The Risks of Technology in Business

Raghavendra Rau

5 June 2023

Introduction

In the last two lectures, I spoke first on how big data needs to be organised to be useful to process. Next, I tried to explain how AI systems work. In this final lecture, I will explain how a blind reliance on these systems may lead to major issues for society.

Technology has revolutionised the world of business in many ways. One of the major ways it has done this is to reduce the asymmetric information it has about its stakeholders such as customers and workers. With the help of technology, businesses can track consumer behaviour, preferences, and reactions to various products and services. They can use this to tailor their products and services to meet the needs of their target audience, hoping that this will lead to increased sales and customer satisfaction.

Specifically, technology helps firms solve problems of information asymmetry, in particular adverse selection and moral hazard. Adverse selection occurs when one party has more information than the other in a transaction, leading to an unfair advantage. For example, when you buy an insurance policy, you have more information about your health and lifestyle than the insurance company. Knowing you have a higher risk of getting a certain disease might make you more likely to buy health insurance. This creates adverse selection because the insurance company may end up covering a higher number of high-risk individuals, which can lead to increased costs for the company. The same issue holds with car insurance companies who would like only safe drivers. As an example, in previous lectures, I discussed how Root Insurance tries to solve this problem by monitoring your driving before making you a quote.

Moral hazard refers to situations where one party takes on more risk because they know the other party will bear the cost of their actions. For example, an insurance company may want to charge higher premiums to a customer engaging in risky behaviour because they know the customer is more likely to file a claim. But the customer can show the insurance company that they are not very risky and change their behaviour after they get the contract. After getting health insurance, for example, they may be more likely to engage in risky behaviours that could lead to health problems because they know that the insurance company will cover the cost of their medical treatments. They may be more likely to engage in high-risk activities such as extreme sports or smoking because they know that their health insurance policy will cover any resulting medical costs. In previous lectures, I discussed examples such as Lemonade which try to solve the moral hazard problem through technology.

In two of my lectures, I discussed what would happen if you did not have access to this technology. How would you equalise the playing field? How could you create trust when the other party holds the records? One way to address this concern is to use distributed ledger technology, such as blockchain, to record information. Distributed ledger technology provides a secure, tamper-proof method of recording transactions. Each transaction is verified by multiple parties and added to a chain of blocks, creating a permanent and immutable record of the transaction. Each party can see the details of the transaction, including the amount, date, and time. This ensures that both parties have access to the same information, reducing the likelihood of disputes. Transactions are verified by multiple parties, creating a trustless system where no single party has control over the network. This reduces the risk of fraud or corruption and provides greater control to individuals engaging in transactions.

In this final lecture, I discuss the issues that can arise when we trust all these technologies.

Making inferences from information

Businesses (and economists) cannot see what you think, only what you do. For example, if I offer you an apple or a banana at the same price, and you pick the banana what do I infer? Presumably that you like bananas more than apples. But what are you reacting to? Suppose you prefer:

- apples to bananas
- organic to regular and
- ripe to green.

While I think I am asking you to choose between a banana and an apple, you are actually choosing between ripe organic bananas and regular green apples. What dimension is more important to you? What happens if there are more dimensions (how and where it was grown, its sugar content, nutritional value, and shelf life)? So, business decisions are based on information on actions that regardless of the technology used, may not reveal information about our actual preferences.

Unfortunately, besides the fact that it is exceedingly difficult to get at information about preferences from the knowledge of our actions, we are ourselves unbelievably bad at processing more than a few dimensions. Except for visual patterns, the human brain isn't particularly good at processing huge amounts of information. We are only able to juggle about half a dozen distinct pieces of information at the same time—we can't even compare three characteristics of three distinct products. Early attempts at manipulating us stemmed from broad crude approximations of our behaviour.

Does tech help us make better inferences?

Tech allows a personalised business approach based on your specific preferences. It allows businesses to run specific experiments to test what people will respond to. For example, a business can run what is called A/B testing or split testing. This is a process used to compare two different versions of a product or website to determine which performs better. Companies first define a specific objective (for example, increasing conversions, improving user engagement, or increasing sales). The next step is to identify the variables to test. This could include anything from the colour of a button to the placement of a form. The test usually focuses on changing just one variable at a time to accurately measure its impact. The business then creates two versions of the product or website to test, with both versions differing only in the variable being tested. For example, if testing the colour of a button, one version would have a red button and the other version would have a blue button. Users are then randomly assigned to one of the two versions, either by using software tools or by simply splitting the traffic evenly between the two versions. The company then measures outcomes such as conversion rates, click-through rates, or time spent on the page to determine which version performed better. This can be done using statistical methods to ensure that the results are statistically significant.

The most famous example of a company using technology to make better inferences was Cambridge Analytica. Cambridge Analytica argued that traditional political campaign strategies, such as television advertising and door-to-door canvassing, were not effective at reaching certain groups of voters. They claimed that they had developed a proprietary algorithm that could analyse data from a variety of sources, including social media, surveys, and consumer data, to build detailed profiles of individual voters. This algorithm could then be used to identify patterns and correlations in the data that could be used to predict how individual voters were likely to vote. Cambridge Analytica claimed that they could identify the most persuadable voters and deliver targeted messages to them through social media and other digital platforms. They claimed that they had successfully used this approach to help political campaigns in the US, UK, and other countries win elections. They cited examples of campaigns that had used their services to target specific groups of voters with messages that were tailored to their interests and beliefs, resulting in increased voter turnout and support for their candidate. However, the company's methods and ethics came under scrutiny following the 2016 US presidential election, leading to a scandal and the eventual closure of the company.

How good are these inferences really?

Let's take a simple example of a business model that claims to use tech to predict your preferences – robo-advising. Robo-advising is a type of financial advisory service that uses algorithms and automation to provide investment advice to clients. The first step in the robo-advising process is for the clients to provide information about their investment goals, risk tolerance, and other financial information through an online questionnaire or interview process. The robo-advisor algorithm constructs a portfolio that is diversified across different asset classes and tailored to the client's investment goals and risk tolerance. The robo-advisor algorithm goes on to select specific investments for the portfolio, such as stocks, bonds, or exchange-traded funds (ETFs), based on factors such as performance, cost, and risk. The algorithm continuously monitors the portfolio and periodically rebalances it to maintain the desired asset allocation. For example, if a particular asset class has performed well and has increased in value, the robo-advisor may sell some of the holdings in that asset class and use the proceeds to purchase holdings in another asset class to maintain the desired allocation.

The innovation here is that without human intervention, the fees charged by the robo-advisor are tiny compared to that of a human advisor. So, it is crucial that the robo-advisor does an excellent job of measuring your risk. To determine your risk appetite, robo-advisors often use risk questionnaires. Clients are asked a series of questions about their investment goals, time horizon, and risk preferences, and the robo-advisor algorithm uses the responses to determine an appropriate asset allocation for the client's portfolio. Some robo-advisors also use behavioural finance analysis. This involves looking at factors such as the client's age, income, occupation, and past investment experience to determine their risk appetite. A third method is to analyse the client's investment history. The robo-advisor may look at the client's past investment performance, the types of investments they have made, and the level of risk they have taken in the past to determine their risk tolerance. However, in the end, most of the suggestions are pretty random. From personal experience, I have yet to find a robo-advisor that precisely measures my own risk tolerance.

But perhaps this is because we ourselves do not know what level of risk we are willing to accept. For example, most of us may never have been through a financial crisis. Therefore, when we are asked what level of risk, we're willing to accept when faced with a market downturn, our answers will not necessarily have *anything* to do with how we would actually behave during a crisis. If we ourselves don't know how much risk we are willing to accept, how will any algorithm be able to determine that level of risk?

Perhaps tech will do better in helping us stave off discrimination? Discrimination has been shown to exist in a wide variety of situations. For example, Bertrand and Mullainathan (2004) sent fictitious resumes to help-wanted ads in Boston and Chicago newspapers. To manipulate the perceived race, they randomly assigned either African American or White-sounding names to the resumes. It turns out that white names received 50 percent more callbacks for interviews. Callbacks were also more responsive to resume quality for white names than for African American ones. This racial gap was uniform across occupation, industry, and employer size. Similarly, Edelman et al (2017) ran an experiment on Airbnb, where they applied for accommodation using either African American or white names. They found that applications from guests with distinctively African American names were 16% less likely to be accepted relative to identical guests with distinctively white names.

By using digital platforms, prospective borrowers can upload their financial information, customise loan criteria, and search for interest rates, *without* showing their faces. The software processes the data and approves the loan if the numbers check out. Examples of firms that have adopted these types of platforms include Quicken Loans, Bank of America, and fintech startups like Roostify and Blend. These financial institutions are using artificial intelligence (AI) algorithms to assess the creditworthiness of loan applicants. While this can reduce lending disparities by allowing banks to process vast amounts of data about applicants' finances, spending habits, and preferences, it could also introduce more bias. Many factors that appear neutral, such as where someone shops or takes vacations, could double for race, which goes against the Fair Housing Act of 1968. Worse, hidden relationships across credit interactions can inject biases across a number of different areas. For example, if a person is charged more for an auto loan, they could also be charged more for a mortgage.

So, what is the problem?

Observed behaviour is not the same as preferences. And basing ads on what we think are real preferences can have important real-world consequences. As an example, pharmaceutical companies use Facebook to target and advertise to people with certain medical conditions. Lecher (2021) used a tool called Citizen Browser to collect data on Facebook advertisements and found that pharmaceutical companies use specific language and images to target users with medical conditions such as diabetes, arthritis, and cancer. Obviously, this type of targeted advertising can be harmful, as it can lead to users receiving biased information and potentially inappropriate treatments.

Worse, it is not clear who has your data. Flo, a menstrual cycle tracking app, that was used by over 100 million women to keep track of their menstrual cycles, claimed that it would never divulge details of the cycles, pregnancy, symptoms notes and other information that is entered by its users to anyone. According to the New York Times, it did (Gupta and Singer, 2021). The Federal Trade Commission filed a complaint in January 2021, alleging that from 2016 to 2019, Flo passed on intimate health details of its users to marketing and analytics companies like Facebook and Google.

It is also not clear how your data is being used. Hsu (2019) discusses the rise of product placement in streaming shows and movies. As traditional ad-supported television declines, streaming services like Netflix, Hulu, and Amazon are apparently using subtle and sophisticated methods of integrating brands into programming. Advertisers are drawn to the fact that product placement can be more targeted, measurable, and flexible than traditional ads.

Algorithms can also ensnare people who have not interacted with the systems. For example, smartphones are expensive in India. To buy a new smartphone, retailers depend on collateralising the smartphone. For example, as a first-time borrower, you can take high-interest payment plans financed by a loan company to buy a smartphone - but only after you install an undeletable app at the point of sale. The apps can then monitor repayment behaviour throughout the duration of the loan. Some of these apps not only require extensive access to personal data on the smartphone, but they can also remotely control the device. This type of app can allow lenders to harass users or even shut down their phones, causing significant disruptions to their daily life. What is the problem here? The coercive repayment tactics built into their devices ensnare people buying second-hand phones where the original buyer sells the phone on without notifying the buyer of the pending loan.

Why does tech make mistakes in gauging your preferences?

Depending on algorithms to make decisions has several problems. Consider, for example, the detection of Covid-19. Heaven (2021) argues that despite the hundreds of AI tools built to catch covid, none helped in real life. Why not?

One major reason was linked to the poor quality of the data that researchers used to develop their tools. Information about Covid patients, including medical scans, was collected and shared in the middle of a global pandemic, often by the doctors struggling to treat those patients. Many tools were built using mislabeled data or data from unknown sources, which meant that some tools end up being tested on the same data they were trained on, making them appear more accurate than they are. In addition, some AIs were found to be picking up on irrelevant features such as text font and position of patients in scans, which resulted in the incorrect prediction of serious Covid risk. Incorporation bias, or bias introduced at the point a data set is labeled, is also a problem as many medical scans were labeled according to whether the radiologists who created them said they showed Covid, incorporating their biases into the ground truth of a data set.

A second reason is based on Goodhart's law - once a useful number becomes a measure of success, it ceases to be a useful number. There have been lots of examples of this in history. Textile factories were required to produce quantities of fabric that were specified by length, and so looms were adjusted to make long, narrow strips. Uzbek cotton pickers were judged on the weight of their harvest. So, they would soak their cotton in water to make it heavier. When America's first transcontinental railroad was built, in the eighteen-sixties, companies were paid per mile of track. So, a section outside Omaha, Nebraska, was laid down in a wide arc, rather than a straight line, adding several unnecessary (but profitable) miles to the rails. In the NHS 2005 reform, doctors would be given a financial incentive to see patients within forty-eight hours. This often led to GPs focusing on meeting the 48-hour target at the expense of providing high-quality care,

which resulted in misdiagnosis and poor clinical outcomes for some patients. The initiative also created a culture of "box-ticking" in primary care, where GPs were incentivised to meet targets rather than providing personalised care to patients. If the GP could not see you in 48 hours, you would not be given an appointment.

When applied to tech, this meant that programmers were faced with the problem of how to communicate an objective to an algorithm when the only language they had in common is numbers. Consider for example, commonly used algorithms to recommend the release of incarcerated people awaiting trial. Angwin et al (2016) discuss how the use of predictive algorithms in the criminal justice system can perpetuate racial biases and lead to unfair treatment of defendants. These algorithms, used to assess the risk of a defendant committing a future crime, are based on factors such as criminal history, age, and employment status, but they often don't consider other important factors such as poverty and racial disparities in the criminal justice system. There are numerous examples of people being deemed high risk by the algorithms, resulting in harsher sentences, even though they had not committed any crimes in the past.

A third problem is that the data on us that is generated is not complete. We all generate data on ourselves everywhere we go: FICO credit scores for our credit history, our posts on Facebook, teachers are rated on RateMyProfessors.com, authors have Amazon scores, Airbnb hosts and guests are ranked on cleanliness, TaskRabbit, Deliveroo, and Uber drivers are all ranked (and they rank you), company health plans use Fitbit scores and so on.

The most ambitious use of this data is in China where the government announced a new Social Credit Law that aims to improve social trust and prevent dishonest behaviour by its citizens (Yang, 2022) The law is supposed to cover individuals and organisations and will assess them on a range of factors, including credit history, tax payment, social media behaviour, and environmental protection, among others. Those who score poorly on the assessment may face restrictions, including limited access to transportation, education, and job opportunities. The law has been criticised for its potential to violate citizens' privacy and limit freedom of expression. Unfortunately (or fortunately), as Yang (2022) notes, the attempts to gather this data are decidedly low-tech.

The next problem is privacy. How do we know that all this data collected on us will be kept secure? We don't. The Aadhaar database in India is a biometric identification system that assigns a unique 12-digit identification number to every resident of India. The program was launched by the Indian government in 2009 with the aim of providing a universal identity infrastructure to all residents, enabling access to various government services. The database includes personal information such as name, address, gender, date of birth, and biometric data such as fingerprints and iris scans. It is the world's largest biometric identification system, with over 1.25 billion enrolled members as of 2021. In 2017, the Indian government confirmed that a leak occurred in the Aadhaar database. The leak reportedly allowed access to the Aadhaar database by simply purchasing login details for as little as 500 rupees ($7).

The problem of interacting algorithms

The final problem I discuss today is the problem of interacting algorithms. Consider two different types of algorithms:

1. Credit-reporting algorithms: Allow access to private goods and services like cars, homes, and employment.
2. Algorithms adopted by government agencies: Affect access to public benefits like health care, unemployment, and child support services.

Who uses algorithm type 1? Consumer reporting agencies, including credit bureaus, tenant screening companies, or check verification services. They get their data from a wide range of sources such as public records, social media, web browsing, banking activity, app usage, and so on. They use their algorithms to assign people "worthiness" scores, which figure heavily into background checks performed by lenders, employers, landlords, and even schools.

Who uses algorithm type 2? Government agencies that want to modernise their systems. The problem here is that the software procurement process is rarely transparent and lacks accountability. Public agencies often buy automated decision-making tools directly from private vendors usually going for the lowest bidder. There is very little public vetting. Hence, when things go wrong (which happens quite a lot), the defendants and their lawyers have no way to prove their innocence, let alone establish why they were accused.

Similarly, credit-reporting algorithms can also interact with information algorithms that are used by media services such as Facebook, LinkedIn, Tiktok, or Twitter to direct your attention to particular news items. The interaction between credit-reporting algorithms and information algorithms can lead to the amplification of biased and discriminatory information, leading to significant negative consequences. These algorithms can easily target vulnerable individuals who may already have financial difficulties, making them more susceptible to misleading or fraudulent advertisements and information. Moreover, biased information can impact people's decision-making process, including financial decisions, and may also lead to a reinforcement of social inequalities.

Over this set of six lectures, I have attempted to discuss how tech has fundamentally changed how business works. We have discussed blockchains, decentralised finance, and distributed ledger technologies. We have discussed big data and AI. Today we discussed the problems we face in applying tech blindly to our lives.

Next year, I will turn to a completely different topic – a discussion of the big ideas in finance. Most of these ideas won Nobel Prizes. I will discuss why these ideas are important and why they were seminal ideas in finance. I hope you enjoyed these seminars and see you next year!

References and Further Reading

Angwin, Julia, Jeff Larson, Surya Mattu and Lauren Kirchner, 2016, Machine Bias, ProPublica, May 23, 2016 (available at https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing).

Bertrand, Marianne, and Sendhil Mullainathan, 2004, Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination, *American Economic Review* 94, 991-1013.

Christopher, Nilesh, 2021, Loans that hijack your phones are coming to India, *Rest of World*, 17 March 2021 (available at https://restofworld.org/2021/loans-that-hijack-your-phone-are-coming-to-india/)

Edelman, Benjamin, Michael Luca, and Dan Svirsky, 2017, Racial discrimination in the sharing economy: Evidence from a field experiment, *American Economic Journal: Applied Economics* 9, 1-22.

Gupta, Alisha Haridasani, and Natasha Singer, 2021, Your App Knows You Got Your Period. Guess Who It Told?, *New York Times*, Jan 28, 2021 (available at https://www.nytimes.com/2021/01/28/us/period-apps-health-technology-women-privacy.html).

Heaven, William Douglas, 2021, Hundreds of AI tools have been built to catch covid. None of them helped, *MIT Technology Review*, July 30, 2021 (available at https://www.technologyreview.com/2021/07/30/1030329/machine-learning-ai-failed-covid-hospital-diagnosis-pandemic).

Hsu, Tiffany, 2019, You see Pepsi, I see Coke: New Tricks for product placement, *New York Times*, Dec 20, 2019 (available at https://www.nytimes.com/2019/12/20/business/media/streaming-product-placement.html).

Lecher, Colin, 2021, How Big Pharma Finds Sick Users on Facebook, *Markup*, May 6, 2021 (available at https://themarkup.org/citizen-browser/2021/05/06/how-big-pharma-finds-sick-users-on-facebook).

Yang, Zeyi, 2022, China just announced a new social credit law. Here's what it means, *MIT Technology Review*, November 22, 2022 (available at https://www.technologyreview.com/2022/11/22/1063605/china-announced-a-new-social-credit-law-what-does-it-mean/).