

Trustworthy Machine Learning and Artificial Intelligence

—

Kush R. Varshney
krvarshn@us.ibm.com | @krvarshney

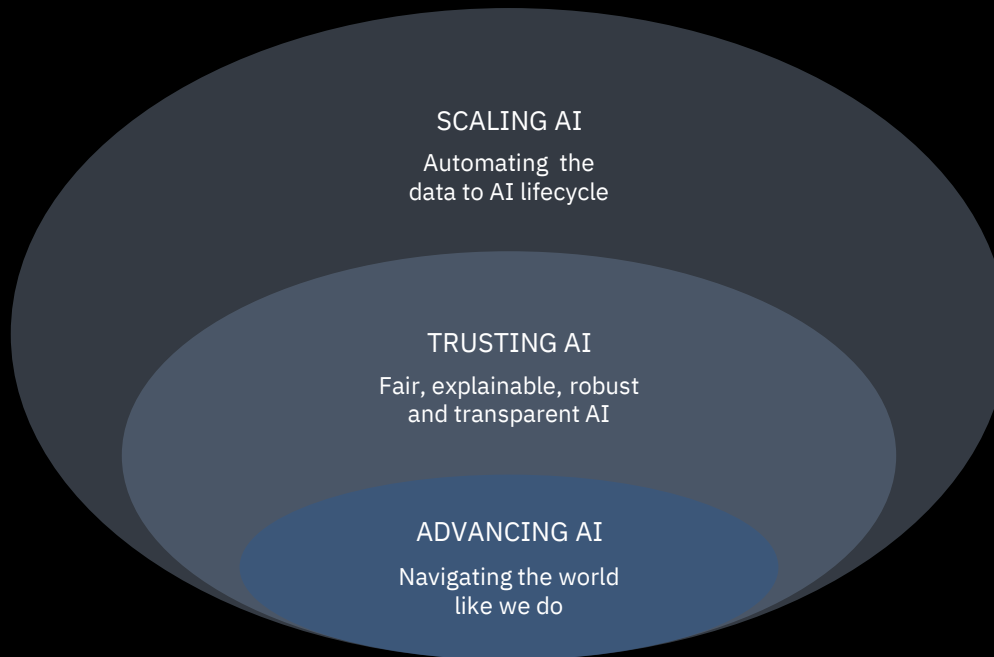
IBM Research AI

Our mission

Advance the
field of AI from
narrow to broad

Narrow AI
Single task single domain
Deep Learning heavy
Data & compute hungry

Broad AI
Multi-task multi-domain
Learn with less data & compute
Trustworthy



Address enterprise
challenges in
scaling AI

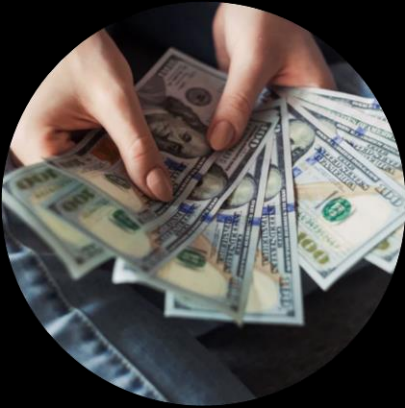
Skills Challenge

Data Challenge

Operations Challenge

Governance Challenge

AI is powering critical workflows and trust is essential



loan
processing



employment



customer
management



quality control

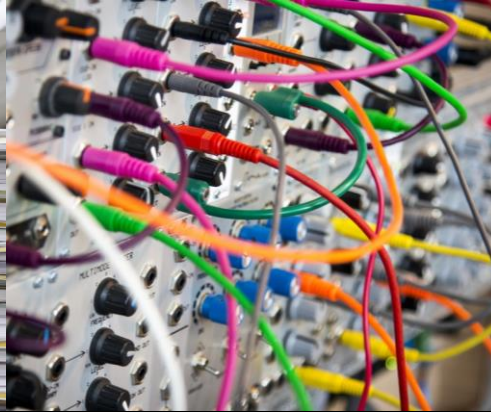
Multiple factors are placing trust in AI as a top business priority



brand reputation



increased regulation



complexity of AI deployments



focus on social justice

What does it take to trust a decision made by a machine?

is it accurate?



is it fair?



did anyone tamper with it?



is it easy to understand?



is it accountable?



is it transparent?



International Time Recording Company
Dayton Scale Company
International Scale Company
Home Office: 270 Broadway
New York, N. Y.



For thirty-one years, the gatherings and conventions of our IBM workers have expressed in happy songs the fine spirit of loyal cooperation and good fellowship which has promoted the signal success of our great IBM Corporation in its truly International Service for the betterment of business and benefit to mankind.

In appreciation of the able and inspiring leadership of our beloved President, Mr. Thomas J. Watson, and our amiable staff of IBM executives, and in recognition of the noble aims and purposes of our International Service and Program, 1931, all purposes of our International Service and Program, 1931, all purposes of IBM songs solicit your vocal approval by hearty cooperation in our song-fests at our conventions and fellowship gatherings.

Yours in International Service,
HARRY S. EVANS

"Progressive Men Employ Progressive Methods"

THINK

REFLEXION

HISSEZ

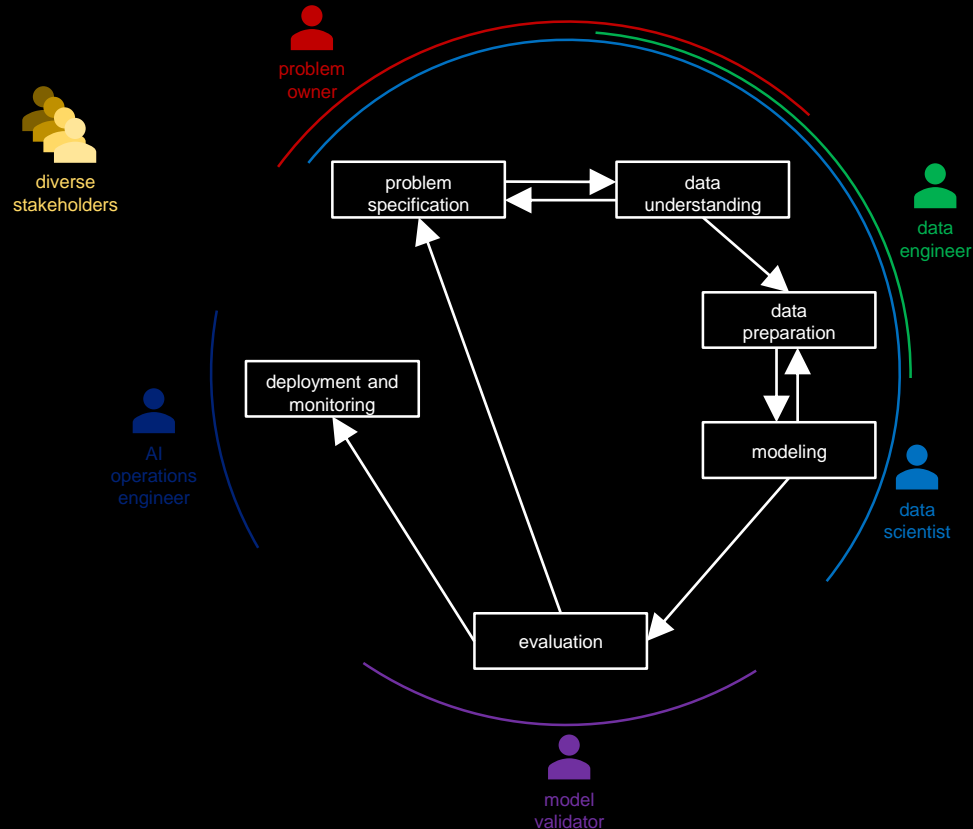
思維

“The toughest thing about the power of trust is that it’s very difficult to build and very easy to destroy.”

—Thomas J. Watson, Sr., CEO of IBM

SONGS
of
The I.B.M.

Machine learning lifecycle



Example – pay-as-you-go solar power

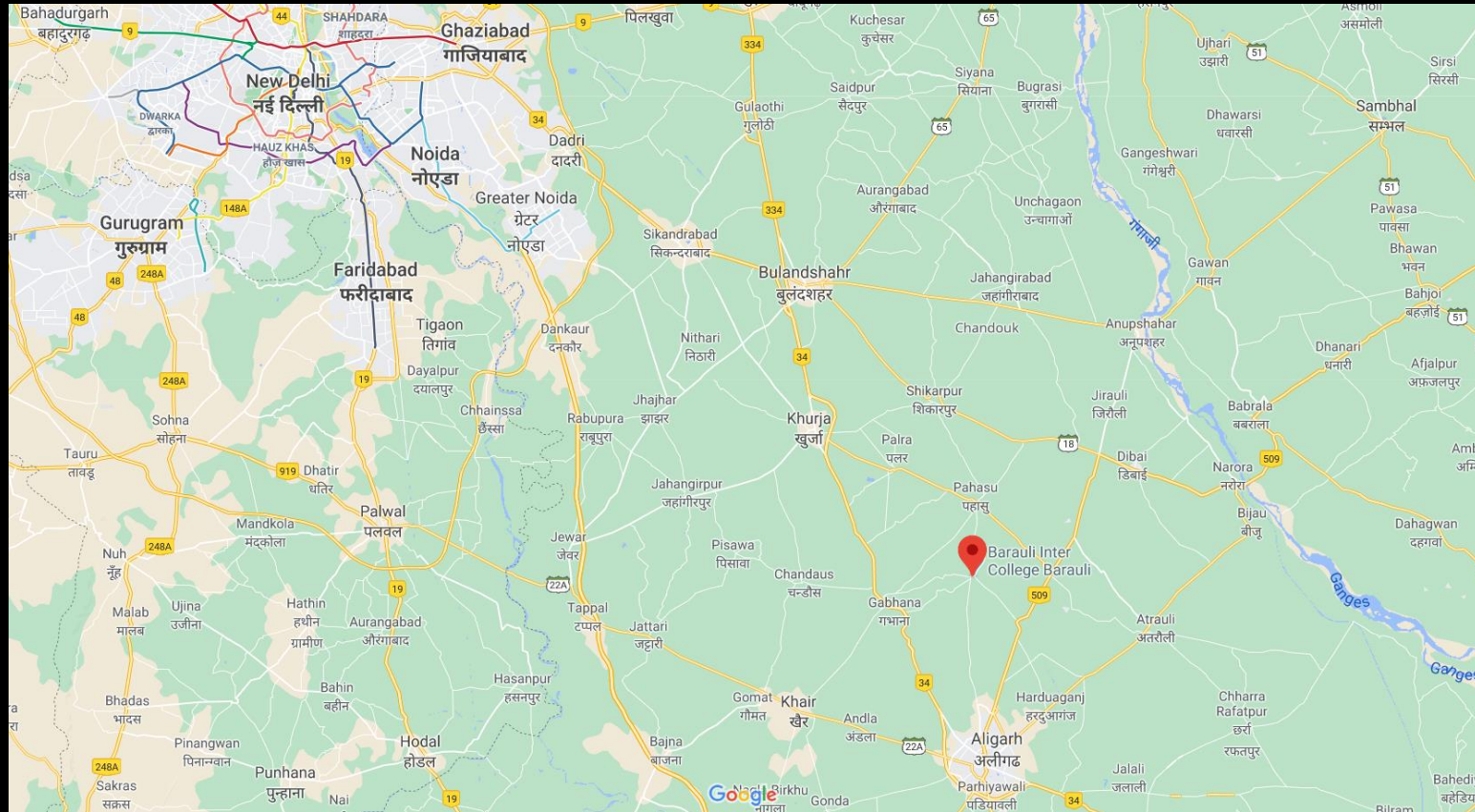
H. Gerard, K. Rao, M. Simithraaratchy, K. R. Varshney, K. Kabra, and G. P. Needham. "Predictive Modeling of Customer Repayment for Sustainable Pay-As-You-Go Solar Power in Rural India." *Data for Good Exchange*, New York, New York, USA, September 2015.

K. R. Varshney. "Signal Processing for Social Good." *IEEE Signal Processing Magazine*, vol. 34, issue 3, p. 112, 108, May 2017.



Predict which applicants will fully pay for (unlock) their system and which ones will have to be repossessed.

Example – pay-as-you-go solar power

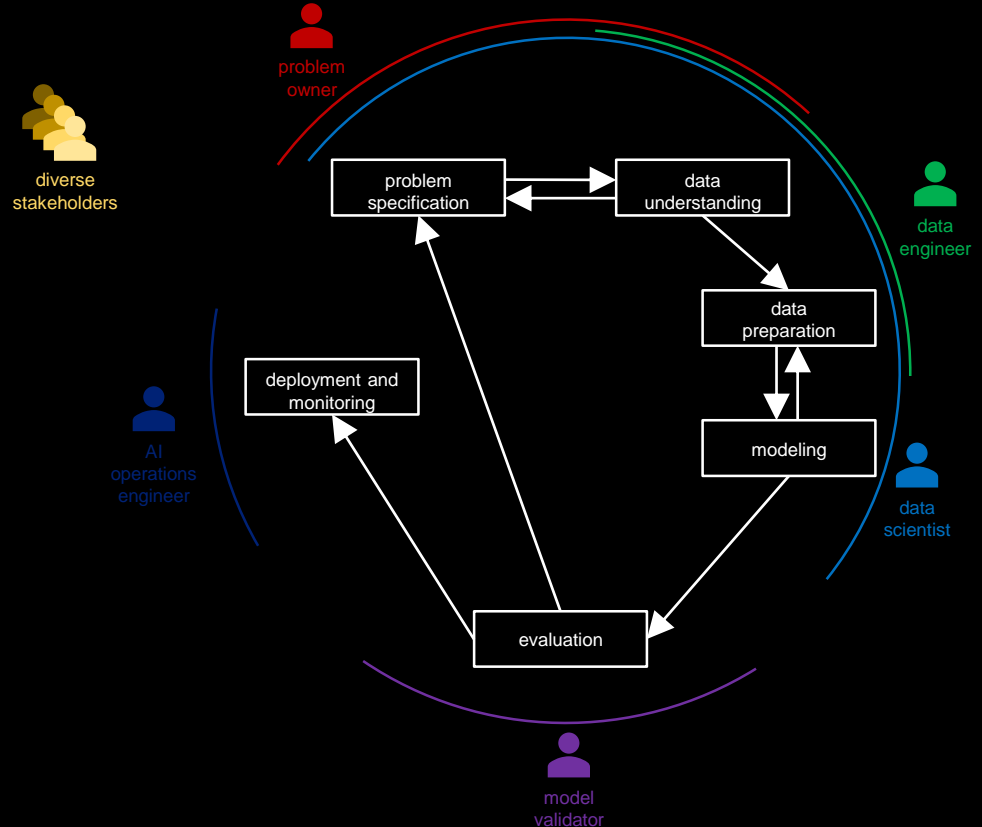


Problem specification

Is the problem the right one to solve?

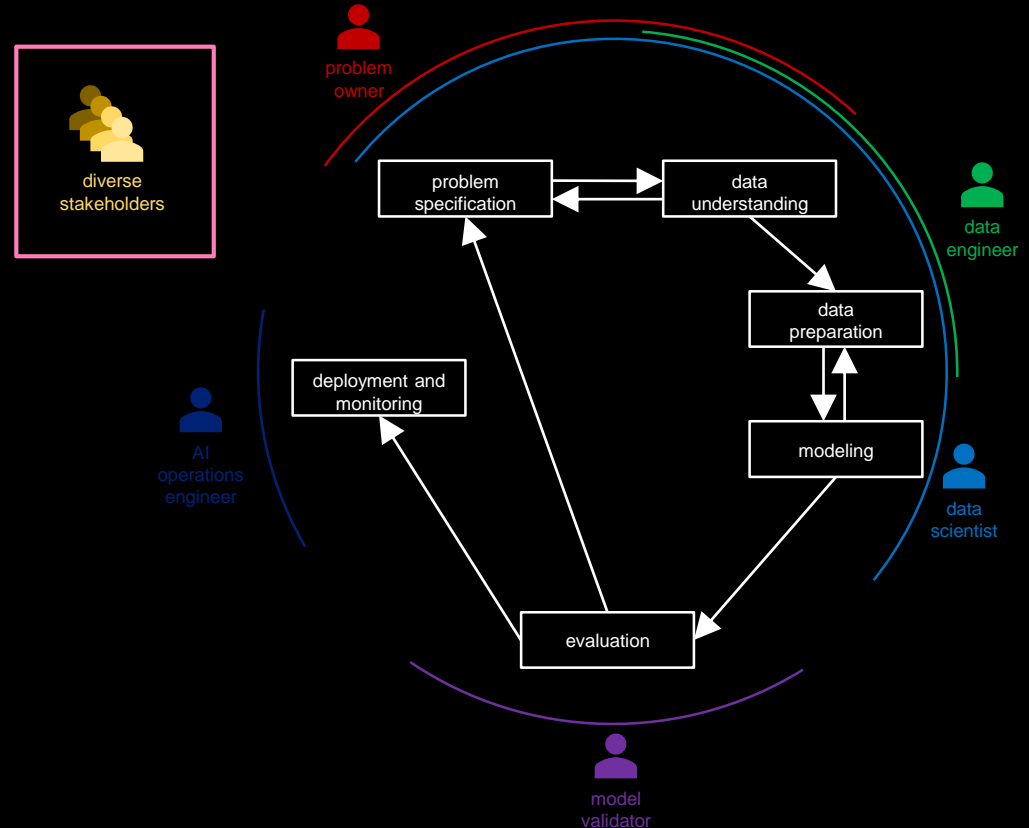
Will solving it cause any injustice?

What are the objectives?

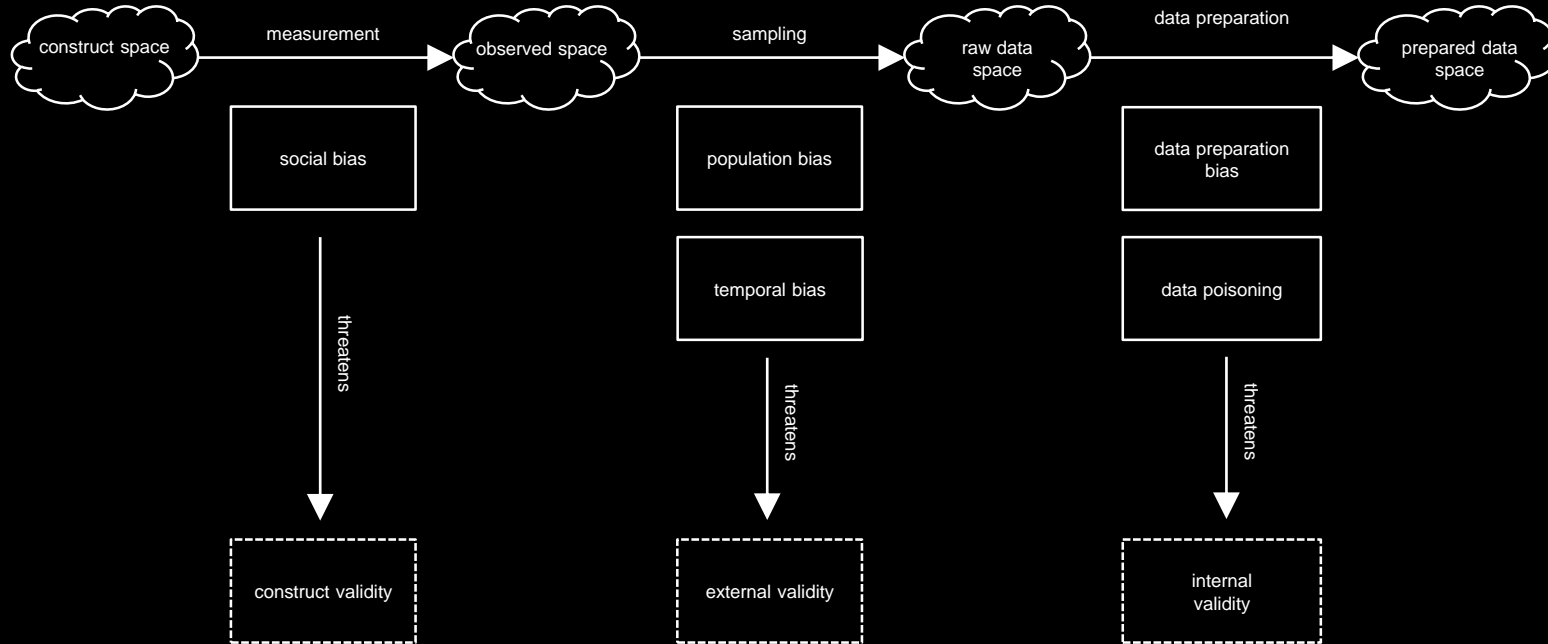


Problem specification

Take advice from a panel
of diverse voices



Data understanding and preparation



Data understanding and preparation

Introducing any new biases?

Recoding features

Sensitive features

- Applicant's last name

Outcome variable

- Care management by health plans

K. R. Varshney and H. Alemzadeh. "On the Safety of Machine Learning: Cyber-Physical Systems, Decision Sciences, and Data Products." *Big Data*, vol. 5, issue 3, p. 246-255, September 2017.

M. Singh and K. Natesan Ramamurthy. "Understanding Racial Bias in Health using the Medical Expenditure Panel Survey Data." *NeurIPS Workshop on Fair Machine Learning for Health*, Vancouver, BC, Canada, December 2019.

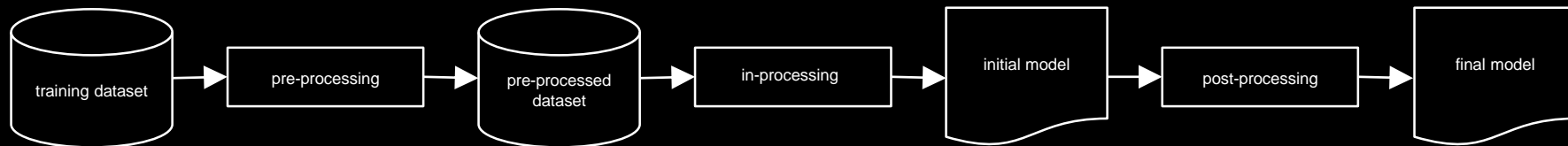
Nature of Business	Recoded Value
Animal husbanded	Livestock Business
Animals business	Livestock Business
Ara Machine	Skilled Labour/Driver
Arif Vairayte pailesh	Restaurant and Hotel
Ata Chakki	Grains/Fruits
Atta Chaki	Grains/Fruits
Atta Chakki	Grains/Fruits
Auto	Vehicle Sales and Repair
auto parts shop	Vehicle Sales and Repair
Auto Dealer	Vehicle Sales and Repair
Auto driver	Skilled Labour/Driver
auto mobile shop	Vehicle Sales and Repair
Auto parts	Vehicle Sales and Repair
Auto Parts Repairing	Vehicle Sales and Repair
Auto parts shop	Vehicle Sales and Repair
Auto spare parts	Vehicle Sales and Repair
Auto(Transport)	Skilled Labour/Driver
Automobiles shop	Vehicle Sales and Repair
Autoparts	Vehicle Sales and Repair
B.K satring Matarial Shop	Construction Business
Bafallo Sale Business	Livestock Business
Bajaj Agency	Vehicle Sales and Repair
Bajaz Dealer	Vehicle Sales and Repair
band	Tent House and Band
Band Business	Tent House and Band
Band Shop	Tent House and Band
banke bihari general store.	General Business
Baraber	Beauty
Barber	Beauty
bartan ki shop	Jewelry and Pots

Commodity	Recoded Value
Barley	Other
Dhaan	Paddy
Dhan	Paddy
Jo	Other
Maize	Other
Makka	Other
Millet	Other
mustard	Other
Others	Other
Paddy	Paddy
Patato	Potato
Peace	Other
peas	Other
Potato	Potato
Sarson	Other
Sorghum	Other
Sugarcan	Other
Sugarcane	Other
Sunflower	Other
Urad Dal	Other
weat	Wheat
Whaet	Wheat
Wheat	Wheat
Wheet	Wheat
Whhet	Wheat

Modeling

R. K. E. Bellamy, K. Dey, M. Hind, S. C. Hoffman, S. Houde, K. Kannan, P. Lohia, J. Martino, S. Mehta, A. Mojsilović, S. Nagar, K. Natesan Ramamurthy, J. Richards, D. Saha, P. Sattigeri, M. Singh, K. R. Varshney, and Y. Zhang. "AI Fairness 360: An Extensible Toolkit for Detecting and Mitigating Algorithmic Bias." *IBM Journal of Research and Development*, vol. 63, issue 4/5, p. 4, July/September 2019.

Bias mitigation algorithms



Machine learning lifecycle

