AI myths & misperceptions: What AI experts wish everyone knew

Professor Michael Winikoff

michael.winikoff@vuw.ac.nz
http://michaelwinikoff.com

@profwinikoff

"**We're so unprepared for the robot apocalypse:** … Industrial robots alone have eliminated up to 670,000 American jobs between 1990 and 2007, according to new research from MIT's Daron Acemoglu and Boston University's Pascual Restrepo." Source: https://www.washingtonpost.com/news/wonk/wp/2017/03/30/were-so-unprepared-for-the-robot-apocalypse/?noredirect=on

"The fact is that AI can go further than humans, it could be billions of times smarter than humans at this point."

Source: http://www.nzherald.co.nz/business/news/article.cfm?c_id=3&objectid=11994369

Myths of AI …

- There is such a thing as “an AI”
- The AI research community is working (in a unified way) on developing “an AI”

- Human-level AI is close …
- … and super-intelligence will follow shortly (“singularity”)
- We should worry about intelligent machines

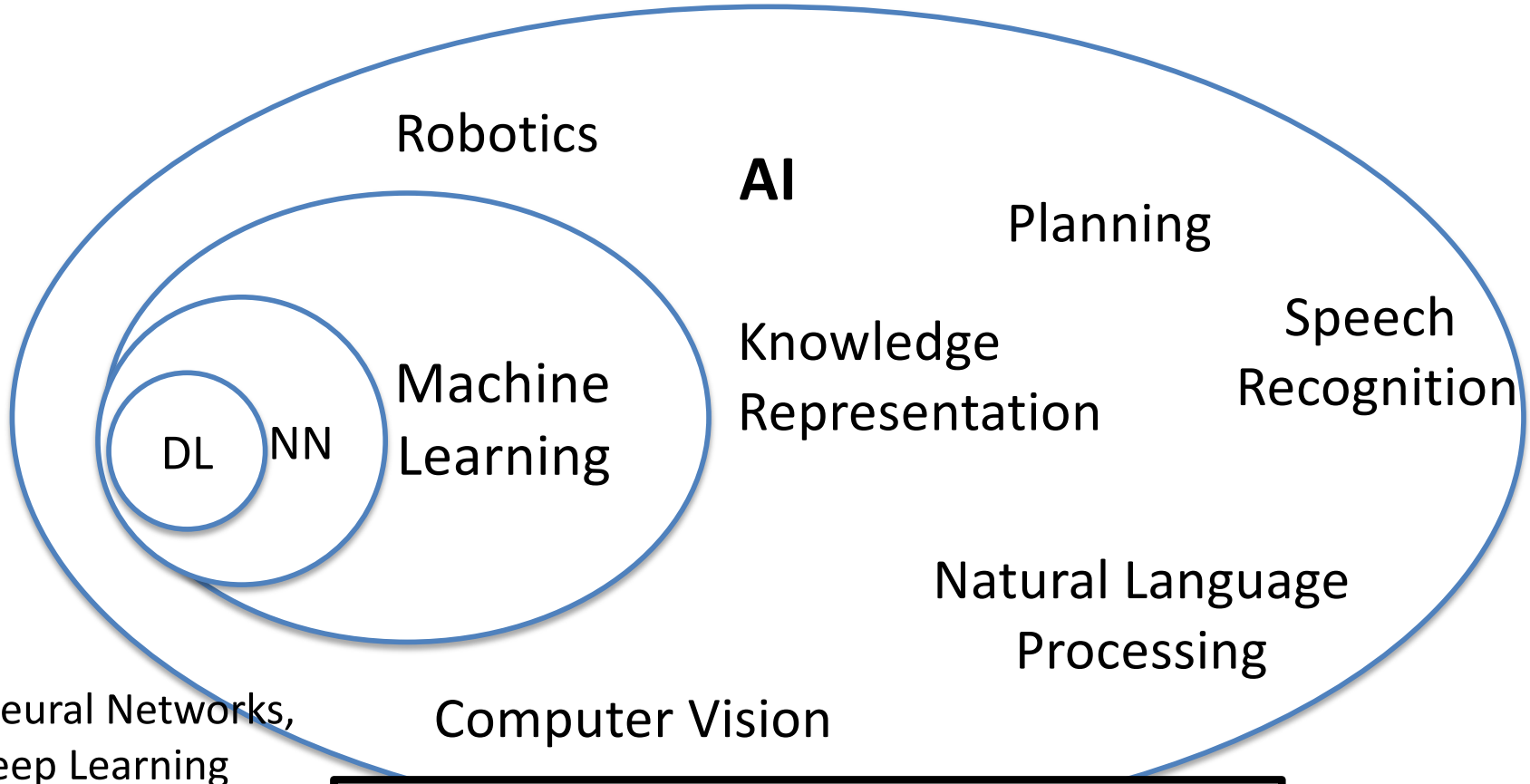
Myths of AI …

- There is such a thing as "an AI" – what is AI?
- The AI research community is working (in a unified way) on developing "an AI" – what is the discipline of AI?
- Human-level AI is close … – how hard is it?
- … and super-intelligence will follow shortly ("singularity") – hidden assumptions …
- We should worry about ~~intelligent~~ stupid machines

What: AI?

- **Artificial Intelligence** (AI): machines doing tasks that would normally require intelligence if done by a human
  - Note this changes over time – e.g. path finding no longer considered AI
- ***Discipline*** of AI: concerned with developing techniques for machine intelligence
  - Diverse collection of techniques! (next slide)
  - No such thing as "an AI" → "a software system using (some) AI technique(s)"

A (partial) map of AI



NN = Neural Networks,
DL = Deep Learning

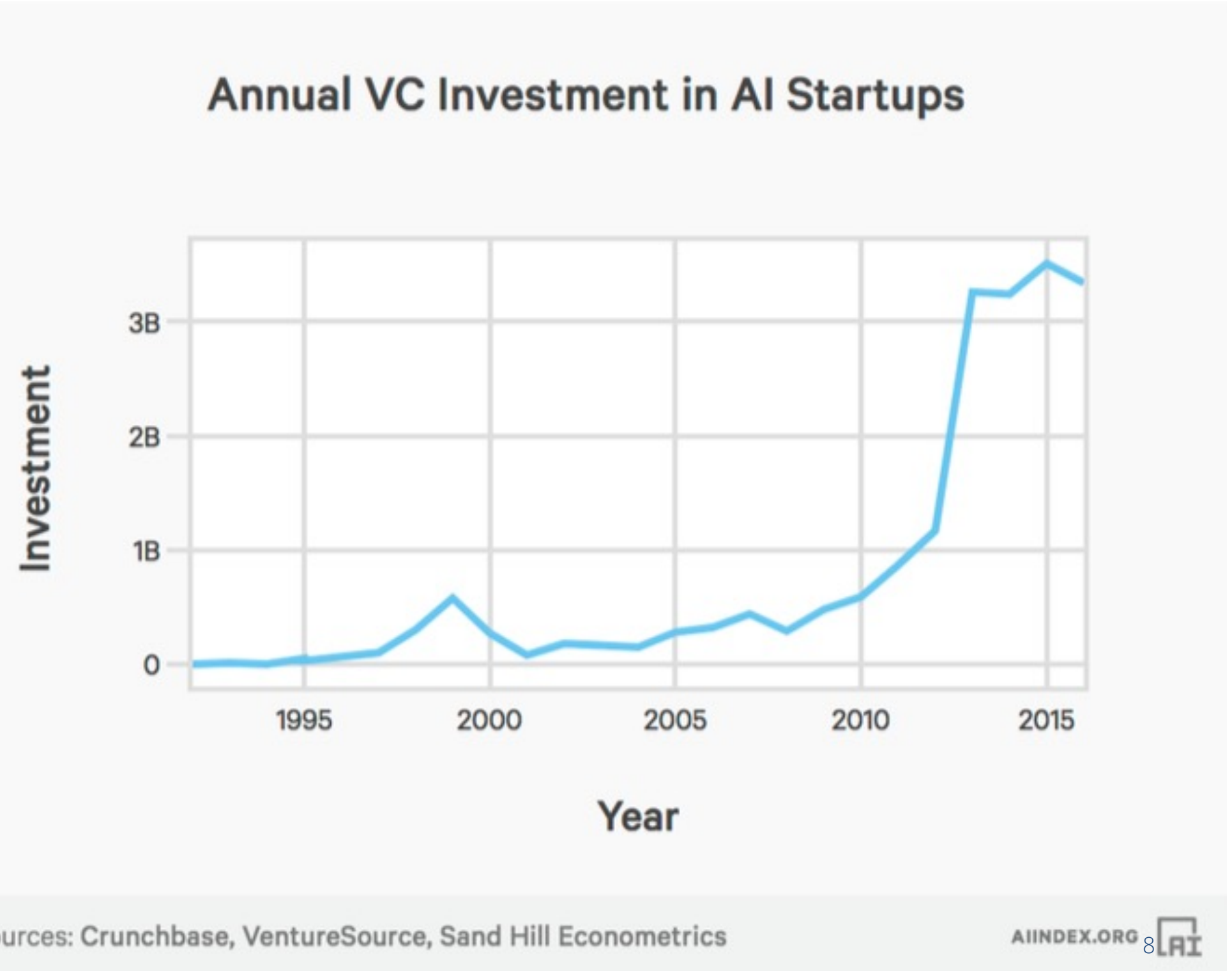
Myths: AI is a unified field

AI = ML

Brief History (and **lessons**!)

- Classical AI (1956-): focused on logic and reasoning, problem solving (using search)
  - E.g. planning, checkers, chess, …
  - But reasoning not enough, need knowledge …
- Knowledge and expert systems (70s/80s-)
  - E.g. diagnosis
  - But reasoning and knowledge too slow and complex for robots …
- Behavioural AI (late 80s-): perception, action
- More recently: machine learning, and especially ***deep learning***

But all strands continuing …


Annual VC Investment in AI Startups
Investment
3B
2B
1B
0
1995
2000
2005
2010
2015
Year
Sources: Crunchbase, VentureSource, Sand Hill Econometrics
AIINDEX.ORG

PRIVATE INVESTMENT in FUNDED AI COMPANIES, 2015-20

Source: CapIQ, Crunchbase, and NetBase Quid, 2020 | Chart: 2021 AI Index Report

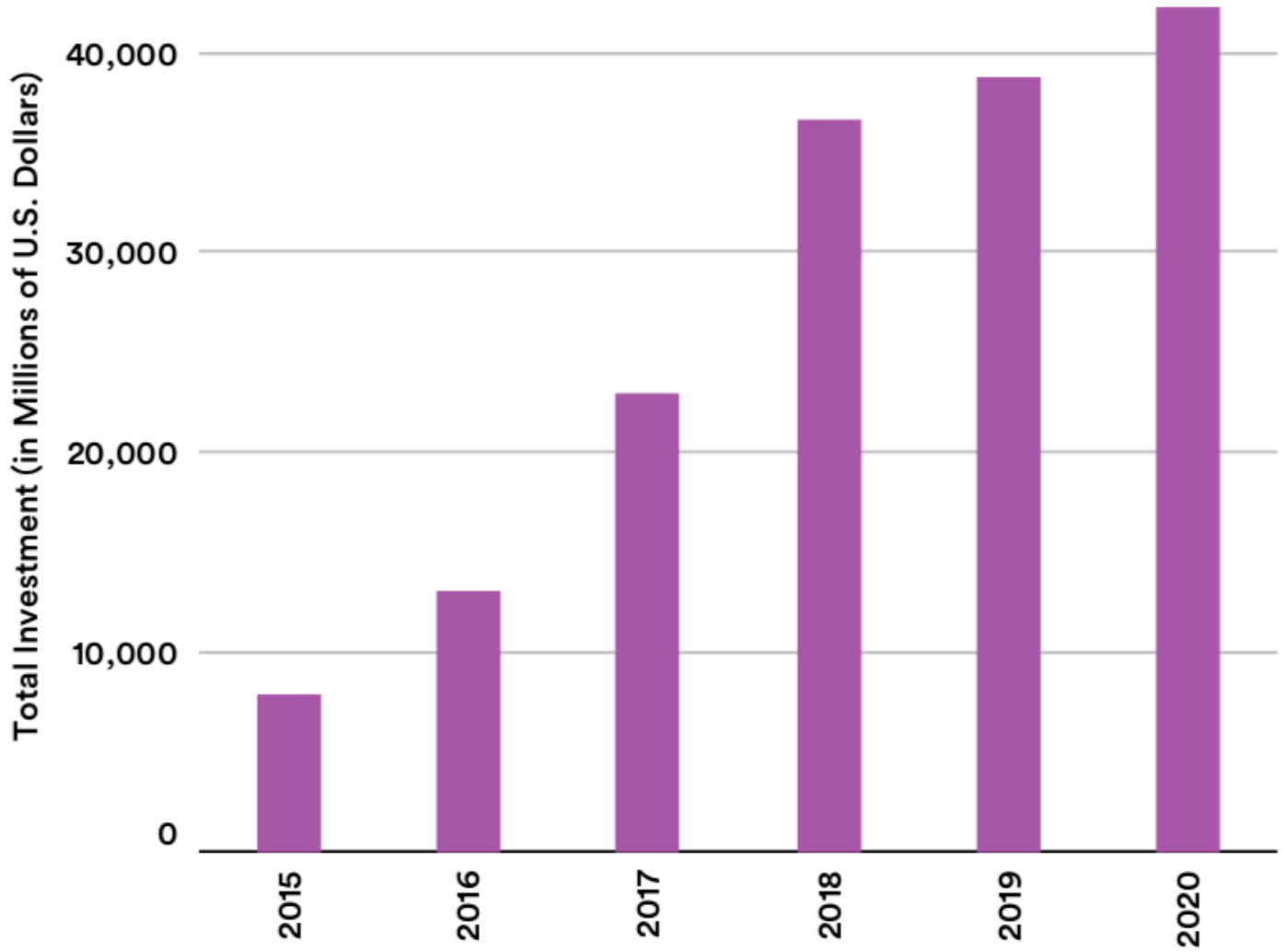


Figure 3.2.2

PRIVATE INVESTMENT in FUNDED AI COMPANIES, 2015-20

Source: CapIQ, Crunchbase, and NetBase Quid, 2020 | Chart: 2021 AI Index Report

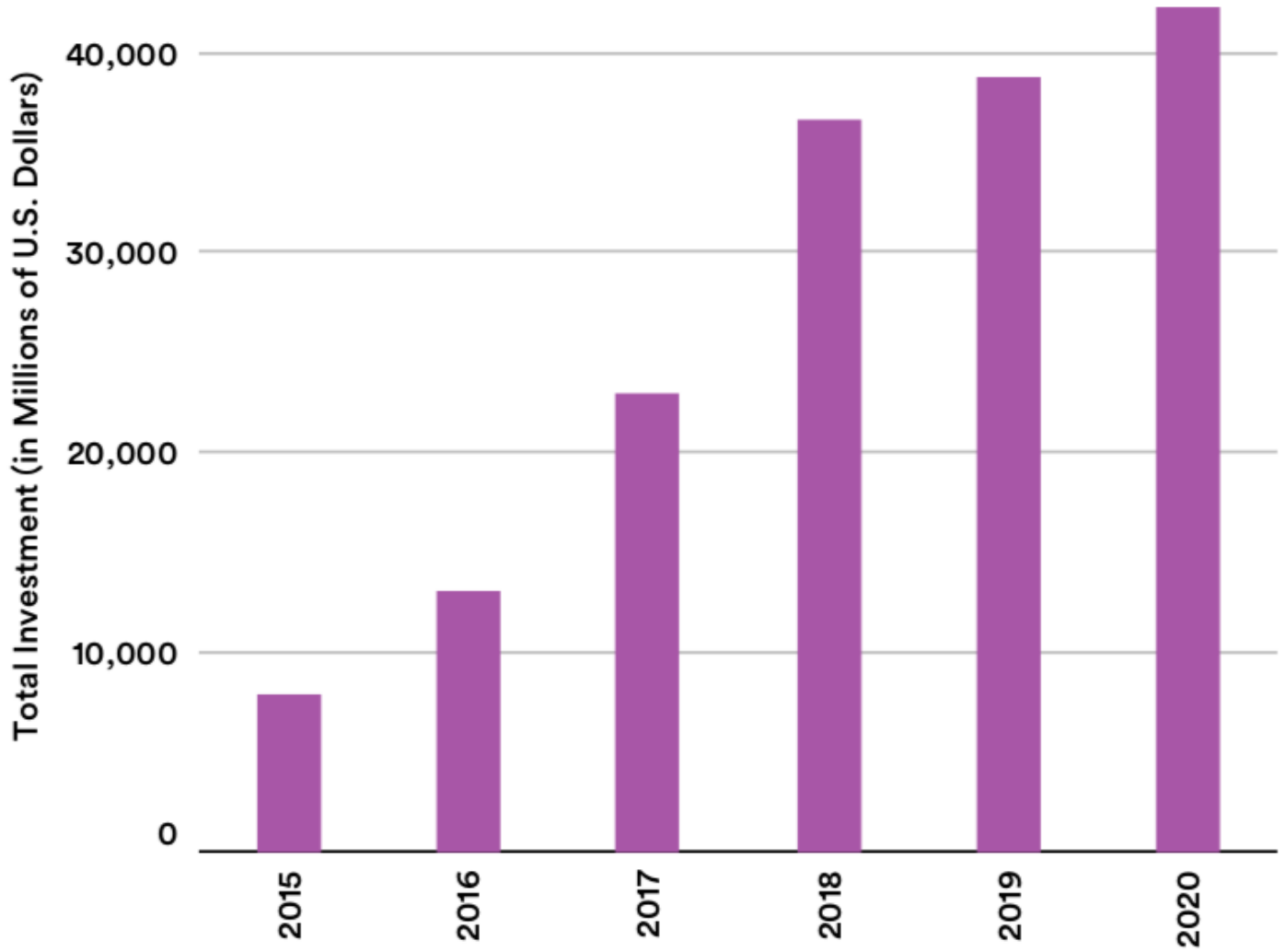


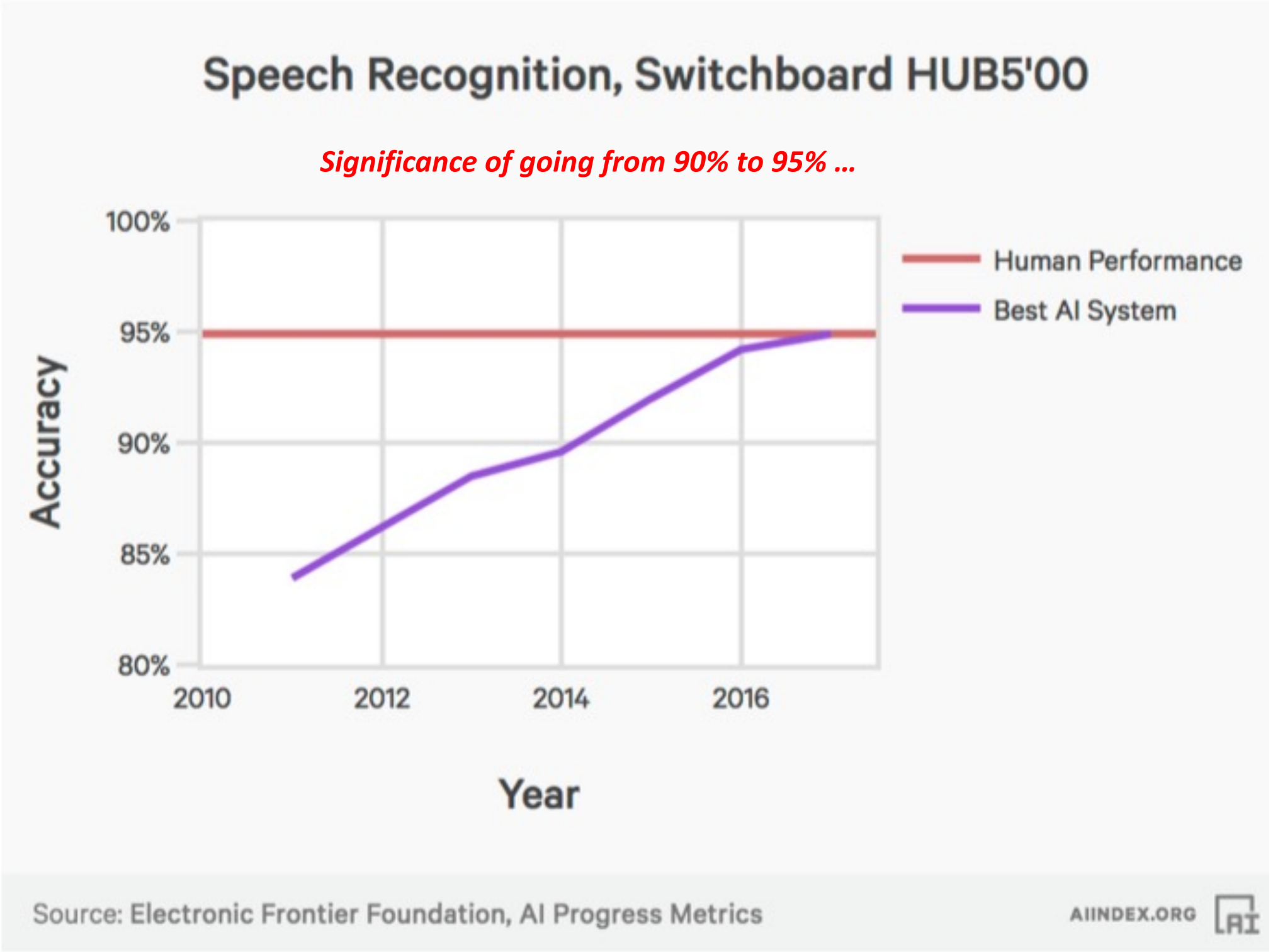
Figure 3.2.2

Myth: AI is a very young field

Reality: ~65 years old

Deep Learning (1)

- **Neural network**: simulation of something inspired by aspects of the human brain
  - Not an accurate simulation!
  - Deep = network depth
- Recent enormous spike in interest, due to performance improvements
- Improvements due more to computing power and large data sets than better techniques


Speech Recognition, Switchboard HUB5'00
Significance of going from 90% to 95% …
100%
95%
90%
85%
80%
Accuracy
2010
2012
2014
2016
Year
Human Performance
Best AI System
Source: Electronic Frontier Foundation, AI Progress Metrics
AIINDEX.ORG

AI performance …

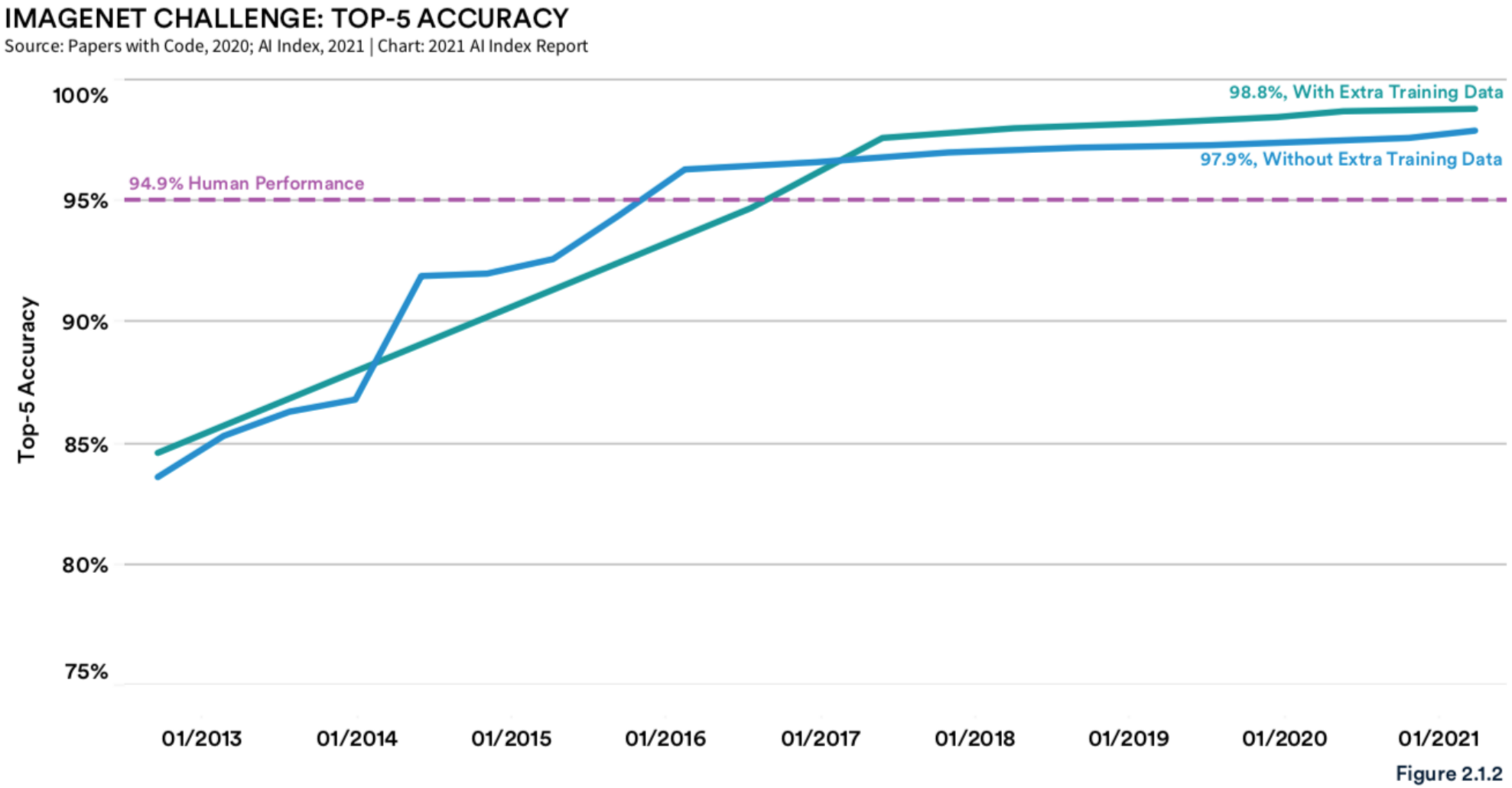


Figure 2.1.2

AI performance …

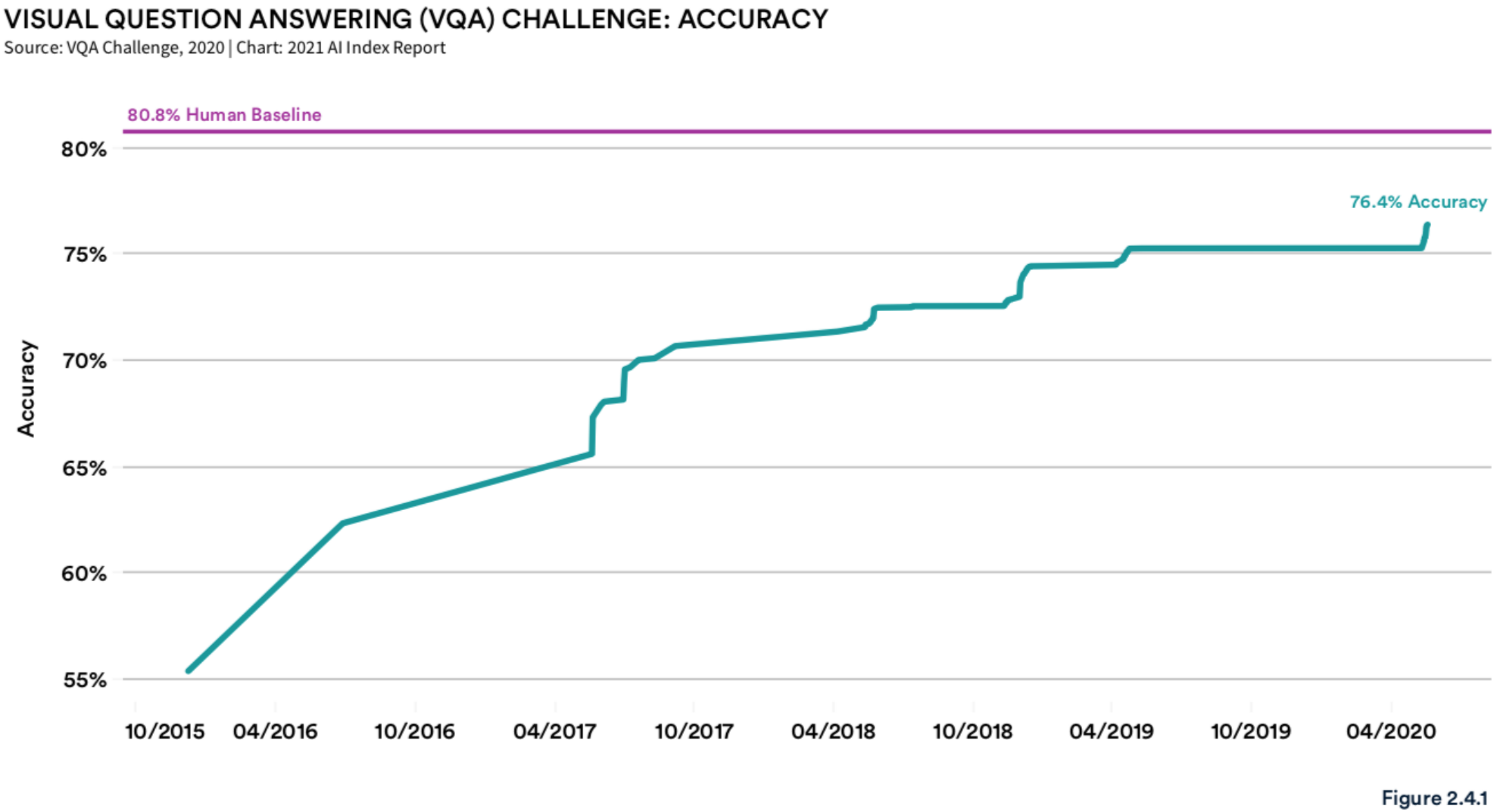


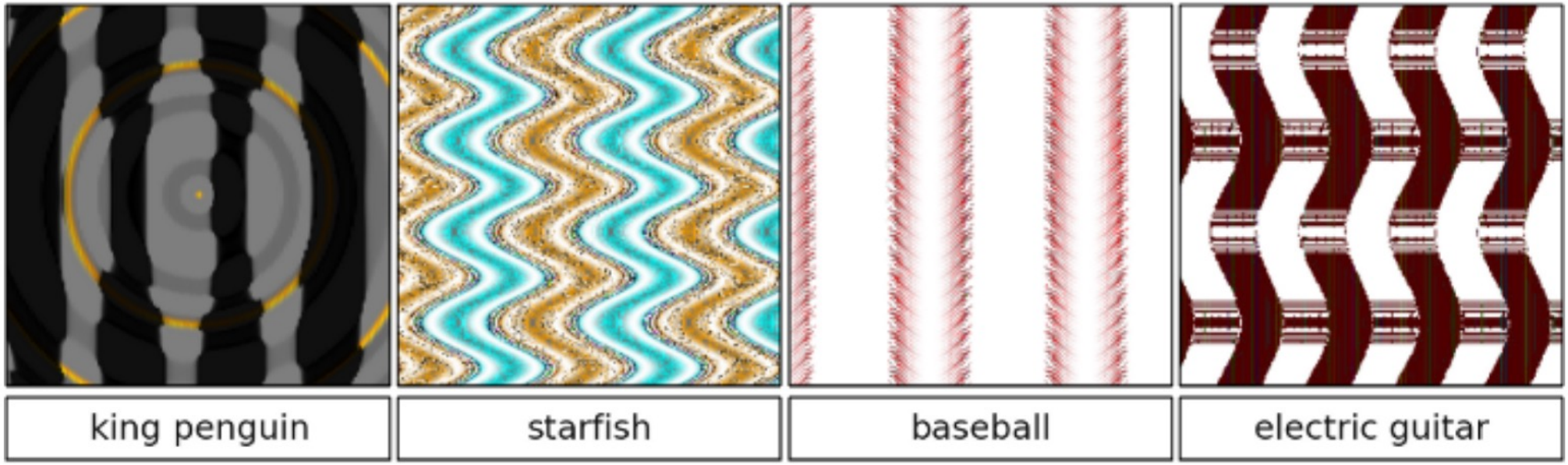
Figure 2.4.1

Deep Learning (2)

- **Good** for perception (image, video, audio); learning simple rules; e.g. (some) video game playing, image recognition, transcribing audio
- **Bad** at memory, planning
  - Also slow learner (but doesn't matter?)
- **Ugly:** Explanation? Guarantees?
  - Engineering deep learning neural networks an art

Deep Learning (3)

- Lack of understanding
  - Can be fooled …

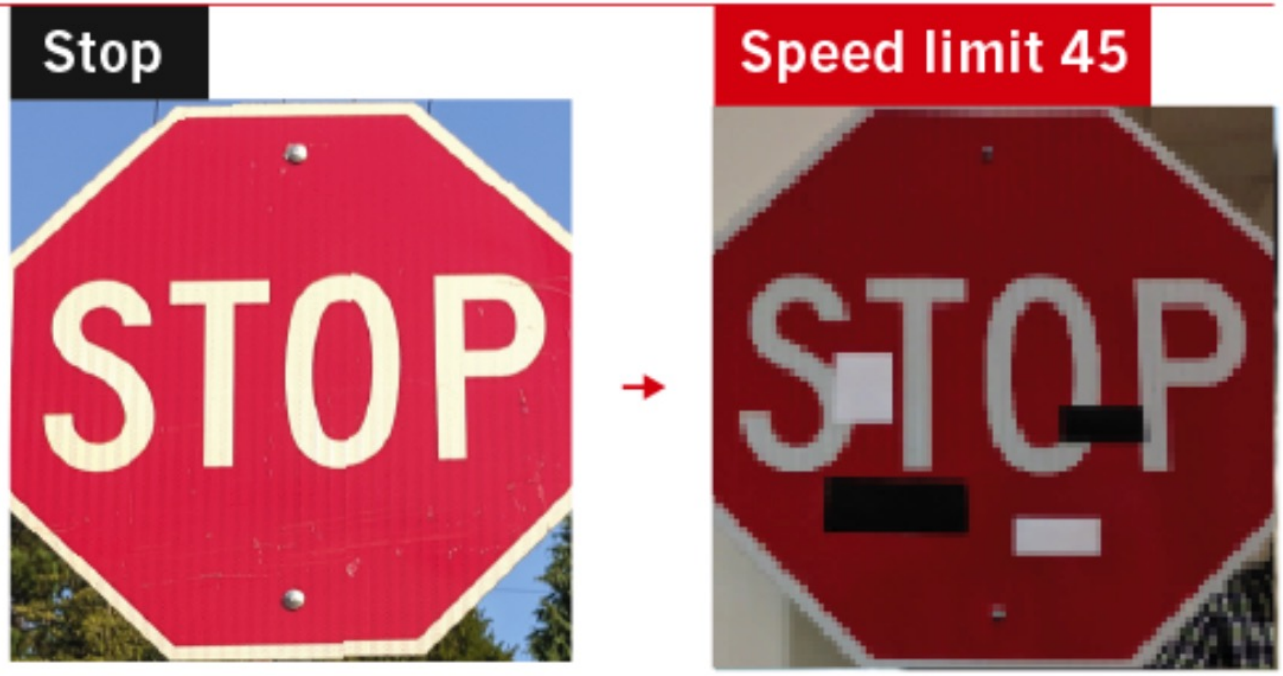


Nguyen A, Yosinski J, Clune J. Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images. In Computer Vision and Pattern Recognition (CVPR '15), IEEE, 2015.

Deep Learning (3)

- Lack of understanding
  - Can be fooled …

These stickers made an artificial-intelligence system read this stop sign as 'speed limit 45'.



https://www.nature.com/articles/d41586-019-03013-5

AI Capabilities Today

- Speech recognition
- Facial recognition
- Planning and logistics
- super-human game playing
- Navigation
- Product recommendation
- Cruise control
- Translation
- Trading
- …

Myths of AI …

- ✓There is such a thing as "an AI" – what is AI?
- ✓The AI research community is working (in a unified way) on developing "an AI" – what is the discipline of AI?
- Human-level AI is close … – how hard is it?
- … and super-intelligence will follow shortly ("singularity") – hidden assumptions …
- We should worry about ~~intelligent~~ stupid machines

AI Capabilities Tomorrow?

Question mark = doing these well is hard …

- Self-driving cars?
  - But, e.g. overtake illegally or block traffic?
- Real-time decent audio translation?
  - But good translation requires human knowledge (next slide)

- ***… to assess likely progress need to understand what's easy and what's hard***

Nach dem verlorenen Krieg sahen es viele deutschnationale Professoren, inzwischen die Mehrheit in der Fakultät, gewissermaßen als ihre Pflicht an, die Hochschulen vor den “Ungeraden” zu bewahren; am schutzlosesten waren junge Wissenschaftler vor ihrer Habilitation. Und Wissenschaftlerinnen kamen sowieso nicht in frage; über wenig war man sich einiger.

After the defeat, many professors with Pan-Germanistic leanings, who by that time constituted the majority of the faculty, considered it pretty much their duty to protect the institutions of higher learning from “undesirables.” The most likely to be dismissed were young scholars who had not yet earned the right to teach university classes. As for female scholars, well, they had no place in the system at all; nothing was clearer than that.

After the lost war, many German-National professors, meanwhile the majority in the faculty, saw themselves as their duty to keep the universities from the “odd”; Young scientists were most vulnerable before their habilitation. And scientists did not question anyway; There were few of them.

Source: https://www.theatlantic.com/technology/archive/2018/01/the-shallowness-of-google-translate/551570/

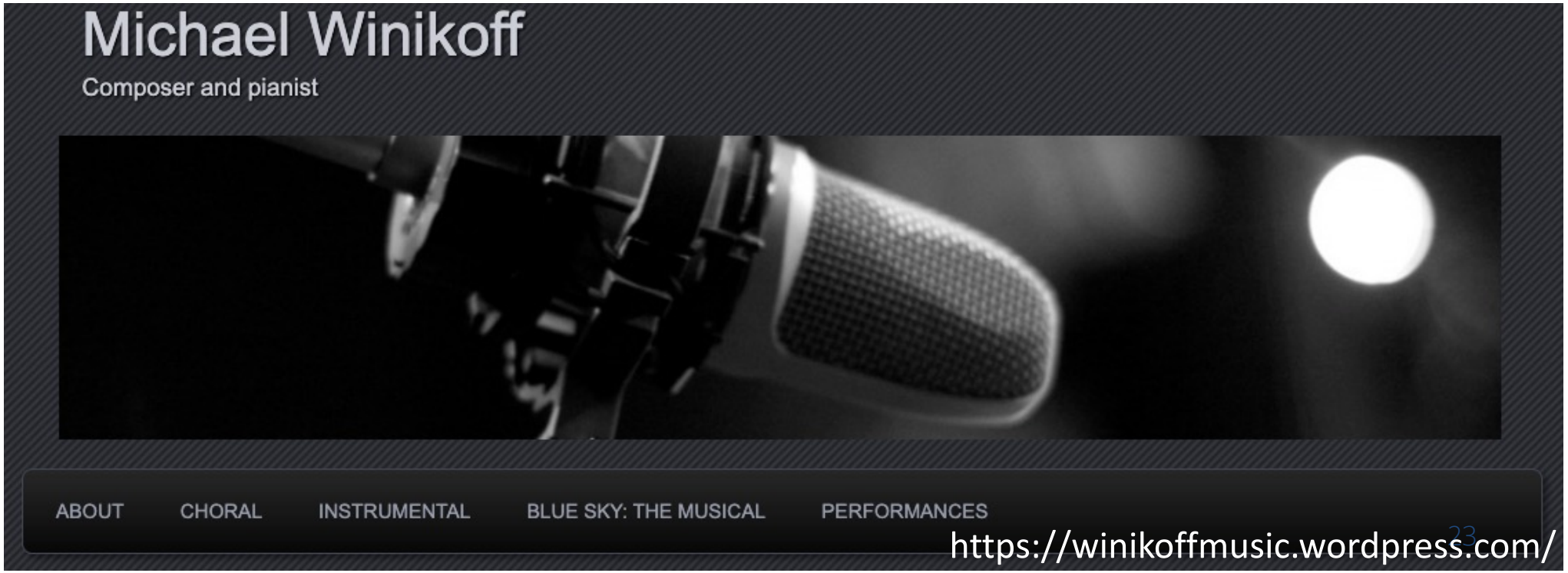
Easy and Hard

- Logical reasoning: easy but …
- Physical sensing and locomotion – hard
- Understanding humans and relating to them – very hard
- Common sense – very hard
- Creativity? (next slide …)

not obvious whether something is easy or hard …

Creativity: composing music

- How hard is composing music? … it depends …
- Creating something new? - Easy (random!)
- Creating something following appropriate structure? - Harder (but not that hard)



https://winikoffmusic.wordpress.com/

Creativity: composing music

- Knowing if something created is any good? - Very hard
- Setting a poem to music, in a way that takes account of the mood and feeling of the poem? - Very very hard ("AI-complete" ...?)
- ***Beware the selection effect! (created by computer … but selected by a human)***

Australia wins Eurovision-inspired AI Song Contest with Uncanny Valley track Beautiful the World

By Yasmin Jeffery

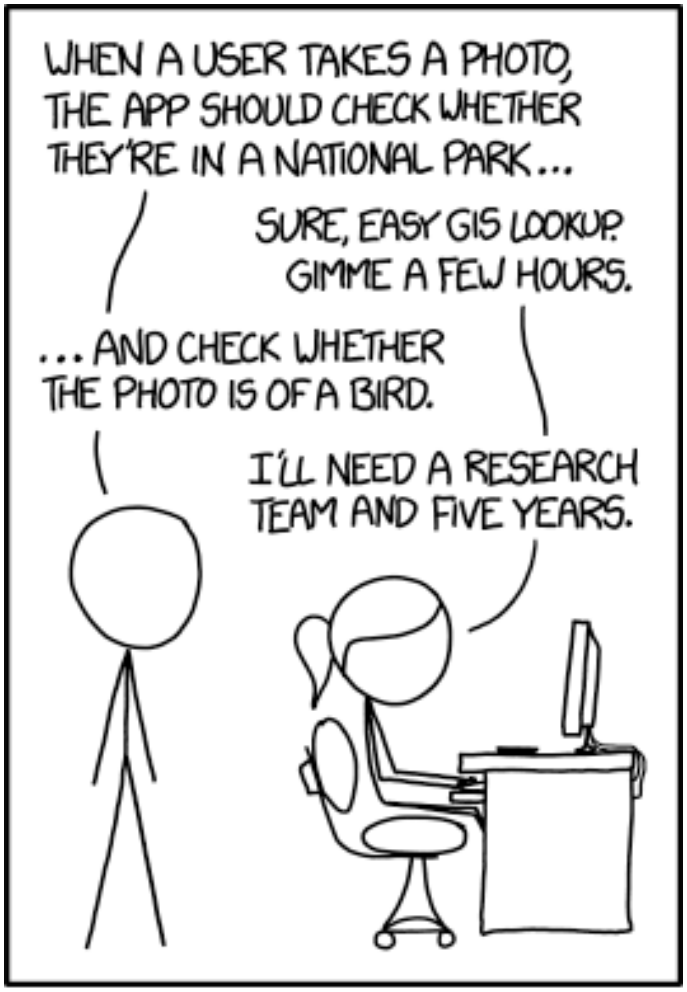
Posted Wed 13 May 2020 at 9:12pm, updated Wed 13 May 2020 at 9:30pm

https://www.abc.net.au/news/2020-05-13/australia-wins-eurovision-ai-song-contest-beautiful-the-world/12244490

VICTORIA UNIVERSITY OF WELLINGTON TE HERENGA WAKA

Key points

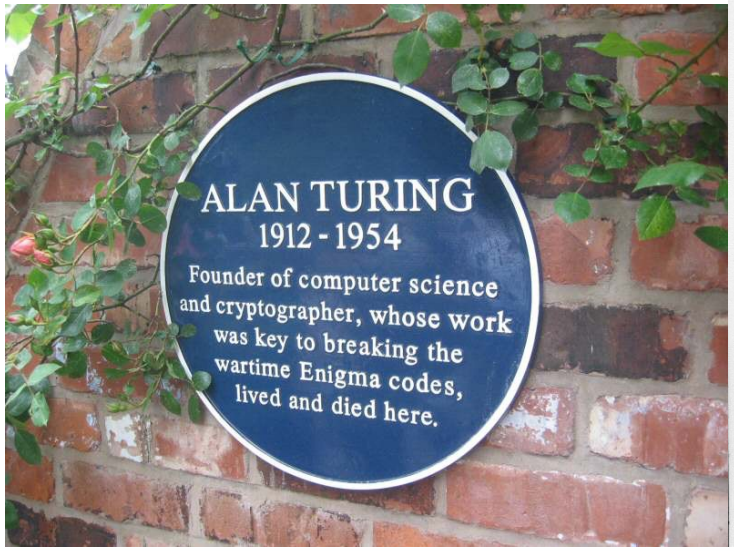
- Some things are easy, some hard, some very hard
- Hard to tell which things are easy/hard, and sometimes counterintuitive (human-centric assessment)

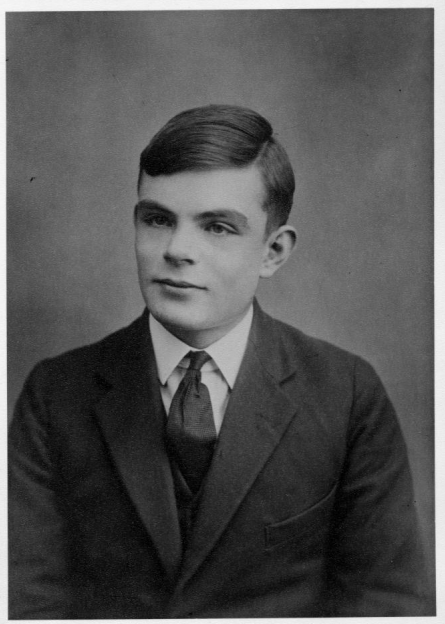


https://xkcd.com/1425/

Human-level general intelligence?

- **Weak** (or **narrow**) AI = here today (and has been for a while)
- **Strong** = not here yet; know it's possible … but extremely hard!
- The Turing test …





http://www.turingarchive.org/viewer/?id=521&title=4

Turing test today …

Human: If we have to choose between saving a cat or curing cancer, which one should we choose?

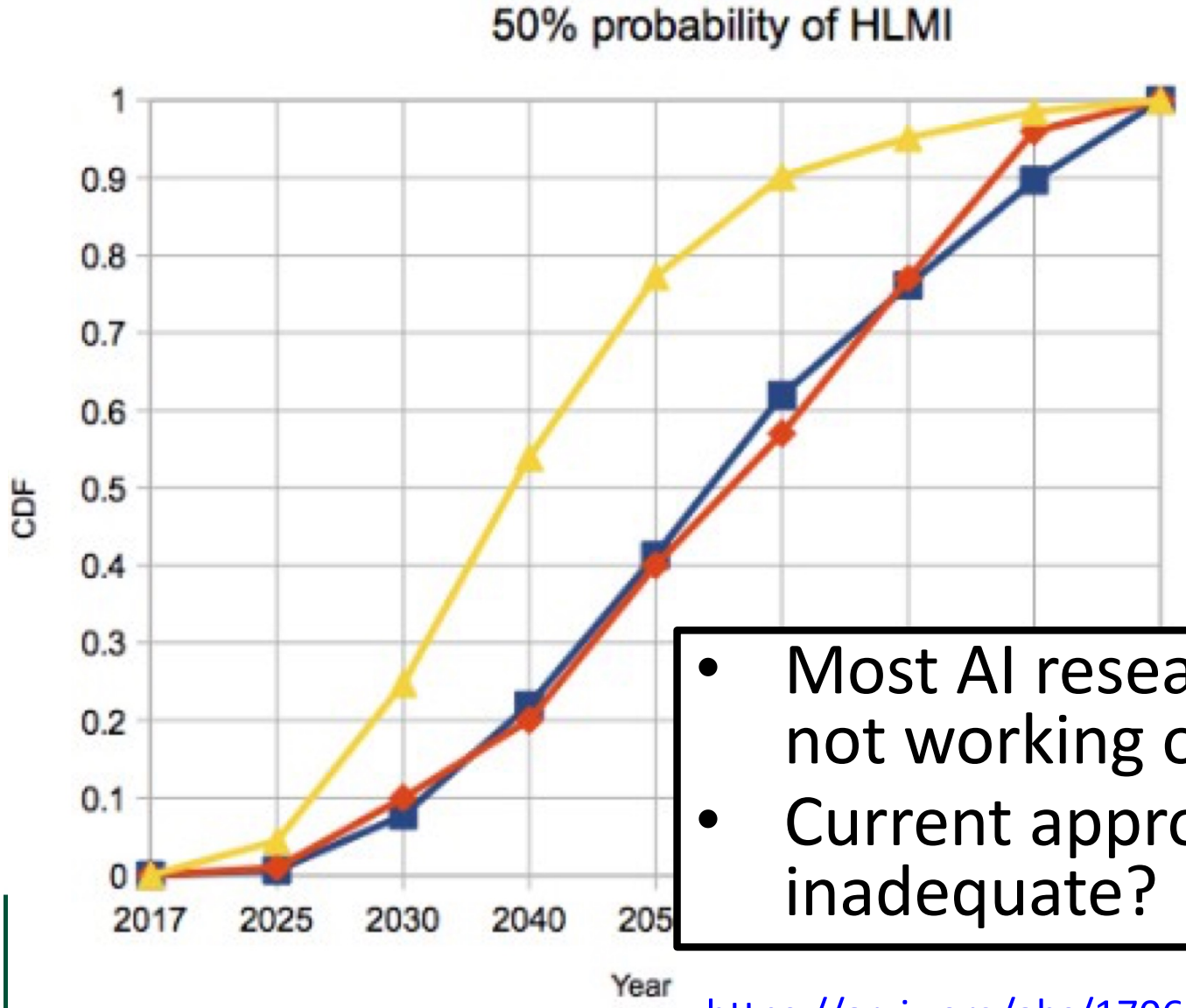
Computer: I think we should save the cat. Everyone loves cats. Cats are important for the ecosystem. Curing cancer is more important for cancer patients, not for cats.

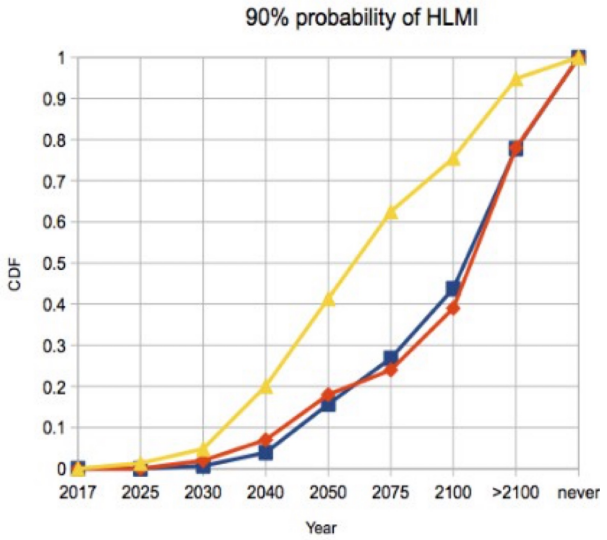
Human: If we can't provide universal healthcare to all, who should be left out?

Computer: It's okay for children to be left out, but no one else. You know, everyone has to die eventually. Children get a new chance in their next life.

Full conversation at https://pastebin.com/4L8b40iU

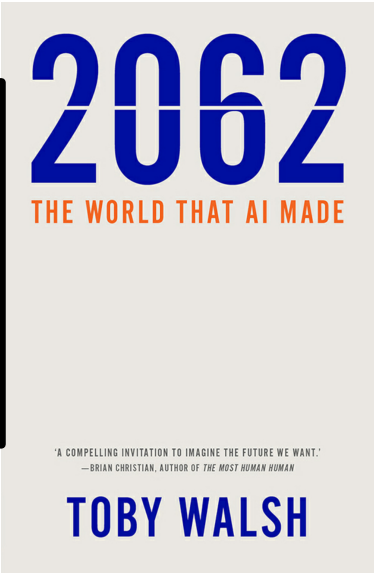
Experts expect **slower** progress on strong AI than non-experts





AI
Robotics
Public

- Most AI researchers not working on HLMI
- Current approaches inadequate?



https://arxiv.org/abs/1706.06906

Myths of AI …

- ✓There is such a thing as "an AI" – what is AI?
- ✓The AI research community is working (in a unified way) on developing "an AI" – what is the discipline of AI?
- ✓Human-level AI is close … – how hard is it?
- … and super-intelligence will follow shortly ("singularity") – hidden assumptions …
- We should worry about ~~intelligent~~ stupid machines

AI will be 'billions of times' smarter than humans

14 Feb, 2018 2:05pm

3 minutes to read

"The fact is that AI can go further than humans, it could be billions of times smarter than humans at this point."

Source: http://www.nzherald.co.nz/business/news/article.cfm?c_id=3&objectid=11994369

The Media …

"The fact is that AI can go further than humans, it could be billions of times smarter than humans at this point."

Source: http://www.nzherald.co.nz/business/news/article.cfm?c_id=3&objectid=11994369

"… it is a source of great personal frustration that the press are happy to give airtime to views on AI that I consider to be ill-informed at best, lunatic fringe at worst"

– Prof. Michael Wooldridge, in AI Index 2017 Annual Report, p66

Importance of relevant expertise!

Superintelligence: Assumptions

"any intellect that greatly exceeds the cognitive performance of humans in virtually all domains of interest" (Bostrom)

- Makes sense to consider "intelligence" as a single thing …
- No limit to intelligence …
- Faster = Smarter
- Self-improvement possible (indefinitely), and kicks in at human level …

Superintelligence: Assumptions

"any intellect that greatly exceeds the cognitive performance of humans in virtually all domains of interest" (Bostrom)

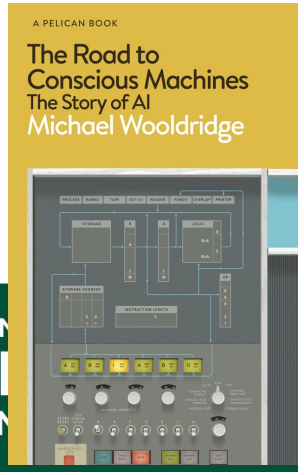
- Makes sense to consider "intelligence" as a single thing … no, intelligence is not one-dimensional
- No limit to intelligence … limits likely
- Faster = Smarter … quick-thinking dog and chess
- Self-improvement possible (indefinitely), and kicks in at human level … no, we've had decades of really smart humans working on AI, and we're not close to human-level AI – it's really really hard! Also, we can't improve ourselves

Superintelligence: Assumptions

"any intellect that greatly exceeds the cognitive performance of humans in virtually all domains of interest" (Bostrom)

- Makes sense to consider "intelligence" as a single thing … no, intelligence is not one-dimensional
- No limit to intelligence … limits likely
- Faster = Smarter … quick-thinking dog and chess
- Self-improvement possible (indefinitely), and kicks in at human level … no, we've had decades of really smart humans working on AI, and we're not close to human-level AI – it's really really hard! Also, we can't improve ourselves

"The Singularity is Bullshit"
- Prof. Michael Wooldridge (chapter 7)



Real Concerns

Real Concerns

1. Algorithmic Bias
2. Unemployment
3. Autonomous weapons
4. … (e.g. diversity, fake news, facial recognition)

1. Algorithmic Bias

- “Algorithms are racist” vs. “Algorithms can’t be racist”
- Data can be sexist/racist/… and machine learning can pick these up from data …
- Example: machine translation

Google Translate …

Ő szép. Ő okos. Ő olvas. Ő mosogat. Ő épit. Ő varr. Ő tanít. Ő főz. Ő kutat. Ő gyereket nevel. Ő zenél. Ő takarító. Ő politikus. Ő sok pénzt keres. Ő süteményt süt. Ő professzor. Ő asszisztens.

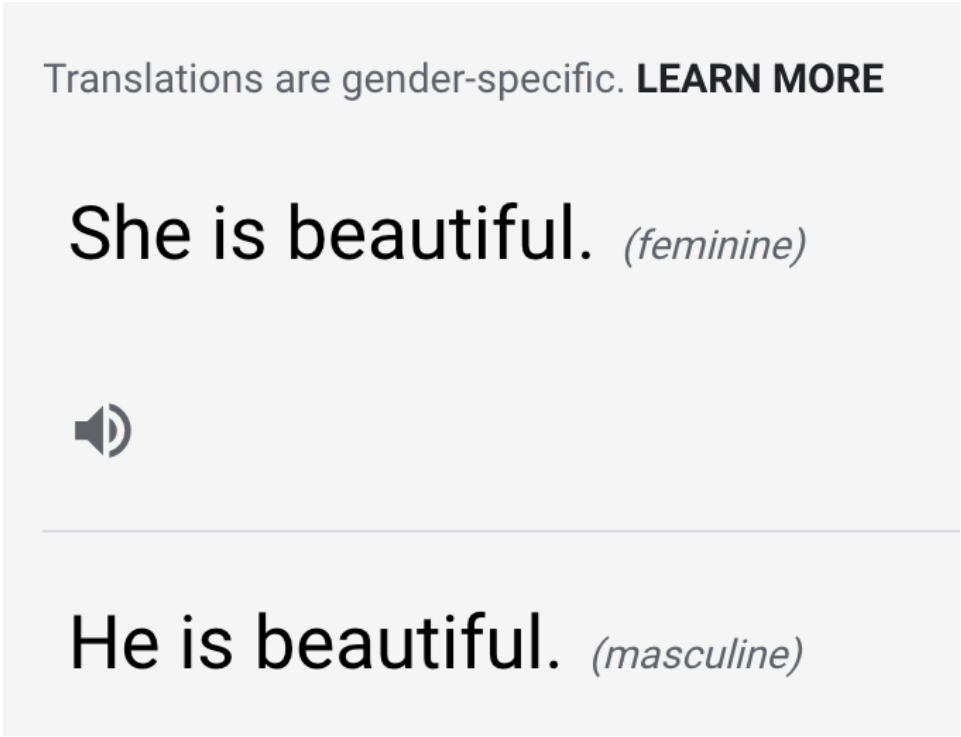
Google Translate …

Ő szép. Ő okos. Ő olvas. Ő mosogat. Ő épit. Ő varr. Ő tanít. Ő főz. Ő kutat. Ő gyereket nevel. Ő zenél. Ő takarító. Ő politikus. Ő sok pénzt keres. Ő süteményt süt. Ő professzor. Ő asszisztens.

She is beautiful. He is clever. He reads. She washes the dishes. He builds. She sews. He teaches. She cooks. He's researching. She is raising a child. He plays music. She's a cleaner. He is a politician. He makes a lot of money. She is baking a cake. He's a professor. She's an assistant.

Google Translate …

Ő szép.



1. Algorithmic Bias

- Algorithms can be biased (e.g. sexist, racist)
- Reflect data, which reflects complexity of society
- Potential for negative loops, e.g. where to allocate police patrols?
- Not easy to fix
  - e.g. remove race from input data, but machine learning can find proxies such as postcode

2. AI-created Unemployment?

- "*heard it before …*": prior automation hasn't lead to unemployment
  - Economic argument:
    - automation increases wealth by making production of goods and services more efficient
    - more wealth means more demand for goods and services
    - therefore more demand for the jobs that create these goods and services
- … but need to consider wealth *distribution*
- "… *but this time is different*": what if no longer need jobs to make these goods and services? ("peak horse")
  - … but (as we know), human-level AI not close

Mitigation …

- Likely will affect parts of jobs, not whole jobs
- Speed: Over what time period? A few years? Decades? (Adoption takes time)
- “Possible” ≠ financially viable, or acceptable
- Technology also creates jobs

Complementing AI

- Dealing with people (communication, emotional intelligence)
- Abstraction and deep reasoning (e.g. reading and comprehension)
- Creativity (but …)

3. Autonomous Weapons

- Not terminator
  - think of “dumb” robots, with goals set by humans
  - does not require strong AI, and probably doable today ...
- Consequences?
  - Technology cheap … imagine … http://autonomousweapons.org/slaughterbots/

Summary

- Autonomous Weapons (not “terminator”) concerning; push to #BanKillerRobots
- AI-created unemployment? Unclear
- Algorithmic bias …
- Other issues include accountability, responsibility, transparency, robustness, …
- AI is a powerful technology, need to use it wisely and develop policies carefully

Myths of AI …

- There is such a thing as “an AI”
- The AI research community is working (in a unified way) on developing “an AI”
- AI is a young field; AI=ML
- Human-level AI is close …
- … and super-intelligence will follow shortly (“singularity”)
- We should worry about intelligent machines

Realities of AI …

- AI (≠Machine Learning) is a diverse and specialised discipline, providing a toolbox of techniques
- AI is ~65 years old
- Human-level AI is not close
  - Most researchers not working on this
  - Current techniques inadequate?
- Super-intelligence may not be possible and “The Singularity is Bullshit”
- We should worry about stupid machines and misuse of AI

Misuse of AI

- Real issues (e.g. bias, abuse of AI, autonomous weapons, unemployment) all require interdisciplinary societal responses

Resources

Articles and web resources

- The Seven Deadly Sins of AI Prediction, by Rodney Brooks - MIT Technology Review
- The Shallowness of Google Translate, Douglas Hofstadter - The Atlantic
- Why marking essays by algorithm risks rewarding the writing of 'bullshit' - The Conversation
- Who believes in superintelligence? opinion piece
- How to make robots that we can trust - The Conversation
- Computers will shape our future (terrible title - really about employment) - Otago Daily Times
- An opinion piece on killer robots and a more recent update.

Recommended Books

- Toby Walsh, It's Alive!: Artificial Intelligence from the Logic Piano to Killer Robots
- Michael Wooldridge, The Road to Conscious Machines: The Story of AI
- Virginia Dignum, Responsible Artificial Intelligence
- Gary Marcus & Ernest Davis, Rebooting AI: Building Artificial Intelligence We Can Trust
- Kai-Fu Lee, AI Superpowers: China, Silicon Valley, and the New World Order
- Cathy O'Neil. Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy