

Microsoft AI: Intro and Ethics

Ben Haklai, Adv.
Commercial (Legal) Lead – Microsoft Israel

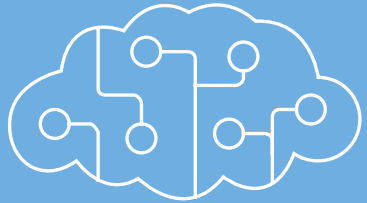






AI, Machine Learning and Deep Learning

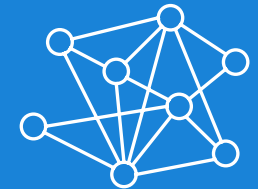
Artificial
Intelligence



Machine
Learning



Deep
Learning



Breakthroughs Enabling Intelligence



Big Data



Powerful
Algorithms



Cloud
Computing

Computers Understanding the World



Deep learning advancements

96%

RESNET vision test
152 layers

5.1%

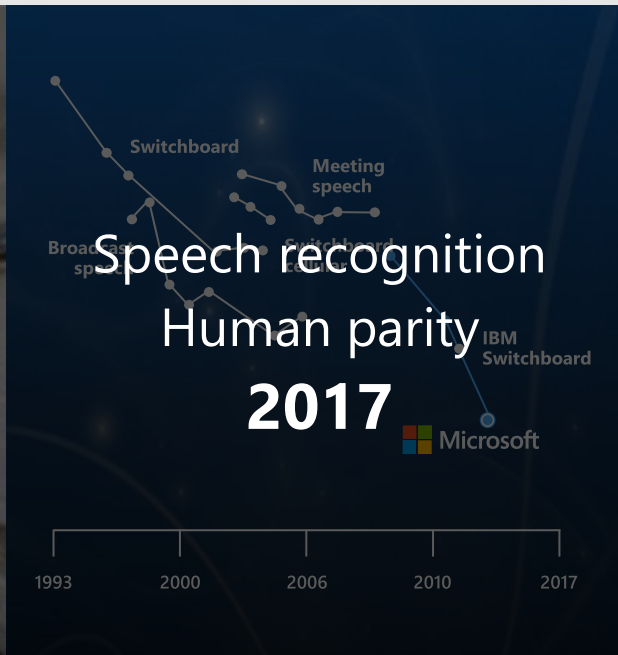
Switchboard speech
recognition test

88.493%

SQuAD reading
comprehension test

69.9%

MT
research system



Microsoft CaptionBot June 19, 2016

I am not really confident, but I think it's a couple of brown bears standing next to a cat.



Microsoft CaptionBot June 20, 2016

I think it's a close up of a cat looking at a dog.



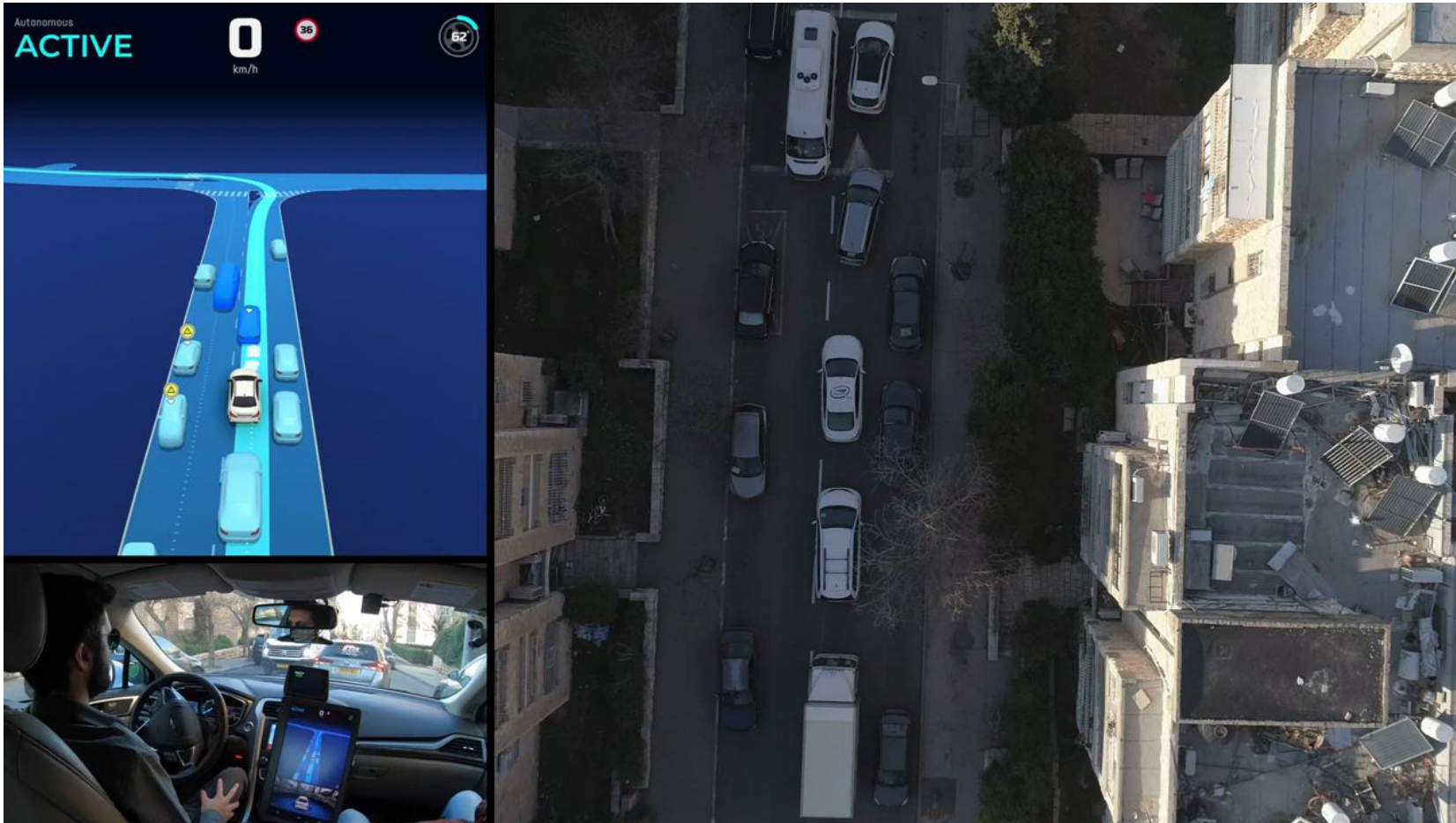
Microsoft CaptionBot June 11th 2018



I think it's a brown and white dog looking at the camera.



Where are we today?



Design AI to Earn Trust



Fairness



Reliability
& Safety



Privacy &
Security



Inclusiveness



Transparency



Accountability

Fairness

Understand how bias can be introduced and affect recommendations

Attract diverse pool of AI talent

Develop analytical techniques to detect and mitigate bias

Human review and domain expertise

The New York Times

Facial Recognition Is Accurate, if You're a White Guy

By STEVE LOHR FEB. 9, 2018



Gender was misidentified in up to 1 percent of lighter-skinned males in a set of 385 photos.

Gender was misidentified in up to 12 percent of darker-skinned males in a set of 318 photos.



Gender was misidentified in up to 7 percent of lighter-skinned females in a set of 296 photos.

Gender was misidentified in 35 percent of darker-skinned females in a set of 271 photos.

Why **diversity**
is an even bigger
imperative



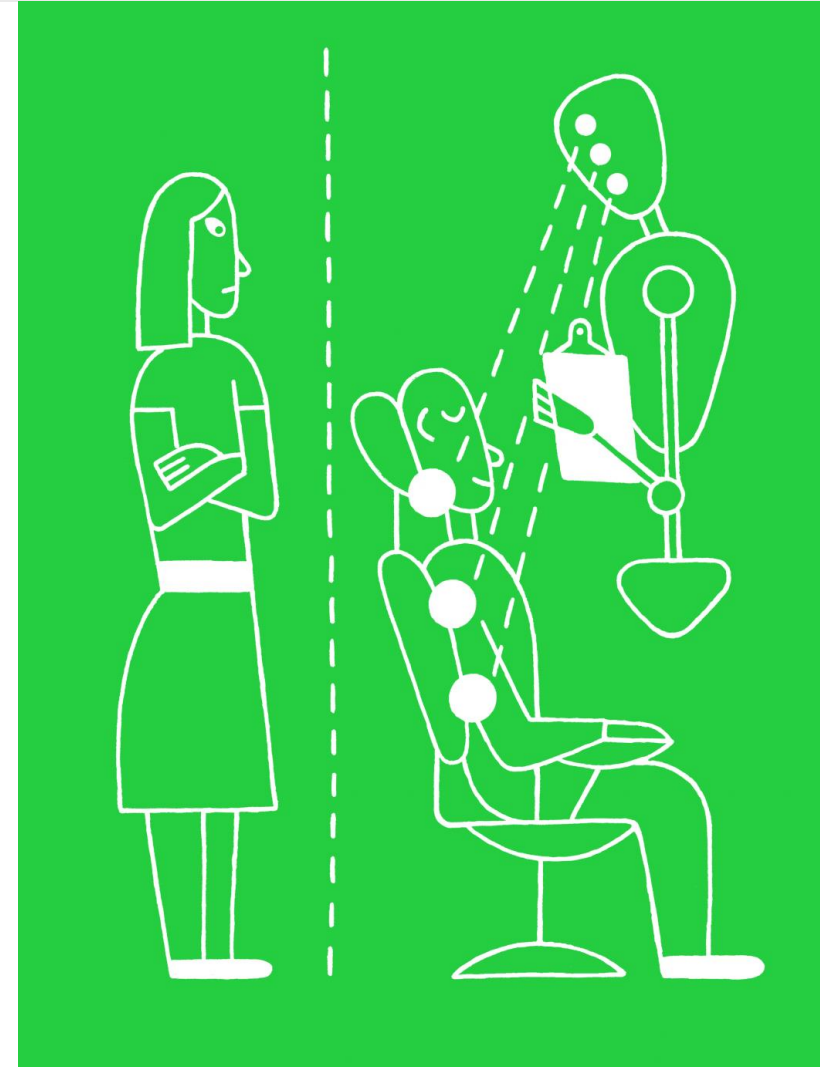
The Dangers of Tech-Bro AI

Tabitha Goldstaub, a cofounder of [CognitionX](#), which helps companies deploy AI, says that diversifying the field is necessary to make sure products actually work well.

October 24, 2017

If you think about artificial intelligence and the core components of what makes an AI, you have the fact that it's given a goal and then will find a way to reach that goal. It's then often very un-transparent as to how it reached that goal. If you're building into a machine some unconscious bias, you might not know that it's there; the output could be detrimental to women and it's very tough to work out exactly why that has happened.

In traditional technology, you can see [what has happened]: **women**



MIGUEL PORLAN

Reliability & Safety

Evaluate [training data](#)

[Test](#) extensively (and enable a user feedback loop)

[Monitor](#) ongoing performance

Design for [unexpected](#) circumstances—including nefarious [attacks](#)

[Human](#) in the loop



Privacy & Security

Existing privacy laws (e.g. the [General Data Protection Regulation](#)) apply

Provide [transparency](#) about data collection and use, and good [controls](#) so people can make choices about their data

Design systems to [protect](#) against bad actors

Use [de-identification techniques](#) to promote both privacy and security



Inclusiveness

Inclusive design practices to address potential barriers that could unintentionally exclude people

Enhances opportunities for those with disabilities

Build trust through contextual interaction

EQ in addition to IQ



Transparency

People should **understand** how decisions were made

Provide **contextual explanations**

Make it easier to raise awareness of **potential bias**, errors and unintended outcomes



Accountability

People must be **accountable** for how their systems **operate**

Norms should be observed during **system design** and in an **ongoing** manner

Role for **internal review boards**



Ethical AI Decision Framework

Four questions we should always ask

- 1 System Purpose**

Will the technology be used to augment the abilities of individuals and result in a positive impact on people and society? If no, seek guidance.
- 2 Technology Capability**

Are the AI technology and people capable of performing the tasks that the AI system should do? If no, seek guidance.
- 3 Quality & Reliability**

Will the technology be effectively designed, operated and maintained by a responsible party? If no, seek guidance.
- 4 Sensitive Uses**

Is the system in a sensitive category which may have a broad impact on people and society? If yes, seek guidance.

Sensitive use categories

At Microsoft, we recognize that there are sensitive categories of Artificial Intelligent systems which are involved in making automated decisions and recommendations that may have a broad impact on people and society in the following ways:



Denial of Consequential Services



Risk of Harm



Personal Freedom

Denial of Consequential Services

Artificial Intelligent systems which are involved in making automated decision or recommendations that may directly result in the denial of consequential services or support.



Risk of Harm

Artificial Intelligent systems which are involved in making automated decisions or recommendations where these systems may create a significant risk of physical or emotional harm.



Personal Freedom & Privacy

Artificial Intelligent systems which are involved in making automated decisions or recommendations which create a significant restriction of personal freedom or privacy.



Ensuring AI Empowers Us All



Staying true to **timeless values**

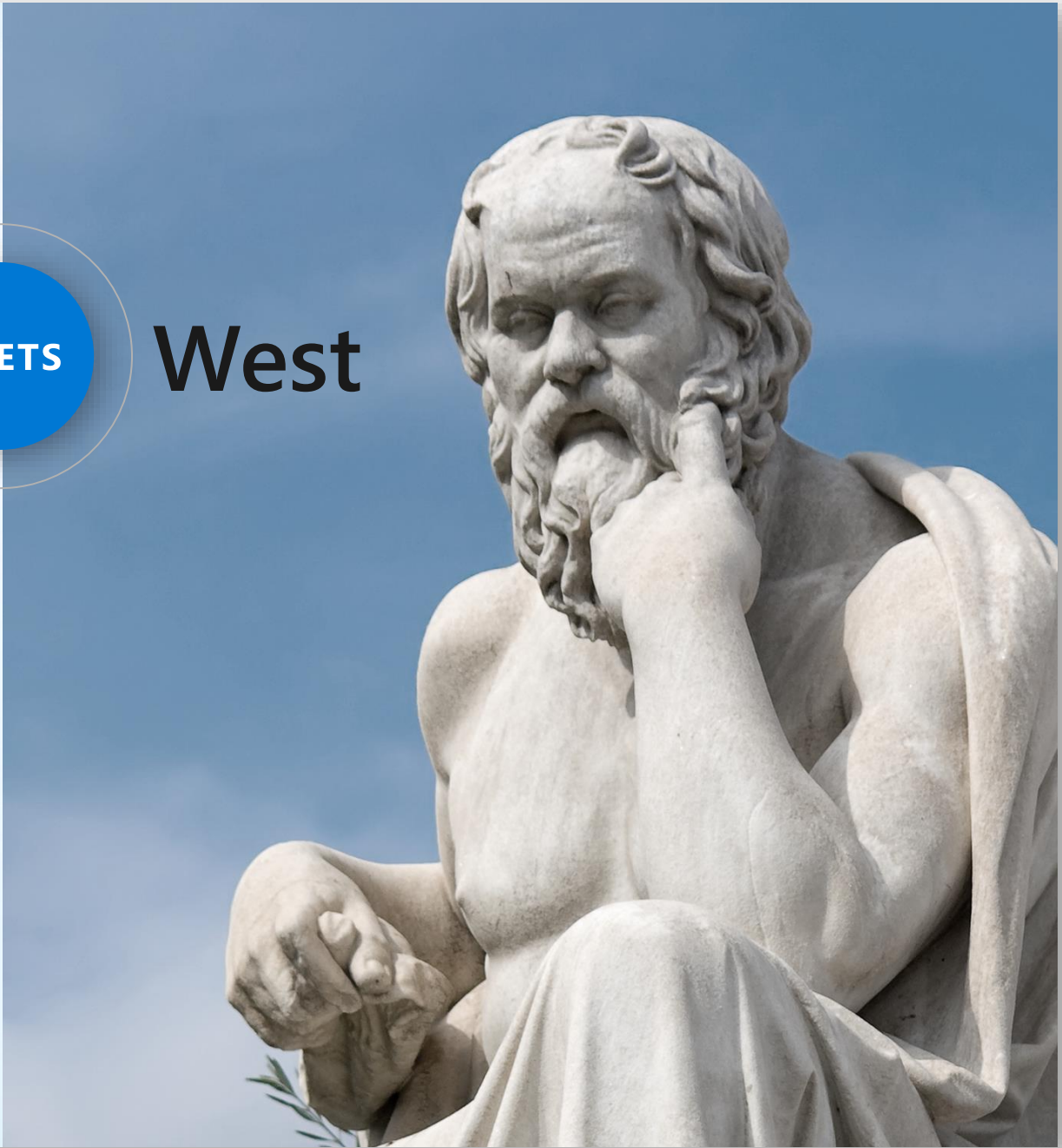
But whose values?



East



West



Thank You!

Ben Haklai @ LinkedIn

[linkedin.com/in/benhaklai](https://www.linkedin.com/in/benhaklai)

