments from an organization side including marketing and sales material, servicing interactions, proactive contact, or any other potential interaction.
Trusted Data and AI Implications	• Privacy and Data Governance • Diversity, Non-discrimination and Fairness • Transparency • Societal and Environmental Well-being
Prioritization, Execution, and Governance Best Practices	NBA models are becoming increasingly more important for large institutions as customer expectations change. While many large FIs are structured in a product-centric way, customers expect holistic service and single points of interaction for their entire relationships. For example, a customer who recently received an unexpected fee for their checking account is unlikely to want a sales offer the next day. Greater customer expectations pose an interesting challenge, especially as it relates to NBA. There are a few key considerations here: • Gaining consensus on 'best' across lines of business (is one line ready to give up short-term profit for long-term customer relationship?) • Incredibly large and detailed amounts of data from all interaction channels as well as products and transaction history • Integration with current interaction points (how does this information get to the phone channel, branch network, digital team, etc.?)

Loan and Deposit Pricing

Developing Maturity

Description	AI allows us to get more prescriptive about loan and deposit pricing for progressively smaller clusters of customers. The move from average pricing policies to risk-based policies, and now to true precision pricing optimization, has improved profitability, customer retention, and lowered risk for large institutions. AI allows us to sift through massive amounts of data related to customer behavior, elasticity, and product preferences to go beyond pure credit score pricing and optimize for customer profitability as opposed to pure risk management. Optimization algorithms are used to allow for a profit maximizing policy given a series of constraints such as increased market share, lower default, higher balances, etc.
Trusted Data and AI Implications	• Diversity, Non-discrimination and Fairness • Societal and Environmental Well-being • Transparency
Prioritization, Execution, and Governance Best Practices	Improved loan and deposit pricing should be treated as a high priority in retail banking and can be tackled iteratively. Each step made toward smaller customer groups and better AI-based optimization models will produce lower risk, higher profit books of business. A few key considerations when implementing new pricing policy: • How do you manage the current book of business? Pricing changes for new customers are easy, but retroactively changing is more challenging. • What is your appetite for risk? • Can your systems handle a more detailed pricing scheme?

Credit Adjudication

Advanced Maturity

Description	AI models let us make better inferences and predictions around customer risk. While credit scores are a useful tool for making credit risk decisions, they do not always represent an institution's holistic risk appetite nor a customer's true risk. Not only does AI let an organization make better adjudication decisions, it can also speed up adjudication time. Many vanilla credit applications can be adjudicated instantly by machine, freeing up human time for examining more complicated customer profiles with higher risk.
Trusted Data and AI Implications	• Diversity, Non-discrimination and Fairness • Societal and Environmental Well-being • Transparency
Prioritization, Execution, and Governance Best Practices	Institutions are able to make great credit decisions with the tools currently available. As an organization increases its analytical maturity, taking on more complex adjudication models can help with reducing the overall risk of a portfolio and time to adjudicate while increasing profit. This is a medium priority use case, but will become more important as more large institutions improve their ability to identify and price risk. As consumers move to an always-online world, being able to make instant adjudication decisions can be a competitive advantage.

Customer Sentiment Tracking

Advanced Maturity

Description	Canada's large FIs have different names for tracking customer sentiment, but all culminate in a customer survey. While surveys are useful tools for data collection, the way that data is analyzed can create a competitive advantage. Historically, organizations have relied on basic statistics on Likert scales and human-processed verbatims to hit a reasonable sample size. Today's AI allows for natural language processing of customer verbatims, meaning an organization can understand the sentiment of a customer's comment, why a customer scored certain areas in certain ways, and specific complaints that went poorly. With the right technology, this could be piped into a dashboard allowing for near real-time calculation of Net Promoter Score with instant insight into what is driving a change.
Trusted Data and AI Implications	• Human Agency and Oversight • Privacy and Data Governance • Societal and Environmental Well-being • Accountability
Prioritization, Execution, and Governance Best Practices	Current sentiment tracking capabilities are sufficient for many needs, leaving AI-based customer sentiment tracking as a nice-to-have for many organizations. That being said, getting ahead on customer sentiment has a number of benefits, from increased customer loyalty to increased revenue and reduced risk management. Given the drive to use AI in more areas of the business and the increased focus in privacy laws, understanding how a customer feels about all interactions (including the AI-based ones) will allow an organization to stay on the frontier of AI development.

Customer Lifetime Value

Developing Maturity

Description	Customer lifetime value (CLV) is an estimate of the value that a customer will bring to a business over the entire length of their relationship with the organization. AI, and specifically ML is, at its core, a predictive tool which can be applied to improve the quality of CLV predictions. Many organizations today rely exclusively on monetary sales to predict CLV, but ML can be trained on the same transaction data and is able to recognize more complex patterns in purchase recency and purchase frequency, in addition to the monetary value of those purchases. More data and better prediction tools generate a more accurate prediction of customer CLV.
Trusted Data and AI Implications	• Privacy and Data Governance • Diversity, Non-discrimination and Fairness • Societal and Environmental Well-being
Prioritization, Execution, and Governance Best Practices	Using AI to improve the quality of CLV predictions allows a financial services institution to more efficiently manage its existing customer base, whether this be increasing investment to retain profitable customers or generating new sales strategies to develop those customers. Using ML on customer transaction data has significant privacy implications though; organizations must ensure they have informed consent from customers for use of their data. But the ethical considerations go beyond adhering to privacy laws; ML can be used to unearth details and potentially sensitive information about customer behaviors, so an organization using AI for CLV must have in place ethical guidelines as to what information is acceptable to use in CLV predictions and what is not, which is something privacy law cannot always tell you. Bias and fairness have been, and continue to be an important consideration of CLV predictions; fair and equal targeting must play a role in a firm's marketing strategy, regardless of whether CLV is predicted by an AI-based application or not.

Customer Segmentation

Developing Maturity

Description	Pre-AI, customer segmentation was quite generic and focused on categorizing customers based on demographic attributes such as age, geography, and marital status. At a foundational level, AI-based clustering models can be introduced to develop better segmentations; the unsupervised learning models can help you uncover unknown patterns in your customer data to uncover new segments of customers. Once this foundation is built, clustering models can then be used to create micro-segments: highly specified groups of customers based on not only their product affinity, but also their preferred marketing messages, channels, frequencies. This micro-segmentation increases customer engagement, which creates significant efficiencies in marketing spend. More advanced applications of AI in this space now include real-time segmentation, where an AI can learn the changing preferences of a customer given each interaction with the bank.
Trusted Data and AI Implications	• Privacy and Data Governance • Diversity, Non-discrimination and Fairness
Prioritization, Execution, and Governance Best Practices	Just over half of North American banks are already using some form of AI for customer segmentation in their marketing department. The technology it is a time-tested way to improve targeting, personalization, and engagement of your existing marketing initiatives. Like customer lifetime value, the biggest barrier to adoption for AI-based customer segmentation is data availability. In order for a clustering model to generate a customer segmentation recommendation, it must be trained on detailed customer marketing data. Gathering this data is easier said than done, especially when we acknowledge the legacy systems present in many financial institutions today. The best practice involves creating a data lake, where customer information from across the bank can be aggregated and stored to develop a robust picture of each customer. This data lake can, of course, be used by many other AI applications across the bank, not just for customer segmentation. The more data available, the better the AI decisions will be; so ensuring there is a robust feedback-loop to collect data as the AI-based segmentations are implemented will further improve the AI learning process to provide even more refined suggestions over time.

High-Frequency Trading/Robo-Advisors

Basic Maturity

Description	AI has changed the dynamic for the securities arms of FIs. AI has actually been applied to trading for upward of ten years now, in what is commonly referred to as high-frequency and algorithmic trading. High-frequency trading is executed by fully autonomous AI systems (without human intervention); all data from every trade is captured and fed back into the ML system which uses it to analyze, recommend, and execute future trades almost instantly. This application of AI has completely changed market dynamics and effectively developed a new type of trading. Similar trading algorithms are now being applied to less frequency trading executions, at both the institutional and personal portfolio management, a service now often referred to as robo-advising. In place of human-generated trading strategies, ML is used to provide personalized investment recommendations. At its most basic level, an individual can preselect a type of trading strategy backed by an AI-based trading algorithm, which automatically executes ongoing trades given a set of investment guidelines. With greater access to customer risk preference data, an AI system can be trained to provide hyper-personalized investment recommendations.
Trusted Data and AI Implications	• Human Agency and Oversight • Societal and Environmental Well-being
Prioritization, Execution, and Governance Best Practices	In addition to reducing the cost of portfolio management for financial services organizations, AI-based portfolio management services also offer consumers greater accessibility to financial advice compared to traditional human-based portfolio management services as the technology reduces portfolio management fees. There is a wide range of autonomy in robo-advising services: completely autonomous systems at the institutional level used for high-frequency trades, semi-autonomous systems, and low-autonomy systems used as tools by human investment managers. The level of autonomy selected for a given application must be a conscious choice for an organization; one that must take into consideration the promotion of human values, and implications given potential workforce displacement. A new fintech starting from the ground up may decide to implement a fully autonomous robo-advising system to offer a competitively priced option for an under-served market, while a semi-autonomous application may be a better fit for an established investing group with several hundred investment managers. There are appropriate settings for all levels of autonomy, but a responsible FI must ensure a people-first approach when designing the application.

Risk Use Cases

Cybersecurity

Developing Maturity

Description	As cyber criminals become more sophisticated, FIs will rely heavily on AI tools to detect and prevent network intrusions and protect customer data. AI is absolutely critical here for sifting through incredibly large amounts of data, reacting quickly, and being very accurate. Today's AI models can sift through server logs and identify anomalous patterns or tie those patterns back to zero-day events posted on the internet. Once identified, suspicious events can be blocked and a human alerted. These models work not only for external attacks, but internal challenges as well.
Trusted Data and AI Implications	• Technical Robustness and Safety • Privacy and Data Governance • Societal and Environmental Well-being
Prioritization, Execution, and Governance Best Practices	Improved cybersecurity is a high priority, always on project. More sophisticated attacks are being developed every day and it is the responsibility of an FI to protect both their own and their customer's data. FIs should consider elevating their cybersecurity team in the organization and linking closely with data science and AI teams. Governance should be focused on increased data access and heavy investment in the right talent.

Fraud Detection

Basic Maturity

Description	AI has evolved the world of fraud detection from a rules-based approach to highly accurate predictions in real time. Using historical data from customer transactions, AI algorithms can quickly find outliers and flag them for further review, or identify the probability of fraud for any given transaction. With increased customer data and accuracy, more prescriptive decisions can be made in real time including real time transaction decline, waiting period changes for e-transfers, and elevated risk processes.
Trusted Data and AI Implications	• Technical Robustness and Safety • Privacy and Data Governance • Safety and Security • Diversity, Non-discrimination and Fairness • Societal and Environmental Well-being
Prioritization, Execution, and Governance Best Practices	Fraud prevention continues to be a high priority both from a cost containment and customer protection point of view. Without an investment in fraud prevention, the FI will quickly become a target for more sophisticated fraudsters. Execution largely involves building in more sophisticated techniques to current fraud prevention techniques. Governance implications include: • Level of risk tolerance for losses vs. customer satisfaction impact of stricter controls • Customer segmentation for different fraud policies • Investment from senior executives for new technology/collaboration tools (cross-FI)

Anti-Money Laundering

Basic Maturity

Description	Anti-money laundering (AML) is particularly well-suited to AI applications. Large FIs need to sift through a massive number of transactions with the hope of finding something suspicious. Without AI, AML teams are largely based on transaction rules and topical news. AI models can automatically sift through all transactions and identify anomalous patterns, allowing human reviewers to look through a smaller number of alerts with a higher probability of AML concerns.
Trusted Data and AI Implications	• Technical Robustness and Safety • Societal and Environmental Well-being • Accountability
Prioritization, Execution, and Governance Best Practices	AML reporting is a regulatory requirement for all FIs and should remain a high priority from a risk mitigation perspective. Most FIs already have sophisticated governance processes around AML considerations—the key change with AI tools is training operators, releasing large quantities of data, and ensuring models stay up-to-date.

Model Validation and Bias Detection

Advanced Maturity

Description	As AI techniques get more sophisticated across all areas of an FI, so do the risks of algorithmic bias, unfairness, and opaque decision making. Fortunately, modern AI techniques are able to assess bias both in underlying data sets and in an AI itself. These tools typically need access to massive data in the clear (e.g., to assess for gender bias, the tool needs to see gender).
Trusted Data and AI Implications	• Technical Robustness and Safety • Transparency • Diversity, Non-discrimination and Fairness • Accountability
Prioritization, Execution, and Governance Best Practices	Implementing bias checking and explainability tools needs to be a high priority if an FI is adopting any of the other AI use cases. Customer sentiment requires explainable decisions and fairness, meaning there needs to be a checking mechanism for every AI implemented. Governance considerations include: • Which group owns and checks all models • When a model needs to be checked for bias • How to remain agile when model validation is required

Operations Use Cases

Robotic Process Automation

Developing Maturity

Description	AI and robotic process automation (RPA) techniques are allowing FIs to free up expensive human talent by automating repetitive tasks. Humans can then be redeployed on more complicated, specialized tasks of which machines are not yet capable. New AI applications not only speed up processes, but can actively learn and adapt to changing environments. For example, leading FIs are using RPA to help with processing insurance claims, automating back office asset management operations, KYC documentation processes, and much more. The AI robots behind this can complete simple processes faster and more accurately than humans.
Trusted Data and AI Implications	• Human Agency and Oversight • Technical Robustness and Safety
Prioritization, Execution, and Governance Best Practices	RPA presents an interesting cost take-out and efficiency play. It can be implemented in small pilots across the organization to start with relatively low cost. Investment in RPA today can have profound implications on an FI going forward. Execution and governance considerations should include which areas will benefit most, how to re-deploy resources that have extra time, and how to up-skill resources to work through the much more complicated cases that cannot be automated.

Operational Efficiencies

Basic Maturity

Description	Recent advances in both computer vision and natural language processing have allowed for vast improvement in automated document digitization capabilities. Computer vision, specifically optical character recognition, can be used to interpret printed text and documents. The content is translated into digital form which can then be stored and/or analyzed by other AI applications through the use of natural language processing. Today, many firms use third-party human-based digitization services but AI offers a faster and more cost-effective solution. Document digitization can be applied to virtually any part of a business, but some popular applications include the digitization of historical records, automated invoice processing, automated loan application entry. One particularly interesting application is the digitization of the mailroom; instead of using a team of indexers to sort and route mail, AI can be used in their place to read and route mail 24/7/365. This automation then allows mailroom personnel to execute other high-value tasks.
Trusted Data and AI Implications	• Human Agency and Oversight • Privacy and Data Governance
Prioritization, Execution, and Governance Best Practices	With a vast number of potential applications, document digitization can be applied to virtually any existing human-based documentation process within a financial services organization. As these potential applications involve a significant change to the level of human involvement in a given process, the adoption of this kind of AI must be accompanied by a clear people strategy, centered on upholding human values. A responsible financial services organization will ensure they have a clear vision of how the displaced employees will be used to accomplish other tasks prior to implementing any form of AI for operational efficiency. Privacy also becomes of greater importance when gathering and storing historical customer data; FIs must be sure that they are following the relevant privacy laws and pay particular attention to informed consent, ensuring consumers have explicitly consented for their data to be translated digitally, stored, and used in the proposed manner.

Expense Management

Basic Maturity

Description

Many vendors today offer end-to-end expense management solutions built on AI. From the initial receipt submission to the payout, several types of AI are applied to increase the accuracy, speed, and cost-effectiveness of the process. Employees can submit receipts to the AI-backed system, where computer vision and natural language processing is used to generate automatic expense reports. The report is then sent back to the user in real time for a quick validation before being sent through for approval. ML is then applied to root out issues including duplications, false claims, or policy exceptions given a company's individual expense policies. A human supervisor intervenes where necessary to provide direction on these problem cases, but the vast majority of cases are automatically approved and paid out the same day.

Twenty-four hours is the average expense turnaround promised by AI-based vendors, a significant decrease from the average two weeks it takes for a human-run expense management system. This significant reduction in expense turnaround not only increases employee satisfaction, but comes with an increase in expense accuracy; suspicious transactions are, in fact, more than twice as likely to be caught by an AI-based system than by a human.

Trusted Data and AI Implications

- Human Agency and Oversight
- Accountability
- Diversity, Non-discrimination and Fairness

Prioritization, Execution, and Governance Best Practices

You should be looking to a third-party vendor for an expense management system rather than to your internal team for development. Both existing expense management solution vendors as well as new AI-first vendors offer product options, but there is a vast range in the scale and degree of process integration across vendors, so a robust comparison based on your organization's size and needs is warranted.

The introduction of an AI-based expense management solution will help decrease the time it takes employees to submit expenses, which of course is a positive, but this type of system may reduce the need for certain accounting roles. A responsible organization will have a clear people strategy on how those displaced employees will be repositioned or retrained for other activities. Using AI from a third-party vendor also comes with responsibility and accountability considerations; you must ensure the vendor is upholding the same ethical standards you have in place related to AI, and you both must align on clear lines of accountability should any ethical or other issues arise from the use of the AI application.

Corporate Functions

Talent Acquisition

Leading Maturity

Description	AI-based recruitment techniques are being tested out by some financial services organizations to refine the talent acquisition process, particularly for positions with high turnover and application rates, like tellers and customer service representatives. Natural language processing is used to interpret applicant resumes and then an ML algorithm, trained on historical resumes, determines whether a resume should be approved for the next round. This kind of resume review application is typically still in the test and learn phase as it involves the automation of highly complex judgement-based decisions, which are often challenging for organizations to align and agree upon.
Trusted Data and AI Implications	• Human Agency and Oversight • Privacy and Data Governance • Diversity, Non-discrimination and Fairness • Transparency • Societal and Environmental Well-being
Prioritization, Execution, and Governance Best Practices	The modeling process itself is fairly straightforward when it comes to any AI-based resume readers, once an organization determines what attributes it is looking for in potential employees. The data collection and governance implications are what present significant potential challenges: resumes of both current employees as well as unsuccessful applicants need to be used to train the ML algorithm, and an organization must obtain informed consent from all parties. Explainability also becomes an important consideration; many applicants will desire to understand why they did not move forward in the hiring process, and the ML model used in the application must be able to provide the necessary level of explainability, which may require some sacrifice in accuracy. In a similar vein, applicants may demand to be evaluated by a human rather than an AI system, an option that is mandated to be made available to individuals in the EU, so FIs must consider how they will integrate humans into the decision-making process. A people-first strategy must consider this, along with the displacement of employees currently involved in the resume review process. Bias and fairness are also a major risk in this type of application due to the inclusion of many protected attributes including gender, race, and nationality in the resume training data. Amazon, an organization well-regarded as highly evolved in their AI journey, had significant gender bias issues with an AI-based acquisition tool it developed. So, although there is potential for significant cost savings, the complexities and potential ethical implications of AI-based talent acquisition tools make it an initiative likely best undertaken by an FI further along in its AI ethics journey.

Talent Retention

Leading Maturity

Description	The war for talent heavily impacts large FIs, and retaining top talent will continue to be a priority. New AI tools allow for a better understanding of the relationship between performance, compensation, sentiment, and retention. An AI model can pull information from employee surveys, performance reviews, compensation, and market demand to make informed decisions on retaining talent.
Trusted Data and AI Implications	• Privacy and Data Governance • Transparency • Diversity, Non-discrimination and Fairness • Societal and Environmental Well-being
Prioritization, Execution, and Governance Best Practices	Talent retention should be a higher priority than talent acquisition. Again, the AI tools in this space are fairly robust and easy to implement, though privacy and bias remain significant concerns. Regulatory concerns must be investigated as well, as Canada and the US have different regulations regarding what an organization can do with data.

Audit

Leading Maturity

Description	AI models allow for more accurate, faster, and more frequent auditing of critical processes. A major auditing constraint is the reliance on humans manually checking processes against process documents. AIs, with access to process documentation and the output of those processes, can perform always-on auditing on a much broader scale, freeing up human capacity for more complicated or critical processes.
Trusted Data and AI Implications	• Technical Robustness and Safety • Privacy and Data Governance
Prioritization, Execution, and Governance Best Practices	AI-based auditing can be implemented piece-meal and prioritized by critical areas (e.g., auditing sales conduct or AML functions). The complexity of AI-based auditing on a large scale means it is important to prioritize, build efficiency, then scale. Governance considerations include: • How to get access to the right data at the right time • How to fix and operationalize audit findings

Collections

Developing Maturity

Description	AI has changed the collections process from getting ahead of customer defaults to better solutions for customers in financial distress. AI-based tools using customer transaction, wealth, and FI interactions data are able to predict when a customer may be in need of financial support (allowing an FI to prevent a customer from ending up in collections), how to best assist the customer, and how to best get in touch with the customer. A combination of better risk assessment and understanding of a customer's journey can make collections more efficient and a significantly better customer experience.
Trusted Data and AI Implications	• Human Agency and Oversight • Privacy and Data Governance • Diversity, Non-discrimination and Fairness • Societal and Environmental Well-being
Prioritization, Execution, and Governance Best Practices	An investment in better collections processes presents options for a better customer experience and reduced cost.

Customer Service

Developing Maturity

Description	The ability to serve customers faster, more efficiently, and on their terms will continue to be a differentiator for FIs. While continued focus on physical networks is important, the move to digital services can reduce costs and increase customer satisfaction. Implementing AI can help focus an FI's efforts on what matters most to the customer. For example, the implementation of chat bots allows customers to access the information they need faster or to get deeper insights on their own behaviors. Predictive AI can help identify which web pages will drive call volumes instead of self-serve behaviors. Better understanding of a customer's behavior will allow an FI and an FI's front-line employees to better serve all customers.
Trusted Data and AI Implications	• Privacy and Data Governance • Diversity, Non-discrimination and Fairness • Societal and Environmental Well-being
Prioritization, Execution, and Governance Best Practices	Customer-facing applications of AI should be a high priority for FIs. Shifting customers to digital channels and optimizing points of contact will continue to be a differentiator. As well, AI tools in this space are readily available and easily implementable at relatively low cost. Governance considerations include: • Testing of customer-facing features prior to implementation is critical • Training and maintenance of the platforms • Executive sponsorships across all channels

Powered by IEEE SA

IEEE