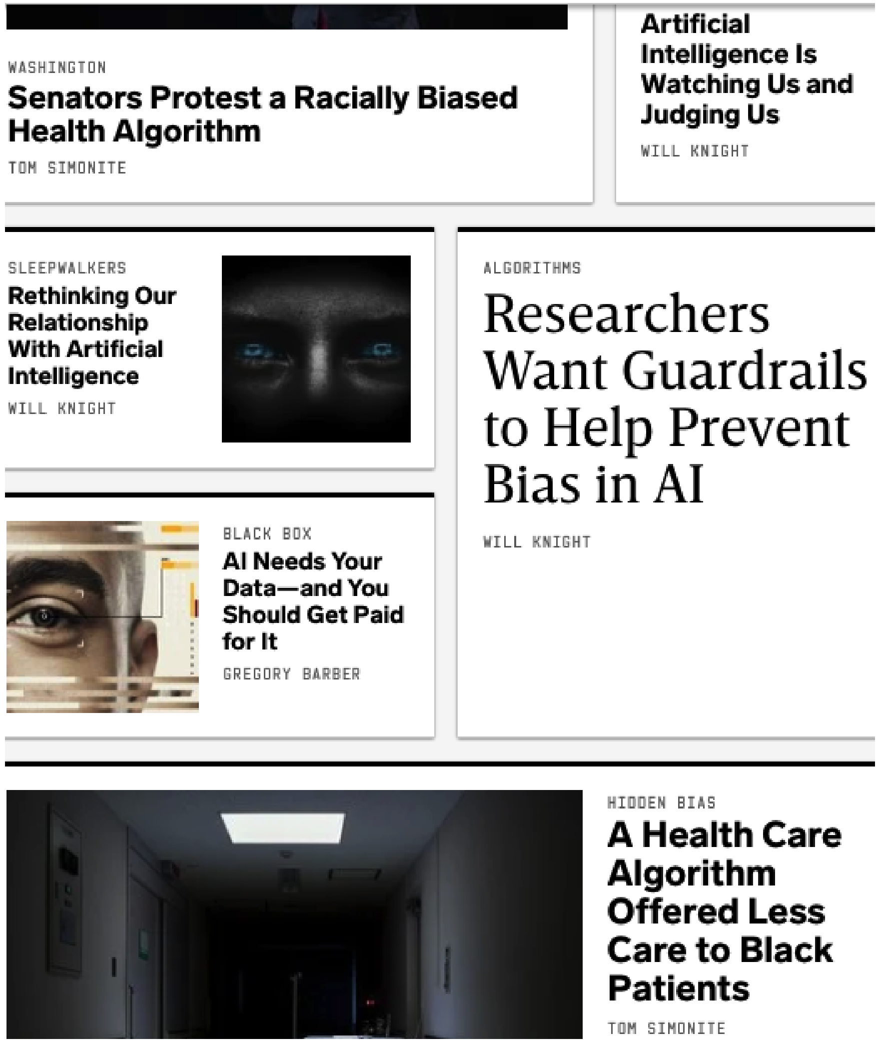
Ethics of Artificial Intelligence & Learning Environments

Rebekka Darner, *School of Biological Sciences & CeMaST*
Elahe Javadi, *School of Information Technology*
Allison Antink-Meyer, *School of Teaching and Learning*



WASHINGTON
Senators Protest a Racially Biased Health Algorithm
TOM SIMONITE

Artificial Intelligence Is Watching Us and Judging Us
WILL KNIGHT

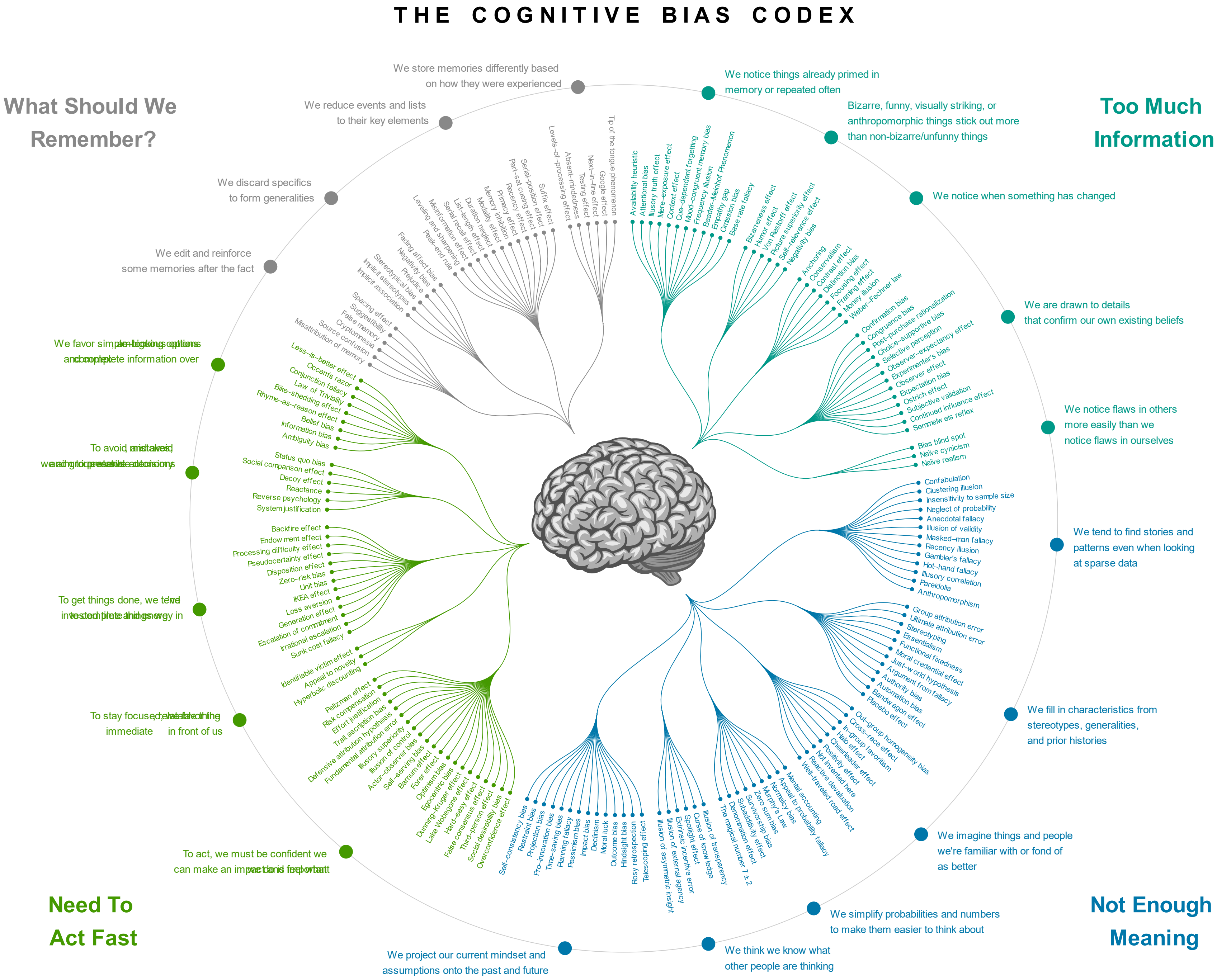
SLEEPWALKERS
Rethinking Our Relationship With Artificial Intelligence
WILL KNIGHT

ALGORITHMS
Researchers Want Guardrails to Help Prevent Bias in AI
WILL KNIGHT

BLACK BOX
AI Needs Your Data—and You Should Get Paid for It
GREGORY BARBER

HIDDEN BIAS
A Health Care Algorithm Offered Less Care to Black Patients
TOM SIMONITE

Chihuahua or Muffin?



THE COGNITIVE BIAS CODEX
What Should We Remember?
We store memories differently based on how they were experienced
We reduce events and lists to their key elements
We discard specifics to form generalities
We edit and reinforce some memories after the fact
Tip of the tongue phenomenon
Google effect
Next-in-line effect
Testing effect
Absent-mindedness
Levels-of-processing effect
Suffix effect
Serial-position effect
Part-set cueing effect
Recency effect
Primacy effect
Memory inhibition
Modality effect
Duration neglect
List-length effect
Serial recall effect
Misinformation effect
Leveling and sharpening
Peak-end rule
Fading affect bias
Negativity bias
Prejudice
Stereotypical bias
Implicit stereotypes
Implicit associations
Spacing effect
Suggestibility
False memory
Cryptomnesia
Source confusion
Misattribution of memory
Too Much Information
We notice things already primed in memory or repeated often
Bizarre, funny, visually striking, or anthropomorphic things stick out more than non-bizarre/unfunny things
We notice when something has changed
We are drawn to details that confirm our own existing beliefs
We notice flaws in others more easily than we notice flaws in ourselves
Availability heuristic
Attentional bias
Illusory truth effect
Mere-exposure effect
Context effect
Cue-dependent forgetting
Mood-congruent memory bias
Frequency illusion
Baader-Meinhof Phenomenon
Empathy gap
Omission bias
Base rate fallacy
Bizarreness effect
Humor effect
Von Restorff effect
Picture superiority effect
Self-relevance effect
Negativity bias
Anchoring
Conservatism
Contrast effect
Distinction bias
Focusing effect
Framing effect
Money illusion
Weber-Fechner law
Confirmation bias
Congruence bias
Post-purchase rationalization
Choice-supportive bias
Selective perception
Observer-expectancy effect
Experimenter's bias
Observer effect
Expectation bias
Ostrich effect
Subjective validation
Continued influence effect
Semmelweis reflex
Bias blind spot
Naïve cynicism
Naïve realism
Need To Act Fast
We favor simple-looking options and complete information over
To avoid mistakes
To get things done, we tend
To stay focused, we favor the immediate
To act, we must be confident we can make an impact and feel what
Less-is-better effect
Occam's razor
Conjunction fallacy
Law of Triviality
Bike-shedding effect
Rhyme-as-reason effect
Belief bias
Information bias
Ambiguity bias
Status quo bias
Social comparison effect
Decoy effect
Reactance
Reverse psychology
System justification
Backfire effect
Endowment effect
Processing difficulty effect
Pseudocertainty effect
Disposition effect
Zero-risk bias
Unit bias
IKEA effect
Loss aversion
Generation effect
Escalation of commitment
Irrational escalation
Sunk cost fallacy
Identifiable victim effect
Appeal to novelty
Hyperbolic discounting
Peltzman effect
Risk compensation
Effort justification
Trait ascription bias
Defensive attribution hypothesis
Fundamental attribution error
Illusory superiority
Illusion of control
Actor-observer bias
Self-serving bias
Barnum effect
Forer effect
Optimism bias
Egocentric bias
Dunning-Kruger effect
Lake Wobegone effect
Hard-easy effect
False consensus effect
Third-person effect
Social desirability bias
Overconfidence effect
Not Enough Meaning
We tend to find stories and patterns even when looking at sparse data
We fill in characteristics from stereotypes, generalities, and prior histories
We imagine things and people we're familiar with or fond of as better
We simplify probabilities and numbers to make them easier to think about
We think we know what other people are thinking
We project our current mindset and assumptions onto the past and future
Confabulation
Clustering illusion
Insensitivity to sample size
Neglect of probability
Anecdotal fallacy
Illusion of validity
Masked-man fallacy
Recency illusion
Gambler's fallacy
Hot-hand fallacy
Illusory correlation
Pareidolia
Anthropomorphism
Group attribution error
Ultimate attribution error
Stereotyping
Essentialism
Functional fixedness
Moral credential effect
Just-world hypothesis
Argument from fallacy
Authority bias
Automation bias
Bandwagon effect
Placebo effect
Out-group homogeneity bias
Cross-race effect
In-group favoritism
Halo effect
Cheerleader effect
Positivity effect
Not invented here
Reactive devaluation
Well-traveled road effect
Mental accounting
Appeal to probability fallacy
Normalcy bias
Murphy's Law
Zero sum bias
Survivorship bias
Subadditivity effect
Denomination effect
The magical number 7 ± 2
Illusion of transparency
Curse of knowledge
Spotlight effect
Extrinsic incentive error
Illusion of external agency
Illusion of asymmetric insight
Telescoping effect
Rosy retrospection
Hindsight bias
Outcome bias
Moral luck
Declinism
Impact bias
Pessimism bias
Planning fallacy
Time-saving bias
Pro-innovation bias
Projection bias
Restraint bias
Self-consistency bias

Implicit bias	Sampling bias	Automation bias
Reporting bias	In-group bias	Coverage bias
Out-group homogeneity bias	Non-response bias	Confirmation bias

Neutrality

Biases

Mechanics of AI

Stakeholders

Errors

Regulations

AI learning cycles

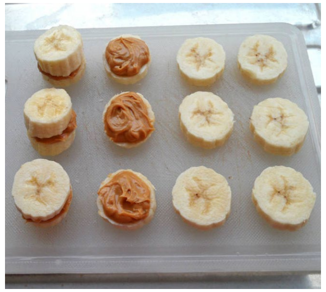
An Algorithm for the best PB&J sandwich?

input

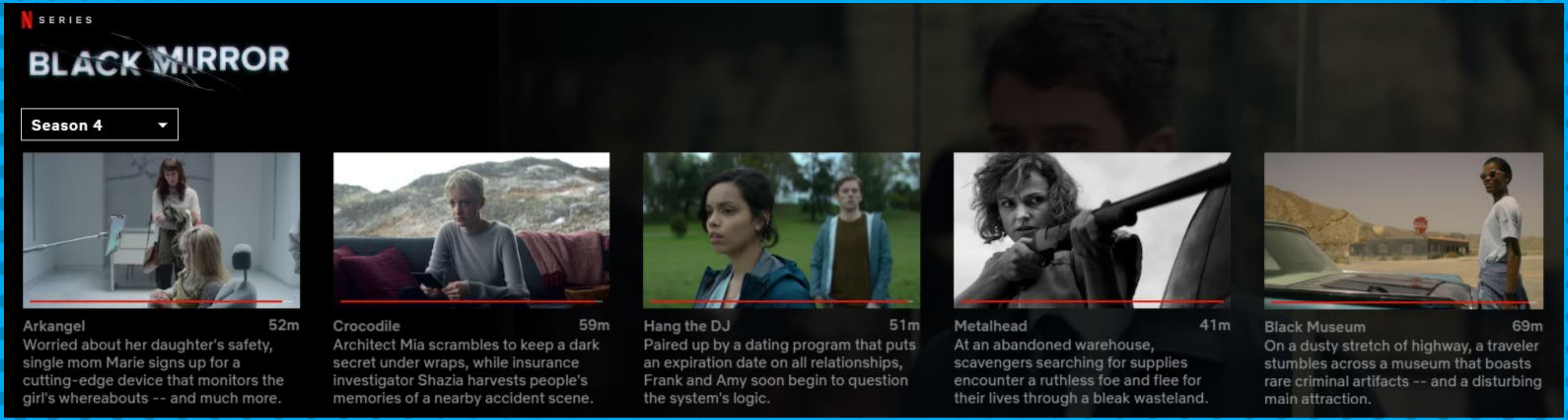
Steps to change input

output

	Taste	Nutrition	Cost
Child			
Parent			
Doctor			



The “best” PB&J sandwich could mean a lot


SERIES
BLACK MIRROR
Season 4
Arkangel 52m
Worried about her daughter's safety, single mom Marie signs up for a cutting-edge device that monitors the girl's whereabouts -- and much more.
Crocodile 59m
Architect Mia scrambles to keep a dark secret under wraps, while insurance investigator Shazia harvests people's memories of a nearby accident scene.
Hang the DJ 51m
Paired up by a dating program that puts an expiration date on all relationships, Frank and Amy soon begin to question the system's logic.
Metalhead 41m
At an abandoned warehouse, scavengers searching for supplies encounter a ruthless foe and flee for their lives through a bleak wasteland.
Black Museum 69m
On a dusty stretch of highway, a traveler stumbles across a museum that boasts rare criminal artifacts -- and a disturbing main attraction.

RESEARCH ARTICLES

ECONOMICS

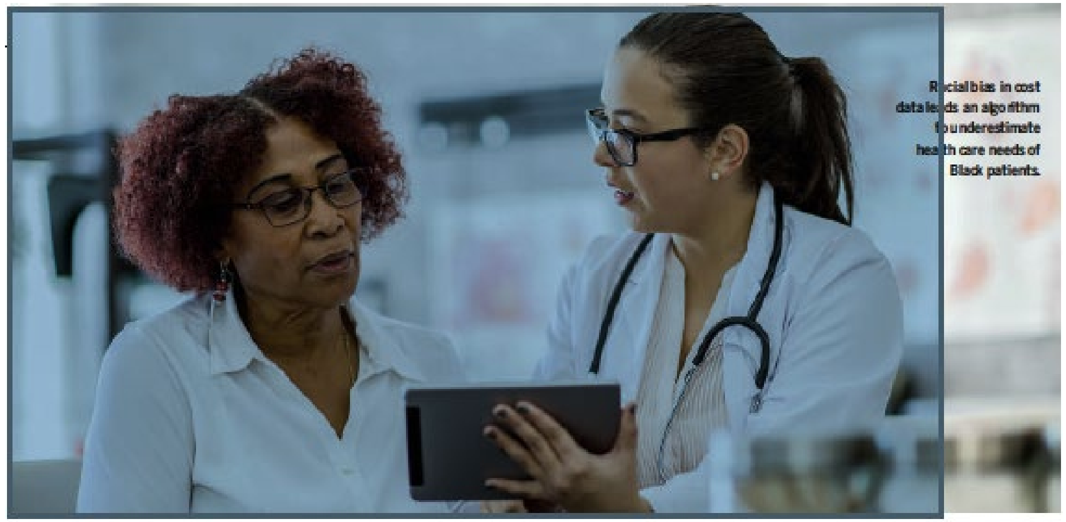
Dissecting racial bias in an algorithm used to manage the health of populations

Ziad Obermeyer[1,2]*, Brian Powers[3], Christine Vogeli[4], Sendhil Mullainathan[5]*†

Health systems rely on commercial prediction algorithms to identify and help patients with complex health needs. We show that a widely used algorithm, typical of this industry-wide approach and affecting millions of patients, exhibits significant racial bias: At a given risk score, Black patients are considerably sicker than White patients, as evidenced by signs of uncontrolled illnesses. Remedying this disparity would increase the percentage of Black patients receiving additional help from 17.7 to 46.5%. The bias arises because the algorithm predicts health care costs rather than illness, but unequal access to care means that we spend less money caring for Black patients than for White patients. Thus, despite health care cost appearing to be an effective proxy for health by some measures of predictive accuracy, large racial biases arise. We suggest that the choice of convenient, seemingly effective proxies for ground truth can be an important source of algorithmic bias in many contexts.

Overbooked and Overlooked: Machine Learning and Racial Bias in Medical Appointment Scheduling

32 Pages • Posted: 23 Oct 2019



Racial bias in cost data leads an algorithm to underestimate health care needs of Black patients.

SOCIAL SCIENCE

Assessing risk, automating racism

A health care algorithm reflects underlying racial bias in society

By Ruha Benjamin

era, the intention to deepen racial inequi-

beyond the algorithm developers by con-

Type I and Type II errors & stakeholders

Discussion:

1. If there are two tests and they have all numbers similar except for False Positives, which one would you prefer, the one with higher or lower number of False Positives?
2. If there are two tests and they have all numbers similar except for False Negatives, which one would you prefer, the one with higher or lower number of False Negative?
3. If we're comparing two tests and they have different numbers for both False Negatives and False Positives. Would you pick the one with lower False Positives or the one with lower False Negatives?

Can your answer depend on the system we're discussing or depend on what groups of stakeholders we are representing? See the ethical matrix below and use a [-2,+2] scale (-2,-1,0,+1,+2) to show a group may be negatively /positively impacted by an outcome; 0 if they may not be impacted much or be.

	False Positive	**False Negative**
patient		
doctors		
Insurance company		

H.R.2231 - Algorithmic Accountability Act of 2019

116th Congress (2019-2020) | Get alerts

BILL Hide Overview ✖

Sponsor: Rep. Clarke, Yvette D. [D-NY-9] (Introduced 04/10/2019)

Committees: House - Energy and Commerce

Latest Action: House - 04/10/2019 Referred to the House Committee on Energy and Commerce.

S.2637 - Mind Your Own Business Act of 2019

116th Congress (2019-2020) | Get alerts

BILL Hide Overview ✖

Sponsor: Sen. Wyden, Ron [D-OR] (Introduced 10/17/2019)

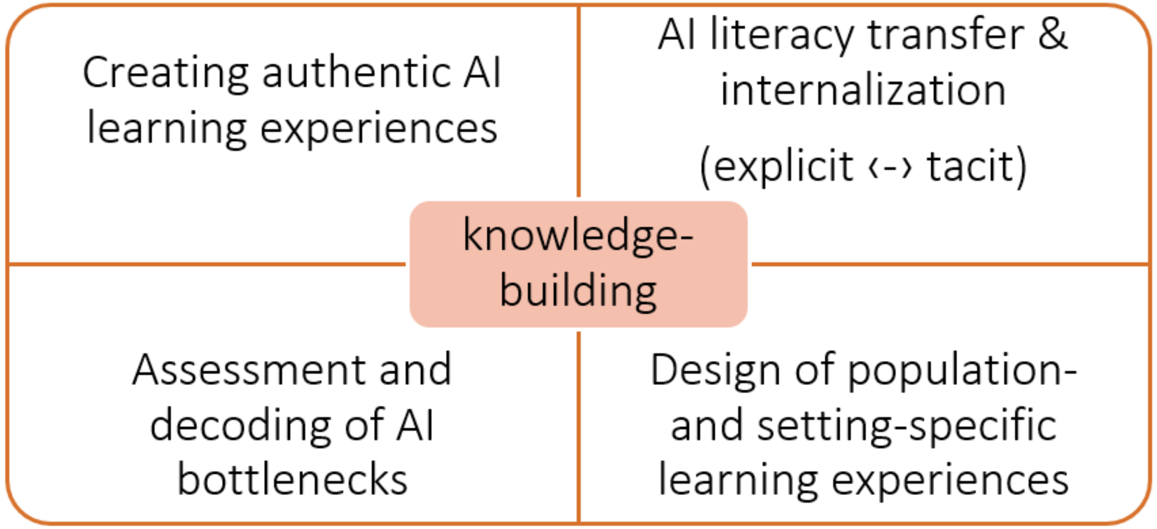
Committees: Senate - Finance

Latest Action: Senate - 10/17/2019 Read twice and referred to the Committee on Finance.

Artificial Intelligence Learning Cycle

Artificial Intelligence Learning Cycle (AILC; Antink-Meyer & Arias, in press)	
AILC Phase	**Resulting from this phase, learners:**
Empathy	• become familiar with a problem that is embedded in a context/storyline • built personal connections with the context/storyline
Engage	• become familiar with a AI technique, tool, or service that they will need in the AILC • developed an understanding of the nature of the problem • identified parameters involved in the problem
Explore I	• explored concepts related to the problem • experienced practices needed in the AILC including collection and analysis data needed in the AILC
Explain	• self-assessed knowledge of concepts and practices • developed understanding about the skills needed to create a solution to the problem • improved knowledge about the concepts related to the context/storyline and problem
Explore II	• prototyped (e.g. computer programs, simulated models, investigation of design elements) • analyzed potential design solutions and justified their designs using their knowledge of concepts and skills
Elaboration	• application of evidence from previous AILC phases to a unique design solution • analysis of design solution performance • proposed improvements based on performance analyses

Cascading teaching-learning model



Google’s Teachable Machines

- https://medium.com/tensorflow/real-time-human-pose-estimation-in-the-browser-with-tensorflow-js-7dd0bc881cd5

Acknowledgement

Some content and some images on the slides have been taken from the web including the following website(s):

- www.codeproject.com
- http://www.inf.u-szeged.hu/~ormandi/ai2/06-naiveBayes-example.pdf
- http://www.internetbillboards.net/2015/12/15/how-to-get-more-from-online-course-discussions/
- http://www.ellenhartson.com/do-you-have-an-agenda/
- http://info.growingyourleaders.com/blog/peer-led-learning-%E2%80%93-the-future
- http://healthymamamagazine.com/teal-tick/

References

- Almstrum V. M., Henderson P. B., Harvey V., Heeren C., Marion W., Riedesel C., Soh L-K, & Allison Tew E. (2006). Concept inventories in computer science for the topic discrete mathematics. SIGCSE Bull. 38 (4):132-145. DOI: https://doi.org/10.1145/1189136.1189182
- Antink-Meyer, A. & Arias, A.M. (accepted, in press). Teaching K-8 teachers about integrating science and engineering: An engineering learning cycle model and acoustics example. Journal of College Science Teaching.
- Allison Antink-Meyer & Brown, R.A. (2019). Nature of Engineering Knowledge: An Articulation of NOEK for Science Learners with NOS Understandings. Science & Education, 28(3): 539-559.
- Angwin J. , A. Tobin, M. Varner (2017), Facebook (still) letting housing advertisers exclude users by race. ProPublica, 21 November 2017; www.propublica.org/article/ facebook-advertising-discrimination-housing-race-sexnational-origin.
- Benjamin R. (2019). Race After Technology: Abolitionist Tools for the New Jim Code, 1st edition, Polity Press.
- Benjamin, R (2019). Assessing risk, automating racism, Science, 366 (6464): 421-422.
- Cross, J., Hamner, E., Zito, L., Nourbakhshh, I., & Bernstein, D. (2016). Development of an assessment for measuring middle school student attitudes towards robotics activities. In Frontiers in Education Conference (FIE), 1: 1-8.
- Cusumano-Towner M. F, Saad F. A., Lew A. K., and Mansinghka (2019). Gen: a general-purpose probabilistic programming system with programmable inference, Proceedings of the 40th ACM SIGPLAN Conference on Programming Language Design and Implementation, Phoenix, AZ, USA, 221-236.
- Chouldechova A. and A. Roth (2018). The Frontiers of Fairness in Machine Learning. http://www.cis.upenn.edu/~aaroth/cccreport.html
- AI4All to launch in September, now phase 2 beta testing
- Danielsiek H., Paul W., & Vahrenhold J. (2012). Detecting and understanding students' misconceptions related to algorithms and data structures. Proceedings of the 43rd ACM technical symposium on Computer Science Education (SIGCSE '12). ACM, New York, NY, USA, 21-26. DOI: https://doi.org/10.1145/2157136.2157148
- Deniz, H. (2011). Examination of changes in prospective elementary teachers' epistemological beliefs in science and exploration of factors meditating that change, Journal of Science Education and Technology 20 (6): 750-760
- Drobnis, A. (2010). Girls in computer science: A female only introduction class in high school (Master's Thesis). Retrieved from http://mars.gmu.edu/bitstream/handle/1920/5806/AWD-Dissertation.pdf?sequence=1&isAllowed=y
- Enochs, L. G., & Riggs, I. M. (1990). Further development of an elementary science teaching efficacy belief instrument: A preservice elementary scale. School Science and Mathematics, 90: 695–706.

References

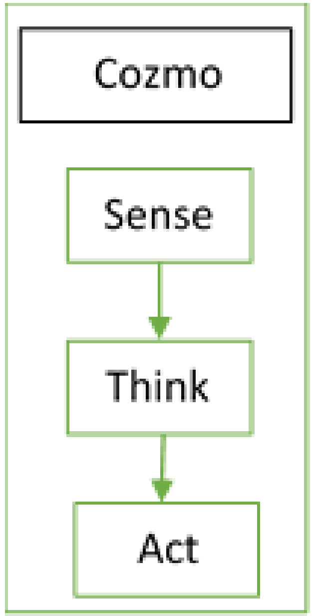
- Enochs, L. G., Smith, P. L. and Huinker, D. (2000), Establishing Factorial Validity of the Mathematics Teaching Efficacy Beliefs Instrument. School Science and Mathematics, 100: 194-202. doi:10.1111/j.1949-8594.2000.tb17256.x
- Eubanks V. (2018). Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. St. Martin's Press.
- Gardner-McCune C., Touretzky D. S., Martin F., and Seehorn D. (2019). AI for K-12: Making Room for AI in K-12 CS Curricula. Proceedings of the 50th ACM Technical Symposium on Computer Science Education (SIGCSE '19). ACM, New York, NY, USA, 1244-1244. DOI: https://doi.org/10.1145/3287324.3293729
- Goldman K., Gross P., Heeren C., Herman G. L., Kaczmarczyk L., Loui M. C., and Zilles C. (2010). Setting the Scope of Concept Inventories for Introductory Computing Subjects. Trans. Computer Education. 10 (2): 1:29. DOI=http://dx.doi.org/10.1145/1789934.1789935
- Hemment D., Aylett R., Belle V., Murray-Rust D., Luger E., Hillston J., Rovatsos M, & Broz F. (2019). Experiential AI. AI Matters. 5 (1): 25-31. DOI: https://doi.org/10.1145/3320254.3320264
- Kelley, T.; Sung, E. (2017). Examining Elementary School Students' Transfer of Learning through Engineering Design Using Think-Aloud Protocol Analysis, Journal of Technology Education, 28 (2): 83-108.
- Kier, M. W., Blanchard, M. R., Osborne, J. W., & Albert, J. L. (2014). The Development of the STEM Career Interest Survey (STEM-CIS). Research in Science Education. 44 (3): 461-481. 10.1007/s11165-013-9389-3.
- Lake, V. & Jones, I. (2008). Service-learning in Early Childhood Teacher Education: Using Service to Put Meaning Back into Learning. Teaching and Teacher Education, 4(8): 2146-2156
- Liao, H-T & Wang, Z. & Wu, X. (2019). Developing a Minimum Viable Product for Big Data and AI Education: Action Research Based on a Two-Year Reform of an Undergraduate Program of Internet and New Media. ICBDC: Proceedings of the 2019 4th International Conference on Big Data and Computing, 42-47. 10.1145/3335484.3335509.
- Mansour, N. (2009). Science teachers' beliefs and practices: issues, implications and research agenda. International Journal of Environmental and Science Education, 4(1): 25-48.
- Miyake Y. (2016). Square Enix AI academy: AI workshop for blackboard architecture. SIGGRAPH ASIA 2016 Symposium on Education: Talks (SA '16). ACM, New York, NY, USA, 10: 4. DOI: https://doi.org/10.1145/2993363.3006043
- Munley, M. E., & Rossiter, C. (2014). Girls, Equity, and STEM in Informal Learning Settings: A Review of Literature. National Science Foundation under Grant Number (DRL-1421427).
- New, J., and Castro D. (2018). How Policymakers can Foster Algorithmic Accountability, Center for Data Innovation, https://www.datainnovation.org/2018/05/how-policymakers-can-foster-algorithmic-accountability/

References

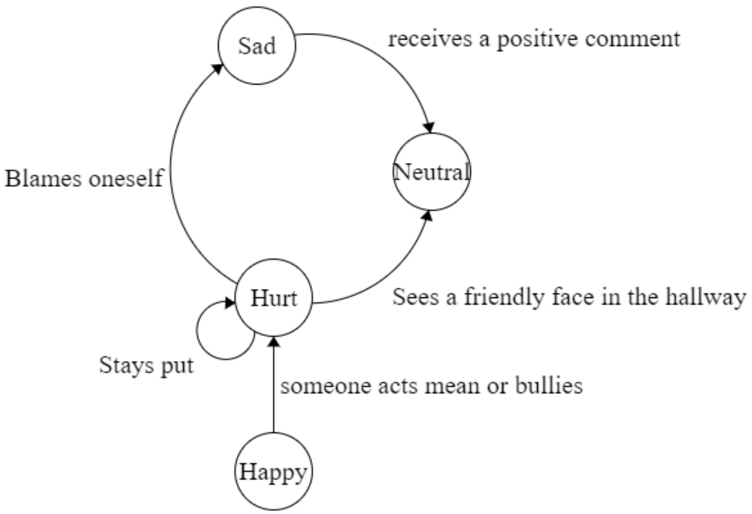
- Nonaka, I. (1991), The Knowledge-Creating Company, Harvard Business Review, 69: 96–104
- Novak, J. D., & Musonda, D. (1991). A twelve-year longitudinal study of science concept learning. American Educational Research Journal, 28(1): 117-153.
- Obermeyer, B. Powers, C. Vogeli, S. Mullainathan (2019), Dissecting racial bias in an algorithm used to manage the health of populations, Science. 366, 447.
- Pace, D. (2017). The Decoding of the Disciplines. Indiana University Press.
- Payne, B. H. (2019). An Ethics of Artificial Intelligence Curriculum for Middle School Students, with support from the MIT Media Lab Personal Robots Group, directed by Cynthia Breazeal.
- Peters D., Calvo R. A., & Ryan R.M. (2018). Designing for Motivation, Engagement and Wellbeing in Digital Experience, Frontiers in Psychology, 9 (797): 1-15.
- Pariser, E. (2011). The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think. Penguin Press.
- Robert, L. P. (2019). The Future of Pedestrian-Automated Vehicle Interactions, XRDS: Crossroads, 25(3): 30-33.
- Rose, M. A.; Carter, V.; Brown, J.; Shumway, S. (2017). Status of Elementary Teacher Development: Preparing Elementary Teachers to Deliver Technology and Engineering Experiences. Journal of Technology Education, 28 (2): 2-18.
- Samorani, M., Harris, S., Blount, L, Lu, H., & Santoro, M. A. (2019). Overbooked and Overlooked: Machine Learning and Racial Bias in Medical Appointment Scheduling. Available at SSRN 3467047.
- Tannu S. S. and Qureshi M. (2019), Ensemble of Diverse Mappings: Improving Reliability of Quantum Computers by Orchestrating Dissimilar Mistakes. (MICRO-52). MICRO-52, Columbus, OH, USA, https://dx.doi.org/10.1145/3352460.3358257
- Tarmo, A. (2018). Science teachers' beliefs and teaching practices in Tanzanian secondary schools, Cogent Education, 3(1): 1-20
- Touretzky D. S. and Gardner-McCune C. (2018). Calypso for Cozmo: Robotic AI for Everyone (Abstract Only). Proceedings of the 49th ACM Technical Symposium on Computer Science Education (SIGCSE '18). ACM, New York, NY, USA, 1110-1110. DOI: https://doi.org/10.1145/3159450.3162200
- Weinberg, A. E., Basile, C. G., & Albright, L. (2011). The effect of an experiential learning program on middle school students’ motivation toward mathematics and science. RMLE Online, 35(3), 1-12.
- Williams R., Park H. W., & Breazeal C. (2019). A is for Artificial Intelligence: The Impact of Artificial Intelligence Activities on Young Children's Perceptions of Robots. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19). ACM, New York, NY, USA, 447: 1-11. DOI: https://doi.org/10.1145/3290605.3300677
- Wellman H. M. and Liu D. (2004). Scaling of theory-of-mind tasks. Child development. 75 (2): 523–541
- You, S. and Robert, L. P. (2019). Trusting Robots in Teams: Examining the Impacts of Trusting Robots on Team Performance and Satisfaction, Proceedings of the 52th Hawaii International Conference on System Sciences, 244 – 253. Jan 8-11, Maui, HI. http://hdl.handle.net/10125/59465

Example: Finite State Machines

Description	Time
Students will discover that while Cozmo can sense and act, Cozmo's brain and thinking process can be modified by them using code blocks Calypso. Students are introduced to states, transition functions, and state diagrams. Students discuss states and transitions in daily life scenarios and are asked to build a state machine diagram for their emotions. Students start with a set of identical emotions, then compare their unique state machine diagrams with their peers.	~ 2 hrs



Emotional States	Hurt Sad Happy Neutral
Events	Someone acts mean or bullies Hears a friend's empathetic note Sees a friendly face in hallway Receives a positive comment Blames oneself with no grounds Stays put
Actions	Leaves the negative environment Responds and reasons Looks at the bully puzzled & shocked



States	Obstacle: block in sight Sound: sound audible
Events	Recognize which block Recognize
Actions	Move straight Change direction Say something