

Ethics in AI

Discussion points

- **Why** do we need to worry about this?
- **What** is required?
- **Who** do we need to enable?
- **How? Which** tools & frameworks?

Pillars of Trustworthy AI

- Privacy
- Robustness
- Fairness
- Explainability
- Transparency

Open Source Tools

- Adversarial Robustness Toolkit
- AI Fairness 360
- AI Explainability 360
- AI Factsheets 360

Let's connect!

- [Twitter.com/naisofly](https://twitter.com/naisofly)
- [Linkedin.com/in/naiyarah](https://linkedin.com/in/naiyarah)
- [Github.com/naisofly](https://github.com/naisofly)

Digital Developer Conference

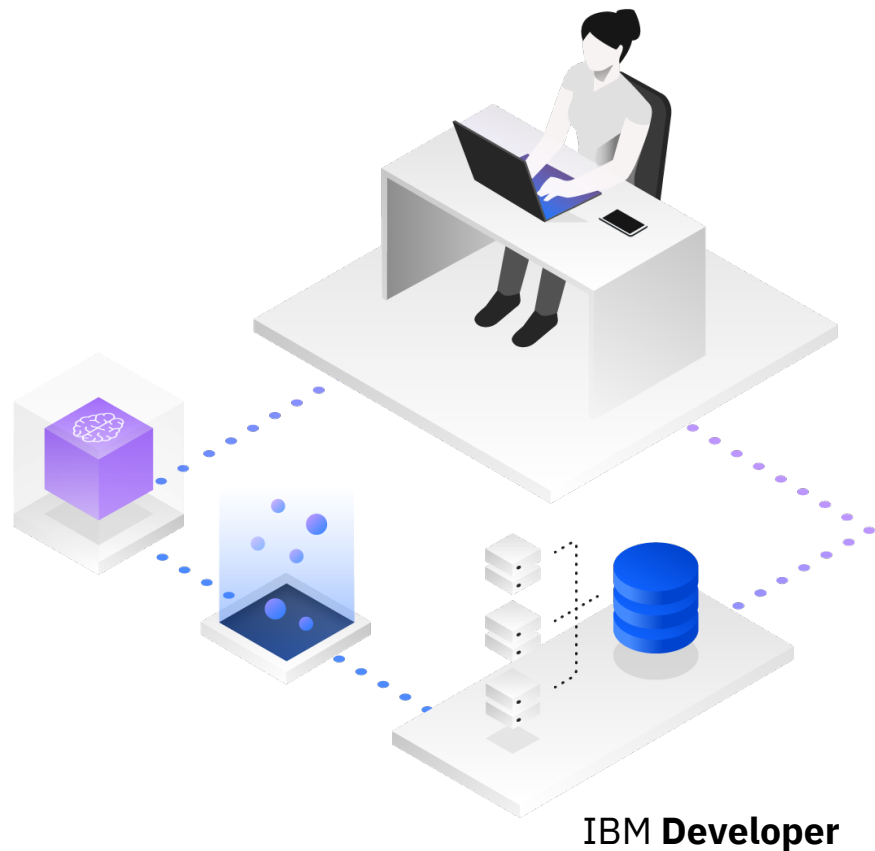
Data & AI

Trustworthy AI

- **Track 1: Trustworthy AI**
- **Track 2: Demand Proof: The urgent need for Data Lineage and Provenance**
- **Track 3: Beyond Accuracy : What does it take to trust AI decisions**
- **Track 4: Uncertainty Quantification 360 - Learn how to apply the UQ360 toolkit**

📅 June 8th 2021

ibm.biz/devcon-ai



↳ AI is changing the World



AI is used in many decision-making applications



Credit



Employment



Admission



Sentencing



Healthcare

Deploying AI requires Careful Management. Why?

“Responsible AI focuses on ensuring the ethical, transparent and accountable use of AI technologies in a manner consistent with user expectations, organizational values and societal laws and norms.”

Accenture, 2018

Why Responsible AI?

What's “right” and “wrong”

Ethical debates are well underway about what's “right” and “wrong” when it comes to high-stakes AI applications such as autonomous weapons and surveillance systems.

Transfer of Human Bias to AI Systems

There's an outpouring of concern and skepticism regarding how we can **imbue AI systems with human ethical judgment**, when moral values frequently vary by culture and can be difficult to code in software.

News about AI Bias

Numerous instances of **AI bias, discrimination, and privacy violations** have already littered the news, leaving leaders rightly concerned about how to ensure that nothing bad happens as they deploy their AI systems

The Apple Card Didn't 'See' Gender—and That's the Problem

The way its algorithm determines credit lines makes the risk of bias more acute.

MONEYBOX

Amazon Created a Hiring Tool Using A.I. It Immediately Started Discriminating Against Women.

By JORDAN WEISSMANN

OCT 10, 2018 • 4:52 PM

MIT Researcher Exposing Bias in Facial Recognition Tech Triggers Amazon's Wrath

By Matt O'Brien | April 8, 2019



Jerome Pesenti
@an_open_mind

#gpt3 is surprising and creative but it's also unsafe due to harmful biases. Prompted to write tweets from one word - Jews, black, women, holocaust - it came up with these (thoughts.sushant-kumar.com). We need more progress on #ResponsibleAI before putting NLG models in production.

Five Focus Area for Ethical AI

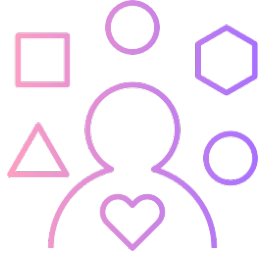
Get Started Here: <https://ibm.biz/MEA-DataAI>

ACCOUNTABILITY



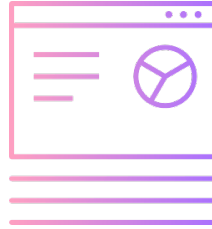
AI designers and developers are responsible for considering AI design, development, decision processes, and outcomes.

VALUE ALIGNMENT



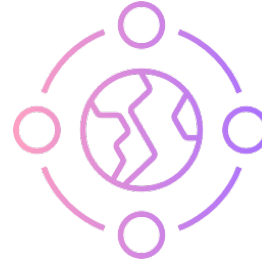
AI should be designed to align with the norms and values of your user group in mind.

EXPLAINABILITY



AI should be designed for humans to easily perceive, detect, and understand its decision process.

FAIRNESS



AI must be designed to minimize bias and promote inclusive representation.

USER DATA RIGHTS

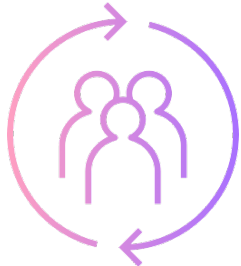


AI must be designed to protect user data and preserve the user's power over access and uses.

How can you adopt responsible AI?

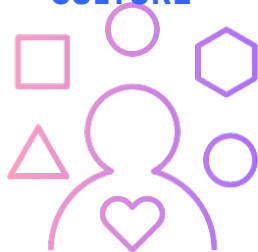
Get Started Here: <https://ibm.biz/MEA-DataAI>

PEOPLE



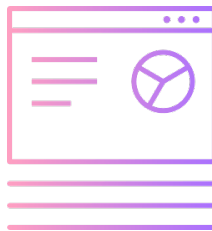
Organization with diverse team do better with ensuring diverse representation in their data and AI pipeline and avoiding bias.

CULTURE



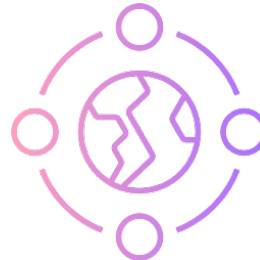
Imbibing a culture of accountability, teams must ensure there are ethics AI Practitioner on their AI team to ensuring the five focal points are covered

DESIGN FRAMEWORK



Having a framework that works for your team and covers the five focal point of accountability

END USERS



Making use of Design Thinking with the end user in mind, understanding their needs and pain points and designing what is usable for the user

TOOLS



Making use of Open Source and commercial tools to address fairness and explainability

Trusted AI Lifecycle through Open Source

Pillars of trust, woven into the lifecycle of an AI application

Did anyone
tamper with it?



ROBUSTNESS

Is it fair?



FAIRNESS

Is it easy to
understand?



EXPLAINABILITY

Is it accountable?



LINEAGE

Adversarial
Robustness 360

↳ (ART)

github.com/IBM/adversarial-robustness-toolbox

art-demo.mybluemix.net

AI Fairness
360

↳ (AIF360)

github.com/IBM/AIF360

aif360.mybluemix.net

AI Explainability
360

↳ (AIX360)

github.com/IBM/AIX360

aix360.mybluemix.net

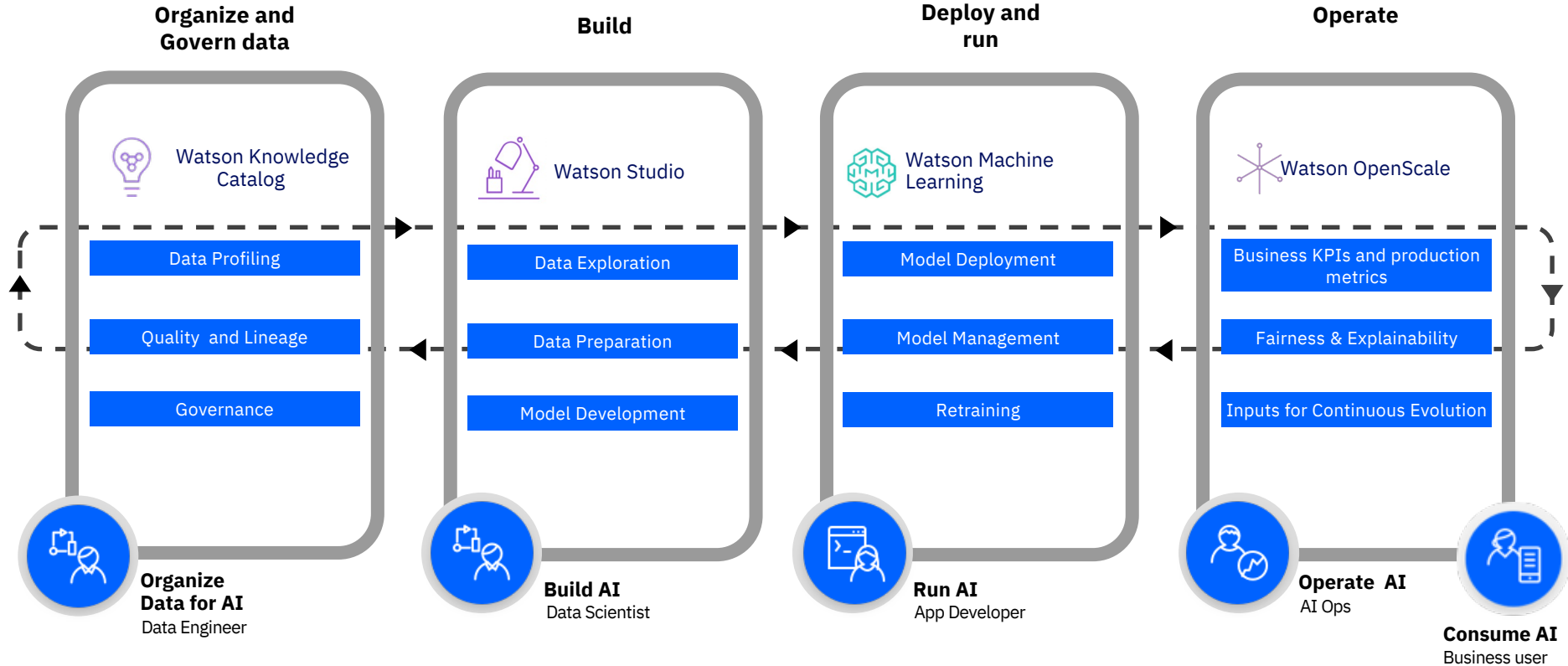
AI FactSheets
360

↳ (AIFS360)

aifs360.mybluemix.net

AI pipeline

Get Started Here: <https://ibm.biz/MEA-DataAI>



Example: credit risk

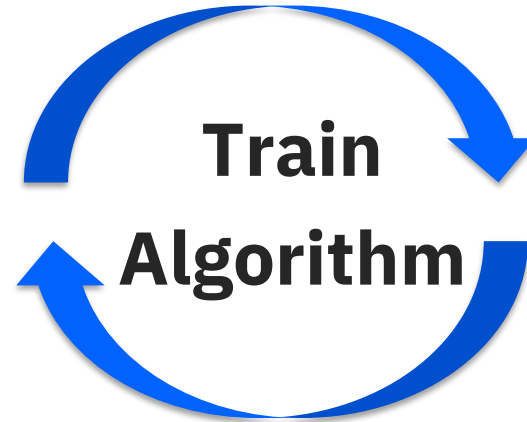
Historical Loans

John X

- **credit_score=800**
- **age=25**
- **income=\$900,000**
- **works in Oil & Gas**

Label

No Risk



Output

**Credit
Risk
Model**

Example: credit risk

New Applicant

James Y

- **credit_score=900**
- **age=55**
- **income=\$1,200,000**
- **works in Insurance**



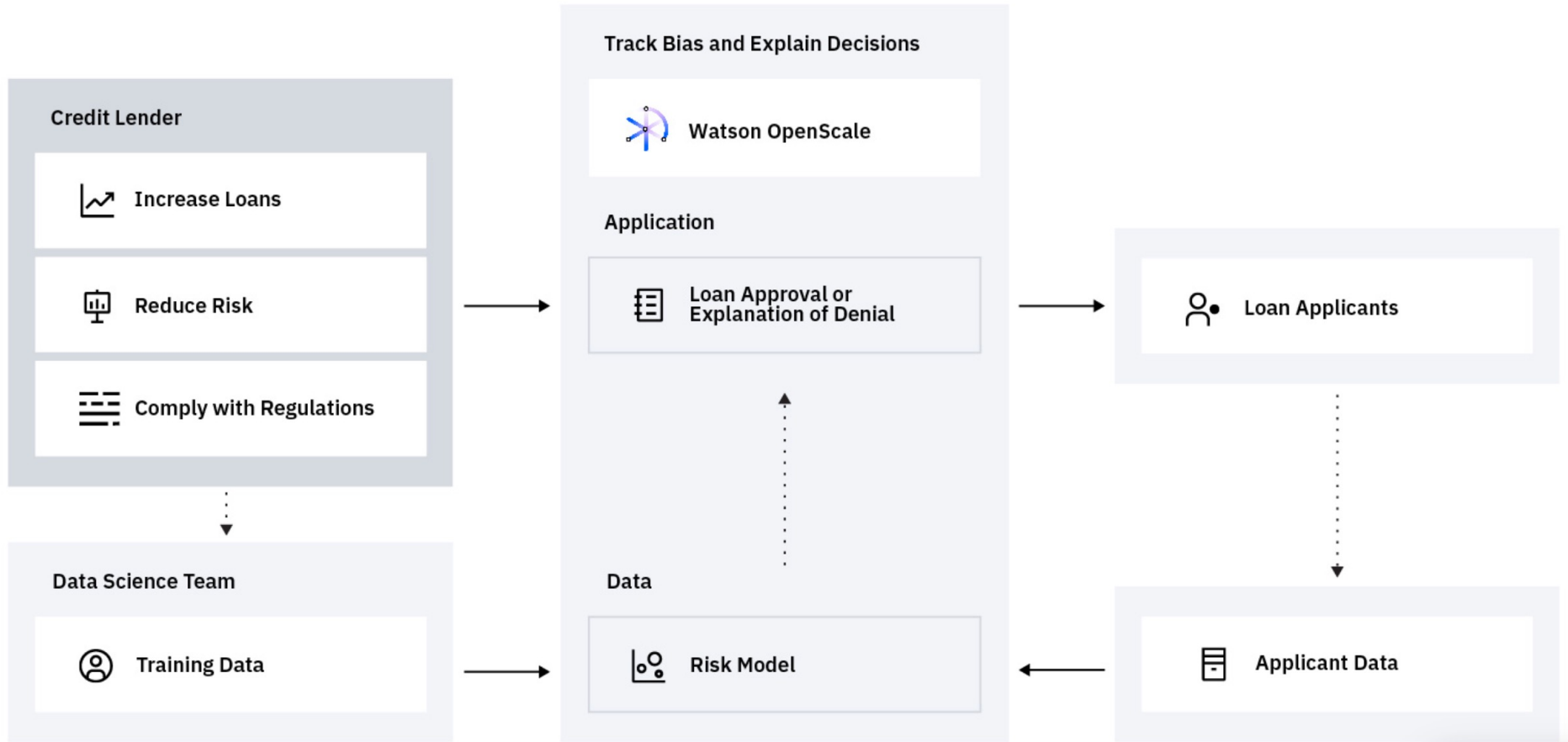
**Credit
Risk
Model**

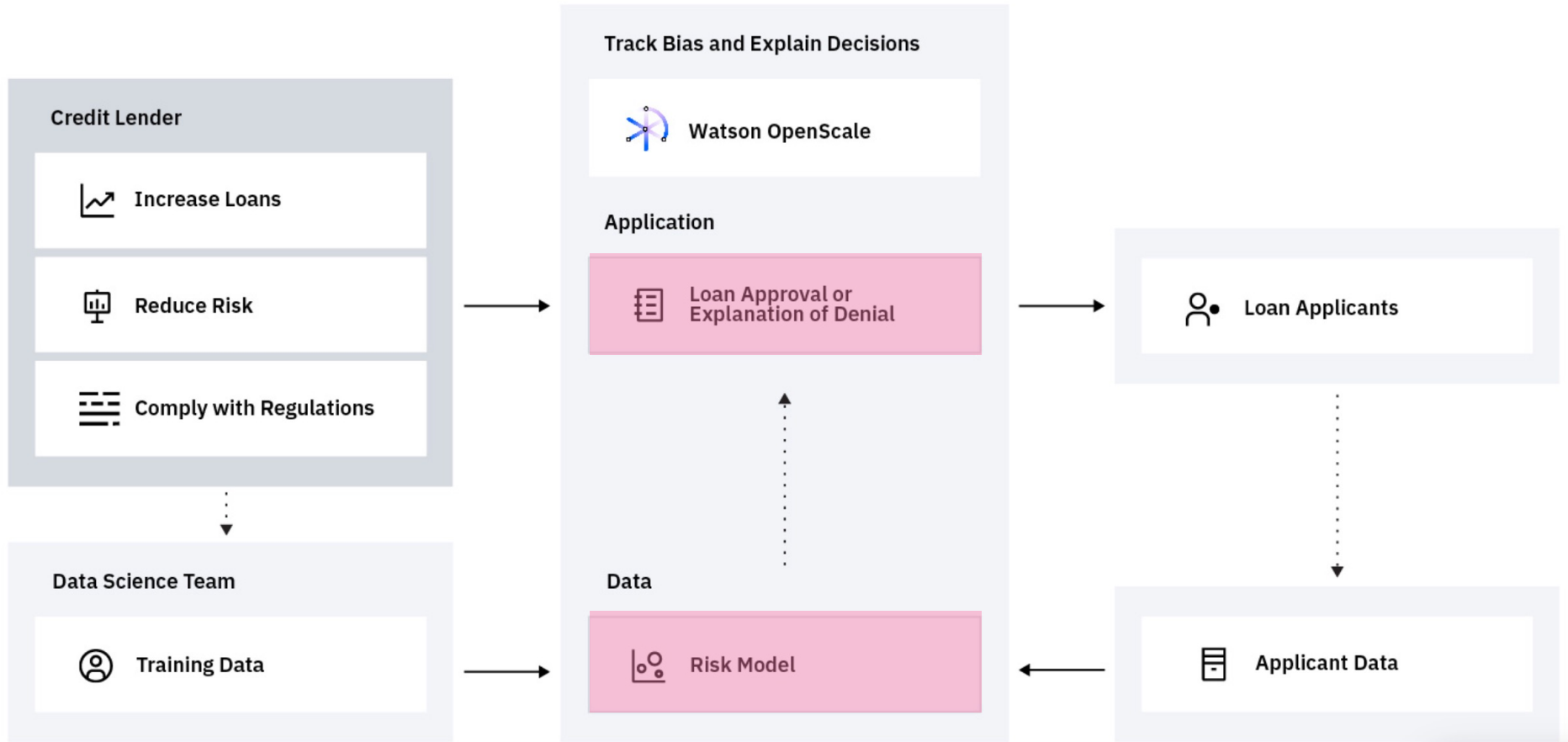


Output

No Risk

Risk





Key Takeaways

- **Ethics of AI is new muscle** we all need to develop
- **Everyone in tech is accountable** and needs to own building responsible and ethical AI systems.
- **Cross Functional** teams are key. Inclusion of **non-traditional disciplines** ensures that tech builders have a broader perspective from inception
- **Introduce User Research to your teams** so that they can build **empathy** for their users and better understand their needs and pain points
- Define **accountability** for your team

