Today’s Agenda

1. Case study discussions: Luttig, Tsai, Israni

2. Introduction to Algorithmic Decision-Making: Promise & Perils
INDIVIDUAL PROFILE

JOHN LUTTIG
How would you describe the ethical or social responsibilities of the tech sector as a whole and, specifically, of those individuals working in venture capital? Are they distinct from those of other sectors?

I don’t think the social, ethical responsibilities of tech are distinct from other sectors. Generally speaking, when you’re a venture capital firm that makes bets that others don’t, you’re the difference between the life and death of a start-up. Therefore, your involvement in a company can actually change the trajectory of technology. Consequently, I do think you have some sort of ethical responsibility for what your portfolio companies do.

Some people would say that ethics are quite subjective and that it is thus hard to mandate them in a formal sense. In our case though, there are certain companies we have passed on because they encouraged things we found morally questionable. We thought they would make money, and from a purely capitalistic standpoint we would have invested, but we didn’t because we had moral or ethical objections to their products.

For example, there was one company that made highly addictive products. We thought it was quite a good business with a really strong team. We considered investing and ultimately passed because of the ethical concerns we had about their addictive outputs. There was, however, no formal requirement for us to pass on them. Contrastingly, some funds have concrete vice clauses dictating that they not invest in alcohol, tobacco, drug or sex related companies. Our firm is relatively lax on that front. However, when we find something morally objectionable, we take that into consideration. For context, other venture capital firms also passed on the aforementioned company selling addictive products. Some passed on it purely for technical reasons, as investing in it would have violated their vice clause. Others passed on it for moral reasons: they felt an objection to investing in the company on ethical grounds.
What ethical standard(s) should a person use in deciding whether to fund a new company or create a new company/product?

- Is the standard subjective or objective?
- Is it a matter of personal or professional ethics?
- Or neither? A matter of political ethics or externally imposed rules?
- Who gets to decide?
INDIVIDUAL PROFILE

HENRY TSAI
What do you think the role of technologists and technology in addressing social and civic problems is?

Technologists certainly should play a role, both from an opportunity and a responsibility standpoint. More specifically, when we think about the present moment, there is an ecosystem of big companies, startups, and investors, and each should actively participate. If we step back, the way economists define technology fits in interestingly here. Economists define technology as something that makes you more productive and that helps push out your efficiency frontier. Therefore, creating technology is, by definition, tied to social progress – so there whether we like it or not technologists are going to be participants in social issues. But technologists should not think of themselves as the saviors to all the world’s problems. At the end of the day, technology is just a tool. Problems are never going be entirely solved by technology alone.
Questions about Tsai

Technology as increasing social productivity, “pushing out the efficiency frontier.”

- Why is efficiency a good thing?
- Is efficiency always to be preferred over inefficiency?
INDIVIDUAL PROFILE

ELLORA ISRANI
Another concern was that I worked on products around elections. We always understood that elections were going to be this sensitive space because there was so much potential for things to go wrong. I have a very distinct memory, it must have been the end of 2015 or the beginning in 2016 when the civic engagement team came together. We sat down and made a clear list of the values that were important to us and that we thought should drive product decision-making. The idea was that as we were building these products for elections, we were obviously going to have opportunities to build a product in a way it was more or less fair, and more less accessible or equitable. I think in other products dimensions, the risk of saying let’s give this product to fifty-percent of users and see how metrics are affected – this A/B testing happens all the time – is not significant. But with elections, we said if we’re going to build a registration tool, everyone has to have access to it; we don’t want to end up in a position where it sways people one way or the other.

Similarly, we had a lot of conversations around to what level should we raise our own consciousness? What digging should we do to make sure that we’re not making things unfair? Our core data science team did a lot of digging into
Questions about Israni

1. If FB builds election products, how should FB’s voter registration tool value efficiency vs equity/fairness?

2. If FB discovers that the tool increases voter registration disproportionately for Republicans (or for Democrats), what should happen? Who should decide?
Did your Stanford education prepare you to think through the social, ethical impacts or your professional work in tech?

In some ways yes and in some ways no. It was always in the background of things. At least when I was a computer science major, we had one required ethics class. Maybe there were one or two options, but the vast majority of people took this one class about the ethics of computer science. I think it was seen as a requirement that everyone had to get through and I don’t know that anyone paid it serious attention. So, it was always in the background of things. Nonetheless, we definitely could have and probably should have paid more attention to it. I think all that is to say that while ethics was in the background of my Stanford education, looking back on it, it should have been more in the foreground and it should have been more explicitly thought about and considered. I’m not trying to levy a criticism of the Stanford curriculum. I think it is incumbent on all of us – the University, the departments, the professors, and all students – to think about these things.
Today’s Agenda

1. Case study discussions: Luttig, Tsai, Israni

1. Introduction to Algorithmic Decision-Making: Promise & Perils
What is an algorithm?

*Answer*: a list of rules to follow in order to solve a problem

A *computer algorithm*: software code that produces an output from given inputs.

Obvious benefit: with big data, computer algorithms can more accurately and more efficiently perform tasks than humans can.
If computer algorithms can more accurately and more efficiently outperform human at certain kinds of problem-solving, then algorithmic decisionmaking has enormous potential for good.

O’Neil:

“This is the Big Data economy, and it promised spectacular gains. A computer program could speed through thousands of résumés or loan applications in a second or two and sort them into neat lists, with the most promising candidates on top. This not only saved time but was also marketed as fair and objective. After all, it didn’t involve prejudiced humans digging through reams of paper, just machines processing cold numbers.”
So what’s the connection between algorithms and values like fairness or transparency?

If an algorithm is a mathematical function to translate inputs into a determinative output, what do values like fairness have to do with math or logic?

What does fairness have to do with $2 + 2 = 4$?

Let’s take an example of some immediate relevance to you...
Assignments and Grading Breakdown

In addition, the course includes four assignments and a final exam. You will receive more information about each of the assignments well in advance of their due dates.

- Technical assignment on algorithmic decision-making – **due January 25**
- Philosophy paper on privacy – **due February 8** [NOTE: WIM students will have an additional revision due later]
- Group policy assignment on autonomous vehicles – **due February 27**
- Technical assignment on platforms and social networks – **due March 15**
- Final exam — **held on March 20**

Grades will be calculated as follows:

<table>
<thead>
<tr>
<th>Non-WIM Students</th>
<th>WIM Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participation – 20%</td>
<td>Participation – 15%</td>
</tr>
<tr>
<td>Technical Assignment 1</td>
<td>Technical Assignment 1 – 10%</td>
</tr>
<tr>
<td>Philosophy Paper – 20%</td>
<td>Philosophy Paper (including revision) – 30%</td>
</tr>
<tr>
<td>Policy Assignment – 20%</td>
<td>Policy Assignment – 20%</td>
</tr>
<tr>
<td>Technical Assignment 2</td>
<td>Technical Assignment 2 – 10%</td>
</tr>
<tr>
<td>Final Exam – 20%</td>
<td>Final Exam – 15%</td>
</tr>
</tbody>
</table>
Possible CS 181 Grade Schemes

• Grading on curve: pre-determined number of As. Grades are determined on the basis of performance in class relative to other students in class.

• Grading according to an external and objective standard. Everyone can earn an A, in principle.

• Grading relative to your own past performance: that you show effort and improve over the quarter factors into your grade.
Still More CS 181 Grade Schemes

• Everyone gets an A, assuming you show up and submit work. No relationship to actual quality of work.

• Grades are a function of hours worked on class assignments (monitored by Stanford IT and Canvas). More work = higher grade. Person who works the most gets the highest grade.

• People who come visit my office hours get higher grades.

• Athletes get higher grades (because I love sports)
• Athletes get lower grades (because I hate sports)
What Values are at Stake in the CS 181 Grading Algorithm?

1. **Fairness**: does the grading scheme treat people fairly?

   NOTE: multiple conceptions of fairness

2. **Transparency**: is the grading scheme made transparent? (or is it a black box?)

3. **Privacy**: are grades made public?

4. **Due Process**: is there a fair process for appeal?
Promise and Peril of Algorithms

Algorithms used widely in society.

- Banking/Credit Rating/Scoring
- Hiring (resume scans)
- Medicine and health care (triage; skin cancer)
- Dating, suitability as lovers
- Education (online learning, what to learn next)
- Recommendation systems (movies, books, news)
- Criminal justice (recidivism, criminal risk assessment, sentencing recommendations, bail)
Promise and Peril of Algorithms

Algorithms used widely.

**Promise**: improve accuracy and efficiency over human decisionmaking. Humans make mistakes, and they’re slow.

Diminish or eliminate bias, randomness, human fallibility.

**Peril**: widespread adoption encodes bias, threatens fairness, privacy, transparency, and due process.
[Jennifer is] ranked 1,396 out of 179,827 high school students in Iowa. . . . Jennifer’s score is the result of comparing her test results, her class rank, her school’s relative academic strength, and a number of other factors. . . .

Can this be compared against all the other students in the country, and maybe even the world? . . . That’s the idea . . . . That sounds very helpful. . . . And would eliminate a lot of doubt and stress out there.

—Dave Eggers, The Circle
Eric Loomis
They are often under pressure to make decisions quickly and intuitively.

Judge Jerome Frank once said, “justice is what the judge ate for breakfast.” In fact, a study in Israel found that judges were twice as likely to grant a defendant parole at the beginning of a court session, right after they would have eaten a meal or a snack, than at the end of a session, when they were presumably hungry and tired. Other research has shown how race, gender, and age can all contribute to disparities in sentencing -- not to mention the vast disparities between the treatment of defendants from one jurisdiction to another.

So, given that judges left to their own devices can make biased (or at the very least arbitrary) decisions, and criminal-risk-scoring algorithms apparently have their own problems, how do they compare to each other? Do the scores help the judges make decisions in court that are more, or less, biased?
Algorithms:  
Six Sources of Ethical Concern

1. Sample size bias
2. Bias in the data
Man is to Computer Programmer as Woman is to Homemaker?  
Debiasing Word Embeddings

Tolga Bolukbasi\textsuperscript{1}, Kai-Wei Chang\textsuperscript{2}, James Zou\textsuperscript{2}, Venkatesh Saligrama\textsuperscript{1,2}, Adam Kalai\textsuperscript{2}

\textsuperscript{1}Boston University, 8 Saint Mary’s Street, Boston, MA
\textsuperscript{2}Microsoft Research New England, 1 Memorial Drive, Cambridge, MA
tolgab@bu.edu, kw@kwchang.net, jamesyzou@gmail.com, srv@bu.edu, adam.kalai@microsoft.com

Abstract

The blind application of machine learning runs the risk of amplifying biases present in data. Such a danger is facing us with \textit{word embedding}, a popular framework to represent text data as vectors which has been used in many machine learning and natural language processing tasks. We show that even word embeddings trained on Google News articles exhibit female/male gender stereotypes to a disturbing extent. This raises concerns because their widespread use, as we describe, often tends to amplify these biases. Geometrically, gender bias is first shown to be captured by a direction in the word embedding. Second, gender neutral words are shown to be linearly separable from gender definition words in the word embedding. Using these properties, we provide a methodology for modifying an embedding to remove
For those who are not directly involved in the technical development of algorithms for large scale data systems, the end product of such as system can feel like a “black box”—an opaque machine that takes inputs, carries out some inscrutable process, and delivers unexplained outputs based on that process. The technical processes involved in algorithmic systems are typically unknown to a consumer, potential student, job candidate, defendant, or the public as they are often treated as confidential or proprietary to the entities that use them. Some systems even “passively” pre-screen potential candidates without notice as a preemptive effort to streamline decision-making processes at a later date. This lack of transparency means that
Algorithms:
Six Sources of Ethical Concern

1. Sample size bias
2. Bias in the data
3. Privacy
4. Lack of transparency/explainability
5. Lack of due process
6. What does Fairness Mean? What version of fairness to encode in an algorithm?