MITSloan Management Review

WEBINAR

OCTOBER 2019

Sponsored by:

sas

ARTIFICIAL INTELLIGENCE

People-Centered Design Principles for AI Implementation

A presentation by R. "Ray" Wang, CEO of Constellation Research, and David A. Bray, executive director of the People-Centered Internet coalition

R “Ray” Wang is CEO of Constellation Research in Palo Alto, California.

David A. Bray, Ph.D., is the executive director of the People-Centered Internet coalition and senior fellow at the Florida Institute for Human & Machine Cognition.

Allyson MacDonald . is Senior Associate Editor, Digital, at *MIT Sloan Management Review*. She will moderate the session

People-Centered Design Principles For AI Implementation

How To Build Data-Driven, People-Centered Business Models

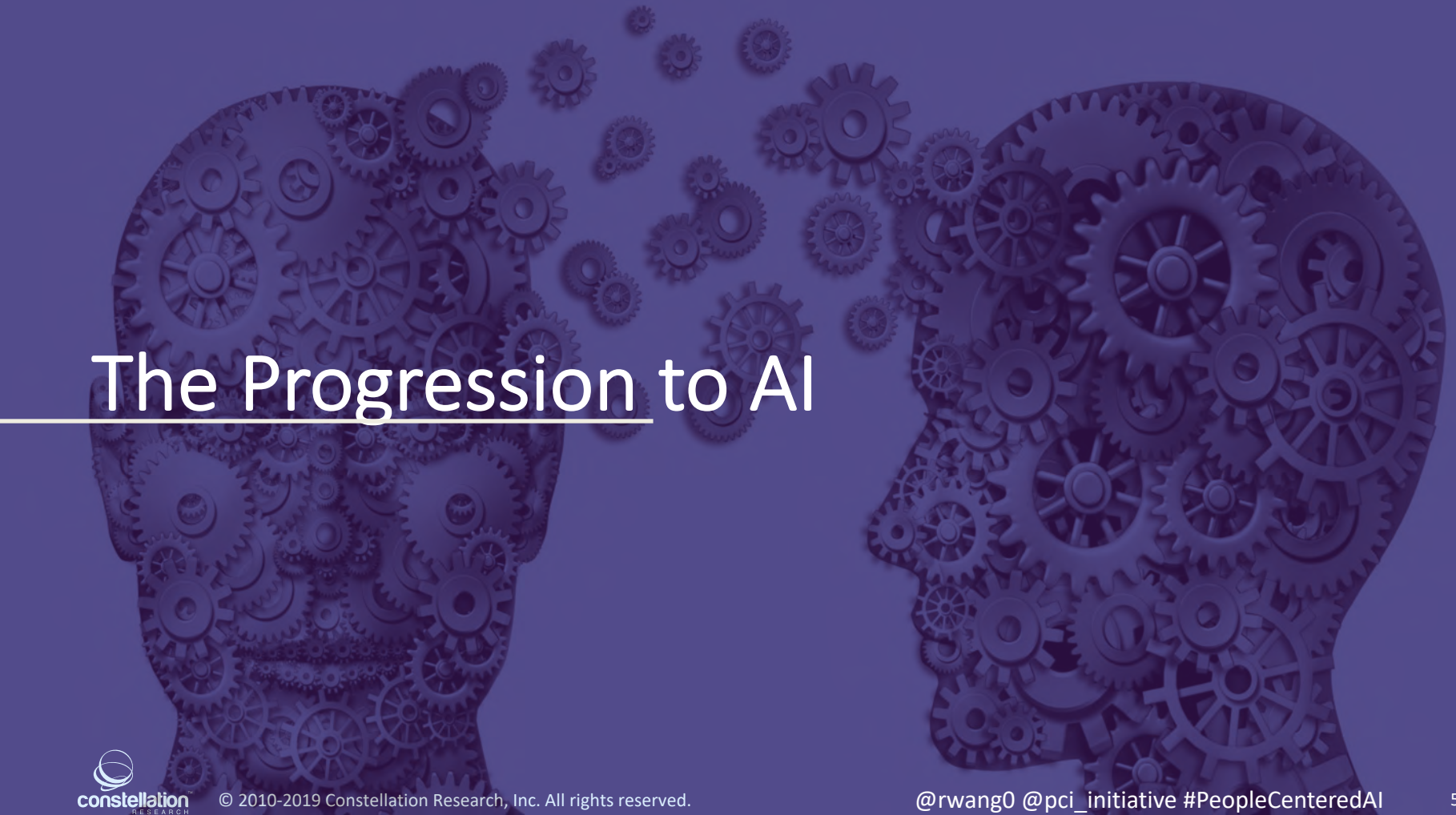
OCTOBER 2019

R "**RAY**" **WANG** (@RWANG0)
PRINCIPAL ANALYST
& CHAIRMAN, CONSTELLATION

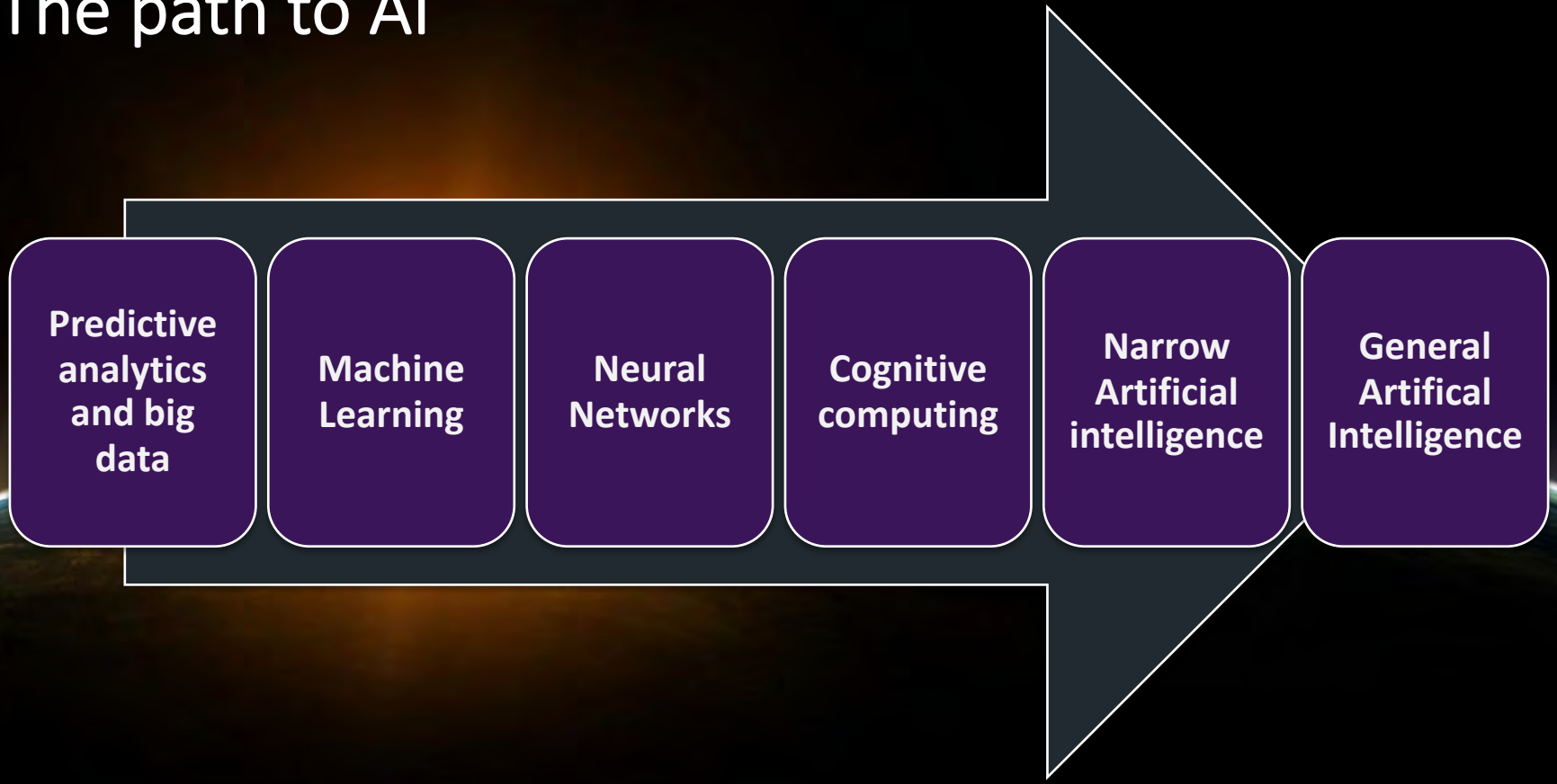
&

DR. DAVID BRAY (@PCI_INITIATIVE)
EXECUTIVE DIRECTOR, PCI
& SENIOR FELLOW, IHMC

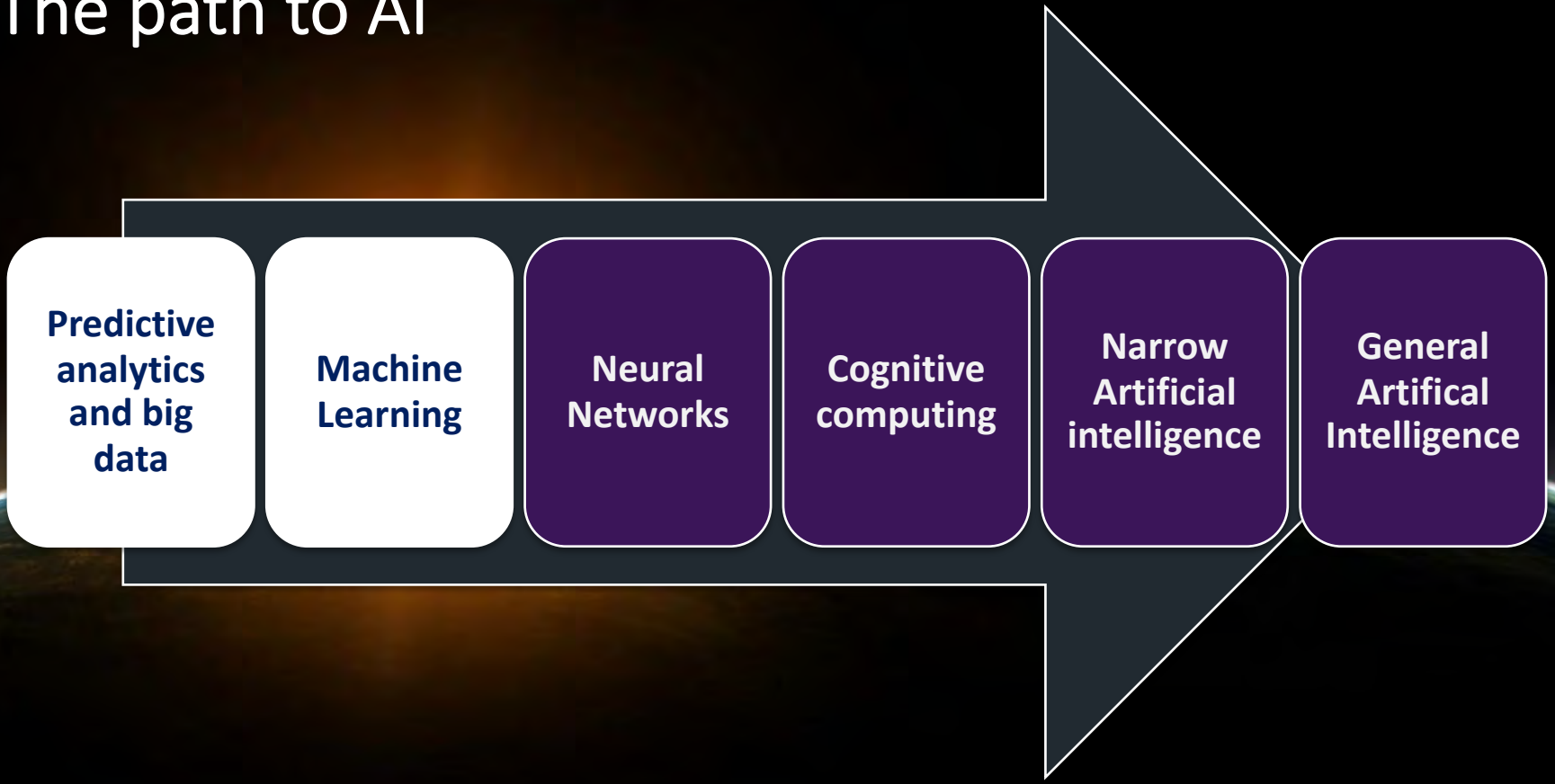
The Progression to AI



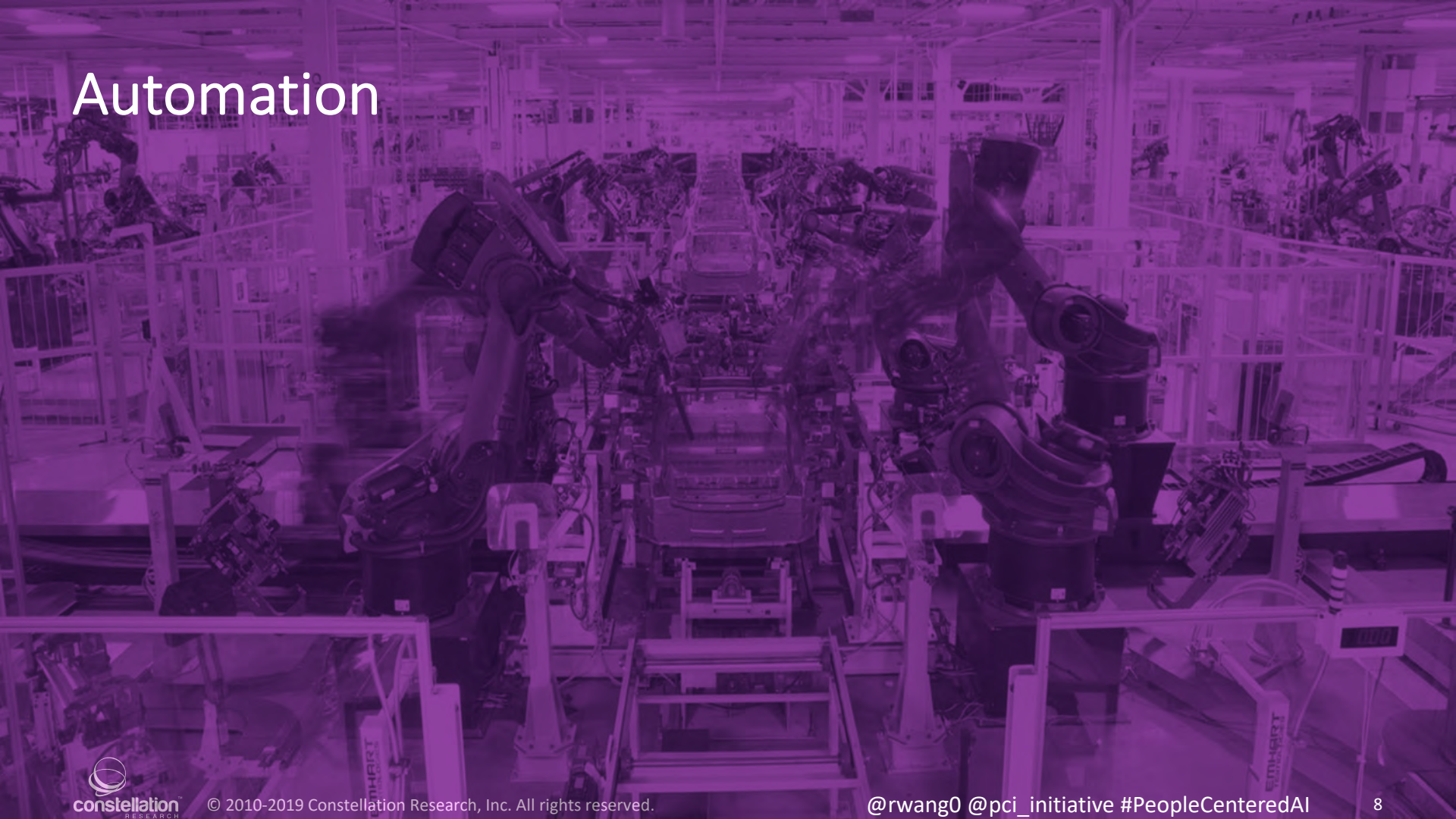
The path to AI



The path to AI

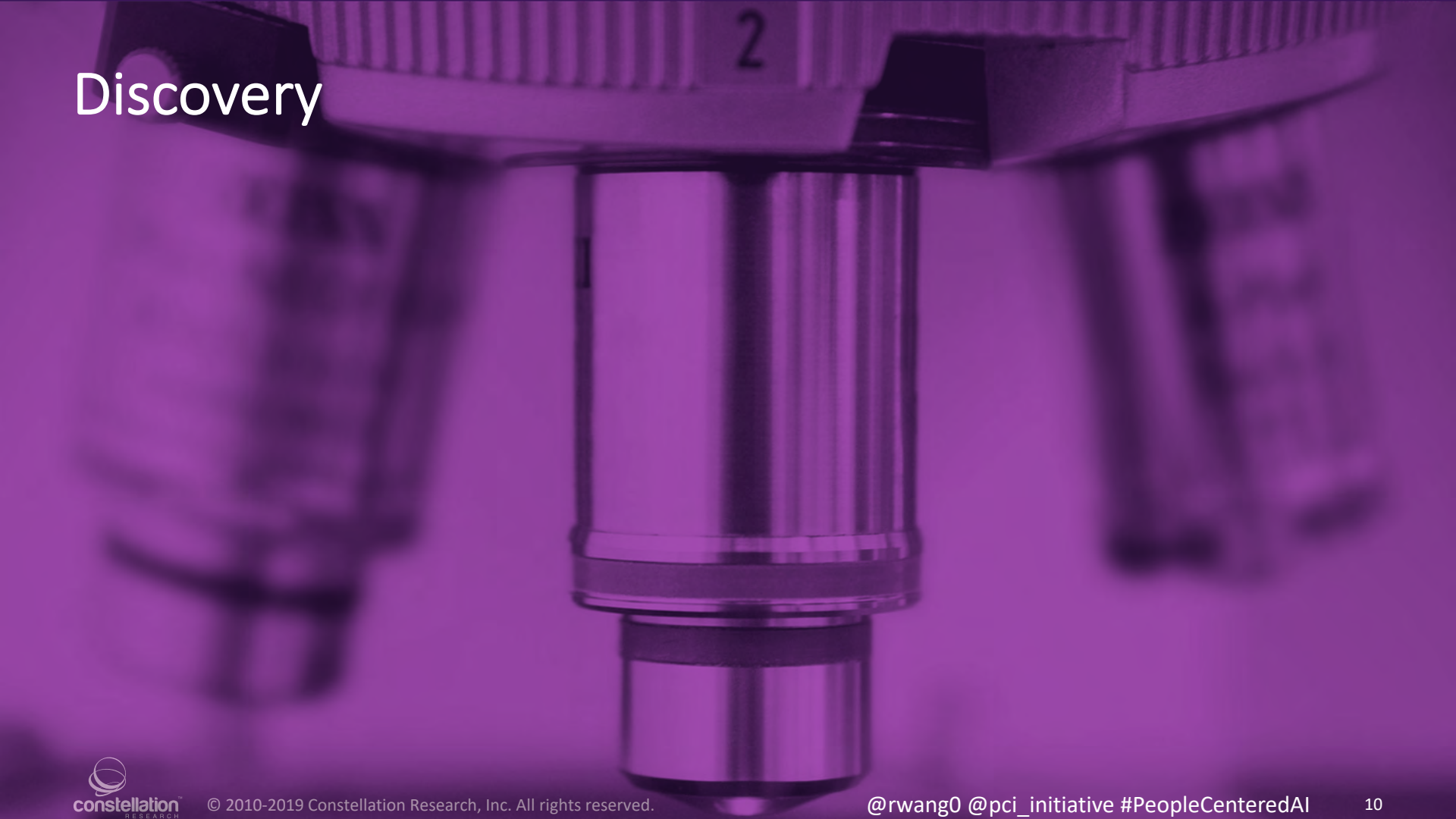


Automation



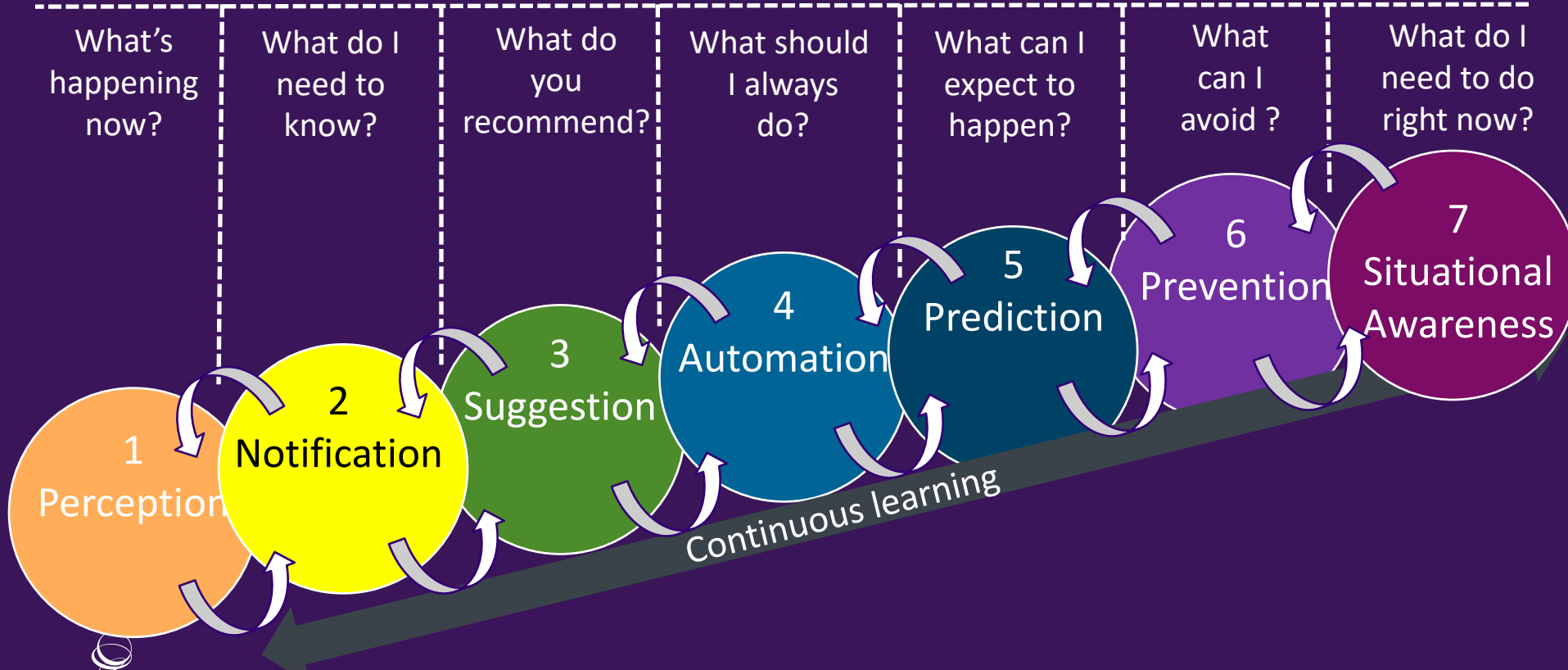
Augmentation

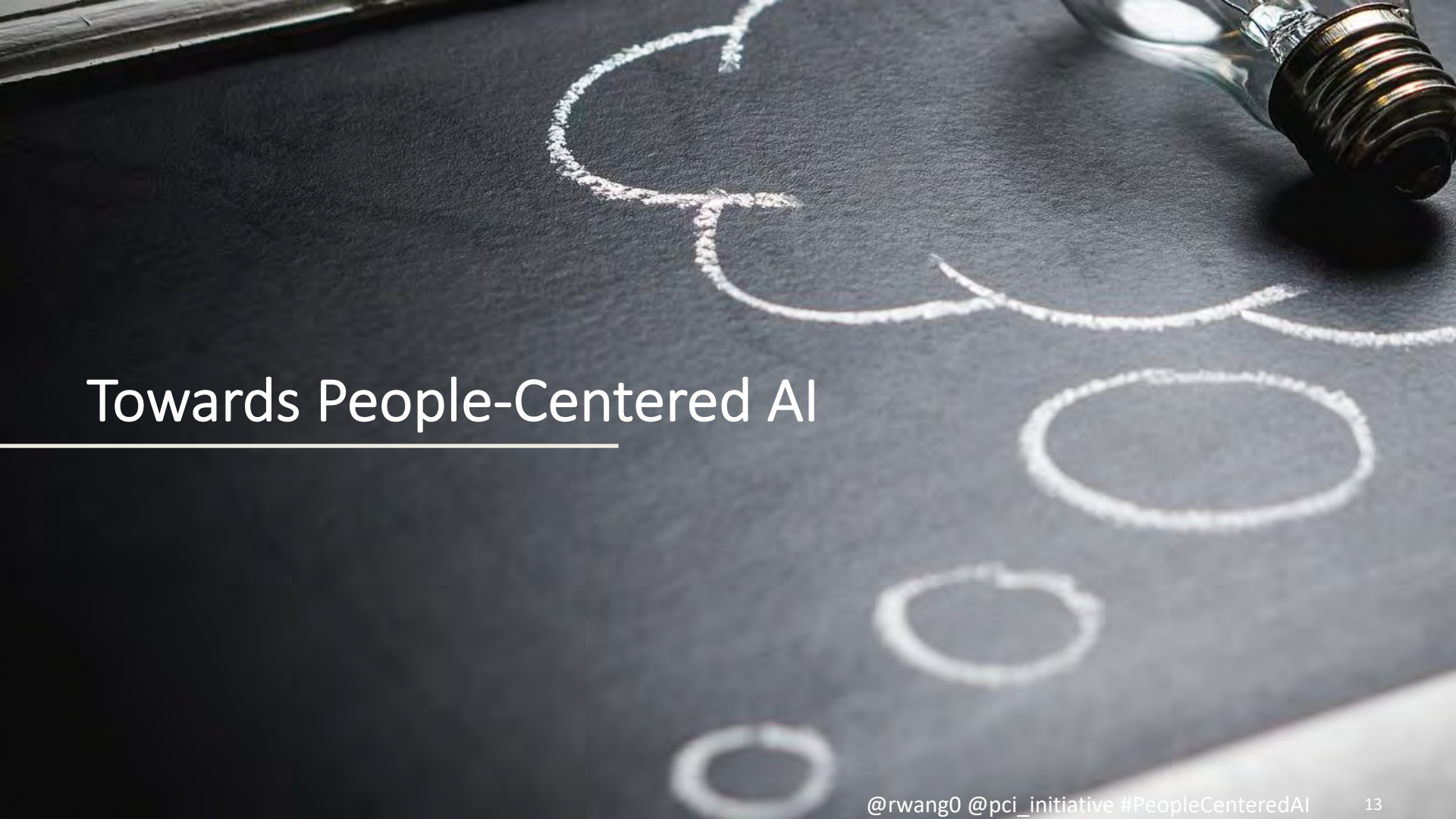
Discovery



Risk mitigation and compliance

Assign outcomes, amplify with artificial intelligence





Towards People-Centered AI

The world will not have a universal set of ethics

Transparency

Wherever possible, make the high-level implementation details of your AI project available for all of those involved.

People should understand what AI is, the way it works to include how data sets are used to tune algorithms, and how AI may affect their work.

When intellectual property or other sensitive information might be exposed, an organization may want to include a panel of external stakeholders, keeping in mind that certain data sets might need to be protected from disclosure if they contain sensitive information or there are privacy concerns.

Explainability

Employees within an organization and external stakeholders, to include potential customers, should be able to understand how any AI system arrives at its contextual decisions.

The focus here is less on an explanation of how the machine reached its conclusions — as AI often cannot be explained at that level of detail — and more on the level of what method was used to tune the algorithm(s) involved, what data sets were employed, and how human decision makers decided to use the algorithm's conclusion.

Reversibility

Organizations also must be able to reverse what a deep learning effort "knows."

Think of it as the ability to unlearn certain knowledge or data, which helps protect against unwanted biases in data sets. Reversibility is something that must be designed into the conception of an AI effort and often will require cross-functional expertise and support.

Create data advocates

To reduce the risk of poorly tuned AI implementations, organizations can create a data advocate or ombudsman function that brings together human stakeholders from different business units (with potential to include outside stakeholders as well).

Data advocates are responsible for ensuring the data sets are both appropriate for the questions being asked of any AI implementation and sufficiently diverse for optimal tuning.

Create data advocates

Consider a people-centered code of ethics.

Have fair ways for people to elevate potential concerns with your products, data use, or other services.

Create data advocates

In working towards more People-Centered approaches to data and AI, all of us have roles to play.

We are the Calvary!

Mindful monitoring

Another way to reduce risk is for organizations to establish a mindful monitoring system to test data sets for biases. This technique requires identifying three pools of data sets: (1) trusted data — the “trusted pool”; (2) potentially worthwhile data — the “queued pool”; and (3) problematic or unreliable data — the “naysayer pool.”

Data outputs from an AI implementation — which are tuned on a queued pool of data (yet to be fully vetted or accepted) — are compared with the outputs from the trusted pool of data.

Mindful monitoring – Example, part 1

Data pool	What data lives here	Monitoring actions for company
Trusted	This vetted data is fit for training AI systems.	Regularly assesses if previously approved data for the organization might now be obsolete, problematic, or unreliable.

Mindful monitoring – Example, part 2

Data pool	What data lives here	Monitoring actions for company
Trusted	This vetted data is fit for training AI systems.	Regularly assesses if previously approved data for the organization might now be obsolete, problematic, or unreliable.
Queued	Data that may be useful for training AI but has not been vetted.	Regularly assesses if this data can improve the company's existing pool of trusted data.
Naysayer	Data that is unfit for training AI. This data pool monitors data in other pools for outdated, inaccurate data.	Regularly assesses the robustness and diversity of the data used to train the deep learning system.

Bounded expectations

Organizations also should clearly specify how data sets will be used to train AI networks, and explain to external stakeholders and internal employees what the accepted norm will be for how the company relies on the data gathered with deep learning.

This method requires a clear list of what the organization can do with the data it generates or acquires, along with what it cannot do. Companies should also make clear the steps that have been taken to verify these bounds — ideally through a third party, such as an outside compliance review.

Bounded expectations – OARS Framework, part 1

Obligations to Society?	Acknowledged Unknowns?

Bounded expectations – OARS Framework, part 2

Obligations to Society?	Acknowledged Unknowns?
Responsible Actions?	Safeguards?

Closing thoughts: towards people-centered AI

As you grow, consider a “Ombuds” function for your Board.

Involve your board and outsiders on important questions re: data you are using, producing, and how your AI services impact people.

MITSloan Management Review

Today's webinar will be available to watch on-demand within several business days.

Watch your inbox for access instructions.

WEBINAR SPONSORED BY: