Module Part 2

Session: Prediction, Future Scanning, Recap (back to the future).

1. Which types of legal issue are amenable to a computational perspective?

2. How might visualization help in understanding (legal) data?

3. What are the various potential interactions between people and machines in the future

4. What are the limits of machine/human interactivity?

- Big Data
- IBM Watson

Workshop Session: Creating Web sites by structuring and styling your pages with HTML & CSS
https://www.codecademy.com/learn/web

Supporting files can be downloaded from here:

http://scisweb.ulster.ac.uk/~kevin/legal

Direct links to files are:

http://scisweb.ulster.ac.uk/~kevin/legal/LawTech Module 2 Notes.pdf

http://scisweb.ulster.ac.uk/~kevin/legal/LawTech Module 2 Prezi PDF notes.pdf

Reading Material

Contents

1.1 Electronic Discovery

Electronic Discovery also known as e-discovery is the process of discovery in civil litigation that is carried out in electronic formats. It covers what most often is referred to as electronically stored information or ESI for short. Some examples of ESI included are e-mails, instant messaging chats, documents, accounting databases, CAD/CAM files (Computer-aided design/ Computer-aided manufacturing), websites and any other electronic information that could be relevant evidence in a law suit. Also included in e-discovery are "raw data" and "metadata," which forensic investigators can review for hidden evidence. E-discovery can be carried out offline on a particular computer or it can be done in a network, court ordered or government sanctioned hacking for the purpose of obtaining critical evidence is also a type of e-discovery.

There are six stages of the e-discovery Process:

1. **Identification**

This stage is when potentially responsive documents are identified for further analysis and review. People who are in possession of potentially relevant information or documents are identified, to ensure a complete identification of data sources, data mapping techniques are often used. As the scope of data can be overwhelming at this phase, attempts are then made to reduce the overall scope, such as limiting the identification of documents to a certain date range or search terms to avoid an overly tedious request.

2. **Preservation**

A duty to preserve begins upon the reasonable anticipation of litigation. During this stage, data identified as potentially relevant is placed in a legal hold. This ensures that the data cannot be destroyed it is ensured that this process is defensible, while the end goal is to reduce the possibility of any data being spoiled or destroyed.

3. **Collection**

At this stage now the documents have been preserved, collection can begin. Collection is the transfer of data from a company to their legal counsel, who will then determine the relevance and disposition of the data. Companies that deal with litigation frequently have software in place to quickly place holds on certain custodians when an event such as legal notice is triggered and begin the collection process immediately. Other companies who deal with litigation less frequently may need to call in digital forensic experts to prevent any data getting spoiled. The size and scale of this collection is determined by the identification phase.

4. **Processing**

During this stage, native files are prepared to be loaded into a document review platform. This phase often involves the extraction of text and metadata from the native files. Various data selection techniques are used during this phase, such as deduplication. Sometimes native files will be converted to a paper-like format such as PDF to allow for easier redaction and bates-labelling. Modern processing tools can also employ advanced analytic tools to help document review attorneys more accurately identify potentially relevant documents.

5. **Review**

The review stage is where documents are reviewed for responsiveness to discovery requests and for privilege. Different document review platforms can assist in many tasks related to this stage, including the rapid identification of potentially relevant documents, and reducing the amount of documents according to various criteria such as keywords. Most review tools also make it easy for large groups of document review attorneys to work on cases, to speed up the reviewing process and eliminate work duplication.

6. **Production**

At the production stage documents are turned over to opposing counsel, based on agreed specifications. Often this production is accompanied by a load file, which is used to load documents into a document review platform. Documents can be produced either as native files or in paper-like format again such as PDF alongside metadata.

The key to being successful with e-discovery is to have a policy that stresses:

- Information without business value must be disposed of according to policy and in the normal course of business.
- Records that have value to the organization must be stored and managed properly, under the control of the organization.
- A classification scheme providing an information and records management structure must be implemented for consistency and control.
- Records that are no longer needed must be destroyed in a systematic and documented way.

Common issues in e-discovery:
In an e-discovery project there may be a number of different people involved such as lawyers for both parties, forensic specialists, IT managers and records managers among others. Forensic examination often uses specialized terminology for example, "image" refers to the acquisition of digital media. This can lead to confusion. While attorneys involved in case litigation try their best to understand the companies and organization they represent, they may fail to understand the policies and practices that are in place in the company's IT department. As a result, some data may be destroyed after a legal hold has been issued by unknowing technicians performing their regular duties. To avoid this trend many companies are deploying software which properly preserves data across the network, preventing inadvertent data spoliation. Given the complexities of modern litigation and the wide variety of information systems on the market, electronic discovery often requires IT professionals from both the attorney's office (or vendor) and the parties to the litigation to communicate directly to address technology incompatibilities and agree on production formats. Failure to get expert advice from knowledgeable personnel often leads to additional time and unforeseen costs in acquiring new technology or adapting existing technologies to accommodate the collected data

Tools for e-discovery:
SonaVault eDiscovery Software and eDiscovery Tools
Business and organizations faced with eDiscovery legal challenges need effective tools to gain compliance quickly and thoroughly. SonaVault eDiscovery Software Tools, which are a part of SonaVault Email Archiving Software, are powered with Microsoft SQL Server Standard Edition. This allows enterprise-class searches and data management for compliance purposes. SonaVault automatically places legal holds as part of the retention policy. Flagged email is automatically held in a case, and the same email can be flagged in different case. (A case refers to an internal dispute or a legal situation that an organization is involved in.) Retention policies determine the lifetime of the emails that remain in the SonaVault Email Archive Server. There are two types of retention policies, global and custom. Global policy is applied on an organizational level and is processed for all emails. For e.g. SEC 17a dictates that emails should remain on the system for seven years. So, global policy can be set to seven years. Custom policies can override global policies and can be used to expire unwanted (for e.g. SPAM or subscription) email prior to the global policy period. SonaVault eDiscovery Software and eDiscovery Tools meets industry-defined retention policy requirements such as Sarbanes Oxley (SOX), Health Insurance Portability and Accountability Act (HIPAA), Gramm-Leach-Bliley Act (GLB Act), Security Exchange Commission (SEC) Requirements, Financial Industry Regulatory Authority (FINRA). With SonaVault eDiscovery Software and eDiscovery Tools, IT administrators can create different groups and define the retention policies according to each group. For instance, if an education center only wanted to hold emails from students for three years, but wanted to retain the emails from faculty and staff for seven years, then it can create multiple groups and assign different retention policies unique to each group.

Nuix E-discovery Workstation
Nuix eDiscovery Workstation is a complete and scalable off-the-shelf solution for processing, investigation, analysis, review and production. It can be installed anywhere, including behind a company firewall and in the cloud. It directly indexes virtually all forms of unstructured information, in any language. It can process complex proprietary formats such as Lotus Notes, Microsoft Exchange, Microsoft SharePoint, webmail and forensic images. Nuix technology identifies and makes searchable the information other software routinely skips. It is quick to deploy and rapidly processes and analyses data in a way that is forensically preserved and defensible. It is simple to use and its graphical displays show clearly who communicated what, to whom and when. It exports to all major review platforms. It incorporates all Nuix's advanced capabilities in one license price and can scale to meet the most demanding caseloads.

1.2 A New Way to Look at Law, With Data Viz and Machine Learning

Source: http://www.wired.com/2014/06/ravel-law/

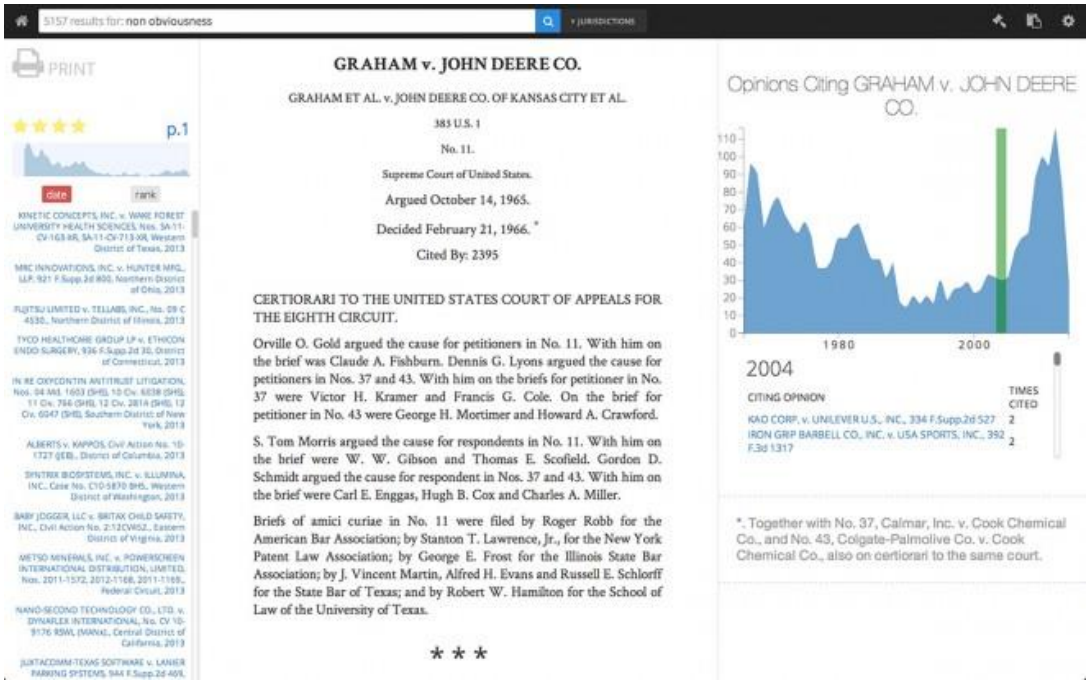
ON TV, BEING a lawyer is all about dazzling jurors with verbal pyrotechnics. But for many lawyers–especially young ones–the job is about research. Long, dry, tedious research. It is that less glamorous side of the profession that Daniel Lewis and Nik Reed are trying to upend with Ravel. Using data visualization, language analysis, and machine learning, the Stanford Law grads are aiming to reinvent legal research–and perhaps give young lawyers a deeper understanding of their field in the process.

Lawyers have long relied on subscription services like LexisNexis and WestLaw to do their jobs. These services offer indispensable access to vast databases of case documents. Lewis remembers seeing the software on the computers at his Dad's law firm when he used to hang out there as a kid. You'd put in a keyword, say, securities fraud, and get back a long, rank-ordered list of results relevant to that topic. Years later, when Lewis was embarking on his own legal career as a first year at Stanford Law, he was struck by how little had changed. "The tools and technologies were the same," he says. "It was surprising and disconcerting." Reed, his classmate there, was also perplexed, especially having spent some time in the finance industry working with its high-powered tools. "There was all this cool stuff that everyone else was using in every other field, and it just wasn't coming to lawyers," he says.

Telling the Story of the Law

Ravel started as the duo's project for LaunchPad, a hugely popular course in Stanford's venerable design school. The site, which has since garnered upwards of $9 million in VC funding, brings a number of powerful tools to the research process. Search results, instead of coming back as a block of text, are rendered as an interactive visualization. The cases take the form of bubbles, arranged by date. Landmark cases are nice and big; lesser cases are smaller. Lines join the circles, showing you how the cases are interrelated. You can filter these visual results in a number of ways, separating out, for instance, which rulings came from district courts, which came from circuit courts, and which were handed down by the Supreme Court itself.

Westlaw and Nexis, Lewis points out, have long reigned simply on the basis of access. They were gatekeepers to all legal history, and a subscription was an obligatory expense for any legal firm or law school. In recent years, however, much more of that information has become freely available. And with it, new tools have become possible. As its creators see it, Ravel's visual search offers myriad improvements over the old columns of text results. It better lets you see how cases evolved over time, and potentially lets you see outliers that could be useful in crafting an argument–cases that would languish at the bottom of a more traditional search. The visualization, Reed insists, "tells a lot more of the story of law than the rank ordered list." (That might be true. When they first showed their visual search to a veteran judge, he looked at the complex map of circles and responded: "This is how my brain works!") Ravel also has some smart touches for processing cases once you've found them. A clean interface makes it easy to skim documents, for instance, with built-in tools for highlighting and annotating text. Early users have reported that Ravel cut their overall research time by up to two thirds, on some occasions.



Ravel's reading interface

Mapping the Law

Ravel's most ambitious features, however, are intended to help with the analysis of cases. These tools, saved for premium subscribers, are designed to automatically surface the key passages in whatever case you happen to be looking at, sussing out instances when they've been cited or reinterpreted in cases that followed. To do this, Ravel effectively has to map the law, an undertaking that involves both human insight and technical firepower. The process, roughly: Lewis and Reed will look at a particular case, pinpoint the case it's referencing, and then figure out what ties them together. It could be a direct reference, or a glancing one. It might show up as three paragraphs in that later ruling, or just a sentence.

Once those connections have been made, they're handed off to Ravel's engineers. The engineers, which make up more than half of the company's ten-person team, are tasked with building models that can identify those same sorts of linkages in other cases, using natural language processing. In effect, Ravel's trying to uncover the subtle linguistic patterns undergirding decades of legal rulings. That all goes well beyond visual search, and the idea of future generations of lawyers learning from an algorithmic analysis of the law seems quietly dangerous in its own way. Still, compared to the comparatively primitive tools that still dominate the field today, Lewis and Reed see Ravel as a promising resource for young lawyers and law students. "It's about helping them research more confidently," Lewis says. "It's about making sure they understand the story in the right way." And, of course, about making all that research a little less tedious, too.

Note: Large list of visualisation tools http://guides.library.harvard.edu/visualization

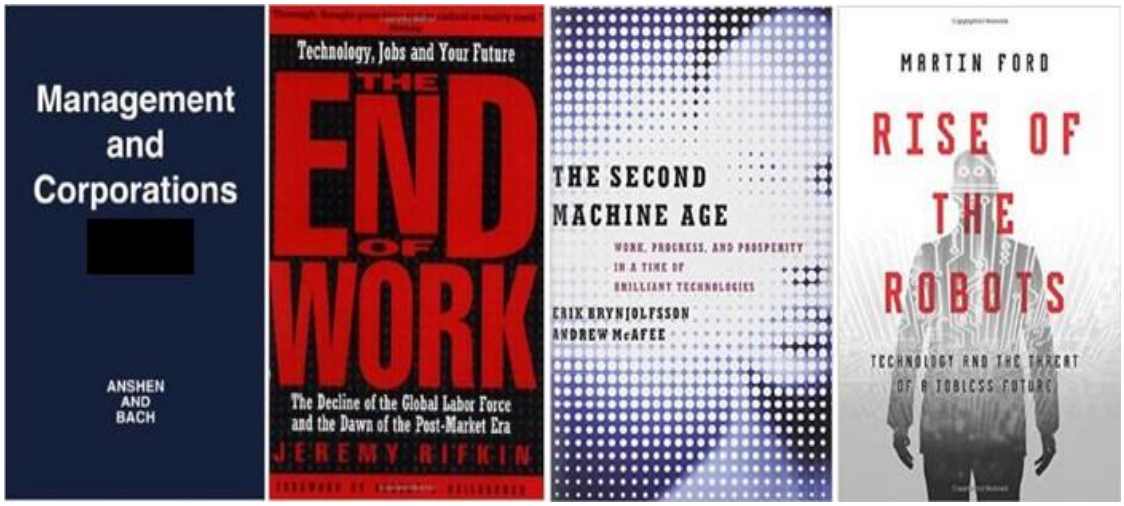
1.3 The 'Jobless Future' Is A Myth

Source: http://www.forbes.com/sites/stevedenning/2015/06/04/the-robots-are-not-coming/#3fdb34b2e392

For connoisseurs of human folly, there is hardly anything more fascinating than the spectacle of the almost universal acceptance of an idea that is logical, obvious and wrong. A current occurrence is the media acceptance of the idea that, because "the robots are coming for your jobs," in future there will soon be "no work for human beings." The idea is simple and frightening. The power of technology is increasing. Technology substitutes for human capabilities. Soon the robots will be doing everything, including the most sophisticated brain work and even management. Therefore there will be steadily less need for human capabilities. Soon, there will be no need for human workers at all. This leads on to even more shocking conclusions. In future, computers will do everything and rule the world. Human beings will become their slaves. The forces driving these trends are inexorable. Resistance is futile. There is little we can do except crawl into a hole somewhere and await our horrifying machine-driven destiny. According to Stephen Hawking "Full artificial intelligence could spell the end of the human race."

The Prophets Of 'The Jobless Future'

Concern about the effect of machines on work and on people has been around ever since machines were first invented. As computers emerged and machines became more powerful and capable, the level of angst has risen. Nobel-Prize winning economist Herbert Simon addressed the issue back in 1960.



Two decades ago, extreme alarm was expressed by futurist Jeremy Rifkin's in The End of Work {1996): "In the years ahead," wrote Rifkin, "more sophisticated software technologies are going to bring civilization ever closer to a near-workerless world."

More recently, Erik Bryjolfsson and Andrew McAfee in *The Race Against The Machine* (2011) and *The Second Machine Age* (2014) argued that with the advance in technology, computers can do it all. Computers now drive cars in traffic, translate between human languages effectively, and beat the best human Jeopardy! player. Computers are rapidly encroaching on skills that used to belong to humans alone. This phenomenon is both broad and deep, and has profound economic implications. Humans aren't keeping up. Although technology isn't destiny, humans are losing the race against the machine. "In domain after domain," they write, "computers race ahead." In calling this, "the second machine age," the authors risk giving the impression that machines are the inevitable victors. Now a new book by Martin Ford, *Rise of the Robots* (May 2015), proclaims even more loudly the arrival of "the jobless future." As technology continues to accelerate and machines begin taking care of themselves, fewer people will be necessary. "Artificial intelligence is already well on its way to making even good jobs obsolete. As progress continues, blue and white collar jobs alike will evaporate, causing massive unemployment and the implosion of the consumer economy itself." Ford is currently giving appreciative interviews left and right, as his message of "the jobless future" sweeps the media. Others such as Bill Gates, Elon Musk, and Steve Wozniak have expressed similar concerns. Many mainstream social scientists have accepted the findings. The conclusions, writes the Wall Street Journal, are "all but inevitable." Inevitable but wrong.

The Flaws In The Reasoning

There are a number of flaws in the reasoning in Ford's book. One flaw is the underlying assumption that whatever is feasible will occur. In science fiction, this may be true, but not in real life.

- The fact that a Concorde can fly faster than a Boeing 767 doesn't mean that Concordes replace 767s. The outcome depends on the costs and benefits of developing and operating the two types of aircraft. Airlines and airline passengers have not embraced Concordes because it's not economical to do so.
- Similarly, the fact that a computer can do something better than a human being doesn't mean that the computer will replace the human being. The market will determine whether it is economical to do so, given the costs and perceived benefits.

A second flaw in the reasoning is the implicit assumption that computers with miraculous performance capabilities can be developed, built, marketed, sold, operated and replicated at practically zero cost and that they will have zero secondary employment effects. In reality, huge teams of people are often necessary to perform these tasks at considerable cost. So it's not obvious either, on the one hand, that the necessary investments will be made or, on the other hand, that the secondary effects of the innovation on employment will be irrelevant or negative..

A third flaw is the failure to consider how the marketplace will react to the computer as a new market entrant. Ford's book assumes that superior performance on one dimension will cause the market to rapidly embrace the computer and abandon the human worker entirely. Actual experience shows that this is never the case with new entrants into a marketplace. The new entrant may appeal to some customers but will not appeal to others. In a free market, prices will determine the eventual proportion of the market shared held by each.

A fourth flaw in the reasoning is to assume that when machines replace human capabilities, as they have been doing for thousands of years, nothing else changes. In reality, as Philip Auerswald, Associate Professor of Public Policy at George Mason University, and author of *The Coming Prosperity*(2012) points out, when machines replace one kind of human capability, as they did in the transitions from hunter/gatherer, from serf, from freehold farmer, from factory worker, from clerical worker, from knowledge worker on to whatever comes next, in each case, new human experiences and capabilities emerged.

Often these new experiences and capabilities were unimaginable in the prior era. The new experiences and capabilities were mostly higher value, and offered more interesting work, than the experiences and capabilities that had been replaced by machines. This was true of agriculture, industrialization, mass production and so on. Why should it be different now?

Why Watson Failed In Jeopardy!

These flaws in reasoning are evident in most of the book's many examples. Let's take one of his most spectacular and most frequently cited examples: the triumph of IBM's Watson over human champion Ken Jennings at "Jeopardy!" It's true that the best human contestant in history was defeated by a computer. But what was the net result in the television marketplace? Nothing. At the end of the show, Jennings quipped, "'Quiz show contestant' may be the first job made redundant by Watson, but I'm sure it won't be the last." Jennings, like the authors of these books, was mistaken. Jeopardy! has not been transformed into a show where rival computers fight each other. Jeopardy! continues to be a show with human contestants. Just ask Aaron Rodgers. The fact that Watson was able to come up with better answers in one competition doesn't mean that future human contestants are replaced. The reality is that the program's audience wants something that Watson doesn't have: a human personality with all the quirks and flaws of real people. The conclusion that Watson is better as a Jeopardy! contestant is correct—except in the one aspect that matters most to the audience: being human.

The book gives an interesting account of at least some of the vast expenditure of money and human manpower over many years by IBM to create, maintain, market and sell Watson. But when the book comes to evaluate Watson's impact, it forgets all this expenditure and declares Watson the absolute winner. The economic reality is that it's not cost-effective for IBM or any other firm to continue submitting Watson or its clones as a contestant on Jeopardy! when there are much cheaper alternative contestants: i.e. human beings, who also happen to be more interesting to the audience. As a commercial enterprise, developing computers as Jeopardy! contestants was a wonderful one-time PR stunt, but it was a commercial failure. Watson has no

future commercial prospects in Jeopardy! Its role in Jeopardy! has been discontinued by IBM and there is no prospect of it being taken up by other firms. IBM has moved Watson and its expensive team of supporting experts on to other business challenges, including food trucks at SXSW in Austin. Time will tell with such business endeavors are any more sustainable than the Jeopardy! exercise. The point is that having more intellectual capability isn't by itself sufficient for success: the use of computers in any setting has to be cost-effective, including all the costs of developing, creating, selling, marketing and maintaining the computer, integrating it with other services, and including the market demand for the offering and the consequent price for which it can be sold.

Is The Jobless Future Already Here?

But surely computers are already replacing humans, doomsayers like Ford cry. We can see it every day! Machines are cheaper and better and more reliable! ATMs are replacing bank tellers! PCs are replacing secretaries! It's obvious! Jobs are already disappearing before our very eyes! "Employment for many skilled professionals," writes Ford, "including lawyers, journalists, scientists, and pharmacists— is already being significantly eroded by advancing information technology."

But is it true that "the jobless future" is already here? The question is, as always, not whether individual jobs are being replaced, but rather what is the net effect of *all* the changes in employment. ATMs and PCs for instance don't instantly appear out of nowhere at no cost. They have to be designed, built, marketed, sold, introduced in the marketplace and in the workplace, and reconciled with all the other things and systems going on. All of these activities require large amounts of human time and effort. The net effect of ATMs and PCs is not simply the removal of the jobs of bank tellers and the secretaries. We have to add in all the other jobs and work that are being created, and then see what the net effect is. Are machines already rapidly replacing labor on a net basis? Not according to current statistics. If it were true, we would be seeing massive gains in worker productivity, i.e. output per worker. But when macro-economists look at the productivity numbers for the last decade, they find exactly the opposite. Productivity gains as measured in conventional terms have been anemic. It seems likely that, as in the past, the apparent loss of human jobs to machines is still being moderated by the creation of new kinds of capabilities needed to tend to the machines and the new experiences being generated by the machines.

We can see how this transition plays out today, by looking at the steps being taken by a firm like General Motors that, like all car manufacturers, is investing heavily in automation. Most of the capabilities it is looking for in future employees would have been unimaginable just a few years ago. Why should we think that the capabilities required in future will be only those we know today? Thus to understand the future employment situation, we would have to not only subtract the human capabilities that are being replaced but also add in the future experiences that will be generated and the human capabilities that will be required to deal with those experiences. When it comes to predicting the future, it's hard to fully imagine what those new experiences and capabilities will be. So it's not surprising that a failure of imagination can lead to a lot of angst about "the looming jobless future."

Will The Future Be Different?

So "the workerless future" hasn't happened in the past. And it hasn't happened yet. What about the future? Ford is confident that, this time, the future will be different. His reason: new technology. *"Up until the moment the first aircraft achieved sustained powered flight at Kitty Hawk, North Carolina, it was an incontrovertible fact— supported by data stretching back to the beginning of time— that human beings, strapped into heavier-than-air contraptions, do not fly. Just as that reality shifted in an instant, a similar phenomenon plays out continuously in nearly every sphere of technology. This time is always different where technology is concerned: that, after all, is the entire point of innovation. Ultimately, the question of whether smart machines will someday eclipse the capability of average people to perform much of the work demanded by the economy will be answered by the nature of the technology that arrives in the future— not by lessons gleaned from economic history."*

Yet technology cannot free itself so easily from the principles of economics. The principles of economics still apply as much to the airline industry as to prior forms of transportation. Planes were a new technology but they did not replace all other forms of transport. There were some types of travel for which planes were the best fit and others for which they weren't. Over time, the marketplace sorted out the price and the demand for each. Some offerings, like the Concorde (or Watson in Jeopardy!) which are superior in one dimension—

speed (or the ability to answer Jeopardy! questions)—have ended up with no market at all. In the same way, the marketplace will sort out the role of machines and humans. A totally jobless future is not among the real possibilities.

The Coup De Grace: Comparative Advantage

Ford suggests the need to "rethink comparative advantage." But before we do this, we need to understand it. Comparative advantage was developed by David Ricardo in 1817 to explain the conundrum: why do countries engage in international trade even when one country's workers are more efficient at producing *every* single good than workers in other countries? Why isn't everything produced by the more developed country? The theory of comparative advantage is regarded as one of the most powerful yet counter-intuitive insights in economics. In 1960, Nobel-Prize-winning economist, Herbert Simon, applied the theory of comparative advantage to explain the related conundrum of employment: if machines can increasingly do whatever humans can do, why won't everything eventually be done by machines?

Simon explained this in a remarkable talk entitled "The Corporation: Will It Be Managed By Machines?" in *Management and the Corporations*. edited by M. L. Anshen, and G. L. Bach. (McGraw Hill, 1960). I am indebted to Philip Auerswald for drawing it to my attention. In his 1960 essay, Simon argued that the content of work will change, but because of the operation of comparative advantage, there will be still be work for human beings who want it:

> "The change in the occupational profile depends on a well-known economic principle, the doctrine of comparative advantage. It may seem paradoxical to think we can increase the productivity of mechanized techniques in all processes without displacing men somewhere. Won't a point be reached where men are less productive than machines in *all* processes, hence economically unemployable?
>
> "The paradox is dissolved by supplying a missing term. Whether man or machines will be employed in a particular process depends not simply on their relative productivity in physical terms, but on their cost as well. And cost depends on price. Hence–so goes the traditional argument of economics–as technology changes and machines become more productive, the prices of labor and capital will adjust themselves as to clear the market of both. As much of each will be employed as offers itself at the market price, and the market price will be proportional to the marginal productivity of that factor.
>
> "By operation of the market price, manpower will flow to those processes in which productivity is high relative to the productivity of machines; it will leave those processes in which its productivity is relatively low…
>
> "We conclude that human employment will become smaller relative to the total labor force in those kinds of occupations and activities in which automatic devices have the greatest advantage over humans; human employment will become relatively greater in those occupations and activities in which automatic devices have the least comparative advantage…
>
> *"…full employment* does not necessarily mean a forty-hour week, for the allocation of productive capacity between additional goods and services and additional leisure may continue to change as it has in the past. *Full employment* means that the opportunity to work will be available to virtually all adults in the society…
>
> "In the entire occupied population, a larger fraction of members than at present will be engaged in occupations where 'personal service' involving face-to-face human interactions is an important part of the job."

The final point is crucial. As technology is introduced, humans move up the value chain from commodities to products to experiences. The possibilities in terms of new capabilities and experiences are literally infinite. We should therefore start from an assumption of abundance, not one of scarcity.

All Problems Are Technological?

To a man with a hammer in his hand, everything can resemble a nail. And Ford, a technologist, sees technology as the cause and the solution to pretty much everything. Among the more extreme examples is his treatment of the shocking gap that has emerged since the 1970s between the increases in worker productivity and the lack of increase in worker compensation.

6. Productivity and real hourly compensation, manufacturing sector, first quarter 1949–third quarter 2010

NOTE: The shaded bars denote National Bureau of Economic Research (NBER)-designated recessions.

To suggest, as Ford does, that the gap is caused by technology is to overlook the well-documented role of top management in redirecting those productivity gains to shareholders and themselves.

Real Issues In Employment

None of this is to suggest that there are not real issues in employment today, including:

- The fact that for the last 25 years firms older than five years have been net job destroyers.
- The fact that management practices flowing from shareholder value theory are inimical to full employment.
- The fact that majority of big firms are implementing obsolete management practices (hierarchical bureaucracy) that cripple investment and innovation and jobs.
- The fact that more than half of the off-shoring decisions taken by US business ever the last several decades were not based on the total cost of off-shoring even at the time they were made.
- The fact that an over-sized financial sector is diverting money and talent away from investment in real products and services and towards zero-sum and negative-sum games that do not help the economy or jobs.
- The fact that middle class jobs are hollowing out, as new jobs tend to be at the high end or the low end, not in the middle..
- The fact that, as the marketplace for work increasingly becomes a freelance economy, public policies are not in place to encourage and moderate the development.

However to understand and resolve these issues, Ford's book offers more confusion than help. We need to stop agonizing about an apocryphal vision of a "jobless future" and to focus on the pressing real issues that we can actually fix.

1.4 Big Data

In 1988, foreshadowing much that is now claimed in the field of 'Big Data', Harvard's Soshana Zuboff made the following claim in her ground-breaking book In the Age of the Smart Machine: 'Information technology not only produces action but also produces a voice that symbolically renders events, objects, and processes so that they become visible, knowable, and shareable in a new way. In more homely terms, she was referring to the value of the great streams of information that are generated as a by-product of computerization. For example, early stock-control systems yielded useful information about customers' buying habits. Zuboff termed this the 'informating power of intelligent technology'. Although the term 'informating' did not catch on, her main insight is now received wisdom—that analysis of large bodies of data generated by our technologies can provide us with valuable new insights, and can help us make more responsible predictions in many fields.

The discipline that has emerged to specialize in this capture and analysis of information is now popularly referred to as 'Big Data'. When this term was first used, it was confined to techniques for the handling of vast bodies of data—for example, the masses of data recorded by the Large Hadron Collider. Now, Big Data is also used to refer to the use of technology to analyse much smaller bodies of information. Some speak instead of 'data analytics', 'data science', and 'predictive analytics', all of which seem to mean roughly the same thing. Specialists in the area, whatever label is preferred, are often called 'data scientists'. There has been no shortage of hype about Big Data. There are commentators who argue, with some justification, that its claims are too extravagant and that its methodology is underdeveloped. What is hard to deny is the volume of data that are swilling around. In 2010, Google's Chairman, Eric Schmidt, claimed that we create as much information every two days as was created between the dawn of civilization until 2003. Following available figures, by 2020 that quantity of information will be generated every couple of hours. This leap can be attributed in part to the quantities of video, images, and audio content that are pouring onto the Web, and in part to the rapid growth of inexpensive sensors. In relation to the latter, on one view, the 'proportion of the world's data that comes from such sensors is expected to increase from 11 percent in 2005 to 42 percent in 2020'.

The upshot of all of this is that great volumes of data are now at large, and the broad aim of data scientists is to develop methods for collecting, analysing, and exploiting these data. Case studies of success in Big Data abound. One (not entirely uncontroversial) illustration is Google Flu Trends, a system that can identify outbreaks of flu earlier than was possible in the past, by identifying geographical clustering of users whose search requests are made up of similar symptoms. Another is provided by Walmart, which analysed the buying habits of its customers prior to hurricanes and found not just that flashlights were in greater demand but so too were Pop-Tarts; and this insight enabled them to stock up accordingly when the next storm came round. Natural language translation systems and self-driving cars are also said to operate on the back of Big Data techniques. While there are many ways in which Big Data is valuable, most specialists in the field would agree with Mayer-Schönberger and Cukier that, '[a]t its core, big data is about predictions...it's about applying math to huge quantities of data in order to infer probabilities...these systems perform well because they are fed with lots of data on which to base their predictions'. More extravagantly, Eric Siegel, a computer scientist, goes further when he speaks of 'computers automatically developing new knowledge and capabilities by curiously feeding on modern society's greatest and most potent unnatural resource: data'. If we combine these views of Big Data, we can see its promise for the professions—as a way of making predictions and as a way of generating new knowledge. To pin this down further, we have to identify the data sources that might be used here. We have in mind the data that are created by professionals in the course of their work. This includes the information they gather and the guidance they provide—in medical records, legal files, financial accounts, tax returns, architectural drawings, consulting reports, and so forth. In the past, before the Internet was widely used, and even in its early days, most professionals expressed little interest in capturing and analysing these bodies of data. Instead, the focus of most professionals was on the particular job at hand. The data generated (the 'data exhaust', as some would name it) were treated as though they were disposable once the file was closed. To some extent, it was largely left to academics, if to anyone, to gather together and study the data that were created as a by-product of the delivery of professional work in the conventional way. The revelation from the field of Big Data is that if this work product is captured, it may well yield patterns, correlations, and insights not previously recognized by professionals. This could be new practical expertise of a sort, and might form the basis of valuable predictions. This is not the equivalent of a professional informally saying, 'we have

seen this before and it is likely that...', although it would be useful to have this formalized. Rather, the use of Big Data should identify trends and unearth knowledge that professionals simply had not noted or known of in the past. In this way, data can be regarded as 'a priceless collection of experience from which to learn'. Already there is analysis of large data sets that link symptoms to diagnoses in medicine, fact patterns to judicial decisions in law, and performance to teaching methods in education. And as techniques become more sophisticated, these data could produce medical diagnoses, legal predictions, and educational insights that human practitioners could not.

The new knowledge that flows from Big Data techniques falls into the category of 'practical expertise'. Whether this is practical expertise or not does not depend on its origins-either a human mind or data and software—but whether it can be used to solve a particular set of problems. Our expectation is that systems based on Big Data techniques will often draw conclusions, offer advice, and provide guidance at the standard of human experts or higher. It is important to stress that these high-performing systems will not simulate or model the way that human beings work. A system that makes a diagnosis on the basis of a comparison between a particular patient's symptoms and a database of 10 million past patients is not carrying out a differential diagnosis like a regular human doctor. Nor does a system that predicts the decision of a court by comparing the facts of a case with a database of hundreds of thousands of past cases operate like a normal practising lawyer. Big Data techniques do not and will not automate professionals' working practices. Instead, by capturing and reusing huge bodies of past experience, this technology provides an approach to professional work that simply was not possible in the past. In the words of Patrick Winston, a leading voice for decades in the world of artificial intelligence, 'there are lots of ways of being smart that aren't smart like us'.

1.5 IBM's Watson

IBM's system Watson, which we regard as a landmark development in artificial intelligence, was not designed to solve problems in the way that human beings do. Watson was developed in part to demonstrate that machines could indeed attain exceptional levels of apparently intelligent performance. Named after the founder of IBM, the system was developed to compete on Jeopardy!—a TV quiz show in the United States. This represented IBM's latest contribution to the branch of AI that in the 1980s was called 'game-playing'. Previously, IBM had developed Deep Blue, a computer system that beat the world chess champion Garry Kasparov in 1997. In the early 1980s such a system had seemed beyond our reach. It was clear to most researchers who were working on chess-playing systems that the very best players in the world seemed to make creative, intuitive, and strategic leaps which were beyond the understanding of the players themselves, never mind that of those who were trying to develop systems. In the end, of course, brute-force computing, fuelled by the exponential growth in processing power, delivered a system that could outperform the best human players, largely by being able to look many more moves ahead—and so not by playing in the same way as humans. While there is some merit in the claim that chess-playing machines are not really engaging in the same pursuit as human chess players, to diminish the systems excessively on this ground is to commit a version of our AI fallacy—the mistaken view that the only way to develop systems that perform tasks at the level of experts or higher is somehow to replicate the thinking processes of human specialists. This error is also found in much of the critical commentary on Watson.

The development of Watson, of course, was a different order of challenge. To compete well on Jeopardy! requires contestants to have deep and wide-ranging knowledge, which is precisely what AI critics said in the 1980s was beyond the scope of computers. As is now well known, on 14 January 2011, on a live TV broadcast of Jeopardy!, Watson beat the two best ever human contestants. This was a truly remarkable achievement. This is a computer system, effectively, answering questions on any topic under the sun, and doing so more accurately and quickly than the best human beings at this task. It is hard to overstate how impressive this is. It represents the coming of the second wave of AI. Here is a system that undoubtedly performs tasks that we would normally think require human intelligence.

The version of Watson that competed on Jeopardy! holds over 200 million pages of documents and implements a wide range of AI tools and techniques, including natural language processing, machine learning, speech synthesis, game-playing, information retrieval, intelligent search, knowledge processing and reasoning, and much more. This type of AI, we stress again, is radically different from the first wave of rule-based expert systems of the 1980s. It is interesting to note, harking back again to the exponential growth of information technology, that the hardware on which Watson ran in 2011 was said to be about the size of the average bedroom. Today, we are told, it runs on a machine that is the size of three pizza boxes, and by the early 2020s Watson will sit comfortably in a smartphone. Contemplate the impact of Watson-like technology when it is applied across the professions. It is a leap that IBM have themselves made. On their website they expressly say that 'Watson has been learning the language of the professions and is trained by experts to work across many different industries'. IBM speak of the 'Watson Ecosystem', a community or organizations that are developing Watson based applications. Lawyers, doctors, bankers, insurers, and educationalist are already involved. While the commercial opportunities are considerable, in early 2014 it was reported that IBM also intends to invest $100 million in a ten-year initiative to use Watson to help with healthcare and education in Africa.

It is notable that one of IBM's main applications of Watson is in the field of healthcare. Already Watson-based systems can perform diagnoses, prepare treatment plans, and conduct research to a high standard. Although IBM and the medical experts working with them are cagey on this point, for some tasks it appears that Watson is already outperforming human beings. Before long, we expect, many more tasks will be executed by Watson. We conclude, from our own study of Watson, that the technologies already exist to support the development of powerful systems in other professions. The day will come, for most professional problems, when users will be able to describe their difficulties in natural language to a computer system on the Internet, and receive a reasoned response, useful advice, and polished supporting documents, all to the standard of an expert professional practitioner.