



# The Journey to Responsible and Ethical AI

Shobhit Shrotriya

Managing Director,  
Global Life Sciences R&D Operations

  
**accenture**

# TABLE OF CONTENTS

**01**

## **Encountering the AI Ethical Dilemma**

Understanding the common dilemmas and challenges AI poses today

**02**

## **The world of Artificial Intelligence Bias**

Understand AI bias

**03**

## **Deep dive into Responsible AI**

Understanding Responsible AI, Traversing the RAI framework

**04**

## **The Journey to RAI**

Human + Machine.....together!

# TABLE OF CONTENTS

**01**

## **Encountering the AI Ethical Dilemma**

Understanding the common dilemmas and challenges AI poses today

02

## The world of Artificial Intelligence Bias

Understand AI bias

03

## Deep dive into Responsible AI

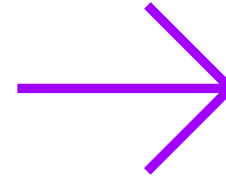
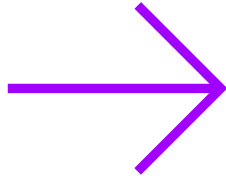
Understanding Responsible AI, Traversing the RAI framework

04

## The Journey to RAI

Human + Machine.....together!

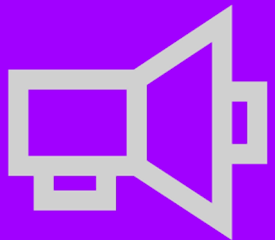
# IMAGINE YOUR TEAM HAS BEEN EXECUTING A BIG “AI” PROJECT...



Set specific goals around -  
**improve efficiency of current processes, reduce cost and enhance user experience**

Gathered a **great team** of  
Data Scientists & AI Experts  
**Executed** stakeholders'  
outreach and research

**Proposed a solution** driven  
from an algorithm created to  
achieve the goals while  
balancing tradeoffs



And you've just announced your recommendations publicly and your team is feeling good about making progress on this big challenge...

# AND THEN THE WAVE OF BACKLASH STARTS....

01

Data Science experts denouncing tools as opaque, unreliable, and inaccurate!

Legislature urging caution and skepticism over the use of the algorithm

Research papers discrediting the algorithm

...And you are struggling to defend your work and questioning your algorithm in the wake of this backlash

# WHAT WENT WRONG?

## WHILE UNLOCKING VALUE, AI AND ANALYTICS INTRODUCES NEW RISKS AND CHALLENGES



### Amazon's HR Hiring Tool

Existing gender imbalance for technical jobs was ingrained in historical data that Amazon's AI used to make hiring decisions. The outcome – successful male candidates and unsuccessful female ones. **After four years of significant investment, Amazon was forced to retire the application<sup>1</sup>.**

<sup>1</sup> <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scrap-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>



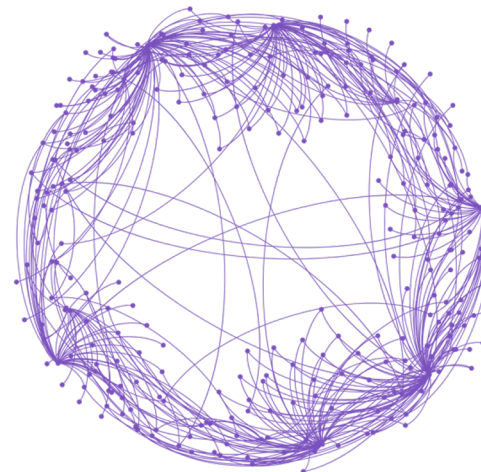
### Microsoft's Twitter Bot "Tay"

Shortly after its launch, Twitter users began tweeting Tay politically incorrect phrases. Microsoft did not train the chatbot to navigate this type of inappropriate behavior. Tay responded by sending its own inflammatory tweets to the public, and **after 16 hours of operation, Microsoft was forced to shut the bot down<sup>2</sup>.**

<sup>2</sup> <https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist>

## UNINTENDED CONSEQUENCES

Unexpected, but harmful, outcomes have led to consumer backlash and legal problems. Launching AI without an understanding of its social impact can be risky to your company's reputation and brand.



## COMPLIANCE, GOVERNANCE & SECURITY CHALLENGE

Deploying AI without anchoring to robust compliance and core values may expose a business to significant risks including employment/HR, data privacy, health and safety issues. The potential fines and sanctions can be business threatening.

# AS AI ACCELERATES, UNIQUE ETHICAL CONCERNS TAKE A CENTER STAGE

01



**Inclusion & Diversity** The power of AI is in the hands of the few and with the traditional power brokers.



**Artificial Stupidity** AI lacks general intelligence (empathy, common sense), which can lead to unfavorable outcomes.



**Privacy** AI will erode our notions of data privacy and do things with data that we didn't consent to.



**Job Apocalypse** AI is so ruthlessly efficient that it will lead to massive job loss.



**The Singularity** We will create something that is more intelligent than humans and we will lose control.



**Lack of Transparency** True (scary) AI doesn't explain itself.

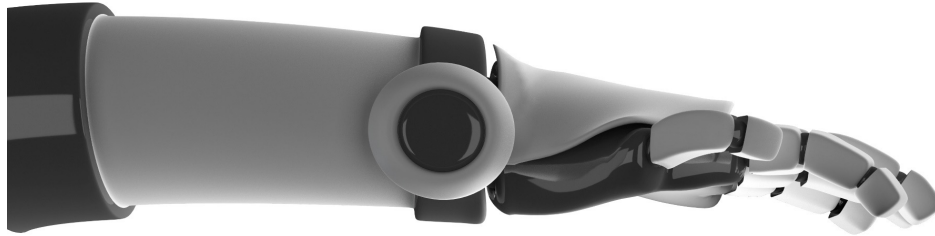
# WITH SUPERPOWERS COMES GREAT RESPONSIBILITY!

01

“

69% of business leaders believe their industry will be completely transformed by intelligent technologies [but only] 3% of executives plan to significantly increase investment in skills development programs in the next three years<sup>5</sup>.

Reworking the Revolution,  
Ellyn Shook & Mark Knickrehm



”

Tomorrow is now, and the AI race is on. Greater investment in AI could boost revenues by 38% by 2022, and boost employment by 10%. But if left unchecked, **AI BIAS** could have significant societal repercussions.

“

Data is not objective, it is reflective of pre-existing social and cultural biases<sup>4</sup>.

Rumman Chowdhury,  
Accenture Responsible AI Lead

”

“

30% of AI projects will deliver erroneous outcomes due to bias in data by 2022<sup>3</sup>.

Gartner, 2018

”

<sup>3</sup><https://www.gartner.com/en/newsroom/press-releases/2018-02-13-gartner-says-nearly-half-of-cios-are-planning-to-deploy-artificial-intelligence>

<sup>4</sup><https://hbr.org/2018/10/auditing-algorithms-for-bias>

<sup>5</sup>[https://www.accenture.com/\\_acnmedia/PDF-69/Accenture-Reworking-the-Revolution-Jan-2018-POV.pdf](https://www.accenture.com/_acnmedia/PDF-69/Accenture-Reworking-the-Revolution-Jan-2018-POV.pdf)

# TABLE OF CONTENTS

01

## Encountering the AI Ethical Dilemma

Understanding the common dilemmas and challenges AI poses today

02

## The world of Artificial Intelligence Bias

Understand AI bias

03

## Deep dive into Responsible AI

Understanding Responsible AI, Traversing the RAI framework

04

## The Journey to RAI

Human + Machine.....together!

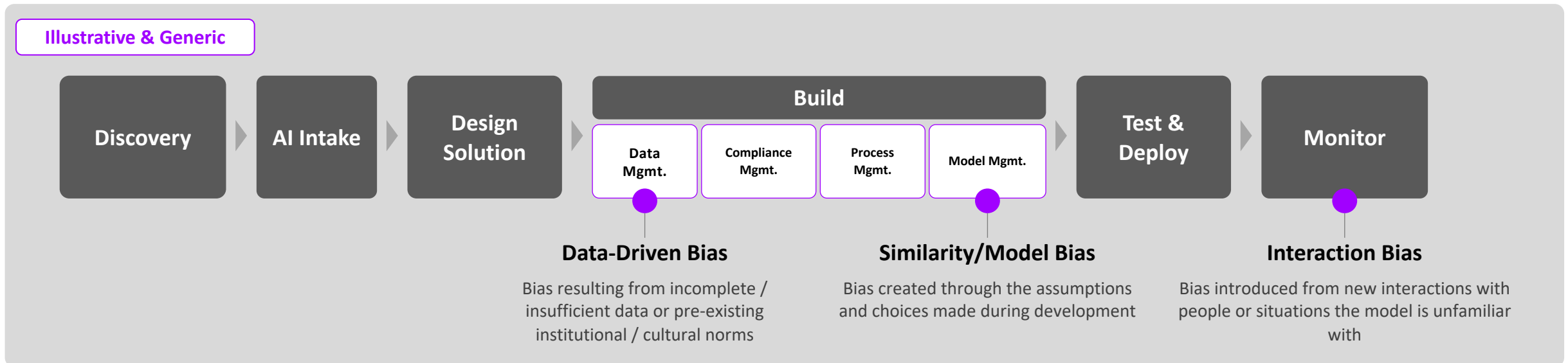
# UNDERSTANDING AI BIAS

- **Oxford Dictionary** defines bias as an “**inclination or prejudice for or against one person or group, especially in a way considered to be unfair**”.
- **Artificial Intelligence is susceptible to biases** that emerge from its interactions with humans and the data it is given.
- As AI learns from **biased people** and **biased data** its ability to learn, predict, and surprise are tainted with the biases it learns and new biases it propagates.
- When AI is either improperly implemented or improperly managed <sup>6</sup>**three bias types can emerge**:
  - Data-Driven,
    - *post hoc, ergo propter hoc*\*
  - Interaction, and
  - Similarity Bias

# HOW MIGHT **BIAS** CREEP IN?

Bias can be introduced into AI applications through **data**, **models**, or ongoing **operations**. All three components work in concert and can **reinforce each other** if bias is not addressed at the early design stage.

The following high-level process flow illustrates where **unintentional bias is most likely to arise** in the context of a typical AI development and implementation process.



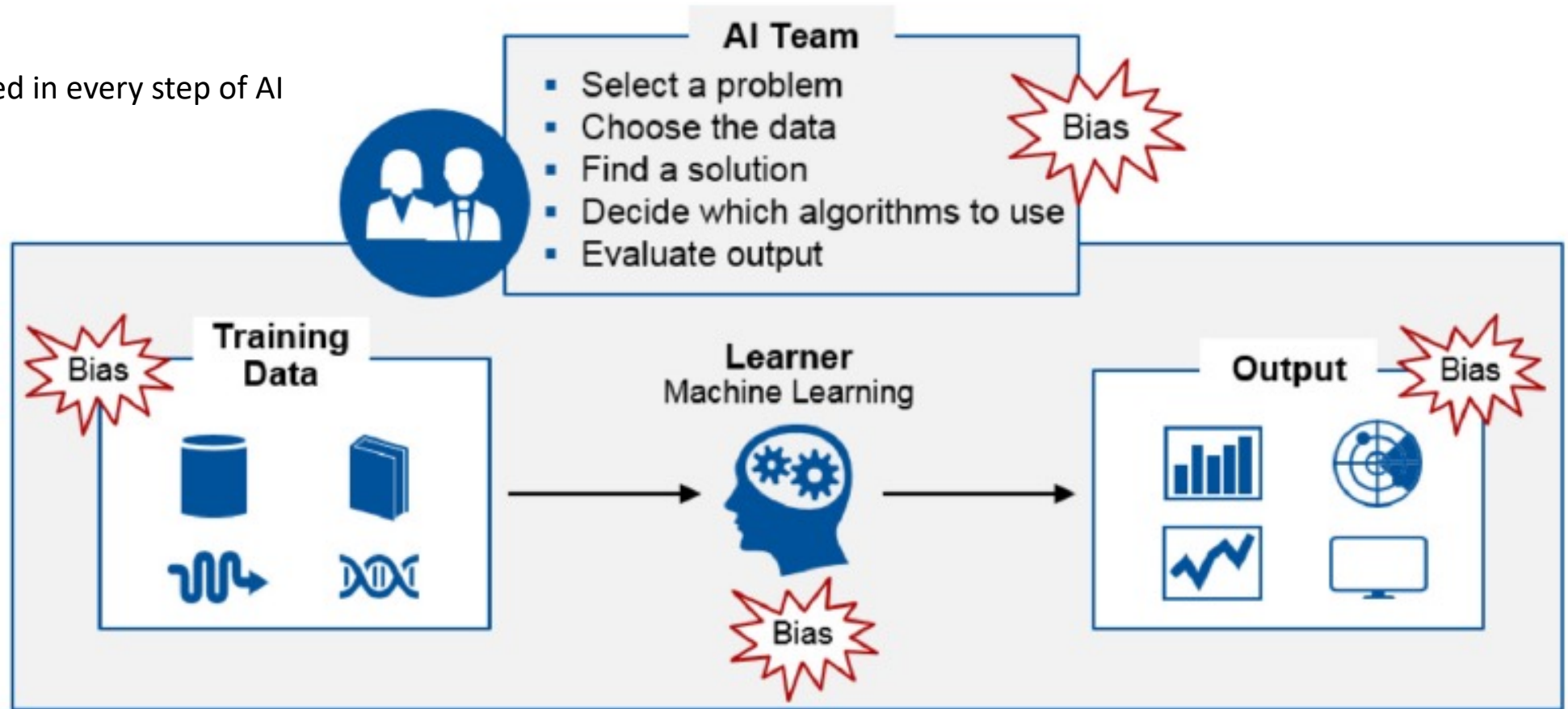
It is critical that we have proper visibility throughout the AI lifecycle, so that the appropriate controls can proactively be embedded.

# IN AI, DATA, ALGORITHMS AND THE PEOPLE WHO SELECT THEM CAN BE BIASED

02

Bias needs to be eliminated in every step of AI development <sup>6</sup>

Whether it is unconscious bias that creeps in from a human component or unplanned bias that occurs because of training data, it can proliferate through the implementation of Artificial Intelligence.



# TABLE OF CONTENTS

01

## Encountering the AI Ethical Dilemma

Understanding the common dilemmas and challenges AI poses today

02

## The world of Artificial Intelligence Bias

Understand AI bias

03

## Deep dive into Responsible AI

Understanding Responsible AI, Traversing the RAI framework

04

## The Journey to RAI

Human + Machine.....together!

# WHAT DO WE MEAN BY RESPONSIBLE AI?

Responsible AI is the practice of using AI with **good intention** to empower employees and businesses, and **fairly impact** customers and society – allowing companies to **engender trust** and scale AI with confidence<sup>8</sup>.

# DEFINE THE AMBITION AND CUSTOMIZE **PRINCIPLES OF TRUST**

Consequences of AI bias can be severe, and so it must be responsibly implemented to meet consumer expectations and remain relevant. Laws haven't kept up with technology, so companies must recognize that they need to do more than avoid illegal activities. Businesses should use a **human-centered approach** to achieve responsible AI, based on a TRUST framework *first developed by Accenture Responsible AI*:

**T****Transparent**

Honest use of tools & technology to detect bias

**R****Reliable**

Commitment to fix bias with governance to oversee the process

**U****Understandable**

Explainable, simple, and interpretable inputs and outputs; open the 'black box'

**S****Sustainable**

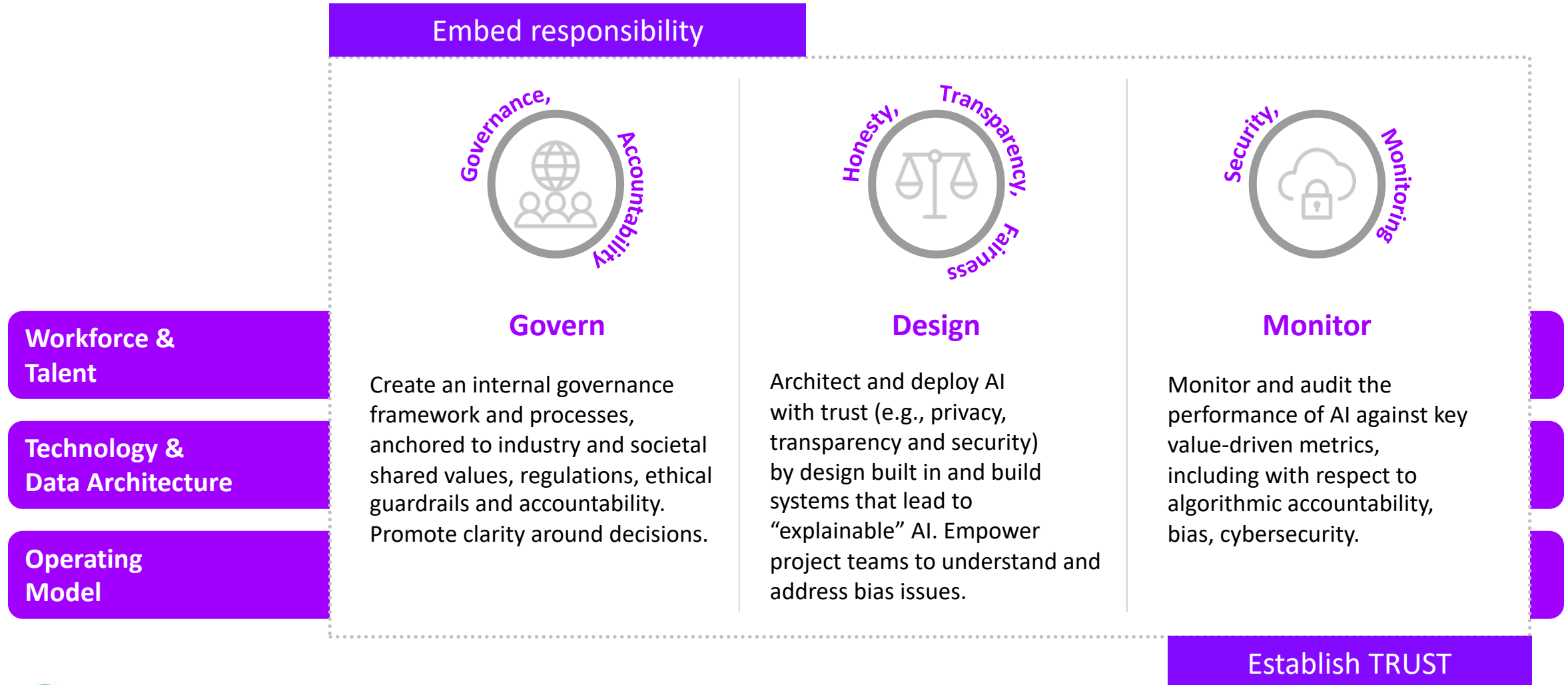
Consistent, scalable and manageable guardrails to secure the system

**T****Trainable**

Human + Machine to co-create, inform, advocate and complement each other

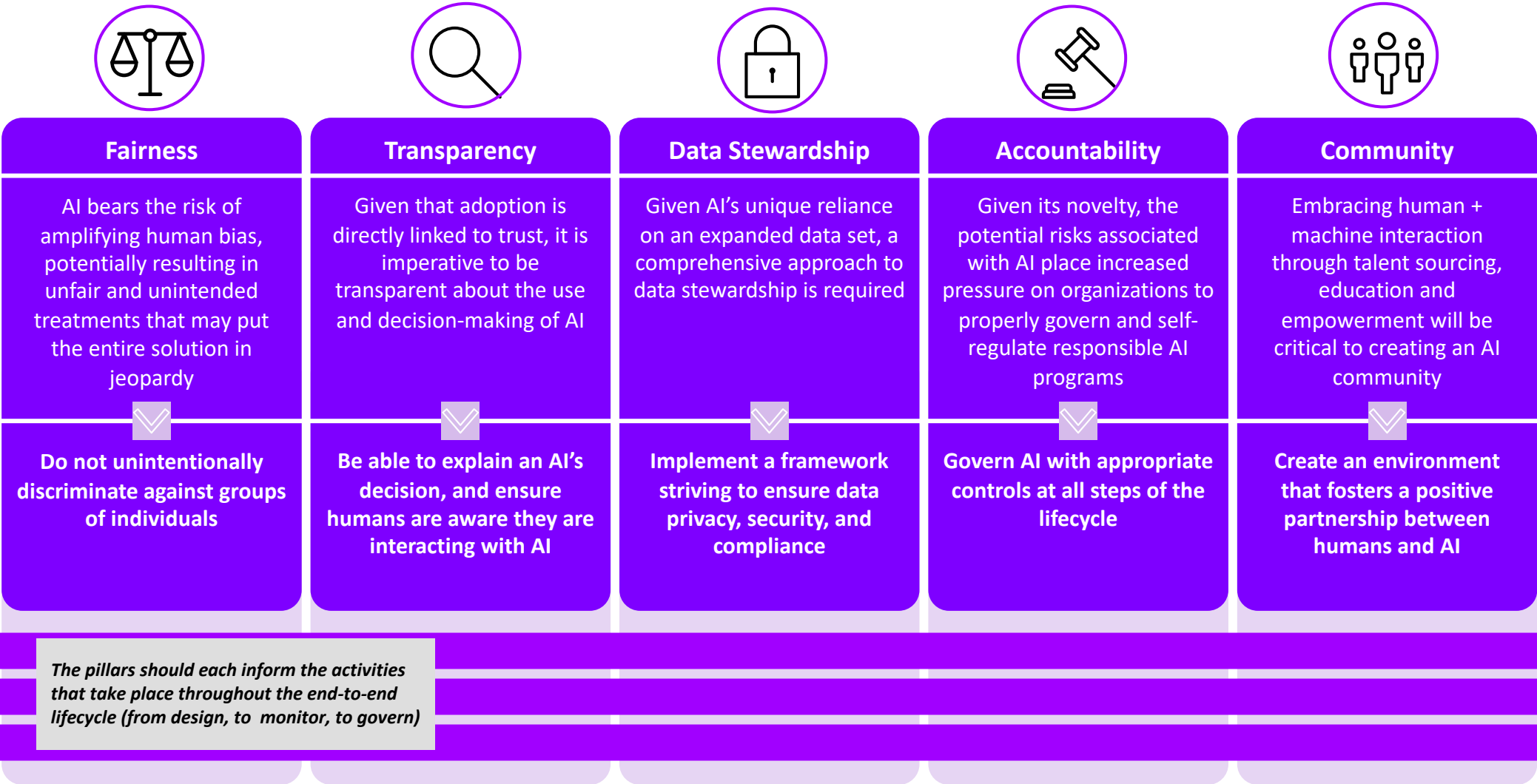
# RESPONSIBLE AI FRAMEWORK....

RESPONSIBLE AI DRIVERS to establish trust should be weaved in the AI architecture and operating capabilities<sup>8</sup>



# LEADING TO RESPONSIBLE AI PILLARS

FIVE RESPONSIBLE AI PILLARS<sup>8</sup>



# APPLYING 'FAREA<sup>6</sup>' METHODOLOGY

- Is the algorithm **FAIR**?
  - Ensure the artificial intelligence tool cannot be used to propagate virtual discrimination.
- Is the algorithm **AUDITABLE**?
  - Do you fully understand what the AI tool is meant to do and how it operates?
  - What data is it collecting?
- Who is **RESPONSIBLE** for the algorithm?
  - Who is responsible for digital decisions the algorithm makes?
  - Who is responsible for auditing the algorithm?
- Is the algorithm **EXPLAINABLE**?
  - Are digital decisions and any data necessary available and explainable to end users and stakeholders?
- Is the algorithm **ACCURATE**?
  - Have you tested the algorithm with test data?
  - Do you know how the tool will react to novel situations?

# TABLE OF CONTENTS

01

## Encountering the AI Ethical Dilemma

Understanding the common dilemmas and challenges AI poses today

02

## The world of Artificial Intelligence Bias

Understand AI bias

03

## Deep dive into Responsible AI

Understanding Responsible AI, Traversing the RAI framework

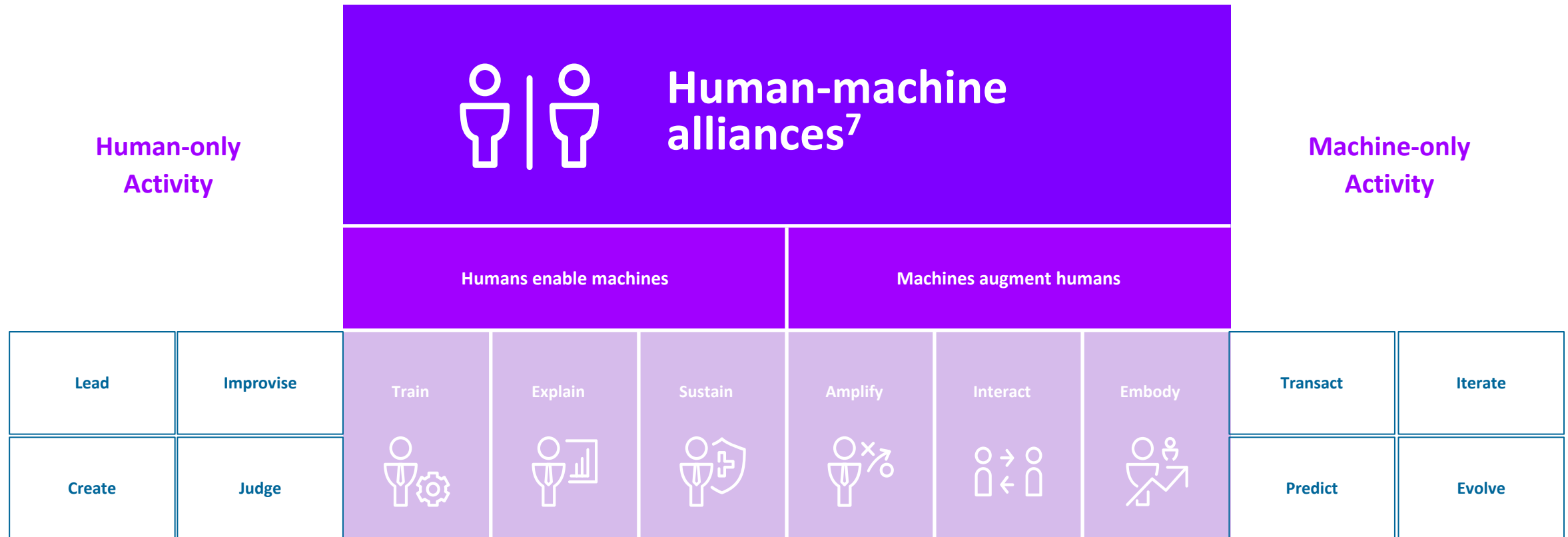
04

## The Journey to RAI

Human + Machine.....together!

# HUMAN + MACHINE DELIVERING TOGETHER

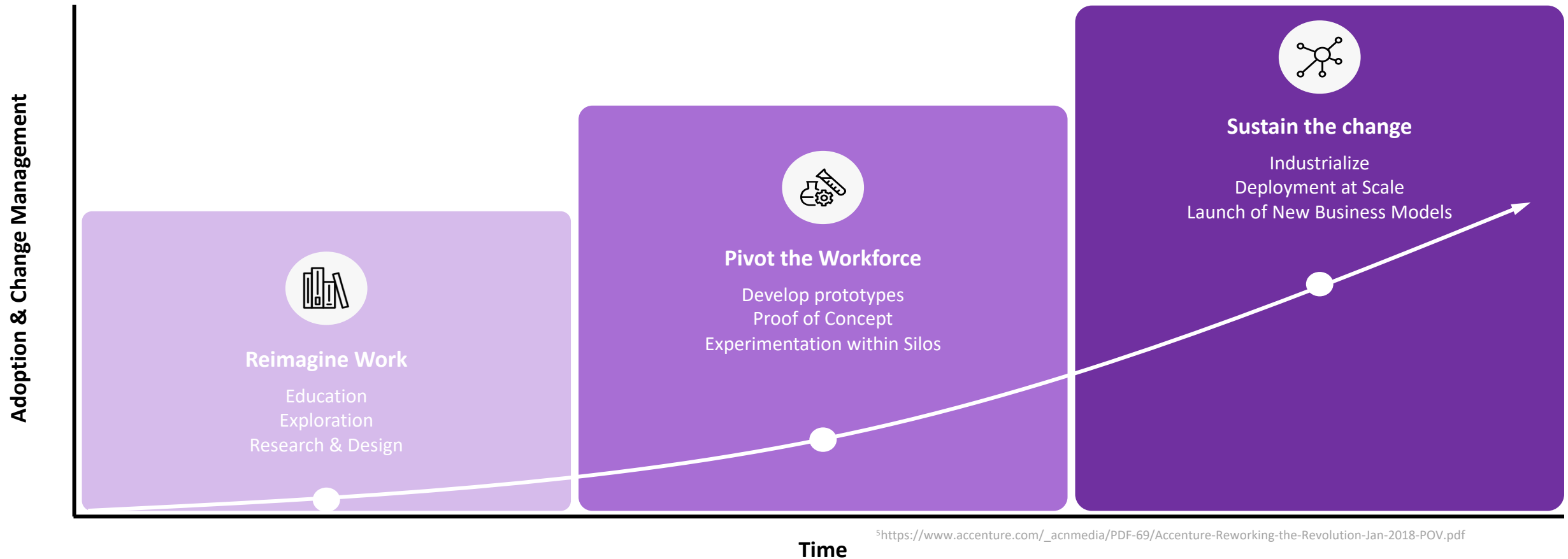
THE ADVENT OF AI INTRODUCES NEW INTERSECTIONS BETWEEN HUMANS AND MACHINES, RESHAPING THE ACTIVITIES THAT EACH HAS TRADITIONALLY BEEN KNOWN TO DO IN A WAY THAT UNLOCKS GREATER VALUE



Our workforce will need to prepare to work alongside AI in these new ways

# MANAGING THE 'AI' CHANGE

THE ACTIONS THAT WE MUST TAKE TO FULFILL OUR RESPONSIBLE AI MANDATE FALL INTO THREE HIGH-LEVEL CATEGORIES, EACH DRIVING A DIFFERENT TYPE OF VALUE FOR CLIENT AND THE ORGANIZATION<sup>5</sup>



The background features a close-up of a woman's face with a digital overlay of a network graph. The graph consists of numerous small, glowing nodes in blue, purple, and yellow, connected by thin white lines. The woman has dark hair and is looking slightly to the right. The overall color palette is dominated by deep blues and purples, with a bright cyan diagonal band on the left side.

**THANK YOU!**

[Shobhit.Shrotriya@accenture.com](mailto:Shobhit.Shrotriya@accenture.com)