Artificial Intelligence History

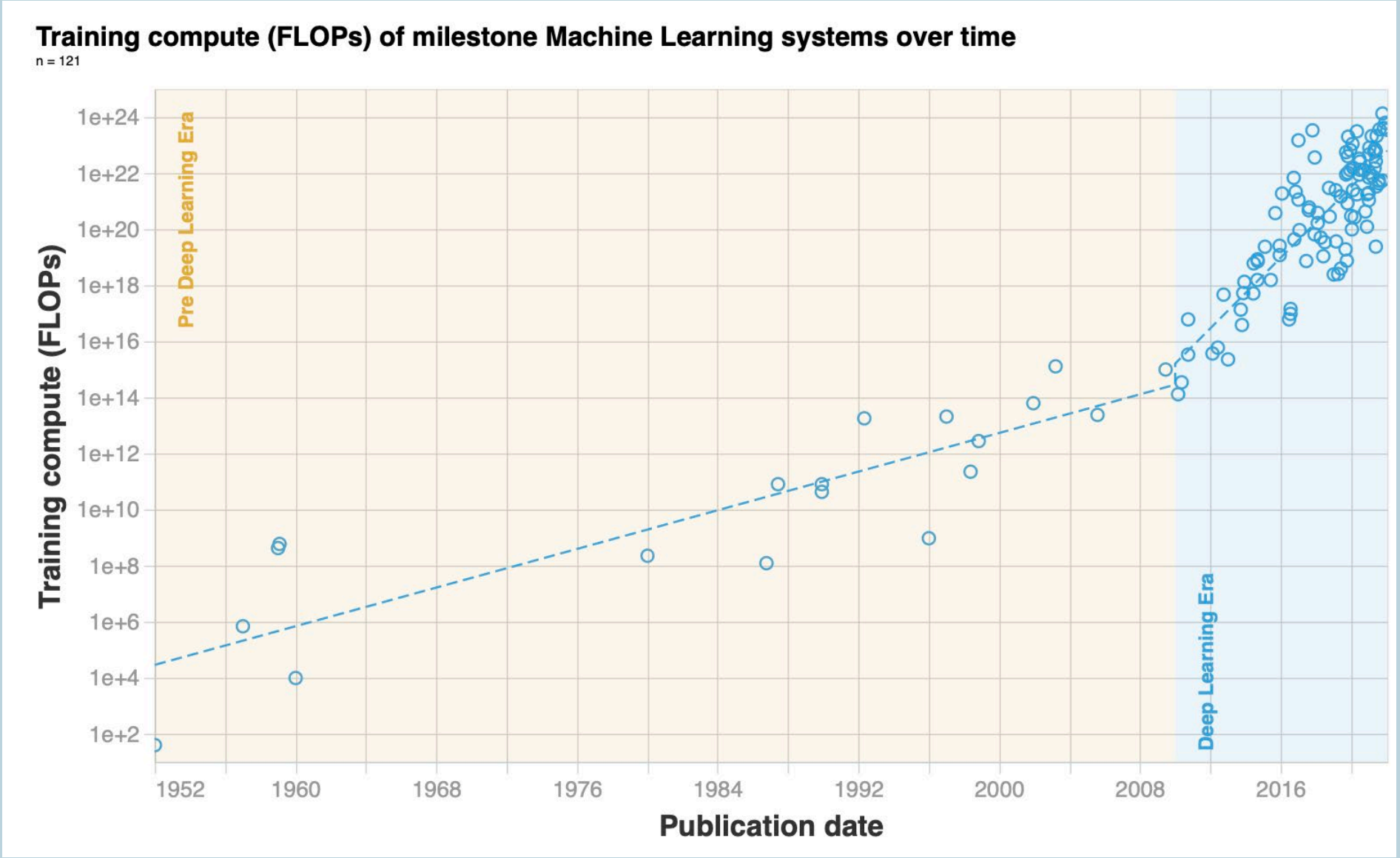
- Constant (over)optimism about progress
- Or, constant underestimation of the difficulty of the tasks
  - Minsky (1967) "Within a generation ... the problem of creating ‘artificial intelligence’ will substantially be solved.” (1970) "In from three to eight years we will have a machine with the general intelligence of an average human being."[
  - Simon and Newell (1958), “…within ten years a digital computer will be the world’s chess champion…” and “…within ten years a digital computer will discover and prove an important new mathematical theorem.” Simon (1965) “…machines will be capable, within twenty years, of doing any work a man can do.”
- This has led to a boom and bust cycle with AI
  - Lots of promise, lots of money
  - Overpromising, money dries up

Computing Revolution

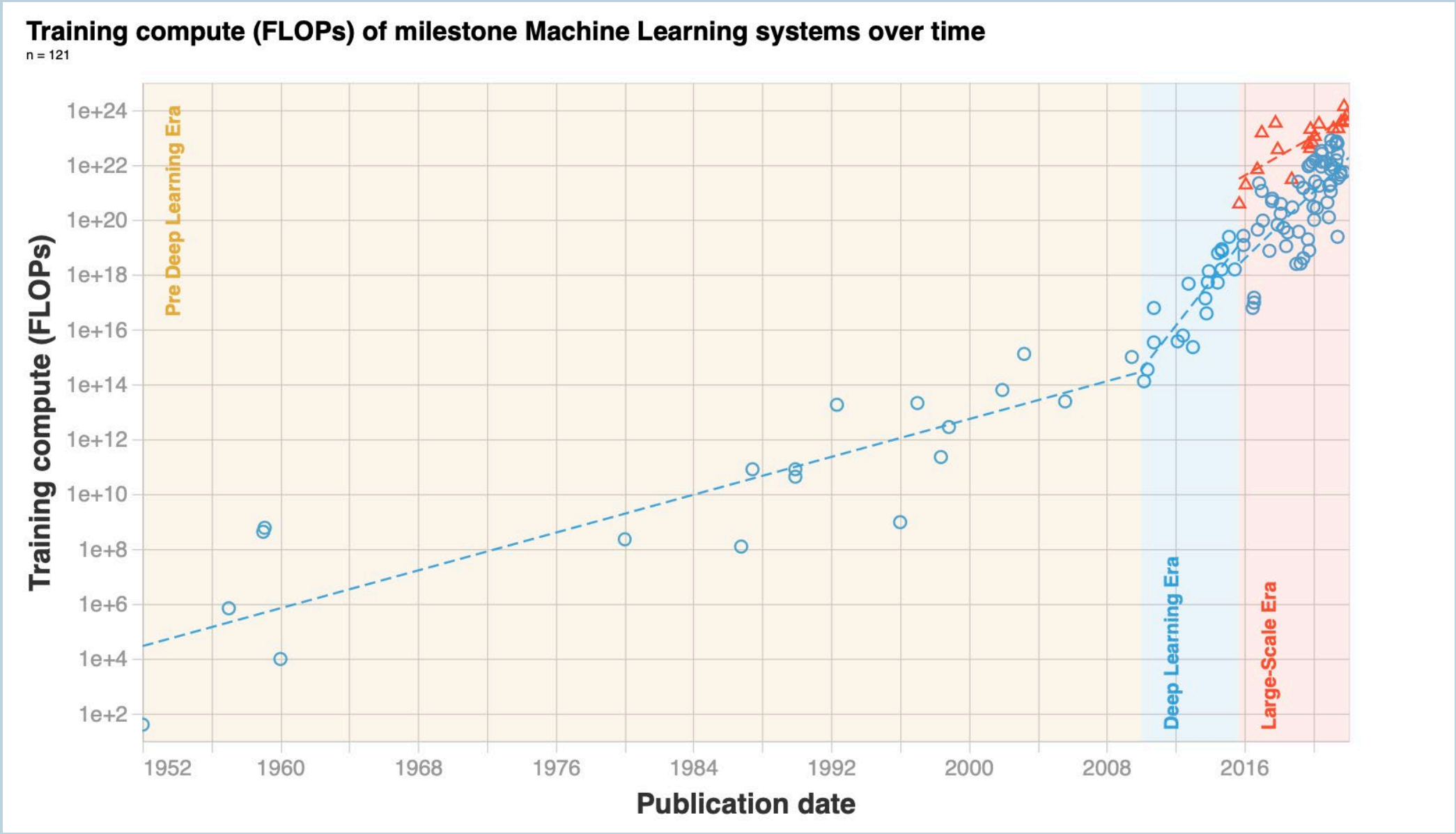
- Much of the history of CS is the realization of the modern computing world by visionaries and realized by engineers/programmers (often the same people)
- Some key advances
  - von Neumann architecture: 1945
  - Transistor: 1947
  - Integrated circuit – Moore's law: 1958/59 - now
  - Timesharing computers – multiprocessing: 1949 - 1964
  - Internet: 1969 - 1982
  - World wide web: 1995
  - Mouse/graphical interfaces: 1968 - 1983
  - Personal computing: 1973 - 1984
  - Distributed systems – cloud: 1960 - 2002
  - Cryptography: 1976 - now
  - Smartphones: 1973 - 2010

Progress in systems enabled the ML revolution

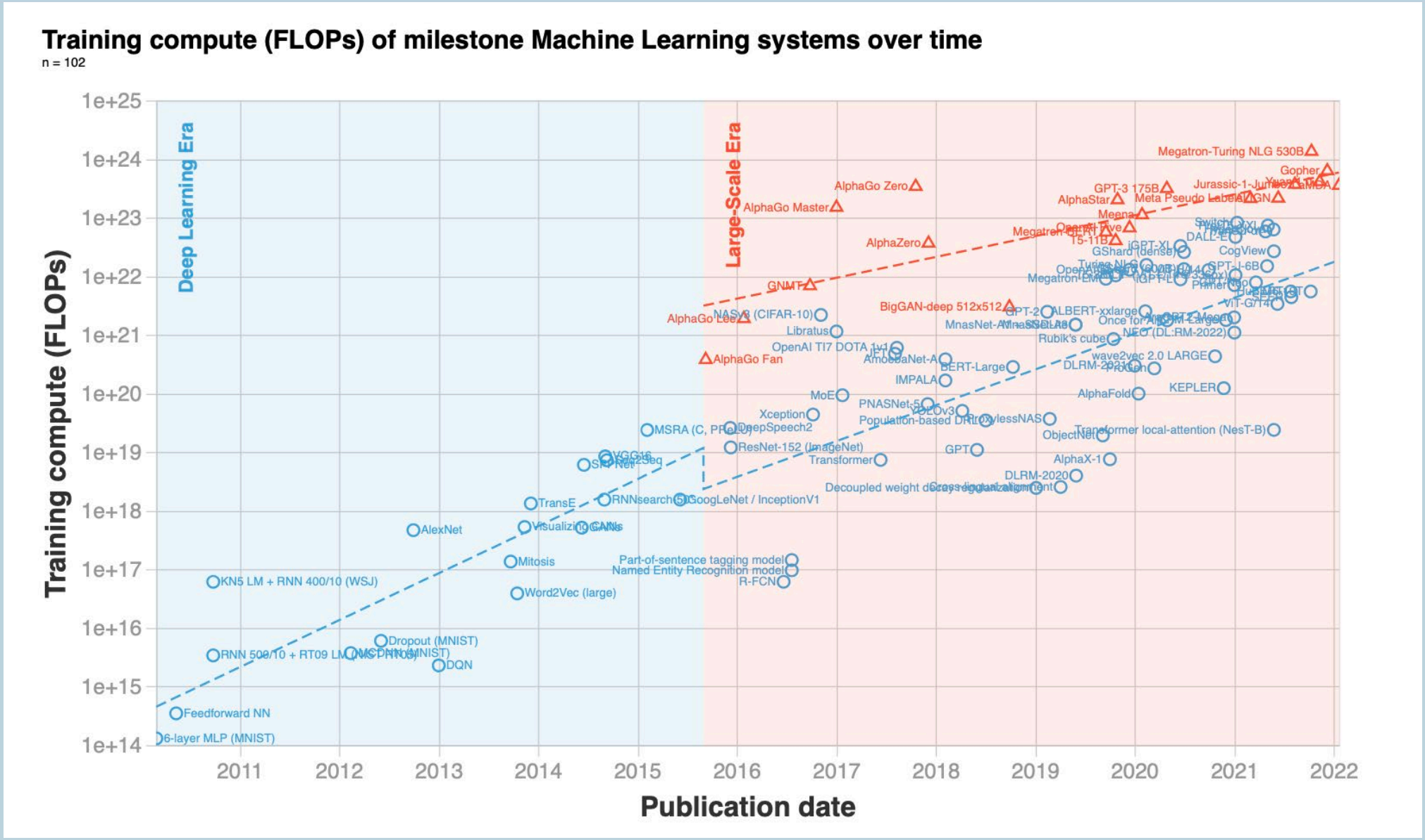
- Deep neural networks are scaled up versions of the neural networks studied since the 40's
- Training data volumes from the Internet have led to qualitative advances in the results produced using these techniques
  - Required massive storage capability
  - Required the Internet
  - Era of "big data" Large LLMs – 200B parameters, 50 TB of data, probably at least 500 TB of training data (GPT-4 100-200T parameters?).
- Neural networks of this size require massive computing resources to train (inferencing not quite as expensive)
  - Massively parallel GPU-based systems are essential
  - (GPU-Graphics Processing Unit)



From: Compute Trends Across Three Eras of Machine Learning https://arxiv.org/pdf/2202.05924.pdf



From: Compute Trends Across Three Eras of Machine Learning https://arxiv.org/pdf/2202.05924.pdf



From: Compute Trends Across Three Eras of Machine Learning https://arxiv.org/pdf/2202.05924.pdf

Period	Data	Scale (start to end)	Slope	Doubling time
1952 to 2010 Pre Deep Learning Trend	All models ($n = 19$)	3e+04 to 2e+14 FLOPs	0.2 OOMs/year [0.1; 0.2; 0.2]	21.3 months [17.0; 21.2; 29.3]
2010 to 2022 Deep Learning Trend	Regular-scale models ($n = 72$)	7e+14 to 2e+18 FLOPs	0.6 OOMs/year [0.4; 0.7; 0.9]	5.7 months [4.3; 5.6; 9.0]
September 2015 to 2022 Large-Scale Trend	Large-scale models ($n = 16$)	4e+21 to 8e+23 FLOPs	0.4 OOMs/year [0.2; 0.4; 0.5]	9.9 months [7.7; 10.1; 17.1]

From: Compute Trends Across Three Eras of Machine Learning https://arxiv.org/pdf/2202.05924.pdf

Computer Science: 3 communities view

- Theory – Establishing clearly (with proofs) what things computers can do and how hard it is to do them.
- Systems – Making software and hardware systems work efficiently and correctly.
- Artificial Intelligence – Making computers do things that humans can do.

Today

- The 3 communities seem to be overlapping much more than in the past.
- Theoretical underpinnings of AI/ML growing rapidly
- AI/ML approaches becoming state of the art in many systems areas
  - Not because they are “better” so much as because the algorithmically tractable parts are well understood and now we’re all tackling much harder problems
  - Systems advances and theoretical understanding help accelerate further advances in AI and vice versa

Tomorrow

- Will we understand what we are doing sufficiently well in time to control it?
- Will we continue to produce algorithmic and systems advances sufficient to support the continued astonishing progress in ML