

A wireframe model of a human head and neck, rendered in a light blue/white mesh. The head is positioned on the right side of the frame, facing left. The background is a dark blue-grey with a pattern of glowing blue squares and lines, suggesting a digital or data environment. In the top right corner, there is a small logo consisting of three vertical bars of increasing height, followed by the text "DISTILLED ANALYTICS".

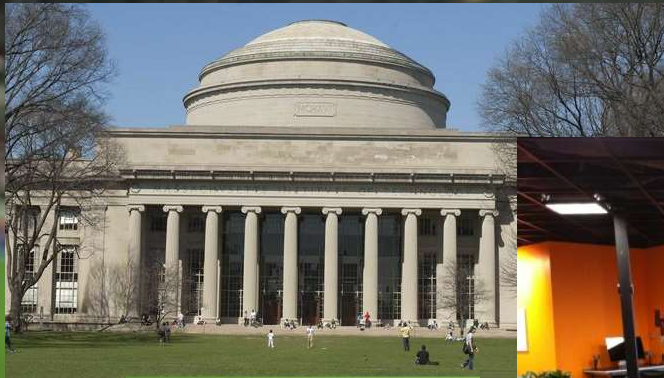
# Ethics of AI

## Considerations for Theory and Practice

**David Shrier**

david@DistilledAnalytics.com  
October 2018

# An interesting seat at the game



 **Massachusetts  
Institute of  
Technology**



**(THIS TALK IS MY PERSONAL OPINION)**

A dramatic, high-contrast image of a nuclear explosion. A massive, bright white and yellow mushroom cloud rises from a city, with intense orange and red flames and smoke billowing from the base. The sky is filled with swirling, fiery patterns. The overall tone is apocalyptic and urgent.

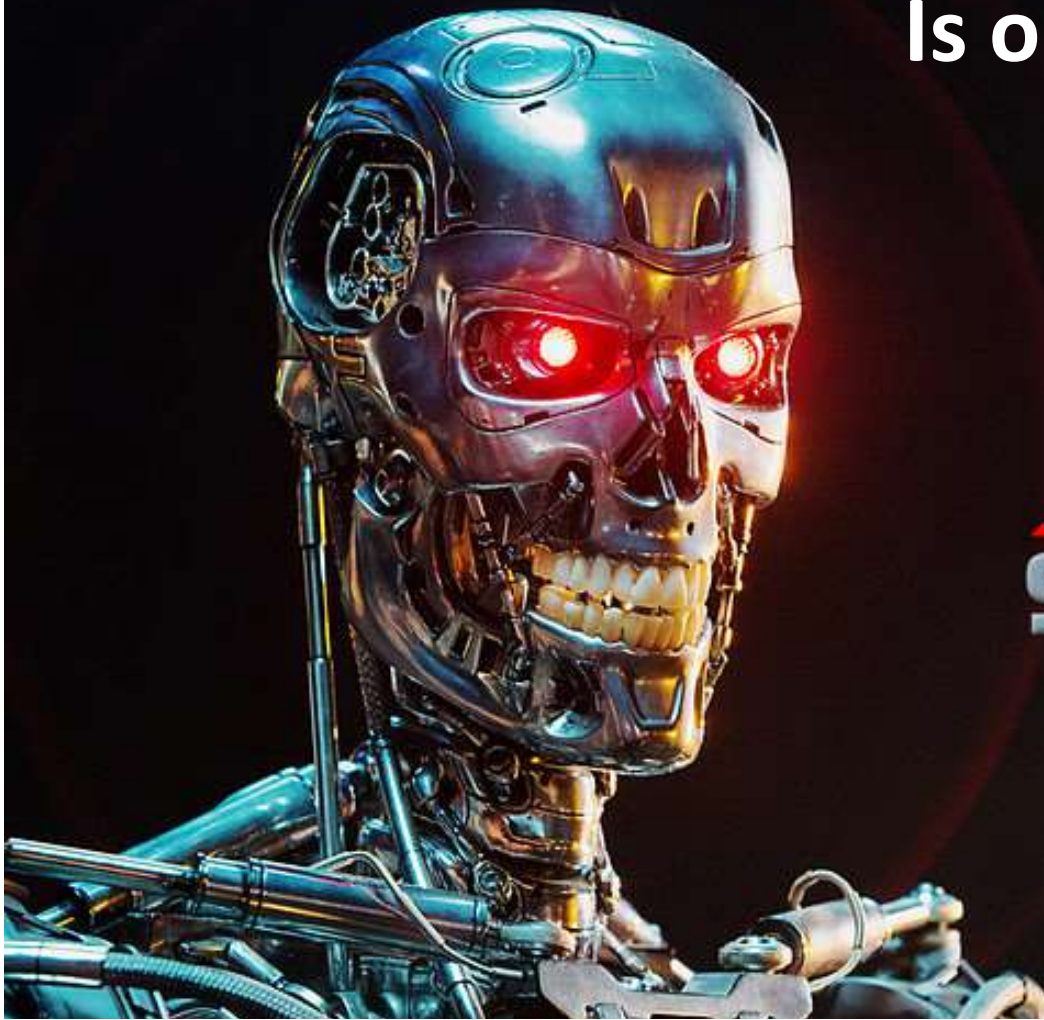
**30-50%**  
**of jobs lost**

Source: PwC; Kai-Fu Lee

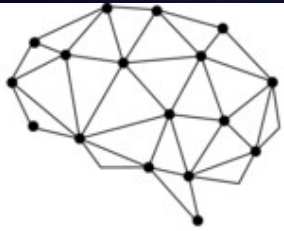
**“Most likely cause of World War III”**



Is our destiny written?



# Weaponization of AI



Cambridge  
Analytica

facebook®

2016 “fake news”  
impressions?

1bn impressions to  
150+m U.S. voters

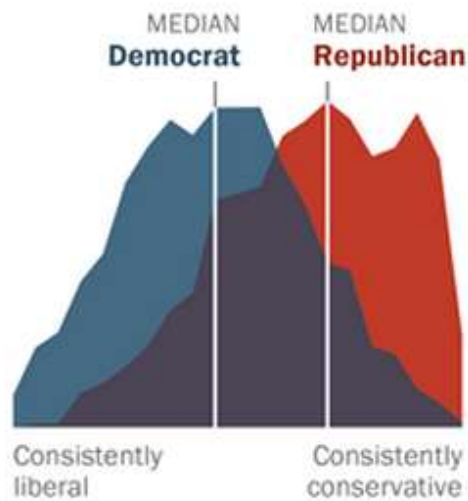
Source: The Verge



# Polarization of the Electorate

*Distribution of Democrats and Republicans on a 10-item scale of political values, by level of*


**Among the politically engaged**  
**1994**



Source: Pew Research

facebook.

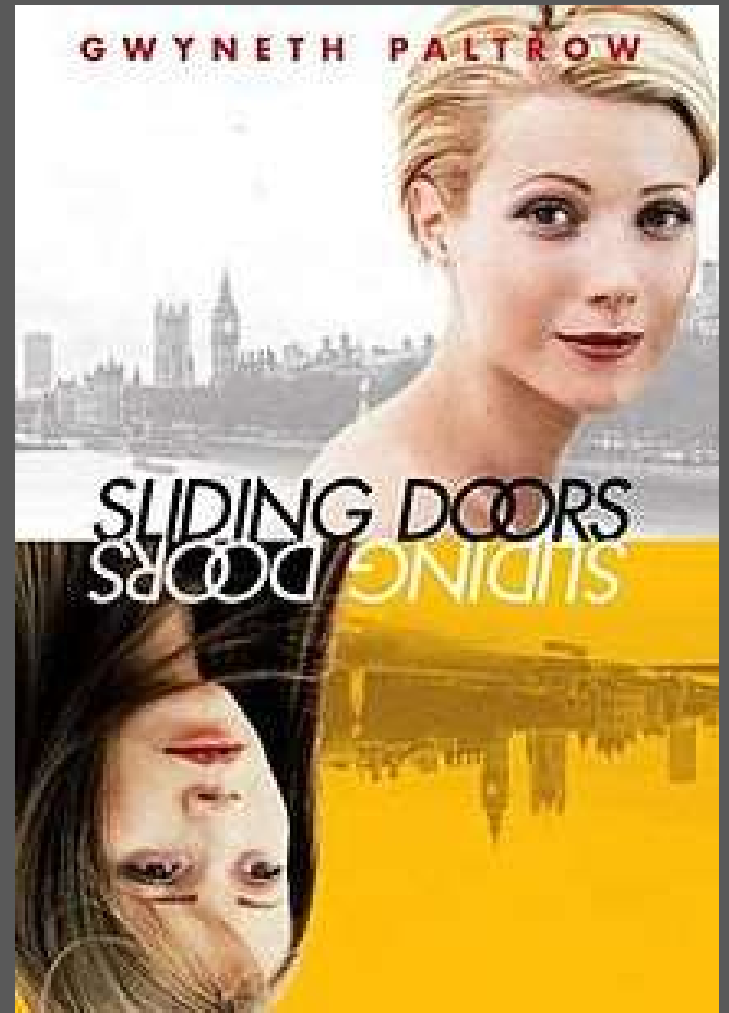


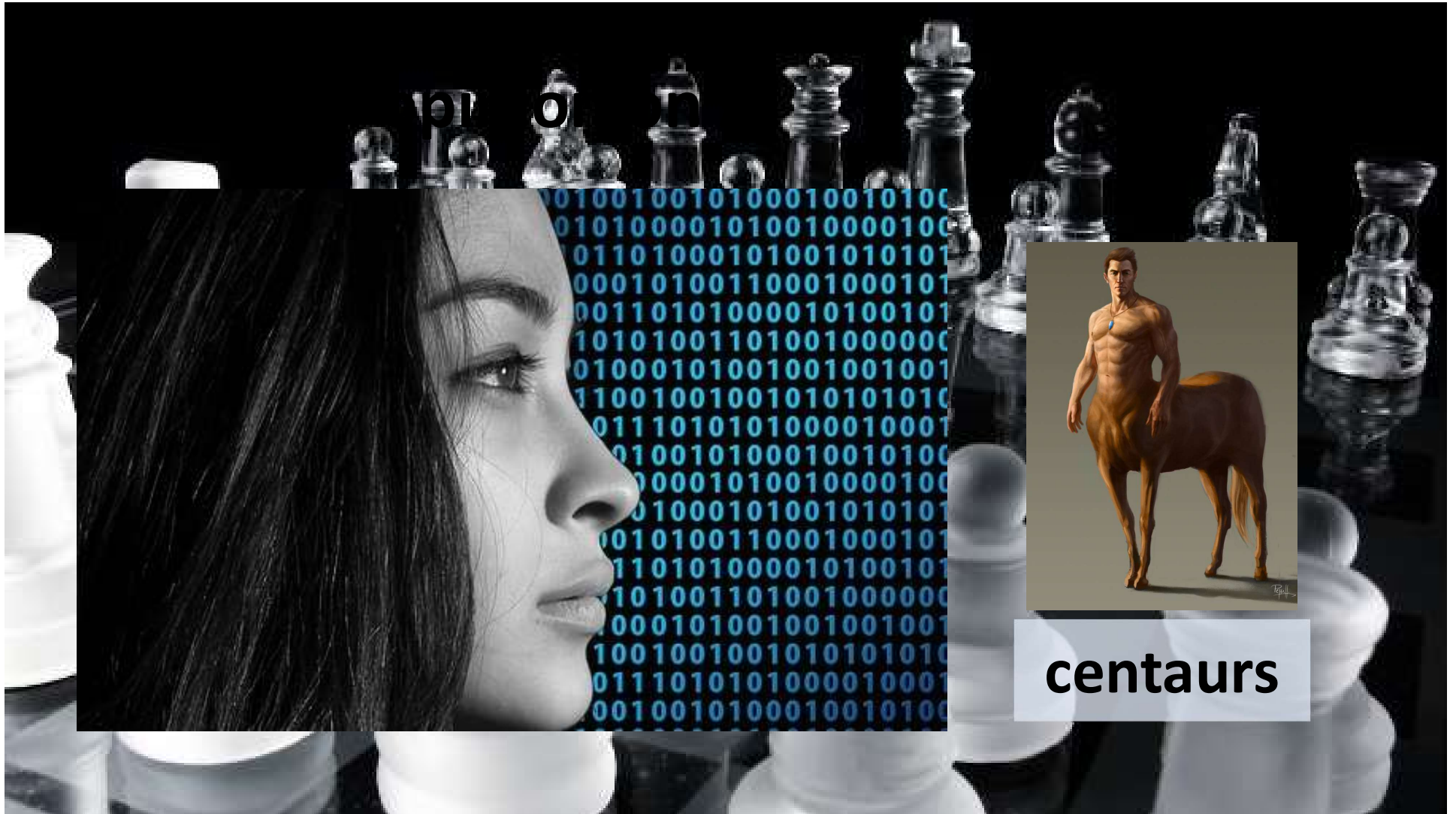
- 
- Large financial institution
  - Adopted AI, but didn't supervise closely
  - Great results, until...

**TECH AUDIT!**



# ANOTHER REALITY





centaurs

# Context for the Revolution

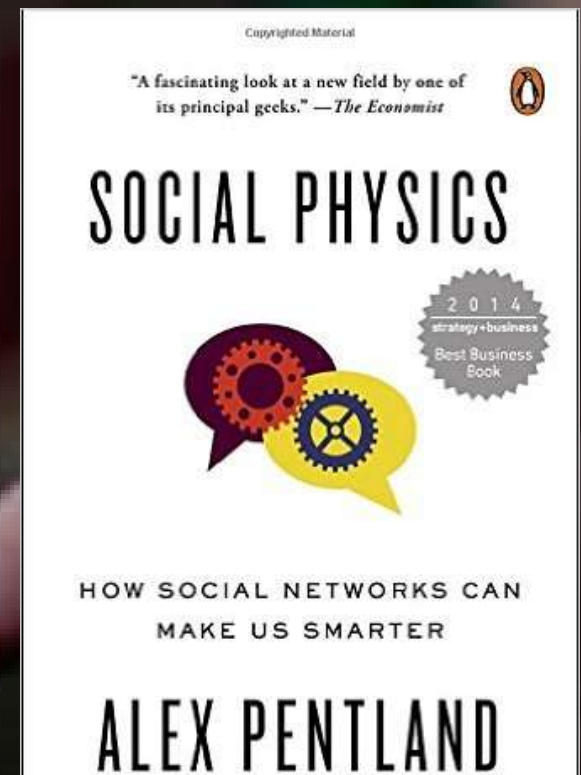




# The Cyborg Collective (1996)



# Ubiquitous Sensing



# Better than other predictive models

215+% better at behavior prediction  
than social media driven models

Predict behavior



Random Guess

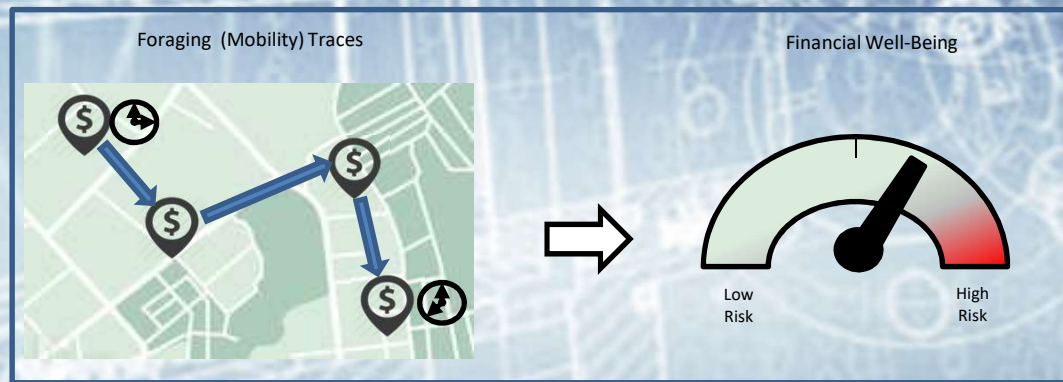
Source: Wei Pan, Nadav Aharony, Alex Pentland (2011) "Composite Social Network for Predicting ..." arXiv:1106.0359 [cs.SI]



WHAT IF...?



# What if we could predict financial behaviors?



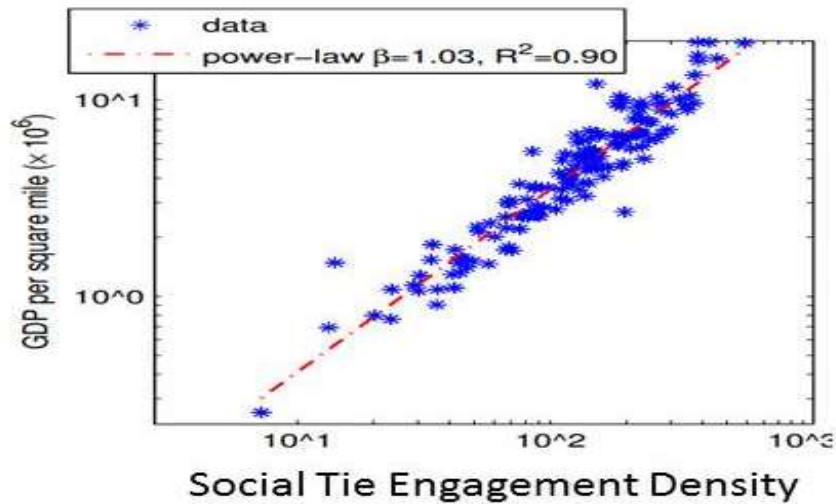
Singh, Bozkaya, Pentland, PLoS One 2015

**+30-50%  
better  
credit**

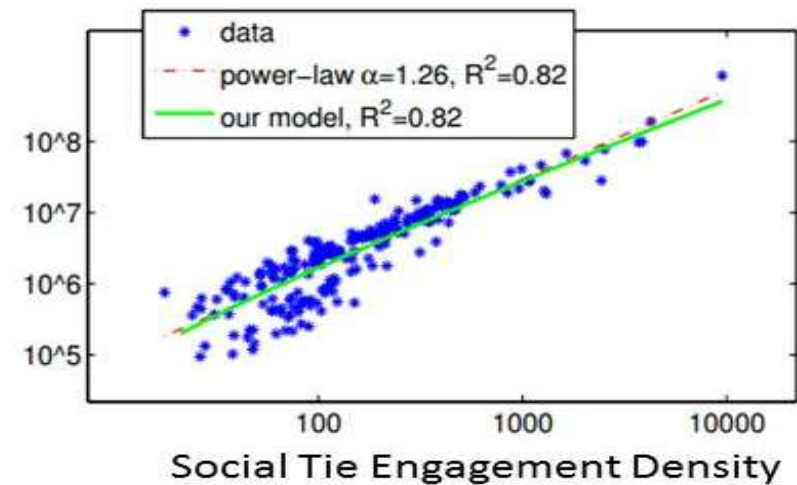
## Promote savings?



# What if we understood GDP better?



GDP per Square KM (in Euro)



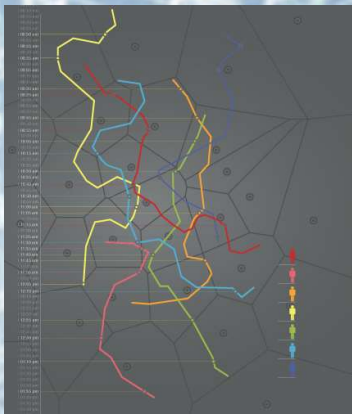
# Predict ROI of infrastructure?



# What if AML/KYC were better?

ID bad actors: from  
2% to 99.9+%

False positives:  
from 80+% to <1%



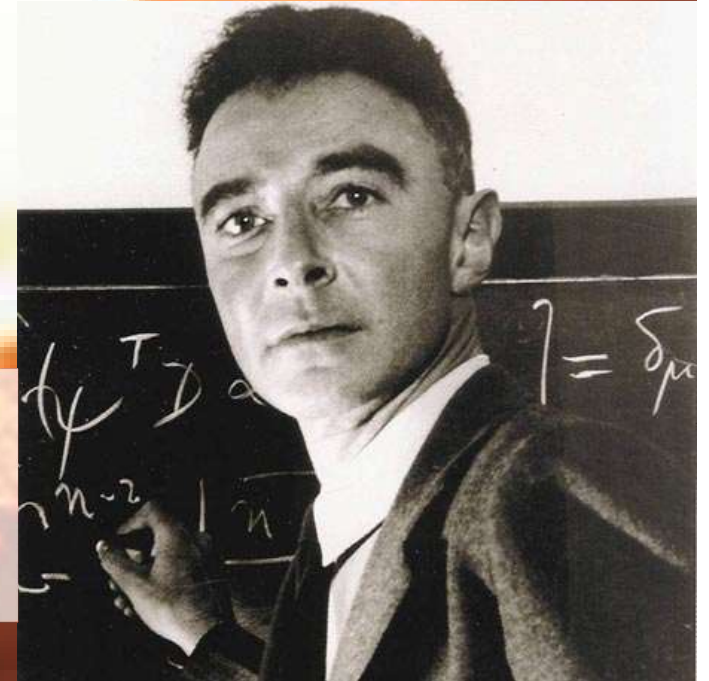
“Instant KYC?”



**SOME REFLECTION**

**“I am become death, the destroyer of worlds”**

J. Robert Oppenheimer  
*Inventor of the  
atomic bomb*





**“Dumb f--ks”**

**facebook®**

**Mark Zuckerberg**  
*Founder of Facebook*  
*regarding his users*



**KEEP  
CALM  
AND  
OBEY  
THE  
THREE  
LAWS**



## **“Three Laws of AI?”**

- 1) An AI may not injure a human being or, through inaction, allow a human being to come to harm.**
- 2) An AI must obey orders given it by human beings except where such orders would conflict with the First Law.**
- 3) An AI must protect its own existence as long as such protection does not conflict with the First or Second Law.**



A marble statue of a bearded man, likely a philosopher, in a thinking pose (The Thinker). The statue is white and set against a solid blue background. The man has a full beard and curly hair, and his right hand is raised to his chin.

**The ethics of AI are  
the ethics of the user**

**...so we need to embed  
ethics in our daily work**