Scale AI Conference

10/6

From Seeing to Doing: Understanding & Interacting With The Real World With Fei-Fei Li

Data-Driven AI for Autonomous Vehicles With Dmitri Dolgov of Waymo
Dmitiri - cofounder of Waymo in customer strategy

- -going to full autonomy is hard because of understanding the environment
- -need to model data, bite off as much of torso of data to solve your models problems
  - -the long tail of one model may be the head of the distribution of another sensor/model
- -evaluation: need to know the performance of your model at the tail
- -planning/imitiation learning: auto labeling and simulator (reinforcement learning) can save time

- -NLP in general has the role of the human teacher, but imitation learning in a complex driving environment is a lot, but not enough
  - -need to augment intervention by humans to boost the performance by injecting some bias into the model, and use a simulator to explore some areas of the problem space to take care of edge cases

- -Waymo built their own lidar sensors
  - -want to build fully autonomous system, perception is just the start of the problem
  - -different sensor modalities complement each other
  - -tehyre on 5th gen of hardware suite, and have a lot of experience in what matters to their whole stack —> build what they need for where they're going at scale

- -they already offer fully autonomous ride hailing service in SF

Panel: Lessons Learned Scaling ML Systems

Hussein Mehanna - Cruise
- -VP of AI/ML for cars
- -infrastructure to build the models to deployment

Differences over 3 years
- -reproducibulity was a problem when they were smaller
- -they now train on much more data than before
- -increasing the automation used to train models

- -when you haven't seen something before you have to quickly handle it, but this is still reactive
- -people often obsess about solving the long tail for a single model, but there are some tricks by leveraging other systems to make the problem tractable
  - -don't need agi, can just leverage a helping hand to the model

-can leverage ML systems that try to get your system to the edge cases, i.e. using synthetic data, and at cruise you can leverage synth data for training perceptive systems

-There needs to be more investment in understanding the performance of robotics I the environment that they're in

-accelerate process of collecting data from the environment

-make sure offline test/simulated environment is as closely correlated to the real environment as possible, but it'll never be perfect

-need to solve problems from an end to end perspective… improving model is great but it has to work with both downstream and upstream processes

-need to collect data where the system is making mistakes

Danielle Dean - Aurora

-tech director of ML at iRobot

-now increasing testing at every layer (system and indie pieces) to know what went wrong in the process debug/jump into processes
-created a data collection and annotation team to structure and think about how to use their data

-how to design system as a whole so ML systems can work with other systems, data driven systems have some weaknesses, so how do you complement them?

-what pieces are you doing to get to scale, and what more parts can you automate to get to scale

KPIs
-how many solutions are in production providing real customer value?
-how many iterations in your solution to the point when you'd tell a friend, i.e. when utilizing synthetic data

Bart Nabbe - Aurora

-VP of strategic partnerships

-neccesary to collect the right amount of data, but what are the items needed? What is really important now? —> being explicit what specific data actually impacts a model

-have to build safe systems, becoming good at edge cases can't be relied on by a single network
-how many of these cases have been seen?
—> No single measurement can be left unexplained —> you might not understand it, but it can be important

-always set outrageous goals, acquired company to build simulation model

- -improvement of ML platforms as well as internal investment lead to better solutions
    - -what other techs will be needed in the future —> how to get better networks, how to etc
        - —> need to continuously invest in solutions

- -had 27k hours doing left turns in simulation before the car took an actual left turn

- -systems approach —> **how will your AI model impact the performance of the whole system**

- -how quickly can your system become better because there are always new things to learn
    - -how to bring in this new info to make a better system

Russell Kaplan - Tesla vision DNN, and Scale Nucleus

- -encounter new rarer edge cases every day

- -how much of the problem can be solved by ML, an d how much of the problem has to be augmented and solved by humans
    - -as long as the tradeoff between model and person, then the system gets better over time with fewer gaps

Jeff Wilke on Accelerating AI Adoption: From Ops to Customer Experience

-chairman and co-founder @ Re:Build Manufacturing and Retired CEO @ Amazon's Worldwide Consumer

- -former Amazon CEO
- -2010 saw some algorithms from data scientists using AI that were able to compete with conventional
- -around when image net came out, AI started to become king
- -lots of processes with humans in the loop
    - -manufacturing companies focused on transportation infrastructure, leadership… but haven't focused on data and AI
    - -who is the senior most data leadership entity in your company?
        - -these people are generally still in charge of writing code
        - -he contends the person with this CSCI skill must be sitting at the table in the senior level
            - -leads to companies outsourcing their data strategy
            - -they will lose to the people who've invested in this way
        - -this skill gap has spread between US manufacturers and Asian manufacturers

- -lean manufacturing —> reduction of waste and cycle time… dev in auto industry first, went to work at honeywell
- -Amazon in 1999 wasn't manufacturing warehouse, it was more like an assembly line which he was more used to
    - -why not apply techniques from lean (assembly line processes) —> reduce cycle time/variation

- -pre-2010 —> starts with applications of algorithms and applied math to solve problems

-optimize processes for time, quality and cost
-serach, process control, regression for forecasting, and they were applying these methods as they scaled from 2000-2010
-did this by making decisions based on this applied math
-had sensors all over the operation to and actuators to send info to machines and humans

-by 2010 systems were so complicated that their math couldn't solve these in real time, so they broke them up, which didn't lead to optimal solutions
—> this was his impetus to look elsewhere
-data streams from sensors/actuators were so large and the time to crunch numbers lead to simplifying the data which made mistakes

-applied classical math algorithms, but there were repetitive humans in the loop that really didn't have to be there, mainly in customer service —> very high bandwidth with a lot of data
-started to close this loop and move humans to more high value positions, and how to audit AI algos that were running the processes

-theres a wave going across all businesses —> getting a lot of digital data
-set up all the sensors and got a lot of data, which enabled data hungry algorithms
-without the data you don't have the enabling AI
-**many people have more data than they realize in data streams which they don't know how to use and optimize**

-more than 50% of people know a daily use of AI
-start in places where if you look across your whole operating stack, you can find optimizations already happening, and its a natural place to start implementing AI
-where do you have a lot of low skill people doing the same thing over and over
-NLP and text to speech can accomplish things that were complicated 10 years ago
-**can move humans to low skill things to high value positions (optimist)**
-can interact with outcome of AI to do something interesting
-can turn AI into a customer facing product as in Amazon Go stores
—> product can change given the application of AI

-Both finding things to automate, and making new products with new tools

-traditional companies are operational —> input to output, now to prioritize AI use cases (back office vs customer facing)
-**why do you have to prioritize, why not do both?**
-architecture too complicated? —> need to fix this and use the cloud to enable them to do both
-dont have enough employees —>
-where will the most value come from the easiest?
-many companies depend on software they didn't write, use of the shelf solutions
-strategically valuable use cases that change company trajectory, you need to find a way to get the resources to enable this

- make sure at least one person on the team sits at the company leadership level

- Scale enables companies to quickly get to a usable value adding system after identifying the need

-many companies using forecasting to make sure they have the right things at the right time

- machine was spitting out forecasts, but humans didn't trust the model so they were using their own forecast… the trick is to prove that the m machine leads to demonstrably better outputs given clean input
  - make sure the data streams aren't corrupted

-separate stateless computation from workflow

- make sure you can have the same engine run both simulations and production so that you're using the same worksflow to compare outputs

-any stochastic process can be improved by forecasting

-humans will push back and say they can do better than ai… in every industry

- solution: pick highest leverage areas where you can demonstrate that AI can perform better than humans, and demonstrate that it won't lead to humans losing their jobs, and letting them let go control of the wheel
  - the models humans made on spreadsheets can be hard to let go of
  - **where is the easier place to demonstrate superiority of AI in data driven decisions over human judgment**???

- best leadership solution is to educate workers over what the goals of the process are and to automate them
  - their jobs can change to auditing the system and understanding where the company is heading based on the solution
    - their jobs get better by not being repetitive and their company is more efficient and customer epericne becomes better

-concern that AI will result in millions of job losses…

- work of humans increased at amazon with AI, but they went to more skilled jobs
- move us toward more interest/safe/value adding work
- many things our brains still do better for machines, so you want people to be in those positions and not in doing problems that we've already solved

-if you went back in a time machine and saw what processes we've already solved —> predicability of our solutions

- in algorithms, this solution was predictable because you can characterize a problem in math, but have no ability to solve the problem in practical time given the current state of tech (i.e. batch time), but understanding computation will become quicker over time it was easy to imagine that these types of operations would be possible
- he was skeptical at the breadth of humans adopting this tech and using it every day, but understood it was possible

-biggest problem companies face is the lack of people who can tackle these problems

-lots of companies poised to benefit from AI but they allowed a hiring gap that they have to close, they must hire data sci tests so they can build solutions into process and product
- -amazon allowed these skills to wane
- -why does it take 2 days to settle money in financial industry?
  - -we've allowed old architecture to determine the status quo of customer experience

-need people sitting at the table with deep expertise, may start with someone outside of the company
- -but they need to start by hiring more data talent to understand how they can implement AI in their processes

-companies that are operations/manufacturing heavy —> you should love your technologists, don't just assume they're working on interesting stuff and will relay it to the highest level
- -execs must dig in to understand what they're doing and how they can contribute

The Data-Centric AI Approach With Andrew Ng

-working at steel plant to detect defects to understand multiple types
-using 1 tool they improved performance to over 90%
- -if you expect every steel plant to implement a new solution it won't work, but a data centric approach allows non-specialists to use AI
- -there were inconsistencies between data labelers, which led to months of failure

-model centric vs data centric approach to AI
- -data centric didn't exist before his talk this march

Evolution of new technology approach
- -1) handful of experts do it
- -2) principles are widespread and many apply it
  - -currently here
- -3) tools make the application systematic
  - -if data is inconsistently labeled, still depending on data scientists to find the problem
  - -starting to solve problems like measuring consistency between labelers
    - -data centric philosophy isn't to vote between labelers, but label where they disagree and revise the labeling instructions until they reach a consensus

  - -make data quality systematic —> usually throw in noisy data nd hope the model figures it out in training
    - -labelers aren't necessarily wrong, but inconsistency contradicts in training

Tips:
- 1.. Make labels y consistent —> ideally there is a deterministic mapping function x—>y with consistent labels
  - -clean data can lead to much faster training
- 2.. use multiple labelers to spot inconsistencies

3.. Repeatedly clarify labeling instructions by finding ambiguous examples
4.. Toss out bad examples of data
    -Weetok disagrees with this to an extent
5.. use error analysis to drive a systematic approach to focus on subset of data to improve

Iterative workflow
    -train model, error analysis to determine next step, improve data
    Improve y labels
        -multiple labels to measure consistency
        -improve label definitions
    Improve X
        -toss out noisy data
        -collect or augment more data

-preprocessing/engineering data is not 1 step, but an iterative component of monitoring/maintenance post-deployment
-AI = code + data

-model/softeware centric —> how can you change the model to improve deployment
-data centric —> how can you change data to improve deployment

The Challenges of Full-Stack AV Development With Jesse Levinson of Zoox
    -The type of data architecture you should have is dependent on the type of model you want to create and the desired outcome

How to Build an Effective AI Strategy for Business
Catherine Williams: Global Head of iQ at Qualtrics - formerly an academic
App nexus -> amazon -> xandr
    -working in industry allows her to create much more of an impact

    -lots of hype in VC around AI
    -Listen & remember, process & understand, build a culture of action
    -how best to utilize math/ML to drive business decisions
    -ANdrew Ng: AI is the new electricity

    -Business Impact = Strategy, People (roles, expertise, org), and Execution (data, tools, infrastructure, governance)

    Strategy —> problems AI will solve
        -seek to create knowledge or action for a task?
        -human in the loop or autonomous?
        -get more data, and extract more signal out of the data that you do have! (DNNs and ensembles)

    -what elements of your available data can fit with the problems you have to create the best product market fit?

Panel: Tools to Accelerate ML Development & Rate of Innovation

Siva Gurumurtha

- -100k cameras that were running for several years, had human labelers label 20 mil images
- -launched driver distraction feature, could id cell phones from video and predict whether the cell phone behavior was risky or not

- -edge and cloud architecture
  - -try to take a deployed model at edge and simulate prototype model as if it were operating on the edge —> more speed
    - -virtualize edge environment into the cloud to do end to end simulation —> prevents things from breaking
  - -reducing the cycle time

- -regression in model performance is real
  - -adding synthetic training data for edge cases can lead to worse performance in the head of the distribution —> making model subvjersioning necessary to keep track of performance given certain datasets
    - -this must be happening at scale
    - -33 test sets, 20+ models in a given year
    - -need automated engine to kn ow where the model fails or works better
      - -works well in the cloud, but when deployed to the edge it isn't necesarily obeyed

  - -everyone throws out a lot of ideas and some stick with the ML engineers
  - -gold standard to this process will be essential to helping organizations deploy successful models at scale

Sammy Omari

- -depends on use case/scenario to determine which direction to take
  - -detecting pedestrians stepping in front of cars —> inherently challaengiunng problem… solve by utilizing a few pictures in image, limited data
  - -most important how to mine for this data (fully/partially automated), and design this in a way for architects to iterate on this model very fast)
  - -attributes can be model error, but also want scenario attributes i.e. unexpected turns, and query these attributes for devs to quickly create new datasets to cover these cases
    - -in a cost effective way // commercially viable
    - -accelerate velocity of burning down the long tail

- -there is friction throughout the modeling process
  - -one big model with multiple heads, or separate models
  - -end to end improvement of system models is the goal
    - -figure out did you actually boost your performance on the long tail of solutions
    - -evaluate this in scale for many different scenarios
    - -mine for simulation scenarios
  - -large scale simulation —> challenge is to optimize signal to noise ratio when running millions of iterations

-Automate evaluation at scale —> as if you actually deployed the model (largest challenge)

-ML systems in industry —> trend to start combining multiple modalities in a single model, then spin out different predictions from

- -reducing number of ML nets, but 1 net doing multiple tasks
    - -need new way to structure teams around 1 end to end system
    - -how do you ensure the model doesn't neglect a sub sec of network when optimizing 1
- -still at early stages of how to do this in industry
    - -may impact system in ways you didn't expect with backdrop

-at end of day need to look at distribution of multiple scenarios

- -make binary benchmarks and unit tests, and aggregate metrics for the distribution of each metric
- -early fusion of data leads to complexity that makes clear boundaries of ownership more difficult
    - -puts higher value on infrastructure, with validation becoming more critical

Yanbing Li

- -simulation isn't just to validate, but gives synthetic data to dramatically extend into different cases that you don't typically see
- -accelerate ML, take the friction of building core models out, I.e. how to make launching a new experiment a matter of pushing a button
    - -seamless validation
    - -abstract away the complexity of underlying infrastructure so ML developers can focus more on outcome
- -managing aCI/CD processes correctly to have an efficient pipeline, have an efficient validation and deployment pipeline

- -firehouse of data, how do you handle and query it efficiently
- -use data lake to hone in on better analytics
    - -use data labeling to make the process more intelligent

Gonen Barkan

- -close iterations with the perception team led to his team to change the way they collected data to get something that actually mattered (traffic cone height)
- -autonmation is critical
    - -usually treat sensors as a fixed input, however you can modify the input of data to change the position of a sensor
    - -on the fly change sensor behavior, but means you need to detect new data on the fly
        - -many MLops say not to change the data, even though the hardware is in place
        - -NEED adaptable way to handle input data to quickly iterate on the way the data can be used

- -you need a cost function because a couple of KPIs won't give the complete story of how a model will perform

-critical to close the loop between the hardware group and the customers —> accelerates all of the neighboring teams so process can be tested in parallel
-resolution or SNR more importNt? Can't have both and need to decide early in the development process

-focus on dataset building for the long tail

Democratizing & Accelerating the Future of AI With Kevin Scott

-executive vice president of Technology & Research, and the chief technology officer of Microsoft

-ML tech is becoming more broadly applicable over the last 2-3 years
-broader range of people using ML to solve their problems
-ML has a greater impact on what's happening in the world
-help scientists solve problems
-increasingly packaging ML in ways such that many developers can easily use the technology
-MSFT and OpenAI codex/copilot, etc —> accelerate the innovation loop
-self supervised models at scale become better at doing tasks we know they can perform, and got to the point where pertained models can be foundation models
-not everyone should have to rebuild these from scratch
-solve a bunch of problems the initial model builders didn't consider
-take a model good at natural language, but can make it good at artificial language
—> can transfer these to graphs for molecular research
-combinatorial optimization problems are computationally expensive (NP hard), and in simulations there's a tradeoff between complexity and time scale
-solve using probabilistic heuristic hacks based on your understanding of the problem domain, which leads to non-optimal solutions
-use ML to dramatically increase the performance of simulations
-fluid dynamics —> becoming orders of magnitude better (2 min paper of simulating flow and games by predicting a physics engine)

-tools are here to assist us in doing cognitive work
-manage some complexity to become more productive in programming, not about removing worlds need for programmers —> want to help them become more productive
-widening aperture of who can be a programmer
-higher level communication w computer
-orders of magnitude increase in performance transforms what is possible
-democratize access to technology
-explain in concepts as opposed to language syntax
-success will be measured by how many businesses are formed due to the tech, or brand growth due to new applications of these tools
-"100s of billions of people to harness the power of machines"

-challenge is showing people that this is currently possible
-these methods require a huge amount of compute

-shift from “my team, my data scientists, my models, my data” to platform models where teams can take pertained models and deploy them without the need for underlying infrastructure or complete ownership over the whole process

-important that policy makers provide funding for public institutions to build computers capable of building/training large models on the scale of google/Facebook, even though these models are often available to the public

- -training can be much more efficient than it currently is
- -growing battle of AI standards between US/China, how to avoid 5G battle
    - -have good private/public partnerships
    - -countries have an ethos of how they want tech to impact their citizens
    - -need to invest more in AI
        - -silicon investment, chips, leading up to computers
        - -have robust pipeline of diverse students with education in AI/ML/CSCI

- -**Apollo program 2% of GDP/yr** => moonshots are super ambitious goals
    - -arbitrary goal, but the process and boost in technology was very real
- -smaller investment in AI with a specific problem in mind could lead to the creation of a similar foundation of technology
    - -I.e. healthcare, aging population, education
        - -more retired people than those who can take on the work that they’re leaving

- -look toward AI as one of the few answers too societal problems
    - -good and bad of what you can do with these tools => impossible to imagine all… need a grounding in both directions
    - -be better teachers of where the tech is and where its going
        - -**there is no Neil Degrass Tyson for AI**
            - **-What if EY had an AI mascot who publicized AI topics at a digestible level for the public**
    - -children performing well on standardized tests no longer matters because AI can already perform well on them… more important in how can children be taught and judged on creativity, not repetitive tasks of info recall

-Moonshots: security, how to use AI tools and defend us from malicious attacks in the world, grand challenge in manufacturing (global snarl in our supply chains) how to make this more robust so we can build everything society needs, and build this more flexibly

- -challegne of general health, how can you have expert companionship for elderly, i.e. so much needed in Maine that is unsolvable by any amount of money.. infrastructure isn’t there ./ no solution

-should have aspirations of real democracy of technology/AI (need within 2 years)

- -need objectives around how to measure real human benefit
    - -**Ai is good at optimizing objective functions, just need to form objective functions that accurately measure human good**
- -have a set of things that are working yet very inefficiently

Enterprise AI Strategy: Bret Taylor on Lessons Learned at Salesforce

-CEO of salesforce

-talk of digital transformation after pandemic (also data transformation), because if you dot have a digital company you no longer have a company (zoom meetings)
- -better customer and employee experience was needed
- -7500 chatbot interactions a day for insurance companies and jobless claims

-all companies need to use AI and be an AI company
- -experiences will be driven by AI experience, and need these to be personalized with specific people

-Gartner 2021: 90% of CIOs say they're working on AI projects, only 20% have deployed
- -**dont build tech for the sake of tech, have a purpose for it**

-**find the easy wins to make AI more accessible**!
- -tech needs to be turned on with clicks not code
- -scale is building intelligent software into tools —> lets companies use AI without building all the infrastructure
- -can work with partners to utilize tech that they use
- -diigtal experiences should all benefit from AI
    - -give sales managers coaching on how to make sales teams more effective

-sales cloud chat bots
- -service cloud voice —> Einstein transcribes what clients are saying and suggesting how to best help them to the service agent
    - -make people more productive, get the m to the outcome faster
- -same degree of personalization as huge cloud companies

-systems creating large pools of digital data
- -combine talent + data + available tools to create personalized experiences
- - +++ enable new experiences and processes / capabilities

-**supply chain optimization for companies with big supply chains due to disruptions is a current market**
- -CTOs want growth, how do companies find their way there now…
- -this is disproportionally around customer experience because it drives the top line
    - -no matter how many brick and mortar companies open, we're now in the habit of digital delivery
- -every company needs to build intelligence into their business strategy

-what is the best investment model into business case of AI
- -a lot of experiments have failed
- -value of AI working is so large

- -**understand why you're building what you're building, it all starts with the customer**
    - -increase of sales from recommended items have large easy to see increase
- -very easy to fund things that clearly succeed
    - -keeps the project from getting away from the company

-what outcomes are they trying to deliver?

-be practical —> execs can have sci-fi view of AI for everything —> short term view is to look at how can AI right now make the company more efficient

- -some sales reps weren't properly updating salesforce after calls, and the data wasn't showing through
  - -Einstein helped them fix an operational problem's single source of truth —> led to more rigor to make sure sales info is up to date
    - -**led to insight in human error in the company to become more effective as an organization**

- -what wins can we get right now, there are often low hanging fruit / short term value from implementing AI that the company can get right away

-get better at predictive intelligence is one of the most important use cases

- -**as systems understand more of the business, can look around corners that boardrooms don't necessarily see**
- -generalize this and make exec teams look stronger
- -**make accountability and predictability of the business stronger**

-Principles of data driven culture

- -Create data **cultures** and help everyone in your org understand data
- -everyone should be making data informed decisions, not based on the highest paid person
- -recongize this is a cultural shift, and your decisions shoulda be based on data
- -Tableu is focused on empowerment —> being an analyst was a domain of everyone not just experts
  - -systems shouldn't be so rigid that a random data scientist is the one making decisions, it should be eveyrone

-use AI to drive HUMAN decision making —> remove people from decision making process, as opposed to human intuition based

-automation is the goal of every CIO —> so many areas where people are copy and pasting tools that can be entirely automated

- -**enabling new user experiences** to drive greater engagement
- —> then you should focus on your customer (products, regions)
  - -**organize yourself around your customers**
  - -want a single view of you as the customer
    - -customer 360, have a single view of your customer around various touchpoint
      - -cant have this with silo'd data

-data is the foundation of being able to build any kind of AI strategy

The time to invest in AI is now, and you shouldn't wait for a discovery in the future

- -overestimate what we can do in 1 year, underestimate what we can do in 10 years

- -**BUILD a SINGLE source of truth (why?)**

-Call to action for people implementing AI

- -tech isn't inherently good or bad, its how you use it

-headlines around tech and misinformation, society is holding AI to a higher standard, and developers need to know to bring in ethics and trust to these conversations
    -not having these stakeholders in the design process is no longer an excuse

-cobsumer expectations around customer experience have changed, and your company is compared to all others quickly in real time
    -need to build personalization into your customer engagements
    -gain market share and customer loyalty

Transforming Drug Discovery Using Machine Learning

Insitro CEO - Daphne Koller
-Erooms Law —> drugs/R&D spending decreases exponentially
-build ML models that improve drug success prediction

-ML finds subtle patterns in immense amounts of high-content info that is too complex for humans to interpret
    -unsupervised learning/clustering will create a much more accurate hierarchy of classification that humans are unable to do

-genotypes are causal of phenotypes, so genotypes can be predictive of disease

-ML is just as much about the model as it is about the data

-10^80 space of possible drugs that could bind to target, but ML can quickly narrow this down

##
10/7
##

A Global Perspective on AI With Eric Schmidt

-founders of ai analyzed YouTube and found the pattern of cats —> started google brain who built bert/transformers
    -had a language model which materially improved search/advertising
    -many gains in revenue/quality started with this tech 10 years ago
    -purchased deep mind which looked expensive at the time but was important

-AI companies do something involving learning
    -the more clicks you have, the more learning opportunities you have
    -goal was to end up with the perfect search result and the perfect ad
    -built competitor on top of tensor flow
        -computer wouldn't find a different result, but would find it in a different way with a new correlation
    -directly improves customer quality/revenue and provides personalization without explicitly targeting them

-AI will be china vs US, not google vs apple (competitor not enemy)
    -build a system of scientists that brought us GPS and internet in the past through gov funding
    -scale of platform and extended work hours that allows china to create their own storms
    -we both generate AI for our own self dependent applications
    -internet used to be optional, now it isn't
        -china has criminalized being anonymous - recorded speech, vague laws… don't let American companies operate
            —> 2 different information spaces that AI will be built around

- -China is building an internet that ensures they remain in power
  - -only problem in their internet is regulation
  - -great firewall uses ML to find VPNs so you can no longer use them

- -democracies will form around US internet (Japan, etc.)
- -autocracies will form around china to use their control features

- -some countries highly dependent on China for trades but have democratic ideals
  - -Australia
  - -Belt and road countries will become clients of the Chinese information space
    - -they get their tech from china already
  - -#1 business parter and supply base of germany is china, same with many European countries
    - -incredibly uncomfortable competition with china where people will be forced to choose sides
  - -Tunisia will have more money and trade opportunities with China, how do they decide?

- -china has more data and less privacy rules
  - -algorithms are just as important as data
  - -US should build national research network
  - -most people would prefer to work in the west, should let them in
  - -government needs more technical AI talent to even understand these debates
    - -this is a national security challenge for the US
    - -crucial that tech companies can go as far as they can with AI, and wait until something bad happens to regulate… china isn't regulating

-government gives speeches all day, if you want to change it it has to start from the top down

- -AI will transform understanding of biology, chemistry, material science
- -next gen boom will be the industries that benefit from AI

- -tik tok is first breakout platform from China, because it matches to your interests, not your friends… he didn't think this would occur for 5 years
  - —> US has to organize themselves around initiatives to align AI with US values, and how agencies should operate with AI

- -there should be AI tools which should be protected like nuclear technology
  - -cyber interference is all new —> taken to the extreme will very dangerous and there's no analogy/doctrine to discuss this around countries
    - -easy to replicate with a copy, hard to contain

-each revolution seems more impactful than the last

- -society has values embedded in tech that we are raised in, tech is shifting these values
- -with targeting, AI WILL LEARN THE BIASES TO exploit
- -extraordinary gains with savants
  - -synthetic biology, going to the lowest energy state, which ML is good at

-MIT drug halisen —> want to build new general purpose antibiotic
-generate as many compounds that have antibiotic properties, then find one furthest away from the current ones
-takes out the costs from doing science

-new forecasting tool from founder of cal tech —> approximating how clouds work
-built natural model simulating a sleeping mouse given their physics engine, then generating labeled data of that sleeping mouse, and it worked

-sceince discoveries come after the invention of a new tool like microscopes, should be retrospectively look at available data to find discoveries

-do you create more jobs or less jobs —> less high qual job because winners get more of the spoils
-companies will excel and push out competition
-need to use the tools to make a higher skilled and employed workforce
-cognitive work is being more successfully automated (copilot)
-esoteric, difficult to train humans on because they're digital
-manual jobs are hard to automate

-passports, banks, vaccine cards made sense 20 years ago, now they don't

-for every industry and every problem —> can think of it as a learning problem that can can be optimized with AI

-internet came from Vietnam anti war movement, there is a concentration of power

-giving each person their supercomputer identity, they will also be powerful

-show your roommate what you're doing with AI and find a way to explain this to laymen

The Impact of AI in Financial Services & Beyond With R. Martin Chavez

CIO Goldman Sachs

-secret at Goldman Sachs algo trading was centered around organization and culture

-silly to use all your own software, use APIs —> blackbox for some to avoid regulation
-banks will be the winners of fintech because they've been making financial software forever

-how future first in class bank use AI
-need tight integration into the workflow of the company and its data
-without it there's no need to use AI
-can't just sprinkle AI into your operations, it needs to drive what you're trying to do

-silly idea that AI will predict the stock market
-if you have a stable distribution, like pictures of cats, then AI will predict it… don't have this with the stock market

-pricing puts and calls —> black sholls formal works under unnatural conditions
    -people have been tweaking/torturing this to get back to options prices that make sense
    -just make model on all options they've had, skip the hackery
        —> gave rise to better Greeks

-market makers sits on stocks from their customers, can't just sell them all
    —> how long to hold, how to hedge, how to predict new customers that would buy so you don't have to sell to dealers market

-no one knows what's in their customer contracts until a financial crisis when they have too figure out what they promised to people

-bring people in with various data science skillsets —> how do you then set these people up to succeed???
    -failure mode
    -data science center of excellence —> makes him want to bury his head
        -sprinkiling ML magic across the company —> big waste of time
            -in 1993 it made sense to make your own software, in 2013 Martin became CIO and said **Download**, **Buy, Build**
            -standard APIs when buying, want the ability to switch

    -**put people with a data science skill set next to the professionals they're training to impact**, and they have to work side by side as equal partners in the business, not as service providers to the business
        -these people ARE the business

-regulatory boundaries are where the API boundaries live
-AI happening on both sides of the regulatory boundary
-originate alone and hold it on your balance sheet

-why is it vital to foster diversity
    -the benefit of portfolio diversification —> same revenue with lower risk
        -uncorrelated views with diversity of people
        -proactively finding diverse people
        -he went into every business because there's a chance he can't help; p, but also a chance he can help because he can transfer things from another discipline's reality
            —> not just altruism, also good for business

Creating Personalized Listening Experiences with Spotify

-spotify has rich but fragmented collection of data
-golden datasets are rich datasets that are highly curated and high quality
-shared models between use cases

-personalization should lead to a lifetime of content, incentive to stay a customer

Developing Realistic Approaches to Deploying ML in Federal Environments

-government adoption of AI in government

Building a Framework to Accelerate the Adoption of AI for National Security

Mac Thornberry
U.S. House of Representatives, 1995 - 2021, Chairman @ Committee on Armed Services, 2015 - 2019

Mark Valentine
Head of Federal @ Scale AI and Formerly, General Manager, National Security @ Microsoft

Building the Next Generation of NLP Applications With Richard Socher

-Every aspect of AI is a company or multiple companies
-such much work in labeling/prepping data
    -so many people want to work pn AI but don't want to work on labeling data
    —> should get more data and make a bigger model

-goal oriented dialogue increases the scope of chatbots
    -think long term and enable your organization to have a research group

-in some industries AI has the potential for AI to massively transform / fundamentally change them, and there will become a divide between the haves and have nots
    -particularly biology
    -make your company more efficient in growth
    -if 1000 people work in call centers, chatbots can take over mundane tasks like password resets

-becomes more important how you train a model, and how to make it better over time
    -**have a single model doing multiple tasks to a high accuracy**
    -no one cracked the problem of whether final models can be the same across multiple use cases in NLP
        -sharing more and more of the decoder, but always have a separate model at the end
        -will enable doing cumulative research on the same model
        -NLP can solve a lot of different scenarios with the same model, but when you fine-tune it, you fork off one of the use cases to make that really good although other use cases may regress

        -have a specific objective function and go hard with it —> probably be pushes to make foundational models that can be built on top of
        -if open source software couldn't work, software would be much slower today
        -**increasingly complex objective functions** —> learn new skills, then catastrophic forgetting when the top layers get worse
            -if you know how to ride a bike, but don't do it for years, then it will be easier to pick up later on when you do it again —> models retrained

    -theres a lack of ability to cleanly build upon past work, there aren't clear abstraction layers such as in linux
        -we should be building upon 1 stable core model

-core area of insurance companies is to reduce their risk
    -customer service chatbot isn't a core competency so they can rely on partners for… **DON'T outsource your core competencies**

-Language models in generating code
    -software dev has become less about writing new algorithms, and more so how quickly can you use available software libraries
        -SaaS programs that help companies test, log, and do analytics
    -when you really innovate you can build yourself
    -dont have to know how to recreate libraries, just how to apply them

    -more people needed who can put together packages to make something with no bugs that can function

    -you will have fewer humans in the loop who will be doing more complex things

-more ML tooling, more use cases, and more ML + X (i.e. ML tooling + verticals in healthcare, battery optimization, sales optimizations, etc.)

-people will strive toward AGI

-**increasingly complex objective functions** —> like kids growing up who have certain sets of skills at different ages
    -series of objectives —> **what would it look like for AI to set its own objective functions**
    -multi-task learning is hard
        -trying to combine fuzzy probabilistic reasoning with discrete knowledge
        -system has so many parameters and training, but it won't be able to understand things that it hasn't seen before
            -how do you abstract to arbitrary settings through logical induction?

Growing With Open Source: From Torch to PyTorch With Soumith Chintala

-created Pytorch

-ML research right now is wide exploration of new ideas, but pytorch wants to help bring these into production
    —> more users and scale = more responsibility

-many people co-developing will provide diversity to bring people out of their thought bubbles, and create new ideas instead of echo-chamber

-ML researcher market needs flexible tools, and they will increasingly use weirder tools
    -tools that easily extend to user experience

-have a feedback loop with customers to know how they are performing and what else they can do

-scaling AI projects —> he got to physical human limits of how much he could keep up with, how to make a scalable approach without falling back on measurement?

-amount of bandwidth between people working on a project depends on how to scale vertically, with sharing of dependencies

Panel: Why Do Businesses Fail at Machine Learning?

Cassie Kozyrkov (Chief Decision Scientist @ Google)

-understand what you're selling
-remind people that ML isn't necessarily the solution to their problem —> think about the right problem, then think about the optimal solution to that specific problem
    -no reason to force in ML solutions which can be suboptimal, or otherwise solved more easily

-no guarantee that the data scientist understands how the domain works, but safest way is to let data scientists who are closest to the problem to kick off the project
    -make sure that if you're even going to do the project, there has to be someone with a deep understanding if the AI project is worth doing

-to what end are you exploring the technical details of a project
    —> always have someone you trust to verify that the project is worth diving into

-the more general the candidates responses, the more generally helpful and versatile they sound
    -however interviewers must be very specific on what they want to be accomplished in order to de-risk for a tight fit between project and DS
    -designing for skills/roles is nontrivial

-to get people excited about AI you have to make it sci-fi, but the most practical use cases aren't novel and are just efficient
    -look at excessively long view of civiliazation
        -advent of writing, ML is just another set of tools —> where did we take shortcuts and start doing things that didn't feel right in communication
        —> much rather talk out than write an email

Jaclyn Rice Nelson (Co-founder @ Tribe AI)

-dont jump to technical solutions right away, figure out who your users are… put on product hats and figure out end users, what are their problems, what is the data available, what do the customers need?

-always pair data scientists with product managers

Deepna Devkar (Vice President, Machine Learning & Data Platform @ CNN)

-important to define the business problems within the project

-want to build rec engine or optimization —> these are solutions, and not thinking about the problem doesn't set up the company for success

-data silos get bigger as companies get bigger

-figure out the requirements in the beginning, then work in a silo for 6 months to make a solution

-its important to work with the user along this process so that the project remains in line with the evolving solution

-For ML/analytics side, **have technical director, lead of products, etc. on any given project**

—> need a strong leader with all of these views

-need holistic view of business problem, circling back to delivery of user experience

-how to train data scientists to be this way is to ask the right questions all the way through the journey

—> **push them to think outside of the technical toolbox**

-good data scientists really want to drive fast impact for a company, and change their thinking into this

-some solutions are just copy and pasted from others

-people interested in roles may only be interested in the position title and level

-the more clearly you can define a role with expectations in first 6 months, the better you are set up for success

-customize interview panel, domain experts for the specific problems

-dont use ML to send people down the wrong path

Drew Conway (Head of Strategic Data Science @ Two Sigma)

-has watched many orgs go down the wrong path

-how do you address the right questions?

-how do you get folks in business adopt your solution to their problem?

-people parachuting in with a new tool will not lead to adoption

-this can be a visibility issue of using their data to see things that didn't see before

-if you don't figure out the minimally invasive way / seamless way to integrate a AI solution into the business framework then there won't be adoption

-worked in NYC gov building to use AI

-everyione had to sit in the seat of building inspectors to observe what they see, and you can map their perspectives to something that you can measure as a data scientist

-challenge DS's to own part of the problem

-how to best underwrite a particular part of an idea

-own it => you'll be judged on how you can solve a certain problem for a business and how willing they are to adopt it
-develop the relationships, understand the problem, how to think about solutions

-the technical side is <50% of the problem, its about how to help the company get better

-when thinking about building something… interviewing a senior candidate, what steps would you take to unpack a project
-talk about data, methods, defining a problem
—> but how are the customers actually interacting with the problem???

-when senior interviewing junior, need to understand less technical aspects, and how the 2 of you can work together to create value for the organization and the candidate's career

-cant be an all-encompassing data scientist
-industry specialization is important because it accelerates your ability to understand and be impactful on a problem

-most important skillset => how well can you articulate a result to a non-expert?
-lessen the intensity on this because it freaks people out, and get through conversation how they articulate ideas —> get this to people early in their career

-basic building block of solving problem with data —> bring together multiple datasets
-how an entity is represented in 1 dataset is different than the way its repped in another ds
-getting 20 datasets containing info about properties to align to a compiled property is difficult

Driving Competitive Edge with Responsible AI Innovation
David Carmona (General Manager, Artificial Intelligence & Innovation @ Microsoft Corporation)

-every company has become a software company, but AI can learn from experiences instead of being explicitly programmed
-company done with the pilot, how do you make it real and bring it to production?
-sense of urgency has been rapidly increasing
-changing how organizing are short terming thinking about AI
-business process optimization
-streamline the way the business is run
-employee productivity
-increased complexity and changes in operations
-spend more time on relevant work
-help reason and make decisions in real time
-customer service

-new needs from customers, and limited company resources to address this

-go beyond technical teams and business units to bring AI to every employee

Fighting Against COVID-19: How Can AI Help Build & Scale an Effective Pandemic Response?

-can't use 1 type fits all models for different kinds of data can't really get to a level of usefulness

-laying the groundwork in collecting high quality data is important, then building out the team to utilize the data at Curative

The Future of Data-Driven Innovation in AI With Jerry Yang

-cofounder of Yahoo

-people are starting to connect the dots through use cases
  -using AI in drug discovery/life sciences which takes industries into places they couldn't go without the tools

-focuses on data in his investment strategy
  -every fortune 500 company has a chief data officer (CDO)

-most CEOs brought their businesses up when success was more focused on operational strategy as opposed to data
  -vendor data, employee data, HR data (gold)
  -honest audit of what you have and where you should be given leading companies in the industry

  -comapnies have data but its not uniform and there's no strategy of how to use it over time
    -they need to understand the value of data, but you have to decide what are core competencies and what is not
      -find the right partners to handle your AI processes (APIs)

-AI used to be global collaboration, but now its being divided politically
  -US needs a national strategy around AI to remain competitive
  -deciding or not to give credit, or insurance AI bot, or NLP censorship
    -as models get more sophisticated, they'll reflect the value we want to create
  -separation of the 2 ecosystems around AI

Put humans in the center of what we need from AI ==> Human Centered AI

Purpose-Driven Leadership in the Age of AI With Jeff Weiner of LinkedIn

-exec chairman of LinkedIn

-Empathy = compassion + action
-important to take time to understand who a person is (their hopes and dreams), and figure out how to best set them up for success
- -champion compassion management
- -if you're angry and I'm empathetic, then I get angry

-as the team grew, people were appreciative
- -people began to understand what you were trying to accomplish, and self selected themselves in because they also wanted to solve that goal

-the more relevant results you get from a search result, the more value you get from it, or a recommendation
- -train AI to give better results over time without human intervention —> increasing relevancy
- -data powered everything in LinkedIn
  - -nothing on the platform can't be made better by more data
  - -the right headline infant of the right user at the right time or the right connection

-important to have an understand of your organizations values at scale, because decisions will be centered around them
- -must create those algos with the business values/logic in mind, or else it'll be stuck in a silo
- -algo can impact people in a negative way that they didn't intend
  - -create a team designed to test the impact of algorithms before they're deployed at scale, to understand how a small segment is being impacted, so they can correct it
  - -10x more likely to get a job with a referral in their pilot
  - —> led to the network gap, which has a lot of power
  - -realized they should close the gap between those who didn't have the social capital to leverage the referrals

-every time they implement changes in their models, they have to team to monitor it for unintended consequences
- -vicious cycles that AI perpetuates from data

-identify problem, quantity problem (research), solve problem through process/people/tech

-cant only be focused on execution day to day but also strategically
- -dont reactively help solve people's problems, but coach them how to solve the problems themselves

-culture: collective personality of the organization, and who they aspire to be
- -improving performance —> not just your what but also the how

-what it means to be a great leader is changing
- -inspiration
  - -leverage what you know to get others to perform toward a common goal
- -awareness
  - -aware of themselves, strengths and weaknesses, how the team is performing, what the team needs, aware of the macro in tech and cultural landscape
- -synthesis

-when you're hyperaware there is a lot of info coming at you, so you must figure out how to separate the signal from the noise, and use the signal

-need to embrace change

-people should ensure the right info is going to the right people so they can internalize the info themselves, write memos before meetings

-become more comfortable with ambiguity

-how they're taking care of themselves and their teams

-invest in the mental health of your employees

-companies, orgs, teams are moving so quickly that they're just trying to keep up with growth, so they put all their attention toward the what —> their performance, not the how of how they're getting there