Machine Learning, Market Structure & Competition

Carl Shapiro and Hal Varian

NBER Economics of Artificial Intelligence
Toronto, Canada
13 September 2017

Quick Review of Machine Learning

What is Machine Learning?

- **Predict Labels as Function of Features**
  - **Classic Approach: Construct Numerical Features, Construct Rules**
  - **Deep Learning: Use Raw Data, Learn Directly**
    - **Images: Pixels**
    - **Translation: Paired Documents**
    - **Transcription: Voice and Text**
  - **Requires Labeled Data (OpenImages), Hardware (GPU, TPU), Software (TensorFlow), Expertise**
- **Optimize Using Reinforcement Learning**
  - **Multi-Armed Bandits**
  - **Chess, Go, Atari Games etc.**

What Can Machine Learning Do?

- **Kaggle Predictions**
  - Passenger Threats; Home Prices; Traffic to Wikipedia Pages; Personalized Medicine; ImageNet; Taxi Trip Duration; Product Purchases; Clustering Questions; Rental Listing Interest; Lung Cancer Detection; Click Prediction; Inventory Demand
- **Demand: Match Customer & Product**
- **Supply: Reduce Cost and Waste**
- **Substitute and Complement Humans**
  - Reduced Demand: Cashiers, Drivers, Translators
  - Increases Demand: Analytic Skills

What ML Inputs Are Scarce?

- **Data Infrastructure: Critical Prerequisite**
  - **Collection, Manipulation, Storage & Retrieval**
  - **System Integrators Can Play Big Role**
- **Software: Open Source & In Cloud**
- **Hardware: Can Be Purchased in Cloud**
- **Expertise: Scarce But Growing Rapidly**
- **Firm-Specific Labeled Data: Key Input**

Obtaining Labeled Data

- **Multiple Ways to Obtain Needed Data**
  - As By-Product of Operations
  - By Offering a Service (GOOG411, Flickr)
  - Hiring Humans to Label Data
  - Buying Data from Provider
  - Sharing Data (Perhaps Mandated)
  - Data from Governments and/or Consortia
- **Data is Non-Rival, Partially Excludable**
  - Rights, Permissions, Licensing, Regulation
  - "Ownership" Too Narrow a Concept for Policy
  - Example: Who Control Driverless Car Data?

Big Data, ML and Public Policy

- **Does Access to Data Give Incumbents a Major Competitive Advantage?**
  - Entrants Must Build or Acquire Necessary Data
  - But: Entrants May Have Data From Adjacent Markets
- **Incumbents Also Learn How to Improve Algorithms and Business Processes**
  - Shape of the "Machine Learning Learning Curve"
  - Domain Knowledge Can Be Important
- **Apply Essential Facility Doctrine to Data?**
  - Scope of "Essential" Data that Must be Shared?
  - How to Regulate Terms & Conditions of Data Access?

Machine Learning Meets Good Old Industrial Organization

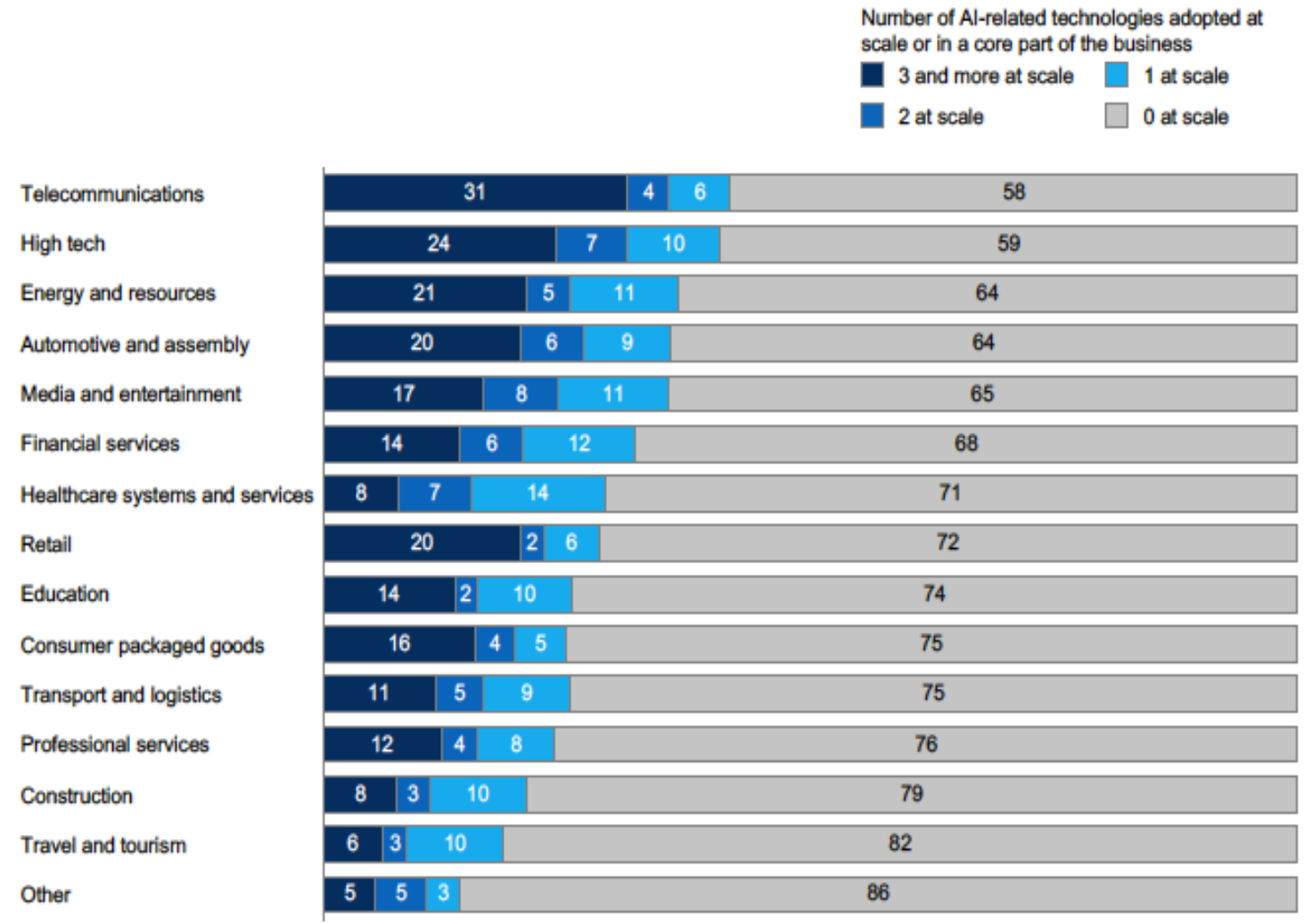
Adoption of ML Technology

- **Which Firms and Industries Will Successfully Adopt Machine Learning?**
  - **Large Heterogeneity in Timing of Adoption & Ability to Use ML Effectively**
- **Can Later Adopters Imitate Early Adopters?**
  - **Patents & Trade Secrets; Firm-Specific Routines**
- **Role of Geography in Adoption Patterns**
- **Very Large Competitive Advantage for Early, Successful Adopters**
  - **Large Firms? New Firms? Disruptive Aspects**

Evidence on AI Adoption

- **McKinsey Global Institute Survey**
  - 3000 "AI Aware C-Level Executives" in 10 Countries
  - 20% Are "Serious Adopters"
  - 40% are Experimenting or are "Partial Adopters"
  - 28% Feel Their Firms Lack the Technical Capabilities to Implement AI
- **Key Enablers of AI Adoption**
  - Leadership, Technical Ability, Data Access

AI Adoption by Industry (McKinsey)



Key Research Question: Machine Learning & Vertical Integration

- **How Will Machine Learning Tools and Data Be Combined to Create Value?**
  - **Within or Across Corporate Boundaries?**
- **Will ML Users Develop Their Own ML Capabilities or Purchase ML Solutions from ML Vendors?**
  - **Classic Make vs. Question, Key for Industry Analysis**
- **One-Stop Shopping in the Cloud is Happening**
  - **Data Labeling, Software, Algorithms, Consulting**
  - **Special-Purpose Hardware: Tensor-Processing Units (TPUs) Create Cost Advantage**

Machine Learning and Vertical Integration: Some Public Policy Questions

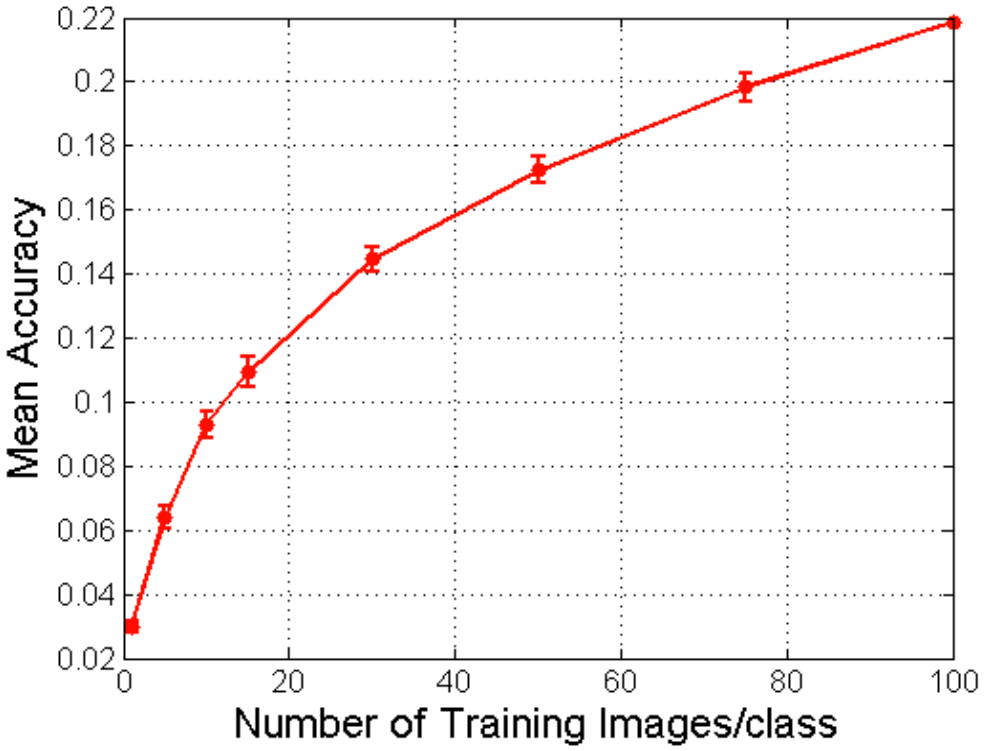
- **Privacy Regulations May Limit Ability of ML Vendors to Combine Data from Multiple Sources**
  - **Limits on Transfer of Data Across Corporate Boundaries and/or Sale of Data**
  - **Privacy Concerns vs. Growth of Markets for Data Used for Machine Learning**
- **Mandated Data Sharing May Promote Vertical Disintegration**
- **Treatment of Vertical Mergers Between ML Vendors and ML Users**

Machine Learning Vendors

Structure of the ML Industry

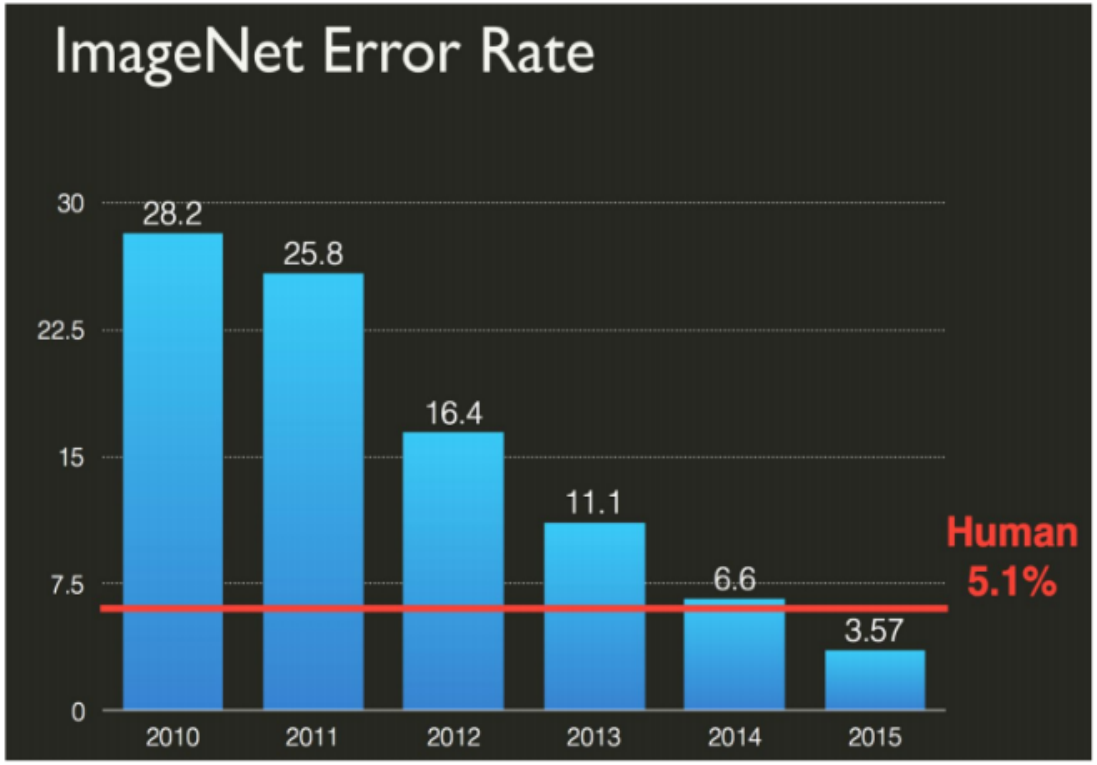
- **ML Vendors Offer Several Services**
  - Data Centers, Containers, Dockers
  - Labeling Services, System Integration, Consulting
- **ML Vendor Could Specialize in ML and Purchase Data Processing/Storage in Cloud**
- **Industry Structure is Oligopolistic**
  - Leaders: Amazon, Google, Microsoft, Salesforce
  - Other Suppliers: IBM? Who Will Be Next?
- **Will Industry Become More Fragmented?**
  - Specialists by Industry?

Diminishing Returns to Scale



Source: Stanford Dogs dataset

ImageNet Progress: 2010-2015



Data Size Held Constant
Improvement Due to Hardware and Software
Source: Stanford ImageNet

Pricing of ML Services

- **Large Fixed Costs, Low Marginal Cost**
  - **Undifferentiated Services & Bertrand Trap?**
  - **Size of Customer Switching Costs**
  - **Containers & Dockers**
- **Learning by Doing for ML Vendors**
- **Multi-Product Offerings and Bundling**

Pricing of ML Services: Google

Feature	Price per 1000 units		
	First 1000 units/month	Units 1001 - 5,000,000 / month	Units 5,000,001 - 20,000,000 / month
Label Detection	Free	$1.50	$1.00
Text Detection	Free	$1.50	$0.60
Safe Search (explicit content) Detection	Free	Now free with Label Detection	Now free with Label Detection
Facial Detection	Free	$1.50	$0.60
Landmark Detection	Free	$1.50	$0.60
Logo Detection	Free	$1.50	$0.60
Image Properties	Free	$1.50	$0.60

Pricing of ML Services: Amazon

Amazon Rekognition API Pricing

US-East (N. Virginia) | US-West (Oregon) | EU (Ireland) | AWS GovCloud (US)

Image Analysis Tiers	Price per 1,000 Images Processed
First 1 million images processed* per month	$1.00
Next 9 million images processed* per month	$0.80
Next 90 million images processed* per month	$0.60
Over 100 million images processed* per month	$0.40

Impact of Machine Learning on Downstream Markets

Impact of ML on Minimum Efficient Scale?

- **Will ML Generally Increase Minimum Efficient Scale by Transforming Variable Costs into Fixed Costs?**
  - Fixed Cost of Developing a ML Solution
  - Substitutes for Variable Labor Costs
- **Not if the Fixed Costs of ML are Small**
  - Off-the-Shelf Generic ML Capabilities vs. Need to Develop a Specialized Solution
  - See: Pricing Structure for ML Solutions
- **ML Could Lower Minimum Efficient Scale**
  - Reduce or Eliminate Certain Fixed Costs

How to Start Up a Startup

- Fund Your Project on Kickstarter
- Hire Employees Using LinkedIn
- Purchase Cloud Computing Services from Amazon
- Use Open Source Software: Linux, Python, Tensorflow
- Set up a Kaggle Competition for Machine Learning
- Communicate Using Skype, Gmail, Google Docs
- Use Nolo for Legal Documents
- Market Your Product or Service Using AdWords
- User Support Provided by ZenDesk

Use of ML for Downstream Pricing

- **Far Greater Price Discrimination?**
  - Yield Management Goes Bananas
  - Auctions and Other Mechanisms
  - But: Customers Can Use ML Counterstrategies
- **Group Discrimination – Many Groups!**
  - More Data on Which to Condition Prices
  - Blurs Line Between Individual and Group Pricing
- **Self-Selection & Product Differentiation**
  - Customized Products
  - But: Competition + Low Consumer Search Costs

Algorithmic Collusion: Economist Catnip

- **Classic Issue of Dynamic Oligopoly Pricing**
- **Rapid Response Equilibria**
  - **In Markets with Transparent Prices**
  - **Firms Move Far Faster Than Consumers**
- **Evolution of Machine Cooperation**
  - **Can Machines Find a Better Way to Coordinate?**
  - **Taking MFNs & MCCs to the Next Level?**
- **Instructive Examples**
  - **NASDAQ; ATPCO; Spectrum Auctions**
  - **Machines Learning Cryptographic Code**
- **Antitrust Implications: Who Goes to Jail?**