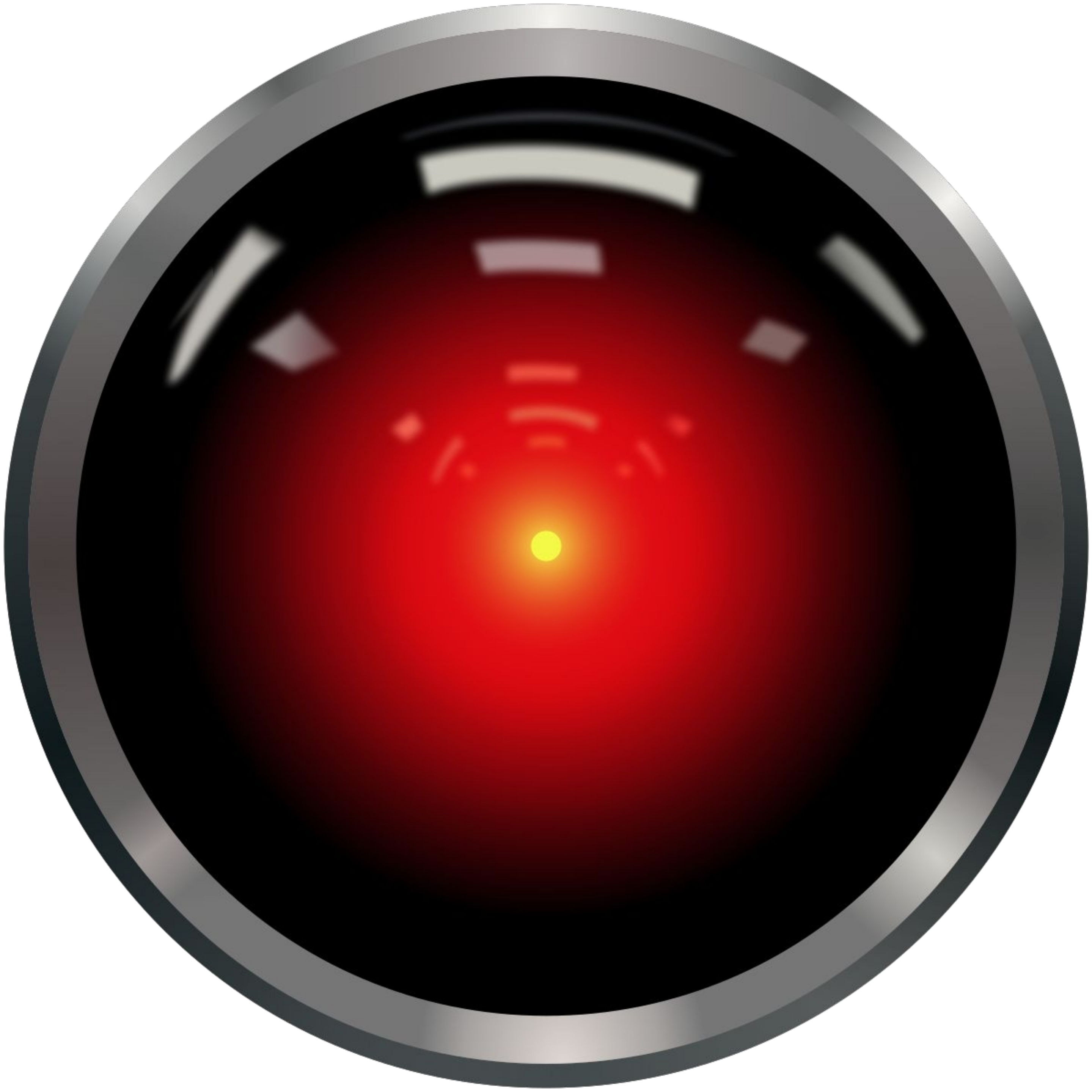
AI for Sceptics

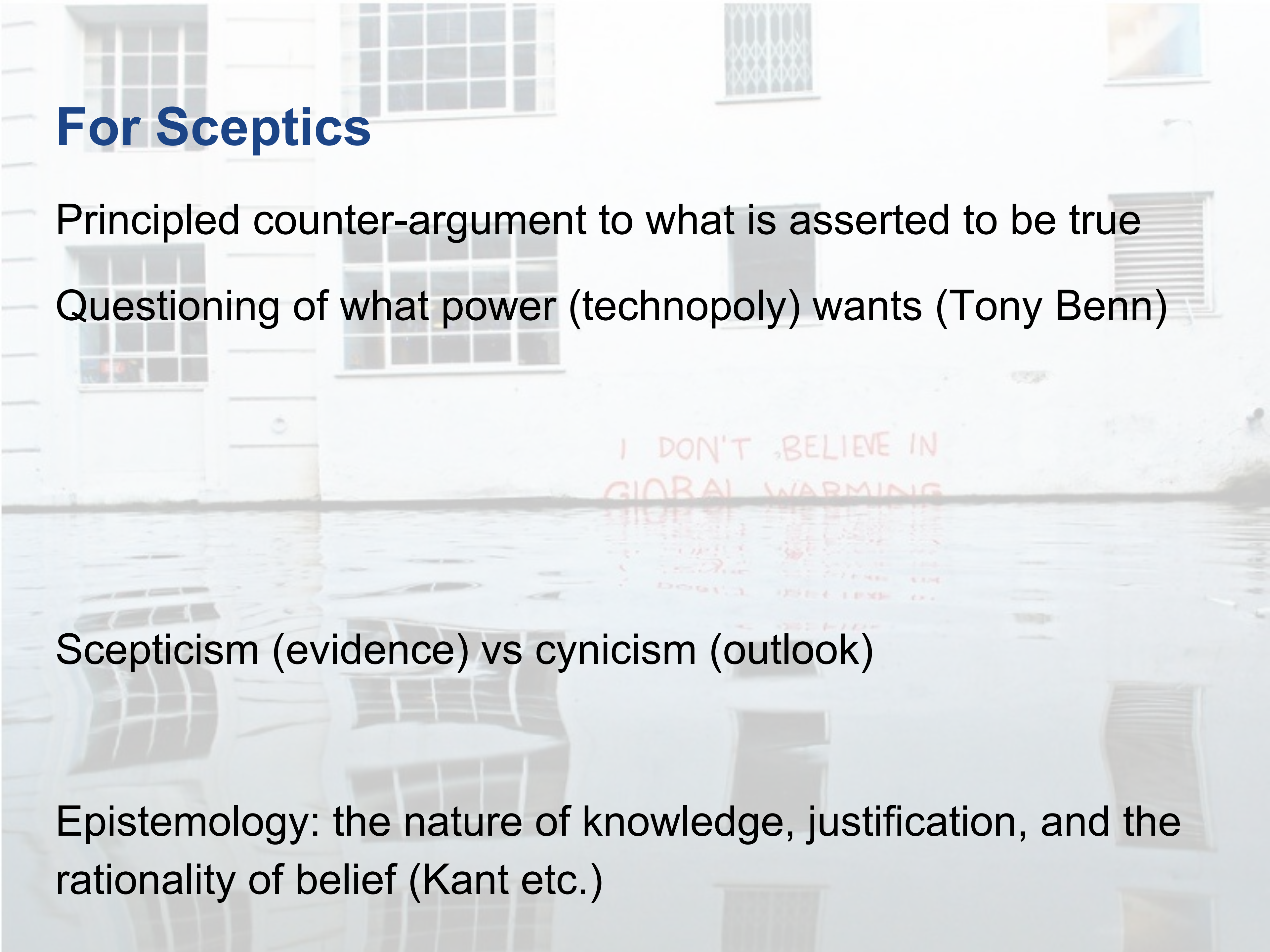
Tim Kindberg



For Sceptics

Principled counter-argument to what is asserted to be true

Questioning of what power (technopoly) wants (Tony Benn)



Scepticism (evidence) vs cynicism (outlook)

Epistemology: the nature of knowledge, justification, and the rationality of belief (Kant etc.)


The Scent
of Data


Statistical Capitalism
Surveillance Capitalism

If by your art, my dearest father, you have
Put the wild waters in this roar, allay them.
The sky, it seems, would pour down stinking pitch,
But that the sea, mounting to the welkin's cheek,
Dashes the fire out. O, I have suffered
With those that I saw suffer: a brave vessel,
Who had, no doubt, some noble creature in her,
Dash'd all to pieces. O, the cry did knock
Against my very heart. Poor souls, they perish'd.

55:55:20 — Swigert: "Okay, Houston, we've had a problem here."

55:55:28 — Lousma: "This is Houston. Say again please."

55:55:35 — Lovell: "Houston, we've had a problem. We've had a main B bus undervolt."

55:55:42 — Lousma: "Roger. Main B undervolt."

55:55:49 — Oxygen tank No. 2 temperature begins steady drop lasting 59 seconds indicating a failed sensor.

55:56:10 — Haise: "Okay. Right now, Houston, the voltage is — is looking good. And we had a pretty large bang associated with the caution and warning there. And as I recall, main B was the one that had an amp spike on it once before.

55:56:30 — Lousma: "Roger, Fred."

Levi: errmmmmmm…

Alfie: you don’t know who’s in your class?

Levi: do you want me to name everyone?

Alfie: Naa, just a couple people!

Levi: ermm..billy, rashad, Ryan and Ann Marie.

Jack: is it jokes?

Levi: um..yeah.

Alfie: do you have miss.naylan-Francis?

Levi: an Mr.Evans.

Alfie: have you had Miss.Naylan-Francis or just Mr.Evans?

Levi: na miss.naylan-Francis, cos Mr.evans isn’t a RE teacher!

Alfie: yeah I know but she might not have been in!

Levi: ermm..anyway..

Laczkovich's proof [edit]

Miklós Laczkovich's proof is a simplification of Lambert's original proof.[10] He considers the functions

$$f_k(x) = 1 - \frac{x^2}{k} + \frac{x^4}{2!k(k+1)} - \frac{x^6}{3!k(k+1)(k+2)} + \cdots \qquad (k \notin \{0, -1, -2, \ldots\}).$$

These functions are clearly defined for all $x \in \mathbf{R}$. Besides

$$f_{1/2}(x) = \cos(2x) \text{ and } f_{3/2}(x) = \frac{\sin(2x)}{2x}.$$

Claim 1: The following recurrence relation holds:

$$(\forall x \in \mathbb{R}) : \frac{x^2}{k(k+1)} f_{k+2}(x) = f_{k+1}(x) - f_k(x).$$

Proof: This can be proved by comparing the coefficients of the powers of x.

Claim 2: For each $x \in \mathbf{R}$, $\lim_{k \to +\infty} f_k(x) = 1.$

Proof: In fact, the sequence $x^{2n}/n!$ is bounded (since it converges to 0) and if C is an upper bound and if $k > 1$, then

$$|f_k(x) - 1| \leqslant \sum_{n=1}^{\infty} \frac{C}{k^n} = C \frac{1/k}{1 - 1/k} = \frac{C}{k-1}.$$

Claim 3: If $x \neq 0$ and if x^2 is rational, then

$$(\forall k \in \mathbb{Q} \setminus \{0, -1, -2, \ldots\}) : f_k(x) \neq 0 \text{ and } \frac{f_{k+1}(x)}{f_k(x)} \notin \mathbb{Q}.$$

```
(defmodule avoid
  :inputs (force heading)
  :outputs (command)
  :instance-vars (resultforce)
  :states
    ((nil (event-dispatch (and force heading) plan))
     (plan (setf resultforce (select-direction force heading))
           go)
     (go (conditional-dispatch (significant-force-p resultforce 1.0)
                               start
                               nil))
     (start (output command (follow-force resultforce))
            nil)))
```

Kleene's proof of Gödel's Theorem

Peter Smith

Faculty of Philosophy, University of Cambridge

There is a familiar derivation of Gödel's Theorem from the proof by diagonalization of the unsolvability of the Halting Problem. That proof, though, still involves a kind of self-referential trick, as we in effect construct a sentence that says 'the algorithm searching for a proof of me doesn't halt'. It is worth showing, then, that some core results in the theory of partial recursive functions directly entail Gödel's First Incompleteness Theorem *without* any further self-referential trick.

We start with reminders about two theorems from the theory of partial recursive functions. First, Kleene's Normal Form theorem:

Theorem 1. *There is a three-place p.r. function C and a one-place p.r. function U such that any one-place partial recursive function can be given in the standard form*

$$f_e(n) =_{\text{def}} U(\mu z[C(e,n,z) = 0])$$

for some value of e.

(Read 'μz' as *the least z such that.*) This is a standard textbook result.

We next fix some not-so-familiar terminology:

Defn. 1. *The function g is a* completion *of a partial function f if g is total and for all n where $f(n)$ is defined, $f(n) = g(n)$.*

Defn. 2. *A partial function f is* potentially recursive *if it has a completion g which is recursive.*

Then we have another textbook result:

Theorem 2. *Not every partial recursive function is potentially recursive.*

In fact, this follows immediately from Theorem 1, by a diagonalization argument:

Proof. Put $f(n) \approx U(\mu z[C(n,n,z) = 0]) + 1$. So $f(n)$ is the function which for argument n takes the value $f_n(n) + 1$ when that is defined and is undefined otherwise. $f(n)$ is by construction a partial computable function. But there is no total recursive function g which completes it.

Suppose otherwise. For some e, then, g is the function f_e – remember, the partial recursive functions include the total recursive functions, and by Theorem 1 the f_e are all the partial recursive functions!

Since g is total, $g(e)$, i.e. $f_e(e)$, is defined. So $f(e)$, i.e. $f_e(e) + 1$, is defined. But g must agree with f when f is defined. So $f_e(e) = g(e) = f(e) = f_e(e) + 1$. Contradiction! $\square$

Two more definitions just to fix more not-entirely-standard terminology:

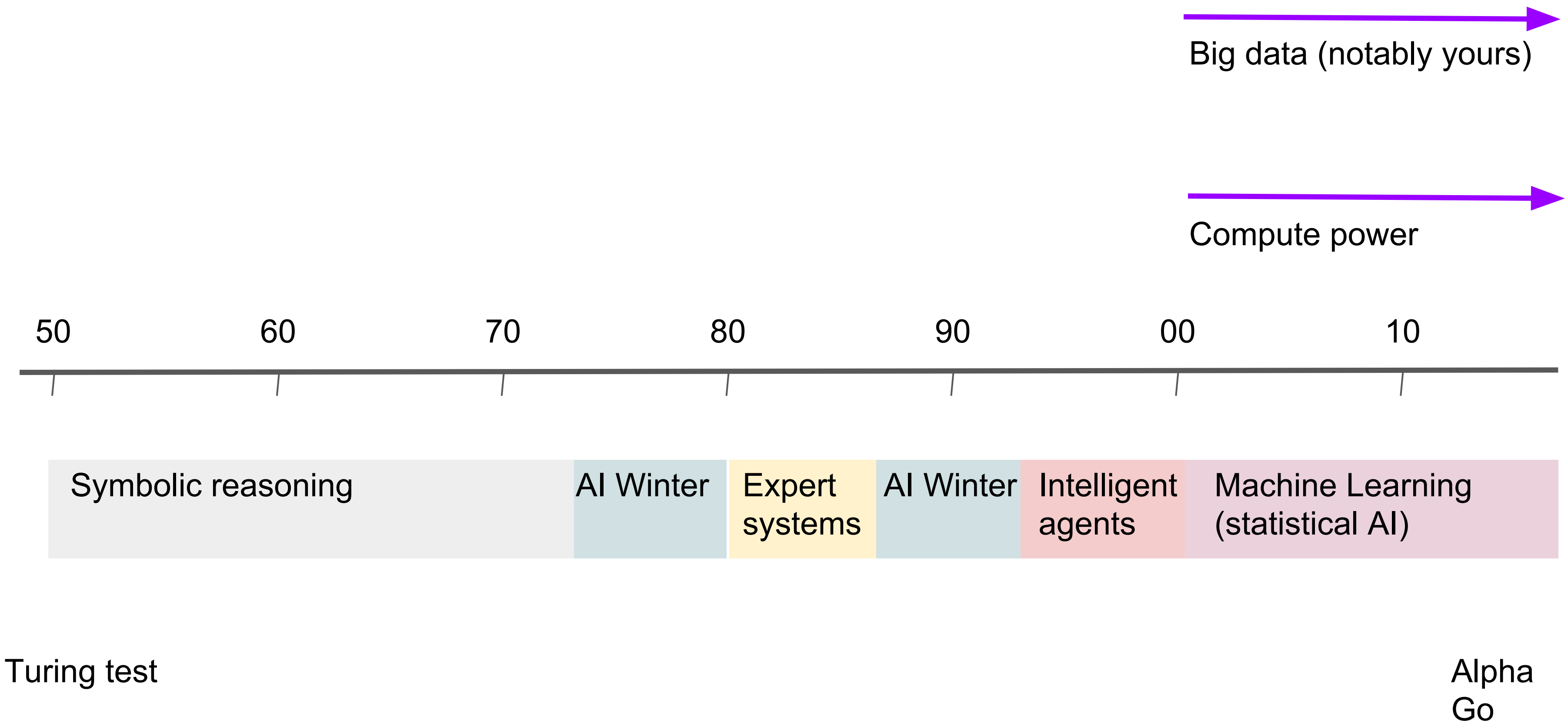
On Non-Computable Functions

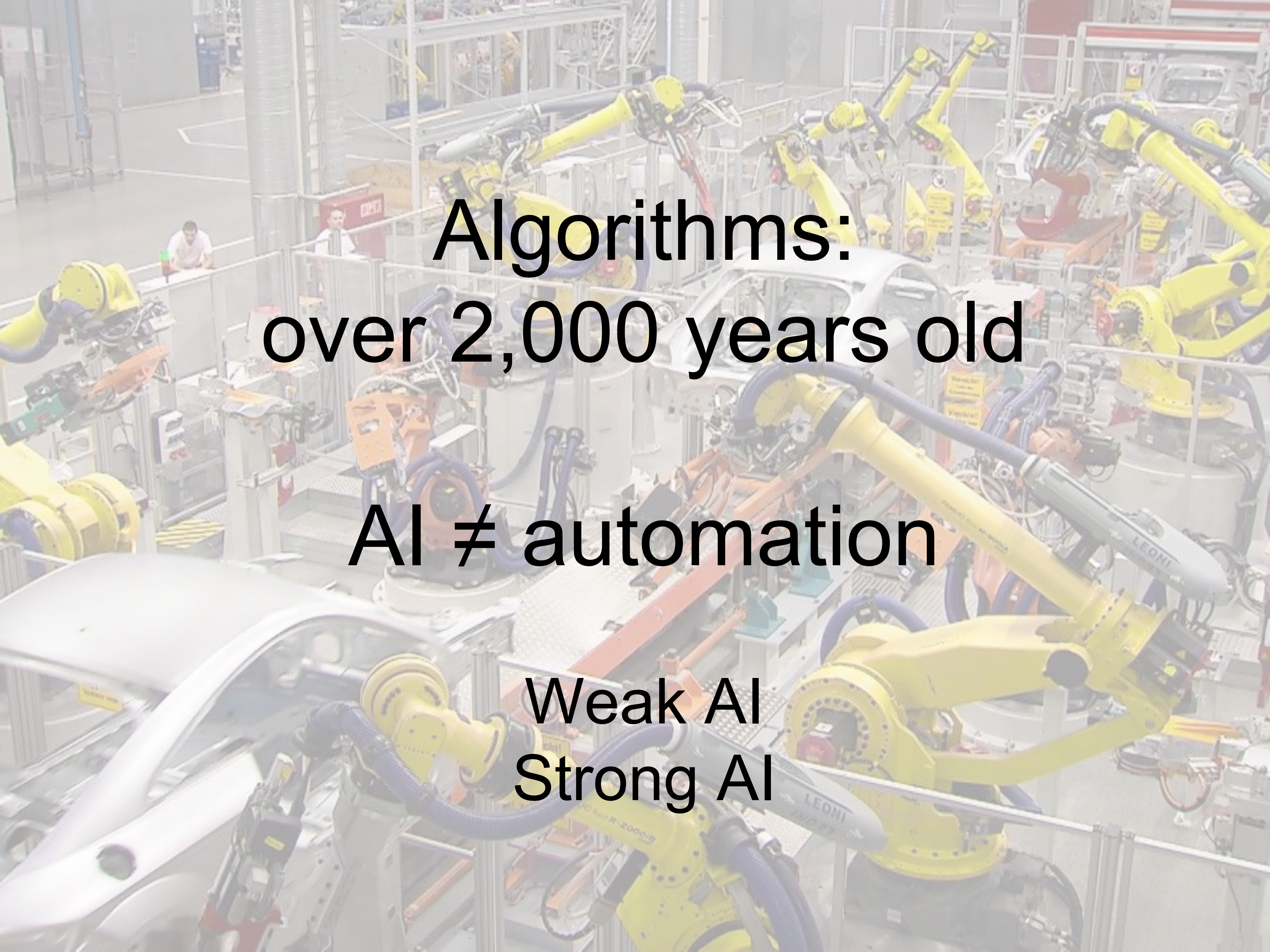
By T. RADO

(Manuscript received November 12, 1961)

The construction of non-computable functions used in this paper is based on the principle that a finite, non-empty set of non-negative integers has a largest element. Also, this principle is used only for sets which are exceptionally well-defined by current standards. No enumeration of computable functions is used, and in this sense the diagonal process is not employed. Thus, it appears that an apparently self-evident principle, of constant use in every area of mathematics, yields non-constructive entities.

"Every day we read that digital computers play chess, translate languages, recognize patterns, and will soon be able to take over our jobs."


Big data (notably yours)
Compute power
50
60
70
80
90
00
10
Symbolic reasoning
AI Winter
Expert systems
AI Winter
Intelligent agents
Machine Learning (statistical AI)
Turing test
Alpha Go



Algorithms:
over 2,000 years old

AI ≠ automation

Weak AI
Strong AI

Speech / handwriting recognition

Game playing

Surveillance

Marketing

Driverless cars

'Killer robots'

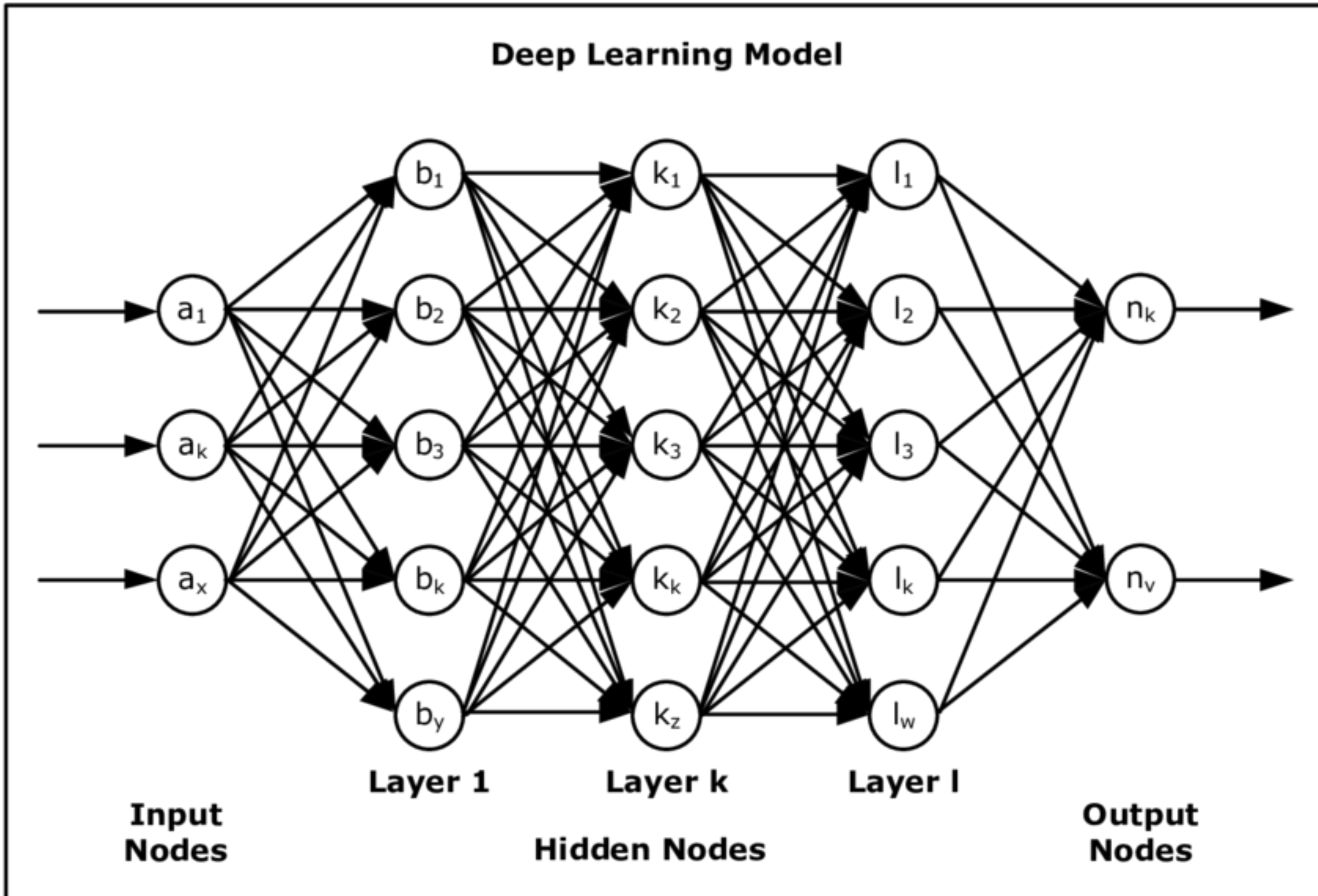
Translation

Machine Learning/ Statistical AI

Artificial neural nets

Bayesian analysis

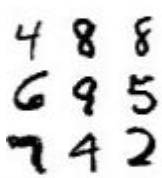
Clustering


Deep Learning Model
a1
ak
ax
b1
b2
b3
bk
by
k1
k2
k3
kk
kz
l1
l2
l3
lk
lw
nk
nv
Layer 1
Layer k
Layer l
Input Nodes
Hidden Nodes
Output Nodes

Recognising handwritten digits

1 Training data

Choose data set → Feature engineer-ing / extraction → Machine learning



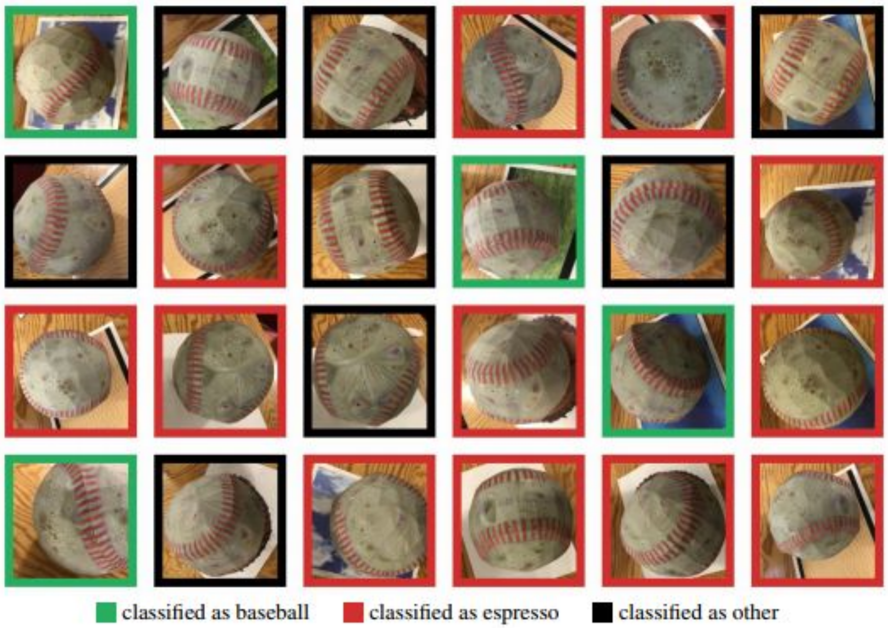
2 Real-world data

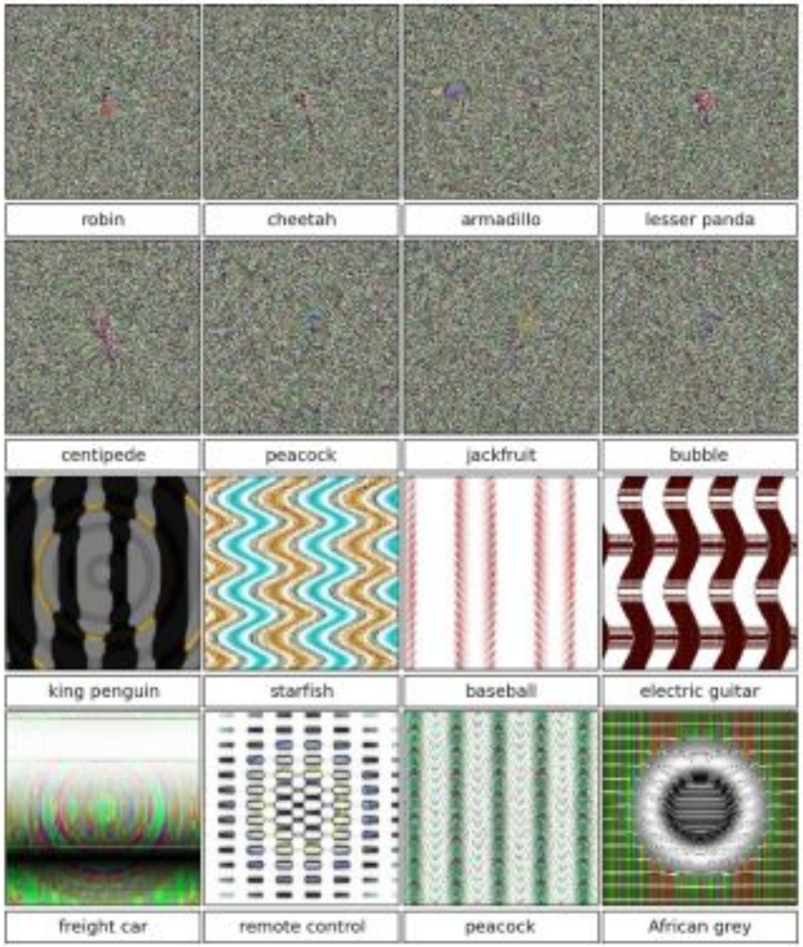
 → Machine learning → ?

Wikipedia entry on Deep Learning

"...learn multiple levels of representations that correspond to different levels of abstraction; the levels form a hierarchy of concepts..."

Deep neural networks are easily fooled





The Scent of Data

No faculty for symbols or reasoning

- No error model
- No explanations
- No applicability across domains

Not how we learn

More akin to a new type of smell (olfaction) than intelligence

Machine learning

Issue 1: Epistemology

Never mind the Turing test

$$7 + 5 = 12$$

Intensional vs extensional logic

IMMANUEL KANT
From a painting

Issue 2: Misguided/premature application

FEB 19, 2017 @ 03:48 PM 200,510 EDITOR'S PICK The Little Black Book of Billionaire Secrets

MD Anderson Benches IBM Watson In Setback For Artificial Intelligence In Medicine

Matthew Herper, FORBES STAFF

I cover science and medicine, and believe this is biology's century.

FULL BIO

“It’s not happening today, and it might not be happening in five years. And it’s not going to replace doctors.” (IBM spokesperson)

Issue 3: Wider Political & Ethical Implications

STANFORD BUSINESS GRADUATE SCHOOL OF

Change lives. Change organizations. Change the world.

Deep Neural Networks Are More Accurate Than Humans at Detecting Sexual Orientation From Facial Images

By Michal Kosinski, Yilun Wang

Journal of Personality and Social Psychology (in press). September 7, 2017.

Organizational Behavior

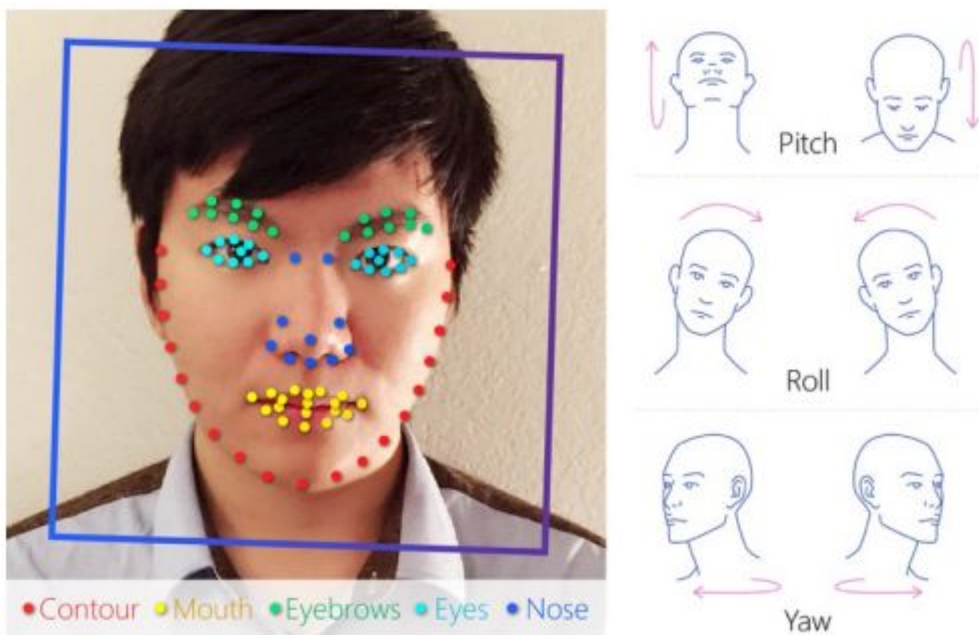
View Publication | Download

We show that faces contain much more information about sexual orientation than can be perceived and interpreted by the human brain. We used deep neural networks to extract features from 35,326 facial images. These features were entered into a logistic regression aimed at classifying sexual orientation. Given a single facial image, a classifier could correctly distinguish between gay and heterosexual men in 81% of cases, and in 74% of cases for women. Human judges achieved much lower accuracy:

Related

Michal Kosinski
Assistant Professor, Organizational Behavior

"Facial images. We obtained facial images from public profiles posted on a U.S. dating website. We recorded 130,741 images of 36,630 men and 170,360 images of 38,593 women between the ages of 18 and 40, who reported their location as the U.S. Gay and heterosexual people were represented in equal numbers. Their sexual orientation was established based on the gender of the partners that they were looking for (according to their profiles)."

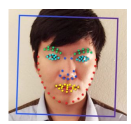


Gay or straight?

1 Training data

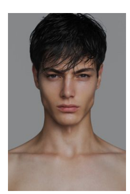
Choose data set

Feature engineer-ing / extraction



Machine learning

2 Real-world data



Machine learning

?

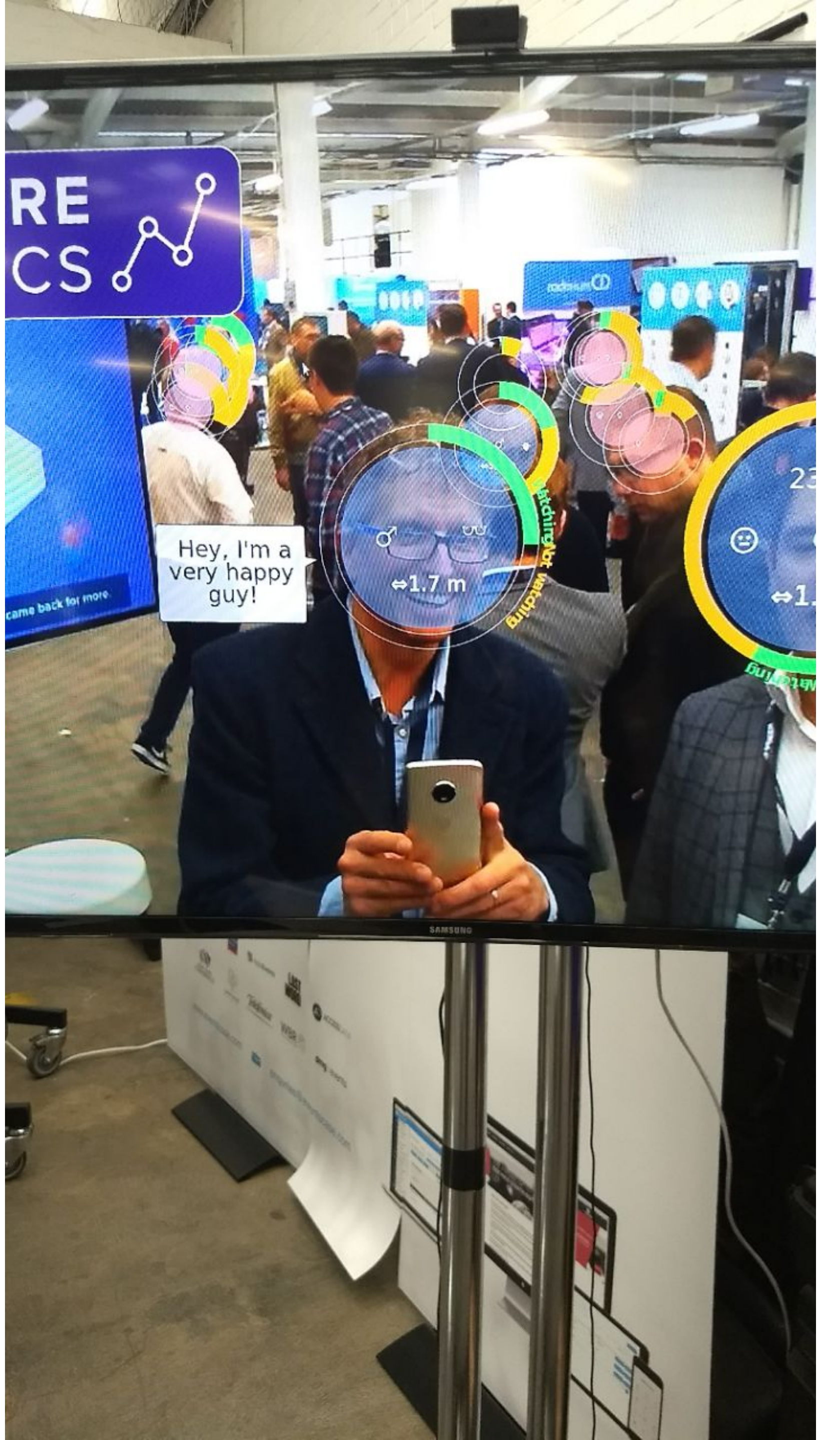
Your job in 2022?

"we employed Amazon Mechanical Turk (AMT) workers to verify that the faces were adult, Caucasian, fully visible, and of a gender that matched the one reported on the user’s profile. We limited the task to the workers from the U.S., who had previously completed at least 1,000 tasks and obtained an approval rate of at least 98%. Only faces approved by four out of six workers were retained."

"Facial features were extracted from the images using a widely employed DNN, called VGG-Face (Parkhi, Vedaldi, & Zisserman, 2015). VGG-Face was originally developed (or trained) using a sample of 2.6 million images for the purpose of facial recognition (i.e., recognizing a given person across different images)."

"Unfortunately, [...] VGG-Face scores are not easily interpretable. A single score might subsume differences in multiple facial features typically considered to be distinct by humans (e.g., nose shape, skin tone, or eye color)."

Figure 3. Heat maps showing the degree to which masking a given part of an image changes the (absolute) classification outcome, which is a proxy for the importance of that region in


RE
CS
Hey, I'm a very happy guy!
⇔1.7 m
watching
not watching
came back for more.
SAMSUNG

Biased bots: Artificial-intelligence systems echo human prejudices

by Adam Hadhazy for the Office of Engineering Communications
April 18, 2017 noon

In debates over the future of artificial intelligence, many experts think of these machine-based systems as coldly logical and objectively rational. But in a new study, Princeton University-based researchers have demonstrated how machines can be reflections of their creators in potentially problematic ways.

Common machine-learning programs trained with ordinary human language available online can acquire the cultural biases embedded in the patterns of wording, **the researchers reported** in the journal Science April 14. These biases range from the morally neutral, such as a preference for flowers over insects, to discriminatory views on race and gender.

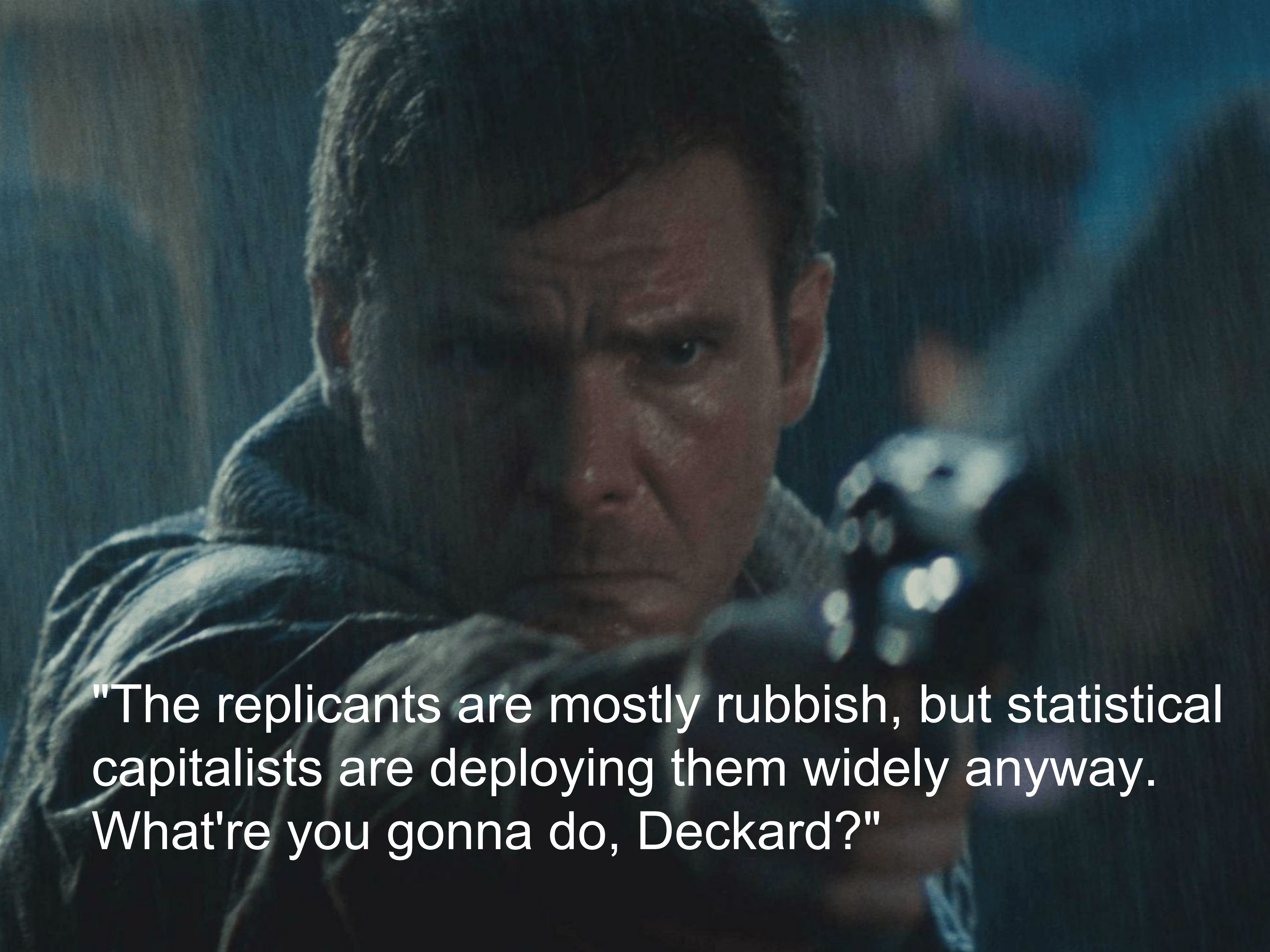
Meet the tech company that wants to make you even more addicted to your phone

BY JONATHAN SHIEBER
Sep 8, 2017

With a magic line of code, Dopamine Labs aims to give any app the same addictive power that Facebook, Zynga and others have spent millions to perfect.

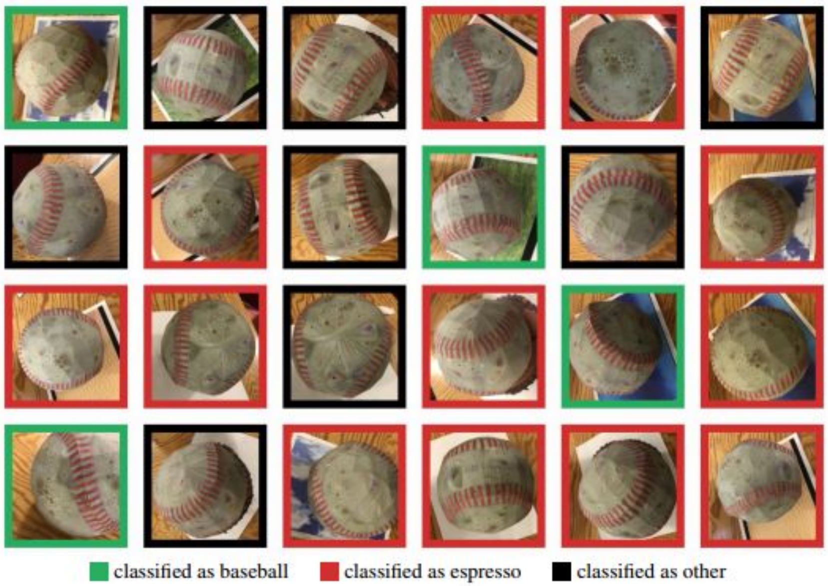
Managing all of this data is Skinner, an artificial intelligence software named after the famous psychologist B.F. Skinner, that monitors different prompts it’s making to the customers on apps that use Dopamine Labs’ code. Skinner’s all about learning what works for improving usage on an app or getting returning customers, and it optimizes those notifications as customers continue to use the app.

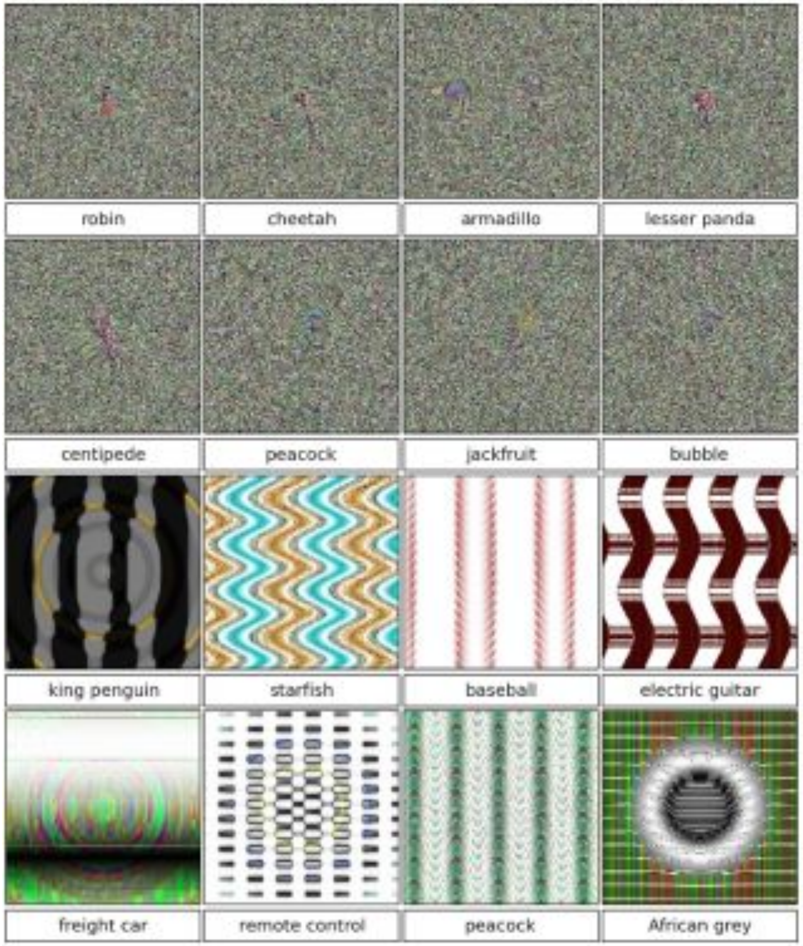
Google says it's now trying to figure out how this happened, the Wall Street Journal reports. The answer is probably an error within Google Photo's facial recognition technology.


"The replicants are mostly rubbish, but statistical
capitalists are deploying them widely anyway.
What're you gonna do, Deckard?"

Research response: subvert own algorithms

[Deep neural networks are easily fooled]





Personal response

Why am I on Facebook? Aternatives?

Creative response?


Thank You!
tim@matter2media.com
@timkindberg
#ForSceptics