

# What Society Must Require from AI

©Ronald M. Baecker, 2019, 2020

MGIEP, 7 February 2020

1

## My background

- Interactive graphics and animation 1966-1987
- Software visualization 1972-1992
- HCI 1985-2005
- Multimedia, CSCW, CSCL 1990-2010
- Technology for seniors ... TAGlab 2002-2016
- *Computers and Society: Modern Perspectives* (2019)
  - Social, political/legal, ethical/moral issues



2



AI

- **Good Old Fashioned Artificial Intelligence (GOFAI)**
- **Machine learning**

3



GOFAI

- **Symbolic knowledge representations**
- **Inference engines**
- **Knowledge elicitation**
- **Expert systems**
- **Seems to have run out of steam**
  - **Difficult to create**
  - **Brittle**

4

Machine  
learning, esp.  
deep learning

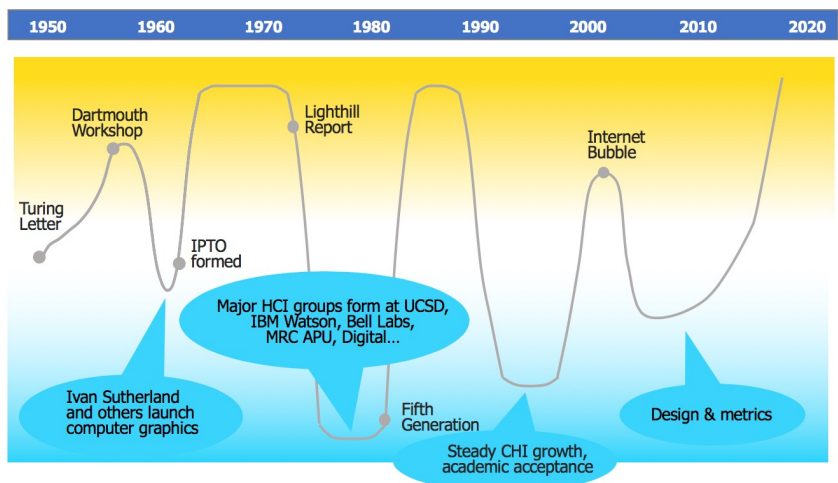
- Connectionist models, neural networks
- Recognizers often trained via supervised learning on big data
- Tens of billions of \$ now being invested annually worldwide

5

AI winters and  
summers ...  
figure courtesy  
of Jonathan  
Grudin



## HCI and AI Summer & Winters



6

## Shanghai invests \$15B

**TOP PICKS**

**News**  
"Serious concerns" over Meng case due to Trump's comments, says her lawyer  
7 Mar 2019

**Economy**  
Almost 1,000 Chinese cities are shrinking, satellite imagery reveals  
19 Mar 2019


**Economy**  
Taiwanese companies are turning their backs on China to return home  
21 Mar 2019

Companies

### Shanghai aims to raise US\$15 billion in funds to gain an upper hand in AI development

The city government is looking at how to draw investors to raise capital and talent to manage the funds in the quest to become a tech hub

**Daniel Ren**  
Published: 6:03am, 5 Jul, 2018



7

☰ 🔍 👤 Inc.
China Is Leading in Artificial Intelligence...
APPLY FOR THE INC. 5000 >

## China aims to be #1

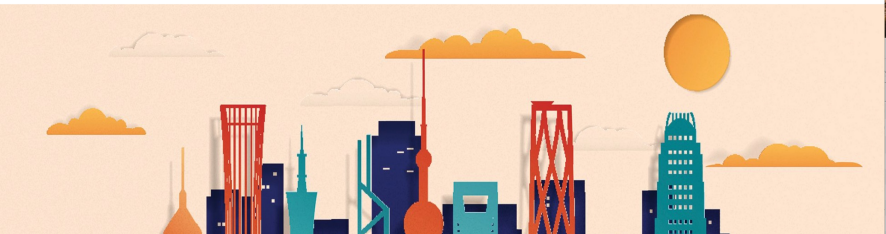
UP NEXT

### China Is Leading in Artificial Intelligence--and American Businesses Should Take Note

If you think of China as a country that copies rather than innovates--think again.

in f t

**By Amy Webb** Founder, Future Today Institute [@amywebb](#)



8

AI Day/Journée de l'IA Registration, Mon, 4 Mar 2019 at 8:30 AM ...
https://www.eventbrite.ca/e/ai-dayjournée-de-lia-registration-5528

## Canadian government enthused

Eventbrite, and certain approved third parties, use functional, analytical and tracking cookies (or similar technologies) to understand your event preferences and provide you with a customized experience. By closing this banner or by continuing to use Eventbrite, you agree. For more information please review our [cookie policy](#).

**eventbrite**
X Search for events
Browse Events
Help ▾
Sign In
Create Event

### AI Day/Journée de l'IA

Treasury Board Secretariat of Canada / Secrétariat du Conseil du Trésor du Canada  
 Monday, 4 March 2019 from 8:30 AM to 4:30 PM (EST)  
 Ottawa, ON

Registration Information			
TYPE	END	QUANTITY	
General Admission	Ended	Free	Sold Out

[Add Me to the Waitlist](#)

**Save This Event**

**When & Where**

9

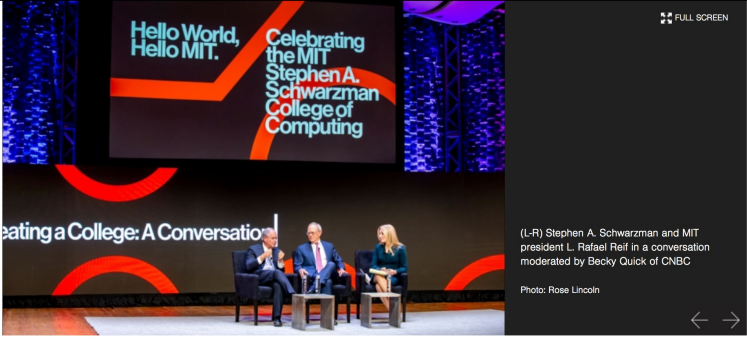
## New college at MIT

# MIT News

ON CAMPUS AND AROUND THE WORLD

Browse
or

Search



(L-R) Stephen A. Schwarzman and MIT president L. Rafael Reif in a conversation moderated by Becky Quick of CNBC  
 Photo: Rose Lincoln

### For founders of new college of computing, the human element is paramount

Stephen A. Schwarzman and MIT President L. Rafael Reif discuss the Institute's historic new endeavor.

Peter Dizikes | MIT News Office  
 March 1, 2019

Press Inquiries

PRESS MENTIONS

10

## 2019 Turing Award



From left to right: Yann LeCun | Photo: Facebook; Geoffrey Hinton | Photo: Google; Yoshua Bengio | Photo: Botler AI

11

## AIforK12 ???

### AI in K-12 Guidelines Initiative



- **Developing national guidelines for teaching AI in K-12**
  - Modeled after the CSTA standards for computing education.
  - What should students know?
  - What should students be able to do?
  - Across four grade bands: K-2, 3-5, 6-8, and 9-12
- **Develop a curated AI resource directory for K-12 teachers**
- **Foster a community of K-12 AI resource developers**

**It's time for all of us to think about AI in K-12.**

**Visit us**  
<http://AI4K12.org>

**Join the mailing list:**  
 Send mail to  
[ai4k12@aaai.org](mailto:ai4k12@aaai.org)

12

## AI's game playing achievements

- Checkers 1994
- Chess 1997
- Jeopardy 2011
- Go 2017
- Poker 2017
- Bridge?

13

## Not-so-consequential AI:

Speech (Siri)

Language (Google)

Image recognition

Research

JAMA | Original Investigation | INNOVATIONS IN HEALTH CARE DELIVERY

## Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs

Varun Gulshan, PhD; Lily Peng, MD, PhD; Marc Coram, PhD; Martin C. Stumpe, PhD; Derek Wu, BS; Arunachalam Narayanaswamy, PhD; Subhashini Venugopalan, MS; Kazumi Widner, MS; Tom Madams, MEng; Jorge Cuadros, OD, PhD; Ramasamy Kim, OD, DNB; Rajiv Raman, MS, DNB; Philip C. Nelson, BS; Jessica L. Mega, MD, MPH; Dale R. Webster, PhD

**IMPORTANCE** Deep learning is a family of computational methods that allow an algorithm to program itself by learning from a large set of examples that demonstrate the desired behavior, removing the need to specify rules explicitly. Application of these methods to medical imaging requires further assessment and validation.

**OBJECTIVE** To apply deep learning to create an algorithm for automated detection of diabetic retinopathy and diabetic macular edema in retinal fundus photographs.

Editorial pages 2366 and 2368

Supplemental content

14



## Siri being stupid

What's the meaning of life

42.

42 what

You were saying...

Sorry, I couldn't hear what you said.

Why don't you have knowledge that you don't understand

Oh, not again.

You've experienced this before

Seems like I cut you off. Can you please repeat that?

I enough for now

I don't know what that means. If you like, I can search the web for 'I enough for now'.

15

## Consequential AI

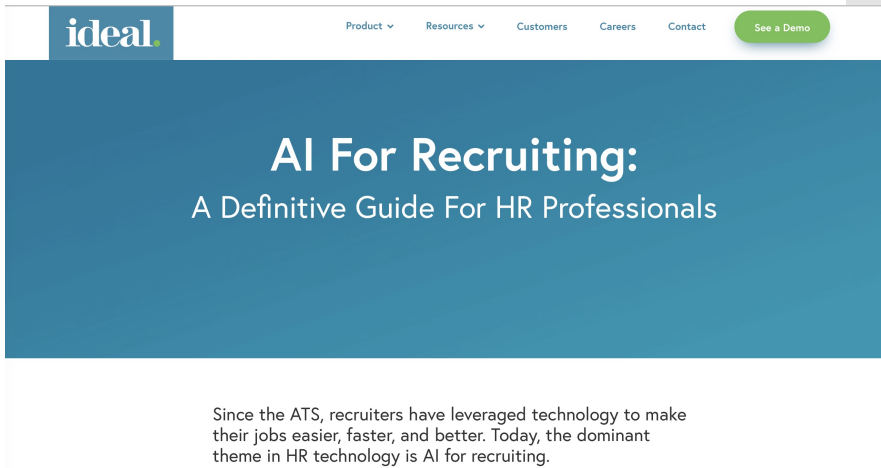
- Filtering résumés
- Informing judicial decisions
- Diagnosing medical conditions
- Recognizing faces in large crowds
- Caring for seniors
- Self-driving cars
- Drones, autonomous weapons
- Intelligent tutors

16



## Consequential AI

- Filtering résumés

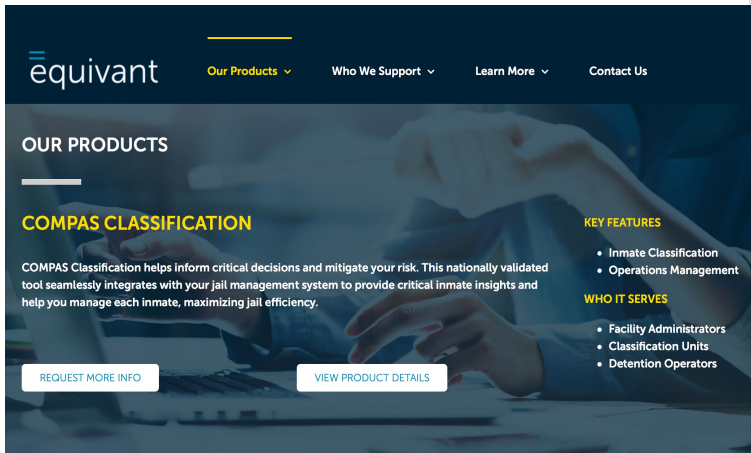


The screenshot shows the Ideal website's landing page. The header includes the 'ideal.' logo and navigation links: Product, Resources, Customers, Careers, and Contact. A green button labeled 'See a Demo' is in the top right. The main heading is 'AI For Recruiting: A Definitive Guide For HR Professionals'. Below this, a paragraph states: 'Since the ATS, recruiters have leveraged technology to make their jobs easier, faster, and better. Today, the dominant theme in HR technology is AI for recruiting.'

17

## Consequential AI

- Filtering résumés
- Informing judicial decisions



The screenshot shows the Equivant website's landing page for COMPAS Classification. The header includes the 'equivant' logo and navigation links: Our Products, Who We Support, Learn More, and Contact Us. The main heading is 'OUR PRODUCTS'. Below this, the 'COMPAS CLASSIFICATION' section describes the tool: 'COMPAS Classification helps inform critical decisions and mitigate your risk. This nationally validated tool seamlessly integrates with your jail management system to provide critical inmate insights and help you manage each inmate, maximizing jail efficiency.' To the right, 'KEY FEATURES' include Inmate Classification and Operations Management. 'WHO IT SERVES' includes Facility Administrators, Classification Units, and Detention Operators. At the bottom, there are two buttons: 'REQUEST MORE INFO' and 'VIEW PRODUCT DETAILS'.

18

## Consequential AI

- Filtering résumés
- Informing judicial decisions
- Diagnosing medical conditions

**IBM**

IBM Watson for Oncology

### IBM Watson for Oncology

Watson for Oncology helps physicians quickly identify key information in a patient's medical record, surface relevant evidence and explore treatment options.

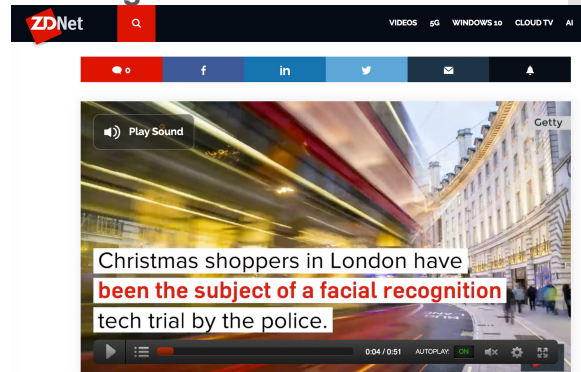
Watch the video (05:07)

Contact Us

19

## Consequential AI

- Filtering résumés
- Informing judicial decisions
- Medical diagnosis
- Recognizing faces in large crowds



London's police will be testing out live facial recognition technology on Christmas shoppers

20

## Consequential AI

- Filtering résumés
- Informing judicial decisions
- Medical diagnosis
- Recognizing faces
- Caring for seniors

The Telegraph

HOME | NEWS

### News

UK | World | Politics | Science | Education | Health | Brexit | Royals | Investigat

News

'Carebots' could take over from NHS medics to save £13bn a year

share

Save



21

## Consequential AI

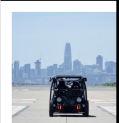
- Filtering résumés
- Informing judicial decisions
- Medical diagnosis
- Recognizing faces
- Caring for seniors
- Self-driving cars

WIRE

SELF DRIVING CARS



ROUNDUP  
How Self-Driving Has Changed Since Uber's Crash and More Car News  
ARIAN MARSHALL



LAWSUITS  
Tesla Sues Zoox Over Manufacturing and Logistics Secrets  
ARIAN MARSHALL



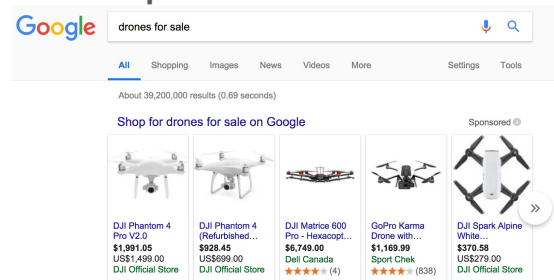
CONVOY  
A Cab's-Eye View of How Peloton's Trucks 'Talk' to Each Other  
ARIAN MARSHALL

ANNIVERSARIES  
1 YEAR AFTER UBER'S FATAL CRASH, ROBOCARS CARRY ON QUIETLY

22

## Consequential AI

- Filtering résumés
- Informing judicial decisions
- Medical diagnosis
- Recognizing faces
- Caring for seniors
- Self-driving cars
- Drones, autonomous weapons



23

## Consequential AI

- Filtering résumés
- Informing judicial decisions
- Diagnosing medical conditions
- Recognizing faces in large crowds
- Caring for seniors
- Self-driving cars
- Drones, autonomous weapons
- Intelligent tutors

24

## What's the difference?

- **Not-so-consequential AI**
  - Pattern recognition
  - ...
- **Consequential AI also needs**
  - Common sense
  - Understanding of context
  - Sensitivity to actors
  - Empathy, compassion
  - Sense of fairness, justice

25

## Outline of talk

- **Digital nightmares ... Lack of**
  - Identity, sensitivity, empathy, compassion
  - Competence, dependability, reliability, consistency
  - Openness, transparency, explainability
  - Fairness, justice, ethical behavior
  - Responsibility, accountability
  - Trustworthiness
- **Other nightmares**
  - Insensitive robot caregivers for seniors
  - Autonomous weapons insufficiently controllable

26

## Common definitions of intelligence

- Capacity to acquire and apply knowledge.  
(*American Heritage Dictionary, 4th ed., 2000*)
- Capacity for learning, reasoning, understanding, and similar forms of mental activity  
(*Random House Unabridged Dictionary, 2006*)
- Ability to acquire and apply knowledge and skills.  
(*Compact Oxford English Dictionary, 2006*)

27

## Yale cognitive psychologist Sternberg's Triarchic theory of intelligence

- Contextual: Purposive adaptation to, and selection and shaping of real-world environments relevant to one's life"
- Experiential: "Ability to deal with novel task and situational demands"

28

Harvard  
cognitive  
psychologist  
and social  
scientist  
Gardner's  
theory of  
multiple  
intelligences

- Typically measured in school, e.g., IQ
  - *Linguistic*
  - *Logical-mathematical*
- Useful in the arts
  - *Musical*
  - *Bodily-kinesthetic*
  - *Spatial*
- *Interpersonal* intelligence, i.e., understanding people's intentions, motivations, desires → ability to work effectively with others
- *Intrapersonal intelligence*, i.e., ability to understand oneself, with an effective working model of oneself → ability to use this effectively in life

29

Outline of talk

- Digital nightmares ... Lack of ...
  - **Identity, sensitivity, empathy, compassion**
  - Competence, dependability, reliability, consistency
  - Openness, transparency, explainability
  - Fairness, justice, ethical behavior
  - Responsibility, accountability
  - Trustworthiness
- Other nightmares
  - Insensitive robot caregivers for seniors
  - Autonomous weapons insufficiently controllable

30



Machines  
mistakenly  
thought to be  
human

- **Anthropomorphism**
  - Interacting with machines as if they were people
  - The uncanny valley

- **Empathy or just pure reason?**
  - Can an algorithm put itself in someone's shoes?
  - Can it feel compassion?

31

Outline of talk

- **Digital nightmares ... Lack of ...**
  - Identity, sensitivity, empathy, compassion
  - **Competence, dependability, reliability, consistency**
  - Openness, transparency, explainability
  - Fairness, justice, ethical behavior
  - Responsibility, accountability
  - Trustworthiness
- **Other nightmares**
  - Insensitive robot caregivers for seniors
  - Autonomous weapons insufficiently controllable

32

## AI algorithms making serious mistakes

- **ML systems today are**
  - Greedy
  - Brittle
  - Opaque
  - Shallow
- **They lack common sense**

33

## One should ask of an AI system that is being hyped ..... (Marcus and Davis)

- **What does the system actually do?**
- **How general is the result?**
- **Can I test my own examples?**
- **If someone claims that this is “better than people”, what people under what circumstances?**
- **Is this an advance toward general AI?**
- **How robust is the system?**

34

## Outline of talk

- **Digital nightmares ... Lack of ...**
  - Identity, sensitivity, empathy, compassion
  - Competence, dependability, reliability, consistency
  - **Openness, transparency, explainability**
  - Fairness, justice, ethical behavior
  - Responsibility, accountability
  - Trustworthiness
- **Other nightmares**
  - Insensitive robot caregivers for seniors
  - Autonomous weapons insufficiently controllable

35

## Algorithmic decisions not understood

- **How do we know what the machine knows?**
  - Does it know what it knows and does not know?
- **Can it explain its decisions lucidly?**
  - How would it explain itself?
- **Is its logic of its decision-making transparent?**
  - If it cannot explain itself, can we derive its logic?
  - Not as a display of millions of neuronal weights
- **DARPA now spending \$75M on these topics**

36

## Outline of talk

- **Digital nightmares ... Lack of ...**
  - Identity, sensitivity, empathy, compassion
  - Competence, dependability, reliability, consistency
  - Openness, transparency, explainability
  - **Fairness, justice, ethical behavior**
  - Responsibility, accountability
  - Trustworthiness
- **Other nightmares**
  - Insensitive robot caregivers for seniors
  - Autonomous weapons insufficiently controllable

37

## Unjust actions and decisions by algorithms

- Bias in “quantified selves” and in machine learning training data
- Can machines be fair and just?

38

## COMPAS: Correctional Offender Management Profiling for Alternative Sanctions

	White	Afro-American
<i>Labeled higher risk, but didn't re-offend</i>	24%	45%
<i>Labeled lower risk, but did re-offend</i>	48%	28%

39

## Outline of talk

- **Digital nightmares ... Lack of**
  - Identity, sensitivity, empathy, compassion
  - Competence, dependability, reliability, consistency
  - Openness, transparency, explainability
  - Fairness, justice, ethical behavior
  - **Responsibility, accountability**
  - Trustworthiness
- **Other nightmares**
  - Insensitive robot caregivers for seniors
  - Autonomous weapons insufficiently controllable

40

No  
responsibility or  
accountability  
for algorithmic  
decisions

- Mishaps are inevitable
- Who or what is responsible and accountable?
  - The programmers? The data scientists?
  - The firm responsible for the code and the data?
  - The individual or the firm that decided to use it?

41

<https://futurism.com/investing-lawsuit-ai-trades-cost-millions>

AI Vendor Is  
Sued, Being  
Held  
Responsible for  
Software  
Screwing Up

## Investor Sues After an AI's Automated Trades Cost Him \$20 Million

The first-of-its-kind case could shape the future of AI legislation.

[Kristin Houser](#) May 6th 2019

### Bad Investment

Back in 2016, Futurism [noted that](#) artificial intelligence was creeping into the world of finance, analyzing market data in real-time to autonomously execute trades for investors.

Now, Hong Kong real estate tycoon Samathur Li Kin-kan is suing a company that used one of those trade-executing AIs to manage his account, causing millions of dollars in losses — a first-of-its-kind court case that could help determine who should be held responsible when an AI screws up.

42

## Outline of talk

- Digital nightmares ... Lack of
  - Identity, sensitivity, empathy, compassion
  - Competence, dependability, reliability, consistency
  - Openness, transparency, explainability
  - Fairness, justice, ethical behavior
  - Responsibility, accountability
  - **Trustworthiness**
- Other nightmares
  - Insensitive robot caregivers for seniors
  - Autonomous weapons insufficiently controllable

43

## Use of AI software we do not trust

- When do we trust other people, beings, or things?
- How can machines demonstrate that they are trustworthy?
- What can we do if we don't trust them?

44

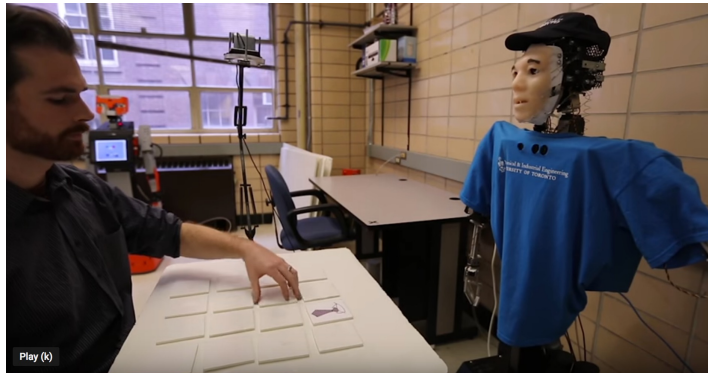


## Outline of talk

- Digital nightmares ... Lack of
  - Identity, sensitivity, empathy, compassion
  - Competence, dependability, reliability, consistency
  - Openness, transparency, explainability
  - Fairness, justice, ethical behavior
  - Responsibility, accountability
  - Trustworthiness
- Other nightmares
  - **Insensitive robot caregivers for seniors**
  - Autonomous weapons insufficiently controllable

45

## Insensitive robot caregivers for older adults



46

## Outline of talk

- Digital nightmares ... Lack of
  - Identity, sensitivity, empathy, compassion
  - Competence, dependability, reliability, consistency
  - Openness, transparency, explainability
  - Fairness, justice, ethical behavior
  - Responsibility, accountability
  - Trustworthiness
- Other nightmares
  - Insensitive robot caregivers for seniors
  - **Autonomous weapons insufficiently controllable**

47

## Autonomous weapons insufficiently controllable



48

## Summary

- **Digital nightmares ... Lack of ...**
  - Identity, sensitivity, empathy, compassion
  - Competence, dependability, reliability, consistency
  - Openness, transparency, explainability
  - Fairness, justice, ethical behavior
  - Responsibility, accountability
  - Trustworthiness
- **Other nightmares**
  - Insensitive robot caregivers for seniors
  - Autonomous weapons insufficiently controllable

49

## Intelligent tutors for social and emotional learning

- AI not ready to do this
- Suggest a focus on teacher training

50

## Some possible objections

- **Perfection is unrealistic and cannot be expected**
- **Humans should oversee AIs**

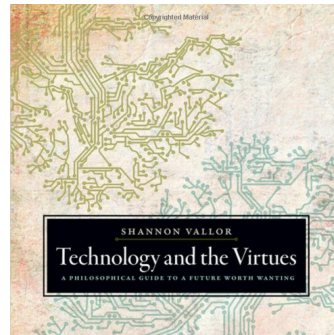
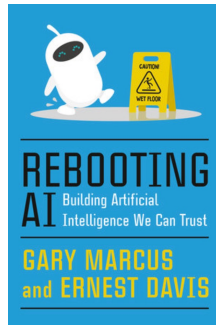
51

## See also these initiatives

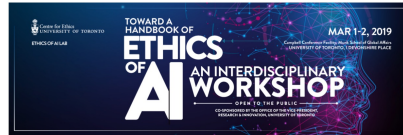
- AAAI/ACM Conference on AI, Ethics and Society
- Fairness, Accountability, and Transparency in Machine Learning Annual Conference
- IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems
- EU Ethics Guidelines for Trustworthy AI
- Council for Big Data, Ethics, and Society
- UN AI for Good Global Summit
- The Montreal Declaration for a Responsible Development of Artificial Intelligence
- AI Now Institute (NYU)
- Ethics and Governance of AI Initiative (MIT Media Lab, Harvard Berkman-Klein Center for Internet and Society)
- One Hundred Year Study on Artificial Intelligence (Stanford University)

52

And some new books ...



The Oxford Handbook of Ethics of AI: Online Supplement



53

Thank you for  
your attention!!!

- Thanks to Saul Greenberg, Jonathan Grudin, Leslie Mezei, Cosmin Munteanu, and Sheila McIlraith for comments on this talk
- “A word uttered from a pure heart never goes in vain” (Gandhi)
- Now it is your turn ... questions and discussion !!

54