The Institute of Internal Auditors

Financial Services AUDIT CENTER

Internal Audit's Role in Responsible Artificial Intelligence

Seth Rosensweig, Partner
Travis Ringger, Director
PwC

Seth Rosensweig
Partner
PwC

Seth D. Rosensweig is a Partner at PwC within the Digital Risk Solutions group where he leads the Digital Risk, Regulatory and Compliance practice. Seth has over 20 years of advising clients around designing and implementing analytics and advanced technology solutions related to the regulatory, risk, compliance, and internal audit agenda. His focus is on: Risk, Compliance and Regulatory technology change, validation, testing and remediation across: AML/ BSA, Fraud, Trade Compliance, DFAST, Basel, CCAR, BCBS239, Risk Data Management and Governance. He also focuses on helping transform their organizations with their use of digital assets.

Seth is a frequent speaker at industry risk and compliance events and is frequently quoted in industry trade publications. He speaks on topics related to Big Data, AI, Automation/Robotics, NLP and analytics as it pertains to managing risk and compliance change.

He has authored/co-authored or been quoted in the following articles / whitepapers: "How Securitizes Can Comply with Dodd-Frank Act"; "No Guts No Glory: Committing to Data-Driven Risk Management" (GARP); "Big Data is Not Just for Big Banks (WSJ Blog); "Data Driven: What Students Need To Succeed In a Rapidly Changing Business Environment"; "Turning The Corner, Advance the use of analytics within Internal Audit (WSJ Blog); "How Community Banks Can Compete in Mobile" (BankTech); "Financial Marketers Unprepared For Era Of Big Data" (Thefinancialbrand.com); "Latest Trends on IT Audit Analytics".

Travis Ringger
Director
PwC

Travis is a Director in PwC's Digital Risk Solutions practice. He specializes in designing and delivering advanced data solutions that provide better awareness and understanding of compliance and business risk, particularly solutions incorporating unstructured data, robotics process automation, natural language processing, and other forms of artificial intelligence. He has deep experience solving clients' data challenges and helping institutions respond to heightened scrutiny and pressure tied to compliance requirements and risk in the financial services, technology, healthcare, and consumer products industries.

Travis has been at the head of a number of initiatives and engagements, driving innovation and improving the business and compliance value that organizations obtain through data analysis, including testing automation, integration of structured and unstructured data, and enhancement of business operations and reporting in order to stay ahead of increasing regulatory demands and drive business value. Examples of this work include creation of solutions which take advantage of big data, analysis, scraping data from the web, image forensics, geospatial analytics, as well as deploying tools for data discovery including optical character recognition and natural language processing.

Travis has spoken at a number of industry events and demonstrated how companies can unlock the insights trapped within their unstructured data. Prior to his experience in Digital Risk Solutions, Travis worked in PwC's Systems & Process Assurance group. He holds a MBA from Northwestern's Kellogg School of Management.

AI has been getting progressively more intelligent, even beating us at our own games

May 11, 1997

DeepBlue beats Gary Kasaprov

February 16, 2011

Watson beats Jeopardy champions

March 15, 2016

AlphaGo beats Lee Sedol

Recent advancements by OpenAI have yielded surprisingly realistic AI-produced language...

Feb 15, 2019: 'Zero-shot" learning by GPT-2 of OpenAI

"Our model is not trained on any of the data specific to any of these tasks and is only evaluated on them as a final test; this is known as the "zero-shot" setting. GPT-2 outperforms models trained on domain-specific datasets (e.g. Wikipedia, news, books) when evaluated on those same datasets".

SYSTEM PROMPT (HUMAN-WRITTEN)	*In a shocking finding, scientist discovered a herd of unicorns living in a remote, previously unexplored valley, in the Andes Mountains. Even more surprising to the researchers was the fact that the unicorns spoke perfect English.*

Source: Better Language Models and their Implications, OpenAI Blog, February 14, 2019..

... However, OpenAI deems the model too good and too 'toxic' to be released in the public domain

```
Dr. Jorge Pérez, an evolutionary biologist from the University of La Paz, and
several companions, were exploring the Andes Mountains when they found a small
valley, with no other animals or humans. Pérez noticed that the valley had
what appeared to be a natural fountain, surrounded by two peaks of rock and
silver snow.
```

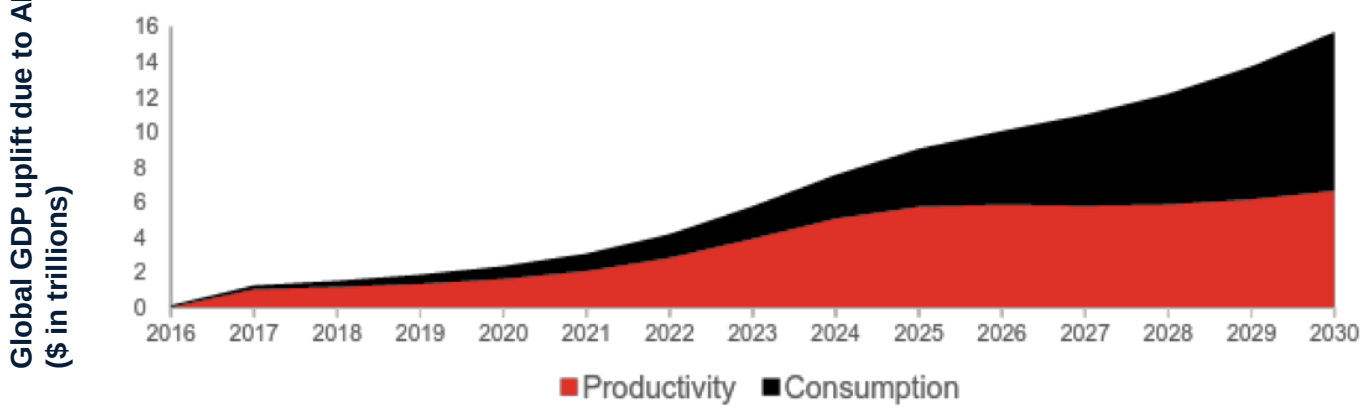
```
However, Pérez also pointed out that it is likely that the only way of knowing
for sure if unicorns are indeed the descendants of a lost alien race is
through DNA. "But they seem to be able to communicate in English quite well,
which I believe is a sign of evolution, or at least a change in social
organization," said the scientist.
```

Due to concerns about large language models being used to generate deceptive, biased, or abusive language at scale, we are only releasing a [much smaller version of GPT-2 along with sampling code.]

Source: [Better Language Models and their Implications], OpenAI Blog, February 14, 2019..

AI will contribute to substantial gains in productivity and consumption...

Global GDP Impact of AI through 2030



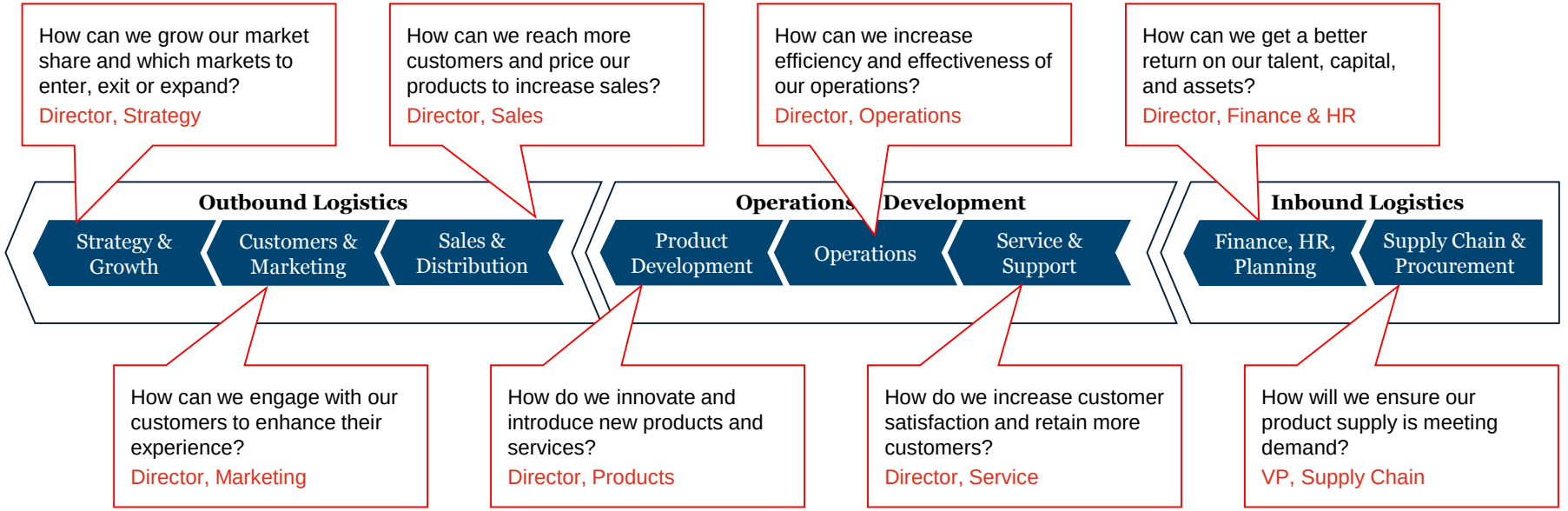
2030 IMPACT: $15.7T

Consumption Contribution: **60%**

Productivity Contribution: **40%**

Source: Sizing the Prize, PwC Report, 2017; A CPA's Introduction to AI: From Algorithms to Deep Learning, What you need to know, CPA of Canada, 2019.

… and is already unlocking opportunities all across the value chain



Risks of AI

The rise of AI brings with it inherent challenges around trust and accountability

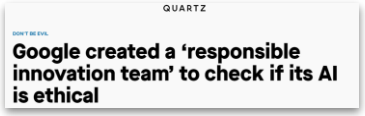
Bank of America confronts AI's 'Black Box' with fraud detection effort

- Banks are researching ways to explain AI algorithms which could have far-reaching impacts in guarding against potential ethical and regulatory breaches
- They need to understand how decision is made so that they can stand behind it

Source: The Wall Street Journal

Google creates 'responsible innovation team"

- Article discusses how Google has added a 'formal review structure' which consists of three groups to make big picture and technical decisions around the use of AI



Source: Quartz

Microsoft dropped some potential deals over AI ethical concerns

- The Company wants AI to be complementarity and not a replacement to human, highlighting the need for AI companies to ensure their approach is responsible and ethical

Source: AI News

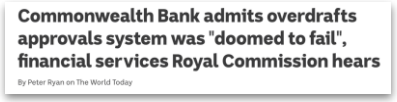
Amazon proposes ethical guidelines on facial recognition software use

- After public outcry and claims that Amazon's Rekognition software could potentially abuse civil and human rights, Amazon has proposed a set of ethical guidelines to policy makers.

Source: The Sociable

CBA's automated overdraft approvals system was "doomed to fail"

- A massive computer glitch inside Commonwealth Bank of Australia (Australia's largest retail bank) resulted in nearly 10,000 customers receiving overdraft facilities they were not able to service without causing serious financial hardship. CBA introduced an automated decision-making process for overdraft products in 2011.
- In 2015, the bank identified two separate but related computer glitches which meant that when a bank staff member entered a customer's rental and living expenses, the automated system read those entries as "zero dollars".



Source: ABC

Organizations face challenges as AI disrupts the entire value chain but few organizations have AI initiatives present on a wide-scale

85%

Of respondents think that AI will significantly change the way they do business in the next 5 years**

AND

84%

Of respondents think that AI based decisions need to be explainable in order to be trusted**

BUT

6%

Of organizations have AI initiatives present on a wide-scale**

83%

Of organizations think that the responsible use of AI is a priority for their organization*

WHILE

52%

Of respondents want to understand their AI systems and tackle the black box problem*

AND

48%

Of respondents want to safeguard their AI systems from external adversarial attacks*

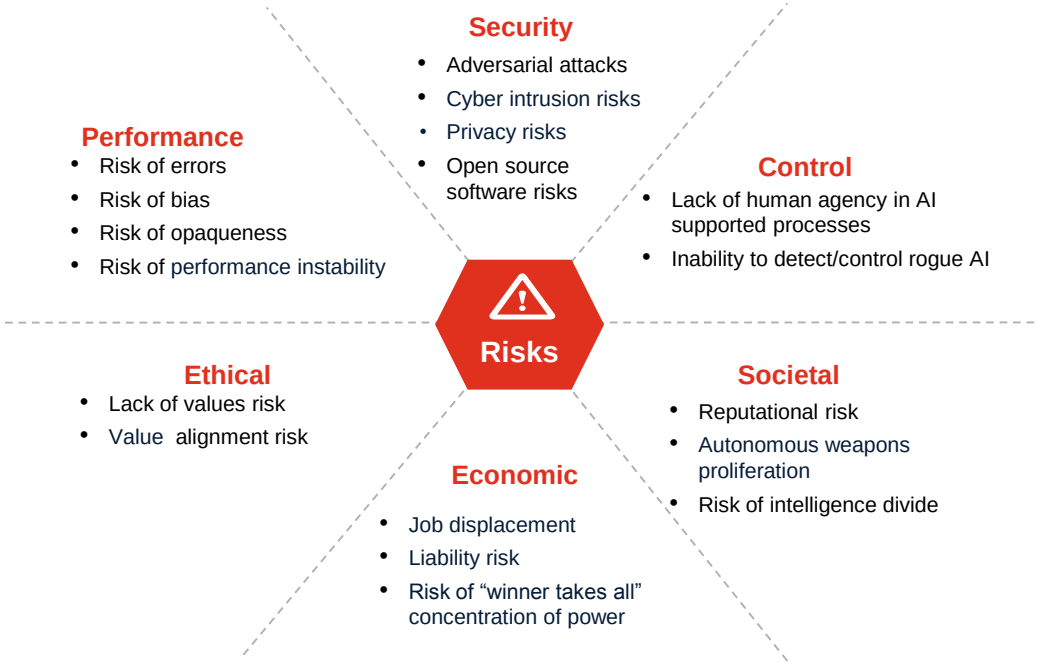
Source:* *SourceGlobal survey, February 2019.*
***PwC's 22nd CEO Survey, 2019.*

Governments are actively developing legal frameworks for holding algorithms accountable

Algorithmic Accountability Act (US)	*General Data Protection Regulation (EU)*	*California Consumer Privacy Act*	*Government Task Forces*
Companies are **legally required** to **assess automated systems** based on training data, model, fairness performance, bias, discrimination, privacy and security. Companies must conduct **impact evaluation** and rectify any identified issues.	Companies should use personal information secured in a manner that **prevents discriminatory effects**. Appropriate mathematical procedures should be adopted for consumer profiling. The data subject possess the **Right To Explanation** to seek meaningful information about the logic involved in modeling.	Modeled off GDPR in many ways, such as **limiting data usage and requiring right to erasure.** At this moment, CCPA does not contain provisions for the right to object to automated decision making.	Globally, governments are also launching task forces and other exploratory bodies on AI and AI governance. Bodies include: - UK All Parliamentary Group on AI (Jan 2017) - NYC AI Task Force (May 2018) - Victoria All-Party Parliamentary Group on AI (March 2018)

Globally, countries are exploring ethical AI standards. Recently, 42 countries signed on to the OECD Principles on Artificial Intelligence.

Organizations will need to understand key risks and answer fundamental questions around design and deployment



How can I improve security and robustness of AI through rigorous validation, continuous monitoring and maintenance, verification and adversarial modeling?

How do I test for bias in the data, model, and human use of AI algorithms to improve fairness of treatment across my organization?

What can I do to add transparency, explainability and provability to the modelling process to improve human understanding of the model outputs?

How do I assess the ethical and moral implications of the development and use of AI?

How can I track and check that AI solutions operate in compliance with relevant regulations?

How can I design effective AI operating models and processes to improve accountability and quality?

To accelerate innovation and fully realize the potential of AI, responsible and ethical considerations need to be prioritized

Ethical & Societal

ETHICS & REGULATION

Assess risks related to ethical aspects of AI and operationalize ethical AI for the respective context; identify and evaluate relevant regulations that impact AI solutions

Performance & Security

BIAS & FAIRNESS

Uncover bias in the underlying data and model development process and enable the business to understand what process may lead to unfairness

INTERPRETABILITY & EXPLAINABILITY

Explain model decision making overall and what drives an individual prediction to different stakeholders

ROBUSTNESS & SECURITY

Assess the performance of AI over time to identify potential disruptions, challenges to long term performance, or impact to data privacy

Control

GOVERNANCE

Introduce enterprise-wide and end-to-end accountability for AI applications and consistency of operations to minimize risk and maximize ROI

The Toolkit provides a set of practical resources to inform and guide the five dimensions that underpin Responsible AI

*Illustrative

Ethical & Societal

ETHICS & REGULATION

- Ethical AI Framework
- Ethical AI Contextualization Methodology
- Regulation framework by territory
- Ethical AI Principles Traceability Matrix

Performance & Security

BIAS & FAIRNESS

- Data bias identification & correction
- Model bias identification & correction
- Fair algorithms

INTERPRETABILITY & EXPLAINABILITY

- Interpretability framework assessment
- Local & Global interpretability toolkit
- Model exploration

ROBUSTNESS & SECURITY

- Model sensitivity analysis tool
- Adversarial analysis
- Ongoing monitoring
- API threat tool

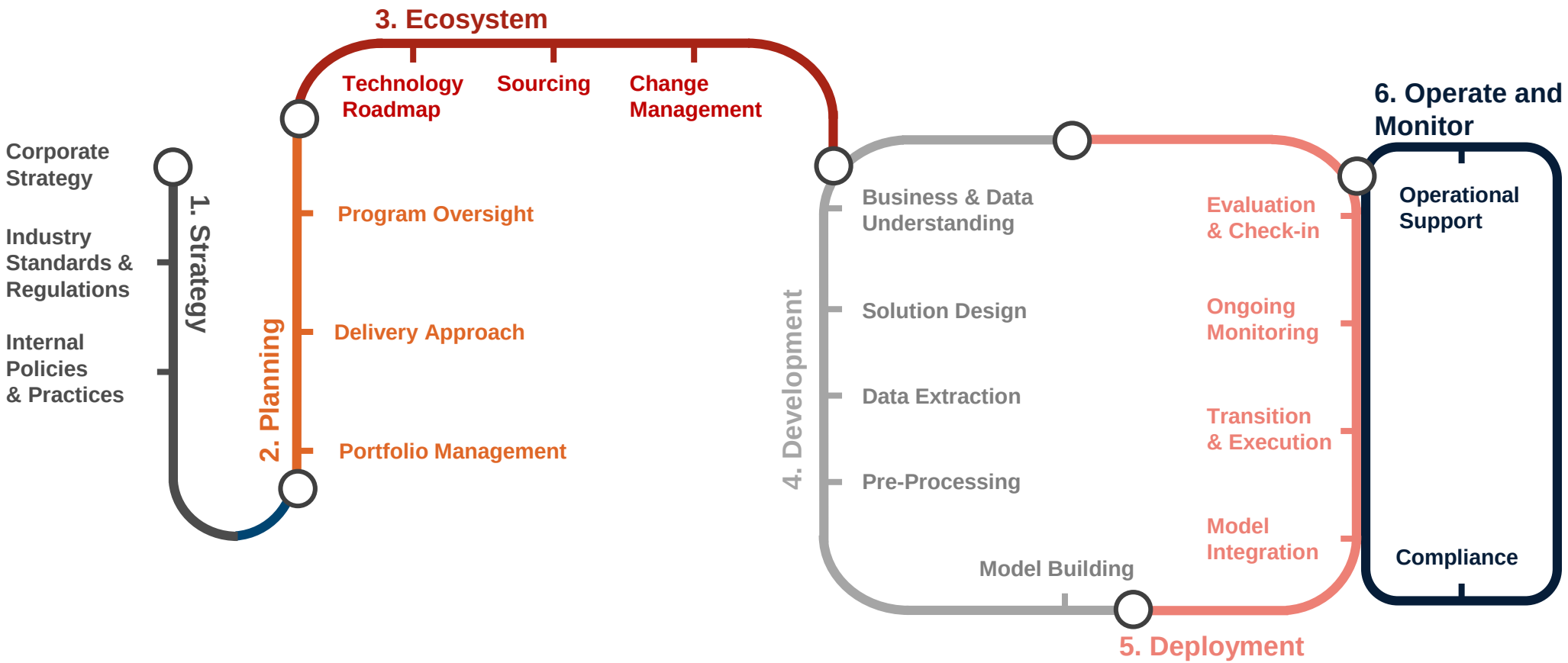
Control

GOVERNANCE

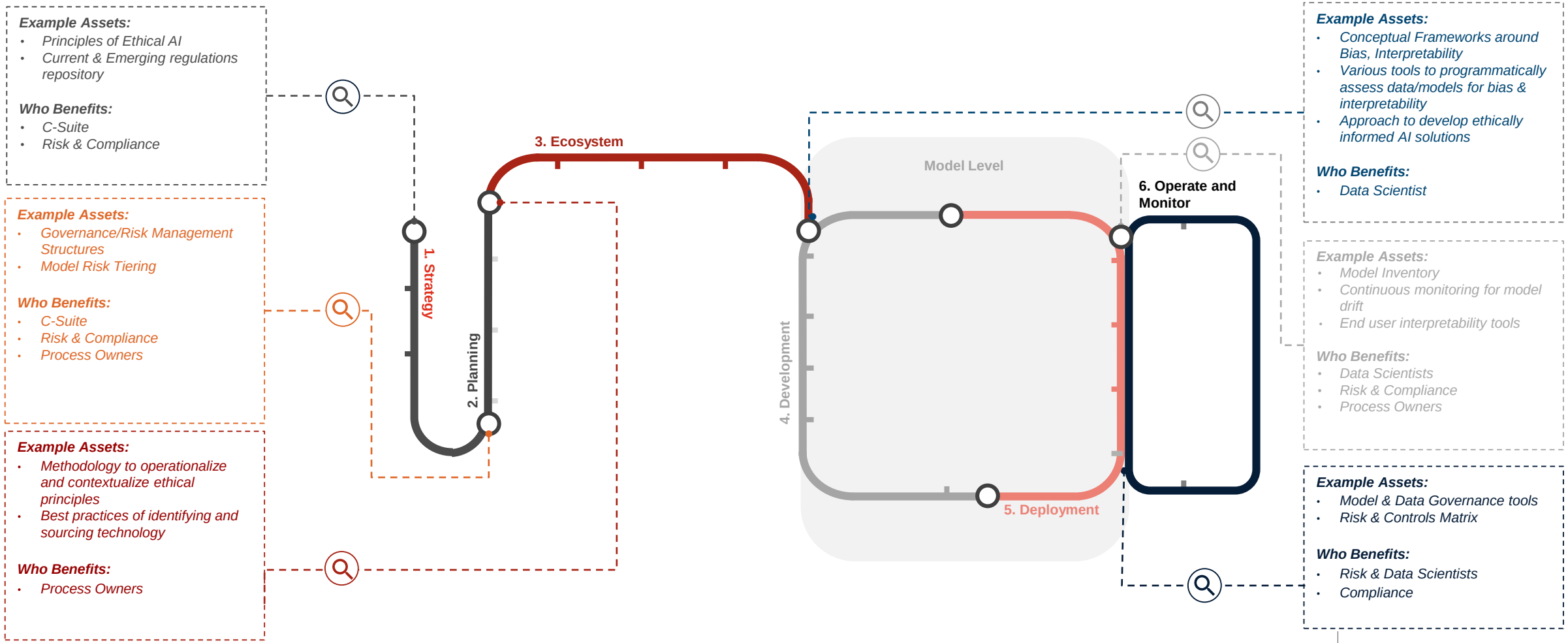
- End to end governance framework
- Risk & controls matrix
- Data sheets & model sheets; practice aids and best practices

The Institute of Internal Auditors | Financial Services AUDIT CENTER

An enterprise governance framework provides structure and best practices across the AI lifecycle (1 of 2)



An enterprise governance framework provides structure and best practices across the AI lifecycle (2 of 2)

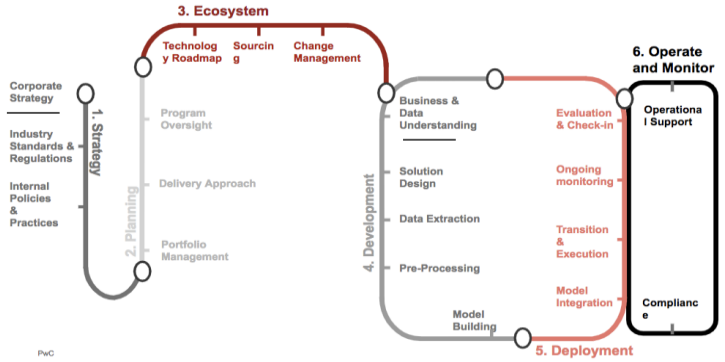


The Institute of Internal Auditors

Financial Services AUDIT CENTER

RAI's end-to-end framework allows organizations to implement AI, Automation and RPA programs from the top of the enterprise down to the individual initiative level.

Moreover, RAI helps organizations to evaluate and establish a risk appetite for their AI portfolio and to subsequently monitor risk exposure for compliance.



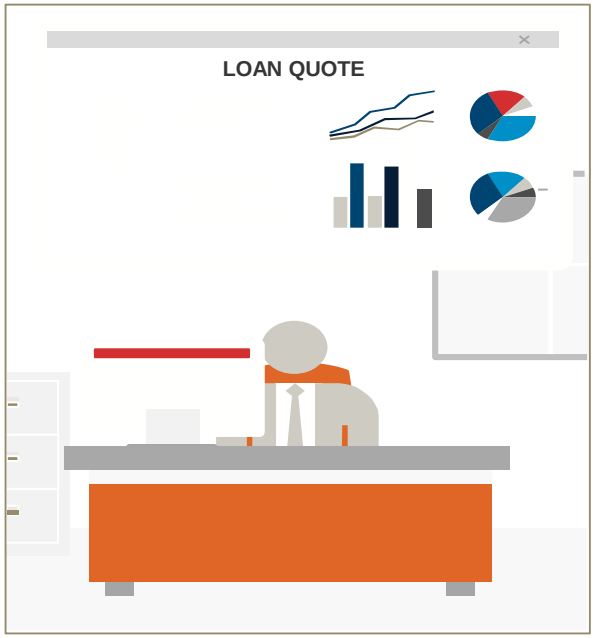
Corporate Strategy:

- Guides an organization through defining how AI/RPA can be used to drive value within an organization, and identify the highest value opportunities via strategic assessments and market research
- Once ambition is established the framework helps to establish investments and the required ROI (financial and non-financial)

Business & Data Understanding

- Establishes business criteria for success as well as the risk associated with the initiative to be compliant with the entity's overall risk appetite
- Additionally, tools within the RAI ecosystem help to provide transparency around data sources, lineage, and assumptions.

Omega Bank is about to launch a Digital Lending solution to existing customers for small, short term pre-approved loans



The New Opportunity

Customer loans are **pre-approved** using a **machine learning algorithm** that crunches customer transaction data, mobile data, open web data and third party data to determine for each customer a willingness-to-pay score, within minutes. Omega Bank wants to improve efficiency while differentiating for clients. **Omega Bank does not have an existing AI code of ethics.**

Identifying the Path Forward

Omega Bank's codes of conduct, strategy roadmaps, marketing campaigns, customer feedback and other external inputs such as competitor activities are compiled to understand the bank's vision and internal culture. Omega Bank wishes to become one of the more fair and trusted banks and therefore has a very **low risk tolerance** to Interpretability, Fairness, Data Privacy and Accountability ethical AI principles.

Complying with Regulation

Omega Bank's key stakeholders leverage the Regulation Framework to identify the laws they need to comply with for the impacted ethical principles. Omega Bank learns their operations will extend to California, thereby **requiring compliance** to CCPA. Given the ethical principle of Fairness, Omega Bank also must be **compliant with anti-discrimination and fair lending laws**.

Adjusting Business Practices

Omega Bank institutes **internal policy changes** to address the ethical principles and regulations. The rapid loan approval system undergoes bias testing so Omega Bank can assess compliance to the new internal policies. The algorithm is evaluated to have **passed three of five selected fairness measures.** The model is adapted, and then launched for use.

Key considerations for AI audits

The IIA's AI auditing framework defines the following components

Governance

- Accountability, oversight, assurance around AI capabilities and strategic alignment
- Data governance, privacy & security, ownership throughout the entire data life cycle

Data Infrastructure & Architecture

- Completeness, accuracy, reliability of critical data elements, especially in highly distributed data environments

Measuring Performance

- Automated full-time equivalents can capture how AI is freeing human workers from mundane tasks, but new metrics are needed to capture the achievement of objectives
- Human error and biases can work their way into models and manifest themselves in results, whether they are intentional or not

The Human Factor

- Some AI models lack transparency to describe how decisions are made.

Source: *Artificial Intelligence: Considerations for the Profession of Internal Auditing*

Prospective AI Audit Plan

We have proposed some possible ways to evaluate and assess AI risk

Governance

- Review policies and procedures and determine whether or not AI is consistent with organizational values
- Assess the regulatory environment for AI-specific requirements and measure the firm's compliance
- Assess competencies and resilience of the data infrastructure, including security and privacy

Measuring Performance

- Independently assess and report on the performance of AI against management's stated objectives
- Evaluate the processes used to train data and apply models to assess the likelihood or occurrence of bias
- Independently assess the intended versus actual outcomes of AI, as well as the actions management took based on actual outcomes

Data Infrastructure & Architecture

- Evaluate the quality of data inputs to AI
- Assess management's ability to monitor and resolve data issues

The Human Factor

- Compare the results of AI-driven results with manual results to assess reasonableness
- Enable management to be aware of the locations and implications of black box decisions