Governing AI

For **Chesapeake Large-Scale Analytics Conference**

2023-11-02 - Marc Donner - marc.donner@gmail.com - NYGeek LLC
(v2.1)

Copyright © 2023 NYGeek LLC

Marc's personal views on AI.

No intentional lies, but definitely personal and incomplete.

AI, real and imaginary, some highlights and lowlights

- 1950 - Alan Turing describes the eponymous 'Turing Test'
- 1956 - the term "Artificial Intelligence" is coined at a conference at Dartmouth College

Combinatorial attacks on games

- 1951 - Christopher Strachey's checkers program
- 1974 - "Chess as Problem Solving" - PhD dissertation by Hans Berliner
- 1979 - Berliner's Backgammon program defeats world Backgammon champion
- 1997 - Deep Blue (ne ChipTest) defeats Garry Kasparov, world chess champion, in six-game match

Neural networks

- 1943 - Perceptron invented by McCulloch and Pitts
- 1969 - "Perceptrons" book by Minsky and Papert
- 2012 - Image recognition breakthrough by U of Toronto team
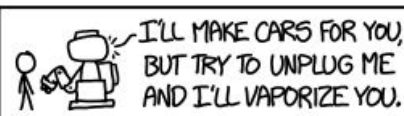
Robotics

- 1920 - "R.U.R." by Karel Čapek
- 1927 - "Metropolis" film by Fritz Lang
- 1942 - Isaac Asimov introduces the "three laws of robotics" in "Runaround"
- 1966 - "Shakey the Robot"
- 2004 - DARPA autonomous vehicle Grand Challenge - first competition
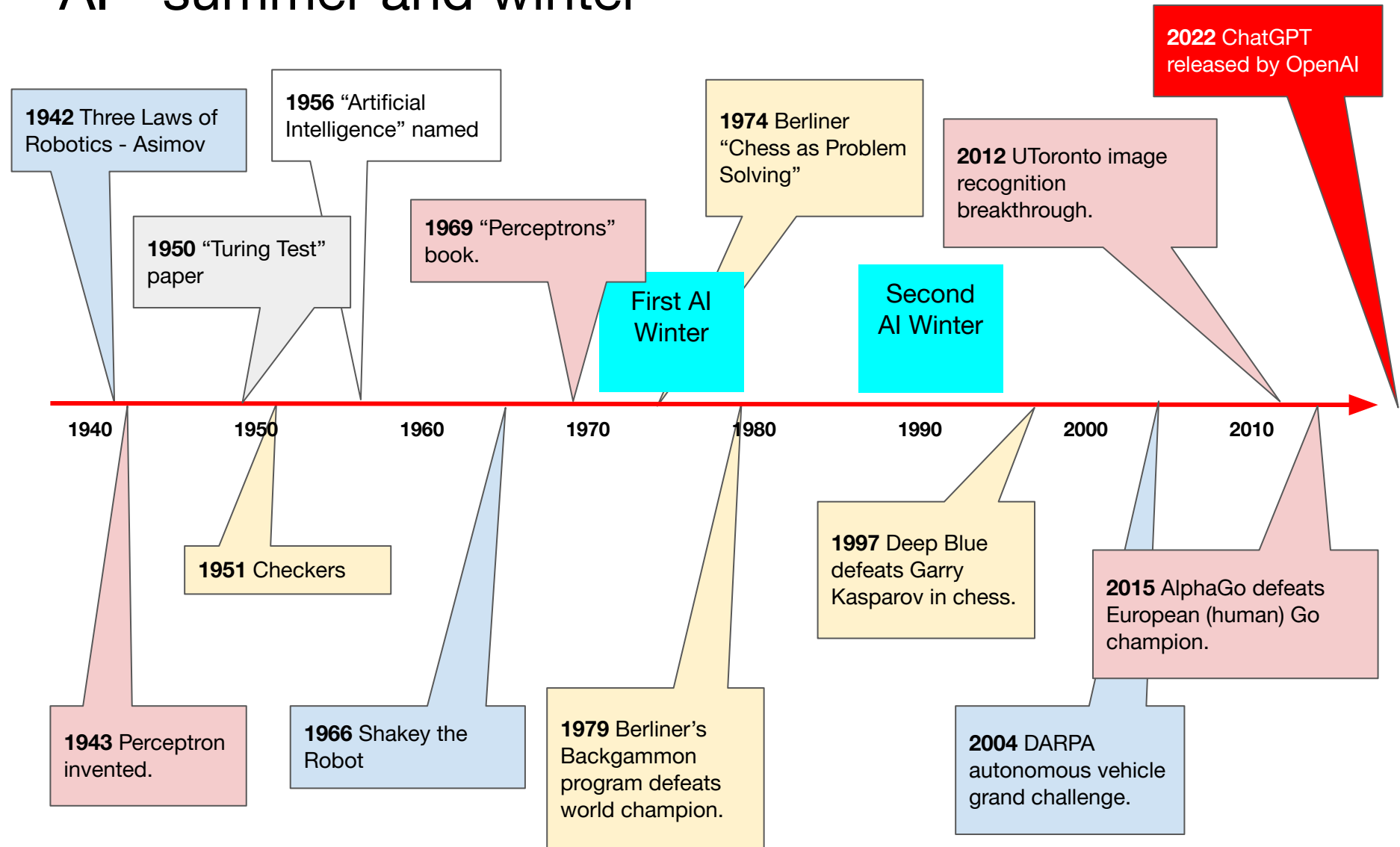
Governing AI before AI

Three Laws of Robotics (Isaac Asimov)

(XKCD.com/1613)

AI - summer and winter

1942 Three Laws of Robotics - Asimov

1950 "Turing Test" paper

1956 "Artificial Intelligence" named

1969 "Perceptrons" book.

1974 Berliner "Chess as Problem Solving"

2012 UToronto image recognition breakthrough.

2022 ChatGPT released by OpenAI

First AI Winter

Second AI Winter

1940   1950   1960   1970   1980   1990   2000   2010

1943 Perceptron invented.

1951 Checkers

1966 Shakey the Robot

1979 Berliner's Backgammon program defeats world champion.

1997 Deep Blue defeats Garry Kasparov in chess.

2004 DARPA autonomous vehicle grand challenge.

2015 AlphaGo defeats European (human) Go champion.

Gartner Hype Cycle? AI Winters!

VISIBILITY

Peak of Inflated Expectations

Plateau of Productivity

Slope of Enlightenment

Trough of Disillusionment

Technology Trigger

TIME

By Jeremykemp at English Wikipedia, CC BY-SA 3.0,
https://commons.wikimedia.org/w/index.php?curid=10547051

?
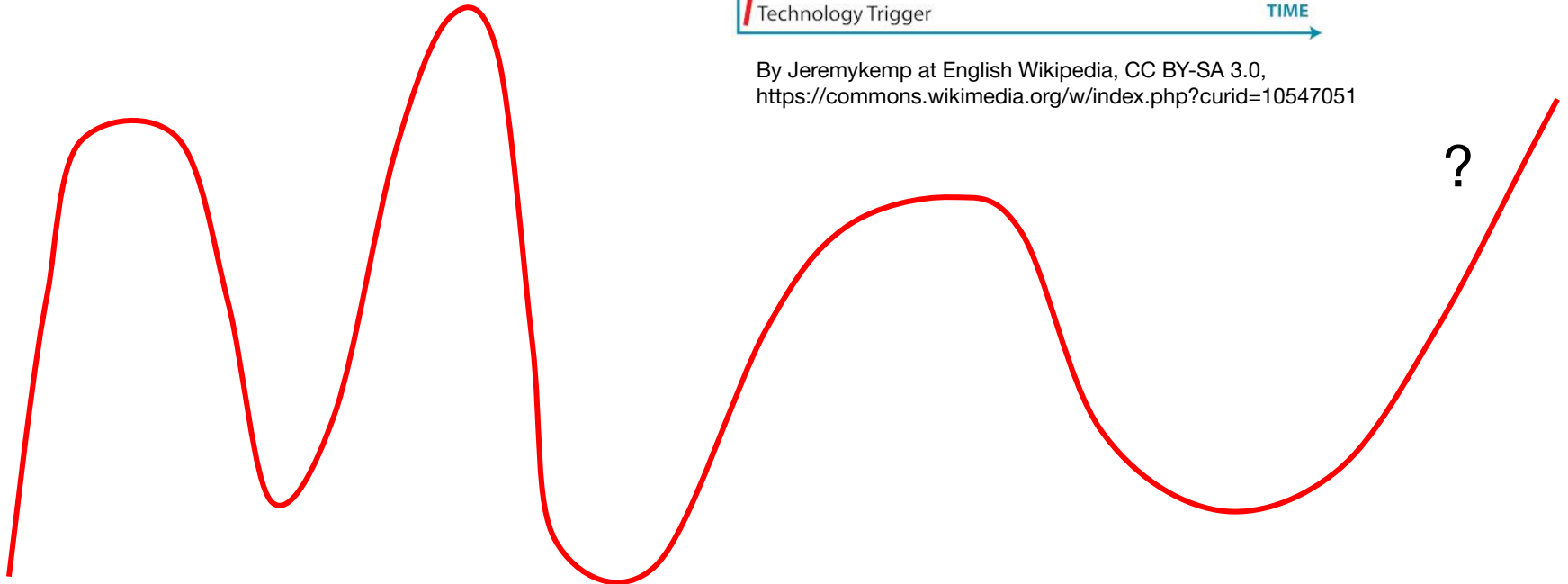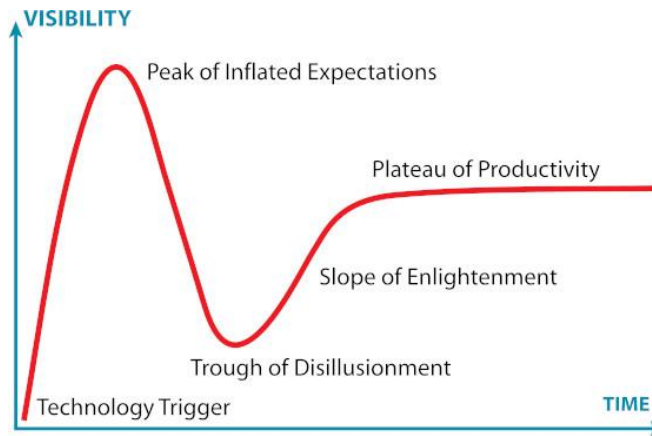
No longer AI ("Not Dead Yet")

AI has been characterized as those areas of computer science for which we do not understand the problems or the solutions well enough to do much science.

Dead ends (and things that just stopped being AI)
- Theorem proving
- Combinatorial attacks on game playing
- Expert systems
- Compilers (!!!)
- Text editors (!!!???)
- Speech recognition
  - Phonetic modeling (dead end)
  - Hidden Markov Models (a huge success, but that success moved speech out of the AI tent)
- Frames
- Knowledge Representation
  - CYC
  - (But the KR work contributed to improved thinking about databases.)

Policy and AI in the modern era

Philosophers (Ethicists) consider governance of AI

- The Trolly Car Problem

The questions of how to govern AI behavior have mostly been ignored by "serious" policy makers *(until very recently - see next slide)*.

1. It's too early and too speculative to waste effort thinking about
2. It's too hard to think about … let's leave it to the industry experts, who understand this stuff

Science Fiction writers consider post-singularity AI: (Verner Vinge 1983)

AI is: Benign, Powerful

- Robert Heinlein, "*The Moon is a Harsh Mistress*," 1966
- Iain M Banks, "*Consider Phlebas*," 1987

AI is: Malign, Threatening

- Stanley Kubrick, "2001, A Space Odyssey", 1968
- Thomas Ryan, "*The Adolescence of P-1*," 1977
- William Gibson, "*Burning Chrome*," 1987

AI is: Dangerous, Omnipotent

- Charlie Stross, "*Singularity Sky*," 2003

This just in …

EU draft AI Law

[1] Risk-based approach

[2] Bans manipulative practices

[3] Bans intent to distort human behavior

[4] Bans physical and psychological harm

[5] Bans discriminatory outcomes

[6] Most law enforcement uses prohibited, exceptions strictly defined and controlled

Writers Guild agreement

Ratified 2023-10-09

Executive Order on Safe, Secure & Trustworthy AI

https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/

China's push into AI

Three laws of AI in China?

1. Don't harm the Communist Party
2. Obey the Communist Party
3. Protect yourself

Policy and AI in the recent past

The **AI as Inventor** challenge

- 2019 patent applications by Stephen Thaler crediting DABUS, an AI system as inventor.
- 2019 USPTO initiates two RFCs:
  - Whether AI may be considered an inventor
  - Potential impact on other IP regimes (copyright, trademark, database protections, and trace secrets)
- Industry submissions (many) and a conference led to the decision:
  - Inventor must be a natural person
  - Author, composer must be a natural person

ChatGPT: "I'll see you and raise you …"

Industry leaders are saying, "Let us regulate ourselves."

Delegating to industry leads to policy failure

Digital Millennium Copyright Act (DMCA)

Fair Use was formally written into US law in the 1976 Copyright Act.

Digital Rights Management (DRM) machinery makes fair use difficult or impossible in practice.

Specifically, anti-circumvention provisions in the DMCA make bypassing DRM, even for fair use, a felony.

Communications Decency Act (CDA)

CDA exempts platform operators from any accountability for misbehavior by the users of the platform.

Social media platforms are usurping the historical roles of news media in providing information to much of the public.

But the absence of accountability on these platforms has resulted in a huge disadvantage for the truth when confronted by lies.

So, what to do if you're no longer an active AI researcher?

(Aside about me)
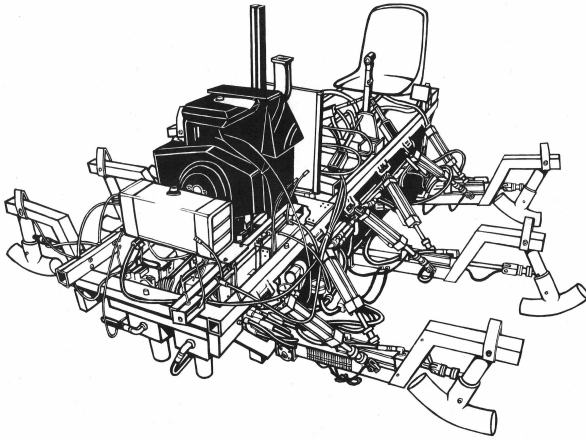
Marc earned a PhD in Artificial
Intelligence (Robotics) in 1984

Educate policy makers with a podcast.

Paid for and produced by the American Bar Association.

Four hosts - two national security lawyers and two computer scientists.

Website (www.mindthegapdialogs.com) (+ professional art)

Rigorous production process

- Mission statement
- Preview discussion with potential guest
- Develop introduction and 10-15 questions
- Preview questions with agreed guest
- Record
- Professional edit (+ professional music)

Guests and Topics (to date)

[1] Dr Stuart Feldman - *Taking the Mystery out of AI*

[2] Prof Avi Goldfarb - *AI as Prediction Tool*

[3] Dr Talia Gillis - *The Future of AI in Lending Decisions*

[4] Dr Rob Erdmann - *How AI Reconstructed a Rembrandt Masterpiece*

[5] Dr Becky Smethurst - *Can AI Uncover the Origins of the Biggest Black Holes*

[6] Judge Rama Elluru + Christian Hannon - *AI or Human: Which is the Inventor?*

[7] Dr Maria Klawe - *AI as Artist's Assistant*

[8] Dr Stanley Tuhrim - *"STAT" AI Helps Stroke Specialists Speed Response*

[9] Prof Edo Berger - *Too Many Stars! Too Many Galaxies! How to choose what to look at?*

[10] Dr Rohan Shad - *Above All, Do No Harm*

[11] Judge Katherine Forrest - *AI's Invisible Hand on the Scales of Justice*

[12] Dr Christina Colclough - *Nothing About Us Without Us*

[13] Lieutenant Colonel Jason Cody - *Game of Drones*

[14] Air Vice-Marshal John Blackburn - *Adding a Dog to the Cockpit*

Some lessons learned

[1] AI is being used widely in fields like medicine, art conservation, risk management, judicial support, astronomy and astrophysics.

[1a] Some fields are using AI responsibly,

[1b] Some are adopting AI without adequate professional review.

[2] AI is much easier to incorporate and apply than people realize.

[3] AI is in daily use in incredibly sensitive areas (courtrooms, hospitals, credit analysis). How confident are we that these AI systems are making fair and unbiased decisions?

[4] Risk management techniques that rely on well-studied populations leave others in the lurch. This is not an AI-specific issue, but big data exacerbates it.

[5] Professionals using AI tools may suffer skills atrophy.

[6] Professionals trained with AI tools will not learn techniques that previous generations depended on. Will this be an issue?

Appendix

Abstract & Bio

Popular understanding of Artificial Intelligence (AI) resembles the fable of the blind men and the elephant. This story recounts the disparate understandings that emerge from the limited experiences of each individual blind man encountering a different part of the elephant - tusk, side, leg, trunk, ear, and tail. AI has permeated much of our technology-augmented work, but like the blind sojourners, we each may have different ideas about what it is and, by implication, what it is not.

The emergence of this new technology raises questions concerning whether the pace of technological innovation has outpaced the law. For instance, US law currently only recognizes humans as creators. It also grants intellectual property (IP) rights only to persons, natural or corporate, regardless of their reliance on AI.

With AI we can get predictions that we can not explain but that we can not match or surpass. Some of these predictions are valuable, but some are flawed due to defects in the training data. Who or what we hold accountable for these flaws, and what incentives we do or do not create for their correction will influence AI's hand in how we work.

In this talk Marc Donner, one of the hosts of the "Mind the Gap: Dialogs on Artificial Intelligence" podcast, will discuss the development of the podcast and some of the lessons learned in the process.

Marc Donner is a native of New York City. He received a BS in Electrical Engineering from Caltech and a PhD in Computer Science (Robotics) from Carnegie Mellon University.

Dr Donner's research career includes time at NASA's Jet Propulsion Lab where he worked on planetary radar, at IBM Research where he worked on ultra-high-resolution displays, real-time systems, AI, robotics, and distributed computing. His industrial career includes work with Morgan Stanley, Union Bank of Switzerland, Google, MSCI, and Uber.

He currently spends some of his time advising small technology companies.

Dr Donner's professional interests cover a broad range of technical areas, including cybersecurity, privacy, software engineering, distributed computing, quantitative finance, AI, robotics, and large scale system management. He has been active in the Usenix Association, the IEEE, and the ACM. He contributes his time pro-bono to a number of charitable, non-profit, and public service organizations, often by serving as their webmaster. His blog may be found at nygeek.net. In addition, he co-hosts the popular ABA podcast "Mind the Gap: dialogs on artificial intelligence" with two national security lawyers and another computer scientist. The podcast website is mindthegapdialogs.com.

His personal interests include photography, bicycling, science fiction, and traveling with his wife.

Metadata

- Memo to: File
- From: **Marc Donner**
- Contact: **marc.donner@gmail.com**
- Date: **2023-09-26**
- For Presentation: **2023-11-02**
- Subject: **Governing AI**
- Copyright: **© Marc Donner 2023**
- Status: **DRAFT**
- DocId: **nygeek-159**
- DerivedFrom: **nygeek-142**
- Version: **V2.1**