REPORT DOCUMENTATION PAGE	Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
03-01-2023	Final Report	1-Sep-2021 - 31-Aug-2022

4. TITLE AND SUBTITLE	
Final Report: AI prospects and Challenges Workshop	5a. CONTRACT NUMBER W911NF-21-1-0357
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER 611102
6. AUTHORS	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES	8. PERFORMING ORGANIZATION REPORT NUMBER
University of California - Los Angeles Office of Contract and Grant Administration 11000 Kinross Avenue, Suite 211 Los Angeles, CA 90095 -1406	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)	10. SPONSOR/MONITOR'S ACRONYM(S) ARO
U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211	11. SPONSOR/MONITOR'S REPORT NUMBER(S) 79057-MI-CF.1

12. DISTRIBUTION AVAILIBILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Stefano Soatto
UU	UU	UU	UU		19b. TELEPHONE NUMBER 310-825-4840

Standard Form 298 (Rev 8/98)
Prescribed by ANSI Std. Z39.18

RPPR Final Report
as of 04-Jan-2023

Agency Code: 21XD

Proposal Number: 79057MICF

Agreement Number: W911NF-21-1-0357

INVESTIGATOR(S):

Name: Stefano Soatto Ph.D.
Email: soatto@cs.ucla.edu
Phone Number: 3108254840
Principal: Y

Organization: **University of California - Los Angeles**
Address: Office of Contract and Grant Administration, Los Angeles, CA 900951406
Country: USA
DUNS Number: 092530369
EIN: 956006143

Report Date: 30-Nov-2022
Date Received: 03-Jan-2023

Final Report for Period Beginning 01-Sep-2021 and Ending 31-Aug-2022

Title: AI prospects and Challenges Workshop

Begin Performance Period: 01-Sep-2021
End Performance Period: 31-Aug-2022

Report Term: 0-Other

Submitted By: Sim-Lin Lau
Email: simlin@cs.ucla.edu
Phone: (310) 825-2303

Distribution Statement: 1-Approved for public release; distribution is unlimited.

STEM Degrees:

STEM Participants:

Major Goals: This workshop is to identify current challenges in the broad field of Artificial Intelligence (AI) as they pertains to applications of interest to ARO, drawing inspiration from related fields from Biology to Cognitive Science, and chart the path for promising directions for the coming decade.

Accomplishments: See attached report.

Training Opportunities: Nothing to Report

Results Dissemination: Nothing to Report

Honors and Awards: Nothing to Report

Protocol Activity Status:

Technology Transfer: No Technology Transfer involved.

PARTICIPANTS:

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Alex Wong
Person Months Worked: 1.00
Funding Support:
Project Contribution:
National Academy Member: N

Partners

,

I certify that the information in the report is complete and accurate:
Signature: Sim-Lin Lau
Signature Date: 1/3/23 10:21PM

AI Prospects and Challenges Symposium

Held on October 28, 2021

Under the aegis of the Army Research Office

Final Report

This report describes the outcomes of a workshop held in-person in the UCLA Campus on October 28, 2021, aiming to discuss prospects and challenges of AI for the coming decade. After a long pause due to the pandemic, the in-person format was designed to foster open and straightforward discussion, as opposed to a string of presentations held over videoconference. Participants were instructed to provide a number of "provocations," consisting of a few slides to set the context, followed by potentially controversial statements on the current challenges, aimed at fostering discussion.

Background

We are currently in a Cambrian phase of AI, with thousands of researchers globally exploring variants of a common paradigm: Big datasets, deep neural networks, empirical validation on benchmarks. Some of this work has had outsize impact in the academic community, industrial domain, and the popular narrative. There are high expectations among the general population, unlikely to be met with the current paradigm. There are many efforts to explore beyond the common paradigm, but still relatively few. These include means of learning with few data: It has been remarked that the relevant limit for learning is not when the volume of training data goes to infinity, but when it goes to zero. Yet current systems are trained with massive datasets. Something appear amiss when the literature describes evidence that humans store about 1.5MB in the process of language acquisition, whereas we train models with billions if not trillions of parameters. Since natural language is a human construct, its complexity is no larger than the complexity of the union of human brains, whose capacity we appear to be approaching, possible having surpassed, with the latest wave of massive models. Accordingly, many have focused on semi-supervised, un-supervised, self-supervised learning, meta-learning, few-shot learning, zero-shot learning. Gaining the ability to learn with few data requires the ability to establish relations among concepts, entities and representation previously learned – ontogenically or phylogenically – from large volumes of data.

While traditional AI has focused on deductive learning, starting from a system of truth and proceeding in the deduction through the tools of logic and probabilistic inference, there is no system of truth in measured data. The vast majority of "AI" work today, focused on deep learning, is inductive, where concepts, classes and entities are learned from data. The connection between inductive (learning) and deductive (reasoning) is as old as AI or older, also related to the so-called "signal-to-symbol barrier" problem. Yet modern deep networks, especially in the area of knowledge, have shown the ability to incorporate relational modeling, relation discovery and basic forms of reasoning. This is an area where developmental biology can provide inspiration.

In practice, human information processing is critical not just as a source of inspiration.

Interaction between humans and learning machines is expected to be part of all complex systems in years to come, from operating complex semi-autonomous machines to analyzing business metrics.

Relations among entities or concepts are widely varying. In particular causality, including causal relation discovery and causal analysis, are particularly relevant for DOD scenarios. This is an area that has been maturing for years, yet some of the foundations are still the subject of debate. In particular, causal analysis limited to an existing directed acyclic graph, with categorical nodes, is particularly limiting since a human designer must specify the graph and the variables.

Once a system is in operation, questions of security, privacy and operational safety become important, especially in DOD scenarios. There is a considerable amount of work, but still little in relative terms, on model security, robustness to various forms of perturbations, including adversarial perturbations, membership inference attacks, and model stealing through distillation. The advent of DeepFakes, spoofing, impersonation has become a challenge to authentication, security and other forms of trust. Cryptographic systems only guarantee consistency, not authenticity, and an encrypted system is as secure as the key written on a post-it note.

Beyond stateless decision machines, as current supervised classification systems, learning-based models are increasingly used on-line, in a closed-loop setting, where they influence the environment. Reinforcement learning, policy learning, stochastic optimal control, sequential decision making are all active areas of investigation.

Finally, ethics, fairness, bias, accountability have received a considerable amount of attention, largely driven by good intentions, but with some red flags raised by academic papers aimed at calling out manufactured problems rather than solving real ones. An example of such a work is described in the slides of one of the presentations.

Format and organization

The workshop was organized around a series of brainstorming sessions focusing on key questions provided ahead of time to the invitees. Each key question or theme had at least one champion and at least one critic, and the format was meant to trigger an open discussion, as summarized in this report.

Invitees

To ensure freedom to openly discuss sensitive topics, the workshop was conducted with invited participants and DOD representative, with no public present. The meeting was kicked off by Dr. **Hamid Krim**, and Dr. **Behzad Kamgar-Parsi**, Program Leads for ARO and ONR respectively, and participants who led the discussion included included **Pietro Perona**, Allen Puckett Professor of Electrical Engineering and Computation and Neural Systems at Caltech, who focused on the role of "small data," communities of experts, and the danger of drawing conclusions from observational studies, specifically in the area of bias and fairness. **Michael Kearns**, National Center Chair Professor, Computer and Information Science, University of Pennsylvania, spoke of privacy, bias and fairness, providing an alternate viewpoint. **Olga Russakovsky** Assistant Professor at Princeton University presented a view of the risks and dangers of current AI systems in the area. **Judea Pearl**, Professor of Computer Science Emeritus, Computer Science Department, UCLA, presented an impassionate defense of causal analysis and reasoning, echoed by **Dan Roth**, Eduardo Glandt Distinguished Professor, University of Pennsylvania, who focused on reasoning and the limitations of current large language models in this domain. **Stefano Soatto**, Professor of Computer Science at UCLA, focused on embodied cognition. **Angela Yu,** Professor of Neuroscience at UCSD, provided a view from the standpoint of biology, Dr. **Alicia Solow-Niederman,** Associate Professor at the University of Iowa and Berklman Klein Center for Internet and Society at Harvard, provided a view from the standpoint of the Law, Prof. **Sidd Srinivasa** provided a view from the standpoint of Robotics, and Prof. **Alex Smola** provided a view from the standpoint of large-scale system infrastructure for Machine Learning. Last but not least, Dr. **Alex Wong**, now Assistant Professor at Yale University, coordinated the discussion and Q&A, and **Stephanie Tsuei** tracked the discussion and contributed to capturing the contemporaneous discussion. A handful of invitees were unable to participate at the last minute due to either sickness or high-risk exposure while still in the midst of the Covid-19 pandemic.

Trigger Questions

The following list was provided to participants prior to the workshop, in order to guide them in their preparation:

- What are the (top-5) technical developments/breakthroughs that occurred in your field in the past 5-10 years that are likely to remain useful/relevant in the next 10-20 years?
- What are the (top-5) "best kept secrets" in your field, meaning discoveries/developments/ideas that are not widely known or talked about, yet likely to have broad influence in years to come?
- What are the (top-5) "ladders to the moon", meaning developments that are likely to take the field in the wrong direction?
- What are the most important questions for your field to address in the next 5 years? 10 years? 15 years?
- What keeps you up at night? What are the issues likely to cause most harm in the next 5-10 years, if not addressed by then?
- What are the areas where AI is expected to have the most beneficial impact to society in the next 10-50 years?

Outcomes

This section captures the contemporaneous discussion with minimal edits to ensure faithfulness. As such, it reads more like edited minutes than a narrative summary. This is intentional, to capture the dynamics of the live discussion. The summary is organized in order of presentation of the provocations, but the discussion often crossed the order of topics, so the theme of the

discussion is not necessarily confined to the theme proposed by each participant, which is listed as the title of the corresponding section as follows:
Hamid Krim -- The view from the command room
Pietro Perona -- What, who, how and why in AI
Dan Roth -- It's Time To Reason
Alicia Solow-Niederman -- AI and the Law: It's Complicated
Judea Pearl -- The tyranny of data-centricity: from deep learning to deep understanding
Angela Yu -- Bridging artificial & natural intelligence
Michael Kearns -- Interpretability research is broken
Olga Russakovsky -- A time for reckoning and reflection in computer vision
Alex Smola -- Learning machine learning
Sidd Srinivasa -- So, where's my robot butler?
Stefano Soatto -- Embodying AI: The challenge from Gibson

Dr. Hamid Krim, Dr. Behzad Kamgar-Parsi
The view from the command room

Drs. Krim and Kamgar-Parsi kicked off the discussion highlighting a certain contrast between the remarkable progress the field of AI has witnessed, and the resulting excitement in academia and in the popular narrative, while at the same time leaving key questions open that are foundational to the utilization of these tools in the context of defense. Questions about uncertainty quantification, prediction of performance, if not performance guarantees, and in general the seeming fragility of systems trained from data. The view was largely shared among all participants, and so was the interest in going beyond the excitement, to solid science that has staying power.

Pietro Perona
What, who, how and why in AI

Dr. Perona pointed to the schism between human learning, which requires few examples, and current methods, and suggested that communities of human experts can play a key role in the development of the field in the decade to come, to address the key "tail problem" where most concepts of interest have few if any training samples.

One of the provocation was that "letting the data speak" is an aphorism. It is the experts speaking through the data, but it is very inefficient. They have to go through data. We have to think of ways in which experts can share knowledge more efficiently. The open discussion offered a counter-challenge: visual knowledge is not just human. Cats and dogs also learn efficiently (maybe) without the language medium. Language is an efficient way for humans to communicate knowledge among each other, but why should it be an efficient way to communicate knowledge to machines? (Response: Cats and dogs are actually quite bad, it is hard to teach them) - but if they are exposed to a threat once, they learn it quickly. (Open for discussion: Are we trying to build intelligent machines where intelligence is a human peculiarity, or intelligence as a broader notion of information gathering and exploitation skills useful for survival (including of the species, that entails some sort to communication skills, both ontogenic — through acoustic, visual, chemical, tactile communication, verbal or not — and phylogenies, communicated through the DNA).

Communities of Knowledge: The machine has to be active in acquire knowledge and has to be able to teach people and acquire information from people. Humans are not oracles. Even if experts, they can disagree. How can machines become the center of communities. How can they dish it out to people. Machine is always on, active agent, always learning from what humans are doing. Industry is building tools for experts to train systems to help them out. Machines can enlarge the volume of the datasets that can be accessed. Experts + Data + Machines.

Next provocation: Correlation vs. Causation. Key aspect is correlation is good for prediction. Not good for action. To make something happen. Example is a straw-man because there are two different events: (A) Patient sick or not (does not change if you remove the doctors) vs. (B) viewer interprets the drawing as patient sick or not (changes if you remove the doctors). So presence of doctors in the drawing is causative for (b) but not for (a).

Next example: observational study (skin color affects face recognition (or perceived gender classification) performance). It is correlational because changing one attribute at a time changes errors. In the specific case of perceived gender classification as as function of different attributes, hair length is causative: Changing hair length at fixed perceived gender vs. Changing perceived gender at fixed hair length. The observational study (Bwolamwini-Gebru) says system does not work well for dark-skinned. Experimental/causative study refutes this claim as, controlling for hair length, there is no statistical difference across skin color. Need to dissect the data in observational studies. There is often insufficient coverage of the dissects. No males with long hair. Very few female with short hair, with exception of African parlamentarians, where chosen dataset has a specific bias: No parliamentarian men with long hair, but African professional women typically wear short haircuts. Dataset is fine except for short-hair women and long-haired men, so bias in the benchmark dataset induces apparent bias in the system.

Challenge: In vision there are millions of variables, how do we do dissects? Example of the weather. Billions/trillions micro-variables, but coarse description with macro-variables. Unexplored piece of the literature. E.g. temperature as macro-variable. Names of storms or meteorological phenomena (e.g. El Niño). Some patterns have names, viable macro-variables. Bias-Fairness considerations require thinking causally, not correlationally.

Q: Is macro-variable just compression/clustering.
A: there is a theorem where causation forces narrowing down set of causative factors. Sometimes you have distinct variables (X causes Y etc.) but in other cases there are dynamic.

Q: What to do if expert knowledge is inaccurate?
A: Experts will disagree, often. Model can look at past disagreements, errors are typically outliers. Experts in a community can be ranked based on the past errors they made, which gives weight to the current label they provide. This gives an estimate of the confidence of a label. Some challenges are more complicated (delineating a cancerous liver cell), but the idea remains: build several representations and give them some confidence score before aggregating.

Q: Are we doing machine learning wrong? Children can learn concepts (cats, dogs) from few examples without being given 'features' (ear shape, …).
A: Humans are very good at one shot learning. The world is a long tail distribution, there will always be categories with few examples. The missing link is intermediate representations. Once we have models that have learned enough different things, we will be able to assemble these representations to learn with fewer examples.
Are we using the wrong models, if we need so many parameters to train a model (and therefore giant data) whereby few-shot learning seems to be incompatible with this kind of models? Are we using the wrong models?

Doubling down on this challenge: Humans use about 1.5MB in acquiring language (https://royalsocietypublishing.org/doi/10.1098/rsos.181393#d3e580) yet we train language models with 1.3Trillion parameters. Is this a symptom that we are using the wrong models? It is not like vision where the scene has infinite capacity/complexity. Language has finite complexity which is defined by humans. So even if the 1.5MB was off by orders of magnitude, there is still something off.

Dan Roth
It's Time to Reason

Provocation: Mixed visual and textual input "n [ice] 2 [meat] u" -> Nice to meet you. Deceptively complex task. This is natural language understanding. Decision are global, but need to make a lot of local decisions driven by different goals, different skills, different context, etc. Requires "knowledge" [define knowledge?]

Natural language meaning: meaning of constituent + rule to assemble. Reasoning: Will we make it to dinner before the movie? Easy task: when is the movie? Hard task: how long is dinner? Hard to train for. Language is a symbolic system, not an invention of AI, symbols are the invariants of communication. Creating symbols is the key to communication and cognition.

Aristotle founded syllogisms: information can prove a point. In NLP today, we have well defined tasks, for which we collect data and train, which is asking: which variables do we want to reason about? Did Aristotle have a laptop: you're answering when did Aristotle die? When was the laptop invented? What happened first? How do we determine this strategy on the fly? How do we train for it?

Aristotle: all beliefs come from syllogism or induction -> we may need to extend that.
There is a lot more than induction to cognitive behavior, we have mathematical theories for generalization. -> still not enough
Reasoning is another major component of cognition, the challenge is to create a unified formulation of learning and reasoning. How do we acquire and incorporate knowledge? Put components together? Deal with inconsistencies?

<u>Knowledge is key</u>
What day is the longest day in New York state? Boston? Melbourne? We have a model of hemisphere so it's easy. How about Yokne'am? -> we know what to look for, we can look where it is and incorporate it in our model. There is no generalization theory here, the examples aren't 'similar'. We don't know how to choose which knowledge is relevant at what time.
Knowledge is used to decompose questions, telling us what we need to learn. Aristotle and a laptop? -> it's about <u>time.</u>
Did warren buffet vote for FDR? Warren was born, but too young to vote for FDR -> time + voting laws. We instantly know what to look for. Implicit decomposition of the question is trivial for humans.
Humans often reply with indirect answers.
Have you eaten? A: I am really hungry -> not an answer, but we understand that they would like to eat. Collecting Q/A to train a model like this will not work, generalization requires another component.

Learning common sense knowledge:
Will we make it to dinner before the movie? Learning time
Assume we can learn distributions of dinner, holiday ,....
How do we assemble all of this together?
First problem: how to temporally order predicates?
Second: How do we exploit declarative and statistical expectations?

<u>Supervision is key</u>

1) To understand the text, we have to solve multiple challenges, understanding the text means identify events, entities, quantities, relations, …
2) We have to understand the question: decompose it in relation to the text

3) Reason about text: combine and manipulate information to answer

We can't provide supervision for everything. What do we provide supervision for?
The model is structure $y = f(g(x), h(x))$ teaching g and h is not tractable.

When and where? Question given a single image. Clip is not the solution -> need reasoning.
Theories of learning are too limited for the sparse annotations we have about the world

Q: Is the aspect of logic fundamental to your reasoning, or can it be substituted with Bayesian reasoning?
A: Bayesian reasoning is sufficient, but you need to bring in the required knowledge for Bayesian inference from somewhere else.
Q: If somebody told you, we had each location in the world as 100 different attributes and we need to classify if the answer is going to be June 21, vs 22. It's a binary classification with 100 variables, only one of which is relevant. Are you trying to say that our current solution is data-inefficient?
A: You invented a solution after you knew the solution. I'll give an example where there isn't a knowledge gap. So we need to sometimes acquire the knowledge and sometimes we need to use the knowledge in an efficient way and know what knowledge is available and when. That is a learning problem.
Q: (Understanding Indirect Answers) How about a simple yes or no?
A: People don't answer directly, therefore we need to understand indirect answers that point to yes or no.
Q: I think this notion of trying to discover how to parse a query so that you figure out the right kind of databases to use… is this a solvable problem? The first point that I want to make is that Joshi was able to translate FDA regulation on blood samples into some kind of temporal logic and these guys were able to do a lot of post processing with it, blood monitors. To me, that suggests that there is a domain [...] that you can use to answer those queries. Second point is that in the early 80s, you have to load a bunch of libraries. These days, we have a lot more memory, but a lot more libraries to learn.
A: Joshi's work focuses mostly on syntax and a little less on semantics. I agree that once you restrict the domain, these questions are a lot easier to answer. There was an AI in the 60s that played blocks world. It was a conversational system where you could say some things to build some shapes to solve a problem. (like Minecraft) If you tried to solve those problems in an open domain, you'd still fail today because their vocabularies were too tiny. The reason we still can't do it today is because of sparsity. The data we see is supposed to represent everything - if we see a discrepancy, we retrain. I was arguing that we need to more than taking the data and sending it to the box. We have to figure out which knowledge we can incorporate and how to compose these box and learn piecewise and with that, we're going to make some progress int he open domain situation.

Q: Do you see language as a prerequisite for reasoning?
A: I think no. We use language as a way to reason, but it's not necessarily the only way.
Q: (reword) Is language necessary for reasoning, and is reasoning necessary for success?
A: (video where pet cat keeps a little kid from climbing off a balcony and falling. Clearly the cat knows some sort of causality.)
A: The reason we don't know that other systems/species reason is because we don't know how they express it.
Q: I'm trying to figure out if the point of your talk is that we're doing it all wrong, or that we need… does it require interaction?
A: With respect to 2, I think there isn't a single dataset that will do it, but 1 dataset can do a slice of it. Reasoned Q&A. This is a step because the composition there is implicit. This by itself requires a lot of knowledge that isn't written explicitly anywhere.
Q: But sketch out a grand challenge.
A: I'm not sure that the only way to make progress. But let's take a dataset and hide 20% of it. And I want you to do well on that 20%
Q: What is in the dataset?
A: The dataset is a collection of examples of the type (did Aristotle have a laptop? Did Bill Nye vote for FDR?) and I want you to give an explanation.
Q: How do you measure the accuracy or score?
A: first: is it correct. 2nd: quality of decomposition. Need a scheme to evaluate the quality of the decomposition relative to some knowledge source. I think we can automate this.
Q: How do we generate those 10000 questions?
A: We cannot generate them automatically. If you generate them automatically using a program, then it is possible to write a learner to learn it and compute it. You need to solicit the information from people. For example, with a select set of Amazon Turkers (about 30), we were able to generate a few thousand questions.
Q: What instructions did you give to the Amazon Turkers?
A: It's tricky. You have to make sure that they don't give you something that is too easy or too complicated. This is just one challenge, though. What can we do now? If there is a piece of text in the question that directly leads to the answer, we can do it. (spam detection). We are really bad when there is no answer. Our systems today are really good at semantic value. They know which places variables (e.g. a person) should be. But if there are multiple people to choose from, then the answer is not there. We don't know how to train for this.
Q: Reasoning is what's lacking right now in AI. My perspective is then perhaps, isn't it crucial, to develop a formal point of view (like harmonics) of semantics. So the formalism can expedite the implementation point of view.
A: We once knew (if you look at AI in the 90s or 2000s), we knew what reasoning was deduction and reduction and that was reasoning. We now think that that was insufficient. We are missing the ability to deal with detail, sparse cases, things that we haven't seen before. We have to make some leaps.

Q: But you (human) seem to parse it into nuggets that you sort of assemble together.
A: That's the first step of this research. Reasoning cannot be decoupled from learning. I have to make decisions on top of components that themselves are learned things. That level of abstraction is undisputed. How exactly am I going to do it? There are probably 70 different ways. But learning is necessary because language is very imprecise and bad.
Q: So these nuggets that you're talking about. Humans have the ability to make up new names, new terms. So how do you expect to do such a thing?
A: I think naming is crucial. And that's why communication with people is so helpful. Because you identify something, it could be complex, you name it, and then it becomes a building block and simplest the computation from that point. I think we should be able to do that too, automatically. I don't know that there is a good solution today.

Q: "Symbols" = identifiers? Identities exist regardless of language. An animal learns a predator as an entity, can detect and recognize it despite intra-class variation (color, size) and nuisance variation (pose, illumination, partial occlusion). The class "predator X" does not need to have an explicit verbal or textual entity, it is just a class?

Q: Do cats or dogs reason? This suggests so: https://www.youtube.com/watch?v=7jB31L7K4qE

"There is no way we are going to learn [to reason, or these concepts] with end-point supervision"
The model needs to be structured [who structures it? Designers? Why not learn structure"?]
People think that neural networks are going to do it [isn't biology existence proof]

Alicia Solow-Niederman
AI and the Law: It's Complicated

Both law and AI are "not clear". It's not clear what we mean when we say "the law" or "AI". This comes out in concrete ways. NIST has proposal on fairness, for example. How should those standards come to be? How should they be codified in law, and what kind of AI should they apply to? Soft vs. Hard: What kind of law? Jurisdiction, coverage (e.g. civilian vs. military), how final (appeals allowed?), how should the law adapt? Who should decide what the law is? Whose voice matters? And how often should it change? Do we want AI-specific law? (i.e. for one specific tool).

Q: "AI -specific law" legislation requires definition. Can we even agree on how to define "AI"? Would some of the existing laws cover some of the most treacherous phenomena (e.g. deep fakes, impersonation, exfiltration etc.)?
A: The speaker prefers legislating for the effects of technology and AI rather than the details of the technology itself. This brings up the problem of forecasting the effect of technology. We

need to think about what information we need to collect from companies working on AI in order to make these forecasts.

Q: How big are the gaps (quick facts) between lawyers and technologists?

Q: Stare decisis is super slow, but AI is fast. How do we handle that?
A: Alludes to what kind of law we want. Stare decisis acts when a judge makes a decision. A regulatory agency is not similarly bound. In theory, it should be able to be more nimble, and why administrative law exists. Of course, there are questions of whether our existing agencies are able to act fast enough. Also, sunset clauses in law.

Q: Legislation uses follows technology. What to do?
A: Brings up ex ante vs. ex post law.

Judea Pearl
The tyranny of data-centricity: from deep learning to deep understanding

Deep understanding = "science of cause of effect" (SCE). Next two missions of SCE = automated scientist and social intelligence. Deep understanding will help you feel that you are in control. Control consists of 5 stages:

- 1 Predict the future
- 2 Predict consequences
- 3 Provide explanations of unanticipated events
- 4 Imagine alternative worlds
- 5 Design experiments to investigate new conjectures

Causal questions (relational statements): Prevent (treatment -> disease); Cause (tax break -> sales); Attribute (health care cost of obesity); Discriminate (hiring record -> prove discrimination/attribution); Regret (quit my job -> regret?) intention.

Equality is a symmetric relation. Cause/effect is not. Asymmetry not fit for language of algebraic equations. Need new logic, inference engine to capture relationships/assignment: Nature looks at atmospheric pressure and assigns a position on the barometer scale, not the other way around)

Now can formalize counterfactual queries. Two fundamental laws of causal inference are:

1. Law of counterfactuals
2. Law of conditional independence (how to interpret relationship of variables in causal models)

Level hierarchy of causal inference (double helix of causal reasoning) is:

- Counterfactuals (cannot answer with randomized controlled trials)
- Intervention (if X, then Y; need randomized controlled trials, cannot answer with just association)
- Association (no causation)

There exists a calculus/inference engine of causality that incorporates available data.

There exists two subcultures:

1. Data centric (data fitting) - knowledge is in the data, goal is to extract it.
2. Science centric (data interpretation) - knowledge is in the process that generates the data. Goal is to find needed features of the process using data and the known properties of that process

ML is data fitting. Seven capabilities of causal science (only in science centric subculture):

1. Encode causal assumptions
2. Predict effects of actions and policies
3. Computing counterfactuals and finding causes of effects (attribution, explanation, susceptibility)
4. Computing direct and indirect effects (mediation; discrimination, inequities, fairness)
5. Integrating data from diverse sources
6. Recovering from missing data (is a causal question, not just a statistical question)
7. Causal discovery (have to know what you're looking for)

Always eventually have to answer counterfactual questions - therefore need to study SCE, don't just manipulate data to get neural networks to do better. Next two missions of science of cause and effects

1. Automated Scientist
    1) Experiment design (aka active learning)
    2) Curiosity
    3) Vetting scientific theory (philosophy)

2. Social Intelligence
    1) Generalize from understanding a domain to understanding other agents
    2) Human-robot communication

Challenges:

- Formalize definition of responsibility. Presents Stanford encyclopedia definition of "moral responsibility", which uses the language of causality

Q: Could human erratic behavior be explained by the absence of lack of SCE?
A: SCE contains several components. The first one is a model of reality. Second is a maxim. The

model of reality is always faulty. Of course, this can account for human fallacy. On the other hand, humans do form consensus in certain questions about counterfactuals. For instance, had Oswald not killed Kennedy, would Kennedy still be dead today? How do we capture and algorithmize this consensus?
Q: MNIST helped focused efforts on deep learning in the 90s and ImageNet in 2010 and helped us measure progress. If you wanted to put out a challenge in progress for SCE for the AI community, what would it be? Especially something for the next 10-15 years. No cheating/correlation possible.
A: Success stories in personalized medicine. Truly personalized medicine. Actually going from input on population to properties of individuals.
Q: What dataset/experimental setup would you put up there so we can concretely see progress and compare methods?
A: The number of people before/after the exercise is administered.
Q: So maybe mice and not people…
A: No - consider selecting patients. Can we improve results for individuals, not just seen on a population level.
Q: Are you not too harsh on ML? Isn't a goal of ML to build causal models, or models in general, and what should we do in a domain where we don't have any relevant knowledge?
A: See slide: What if I don't have a model? It's important to know if you don't have a model, or just have a sketchy model, that we must use SCE to improve this knowledge. Exploit the calculus of causality to see what we can or can't do with the amount of knowledge available. Use SCE to discover a model, either from data or data augmented using experiments.

Q: Is SCE exclusively a human faculty? Are there any papers on causal reasoning in mice?
A: Mice do exhibit reward-neutral curiosity. That means that mice have this restlessness of not reaching a state of understanding. That restlessness we still want to equip machines with. So it's interesting to what degree we have this restlessness. My conjecture is that we (humans) are born with a template, and if the template is unfilled, we feel restless. This restless is curiosity and a lack of control. But, we can simulate it on a machine so that the machine is restless and will reach a level of deep understanding.
Q: What about tools used in birds?
A: I don't know. I need to analyze the data more.
Q: Once you see a behavior that is learned in an animal or a human, is there a test to tell whether a causal reasoning was necessary in achieving the behavior or not. My suspicion is that humans are also not good at causal reasoning.
A: I also pointed out that we reach consensus on counterfactual questions. But we speak the language of counterfactuals while mice don't. I don't know how to test whether a species can do counterfactual reasoning.
Human vs. dog-level AI?

Even in “Extract VQA” (basically a detection problem), we are bad b/c we don’t know how to trust the answer. We can “answer” the question if it is written somewhere (then it is just a detection/find the span), but we don’t know how to evaluate the correctness of the answer.

Angela Yu
Bridging Artificial & Natural Intelligence

SOTA: humans must define tasks, data, features, objective, actions, social communication for robot. Intelligence lies in something other than the computation and implementation of algorithms. Humans do all the thinking. Introduces computational model of intelligence and psychology of intelligence. Very different definitions!

- Fluid intelligence: reasoning, comprehension
- Crystallized intelligence: knowledge from prior learning

Provocation: AI needs more fluid intelligence (comprehension, reasoning and problem solving)
Society values fluid intelligence a lot more. Having to practice 10,000 times is cheating.
Traditional view is that perception is passive, but this doesn’t fit with squinting, echolocation, moving ears in cats - active perception
Actions create knowledge: Helmholtz (1878) proposes we don’t need innate knowledge about world to learn about it**.** E.g. statistical regularities in how scene changes when turning head**.** Experiments with cats from 1963 that demonstrated this**:** The passive cat had no intuition of 3D space without having the tactile experience the active cat had**.** Actions → development of cognitive knowledge (ex: temporal/spatial relationships, hierarchies, etc)**.** Actions are important for the sake of learning**:** In young children, they *want* to explore to understand: their own space/possible actions; causal structural relationships**;** Understand the extent of their influences on environment and other agents**.** Ex: babbling as a part of speech development in babies.
Cannot solely learn from observation**.** Unlabeled Data.
Humans are especially strong at facial recognition compared to things like tumors
Critical period for learning in humans → what are the computational benefits/costs of this trade-off?
Power of Attention

- Overt (position sensors, eye movement) vs covert (internal recall and processing of data)
- Covert (prioritizing recall and processing of data in task-dependent way)
- Neuroscientists think covert attention is a limitation because a single representation can support many tasks, however attention allows for more efficient and accurate processing

Provocation: AI systems need more “attention” -> embed action into perception
Power of Emotions**:** Early on, reasoning and emotions thought to be separate, emotions more primitive.

Speaker’s hypothesis: Emotions are fast sensor fusion and reasoning.
Affect as information: (old) bridge example with attractive woman handing out surveys & a contact number to male passerbys while standing on a bridge
Two settings: low bridge over calm water or high bridge over turbulent water
Results: males more likely to call the number in high bridge and turbulent water case
Interpreted as the high emotional state led to this difference

Provocation: AI systems might be able to use bodily/physiological responses and emotions for fast decisions.
Power of heterogeneity:

- Society requires all types of people
- How much heterogeneity do AIs need?

Power of social Interactions

- Provocations: AIs should learn and exhibit some ability to do social interactions

Kids would not survive if their survival depended on NLP

Q: (on actions and representations). I think this is a point where machine learning is missing a lot. We just train on independently sampled data and don't think about representation. There is not a lot of established theory on how actions should shape representations of the data. Is there any consensus in psychology on how actions and experiences shape representations?
A: Most neuroscience/psychology is experimental. There isn’t much work on internal representation. We should have more work in this area to do this fundamental work. Collecting more data doesn’t solve representation.

Q: Language as a role in learning representation. What sort of literature is there? Is there strong evidence that our ability to speak shapes our representation of visual data?
A: There is some evidence and working hypotheses that facial processing does not tie itself to language. We’re bad at finding words for visually important features, such as “trustworthy”. But, there are exceptions. An example is that people who speak languages with more words for shades of blue are better at distinguishing different shades of blue (which direction is the causal arrow, though?)
Q: I think there is evidence on the relation between language and representation. If you use passive instead of active voice.
A: Not familiar.

Michael Kearns
Interpretability research is broken

Discussion of current “explainability and interpretability”**:** No clear definition for them.

Little discussion of human recipients of these explanations and whether or not they can understand it**.** Much of the research is just proposals of things we can try, and debunking of things that we can try

- Ex: LIME & SHAP
- LIME: Can't explain factors of why a decision boundary is the way it is but we can look locally
- SHAP: give explanation on importance of different inputs to model
- Neither LIME nor SHAP propose a definition of what explanation should look like

We can game these things → deliberately train biased models that can convince methods to generate explanations that don't address these biases. All areas of ML can be gamed by adversarial examples, why is this field different from others? There's no good definition behind these models. Research from Microsoft Research embed interpretability into specific contexts

- Had people predict housing prices based on various factors
- See if people change their predictions after interacting with models varying in complexity, transparency, etc.
- Subjects less likely to use model when it was transparent

Want human subject research that is more specialized and targeted than Amazon Mechanical Turk. A good explanation should help the recipient do something better than they would without the explanation" **.** Worried about anthropomorphizing (attributing human characteristics to other things).

- E.g. there is a cat detector learned from tons of YouTube videos. A whole neuron. It is activated by something that looks like a cat, so the natural explanation is "the activation looks like a cat"
- But what if the activation looked nothing like a cat (to humans), but was equally good at detecting cats in YouTube videos?
- Humans have a preconceived idea of how predictions and decisions are made, and maybe ML doesn't work that way.

Q: That cat detector's activation doesn't look that much like a cat. What if we weren't specifically looking for a cat - would we still identify it as a cat detector? (Dalmation effect)
Q: People need digestible explanations about things they don't understand. (e.g. is the vaccine safe…?) Is there a societal aspect to it as well?
A: It could be that you need an ensemble approach to the definition of a good explanation, and not just targeted at one person. That is fine. There is, however, a need to prevent collusion (in a Byzantine Generals sense). We should only make minimal assumptions about the recipient to prevent collisions or being overly specific. Want general purpose explanations.
Q: Over the last few hundred years, we've developed ways of creating explanations as scientists. Aren't we just reproducing that process?

A: The scientific process itself can be a rigorous form of explainability. This is not the same as explainability to the layperson. What does that look like?
Q: Not convinced that explanations need to be abductive. Should vocabulary change based on the recipient as it does in interpersonal conversations/explanations?
A: "Helps recipients do something better" is just one aspect of the definition of explainability. There could be more, especially representation-based ones. Currently, we just currently propose tons of heuristics without a good definition. These other categories of definitions should have some empirical verifiable way to check if the explanation is good.
Q: Would one define explainability through an interactive process instead of the current one-way street?
A: Yes, that's possible. We've even thought of zero-knowledge proof frameworks, where the recipient couldn't have generated the proof on their own.
Q: You mentioned you are worried about anthropomorphizing. But isn't the purpose to give explanations to people?
A: There is a camp of people that think of people that think we shouldn't give post-hoc explanations, just train models that are easily interpretable to humans. (mostly linear models.) But this feels unsatisfying - explainability shouldn't have to mean simplicity.
Q: Want to be able to give multiple answers, not just a single answer?
A: Yup. Need to.

Olga Russakovsky
A time for reckoning and reflection in computer vision

Where are we with ethics? Current papers in fairness and bias are published in engineering venues (CVPR, ICCV, etc.). Past 2 decades in CV focused on getting it to "work". Now: Get it to "work better". Reflect and understand what has happened: what went right and wrong? Where do we stand? How should we proceed? List of concerns from many different people

- **Tenet 1: GPUs** (problem → working solution)
    - Need lots of money and concentrates power to those who already have it (academia in richer companies can't compete well anymore) → efforts to report GPU time cost; need governmental intervention
    - Environmental impact → governmental intervention
- **Tenet 2: Fast, open-source iteration** (problem → working solution)
    - Limited time for reflection and large leaps - epsilon improvements at a time - not enough time to innovative new ideas → no easy answers, but need to educate students and rethinking incentives. Required reviewing for authors at NeurIPS to decrease load
    - Creates barriers to entry/catch up in the field (reduces diversity of thought and reduces anonymity b/c of focus on large labs) → efforts to welcome newcomers and affinity groups, diversity and inclusion mandates, social media restrictions

    - Increased risk of misuse of open-source code - stuff gets pushed into public before it's ready → Ethics statements required in papers to address this
- **Tenet 3: Big data** (problem → working solution)
    - Creates need for: Common benchmarks that we're overfitting to; Scale (needs $ and effort to build high quality large dataset); Leads to exploitation and undervaluing workers to create these → starting to pay people more, but not all countries say the same
    - Need for scale results in over-reliance on search engines
        - Bias towards people/places that upload and tag photographs
        - privacy/consent violations
        - Poor geographic coverage
            - For something even like soap: our models tend towards U.S. centric objects. Most of the world uses bar soap and the U.S. uses a lot more liquid soap
        - Limits applications to what we can work on - not every application can do big data
    - Some answers (problem → working solution)
        - Need common benchmarks → some opportunities to innovate but not much
        - Need for scale → pay workers more, engage in whole process, value intellectual contribution
- **Tenet 4:** Deep models (problem → working solution)
    - Difficult to interpret
    - Emphasizes accuracy on the most common case, leading to bias (difficult to even properly evaluate bias) → ongoing but difficult research; need better matches between loss and our goal; encouraging works that are tied closer to specific applications (is happening currently)
        - Unintended consequences of patterns in training data that do not hold
        - E.g. if women are commonly next to computer in photos, it thinks all objects next to women are computers
    - Transfer learning might be transfer of bias - we rely on pretrained ImageNet models - when ImageNet was constructed, nobody knew how widely they would be used. Transfer learning is an unintended use of ImageNet → no idea how bias transfers. We need more transparency in intended model design and use.
    - Now things are "working" , what's the next application to tackle? Are we benefiting the select few? → need diversity of thought and perspective! We ourselves are overfitting. Worried that people who are being harmed by neural nets being deployed have no voice.
- Big Questions

- Role of academia?
  - more interdisciplinary work. E.g. defining/studying/mitigating bias with social science departments
  - Education and outreach
- Role of government
  - International landscape
  - Ethical guidelines
  - Funding academia
- Overall, need a lot more work on diversity, equity, and inclusion → need people to agree that more diversity is good and what diversity we should be looking for

Q: This could just be computer vision/ML being new. This is the same as automobiles 200 years ago. Society is equipped to solve these problems and industry does want to solve these problems and some legal protections. So, what in your slides is completely new and specific to AI?
A: That's all true, but the reason why we made progress is because the issues were aired out explicitly with automobiles back then. Therefore, we have to air out the issues with AI. Continuously asking these questions is necessary, and it took a long time to solve it with automobiles.

Q: We resonate a lot with you, as a lab without that many resources. Do you think all this BIG is going to go on forever? Are these problems transient? There are some trends of smaller models that perform well with better priors.
A: Both. Fewer and fewer people can compete in the big model realm. Research that can be directly monetized will probably remain in industry and will continue to be done. This brings up a question of what type of research we value in a community, and what academics should pursue.
Q: Do you think economic pressure can naturally take care of the resource glut in fewer and fewer groups?
A: That research takes a lot of human power as well (data labeling). But, human talent (scientists) is the most expensive thing right now - GPUs are easy to set up. Hard to say.

Alex Smola
Learning machine learning: From Car Mechanics to Driving

Way more demand for ML competent people than supply. Teaching at universities is not scaling up fast enough**.** Solution: make the problem easier. For example, Python+CUDA vs. C++ vs. Minimal Python has made implementing ML models become a trivial task. Tools have gotten easier to install, but we still need to know how to use them**.** Using vs understanding what you are using - covariate shift, Simpson's paradox, algorithms beyond SGD, distributed training are all hard. Automation is difficult, expensive. Ex: training large models is expensive → use a

pretrained network and fine tune. We are currently fine tuning specialized ML components when maybe we really just want a stock, average model that doesn't take forever to set up?
Today, using AutoGluon's pretrained model can get #2 in Kaggle. AutoGluon (a set of pretrained resources) is doing better than a lot of other ML frameworks on Kaggle. Novice ML engineers can't beat it. So are we close to the desirable average setup that's easy to get?
On a scale of ML Automation (0: no automation → 5: full automation, we're currently between "driver assistance" and partial automation (1.5/5). Aspects of automation: model, data, systems, users

- Model: A lot of work on selection, tuning, validation, so we really know how to do this. Wide variety of base learners (trees, kNN, linear, old stuff). Still some issues: Don't know if the new model after k-fold cross validation using full data instead of partial data still works. Seems like this should be a minor issue
- Data: Real data is a mess → have to combine text, images, numerical data. We don't have a good way to do this yet. Statistical testing is ***not*** reliably automated requires parameters). Two-sample tests to monitor shifts and drifts (covariate, label, shift). Independence tests for time series, network data, relevant features, and explainability. But, this is absolutely necessary for automatic model building, automatic preprocessing, feature selection. There are so many extra things that are possible when testing is automated.
- Systems: Know tons of stuff about training. Combine models instead of using a single model. Provides robustness: fails only if all components fail. Stack models by concatenating outputs from individual models and send these as inputs to next layer → looks kind of like DenseNet. Use bagging to ensure conditional independence, but this does have slight label leakage. (becomes an issue with model stacking). Pretraining, combining models. Current SOTA is text naive text embedding, late fusion (there's got to be something better than late fusion, right?). The dumbest things seem to work. Deployment is now really easy with < 5 lines of code, if your system is compatible… Huge explosion of different numerical formats → affects accuracy of model. Hard to answer if they're close enough with certain disturbances → perhaps a place for formal methods
- Users: Productivity tools are mostly closed source (e.g. github, VS Code). There are some good open-source options, such as Jupyter, PyCharm

There's a lot of engineering work that needs to be done in order to make all this reliable

Q: People who worry about automation work in industry, not academia. Perhaps a Kaggle-type competition is a good way to get academia involved? Now suppose that we did it. Would that drive everyone in the right direction?
A: There is a conference coming up on AutoML. Kaggle doesn't force you to make your challenge dataset good, so any good challenge dataset is progress. ML community is especially

weak on multimodal datasets. Absence of such datasets is one cause of such weakness in community.
Q: I think the issue is not data. The issue is that you're giving the Kaggle datasets to experts in ML. See what happens when it's given to novices instead. Will they still achieve good results with a few lines of gluon?
A: Great idea! However, all the datasets were all in very similar formats.
Q: In some of the experiments you were showing, especially those with hybrid data, it seems like we're trying to force the same thing on all the data. Isn't this the right application for preprocessing?
A: We did have to find suitable representations. But shouldn't there be a way to do better than the dumbest thing?

Sidd Srinivasa
So, where's my robot butler?

Sci-fi robot butlers are super amazing, and realistic robots suck**.** So exactly where have robots been useful and robust? Industrial automation (beyond belts and pulleys), such as manipulating packages, Amazon KIVA robotics, Roombas … and more! "Reasonable value" + "not very complex"**.** But still not a robot butler ("high value" + "very complex")

- Tenet 1: Your robot will break. Only the exceptions make it to YouTube. Way more outtakes than good takes. Advice: Avoid scenarios where failure is catastrophic. Want to empower the robot to avoid what it should avoid on its own. Detect impending failure (e.g. roomba detects dog poop)**.** Gracefully handle exceptions - set up classifiers and precision-recall curves so that failures aren't terrible**.** Build virtuous cycles (e.g. robots learn from the exceptions it runs into and human takeovers). Slow is smooth (robots are usually better when they move slowly)**.** Always build the full stack ASAP (integrate early and often

Q: On staying away from catastrophic failures, when there's no risk there's no value (autonomous driving?). Wouldn't this shortchange the impact that robotics can have?
A: It's about stairstepping. The problem is so challenging and vast that we can really only make incremental improvements, especially if Safety First.
Q: Robotics doesn't have the benefit of a simple environment. Can you work so conservatively?
A: We have the benefit of planes not falling out of the sky. I am just frustrated by the inability of robotics to deliver on its promises
Q: We talked about language, intelligence, causality. When you have an embodied agent like a robot, there is always a tangible grounded metric for success. What hints have you learned in recent years to guide the rest of us?
A: Industry has access to the right questions and metrics which should be shared. Two big statements: COMPUTE is making things possible on robots (e.g. model-predictive control). As

for intelligence, robotics is still at a stage where you have to build it to know what can go wrong with it. I think that's a little crude. It's also a little un-democratizing. How do we get out of this loop of having to build the robot before knowing where it could fail? It's incredibly unbounded, because sometimes things randomly fail. We could use some assistance in understanding complex nonlinear phenomena. We'd love to build fewer robots.

Q: Clearly if you do manipulation, you can fail once in a while. But the roomba doesn't do manipulation. What are the key difficulties with the roomba?

A: Roomba is the epitome of a successful robotic system that has found its niche. You asked a good question. There is something about the dynamics of manipulation where failure is catastrophic and you're just a hair's width away from going wrong. Anytime you want an agent to interact physically with the world, you have to deal with phenomena such as friction that are chaotic/sensitive. Humans fumble all the time, but in the space of milliseconds, but don't notice. (need high speed video to see) Tons of closed loop control in a short amount of time makes it very challenging.

Q: My roomba is stupid. It doesn't remember what it vacuumed yesterday and we can't push it to new areas. Why can't we give the roomba a little more intelligence?

Q: How would one measure how smart a roomba is?

A: My bar is low. It just has to not try to vacuum something I'm telling it to not vacuum right this minute. This requires knowledge of context, really, and would be a tough metric to define.

A: Many failures are contextual. Hard to not be annoyed at a robot when it makes the same mistake many times.

A: Difference between comfort and safety. Metric to measure safety may not be comfortable and ideal for a robot. Much easier to make a metric for safety. Safety != intelligent.

Stefano Soatto
Embodying AI: The challenge from Gibson

The current paradigm is, given the data we have, how to train the most effective discriminant function, where effectiveness is measured by averaged error rate in a validation set. This purely inductive setting has a number of limitations, which we have witnessed in the practice: First, generally the average error rate is not a relevant metric. It is ubiquitous because it is easy to use, but individual users care for performance on their use case, their data. Furthermore, in the purely inductive case there is no "error signal" at inference time, no way to truly measure the uncertainty associated to a particular outcome of inference (as opposed to a classifier, again on average, with little practical relevance). This is a distinctly different paradigm from the experiment design setting, whereby – instead of trying to do the best possible job given the data we have, where even the best job may be utterly useless if so is the data (garbage-in-garbage-out) – we try to gather the data necessary to achieve a target performance during inference. For example in visual sensing, we are subject to nuisance variability due to occlusion, scaling, quantization, illumination etc. Simply put, if part of the scene is occluded, we can guess (based on data from different scenes having nothing to do with the present one) “what is on the other side”, but we cannot provide any form of guarantee. Conversely, having the ability to control the data acquisition process, one can reduce the effect of even non-invertible nuisances: To invert occlusion, move around the occlude; to invert scale, move closer; to invert noise, stay put and average (or select). The point is that, given sufficient control authority over the sensing process, we can actively reduce uncertainty to the desired target. This paradigm is not being adopted today.

Summary and Conclusion

The symposium provided a venue for open discussion of questions not often elaborated in the academic literature. Limitations of current systems are often discussed in perfunctory ways as motivations for “novel approaches” in the academic literature, where in reality such novel approaches are often minor increments to the current practice that do not go to the root cause. Fundamental limitations were exposed and discussed in the area of Correlational vs. Causative analysis, reasoning and language, explainability, bias and fairness, scalability, reliability, predictability of behavior, and eventually the possibility of reaching a state of AI where results are stable and reliable, not reliant on manual design, which has moved from the algorithms to the data, with even less transparency and predictability of behavior. Much of the current challenges can be attributed to growing pains of a field that has boomed suddenly, and will naturally evolve over time. To accelerate progress in this area, study on learning with few data, addressing the “long tails”, generalization beyond Machine Learning theory, active learning, uncertainty analysis, causality, and the Law are potentially fruitful avenue of research for the decade to come.