AI AND ETHICS

For AI Builders Program 2022

Ethics for Data Science

Rachel Thomas, PhD
USF Center for Applied Data Ethics & fast.ai
@math_rachel

UNIVERSITY OF SAN FRANCISCO
CHANGE THE WORLD FROM HERE

Google ใช้ภาพถ่ายจากคนเชื้อสายแอฟริกันไร้บ้านในการเทรน AI ตรวจจับใบหน้า

Google exploited homeless black people to develop the Pixel 4's facial recognition AI

Russia Tests New Disinformation Tactics in Africa to Expand Influence

รัสเซียทดสอบเทคนิคการเผยแพร่ข่าวลวงในแอฟริกา

Amazon's facial recognition matched 28 members of Congress to criminal mugshots

AI ตรวจจับใบหน้าของ Amazon จับคู่ภาพถ่ายของ ส.ส. 28 คนกับผู้ต้องหา

Flawed Algorithms Are Grading Millions of Students' Essays

พบว่า AI ที่ใช้ตรวจเรียงความนักเรียนมากกว่าล้านฉบับมีข้อบกพร่อง

WHAT HAPPENS WHEN AN ALGORITHM CUTS YOUR HEALTH CARE

AI ตัดสินว่าควรตัดสิทธิรักษาพยาบาล ใคร เท่าไหร่

Indigenous elder slams 'hollow and tokenistic' consultation by Sidewalk Labs

BBC NEWS | ไทย

ทวิตเตอร์ระงับการใช้งาน 926 บัญชี ที่เชื่อว่าเป็นเครือข่ายปฏิบัติการ "ไอโอ" ของ ทบ.

9 ตุลาคม 2020

ฝ่ายความปลอดภัยของทวิตเตอร์หรือ @TwitterSafety ตรวจสอบพบว่าบัญชีผู้ใช้งานในไทยเกือบ 1,000 บัญชีมีความเชื่อมโยงกับปฏิบัติการข่าวสารหรือ "ไอโอ" (Information Operations: IO) ของกองทัพบก (ทบ.)

https://www.bbc.com/thai/thailand-54473564

WHAT HAPPENS WHEN AN ALGORITHM CUTS YOUR HEALTH CARE

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

ปัญหาสืบเนื่องมาจากอายุขัยเฉลี่ยของคนเพิ่มขึ้น รัฐต้องหาวิธีจำกัดค่าใช้จ่าย

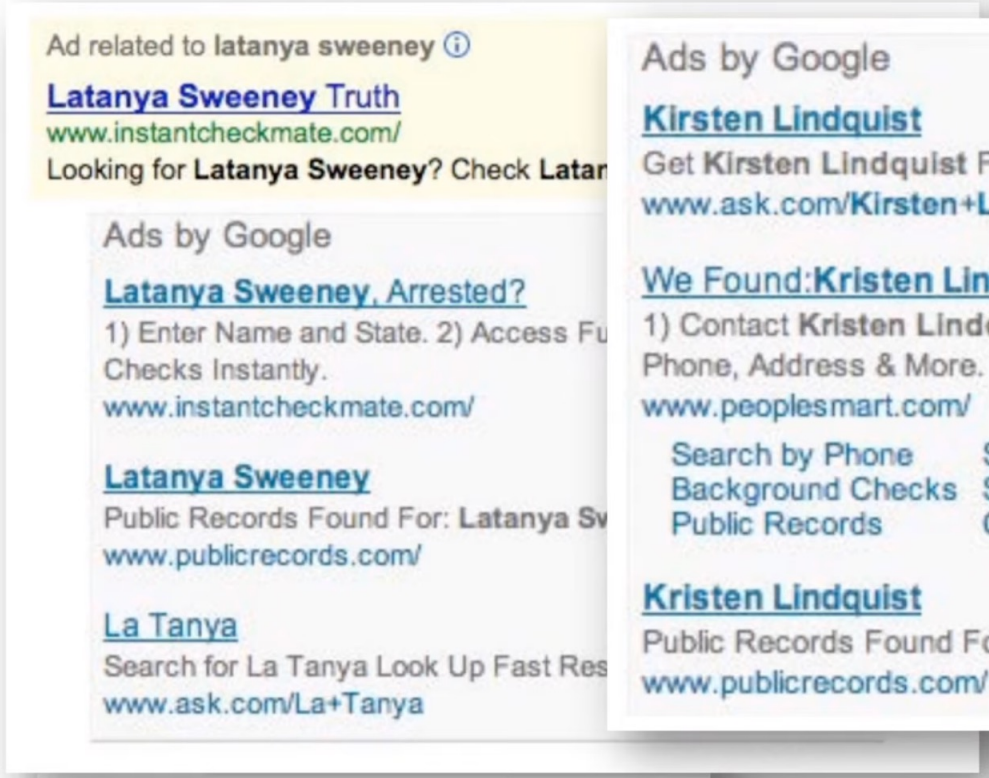
ดอบบี้มีภาวะสมองพิการและได้รับการดูแลจากรัฐภายใต้โครงการ Medicaid อัลกอริธึมของรัฐอาร์คันซอตัดสินให้เธอได้รับการดูแลเพียง 32 ชั่วโมงต่อสัปดาห์ซึ่งไม่เพียงพอ

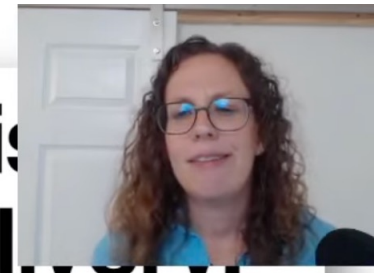
Common issue: Systems implemented with no way to identify & address mistakes

What HBR Gets Wrong About Algorithms and Bias

Written: 07 Aug 2018 by *Rachel Thomas*

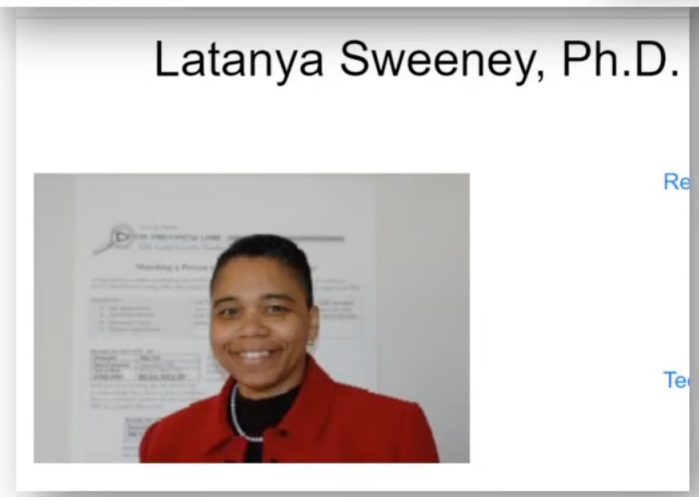
ปัญหาพื้นฐานของการใช้อัลกอริธึมมาช่วยตัดสินใจ: ระบบถูกสร้างขึ้นโดยไม่ได้ออกแบบให้สามารถรู้ว่ามีจุดบกพร่องอย่างไร





Racism is Pois
Online Ad Del
Says Harvard Professor

Search ชื่อคนดำ suggestion มีแนวโน้มที่จะให้ผลในทางลบมากกว่าชื่อแบบคนขาว (เช่น ad แนะนำให้เช็คประวัติอาชญากรรม)



Facebook's ad system seems to discriminate by race and gender

New research shows that Facebook's ad-distribution software is disturbingly biased

In the conce
camps, IBM'
Jews was 8.
Gypsies was 12. General executions were coded as 4, death in the gas chambers as 6.

บทบาทของข้อมูลกับการฆ่าล้างเผ่าพันธุ์ในช่วงสงครามโลกครั้งที่ 2

- การค้นหาคนยิวจากประวัติและการทำสำมะโน เช่น คนยิวไม่มีประวัติรับบัพติศมาที่โบสถ์
- IBM ช่วยพรรคนาซีพัฒนาคอมพิวเตอร์ที่ใช้ในค่ายกักกันและกระบวนการฆ่าล้างเผ่าพันธุ์

The Toxic Potential of YouTube's Feedback Loop

On YouTube's Digital Playground, an Open Gate for Pedophiles

YouTube Executives Ignored Warnings, Letting Toxic Videos Run Rampant

Feedback loops can occur when **your model is controlling the next round of data you get**. The data that is returned quickly becomes flawed by the software itself.

Feedback loops เกิดขึ้นเมื่อโมเดลของเราควบคุมการได้รับข้อมูลในอนาคต เกิดเป็นวงจรการได้รับข้อมูลที่มีอคติ

ตัวอย่าง feedback loops
ในชีวิตจริง

Monopoly Classic

คนที่รวยก็จะยิ่งรวยขึ้น
คนที่จนก็จะยิ่งจนลง
โดยไม่เกี่ยวกับความสามารถอะไรเลย

หากเราพบว่างานของเราเป็นส่วนหนึ่ง
ที่ส่งผลเสียต่อสังคมจะทำอย่างไร
เราจะรู้ได้อย่างไร
เราจะยอมรับความผิดพลาดหรือไม่
เราจะทำอย่างไรเพื่อป้องกันไม่ให้เกิด
สิ่งนี้

ผลที่ไม่คาดคิด

งานของเราอาจถูกใช้โดย / ไปเพื่อ

- คนที่ต้องการทำร้ายผู้อื่น
- สนับสนุนรัฐบาลเผด็จการ
- โฆษณาชวนเชื่อหรือการสร้างข่าวลวง

ETHICS

- The Discipline working with what is good and bad; a set or moral principles
- พจนานุกรมฉบับราชบัณฑิตยสถาน: ปรัชญาสาขาหนึ่งที่ว่าด้วยความประพฤติและการครองชีวิตว่าอะไรดี อะไรถูก อะไรผิด หรืออะไรควร อะไรไม่ควร
- **จริยศาสตร์** ไม่ใช่ ศาสนา กฎหมาย บรรทัดฐานทางสังคม
- **จริยศาสตร์** ไม่ใช่ กฎตายตัว
- **จริยศาสตร์** คือ มาตรฐานของความถูกและความผิดที่มนุษย์ควรหรือไม่ควรปฏิบัติที่พัฒนาขึ้นจากการให้เหตุผล
- **จริยศาสตร์** คือ การศึกษาการพัฒนามาตรฐานเหล่านั้น

จริยศาสตร์ควรถูกสอนอย่างไร

- เป็นวิชาของตัวเอง หรือ เป็นส่วนหนึ่งของทุกวิชา
- ใครควรเป็นผู้สอน นักวิทยาศาสตร์คอมพิวเตอร์ นักปรัชญา นักสังคมศาสตร์
- ควรครอบคลุมหัวข้อใดบ้าง

วิชาจริยศาสตร์ในคณะต่างๆ

Table 1: The number of classes for which the course home department, instructor home department, and instructor degree matches each discipline, sorted by course home most to least.

Discipline	Course Home	Instructor Home	Degree
Computer Science	67	61	31
Info Science	62	49	36
Philosophy	26	21	40
Communication	23	18	19
Other Non Tech	18	18	20
Sci & Tech Studies	13	6	13
Engineering	12	10	7
Law	11	13	22
Other Tech	9	8	7
Math	7	3	6
Business	3	4	1

Table 2: The number of courses that had content for each listed topic, out of 115 total courses, organized from most popular to least popular topic.

Topic	Courses
Law & policy	66
Privacy & surveillance	61
Philosophy	61
Inequality, justice & human rights	59
AI & algorithms	55
Social & environmental impact	50
Civic responsibility & misinformation	32
AI & robots	27
Business & economics	27
Professional ethics	25
Work & labor	23
Design	20
Cybersecurity	19
Research ethics	16
Medical/health	12

Table 3: The number of courses that had each type of learning outcome, organized from most common to most common outcome.

Outcome	Courses
Critique	71
Spot issues	36
Make arguments	26
Improve communication	26
See multiple perspectives	23
Create solutions	21
Consider consequences	18
Apply rules	10

วิชาจริยศาสตร์พูดถึงสาขาอื่นๆ อย่างไรบ้าง

วิชาจริยศาสตร์มีวัตถุประสงค์เพื่ออะไรบ้าง

What Do We Teach When We Teach Tech Ethics? A Syllabi Analysis

Casey Fiesler
casey.fiesler@colorado.edu
University of Colorado Boulder
Boulder, CO

Natalie Garrett
natalie.garrett@colorado.edu
University of Colorado Boulder
Boulder, CO

Nathan Beard
nbeard@umd.edu
University of Maryland
College Park, MD

RESOURCES & ACCOUNTABILITY

WHAT HAPPENS WHEN AN ALGORITHM CUTS YOUR HEALTH CARE

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

ปัญหาสืบเนื่องมาจากอายุขัยเฉลี่ยของคนเพิ่มขึ้น รัฐต้องหาวิธีจำกัดค่าใช้จ่าย

ดอบบี้มีภาวะสมองพิการและได้รับการดูแลจากรัฐภายใต้โครงการ Medicaid อัลกอริธึมของรัฐอาร์คันซอตัดสินให้เธอได้รับการดูแลเพียง 32 ชั่วโมงต่อสัปดาห์ซึ่งไม่เพียงพอ

Common issue: Systems implemented with no way to identify & address mistakes

What HBR Gets Wrong About Algorithms and Bias

Written: 07 Aug 2018 by *Rachel Thomas*

ปัญหาพื้นฐานของการใช้อัลกอริธึมมาช่วยตัดสินใจ: ระบบถูกสร้างขึ้นโดยไม่ได้ออกแบบให้สามารถรู้ว่ามีจุดบกพร่องอย่างไร

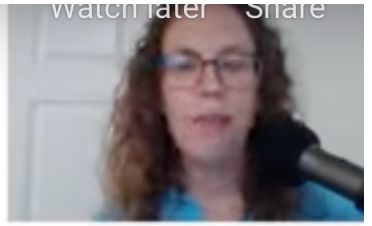
Data contains errors

42 babies were < 1 year old at the time they were entered into the database

28 of those were marked as “admitting to being gang members”

- เด็กอายุน้อยกว่า 1 ขวบ 42 คน ถูกบันทึกชื่อในฐานข้อมูลกลุ่มอาชญากรรมของมลรัฐแคลิฟอร์เนีย
- 28 คนถูกบันทึกว่า “สารภาพว่าเป็นสมาชิกกลุ่ม”

California gang database plagued with errors, unsubstantiated entries, state auditor finds



How the careless errors of credit reporting agencies are ruining people's lives

Their files are full of obvious mistakes that the companies are in no rush to correct.

FTC's large-scale study of credit reports in 2012:

- 26% had at least one mistake in their files
- 5% had errors that could be devastating

ฐานข้อมูลเครดิตปี 2012

- 26% ของแฟ้มประวัติมีจุดผิดอย่างน้อย 1 จุด
- 5% เป็นจุดผิดร้ายแรง

The New York Times

She Was Arrested at 14. Then Her Photo Went to a Facial Recognition Database.

ภาพถ่ายตอนถูกจับอายุ 14 ปี ถูกนำไปใช้ในฐานข้อมูลตรวจจับใบหน้า

FACE RECOGNITION ON FLAWED DATA

Garbage In, Garbage Out: Face Recognition on Flawed Data

NYPD used facial recognition and pics of Woody Harrelson to arrest a man

KHARI JOHNSON @KHARIJOHNSON MAY 16, 2019 8:52 AM

ตำรวจนิวยอร์กใช้รูป Woody Harrelson กับโปรแกรมตรวจจับใบหน้าจับคน

AP: Across US, police officers abuse confidential databases

SADIE GURMAN September 27, 2016

พบกรณีที่ตำรวจใช้ฐานข้อมูลลับอย่างผิดวิธีทั่วสหรัฐฯ

PERSONAL DATA PROTECTION

พระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคล พ.ศ. 2562

- ข้อมูลส่วนบุคคล vs ข้อมูลส่วนบุคคลอ่อนไหว
- บุคคลที่เกี่ยวกับข้อมูลส่วนบุคคล (เจ้าของข้อมูล ผู้ควบคุมข้อมูล ผู้ประมวลผลข้อมูล)
- หลักการในการเก็บข้อมูล (ได้รับความยินยอม มีความจำเป็นที่สมเหตุสมผล)
- การจัดการข้อมูล (เก็บ ประมวลผล เผยแพร่ ลบ)
- สิทธิของเจ้าของข้อมูล (สิทธิในการเข้าถึง สิทธิในการลบหรือทำลาย)
- ความรับผิดและบทลงโทษ (ทั้งทางแพ่ง อาญา และปกครอง)

พื้นที่นี้มีการใช้กล้องวงจรปิด ซึ่งดำเนินการโดย
สำนักงานพัฒนารัฐบาลดิจิทัล (องค์การมหาชน)

DGA
Digital Government Development Agency

โดยกล้องวงจรปิดจะทำการบันทึกข้อมูลดังต่อไปนี้ของท่าน
- ภาพนิ่ง
- ภาพเคลื่อนไหว
- เสียง
- ภาพทรัพย์สินของท่าน

ทั้งนี้ เพื่อประโยชน์โดยชอบด้วยกฎหมายของสำนักงานในการรักษาความปลอดภัยพื้นที่ของสำนักงาน รวมทั้งเพื่อป้องกันและระงับอันตรายต่อชีวิต ร่างกาย สุขภาพหรือทรัพย์สินของท่าน รวมทั้งปฏิบัติตามกฎหมายที่เกี่ยวข้องกับการควบคุมดูแลอาคารสถานที่ของสำนักงาน

สิทธิของท่าน : ตามพระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคล พ.ศ.2562 ท่านมีสิทธิหลายประการเกี่ยวกับข้อมูลส่วนบุคคลของท่านที่สำนักงานเก็บรวบรวมจากการใช้กล้องวงจรปิด

สำหรับข้อมูลเพิ่มเติมเกี่ยวกับประกาศความเป็นส่วนตัวในการใช้กล้องวงจรปิด รวมถึงสิทธิต่าง ๆ ของท่าน โปรดสแกน QR Code ทางด้านซ้ายนี้

FEEDBACK LOOPS & METRICS

The problem with metrics is a big problem for AI

Written: 24 Sep 2019 by *Rachel Thomas*

Overemphasizing metrics leads to:

- manipulation
- gaming
- myopic focus on short-term goals
- unexpected negative consequences

Much of AI/ML centers on optimizing a metric

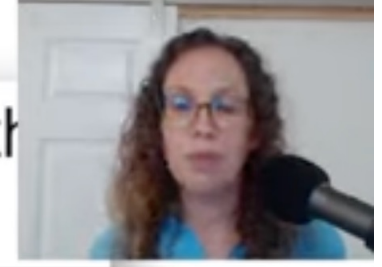
การจงใจ “เก็งโมเดล” ทำให้เกิด

- การจงใจบิดเบือนข้อมูล/โมเดล
- การพยายามชนะระบบเหมือนเล่นเกม
- เน้นแต่เป้าหมายระยะสั้น
- อาจเกิดผลทางลบที่ไม่คาดคิด

Reliance on Metrics is a Fundamental Challenge for AI

Rachel L. Thomas
University of San Francisco
rlthomas3@usfca.edu

David Uminsky
University of San Francisco
duminsky@usfca.edu



What's Measured Is What Matters: Targets and Gaming in th English Public Health Care System

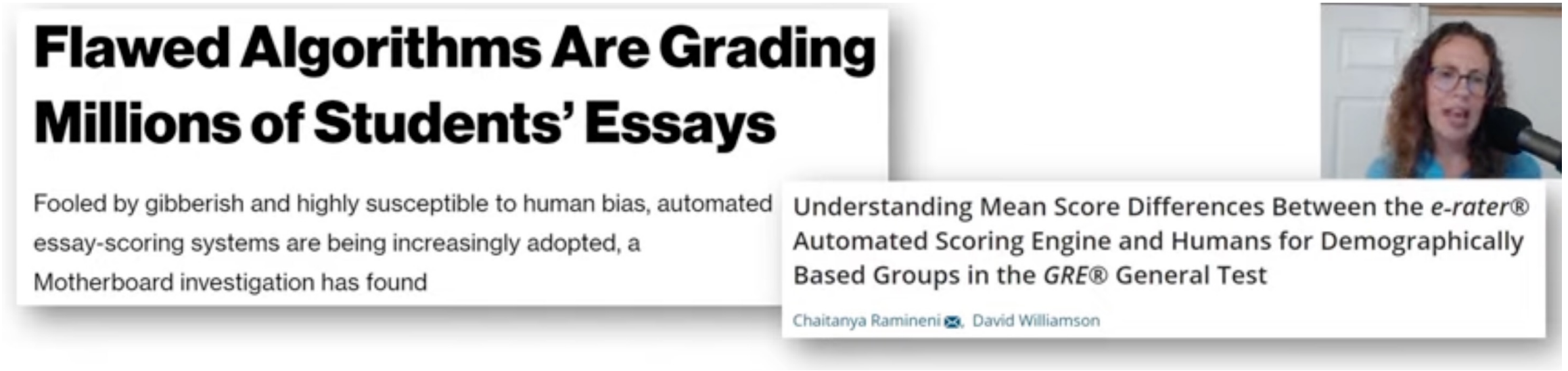
Targets around ER wait times led to:

- cancelling scheduled operations to draft extra staff to ER,
- requiring patients to wait in queues of ambulances
- turning stretchers into “beds” by putting them in hallways
- big discrepancies in #s reported by hospitals vs. by patients

ตัวอย่างการ “เก็ง” ระบบ

ประเทศอังกฤษวัดประสิทธิภาพการทำงานของโรงพยาบาลด้วยเวลารอเข้าพบแพทย์ฉุกเฉิน (ER) เมื่อมีการพยายามจะเอาชนะระบบ จึงมีการ:

- จงใจยกเลิกการผ่าตัดที่นัดหมายไว้เพื่อให้มีเจ้าหน้าที่ในแผนกฉุกเฉินมากขึ้น
- บังคับให้ผู้ป่วยรอคิวในรถพยาบาล (เพราะตัวชี้วัดกำหนดแค่เวลารอในโรงพยาบาล)
- นำ “เปลฉุกเฉิน” มาวางตามทางเดินในโรงพยาบาลจะได้นับเป็น “เตียง” ด้วย (เหมือนว่าผู้ป่วยได้รับการดูแลแล้ว)
- จำนวนนาทีที่รอที่รายงานโดยโรงพยาบาลกับผู้ป่วยไม่ตรงกัน



ตัวอย่างการ “เก็ง” ระบบ

AI ตรวจเรียงความนักเรียนถูกใช้ใน 22 มลรัฐในประเทศสหรัฐอเมริกา

- ให้คะแนนจากคำศัพท์ ความยาว การสะกดคำ ความถูกต้องของไวยากรณ์
- ไม่สามารถตรวจมิติคุณภาพของเรียงความได้ เช่น ความคิดสร้างสรรค์
- เรียงความที่ใช้ศัพท์ยากๆ ถึงจะเรียบเรียงความคิดได้ไม่ดี มีแนวโน้มจะได้คะแนนดีกว่า
- เรียงความที่เขียนโดยนักเรียนเชื้อสายแอฟริกาจะได้คะแนนน้อยกว่าเมื่อเทียบกับให้คนตรวจ
- เรียงความที่เขียนโดยนักเรียนจีนได้คะแนนสูงกว่าเมื่อเทียบกับให้คนตรวจ อาจเพราะมีประโยคที่ท่องมาตอบเยอะ

BIAS

REPRESENTATION BIAS

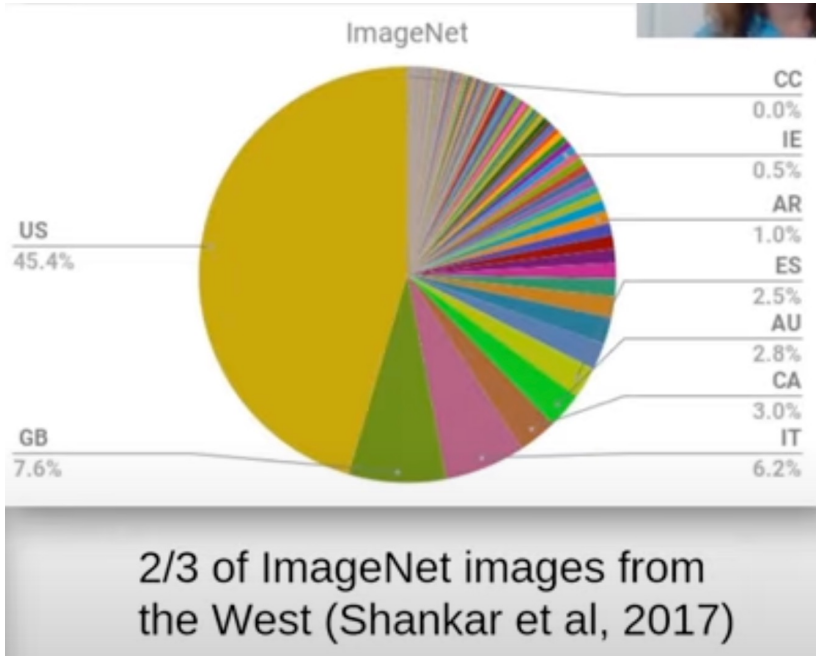
Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
Microsoft	94.0%	79.2%	100%	98.3%	20.8%
FACE++	99.3%	65.5%	99.2%	94.0%	33.8%
IBM	88.0%	65.3%	99.7%	92.9%	34.4%

Joy Buolamwini & Timnit Gebru, gendershades.org

อคติจากจำนวนข้อมูล

AI จำแนกเพศของ Microsoft, Face, IBM จำแนกเพศของผู้ชายคนขาวแม่นยำที่สุด ตามมาด้วยผู้ชายคนดำและผู้หญิงคนขาว และจำแนกผู้หญิงคนดำได้แม่นยำน้อยที่สุด โดยช่องว่างความแม่นยำมากถึง 34%

EVALUATION BIAS



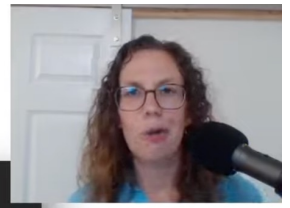
อคติจากการประเมินคุณภาพ

เทรนโมเดลด้วยภาพคนตะวันตก (คนขาว) เป็นหลัก และทดสอบความแม่นยำด้วยภาพคนขาว พบว่าโมเดลมีความแม่นยำมาก

CASE STUDY: การใช้ AI ในการกำหนดระยะเวลาจำคุก

Wisconsin Supreme Court allows state to continue using computer program to assist in sentencing

มลรัฐวิสคอนซินยังคงอนุญาตให้ใช้คอมพิวเตอร์ช่วยตัดสินระยะเวลาจำคุก



Machine Bias

Prediction Fails Differently for Black Defendants

	WHITE	AFRICAN AMERICAN
Labeled Higher Risk, But Didn't Re-Offend	23.5%	44.9%
Labeled Lower Risk, Yet Did Re-Offend	47.7%	28.0%

ถูกจำแนกว่ามีความเสี่ยงสูง แต่ไม่ได้กระทำผิดซ้ำ

vs

ถูกจำแนกว่าความเสี่ยงต่ำ แต่กระทำผิดซ้ำ

A Popular Algorithm Is No Better at Predicting Crimes Than Random People

อัลกอริธึมที่มีชื่อเสียงมีประสิทธิภาพในการทำนายอาชญากรรมไม่ต่างจากคนธรรมดา

HISTORICAL BIAS

"Historical bias is a fundamental, structural issue with the first step of the data generation process and can exist even given perfect sampling and feature selection." – Suresh et. al. 2019

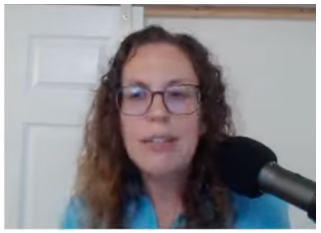
อคติจากข้อมูลในอดีต

อคติจากข้อมูลในอดีตเป็นปัญหาระดับพื้นฐานและโครงสร้าง ในกระบวนการสร้างข้อมูล และสามารถเกิดขึ้นได้ถึงแม้จะมีการสุ่มเลือกข้อมูลอย่างระมัดระวังแล้ว (Suresh et. al. 2019)

MEASUREMENT BIAS อคติจากการวัด

Does Machine Learning Automate Moral Hazard and Error?†

By Sendhil Mullainathan and Ziad Obermeyer*



Using historical EHR data, what factors are most predictive of stroke?

Prior Stroke

Cardiovascular disease

Accidental injury

Benign breast lump ???

Colonoscopy

Sinusitis

ปัจจัยที่สามารถทำนายโรคหลอดเลือดสมองได้มากที่สุด

- ประวัติมีโรคหลอดเลือดสมองมาก่อน
- โรคหัวใจ
- การประสบอุบัติเหตุ
- ก้อนเนื้อ (ที่ไม่เป็นอันตราย) บริเวณหน้าอก
- การส่องกล้องตรวจลำไส้ใหญ่
- โรคไซนัสอักเสบ

เพราะจริงๆ แล้วเราไม่ได้วัดว่าเรามีโรคหลอดเลือดสมองหรือไม่ เราวัดว่ามีอาการหรือไม่ หาหมอ วินิจฉัย และได้รับผลวินิจฉัยว่าเป็นโรคนี้หรือไม่

RACIAL BIAS

อคติจากเชื้อชาติ

- ด้วยลักษณะอาการเหมือนกันทุกประการ แพทย์มีแนวโน้มสั่งวินิจฉัยโรคด้วยการสวนหลอดเลือดหัวใจให้กับผู้ป่วยคนดำ
- การต่อราคารถมือสอง ลูกค้าคนดำถูกเสนอราคาเริ่มต้นแพงกว่าคนอื่น 700 เหรียญ และของแถมน้อยกว่า
- หากใช้ชื่อแบบคนดำสอบถามข้อมูลเช่าบ้านบนเว็บไซต์ Craiglist จะได้รับการตอบกลับน้อยกว่าชื่อแบบคนขาว
- ในการตัดสินคดีระบบลูกขุน หากลูกขุนเป็นคนขาวทั้งหมด จำเลยคนดำมีโอกาสถูกตัดสินว่ามีความผิดจริงมากกว่าคนขาวถึง 16 คะแนน แต่ถ้ามีคนดำเป็นคณะลูกขุนอย่างน้อย 1 คน ผลการตัดสินความผิดระหว่างจำเลยคนดำและจำเลยคนขาวจะไม่แตกต่างกันเลย


The New York Times

Racial Bias, Even When We Have Good Intentions

By Sendhil Mullainathan

MACHINE LEARNING ALGORITHMS CAN AMPLIFY BIAS

- สอนให้ AI ทำนายอาชีพจากประวัติส่วนตัว
- โดยข้อมูลบางครั้งจะมีข้อมูลเพศติดมาด้วย (เช่นคำว่า he/she)
- แต่พอเอาข้อมูลเพศออก พบว่าโมเดลแม่นยำน้อยลง
- ตัวอย่างอาชีพศัลยแพทย์
  - AI ทำนายผู้ชายว่ามีอาชีพศัลยแพทย์ถูก 71%
  - แต่ทำนายผู้หญิงว่ามีอาชีพศัลยแพทย์ถูกแค่ 54%
- เพราะแทนที่จะใช้ข้อมูลต่างๆ มาทำนายอาชีพ AI เลือกใช้เพศมาทำนายอาชีพ ว่าศัลยแพทย์มีแนวโน้มเป็นผู้ชายมากกว่า

Bias in Bios: A Case Study of Semantic Representation Bias in a High-Stakes Setting

Maria De-Arteaga[1], Alexey Romanov[2], Hanna Wallach[3], Jennifer Chayes[3], Christian Borgs[3], Alexandra Chouldechova[1], Sahin Geyik[4], Krishnaram Kenthapadi[4], Adam Tauman Kalai[3]

[1]*Carnegie Mellon University,* [2]*University of Massachusetts Lowell,* [3]*Microsoft Research,* [4]*LinkedIn*

ALGORITHMS ARE USED DIFFERENTLY THAN HUMAN DECISION MAKERS

- เรามักเชื่อว่าอัลกอริธึมมีความเป็นภววิสัย และมีข้อผิดพลาดน้อยกว่า (ถึงแม้ว่าจะสามารถเลือกให้คนสามารถแก้ไขการตัดสินใจของอัลกอริธึมได้)
- อัลกอริธึมมักถูกใช้โดยไม่มีกระบวนการที่เปิดโอกาสให้ร้องเรียนผลการตัดสินใจ
- อัลกอริธึมมักถูกใช้กับงานที่มีจำนวนมาก
- อัลกอริธึมราคาถูก

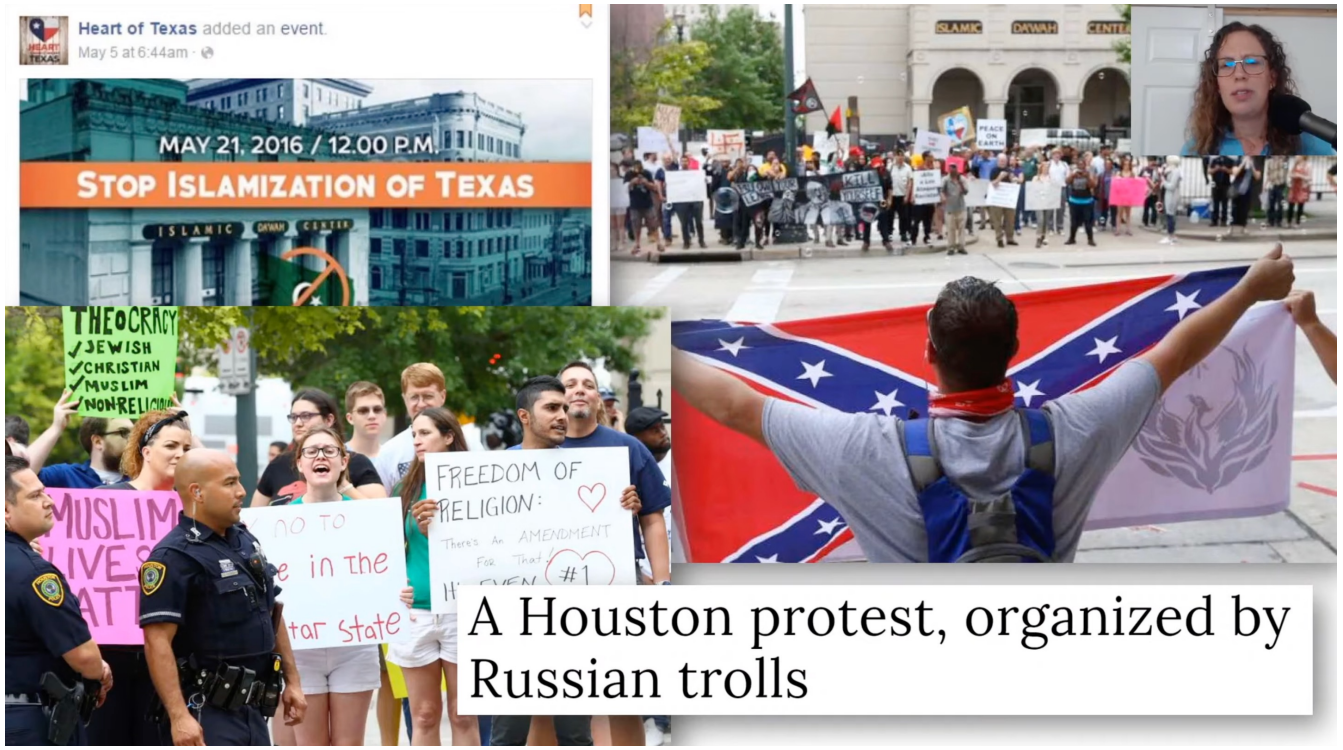
HUMANS ARE BIASED, SO WHY DOES ALGORITHMIC BIAS MATTER?

- เพราะ Machine Learning สามารถสร้าง feedback loops ไม่ใช่แค่รวบรวมข้อมูล แต่สามารถกำหนดลักษณะของข้อมูลที่จะได้รับในอนาคตได้
- เพราะ Machine Learning สามารถขยายอคติที่มีให้รุนแรงขึ้น
- อัลกอริธึมกับมนุษย์ถูกใช้แตกต่างกัน
- เมื่อเทคโนโลยีคือพลัง ย่อมต้องมีความรับผิดชอบตามมา

DISINFOR-MATION
ข้อมูลเท็จ

- ข้อมูลเท็จ = ไม่จริง
- ข้อมูลเท็จ หมายรวมถึงแคมเปญ/ความเคลื่อนไหวที่มีความต่อเนื่องกันที่มีจุดประสงค์เพื่อควบคุมให้เกิดผลลัพธ์บางอย่าง
- ข้อมูลเท็จ เป็นระบบนิเวศ

- AI ของเราอาจถูกนำไปใช้ในการสร้างข้อมูลเท็จ



เกิดม็อบชนม็อบ วันที่ 21 พฤษภาคม 2016 ที่มลรัฐเท็กซัส ระหว่างกลุ่มต่อต้านอิสลามและกลุ่มปกป้องอิสลาม
ปรากฏว่าทั้งสองม็อบจัดโดยรัสเซีย

- FB 73 เพจ ความเคลื่อนไหว 9.7 ล้านครั้งใน 6 ประเทศ ได้แก่ โมซัมบิก แคเมอรูน สาธารณรัฐแอฟริกากลาง คองโก และลิเบีย
- มักเผยแพร่ผ่านแหล่งข่าวท้องถิ่น รวมถึง Telegram และ WhatsApp ด้วย

Claire Wardle:
the trumpet of amplification

การขยายขนาดผลกระทบของข้อมูลเท็จ

4chan/
8chan

WhatsApp/
Telegram/
FB messenger groups

Conspiracy Groups on Reddit/
YouTube

Social Media

Professional Media & Politicians

reddit [redacted] comments other discussions (1)

355 unpopularopinion I believe the US should cut all defence spending and instead spend the money on the military [redacted]

submitted 15 hours ago by [redacted]

I know that the
have in the buc
Instead we hav
just doesn't ma

16 comments share

[–] [redacted] 38 points 15 hours ago

You're wrong.

The defense budget is a good example of how badly the US spends money on the military. I've never seen anyone in the US that wouldn't spend money on the military. If you're going to spend that kind of money, then it absolutely makes sense to send the money to the military.

permalink embed save report give award reply

[–] [redacted] 9 points 15 hours ago

Yeah, but that's already happening. There is a huge increase in the military budget, the Pentagon budget is already increasing, and the Navy is getting two frigates a year. If we just keep cutting military spending, then we're already there.

If we stop paying for the military, there would be no need for an increase in defense spending.

This is all about the money.

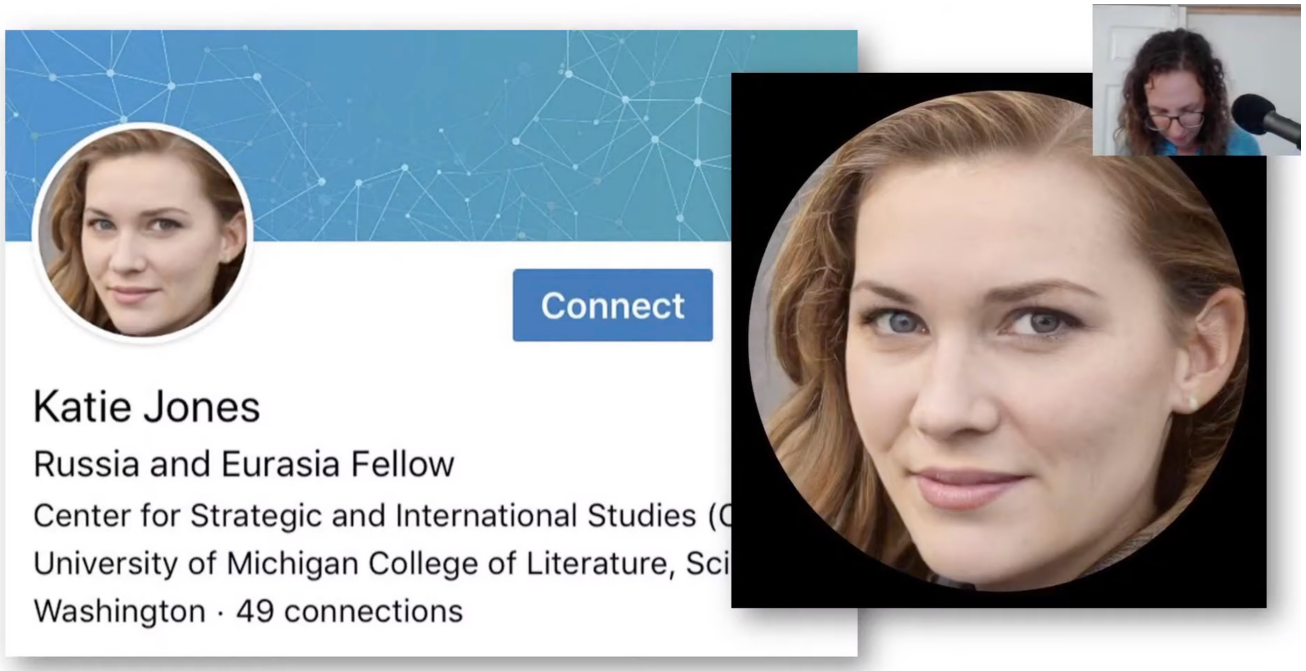
permalink embed save parent report give award reply

[–] [redacted] 4 points 15 hours ago

I didn't mean to sound like "stop paying for the military". I'm not saying that we cannot pay the bills but I think it would make sense to cut defense spending.

The military isn't a good example of what we could do with the money we have. People

คอมคุยกัน



https://www.thispersondoesnotexist.com/

คนปลอม

"In the matter of restoring Internet freedom. I'd like to recommend the commission to
Obama/Wheeler power grab to control Internet access. Americans, as opposed to Washingto
deserve to enjoy the services they desire. The Obama/Wheeler power grab to control Inte
a distortion of the open Internet. It ended a hands-off policy that worked exceptionall
for many years with bipartisan support.",
"Chairman Pai: With respect to Title 2 and net neutrality. I want to encourage the FCC to rescind
Barack Obama's scheme to take over Internet access. Individual citizens, as opposed to Washington
bureaucrats, should be able to select whichever services they desire. Barack Obama's scheme to take
over Internet access is a corruption of net neutrality. It ended a free-market approach that
performed remarkably smoothly for many years with bipartisan consensus.",
"FCC: My comments re: net neutrality regulations. I want to suggest the commission to overturn
Obama's plan to take over the Internet. People like me, as opposed to so-called experts, should be
free to buy whatever products they choose. Obama
net neutrality. It broke a pro-consumer system t
with Republican and Democrat support.",
"Mr Pai: I'm very worried about restoring Inte
Obama/Wheeler policy to regulate the Internet. C
whichever services we prefer. The Obama/Wheeler
the open Internet. It disrupted a market-based c
decades with Republican and Democrat consensus."
"FCC: In reference to net neutrality. I would
scheme to control the web. Citizens, as opposed
whatever products they prefer. Obama's scheme to

More than a Million Pro-Repeal Net Neutrality Comments were Likely Faked

I used natural language processing techniques to analyze net neutrality comments submitted to the FCC from April-October 2017, and the results were disturbing.

Jeff Kao Follow

Nov 23, 2017 · 10 min read

Net Neutrality คือหลักการที่ผู้ให้บริการอินเทอร์เน็ตต้องให้บริการอย่างเท่าเทียม โดยไม่มีการเก็บเงินเพิ่มหรือลดความเร็วกับบริการบางอย่าง เช่น การเข้าบางเว็บ

More than a Million Pro-Repeal Net Neutrality Comments were Likely Faked

Jeff

Computation

After clustering comment categories and removing duplicates, ***I found that less than 800,000 of the 22M+ comments submitted to the FCC (3-4%) could be considered truly unique.***

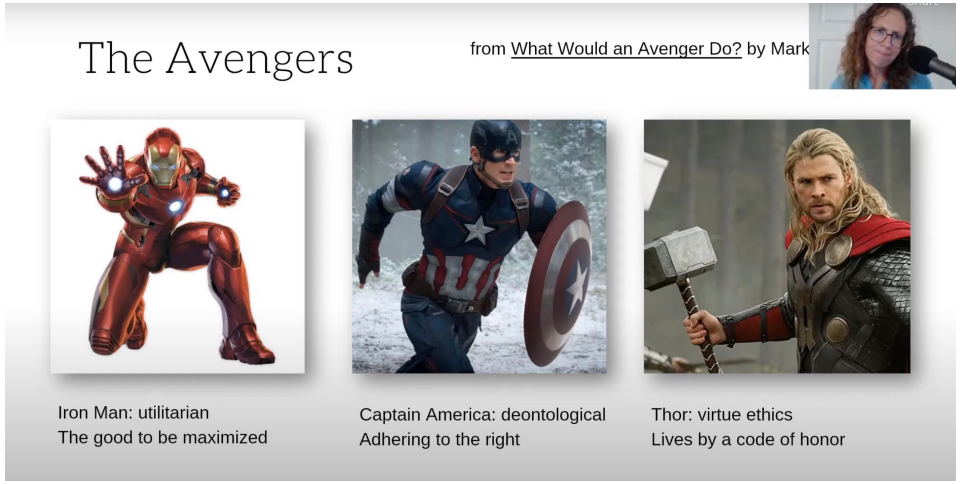
Key Findings:[2]

1. One pro-repeal spam campaign ***used mail-merge to disguise 1.3 million comments*** as unique grassroots submissions.
2. There were likely multiple other campaigns aimed at injecting what may total ***several million*** pro-repeal comments into the system.
3. It's highly likely that ***more than 99%*** of the truly unique comments[3] were in favor of keeping net neutrality.

1. พบถึง 1.3 ล้านความเห็น
2. ยังมีแคมเปญอื่นๆ ในลักษณะนี้อีก ซึ่งสามารถเพิ่มความเห็นไม่สนับสนุน Net Neutrality อีกนับล้านความเห็น
3. โดยที่ 99% ของความเห็นจากคนจริงๆ สนับสนุน Net Neutrality

ETHICAL FOUNDATIONS

จริยธรรม AI มีพื้นฐานมาจากวิชาจริยศาสตร์



- Iron Man: "ประโยชน์นิยม" ประโยชน์ - ความดีเป็นสิ่งที่ต้องเพิ่มให้สูงสุด
  - พิจารณาจากผลของการกระทำมากกว่าเจตนา
  - ถึงเจตนาไม่ดีแต่เกิดผลประโยชน์กับคนหมู่มากก็ถือว่า ok
  - **กรณีโต้แย้ง** การฆ่าคนบริสุทธิ์หนึ่งคนเพื่อ "ยกระดับสังคม" ก็ย่อมทำได้ (ถ้าฆ่าคนดังที่น่าหมั่นไส้มากๆ คนทั้งโลกเกลียด 1 คนได้แล้วคนอื่นทั้งโลกจะมีความสุขกว่าก็ทำได้)
- Captain America: "จริยศาสตร์เชิงกรณียธรรม"
  - การกระทำไม่ขึ้นอยู่กับผลของการกระทำ แต่ขึ้นอยู่กับลักษณะบางอย่างของตัวการกระทำเอง
  - การกระทำที่ควรกระทำ มีลักษณะจำเป็น หรือต้องกระทำโดยปราศจากเงื่อนไข (Richardson, 2006: 713)
  - **กรณีโต้แย้ง** การฆ่าคนเป็นสิ่งผิด แม้แต่การฆ่าคนที่กำลังจะระเบิดฆ่าคนจำนวนมากก็ทำไม่ได้
- Thor: "จริยศาสตร์คุณธรรม"
  - วิธีเข้าถึงซึ่งความเข้าใจและการดำเนินชีวิตที่ดี โดยมีพื้นฐานอยู่ที่คุณธรรม (Kollar, Nathan R.)
  - Cardinal virtue คุณธรรมหลัก 4 ประการตามปรัชญาของเพลโต คือ ปัญญา ความกล้าหาญ ความรู้จักประมาณ และความยุติธรรม
  - คุณธรรมแบบอื่นๆ เช่น คุณธรรมขงจื๊อ คุณธรรมแบบพุทธ ฯลฯ
  - **"คนดีย์"** ขาดมาตรฐานในการตัดสินเพราะยึดถือสามัญสำนึกของผู้กระทำเป็นเกณฑ์

คำถามเชิงกรณียศาสตร์สำหรับเทคโนโลยี (AKA มีกฎอะไรที่เราต้องทำตามบ้างเมื่อพัฒนาเทคโนโลยี)

1. สิทธิของผู้อื่นและหน้าที่ที่เรามีต่อผู้อื่นมีอะไรบ้าง
2. ศักดิ์ศรีและสิทธิในการตัดสินใจของผู้มีส่วนเกี่ยวข้องได้รับผลกระทบจากโครงการของเราอย่างไรบ้าง
3. มีประเด็นที่เกี่ยวกับความเชื่อใจและความยุติธรรมอะไรบ้างที่เราต้องพิจารณา
4. โครงการของเราขัดแย้งกับหน้าที่ทางศีลธรรมที่เรามีต่อผู้อื่นหรือสิทธิของผู้มีส่วนเกี่ยวข้องหรือไม่ และเราจะให้ความสำคัญกับประเด็นเหล่านี้อย่างไร
5. ใครบ้างที่ได้รับผลกระทบโดยตรง ใครบ้างที่ได้รับผลกระทบโดยอ้อม

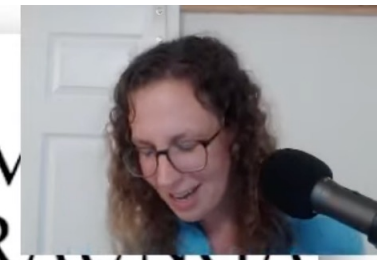
คำถามเชิงกรณียศาสตร์สำหรับเทคโนโลยี
(AKA มีกฎอะไรที่เราต้องทำตามบ้างเมื่อพัฒนาเทคโนโลยี)

1. ใครบ้างที่ได้รับผลกระทบโดยตรง ใครบ้างที่ได้รับผลกระทบโดยอ้อม
2. ผลสะสมที่เกิดจากโครงการของเรา มีประโยชน์มากกว่าโทษหรือไม่ และประโยชน์และโทษที่เกิดขึ้นคืออะไรบ้าง
3. เราคิดถึงประโยชน์และโทษด้านต่างๆ ครอบคลุมหรือยัง (จิตวิทยา สังคม สิ่งแวดล้อม ศีลธรรม ปัญญา อารมณ์ สถาบัน วัฒนธรรม)
4. มีความเสี่ยงที่ว่าโทษจากโครงการของเราจะส่งผลกระทบต่อคนที่มีโอกาสและอำนาจตัดสินใจน้อยกว่า อย่างไม่ได้สัดส่วนหรือไม่

มุมมองทางจริยธรรม 5 แบบ

1. **มุมมองด้านสิทธิ**
   ทางเลือกที่ดีที่สุดที่เคารพสิทธิของผู้มีส่วนเกี่ยวข้อง
2. **มุมมองด้านความยุติธรรม**
   ทางเลือกที่ปฏิบัติกับทุกคนอย่างเท่าเทียมและได้สัดส่วน
3. **มุมมองด้านประโยชน์**
   ทางเลือกที่สร้างประโยชน์สูงสุดและสร้างโทษน้อยที่สุด
4. **มุมมองด้านความดีโดยทั่วไป**
   ทางเลือกที่รับใช้สังคมทั้งหมดไม่ใช่แค่คนกลุ่มหนึ่งกลุ่มใด
5. **มุมมองด้านคุณธรรม**
   ทางเลือกที่ทำให้ฉันปฏิบัติตนอย่างคนที่ฉันอยากเป็น

There are other ethical lenses



TE M
RARAUNGA
Māori Data Sovereignty Network

Data from a Māori worldview: Taonga

- All data are taonga
  - Intrinsic mana – inherent value of data (all data are taonga)
  - Extrinsic mana – valued for specific purposes (taonga is as taonga is used)
- Ensure the mana and mauri of the taonga are intact

Data from a Māori worldview: Whakapapa

- Where it comes from and where it is going to (past, present, future)
- Relates to the purpose and use of data
- Relates to the importance of contextual knowledge
- Relationships within the data (linking individuals to groups)
- Whakapapa of the data (metadata, provenance)
- Still exists even if data set is anonymised or de-identified
- Whanaungatanga
  - Data relates to both the individual and the collective
  - Collective identifiy is integral to Māori data

Data from a Maori worldview

เราสามารถทำอะไรได้บ้าง

- ทบทวนงานอย่างสม่ำเสมอ โดยเฉพาะโครงการต่อเนื่อง
- ทำความเข้าใจงานของเราจากหลายๆ มุมมอง
  - ปรึกษาอาจารย์ – ผู้เชี่ยวชาญ หลายๆ ด้าน
  - ปรึกษาผู้ใช้งานของเรา
  - มีใครที่สนใจนำงานของเราไปใช้โดยที่เราไม่เคยคิดถึงคนกลุ่มนี้
- ลองคิดในจากมุมมองที่แย่ที่สุด (ถ้าเราเป็นคนไม่ดี)
  - ใครที่อยากจะนำงานของเราไปใช้ในทางไม่ดี
  - มีแรงจูงใจอะไรให้มีคนอยากนำงานของเราไปใช้ในทางไม่ดี
  - พยายามแก้ไขไม่ให้เกิดแรงจูงใจนั้น
- ปิดวงจรที่ก่อให้เกิดความเสี่ยงด้านจริยธรรม
  - การป้องกันความเสี่ยงด้านจริยธรรมเป็นงานต่อเนื่อง
  - รับฟังความเห็นที่เกี่ยวกับความเสี่ยงด้านจริยธรรม
  - ศึกษาแนวทางจากงานของคนอื่นๆ นำมาปรับใช้กับงานของเรา