

WHAT MACHINES SHOULDN'T DO

A Necessary Condition for Meaningful Human Control



WHAT MACHINES DO

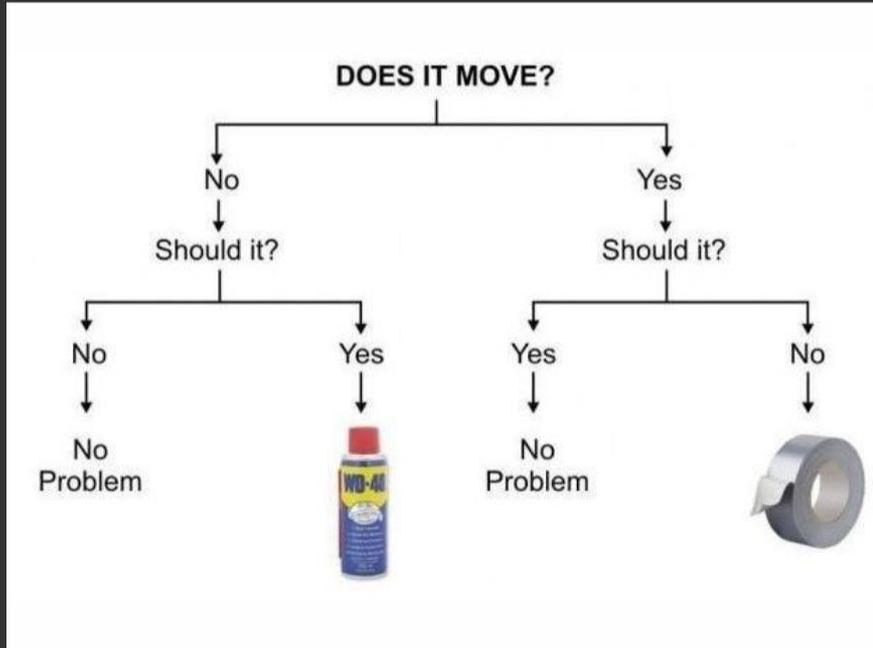
DELEGATING DECISIONS TO MACHINES

WHEN HAVE WE DELEGATED A DECISION TO A MACHINE?



- Reach an output through considerations and weighting NOT given by humans

DELEGATED DECISIONS



Automated



Delegated



PredPol

Predict Crime in Real Time™

PredPol provides targeted, real-time crime prediction designed for and successfully tested by officers in the field.



CLINICAL RECORDS

IMAGING DATA

(GEN)-OMICS INFO

DIAGNOSTIC
TRENDS

SILO.AI

All rise.
 Honorable AI
 judge presiding.

entefy

AI-Powered Fraud Detection

In the US last year, \$118 billion in credit card transactions were declined, but only \$9 billion of this fraud. AI has been able to reduce fraud by 25% without affecting the rate of false positives.



Fraudulent Behavior



Anomaly Detected



Case Prioritization



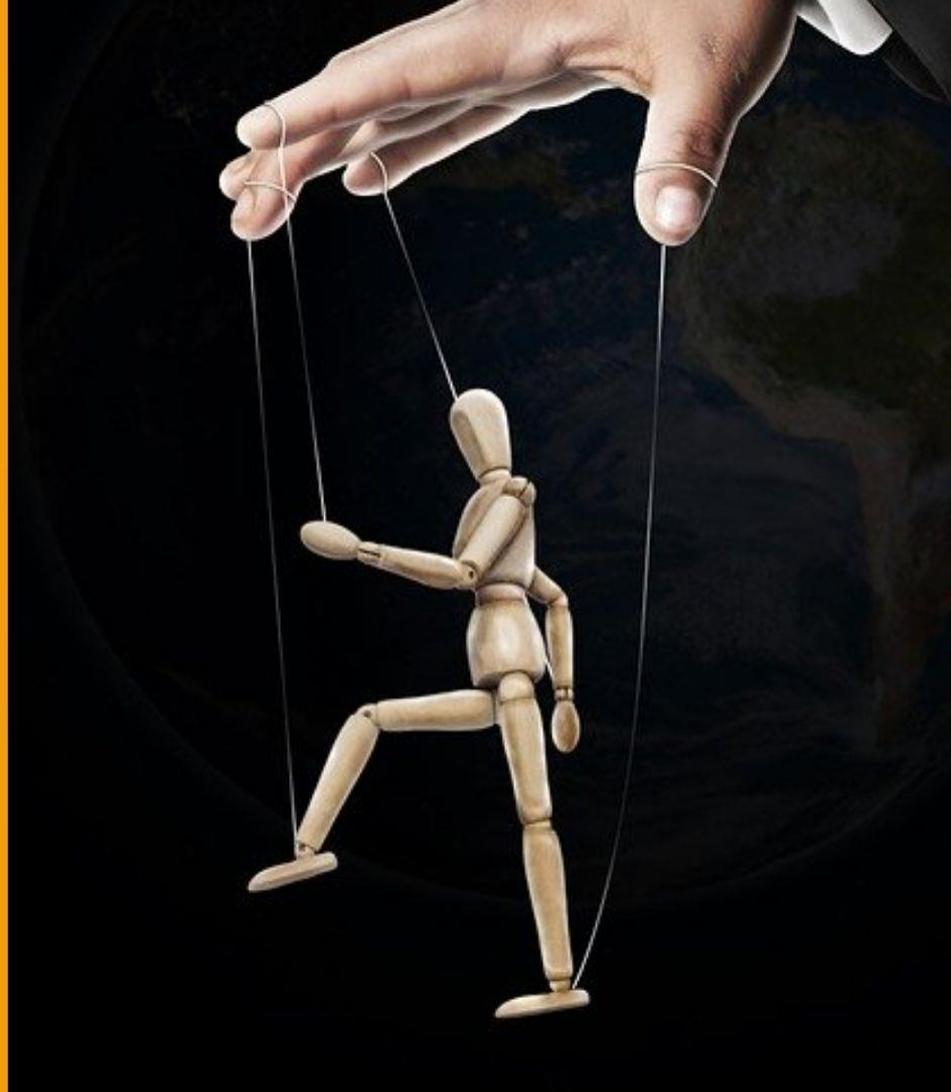
Style It

Take a picture of an item and create matching outfits.

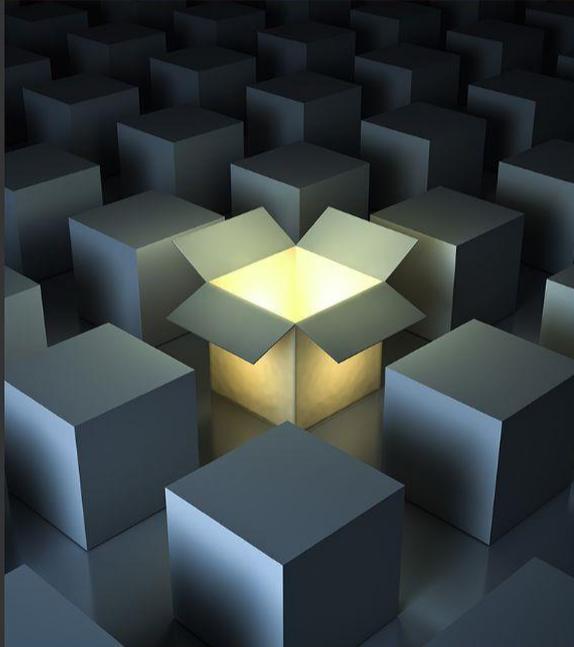
86 Matches

MEANINGFUL HUMAN CONTROL?

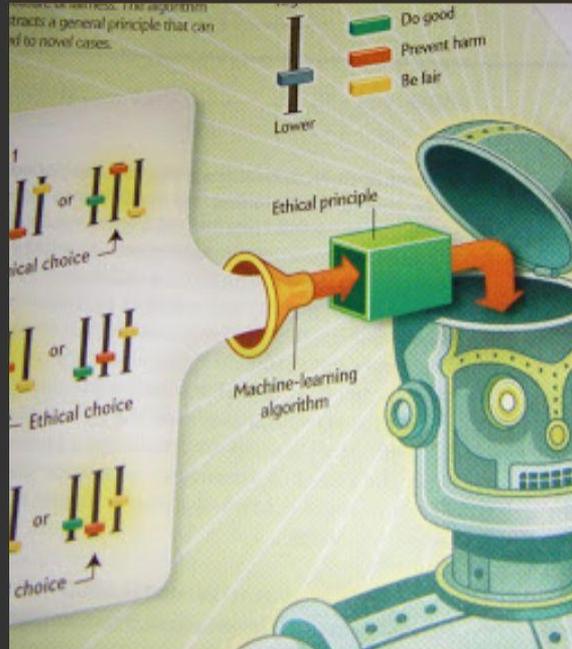
Tech & Human Centered



TECH CENTERED MHC



Explicability



Machine Ethics



Track & Trace

HUMAN CENTERED MHC



On the Loop



In the Loop



REARRANGING DECKCHAIRS

We've already hit the iceberg!

MHC IS ABOUT WHICH DECISIONS SHOULD BE DELEGATED TO MACHINES

Delegate Decision-making to Machine

Train algorithm

Organize Socio-Tech System

**MACHINES SHOULD
NOT HAVE
EVALUATIVE
OUTPUTS**



EVALUATIVE OUTPUTS



(a) Three samples in criminal ID photo set S_c .



(b) Three samples in non-criminal ID photo set S_n .

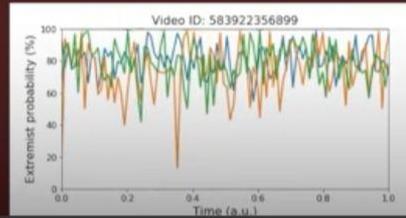
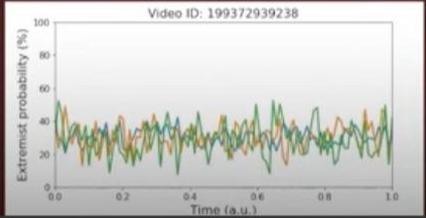
- “Criminal”
person + convicted of crime + bad
- “Suspicious”
person + loitering + bad
- “Beautiful”
person + + good
- “Propaganda”
picture + political message + bad
- “Fake News”
news + false + bad
- “Bad”, “Good”, “Wrong”, “Right”



تنظيم الدولة الإسلامية يعلن قيام الخلافة وتخصيب عبدالله إبراهيم السامرائي أبو بكر البغدادي خليفة



IS propaganda



(a) Three samples in criminal ID photo set S_c .



A.I. Beauty Recognition Technology



Microsoft

BUILDING AI

TO DETECT

FAKE NEWS



HIREVUE VIDEO INTELLIGENCE

GET THE BEST TALENT, FASTER

[SEE HOW](#)

MY
APPROACH
WOULD
BE...

CANDIDATE

FANTASTIC

GREAT

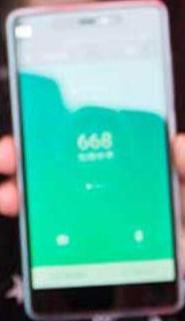
GOOD

OK

How To Hire From 4x The Colleges In 1/4 the Time. [Register Now >](#)

Reimagining Pre-hire Assessments

HireVue delivers custom assessments by blending the power of artificial intelligence with the science of I-O Psychology - all built within an easy to use video interview software platform. Watch the video to see how.



WHY?



Efficacy



Control

EFFICACY: UNKNOWN IN PRINCIPLE



- Every evaluative output is unverifiable
- Suspicious is a judgment
- Even if it turns out this man stole something, that does not mean he was 'suspicious'

GUN?



SUSPICIOUS?



- 1 year ago this kid would have justifiably been labeled 'suspicious'
- CONTEXTS CHANGE!
- Like, for example, a pandemic...

ETHICS: FUNDAMENTAL LACK OF CONTROL

“Good” or
“Bad”
Candidate

Evaluation

e.g. “Years of Experience”,
“Level of Education”, “Number
of Publications”, etc.

Considerations



DYNAMIC: HOW WE GROUND VALUES

What makes a good candidate today may not be what makes a good candidate tomorrow

The screenshot shows the HireVue website homepage. At the top left is the HireVue logo. To the right are navigation links: PRODUCTS, WHY HIREVUE, CUSTOMERS, RESOURCES, COMPANY, a red button labeled SEE A DEMO, and LOGIN. The main content area features a man's face split vertically. The left side is a real-life photo of a smiling man in a plaid shirt. The right side is a dark overlay with white line-art icons: a brain in the top half and a heart in the bottom half. A speech bubble on the right side of the face contains the text "MY APPROACH WOULD BE...". To the right of the face is a vertical bar chart representing a candidate's performance. The bar is filled with diagonal lines and is topped with a star icon and the word "CANDIDATE". To the right of the bar are five performance levels: FANTASTIC, GREAT, GOOD, and OK. A red banner at the bottom of the screenshot contains the text: "How To Hire From 4x The Colleges In 1/4 the Time. Register Now >".

HireVue

PRODUCTS • WHY HIREVUE • CUSTOMERS • RESOURCES • COMPANY • [SEE A DEMO](#) LOGIN

HIREVUE VIDEO INTELLIGENCE

GET THE BEST TALENT, FASTER

[SEE NOW](#)

MY APPROACH WOULD BE...

CANDIDATE

FANTASTIC

GREAT

GOOD

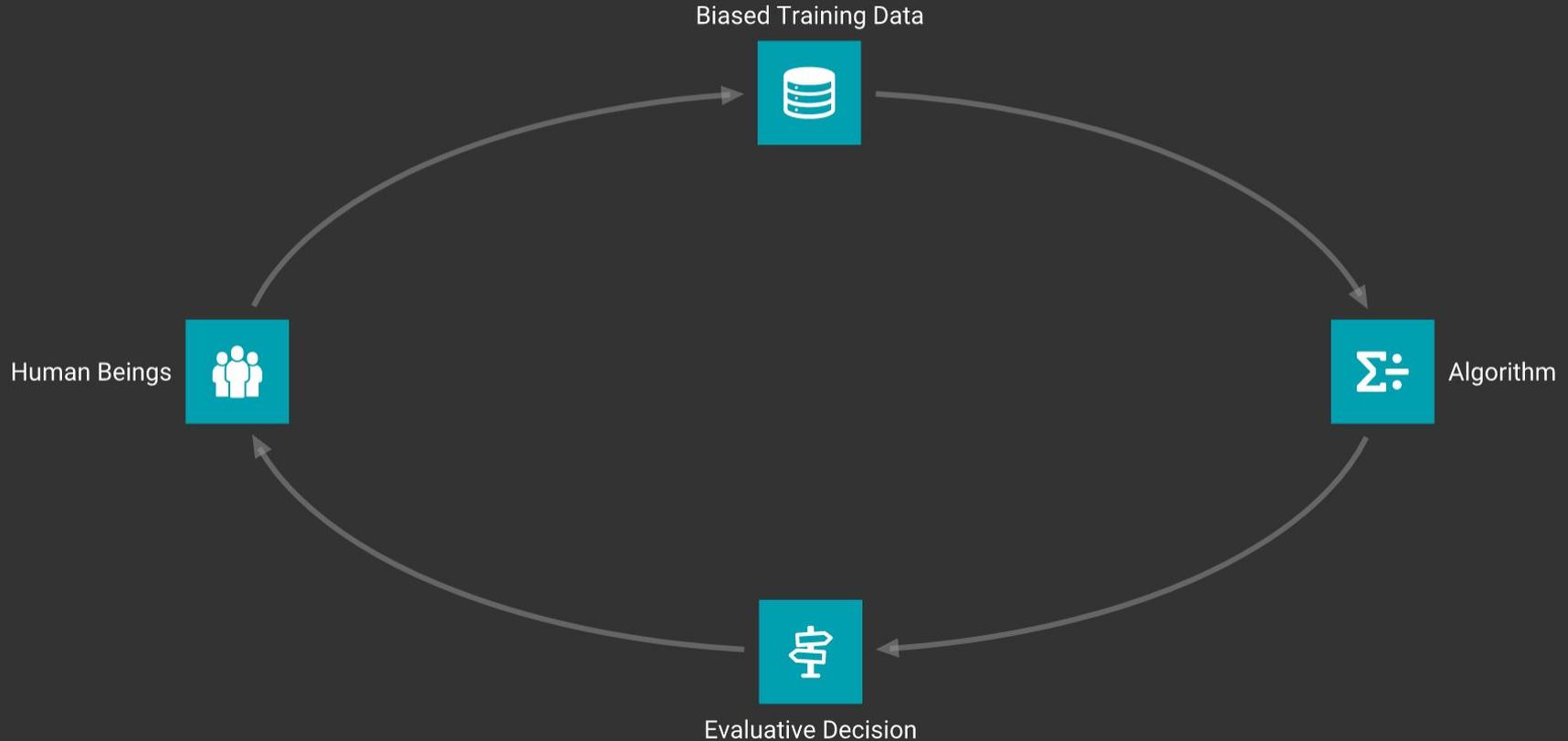
OK

How To Hire From 4x The Colleges In 1/4 the Time. [Register Now >](#)

Reimagining Pre-hire Assessments

HireVue delivers custom assessments by blending the power of artificial intelligence with the science of I-O Psychology - all built within an easy to use video interview software platform. Watch the video to see how.

VICIOUS MACHINE FEEDBACK LOOP



DYNAMIC: PEOPLE'S BEHAVIOR

THE VERGE

TECH ▾

SCIENCE ▾

ENTERTAINMENT ▾

MORE ▾



AD

REPORT \ TECH \ ARTIFICIAL INTELLIGENCE \

These students figured out their tests were graded by AI — and the easy way to cheat

"He's getting all 100s"

By [Monica Chin](#) | [@mcsquared96](#) | Sep 2, 2020, 10:05pm EDT



EVALUATIVE OUTPUTS



Aesthetic



Ethical/Moral



Emotional

AI IS NOT A SILVER BULLET

IT CANNOT EVALUATE BETTER THAN WE CAN



KEEP AI BORING

AI should identify descriptive features that
ground our evaluations



