Actionable Recourse in Linear Classification

Berk Ustun
Harvard University

Joint work with Alexander Spangher and Yang Liu

When you are denied a loan...

Do you know **what** you can do to be approved?

Some lenders are judging you on much more than finances

By JAMES RUFUS KOREN
DEC 19, 2015 | 10:00 AM

Douglas Merrill is founder and CEO of ZestFinance, the parent company of online lender Basix. His firm uses unorthodox metrics to measure creditworthiness but not social media data, which he finds "personally creepy." (Michael Robinson Chávez / Los Angeles Times)

Source: LA Times, 2015

Bloomberg Opinion Subscribe

Technology & Ideas

Own an Android Phone? You Might Not Get That Loan

Algorithms could determine our creditworthiness based on data we didn't know was available or relevant.

By Leonid Bershidsky
May 4, 2018, 6:00 AM EDT
Corrected May 14, 2018, 11:30 AM EDT

Forbes

1,304 views | Apr 25, 2018, 06:00am

Could Personality Tests One Day Replace Credit Scores?

Andrew Josuweit Contributor

Image Courtesy of Student Loan Hero STUDENT LOAN HERO

Recourse in Machine Learning

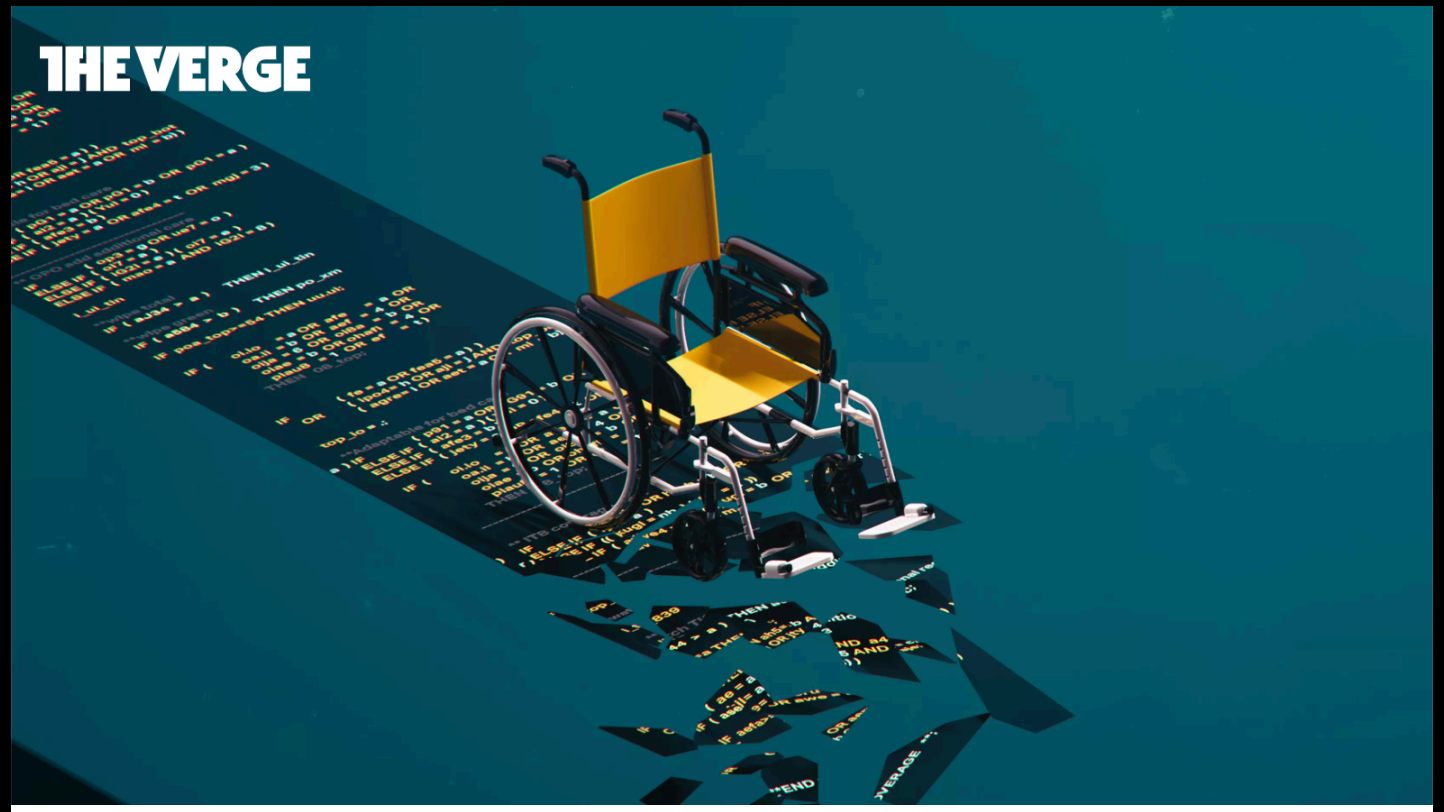
- Recourse = ability to obtain a desired prediction from a model by changing **actionable input variables**

- Recourse ≠ explainability
  - *Why did the model deny the loan? Is this a meaningful reason?*
  - *What can a person do to obtain loan? Is there any feasible action?*

- Recourse = agency in model's decision-making process

1. When should we care about recourse?
2. Why models may not provide recourse
3. Tools to check recourse for linear classifiers
4. Lessons for consumer protection

Public Services



SCIENCE

WHAT HAPPENS WHEN AN ALGORITHM CUTS YOUR HEALTH CARE

By Colin Lecher | @colinlecher | Mar 21, 2018, 9:00am EDT

Hiring

Bartleby

How an algorithm may decide your career

Getting a job means getting past the computer

Source: The Economist, 2018

Insurance

THE WALL STREET JOURNAL.

New York Insurers Can Evaluate Your Social Media Use—If They Can Prove Why It's Needed

New guidance applies to companies operating in New York, but industry consultants say it could have an impact beyond the state's borders

Insurers Want to Know How Many Steps You Took Today

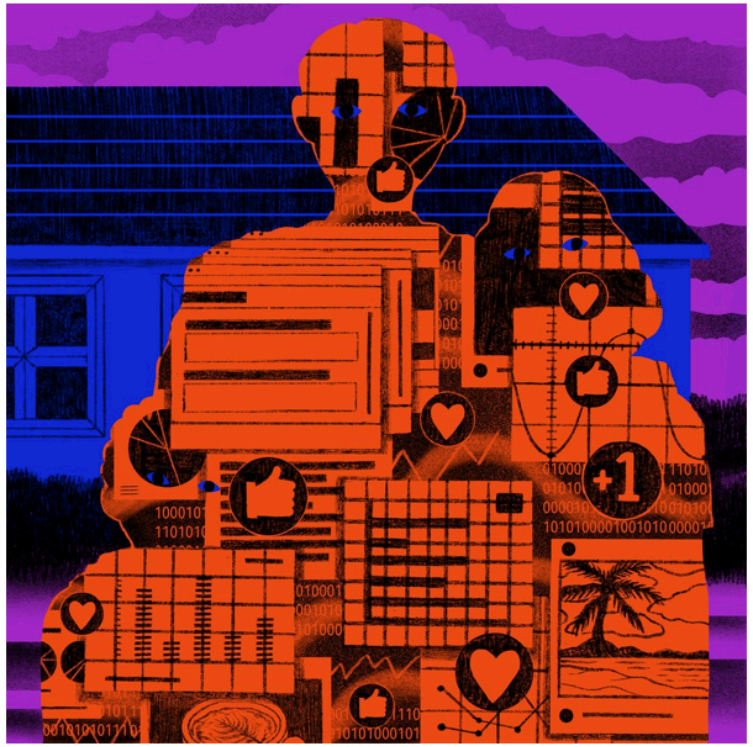
The cutting edge of the insurance industry involves adjusting premiums and policies based on new forms of surveillance.

By Sarah Jeong

Ms. Jeong is a member of the editorial board.

April 10, 2019

448



Lending

Some lenders are judging you on much more than finances

By JAMES RUFUS KOREN
DEC 19, 2015 | 10:00 AM

Insurance

THE WALL STREET JOURNAL.

New York Insurers Can Evaluate Your Social Media Use—If They Can Prove Why It's Needed

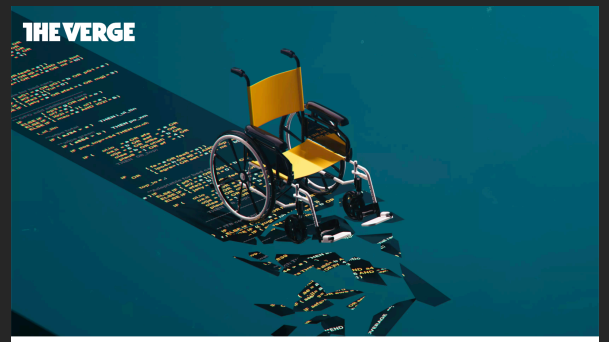
New guidance applies to companies operating in New York, but industry consultants say it could have an impact beyond the state's borders

Hiring

How an algorithm may decide your career

Getting a job means getting past the computer

Public Services



SCIENCE

WHAT HAPPENS WHEN AN ALGORITHM CUTS YOUR HEALTH CARE

By Colin Lecher | @colinlecher | Mar 21, 2018, 9:00am EDT

1. When should we care about recourse?

2. Why models may not provide recourse

3. Tools to check recourse for linear classifiers

4. Lessons for consumer protection

Transparent Models May Not Provide Recourse

1.	Age $\geq$ 60	20 points		0
2.	Income $\geq$ \$50K	10 points	$+$	10
3.	Savings $\geq$ \$5K	10 points	$+$	10
4.	PersonalityType is ENTJ	10 points		0
		SCORE	$=$	20

APPROVE LOAN IF SCORE ≥ 25

Variables that Cannot or Should Not Have to Change

- `HasPhD` ☜ can't just 'un-PhD'
- `Age` ☜ no fountain of youth
- `PersonalityType` ☜ beyond repair
- `AndroidPhone` ☜ shouldn't have to switch
- `MaritalStatus` ☜ shouldn't have to marry

Why not just Regulate Input Variables?

- Minor Differences in Variable Encoding
  - `LatePayment` vs `LatePaymentInLastYear`
- Changes in Deployment Population
  - set of feasible actions
  - missing features
- Superficial Feasibility
  - loan requires increasing `income` by `$10M`

GOAL

evaluate **feasibility & difficulty** of recourse on **deployment population**

1. When should we care about recourse?

2. Why models may not provide recourse

3. Tools to check recourse for linear classifiers

4. Lessons for consumer protection

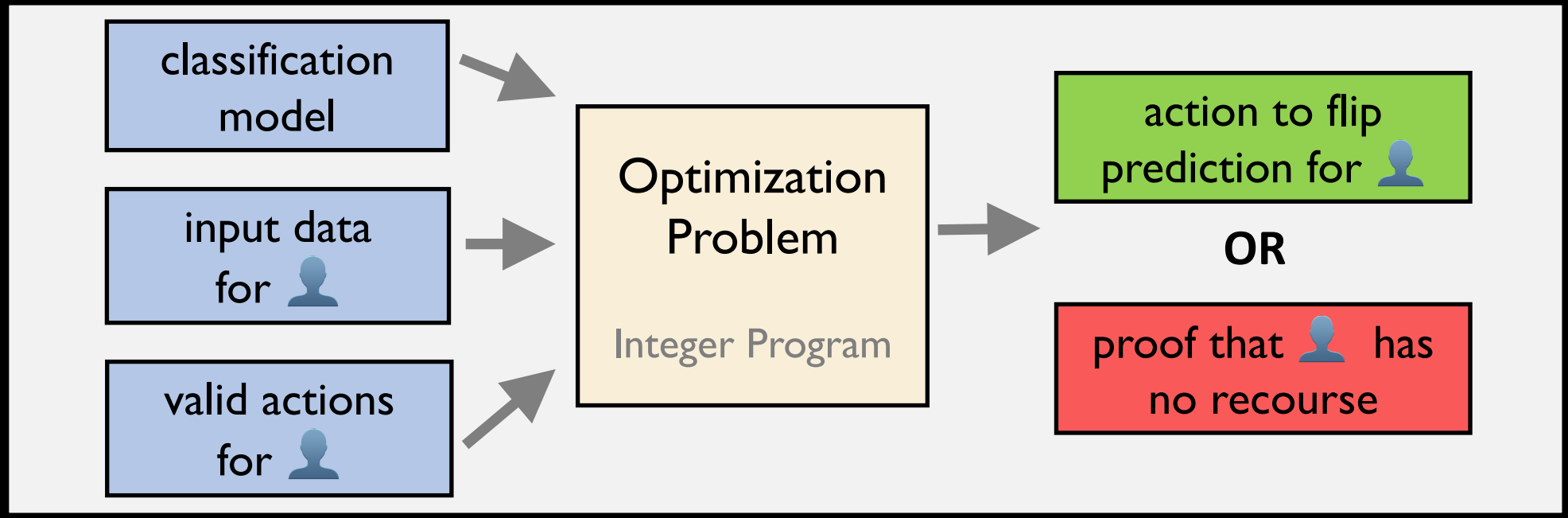
Our Paper

Methods to evaluate recourse without interfering in model development.

Questions that can be answered with our tools:

1. What can a person change to be approved for a loan?
2. What is the feasibility and difficulty of recourse in a population of interest?

Routine to Check Recourse for 1 Person (👤)



- fast ☜ <1 second
- all data types ☜ ordinal, categorical, continuous
- specialized cost functions ☜ to measure / minimize difficulty of actions
- linear classification models ☜ LR, SVMs, decision lists, rule sets

Flipset

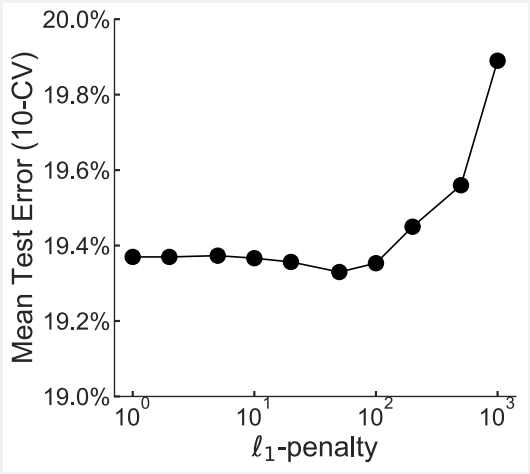
List of actions that a person can change to be approved for a loan

Input Variables to Change	Current Values		Required Values
most_recent_payment	$0	→	$790
months_paid_in_full_in_last_6_months	1	→	4
most_recent_payment	$0	→	$515
months_paid_in_full_in_last_6_months	1	→	2
most_recent_payment	$0	→	$500
months_paid_in_full_in_last_6_months	1	→	2
months_with_low_spending_in_last_6_months	6	→	5

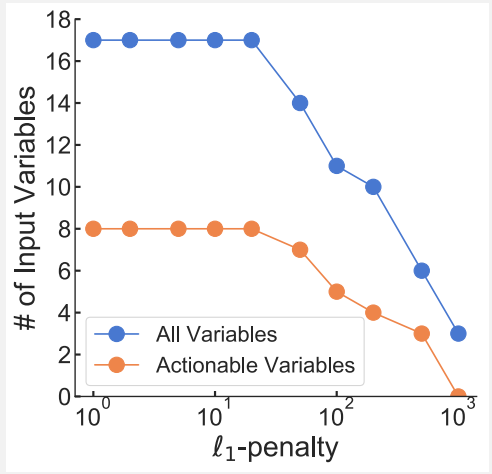
Recourse Audit

Measure feasibility and difficulty of recourse in a population of interest

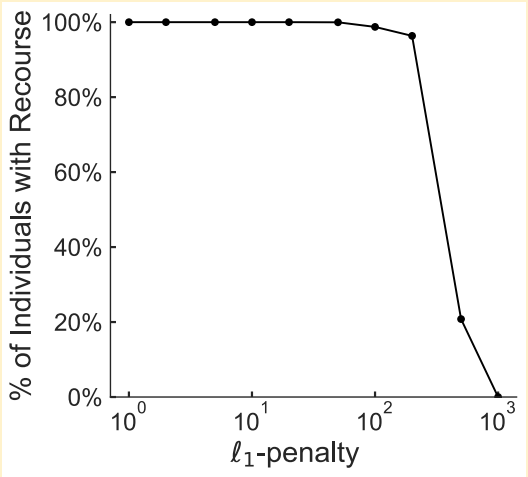
Test Error



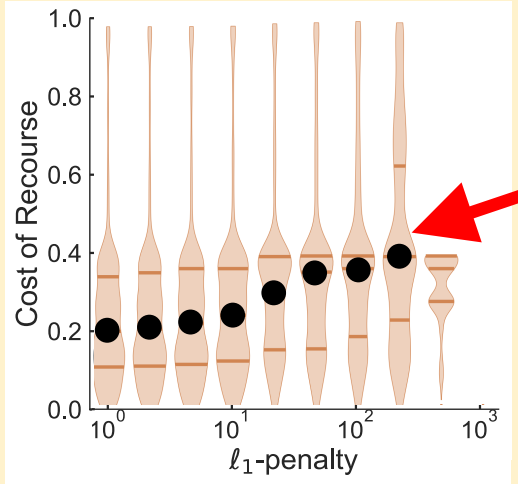
of Input Variables



Feasibility of Recourse



Cost of Recourse

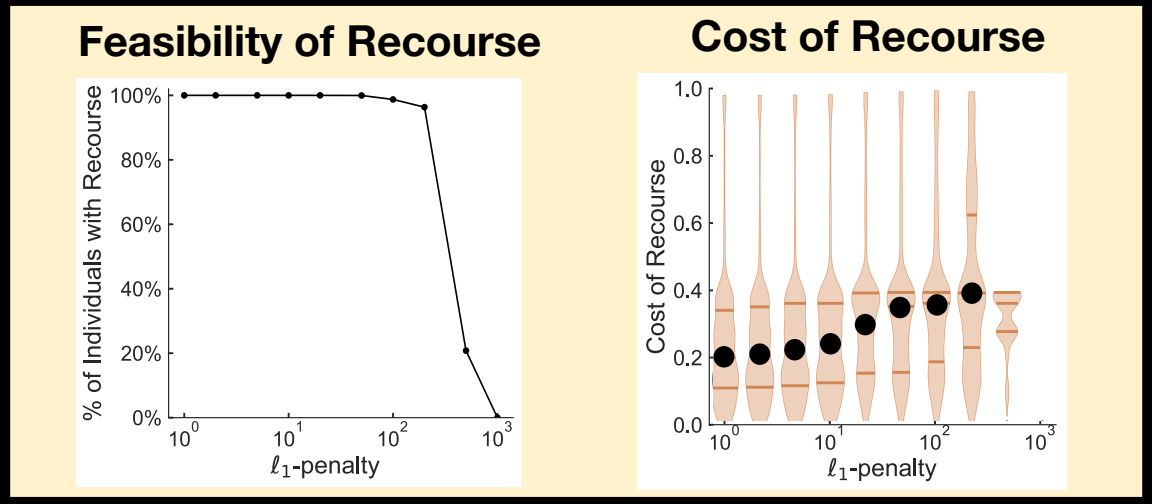


cost = 0.4

⇔

need at least a
40 percentile shift
in any variable
to flip prediction

Recourse Audits



- Model Development
- Model Procurement
- Algorithmic Impact Assessments

Flipsets

Variables to Change	Old		New
most_recent_payment	$0	→	$790
months_paid_in_full_in_last_6_months	1	→	4
most_recent_payment	$0	→	$515
months_paid_in_full_in_last_6_months	1	→	2
most_recent_payment	$0	→	$500
months_paid_in_full_in_last_6_months	1	→	2
months_with_low_spending_in_last_6_months	6	→	5

- Informing Consumers
- Testing in Deployment

1. When should we care about recourse?

2. Why models may not provide recourse

3. Tools to check recourse for linear classifiers

4. Lessons for consumer protection

Lending

Some lenders are judging you on much more than finances

By JAMES RUFUS KOREN
DEC 19, 2015 | 10:00 AM

Insurance

THE WALL STREET JOURNAL.

New York Insurers Can Evaluate Your Social Media Use—If They Can Prove Why It's Needed

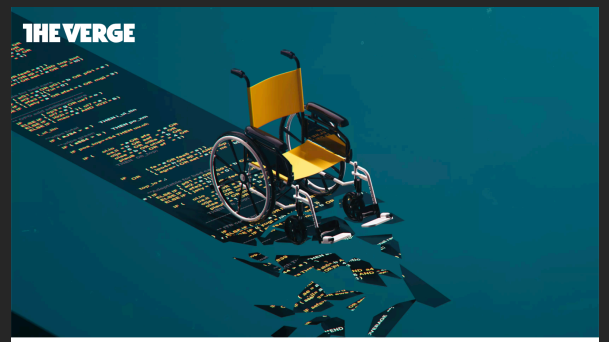
New guidance applies to companies operating in New York, but industry consultants say it could have an impact beyond the state's borders

Hiring

How an algorithm may decide your career

Getting a job means getting past the computer

Public Services



SCIENCE

WHAT HAPPENS WHEN AN ALGORITHM CUTS YOUR HEALTH CARE

By Colin Lecher | @colinlecher | Mar 21, 2018, 9:00am EDT

POLICY | US & WORLD | TECH

A new bill would force companies to check their algorithms for bias

21

By Adi Robertson | @thedextriarchy | Apr 10, 2019, 3:52pm EDT

SHARE

116TH CONGRESS
1ST SESSION

S. _____

To direct the Federal Trade Commission to require entities that use, store, or share personal information to conduct automated decision system impact assessments and data protection impact assessments.

IN THE SENATE OF THE UNITED STATES

Mr. WYDEN (for himself and Mr. BOOKER) introduced the following bill; which was read twice and referred to the Committee on _____

A BILL

To direct the Federal Trade Commission to require entities that use, store, or share personal information to conduct automated decision system impact assessments and data protection impact assessments.

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

SECTION 1. SHORT TITLE.

This Act may be cited as the "Algorithmic Accountability Act of 2019".

SEC. 2. DEFINITIONS.

In this Act:

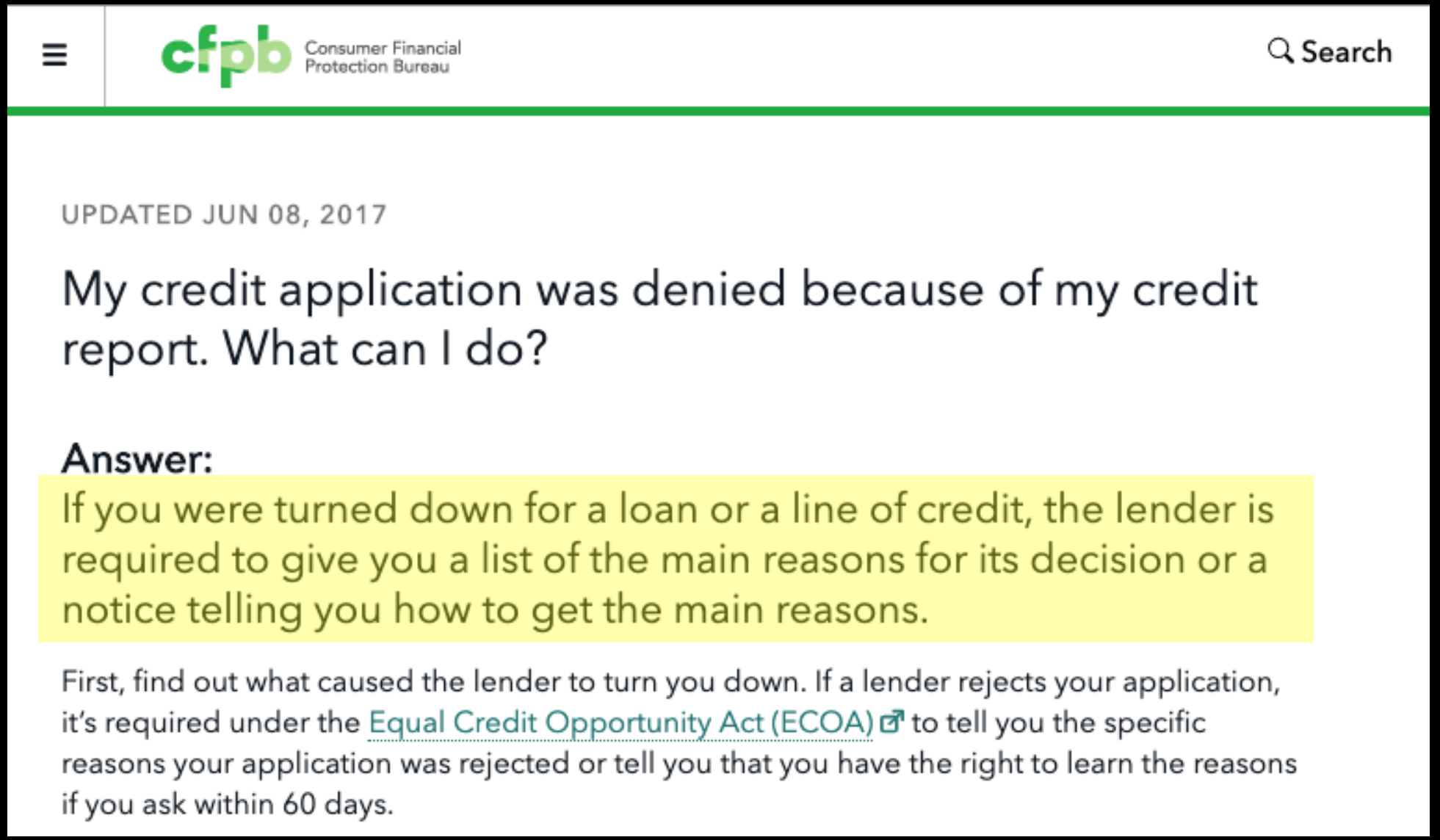
Enslaving the Algorithm: From a "Right to an Explanation" to a "Right to Better Decisions"?

Lilian Edwards, University of Strathclyde [l.edwards@strath.ac.uk]
Michael Veale, University College London [m.veale@ucl.ac.uk]

Published in
IEEE Security & Privacy (2018) 16(3), 46–54, doi:10.1109/MSP.2018.2701152

As concerns about unfairness and discrimination in "black box" machine learning systems rise, a legal "right to an explanation" has emerged as a compellingly attractive approach for challenge and redress. We outline recent debates on the limited provisions in European data protection law, and introduce and analyze newer explanation rights in French administrative law and the draft modernized Council of Europe Convention 108. While individual rights can be useful, in privacy law they have historically unreasonably burdened the average data subject. "Meaningful information" about algorithmic logics is more technically possible than commonly thought, but this exacerbates a new "transparency fallacy"—an illusion of remedy rather than anything substantively helpful. While rights-based approaches deserve a firm place in the toolbox, other forms of governance, such as impact assessments, "soft law," judicial review, and model repositories deserve more attention, alongside catalyzing agencies acting for users to control algorithmic system design.

40 Years of a "Right to an Explanation"



cfpb Consumer Financial Protection Bureau

Search

UPDATED JUN 08, 2017

My credit application was denied because of my credit report. What can I do?

Answer:

If you were turned down for a loan or a line of credit, the lender is required to give you a list of the main reasons for its decision or a notice telling you how to get the main reasons.

First, find out what caused the lender to turn you down. If a lender rejects your application, it's required under the Equal Credit Opportunity Act (ECOA) to tell you the specific reasons your application was rejected or tell you that you have the right to learn the reasons if you ask within 60 days.

What we Knew Back in the 1980s

MEETING THE EQUAL CREDIT OPPORTUNITY ACT'S SPECIFICITY REQUIREMENT: JUDGMENTAL AND STATISTICAL SCORING SYSTEMS

WINNIE F. TAYLOR*

INTRODUCTION

Consumer credit has become an accepted fact of American life. It continues to grow at a phenomenal rate as more and more buyers seek to improve their standard of living by utilizing various financing arrangements. Virtually all home purchases involve some form of mortgage agreement[1] and approximately two-thirds of all consumer automobile purchases are made on an installment payment basis. In addition, many large department stores report that at least half of their business depends on their closed-end credit plans.[2] Total installment credit has risen 68% in the last five years, with consumer installment debt rising by a record $44 billion in 1978.[3]

Americans who are constantly encouraged to become more dependent on credit need to be reminded that credit is available to them as a privilege, not as a legal right. Everyone who wants or needs credit cannot obtain it; each creditor devises its own method of separating those who will receive credit from those who will

ARTICLE

THE EQUAL CREDIT OPPORTUNITY ACT: A FUNCTIONAL FAILURE

JOHN H. MATHESON*

The Equal Credit Opportunity Act was enacted in 1974 as (1) a consumer protection statute designed to provide accurate information to and about consumers involved in credit transactions, and (2) an antidiscrimination statute designed to shield protected classes of consumers from discrimination in the granting of credit. The Federal Reserve Board promulgated regulations to further these statutory goals. Congress intended that the Act would be enforced through both private litigation and public compliance programs. Few private lawsuits have been brought under the Act, however, and public enforcement efforts have neither checked credit discrimination nor halted perpetuation of prior discrimination.

Professor Matheson believes that courts, government enforcement agencies, and consumers should focus on substantive (rather than procedural) violations of the Act and its implementing regulations. The Act should be amended to allow for a minimum damage recovery for successful plaintiffs. The definition of "adverse action" in the regulations should be amended to acknowledge that credit granted on different terms than those requested by an applicant may indicate illegal discrimination. Detailed statistical information must be kept by credit-granting institutions and made available to private litigants and government enforcement agencies to assist them in identifying and eliminating credit discrimination. Professor Matheson believes that these changes will help create a statutory and regulatory framework that will promote better compliance by creditors with the Act's provisions and enhance enforcement efforts by both private parties and public agencies.

Predictions don't have "Principle Reasons"

HYPOTHETICAL CREDIT SCORING SYSTEM

Applicant Characteristics	Allotted Points
Home Phone	
Yes	36
No	0
Own or Rent	
Own	34
Rent	0
Other Finance Company Debt	
Yes	−12
No	0
Bank Credit Card	
Yes	29
No	0
Applicant Occupation	
Professional and Officials	27
Technical and Managers	5
Proprietor	−3
Clerical and Sales	12
Craftsman and Nonfarm-laborer	0
Foreman and Operative	26
Service Worker	14
Farm Worker	3
Checking or Savings Account	
Neither	0
Either	13
Both	19
Applicant Age	
30 or less	6
30+ to 40	11
40+ to 50	8
Over 50	16
Years on Job	
5 or less	0
5+ to 15	6
Over 15	18

PRINCIPAL REASON(S) FOR ADVERSE ACTION CONCERNING CREDIT:

[ ] Credit application incomplete
[ ] Insufficient credit references
[ ] Unable to verify credit references
[ ] Temporary or irregular employment
[ ] Unable to verify employment
[ ] Length of employment
[ ] Insufficient income
[ ] Excessive obligations
[ ] Unable to verify income
[ ] Inadequate collateral
[ ] We do not grant credit to any applicant on the terms and conditions you request
[ ] Too short a period of residence
[ ] Temporary residence
[ ] Unable to verify residence
[ ] No credit file
[ ] Insufficient credit file
[ ] Delinquent credit obligations
[ ] Garnishment, attachment, foreclosure, repossession, or suit
[ ] Bankruptcy
[ ] Other specify: _____

DISCLOSURE OF USE OF INFORMATION OBTAINED FROM AN OUTSIDE SOURCE

[ ] Disclosure inapplicable
[ ] Information obtained in a report from a consumer reporting agency

Name: _____

Address: _____

Telephone Number: _____

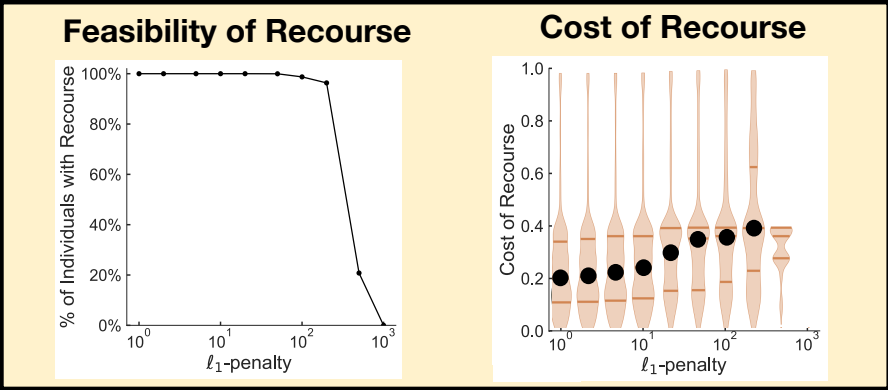
The applicant, rejected under a creditor's scoring system, received a statement of reasons which ... showed the applicant scored lowest in the "time on the job" and "credit references" categories.

Dissatisfied with this response, the rejected applicant wrote for further clarification of the reasons for denial, and for the creditor's minimum requirement for time on the job and the number and type of credit references required.

The creditor responded that the information requested could not be given because there were no minimum standards, and apologetically explained that because different point values are assigned to each factor considered, concrete standards for any one factor could not be established.

Lessons in Designing Effective Consumer Protection

Recourse Audits



Flipsets

Variables to Change	Old		New
most_recent_payment	$0	→	$790
months_paid_in_full_in_last_6_months	1	→	4
most_recent_payment	$0	→	$515
months_paid_in_full_in_last_6_months	1	→	2
most_recent_payment	$0	→	$500
months_paid_in_full_in_last_6_months	1	→	2
months_with_low_spending_in_last_6_months	6	→	5

1. Protect against specific failure modes

2. Specify exact tools and testing procedures
   - Firms comply with minimum requirements
   - Minimize reliance on "expert opinion"

3. Provide multiple avenues to prevent harm
   - Legislation is harder to pass when there is only one way to regulate

Thank you!

Paper

Actionable Recourse in Linear Classification

Berk Ustun, Alexander Spangher, Yang Liu.

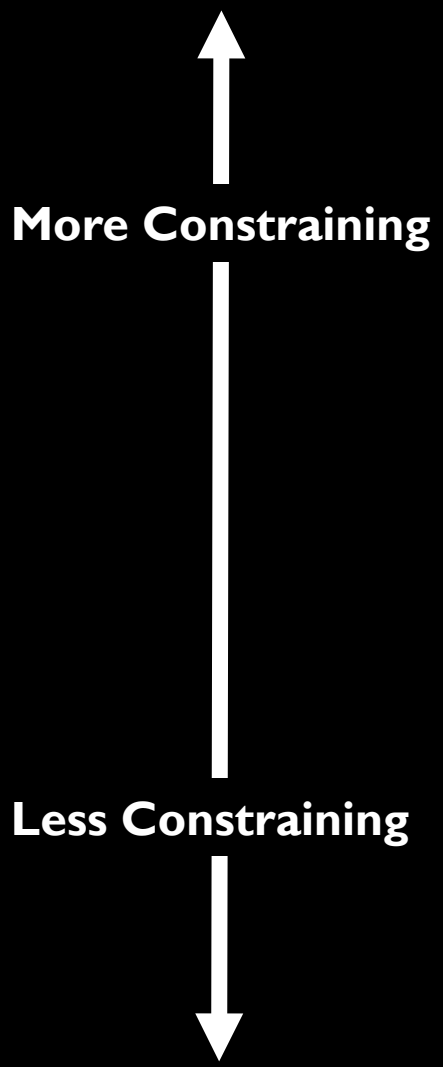
ACM Conference on Fairness, Accountability and Transparency, 2019

Software

actionable-recourse https://github.com/ustunb/actionable-recourse

"Checklist Reason"	"Explanatory Statements"
Credit application incomplete	You failed to list credit references
Insufficient credit references	We require a minimum of three references
Length of employment	We require six (6) months continuous emplyment with one employer
Insufficient income	We require a minimum income of $10,000
Too short a period of residence	We require a minimum of four (4) months at the same residence
Insufficient credit file	We require a minimum of three positive references; your file contains only one

Hierarchy of Legal Norms



Rules: Once a rule has been interpreted and the facts have been found, then the application of the rule to the facts decides the issue to which it is relevant.

Standards: Guide decisions but provide a greater range of choice. Standards define a set of mandatory considerations that are **exhaustive** for adjudication or policy making, .

Principles: Mandatory considerations for judges. Principles identify **some** considerations, allowing one to consider other factors in the decision.

Catalogs: A list of things that are within the legal norm along with a sweepings clause, e.g., "and other things like this."

Discretion: The most flexible option. Relevant legal norm may simply be a secondary rule that grants discretion to an official (frequently a judge).

Source: https://lsolum.typepad.com/legaltheory/2017/12/

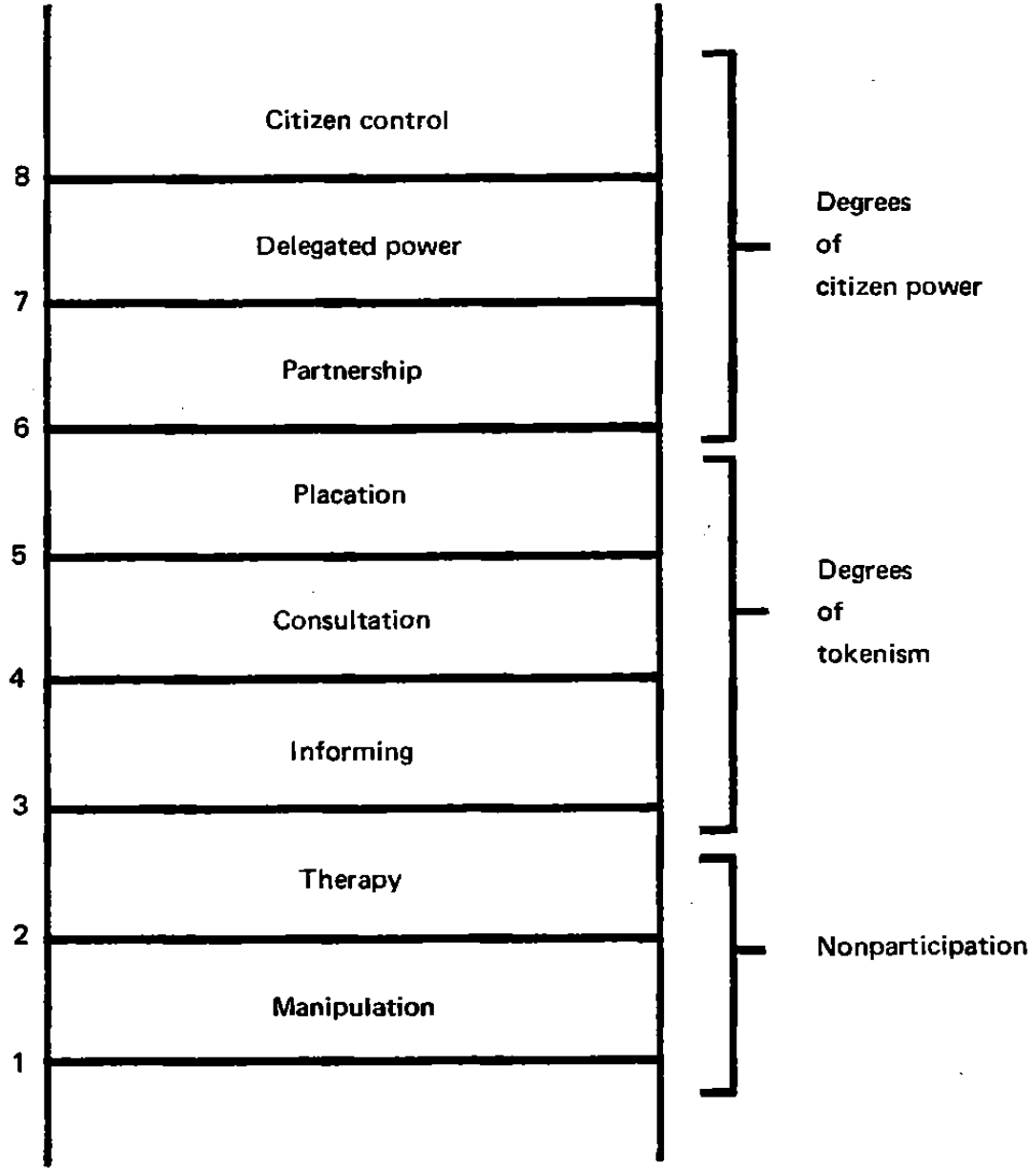


FIGURE 2 *Eight Rungs on a Ladder of Citizen Participation*

Many of the most chilling stories of algorithmic bias don't involve meaningful explanations or a meaningful appeals process

Rachel Thomas

What HBR Gets Wrong About Algorithms and Bias

When algorithms go wrong we need more power to fight back, say AI researchers

The public doesn't have the tools to hold algorithms accountable

By James Vincent | Dec 8, 2018, 2:00pm EST

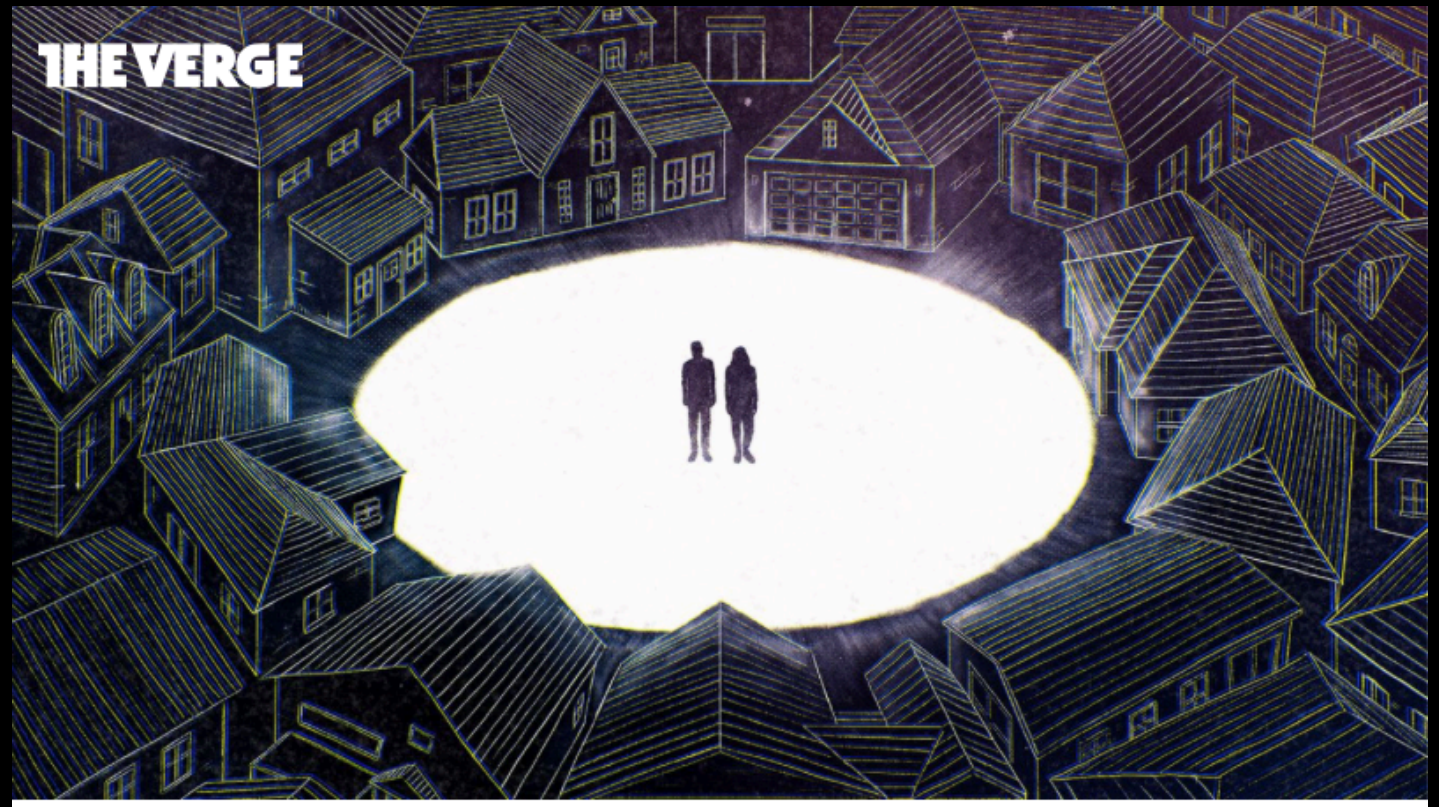
guidelines	The European Commission's High-Level Expert Group on Artificial Intelligence	The Malicious Use of Artificial Intelligence	AI4People	The Asilomar AI Principles	AI Now 2016 Report	AI Now 2017 Report	AI Now 2018 Report	Principles for Accountable Algorithms and a Social Impact Statement for Algorithms	Montréal Declaration for Responsible Development of Artificial Intelligence	Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems	ITI AI Policy Principles	Microsoft AI principles	Artificial Intelligence at Google	Everyday Ethics for Artificial Intelligence	Partnership on AI
key issue	explanations about the idea of a trustworthy AI	analysis of abuse scenarios of AI	meta-analysis about principles for the beneficial use of AI	large collection of different principles	statements on social implications of AI	statements on social implications of AI	statements on social implications of AI	principles of the FAT ML community	code of ethics released by the Université de Montréal	detailed description of ethical aspects in the context of AI	brief guideline about basic ethical principles	short list of keywords for the ethical use of AI	several short principles for the ethical use of AI	IBM's short list of keywords for the ethical use of AI	principles of an association between several industry leaders
privacy protection	x	x	x	x	x	x	x		x	x	x	x	x	x	x
accountability	x	x	x	x	x		x	x	x	x	x	x		x	x
fairness, non-discrimination, justice	x		x	x	x	x	x	x	x	x		x	x	x	x
transparency, openness	x	x		x	x	x	x	x	x	x		x			
safety, cybersecurity	x	x	x	x					x	x	x	x	x		x
common good, sustainability			x	x	x	x			x	x			x	x	x
explainability, interpretabiliy	x		x	x		x		x	x		x			x	
human oversight, control, auditing	x		x	x		x	x	x			x		x		
dual-use problem, military, AI arms race		x		x		x			x	x			x		
solidarity, inclusion, social cohesion			x		x		x		x			x			x
science-policy link		x	x	x	x		x								
field-specific deliberations (health, military, mobility etc.)		x			x	x	x			x					
diversity in the field of AI					x	x	x		x		x				
notes on technical implementations	yes, but very few and superficial	yes, relatively comprehensive	none	none	none	none	none	none	none	yes, but very few and superficial	none	none	none	none	none

HOW AN ALGORITHM KICKS SMALL BUSINESSES OUT OF THE FOOD STAMPS PROGRAM ON DUBIOUS FRAUD CHARGES

H. Claire Brown

October 8 2018, 9:00 a.m.

10



POLICY

AUTOMATED BACKGROUND CHECKS ARE DECIDING WHO'S FIT FOR A HOME

But advocates say algorithms can't capture the complexity of criminal records

By Colin Lecher | @colinlecher | Feb 1, 2019, 8:00am EST

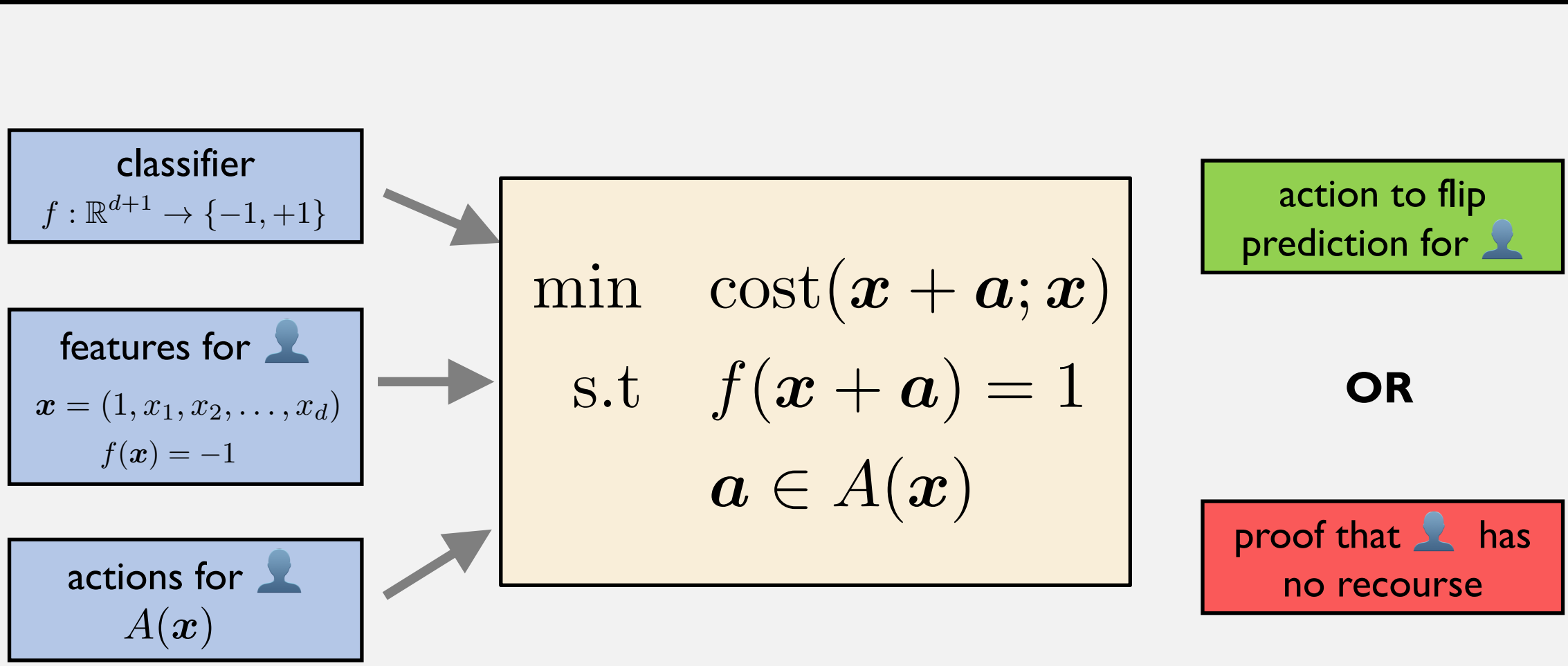
Illustration by Alex Castro

THE WALL STREET JOURNAL.

On Hold for 45 Minutes? It Might Be Your Secret Customer Score

Retailers, wireless carriers and others crunch data to determine what shoppers are worth for the long term—and how well to treat them

Routine to Check Recourse for 1 Person (👤)



Integer Programming Formulation

$$\min \quad \text{cost} = \sum_{j=1}^{d} \sum_{k=1}^{m_j} c_{jk} u_{jk}$$

$$\text{s.t.} \quad 0 \leq \sum_{j=1}^{d} w_j (x_j + a_j)$$

$$a_j = \sum_{k=1}^{m_j} a_{jk} u_{jk} \qquad \forall j$$

$$1 = \sum_{k=1}^{m_j} u_{jk} \qquad \forall j$$

$$u_{jk} \in \{0, 1\} \qquad \forall j, k$$

- Handles all discrete data types
  - binary, ordinal, categorical,
- Discretization guarantees
  - feasibility remains the same
  - costs have controllable discretization error
- Supports diverse cost function
  - use to measure difficulty of actions
- Very fast
  - <1 second

Recourse Audit

Measure feasibility / difficulty of recourse in a population of interest

Input: $\{\boldsymbol{x}_i\}_{i=1}^n$ feature vectors from deployment population

for $i : f(\boldsymbol{x}_i) = -1$ **do**

$\mathsf{IP} \leftarrow \text{RecourseIP}(f, \boldsymbol{x}_i, A(\boldsymbol{x}_i))$

$r_i^* \leftarrow 1$ if IP is feasible else 0

$c_i^* \leftarrow$ optimal cost of IP if IP is feasible else ∞

Output: $\{c_i\}_{i=1}^n$ cost of each sample

Output: $\{r_i\}_{i=1}^n$ feasible of each sample

$$\text{cost}(\boldsymbol{x} + \boldsymbol{a}; \boldsymbol{x}) = \max_{j \in [d]} |Q_j(x_j + a_j) - Q_j(x_j)|$$

Building Flipsets

IP $\leftarrow$ RecourseIP$(f, \boldsymbol{x}, A(\boldsymbol{x}))$ — setup recourse IP

$\mathcal{A} \leftarrow \{\}$ — collection of actions that will flip prediction

repeat

$\boldsymbol{a}^* \leftarrow$ optimal solution to IP

$\mathcal{A} \leftarrow \mathcal{A} \cup \{\boldsymbol{a}^*\}$ — *add $\boldsymbol{a}^*$ to set of optimal actions*

$z_j \leftarrow 1[a_j^* \neq 0]$ — *1 if feature j is altered by $\boldsymbol{a}^*$*

$Z \leftarrow \{j : a_j^* \neq 0\}$ — *indices of features altered by $\boldsymbol{a}^*$*

add constraint to IP to remove actions that alter the same features:

$$\sum_{j \notin Z} z_j + \sum_{j \in Z} (1 - z_j) \leq d - 1.$$

until $\mathcal{A}$ contains enough items **or** IP is infeasible

$$\text{cost}(\boldsymbol{x} + \boldsymbol{a}; \boldsymbol{x}) = \sum_{j:a_j \neq 0} \log\left(\frac{1 - Q_j(x_j + a_j)}{1 - Q_j(x_j)}\right)$$