

<https://technologyandsociety.org/wp-content/uploads/milleropen-1024x577.jpg>

# Traceability, Responsibility, and Ethics in a Hybrid Society

Ray Gallon



@TransformSoc

@RayGallon

ray@transformation society.net

# THE FOURTH INDUSTRIAL REVOLUTION

Or “Industry 4.0” – Extended to other technologies - smart houses, driverless cars, etc.

The 4th Industrial Revolution - „Industry 4.0“

*Many technological innovations  
have claimed that they will  
“change everything.”  
This one really will.*

1782

Power generation

Mechanical automation

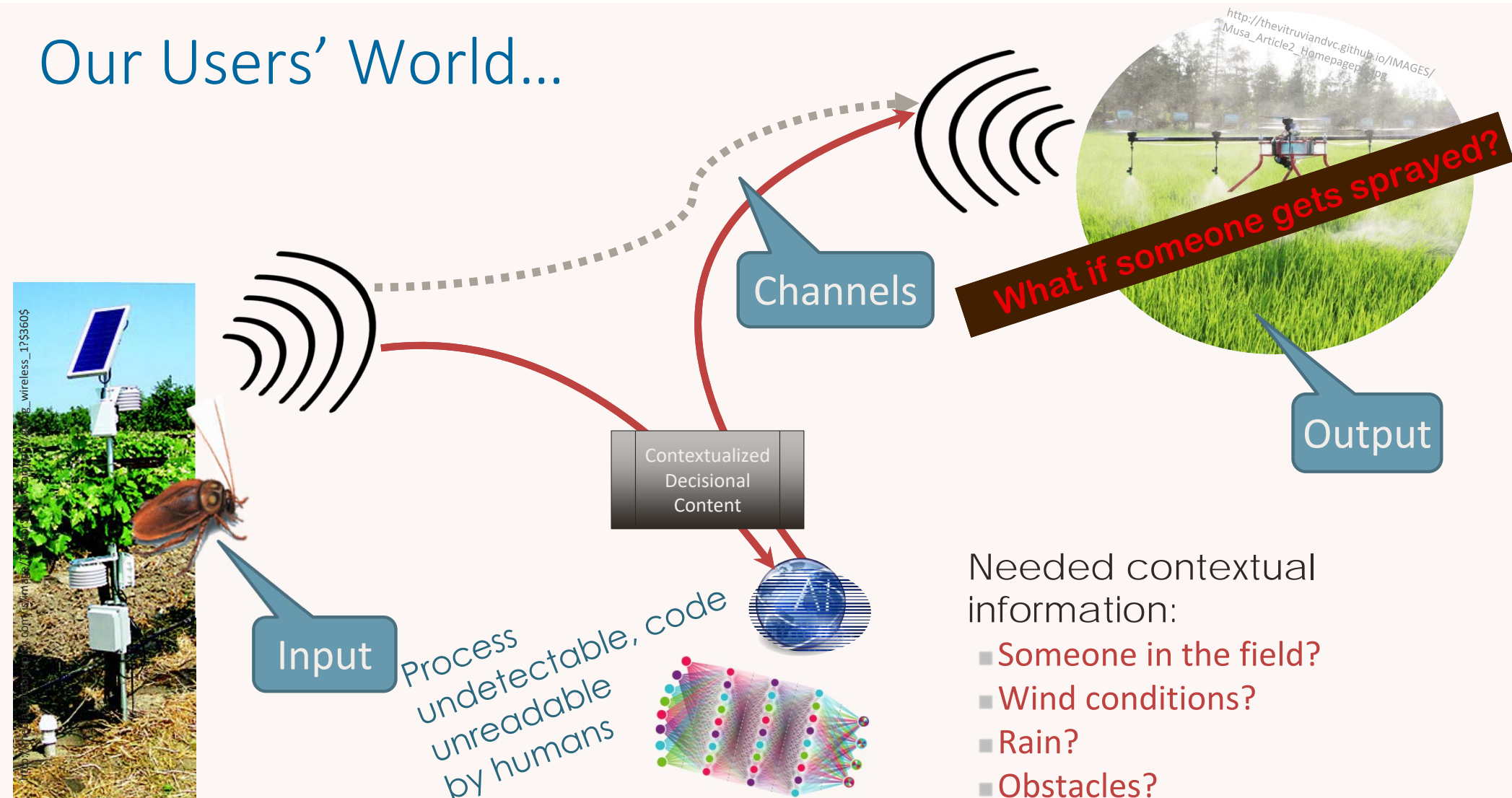
Industrialization

Automation

<http://www.industrialgiletsolutions.com/wp-content/uploads/2015/03/Industry-4.0.png>

- 21<sup>st</sup> Century: Machines decide for humans

# Our Users' World...



What if someone gets sprayed?



## Legal and Social Consequences

- **Traceability**
- **Readability**
- **Responsibility**
- **Accountability**
- **Liability**





Bernard Marr  
Jan 17, 2020

# AI and the Future of Law

- “In the future, is it conceivable that a firm would be charged with legal malpractice if they didn't use artificial intelligence (AI)? It certainly is.”  
-Richard Susskind, OBE
- **First Generation:** Extended courts where we go beyond decisions made by judges to a diagnostic system: guide people – how to assemble evidence and provide alternative ways for dispute resolution.
- **Second Generation:** “*Outcome thinking*” – use technology to help solve disputes without requiring lawyers or the traditional court system. Use AI to predict the outcomes of court decisions based on past decisions by using predictive analytics.

“Imagine if people had the option instead of waiting for a court date (and support from the traditional legal system) to use a machine-learning system to make a prediction about the likely outcome of a case and then accept that as a binding determination.”

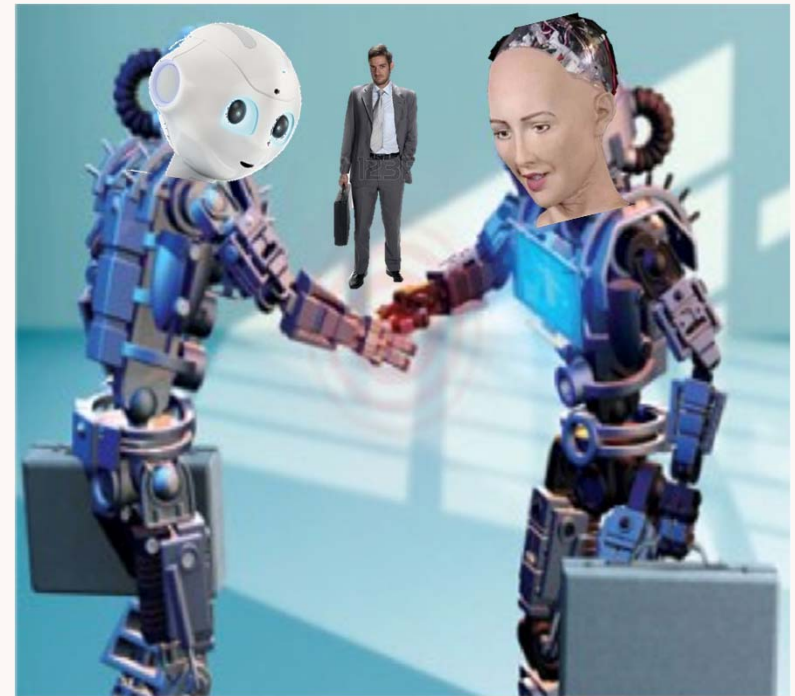
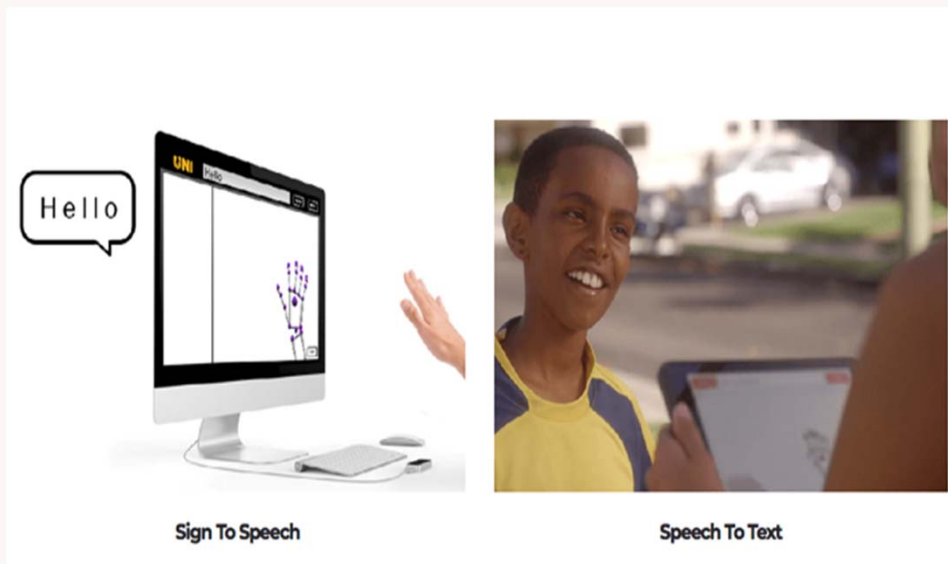
# It's an Increasingly Complex World





# Hybrid Communication

Interactions with tools change  
relationships  
between people



And between people and tools

# Hybrid Legal Practice

- Law is “poised to benefit greatly” from technology because “document-intensive.” AI can take over:
  - Document drafting
  - Document analysis
  - Outcome prediction
- New roles for lawyers:
  - Develop the systems that will solve clients' problems
  - Legal knowledge engineers
  - legal risk managers
  - System development
  - Experts in design thinking

*-Richard Susskind, OBE*



# Hybrid Society



Double amputee controls prostheses with brain waves

Portable Brain-computer interfaces can control just about anything – medical devices, TV's, home heating, computers, or game consoles...

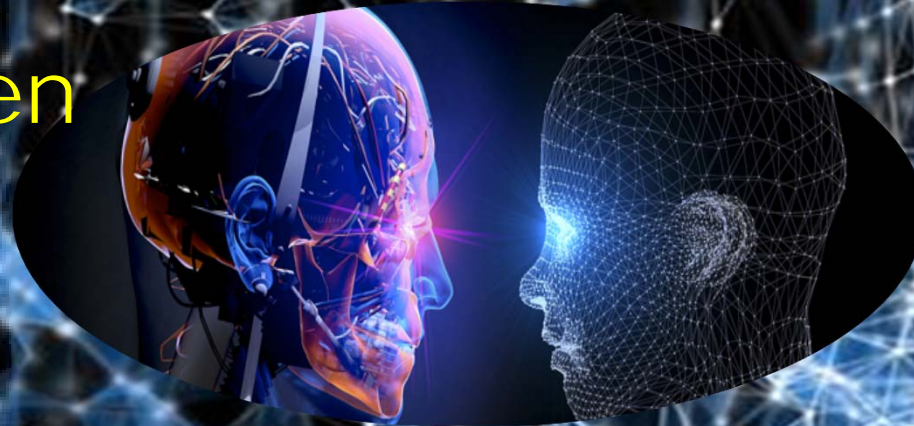


Brain wave controlled drones



# Are we Ready to Interact in a Hybrid Human-Machine Society?

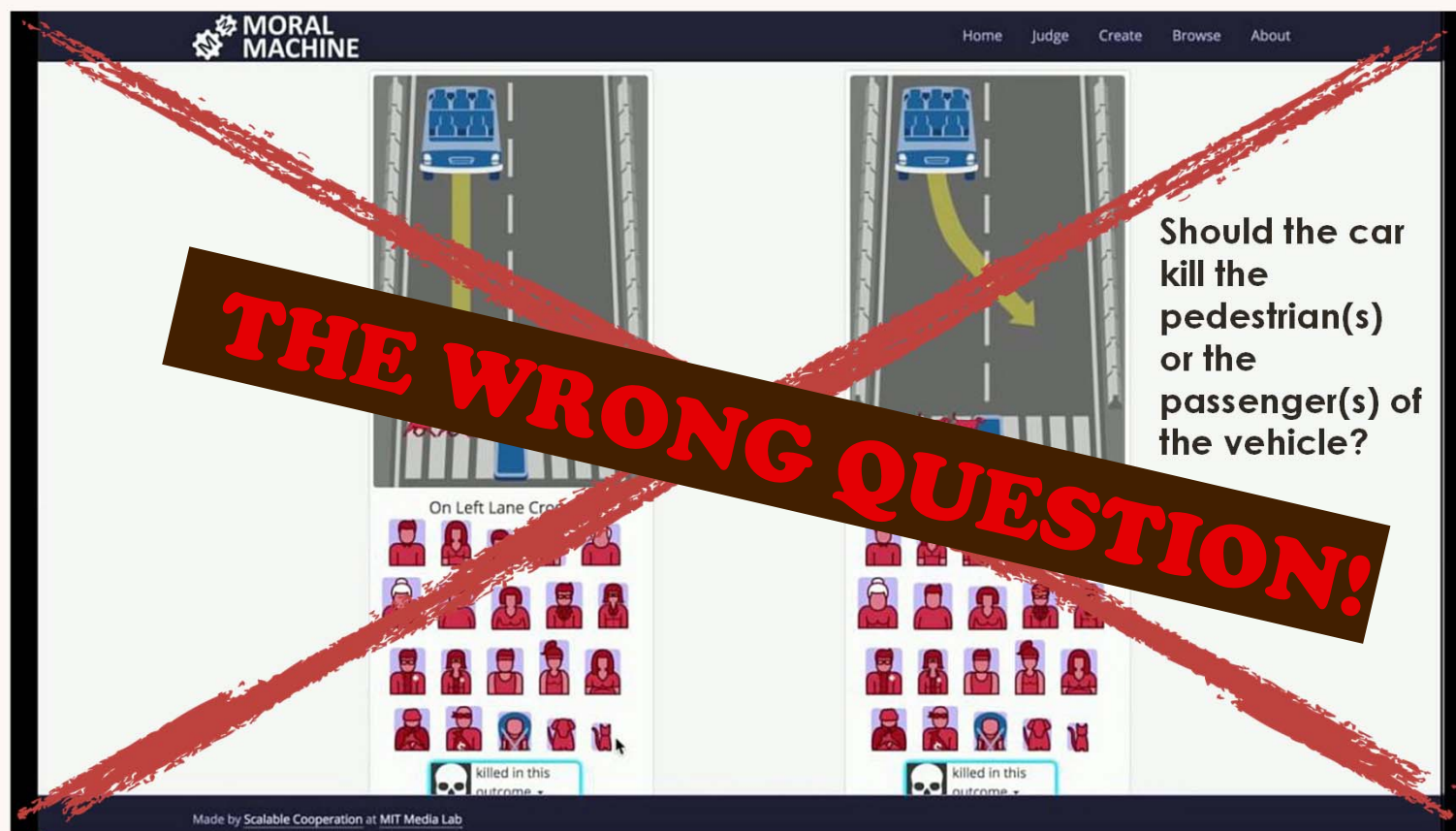
What does it mean to be human, when machines understand human emotions?



<https://i.pinimg.com/564x/39/21/d8/3921d818838f1644d48b66e2a5813137.jpg>



# The Moral Machine



- Developed by researchers at MIT to test possible decision scenarios for autonomous vehicles
- Mostly binary choices such as:
  - Large group or small
  - Older vs. younger people
  - Man vs. woman
  - Pregnant woman or not
  - Human vs. animal

# The Real Moral Question



[Home](#) [Judge](#) [Create](#) [Browse](#) [About](#)

A human would try to avoid killing anyone.  
The car should do the same. Even so,  
sometimes people will die.  
If it can't do that better than humans,  
autonomous vehicles should not be  
allowed on the streets

What part of the ethics question should we communicate to owners of autonomous vehicles?



Driverless cars

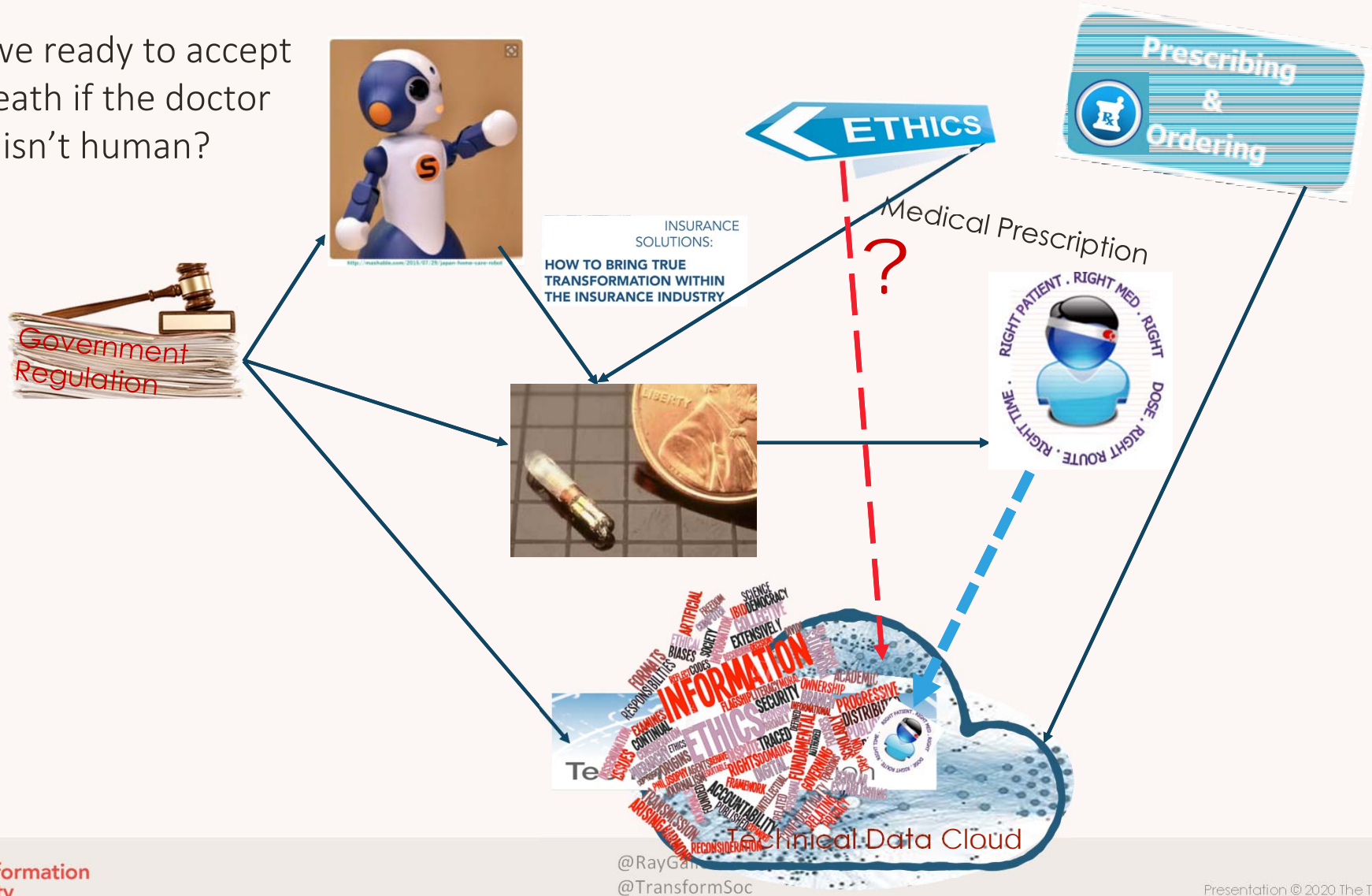


A close-up photograph of a black car key with a silver blade, resting on a white document. The document has the text 'Certificate of Motor Insurance' and 'Policy Schedule' printed on it. A blue arrow points from the left towards the document.

*What should the information offer be?*

# Data Cloud

Are we ready to accept  
a death if the doctor  
isn't human?





Are we ready to accept a death if the doctor isn't human?



What information should we give patients before an operation performed by a robot?



Would you accept an AI Entity for Health Care?

ETHICS

Prescribing & Ordering

INSURANCE SOLUTIONS:  
HOW TO BRING TRUE TRANSFORMATION WITHIN THE INSURANCE INDUSTRY



Medical Prescription



"A group of individuals whose collective behaviour is controlled by a single individual cannot behave in a more complex way than the individual who is exercising the control."

*-Yaneer Bar-Yam,*

*President and Professor at  
the New England Complex  
Systems Institute*

**We are all responsible!**



# Qualia

- Philosophical concept of individual instances of subjective, conscious experience, e.g. "what it is like to taste a specific apple, this particular apple now."
- Four properties that are commonly ascribed to qualia (per Frank Jackson):
  - Ineffable; that is, they cannot be communicated, or apprehended by any means other than direct experience.
  - Intrinsic; that is, they are non-relational properties, which do not change depending on the experience's relation to other things.
  - Private; that is, all interpersonal comparisons of qualia are systematically impossible.
  - Directly or immediately apprehensible in consciousness; that is, to experience a quale is to know one experiences a quale, and to know all there is to know about that quale.

*Source: Wikipedia*



AI can identify their  
*quintessential “dogginess”*

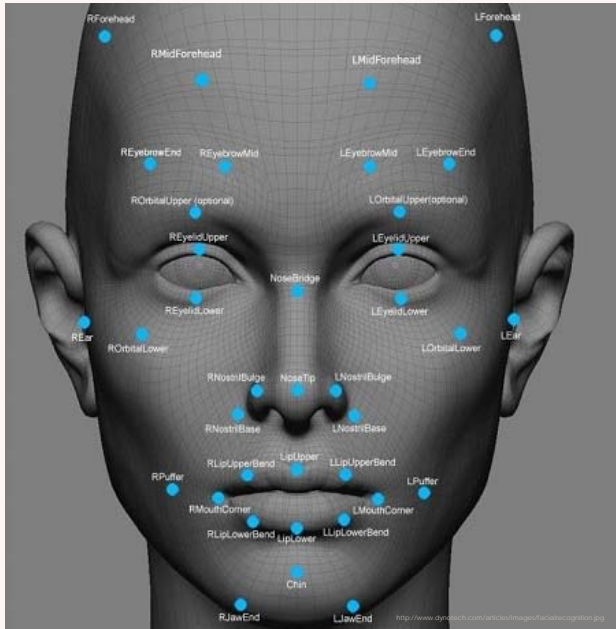
It can even identify this  
particular dog:



**QUALIA!**

But it can never tell you what  
makes this particular dog  
special for you!

# Cognitive Bias



AI algorithm used snow as a characteristic of a wolf

- AI algorithms use statistical data to distinguish objects.
- How it goes about doing this is a mystery, even to the algorithm's designers.
- Facial recognition programmes work similarly using biometric data.
- Can they actually know who you are, or what something really is?

# Cognitive Bias

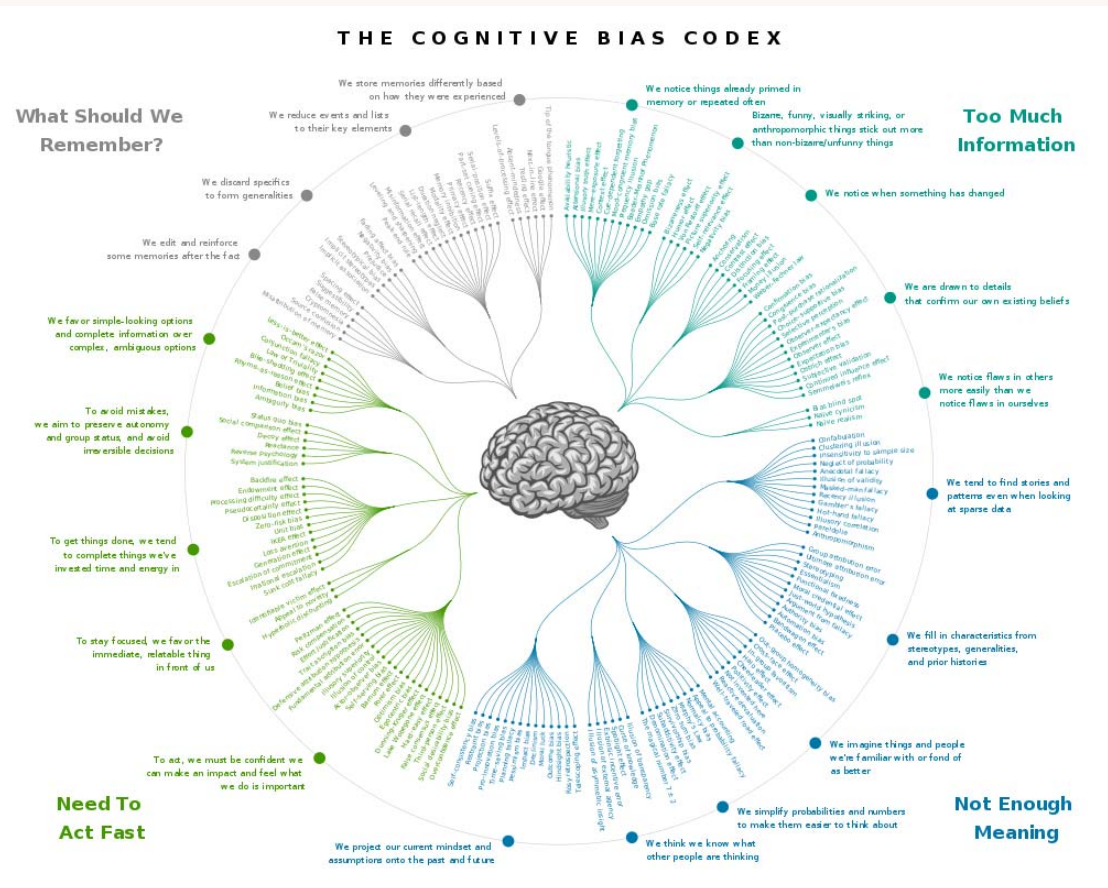


AI algorithm used snow as a characteristic of a wolf

- AI algorithms use statistical data to distinguish objects.
- How it goes about doing this is a mystery, even to the algorithm's designers.
- If you used statistical data to train an AI human resources app, how many women would be hired as coders?



# The Cognitive Bias Codex



Wikipedia lists 193 cognitive biases, grouped by types:

- Decision-making, belief, and behavioral biases
- Social biases
- Memory errors and biases

[https://en.wikipedia.org/wiki/Cognitive\\_bias](https://en.wikipedia.org/wiki/Cognitive_bias)

[https://en.wikipedia.org/wiki/List\\_of\\_cognitive\\_biases](https://en.wikipedia.org/wiki/List_of_cognitive_biases)

# Algorithms are prone to the biases of their creators

---

## Ask

- Who is coding
- How it's coded
- Why it's coded
- How are you testing?
- How are you closing the loop?

## Insist on

- Team diversity
- Code and content audits
- Transparency of code and rational of factors
- Involvement of an ethicist
- Examining data sets
- Explain and justification

*Source: Rahel Anne Bailie*



# AI in the Real World

- Statistics show that most aviation accidents are caused by pilot error
- Most experts agree that autopilots fly planes better and safer than humans
- Are you prepared to board a pilotless plane?



# The Moral Dilemma

**IF PILOTS DON'T GET TO PRACTICE FLYING PLANES, HOW WILL THEY EVER BE ABLE TO KNOW HOW TO LAND ON THE RIVER?**

**ARE SIMULATORS GOOD ENOUGH? AND WHY SHOULD THE PILOTS CARE WHEN INCIDENTS WILL BE SO RARE?**



# Data

# MetaData



# Data

# MetaData





# Encapsulation



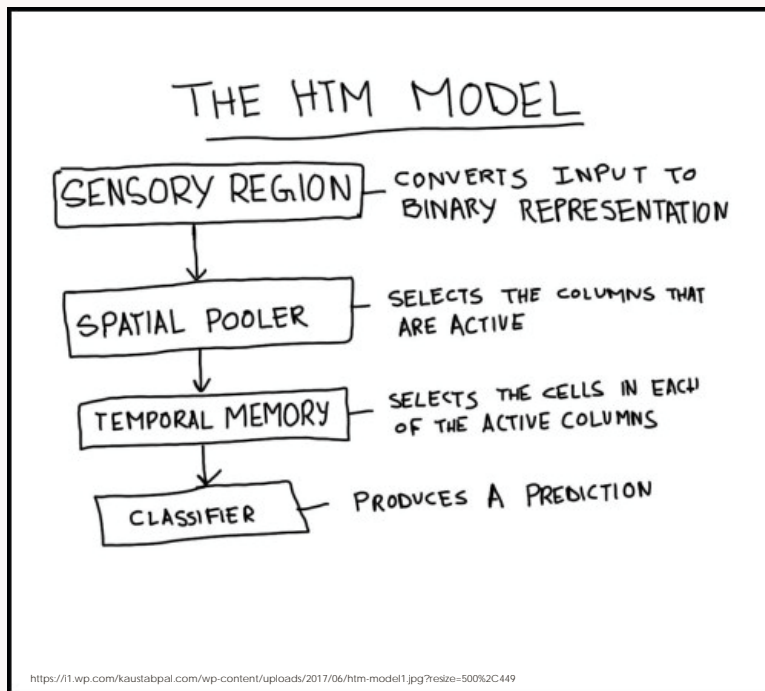
With successive layers of encapsulation, re-encapsulation, recombination with new encapsulations, further re-encapsulation...

- **Where does responsibility lie?**
- **What is the contract between information producer and information user?**

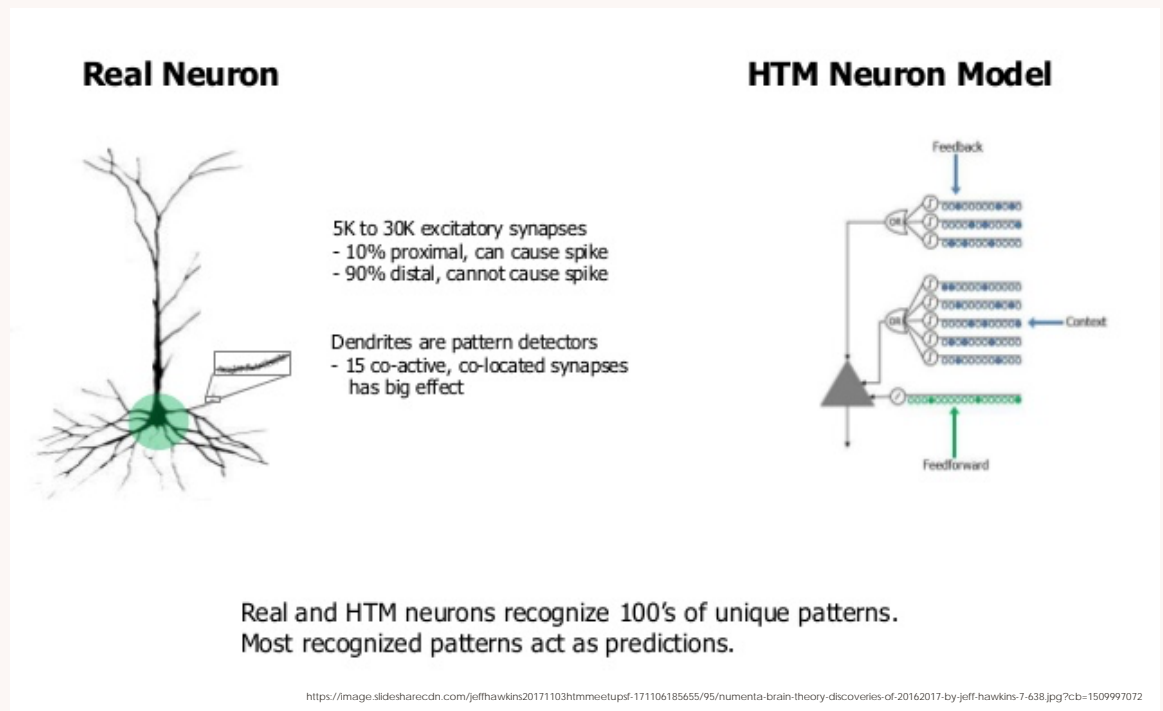


# HTM: Hierarchical Temporal Memory

Jeffrey Hawkins, Numenta



Open source: [github.com/numenta](https://github.com/numenta)













*Process of Generalization*


# Intuition Machine

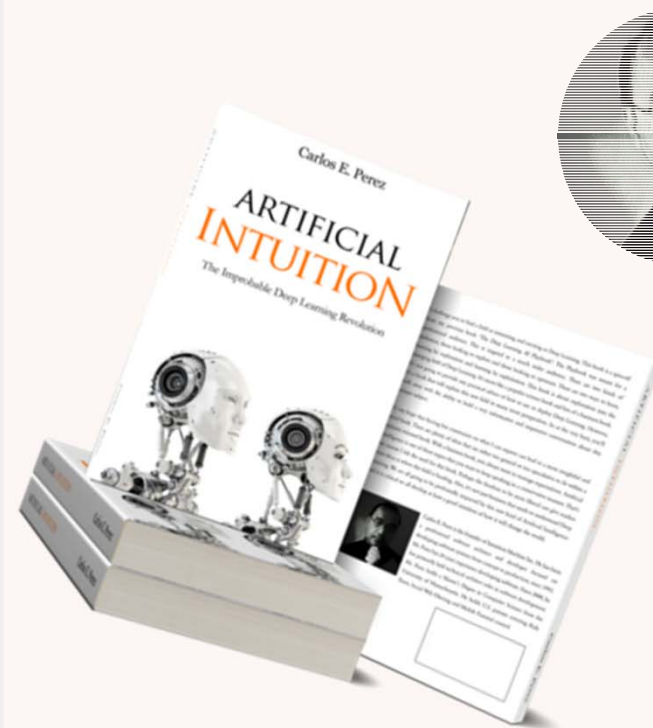
Carlos E. Perez

**The Deep Learning Canvas (v0.5)** Designed for:  Designed by:  Date:  Iteration:

<b>Decision Support</b>  How does proposed features enhance decision making?  Understand the consequences of errors made by automation.  How does the UX improve accessibility and availability?	<b>Features</b>  What kind of problem are we trying to solve?  Classification, Prediction, Translation, Generation, Planning or Optimization?	<b>Value Propositions</b>  What does the customer need to get done?  How can we help him get his job done in the right context?	<b>Cognitive Limits</b>  What combination of cognitive limits are you addressing?  Limits to memory Thinking fast Information overload Lack of meaning	<b>Customer</b>  What are the jobs people do and how can their productivity be amplified by automation?
<b>Data Logistics</b>  What needs to be done to acquire, ingest, fuse and enhance data.	<b>Model Development</b>  What kind of Deep Learning models can we use to implement the features?  How can we synthesize or enhance data to improve a curriculum?	<b>Key Metrics</b>  What can we measure to quantify improvement?  How can we continually evaluate performance?		
<b>Technical Debt</b>  What do we need to track to build a adaptable and sustainable architecture?  How can we democratize our models and datasets?		<b>Context</b>  How messy is the context to perform the job?  Document the activity loop required to perform the job.		

[Intuition Machine](#) by Carlos E. Perez

Licensed under a Creative Commons Attribution-ShareAlike 4.0 International License. 





# ‘Wicked’ Problems

Problems that are difficult or impossible to solve because of incomplete, contradictory, and changing requirements that are often difficult to recognize.

In addition, complex interdependencies make it so that the effort to solve one aspect of a wicked problem may reveal or create other problems.





# From invisible to unknown...?





# Emerging from Chaos

We don't "solve" wicked problems!

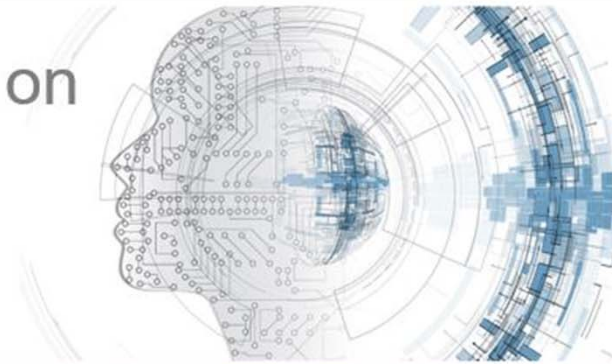


- Complex problems are **entanglements** of many simple threads. They can seem **chaotic**.
- Order **emerges** from chaos in its own way. This might not necessarily be positive.
- Our job is to reiteratively help users find **positive, constructive emergences** from complexity.
- Our primary tool for doing this is a combination of psychology, **linguistics**, and socio-anthropology...

# Actions for Ethics

## The OECD Recommendation on Responsible Innovation in Neurotechnology

embodies nine principles, focusing on:



- |                                       |   |
|---------------------------------------|---|
| 1. Promoting responsible innovation   | 6. Enabling capacity of oversight and advisory bodies                               |
| 2. Prioritising safety assessment     | 7. Safeguarding personal brain data and other information                           |
| 3. Promoting inclusivity              | 8. Promoting cultures of stewardship and trust across the public and private sector |
| 4. Fostering scientific collaboration | 9. Anticipating and monitoring potential unintended use and/or misuse               |
| 5. Enabling societal deliberation     |   |



# European Union Actions

## European Commission (2019)



[https://ai.bsa.org/wp-content/uploads/2019/09/AIHLEG\\_EthicsGuidelinesforTrustworthyAI-ENpdf.pdf](https://ai.bsa.org/wp-content/uploads/2019/09/AIHLEG_EthicsGuidelinesforTrustworthyAI-ENpdf.pdf)



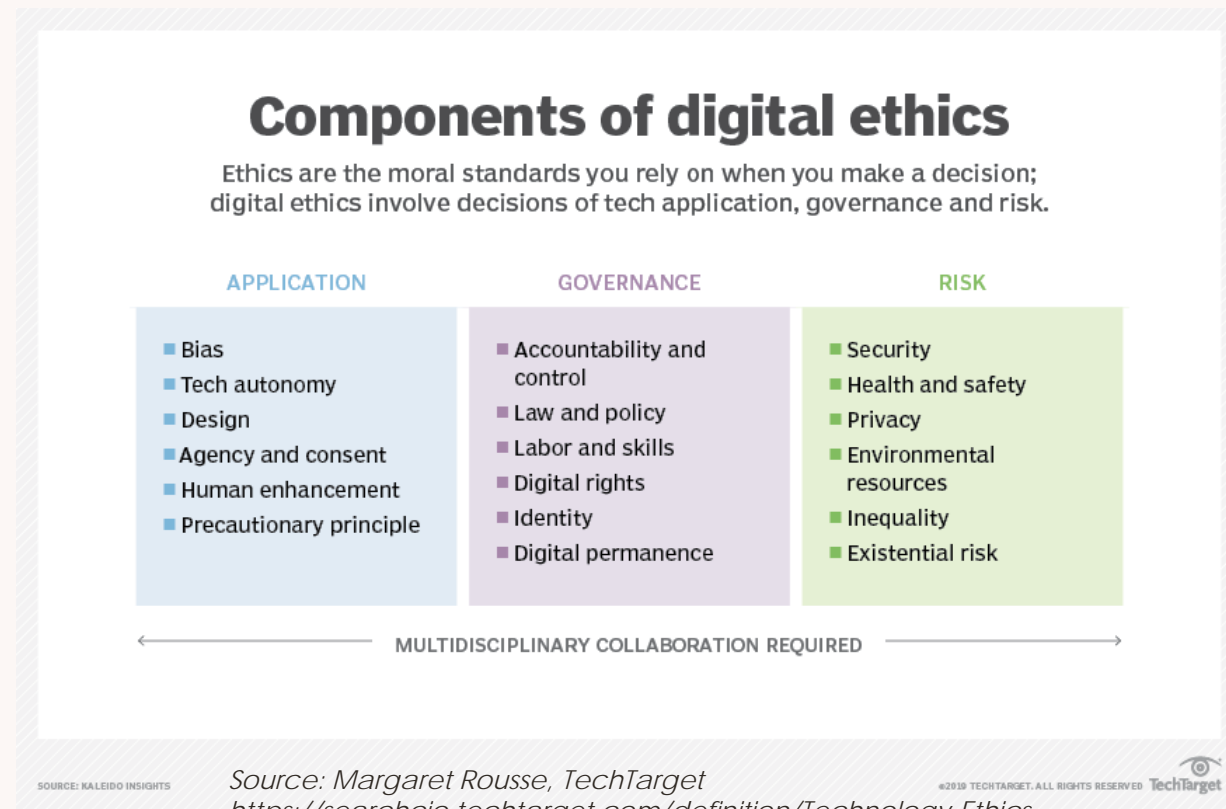
## European Parliament Resolution on Robotics (2017)

<http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+TA+P8-TA-2017-0051+0+DOC+PDF+V0//EN>



# Technology Ethics

- A field of study that seeks to understand and resolve ethical issues surrounding the development and practical application of mechanical and electronic technology.
- Focuses on subjects such as:
  - The relationship between technology and human values and well-being,
  - The condition in which technological advances occur
  - The social repercussions for technological advancements.





# Asilomar AI Code of Ethics – Key Points

- Highly autonomous AI systems should be designed so that their goals and behaviours can be assured to align with human values throughout their operation.
- AI systems should be designed and operated so as to be compatible with ideals of human dignity, rights, freedoms, and cultural diversity.
- The power conferred by control of highly advanced AI systems should respect and improve, rather than subvert, the social and civic processes on which the health of society depends.

## Founders:



- Jaan Tallinn - Skype Co-founder
- Max Tegmark - Cosmologist, Massachusetts Institute of Technology
- Viktoriya Krakovna - Research Scientist, DeepMind
- Anthony Aguirre - Physicist, University of California, Santa Cruz
- Meia Chita-Tegmark - Psychologist, Tufts University

To date, the Principles have been signed by 1583 AI/Robotics researchers and 3447 others

<https://futureoflife.org/ai-principles/>



If Artificial Intelligence cannot help us solve human problems,  
it will become another world problem.

FIND BEAUTY IN COMPLEXITY

Thanks for your attention!

More Information:

<http://transformationsociety.net>  
[ray@transformationsociety.net](mailto:ray@transformationsociety.net)



@RayGallon  
@TransformSoc

Presentation © 2020 The Transformation Society