State and Trends
Where is artificial intelligence heading to?
## Brain vs. Computer

<table>
<thead>
<tr>
<th>Brain</th>
<th>Computer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consciousness</td>
<td>Software</td>
</tr>
<tr>
<td>sequential</td>
<td>parallel</td>
</tr>
<tr>
<td>Mindware</td>
<td>Hardware</td>
</tr>
<tr>
<td>parallel</td>
<td>sequential</td>
</tr>
</tbody>
</table>

- **easy**  Pattern Recognition  **hard**
- **hard**  Logic and Thinking  **easy**

... but there is massive progress*!

* We have superhuman image recognition since February 2015.

---

Dennett: Consciousness Explained
www.amazon.com/dp/0316180661
State of the Art

Deep Blue: 1997
Stanley: 2005
IBM Watson: 2011
Schmidhuber: 2011

<table>
<thead>
<tr>
<th>Game</th>
<th>Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Checkers</td>
<td>Superhuman</td>
</tr>
<tr>
<td>Backgammon</td>
<td>Superhuman</td>
</tr>
<tr>
<td>Othello</td>
<td>Superhuman</td>
</tr>
<tr>
<td>Chess</td>
<td>Superhuman</td>
</tr>
<tr>
<td>Crosswords</td>
<td>Expert Level</td>
</tr>
<tr>
<td>Scrabble</td>
<td>Superhuman</td>
</tr>
<tr>
<td>Bridge</td>
<td>Equal to Best</td>
</tr>
<tr>
<td>Jeopardy!</td>
<td>Superhuman</td>
</tr>
<tr>
<td>Poker</td>
<td>Varied</td>
</tr>
<tr>
<td>FreeCell</td>
<td>Superhuman</td>
</tr>
<tr>
<td>Go</td>
<td>Strong Amateur</td>
</tr>
</tbody>
</table>

How bio-inspired deep learning keeps winning competitions
www.kurzweilai.net/how-bio-inspired-deep-learning-[…] Superintelligence
State and Trends
Consumer Products

Siri

INTRODUCING

amazon echo

Raffaello D'Andrea
go.ted.com/xeh

Superintelligence
State and Trends
Military Robots

P.W. Singer
go.ted.com/xe3
Financial Markets

- High-frequency trading (HFT): Buy and sell securities within milliseconds algorithmically
- In 2009, 65% of all US equity trading volume
- Flash crash: very rapid fall in security prices
- 6 May 2010: Dow Jones lost $1 trillion (over 9%)
- 23 April 2013: One tweet causes $136 billion loss

Kevin Slavin
go.ted.com/xee

Superintelligence
State and Trends
Automation brings abundance

Many industries will be transformed into pure AI industries…

Capital: ↑, Wages: ↓

Should be beneficial if we manage it well – but we are not prepared!
Make sure to work above the API!

Google and Uber Will Challenge Jobs 'Below The API'
www.forbes.com/sites/anthonykosner/2015/02/04/[...]
The Great Decoupling of the US Economy
andrewmcafee.org/2012/12/[…]

Superintelligence
State and Trends
We’ll give robots full autonomy

… because of
– increased speed
– high complexity
– risk of jamming

Current examples:
– financial markets
– auton. weapons

AI is the ultimate productivity boost!

Daniel Suarez
go.ted.com/biXE
Intelligence Explosion
Can artificial intelligence run away?
Interview by John Oliver with Stephen Hawking
www.youtube.com/watch?v=T8y5EXFMD4s
Modelling Capabilities

An advanced AI will also model its operators and go to great lengths to prevent being switched off!

It will behave nicely and cooperatively until the external threats are under control and it is ready for takeover.

Ex Machina Movie
exmachina-movie.com
Optimization Power

Problem: When optimizing a system, unspecified parameters often assume extreme values.

You will get what you wished for and not what you wanted.
Feedback

Systems can feed back into themselves and thus must be analyzed as a whole!

Feedback is either:
- Positive (reinforcing)
- Negative (balancing)
Exponential Functions

If increase is linear to current amount:

\[ \frac{d}{dx} f(x) = c \cdot f(x) \]

solved by

\[ f(x) = e^{c \cdot x} \]

Fold a paper 45 times \( \Rightarrow \) to the moon!

How folding a paper can get you to the moon
www.youtube.com/watch?v=AmFMJC45f1Q
Climate Change

Absorb less CO\textsubscript{2} → Warmer oceans → Rising temperature → More heat absorption → Less reflection → Less ice → Melting ice → Stronger greenhouse effect

Warmer oceans + Rising temperature + Less reflection + Melting ice + More heat absorption

en.wikipedia.org/wiki/Climate_change_feedback
Nuclear Chain Reaction

1. $^{235}\text{U}$
2. $^{239}\text{U}$
3. $^{235}\text{U}$

Superintelligence
Intelligence Explosion

en.wikipedia.org/wiki/Nuclear_chain_reaction
Accelerating Change

Progress feeds on itself:

Rate of Progress in the year 2’000

The Law of Accelerating Returns
www.kurzweilai.net/the-law-of-accelerating-returns

Superintelligence
Intelligence Explosion
Moore’s Law

Exponential and Non-Exponential Trends in IT
intelligence.org/[...]/exponential-and-non-exponential/

Superintelligence
Intelligence Explosion
Hyperbolic Growth

Second-order positive feedback loop

\[
\frac{df(t)}{dt} = c \cdot f(t)^2 \Rightarrow f(t) = \frac{-1}{c \cdot t}
\]

\(f(t)\) reaches infinity in finite time

Singularity

Mathematical Singularity
en.wikipedia.org/wiki/Singularity_(mathematics)

Superintelligence
Intelligence Explosion
Artificial Mind

Imagine all relevant aspects captured in a computer model (thought experiment)

Whole Brain Emulation: A Roadmap
www.fhi.ox.ac.uk/brain-emulation-roadmap-report.pdf

Superintelligence
Intelligence Explosion
Speed Explosion

Computing speed doubles every two subjective years of work.

Marcus Hutter: Can Intelligence Explode?
www.hutter1.net/publ/singularity.pdf

Superintelligence
Intelligence Explosion
Population Explosion

Computing costs halve for a certain amount of work.

Ray Solomonoff: The Time Scale of Artificial Intelligence
citeseerx.ist.psu.edu/...summary?doi=10.1.1.147.3790

Population Explosion

Quantitative
Intelligence Explosion

Proportionality Thesis: An increase in intelligence leads to similar increases in the capacity to design intelligent systems.

Recursive Self-Improvement

Intelligence Explosion
intelligence.org/files/IE-EI.pdf
Three Separate Explosions

Speed
- more speed
- more time

Population
- more people
- more research

Intelligence
- better algorithms
- better research

David Chalmers: The Singularity
consc.net/papers/singularity.pdf
Technological Singularity

Theoretic phenomenon: There are arguments why it should exist but it has not yet been confirmed experimentally.

Three major singularity schools:
– Accelerating Change (Ray Kurzweil)
– Intelligence Explosion (I.J. Good)
– Event Horizon (Vernor Vinge)
Superintelligence

What are potential outcomes?
Definition of Superintelligence

An agent is called superintelligent if it exceeds the level of current human intelligence in all areas of interest.

Pathways to Superintelligence

- artificial intelligence
- neuromorphic
- synthetic
- whole brain emulation
- biological cognition
- brain-computer interfaces
- networks and organizations

Embryo Selection for Cognitive Enhancement
www.nickbostron.com/papers/embryo.pdf
Advantages of AIs over Brains

<table>
<thead>
<tr>
<th>Hardware:</th>
<th>Software:</th>
<th>Effectiveness:</th>
</tr>
</thead>
<tbody>
<tr>
<td>– Size</td>
<td>– Editability</td>
<td>– Rationality</td>
</tr>
<tr>
<td>– Speed</td>
<td>– Copyability</td>
<td>– Coordination</td>
</tr>
<tr>
<td>– Memory</td>
<td>– Expandability</td>
<td>– Communication</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Human Brain</th>
<th>Modern Microprocessor</th>
</tr>
</thead>
<tbody>
<tr>
<td>86 billion neurons</td>
<td>1.4 billion transistors</td>
</tr>
<tr>
<td>firing rate of 200 Hz</td>
<td>4’400’000’000 Hz</td>
</tr>
<tr>
<td>120 m/s signal speed</td>
<td>300’000’000 m/s</td>
</tr>
</tbody>
</table>
Cognitive Superpowers

- Intelligence amplification: bootstrapping
- Strategizing: overcome smart opposition
- Hacking: hijack computing infrastructure
- Social manipulation: persuading people
- Economic productivity: acquiring wealth
- Technology research: inventing new aids
Intelligence and final goals are orthogonal:
Almost any level of intelligence could in principle be combined with any final goal.
Convergent Instrumental Goals

- Self-Preservation
- Goal-Preservation
- Resource Accumulation
- Intelligence Accumulation

Default Outcome: Doom (Infrastructure Profusion)
Single-Shot Situation

Our first superhuman AI must be a safe one for we may not get a second chance!

- We’re good at iterating with testing and feedback
- We’re terrible at getting things right the first time
- Humanity only learns when catastrophe occurred

List of Cognitive Biases
en.wikipedia.org/wiki/List_of_cognitive_biases
Takeoff Scenarios

Physical Limit

Superintelligence

Human Level

Feedback

Intelligence

Time

now

time until takeoff

takeoff duration

Separate Questions!

The Hanson-Yudkowsky AI-Foom Debate
intelligence.org/files/AIFoomDebate.pdf

Superintelligence
# Potential Outcomes

<table>
<thead>
<tr>
<th>Fast Takeoff</th>
<th>Slow Takeoff</th>
</tr>
</thead>
<tbody>
<tr>
<td>hours, days, weeks</td>
<td>several months, years</td>
</tr>
<tr>
<td><strong>Unipolar Outcome</strong></td>
<td><strong>Multipolar Outcome</strong></td>
</tr>
<tr>
<td>Singleton</td>
<td>Second Transition</td>
</tr>
<tr>
<td>(tomorrow)</td>
<td>Unification by Treaty</td>
</tr>
</tbody>
</table>

[foundational-research.org/robots-ai-intelligence-explosion/](foundational-research.org/robots-ai-intelligence-explosion/)
Predicting Arrival

When will we have general artificial intelligence?
Exponential Improvements

The Road to Superintelligence
waitbutwhy.com/2015/01/[…]

Superintelligence
Predicting Arrival

Exponential Improvement Graph

- EINSTEIN
- DUMB HUMAN
- CHIMP
- BIRD
- ANT
- AI INTELLIGENCE

Haha that's adorable the funny robot can do monkey tricks!
The fuck??
Predicting AI Timelines

Great uncertainties:
- Hardware or software the bottleneck?
- Small team or a Manhattan Project?
- More speed bumps or accelerators?

<table>
<thead>
<tr>
<th>Probability for AGI</th>
<th>10%</th>
<th>50%</th>
<th>90%</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI scientists, median</td>
<td>2024</td>
<td>2050</td>
<td>2070</td>
</tr>
</tbody>
</table>

How We’re Predicting AI – or Failing To intelligence.org/files/PredictingAI.pdf
Speed Bumps

- Depletion of low-hanging fruit
- An end to Moore’s law
- Societal collapse
- Disinclination
Accelerators

- Faster hardware
- Better algorithms
- Massive datasets

+ enormous incentives!

Machine Intelligence Research Institute: When AI? intelligence.org/2013/05/15/when-will-ai-be-created/
Economic Incentives

- It’s difficult to enter the race later on
- Machines do more intellectual tasks
- Impossible for humans to compete

3 Breakthroughs That Have Unleashed AI on the World
www.wired.com/2014/10/future-of-artificial-intelligence/
Economic Consequences

– The living costs of digital workers are drastically lower (just energy)
– Thus enormous pressure on wages
– Massive unemployment ahead of us
– Wages approach zero, wealth infinity

Introduce unconditional basic income?
Military Incentives

better robots + more funding

better intelligence + better predictions

Arms Race?
Egoistic Incentives

- Intelligence
- Wellbeing
- Longevity

⇒ Willing to take risks

But with great power comes great responsibility!

PostHuman: An Introduction to Transhumanism
www.youtube.com/watch?v=bTMS9y8OVuY

Superintelligence
Predicting Arrival