ABSTRACT

Major technological changes have a transformative effect on every aspect of human life. Acceleration theories point to an abundance of evidence that technological paradigm shifts occur at a steadily accelerating rate and that future shock is indeed pandemic. We expect a descriptive, let alone normative, theory of technological evolution to answer questions about the impact of potential technology. Most substantially, which if any technology poses existential risk and how can such risk be mitigated?

We explore some of the technologies whose impact is imminent, (e.g. autonomous military robots; algorithmic trading; virtual markets) and possible (e.g. posthumans and ‘uploads’), and examine machine learning approaches claimed to be heading to artificial general intelligence. We conclude with the irreversible impact of a potential intelligence explosion and the notion that artificial intelligence constitutes an existential risk.
ACCELERATING CHANGE

The pace of change of our human-created technology is accelerating and its powers are expanding at an exponential pace (Kurzweil 2006)

We can read the history of modernity as a series of innovations [eg in transportation, communication] in ever increasing time compression

In the information age, not only are the rhythms of life faster, but the rate of change has itself accelerated.

Social theorists have responded in a number of ways to these processes of acceleration. Gane (2006) divides the main responses according to whether they call for [change] to speed-up or for it to slow down....

(Wajcman, "Life in the fast lane", British J. Sociology, 2008)

FUTURE SHOCK

* In the three short decades between now and the twenty-first century, millions of ordinary, psychologically normal people will face an abrupt collision with the future. Citizen’s of the world’s richest and most technologically advanced nations, will find it increasingly painful to keep up with the incessant demand for change that characterizes our time. For them, the future will have arrived too soon. ...

* For the acceleration of change ... is a concrete force that reaches deep into our personal lives, compels us to act out new roles, and confronts us with the danger of a new and powerfully upsetting psychological illness.

(Alvin Toffler, "Future Shock", 1970:12)

THE IMMINENT
**MACHINE LEARNING**

- Reinforcement learning (Arel 2013)
  - Learning by interacting with an environment
  - Learning is driven by rewards
- Theoretical models (Schmidhuber 2013)
  - HSEARCH (Hutter 2001): AIXI: as fast as the fastest algorithm
  - Gödel machines: self-improving, optimally efficient problem solvers
- Implementations
  - Recurring Neural Networks (Schmidhuber 2013)
  - Biologically-inspired Deep machine learning

**VIRTUAL ECONOMY**

- $3bn secondary market ('gold farming')
  - Virtual goods
  - $300,000 ‘games labourers’
  - Second Life, World of Warcraft, ...
- Digital labour (microwork)
  - Mechanical Turk, crowdsourcing
- Crypto-currency: Bitcoin

**CASES IN POINT**

- 3D printers
  - Printed organs for transplants, food, guns...
- Virtual markets
  - Video game industry: ~$80bn global market
  - Autonomous military robots
  - Algotrading

**VIRTUAL ECONOMY**

- ‘Cherry Blossom’ market
  - Facebook likes
  - Twitter followers

---


Anness Eden et al. (eds.), *Singularity Hypotheses, 2013.*

www.swtorstrategies.com
**BIG DOG**

**AUTONOMOUS MILITARY ROBOTS**

- Is the target a Soviet-made T-80 tank?
- **IDENTIFICATION CONFIRMED**
- Is the target located in an authorized fire zone?
- **LOCATION CONFIRMED**
- Are there any friendly units within a 200-meter radius?
- **NO FRIENDLIES DETECTED**
- Are there any civilians within a 200-meter radius?
- **NO CIVILIANS DETECTED**
- Weapons release authorized. No human command authority required

**“BERLIN STATEMENT”**

- Rapid development of armed tele-operated & autonomous robots
- Danger to peace and int’l security
  - accelerate pace & tempo of warfare
  - exacerbate dangers of asymmetric warfare
  - further indiscriminate & disproportionate use of force
  - obscure moral & legal responsibility for war crimes
  - undermines the capacity of human beings to make responsible decisions during military operations

[There is] an urgent need to bring into existence an arms control regime to regulate the development, acquisition, deployment, and use of armed tele-operated and autonomous robotic weapons

**ALGOTRADING**

- **AKA** electronic algorithmic trading, automated trading
- Timeline
  - 2006: 60% of orders LSE (Bloomberg)
  - trading floors ➔ Matching Engines
  - 2008 Flash Crash
  - Affected global financial market
  - 2013: 80% (est.) of US equities
  - LSE matching engines execute orders 0.32ms


International Committee for Robot Arms Control (2010), “Berlin Statement” [There is] an urgent need to bring into existence an arms control regime to regulate the development, acquisition, deployment, and use of armed tele-operated and autonomous robotic weapons
**Algo Trading**

**Share-Trading Programs**
- Shares: Morgan Stanley (MS), Goldman Sachs (GS), Wells Fargo (WFC), JPMorgan (JPM), Bank of America (BAC), ...
- Traders: Mutual funds, pension funds, insurance companies
- Revise prices in milliseconds \( \approx \) 1000 x faster

**Benefits**
- Decrease volatility on the foreign exchange market
- Improve liquidity
- Smooth fluctuations in liquidity

**Types of Algorithms**
- Statistical arbitrage
  - Exploit price fluctuations
- Execution algorithms
  - Break large orders into smaller slices
  - Execute in fixed intervals \( \rightarrow \) predictable
- Algo-sniffers
  - Detect & beat execution algorithms
- Spoofers (illegal)
  - Issue many 'buy' orders below price \( \rightarrow \) demand
  - Sell at a higher price, cancel 'buy' order

**Ultrafast Extreme Events**
- A new financial phenomenon
- Abrupt & extreme price changes
- Typical duration 15-20ms
- Price change > 0.8%
  - Price ticking down (up) ten times before ticking up (down)

**References**
**“FLASH CRASH”**

- 6 May 2008: DowJones drops 6% in 5min
- Overall prices of US shares & futures
- 600 points
- Almost unprecedented rapidity
- Maximum fluctuations: 1–2% a day
- Most price levels recovered < 20min
- Some individual shares:
  - Accenture $40.50 → $0.01
  - Sotheby’s $34 → $99,999.99

Unexplained in traditional terms

+ no ‘new news’ that could account for the huge sudden changes

---

**“FLASH CRASH” FORENSICS**

- Start: large sale $4.1bn
  - Precedence: twice previous year
  - Seller: Kansas City investment managers
  - 75,000 “Futures” × $55,000
  - “Index future contracts”: Track S&P 500 stock-market

- End: ‘Stop Logic Functionality’
  - Safety mechanism in the electronic framework used
  - 5sec trading pause
  - Automated systems stopped throughout US

- 2.41 pm “Hot-potato trading” spasm
  - 14 sec, 27,000 transactions, 9% drop
  - Affected: future-specific market-neutral algorithms
  - Ignore overall market
  - Conclude catastrophe in “Index future contracts”
  - Futures changed hands back and forth

---

**FORESIGHT REPORT**

- Technology and the economy change in the background

**ALGORITHMIC STRATEGIES LIFE CYCLE**

- Performance of Algos (downside)
- Regulatory Changes
- Inefficiencies (of Algos)
- Market changes
- Inefficiencies that allowed strategy to succeed
- Strategy spread

---

**NEW “MARKET ECOSYSTEM”**

- An emerging ecology of competitive machines:
  - Society’s techno-social systems are becoming ever faster… Generating a new behavioural regime as humans lose the ability to intervene in real time (Johnson et al. 2013)

- The universe of computer algorithms is best understood as a complex ecology of highly specialized, highly diverse, and strongly interacting agents (Doyne & Skouras 2012)
SOLUTION?

Conclusion: Regulation
We conclude that immediate regulatory initiative is necessary ...
(1) real-time warning signals for systemic risk;
(2) ... approach to systemic stability policy and to competition policy ...
(3)


DEVELOPING TRADING PROGRAMS

Conclusion: Regulation
We conclude that immediate regulatory initiative is necessary ...
(1) real-time warning signals for systemic risk;
(2) ... approach to systemic stability policy and to competition policy ...
(3)


THE (IM-)POSSIBLE

What is occurring now is, in all likelihood, bigger, deeper, and more important than the industrial revolution ... nothing less than the second great divide in human history, the shift from barbarism to civilization. (Toffler 1970)

INTELLIGENCE AMPLIFICATION (IA)

[The Singularity] will represent the culmination of the merger of our biological thinking and existence with our technology, resulting in a world that is still human but that transcends our biological roots (Kurzweil 2006)
TRANSHUMAN

- A transitional stage between H. Sapiens and Posthumans
- (-ism) The movement dedicated to promote this transition

WHOLE BRAIN EMULATION (UPLOADING)

- Machine intelligence emulating a specific brain
  - Copying: original intellect, memory, and personality
  - Presumably, implemented by a computer
- 'Uploads' can be "real"
  - Inhabit a virtual reality as an avatar, or
  - Inhabit the physical reality as a robotic body
- Uploads can be disruptive
  - The arrival of uploads will result in a sharp transition to an upload-dominated world
  - Uploads would reproduce quickly, and wages would fall
  - Total wealth should rise ...

SOME QUESTIONS

We could all do better by accepting uploads or at worse taxing them rather than trying to delay or segregate them

THE IRREVERSIBLE
Since man’s near-monopoly of all higher forms of intelligence has been one of the most basic facts of human existence throughout the past history of this planet, such developments [as human-level artificial intelligence] would clearly create a new economics, a new sociology, and a new history.

I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted. (Alan Turing, “Computing Machinery and Intelligence”, 1950)

Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any man however clever. Since the design of machines is one of these intellectual activities, an ultraintelligent machine could design even better machines; there would then unquestionably be an intelligence explosion, and the intelligence of man would be left far behind.

Complicated electronic circuits can also make computers act in an intelligent way. And if they are intelligent they can presumably design computers that have even greater complexity and intelligence.

Will computers ever be as smart as humans? Yes, but only briefly (Vernor Vinge, IEEE Spectrum 2008)
EXISTENTIAL RISKS

- Risks that threaten the entire future of humanity
  - Nuclear holocaust
  - Naturally occurring disease
  - Asteroid or comet impact
  - Runaway global warming
  - Genetically engineered biological agent
  - Badly programmed superintelligence
  - Accidental/deliberate misuse of nanotechnology
  - ...


EXISTENTIAL RISK FROM AI

- When we create the first superintelligent entity, we might make a mistake and give it goals that lead it to annihilate humankind, assuming its enormous intellectual advantage gives it the power to do so.

- A well-meaning team of programmers make a big mistake in designing its goal system. This could result, to return to the earlier example, in a superintelligence whose top goal is the manufacturing of paperclips, with the consequence that it starts transforming first all of earth and then increasing portions of space into paperclip manufacturing facilities.

Bostrom, Nick. "Existential Risks", 2002 and "The Ethics of Superintelligent Machines"

EXISTENTIAL RISK FROM AI

More subtly, it could result in a superintelligence realizing a state of affairs that we might now judge as desirable but which in fact turns out to be a false utopia, in which things essential to human flourishing have been irreversibly lost.

We need to be careful about what we wish for from a superintelligence, because we might get it.

Bostrom, Nick. "The Ethics of Superintelligent Machines"

EXISTENTIAL RISKS FROM AI

It seems obvious that major existential risks would be associated with an intelligence explosion, and that the prospect should hence be taken seriously even if it were known to have but a moderately small probably of materializing

The AI pioneers, however—notwithstanding their expectation of rapid breakthroughs and their confidence that human-level artificial intelligence would soon be achieved—did not consider the possibility of greater-than-human artificial intelligence; nor is there any indication that they ever countenanced that there might be risks involved in such an enterprise. They gave no lip service even—let alone serious thought—to the possibility that there might be significant ethical and safety issues involved in creating artificial general intelligence. This was reckless naivety. One must hope that by the time the project eventually becomes technically feasible, our thinking about its implications for humanity’s future will have grown more mature.

Bostrom, Nick. Intelligence Explosion In preparation, 2011
AI OR IA?

IMPLAUSIBILITY OF AI

The panel of experts was overall skeptical of the radical views expressed by futurists and science-fiction authors ... about the prospect of an intelligence explosion as well as of a "coming singularity," and also about the large-scale loss of control of intelligent systems.

("AAAI Presidential Panel on Long-Term AI Futures", 2009)

When the techies are crazier than the Luddites (Jaron Lanier 2011, "An apocalypse of self-abdication")

IMPLAUSIBILITY OF AI

It is hugely unlikely, though not impossible, that a conscious mind will ever be built out of software. (David Gelertner 2011, email)

(The) Singularity will never occur... I am a skeptic. I don't believe this kind of thing is likely to happen, at least for a long time. (Gordon Moore, IEEE Spectrum 2008)

Arthur C. Clark's third law:
When a distinguished but elderly scientist states that something is possible, he is almost certainly right. When he states that something is impossible, he is very probably wrong.

CHANGE DENIALISM

Thus man moves swiftly into an unexplored universe, into a totally new stage of eco-technological development, firmly convinced that "human nature is eternal" or that "stability will return." He stumbles into the most violent revolution in human history, muttering, in the words of one famous, though myopic sociologist, that "the processes of modernization ... have been more or less completed." He simply refuses to imagine the future.

(Toffler 1970:19)
NOTIONS OF A SINGULARITY

- Vernor Vinge (computer scientist, UCSD)
  - 1993: "The Coming Technological Singularity": A change comparable to the rise of human life on Earth
- Hans Moravec (Robotician, CMU)
  - 1998: Mind Children
- Ray Kurzweil (inventor, author)
  - 2006: The Singularity Is Near

Vernor Vinge (computer scientist, UCSD)
- 1993: "The Coming Technological Singularity": A change comparable to the rise of human life on Earth
- Hans Moravec (Robotician, CMU)
- Ray Kurzweil (inventor, author)
- 2006: The Singularity Is Near

NOTIONS OF A SINGULARITY

- 1st Wave: the Agricultural Revolution
  - took thousands of years
- 2nd Wave: the Industrial Revolution
  - took hundreds of years
- 3rd Wave: the Knowledge Revolution
  - It is likely that the Third Wave will sweep across history and complete itself in a few decades

TECHNOLOGICAL SINGULARITY

- Accelerating change
  - Computations/$, growth rate, paradigm shifts, ...
- Discontinuity
  - An essential singularity in the history of the race beyond which human affairs as we know them could not continue (Von Neumann, in Ulam 1958)
  - A change comparable to the rise of human life on Earth (Vinge 1996)
  - A rupture the fabric of human history (Kurzweil 2005)
- Superintelligence
  - IA (Intelligence Amplification)
  - AI (Artificial Intelligence)

AI OR IA?

- I'm sorry Dave, I'm afraid I can't do that
- ?
TO DO

One might think that the singularity would be of great interest to academic philosophers, cognitive scientists, and artificial intelligence researchers. …
The argument for a singularity is one that we should take seriously. And the questions surrounding the singularity are of enormous practical and philosophical concern.

(David Chalmers “The Singularity: A Philosophical Analysis,” 2010)

RESOURCES

BIBLIOGRAPHY

- Bostrom, Nick. Intelligence Explosion, 2011.