Artificial Intelligence—The Revolution Hasn't Happened Yet

By Michael I Jordan

- Key claims:
    a. Academia and (especially) society is too focused on the question of AI with 'human like intelligence' and this is a distraction
    b. The idea that our era is seeing the emergence of an intelligence in silicon that rivals our own distracts us.
    c. AI Engineering does not exist in a systematic planetary scale right now and we need this kind of a discipline:

- The new engineering
    a. Aim is to bring advances of inference to applications safely
        - *"This new discipline aims to corral the power of a few key ideas, bringing new resources and capabilities to people, and to do so safely. "*
        - *"Whereas civil engineering and chemical engineering built upon physics and chemistry, this new engineering discipline will build on ideas that the preceding century gave substance to, such as information, algorithm, data, uncertainty, computing, inference, and optimization."*
    b. Believes that it must include cross disciplinary approaches with the social sciences
        - Moreover, since much of the focus of the new discipline will be on data from and about humans, its development will require perspectives from the social sciences and humanities.
    c. Believes that it is currently not well organised and there are many ad-hoc approaches being applied.
        - *"What we're missing is an engineering discipline with principles of analysis and design."*

- Problems with how we define AI
    a. Today AI actually refers to ML (process data and make predictions) and some data science (building scalable and robust ML systems)
    b. At founding AI had a more specific goal of human-like intelligence
    c. What makes up today's AI comes from low-level pattern recognition and movement control - finding patterns in data and on making well-founded predictions, tests of hypotheses, and decisions.
        - Backprop was from control theory
    d. MJ is not really impressed with the original AI field actually -
        - *Since the 1960s, much progress has been made, but it has arguably not come about from the pursuit of human-imitative AI.*
        - *Document retrieval, text classification, fraud detection, recommendation systems, personalised search, social network analysis, planning, diagnostics,*

> *and A/B testing have been major successes - one could call this AI and that is what has happened…. Such labelling may come as a surprise to optimization or statistics researchers, who find themselves suddenly called AI researchers*

- If AI is too broad, what is a better definition
    a. Human Imitative AI (HAI)- Doesn't actually define this much
    b. Intelligence Augmentation (IA) - e.g. search engine, sound and art generative aids
    c. Intelligent Infrastructure (II)- E.g. self driving cars

- Big Question: "Is working on classical human-imitative AI the best or only way to focus on these larger challenges?... No, we need to solve IA and II problems on their own merits, not as a mere corollary to a human-imitative AI agenda."
    a. HAI has made limited gains towards actually
    b. Success in HAI is not sufficient for the other 2, there are unique challenges
    c. Success in HAI is not necessary for the other 2:
        - "No historical precedent for this argument"
        - "If our goal was to build chemical factories, should we have first created an artificial chemist who would have then worked out how to build a chemical factory?"
    d. Humans are not good at many kinds of reasoning
    e. Human intelligence did not evolve to solve to the kinds of problems that II will have to solve (scale or uncertainty)
    f. The idea that HAI would achieve Human Intelligence, correct its flaws and scale is speculative (to put it nicely)

- So what about HAI
    a. Interesting, but current approach of *"demonstration of systems that mimic certain narrowly-defined human skills—with little in the way of emerging explanatory principle"* deflects from major problems reasoning, causality, long term goal formulation and seeking, representing uncertainty

- Current public dialogue is focused too much on HAI

"Let's broaden our scope, tone down the hype, and recognize the serious challenges ahead."

The Discipline of Machine Learning – Mitchell (2006)

What is the defining question of Machine Learning?

- **"How can we build computer systems that automatically improve with experience, and what are the fundamental laws that govern all learning processes?"**

Mitchell defines *learning* as improvement at some **task** along some **performance metric** on a task following a type of **experience**.

What is the relationship between ML and other fields?

- ML as a "natural outgrowth" of CS and Statistics
    - Manually programming computers vs. how to get computers to program themselves
    - Inferring conclusions from data vs. finding what architectures and algorithms can be used to handle such data and how to combine them into larger systems in a computationally efficient manner
- The study of human and animal learning (in Psychology, Neuroscience, etc.) also has a closely related defining question
    - Mitchell posits that the questions of how computers learn and how animals learn probably have highly intertwined answers
    - However, the influence from these fields is much weaker than the influence of CS and stats – mainly because we understand so little about Human Learning
    - Nevertheless Mitchell expects that the synergy between Human Learning and Machine Learning to grow
- Fields like biology, economics, etc. also have a core interest in the idea of self-adapting systems

What is the current state of machine learning (as of 2006)?

- "It is worth noting that as late as 1985 there were almost no commercial applications of machine learning"
- He lists speech recognition, computer vision, bio-surveillance, robot control, and the acceleration of empirical sciences (e.g. in gene expression, astronomy) as successful applications of ML

What is ML's niche within CS?

- Mitchell states it is best for applications that are a) too **complex** for people to design algorithms themselves and b) requires **customization** to its environment after it is fielded
- ML can reshape our view of CS by emphasizing "the design of self-monitoring systems that self-diagnose and self-repair, and on approaches that model their users, and…take advantage of the steady stream of data flowing through the program"
- Similarly, it can also reshape Statistics "by bringing a computational perspective to the fore, and raising issues such as **never-ending learning**"

What are the current and long-term research questions (as of 2006)? What progress have we made on these questions in the past 17 years?

- Can unlabeled data be helpful for supervised learning?
- How can we transfer what is learned for one task to improve learning in other related tasks?
- What is the relationship between different learning algorithms, and which should be used when?
- For learners that actively collect their own training data, what is the best strategy?
- To what degree can we both have data privacy and the benefits of data mining?
- Can we build never-ending learners?
- Can machine learning and algorithms help explain human learning?
- Can we design programming languages containing machine learning primitives?
- Will computer perception merge with machine learning?

What are the ethical questions raised by machine learning? How do we deal with this as researchers and as a society?

- Data privacy: ML for medicine, security and law enforcement, marketing
- Some other topics of interest in 2023: bias and toxicity, ML for morality (Delphi demo), writing assistants and plagiarism

Statistical Modeling: The Two Cultures by Leo Breiman, 2001 (Notes)

> In the setting of interest: there;s x (predictor variables), y (response variables), and nature (the black box).

> The Data Modeling Culture: starts with assuming a stochastic data model for the inside of the black box. Model validation: Yes–no, using goodness-of-fit tests and residual examination. Estimated culture population: 98% of all statisticians.

> The Algorithmic Modeling Culture
The analysis in this culture considers the inside of the box complex and unknown. Their approach is to find a function fx—an algorithm that operates on x to predict the responses y. Model validation. Measured by predictive accuracy. Estimated culture population. 2% of statisticians, many in other fields.

> When reading problems with data modeling, recollect similar problems with ML!

> Arguments against data modeling:
-Models are too simple, nature is often more complex :: still true
-Heuristics involved: variable selection. Quadratic terms, interactions terms, variable pruning :: dropout, finetuning, normalization, skip connections, momentum
-Trained on all data, no test data
-can data answer intended question :: unclear, the answer is still no, often
-Validation methods: goodness-of-fit can be manipulated by increasing parameter size, has low power, residual analysis does not scale to >3 dimensions :: OOD, generalization
-multiplicity of good models, no emphasis on cross-validation

> There is an old saying “If all a man has is a hammer, then every problem looks like a nail.”
> ML is the latest sledgehammer in town

> Advent of algorithmic modeling: Decision trees and neural nets in mid 1980s, theoretical bounds for SVMs ala Vapnik in 1998, boosting and bagging.., yayyyyyy ml

> Three important lessons from 5 years of algorithmic modeling:
-Rashomon: the multiplicity of good models; still concerning, bagging is a possible fix
-Occam’s dilemma: the conflict between simplicity and accuracy; easy accuracy over simplicity duh, no dilemma
-Bellman: dimensionality—curse or blessing. With the increase in compute, it's a blessing. Higher dimensional sets are more separable.

> A black box for a black box: “So we are facing two black boxes, where ours seems only slightly less inscrutable than nature’s”.