Today’s Agenda

1. Models for Human-AI interaction
   • Human in the loop
   • Human on the loop
   • Human out of the loop
2. Human relationship with AI
   • Do we have a right to know we’re interacting with AI
3. Bringing society in the loop – the role of government
4. Approaches to addressing economic consequences of AI
5. Value Pluralism
Autonomous Vehicles

Google Self-Driving Car in Mountain View, August 2015
Advances in Vehicle Automation

- Level 0: No automation
- Level 1: Driver assistance, such as automated forward (e.g., cruise control) or lateral (e.g., lane keeping) control
- Level 2: Partial automation, such as directional and speed control, but driver must remain engaged (Tesla Autopilot)
- Level 3: Conditional automation -- sustained autonomous driving, but driver must intervene as system requires
  - How does that work out? Video
- Level 4: Fully autonomous driving in certain conditions), such as well-mapped areas at daytime (Waymo cars in some locales)
- Level 5: Fully autonomous driving in any conditions
Human IN the Loop

- AI provides recommendations or suggests actions to human
- Human decides on action that impacts environment
  - Level 1 Autonomous vehicles (e.g., lane drift warnings to drivers)
  - Decision to grant bail (e.g., COMPAS)
    - Algorithm provides risk score. Judge decides.
  - Disease diagnosis (e.g., MYCIN)
    - Algorithm suggests diagnosis. Doctor makes final judgement.
  - Autonomous weapons (scenario 1)
    - AI selects target. Human makes decision to fire/kill.
  - Home mortgages
    - Algorithm computes load default risk. Human makes final decision.

- Are there general principles for situations that should require a human in the loop?
Human ON the Loop

- AI allowed to make decisions that impact the environment
- Human can supervise/override decisions of AI
  - Level 3-4 Autonomous vehicles
    - AI drives. Human intervenes in exceptional situations.
    - Uber collision kills woman – person in drivers seat did not react
    - Tesla on Autopilot collides with truck – driver was watching a movie
  - Airplane autopilot
    - Algorithm flies/lands plan. Human can disengage or override.
  - Autonomous weapons (scenario 2)
    - AI can target and use lethal force. Human may override.
“On October 19, a Waymo Pacifica struck and injured a motorcyclist in California. As is often the case, the collision was caused by a human - in this instance, the safety driver in the Waymo vehicle. In an unusual twist, however, Waymo CEO John Krafcik revealed that if the safety operator had not taken control of the autonomous minivan, then the self-driving software would have avoided a collision.”

-- David Silver, Forbes, Nov. 7, 2018
“In documents made public by the safety board, Asiana acknowledged the likely cause of the accident was the crew's failure to monitor and maintain the plane's airspeed, and its failure to abort the landing when in trouble. The South Korea-based airline said the pilot and co-pilot reasonably believed the automatic throttle would keep the plane flying fast enough to land safely, when in fact the auto throttle was effectively shut off after the pilot idled it to correct an unexplained climb earlier in the landing.”

Human OUT of the Loop

- AI makes decisions that impact the environment
- Human does not (directly) impact AI’s decision-making
- Human is really just part of the environment
  - Advertising and recommendation systems
  - AI game players
  - Level 5 Autonomous vehicles
    - AI drives. Human cannot intervene (e.g., no steering wheel).
    - Not just accident reduction, but potential fuel savings, less resource consumption, and less need for transportation infrastructure
  - Autonomous weapons (scenario 3)
    - AI targets and use lethal force. Human has no control.
Today’s Agenda

1. Models for Human-AI interaction
   • Human in the loop
   • Human on the loop
   • Human out of the loop
2. Human relationship with AI
   • Do we have a right to know we’re interacting with AI
3. Bringing society in the loop – the role of government
4. Approaches to addressing economic consequences of AI
5. Value Pluralism
Knowing When Interacting With AI

- Google Duplex demo

- Recall the Turing Test
  - We become the *interrogated* rather than the interrogator
  - We move from being the subject to the *object*

- Does Google Assistant need to announce itself as an AI?
Knowing When Interacting With AI

• Do we have a right to know when we are interacting with AI?

• Do we have a right to...
  • ...know when a plane is on autopilot?
  • ...book a flight that is guaranteed not to use autopilot?
  • ...choose a non-autonomous rideshare?
    • What if AV rideshares are safer on average than human drivers?
  • ...request a medical diagnosis from a human doctor?
    • What if the AI is more accurate on average?
Today’s Agenda

1. Models for Human-AI interaction
   • Human in the loop
   • Human on the loop
   • Human out of the loop
2. Human relationship with AI
   • Do we have a right to know we’re interacting with AI
3. Bringing society in the loop – the role of government
4. Approaches to addressing economic consequences of AI
5. Value Pluralism
Who is Responsible?

On Wednesday, we asked you who was responsible for addressing the job displacement caused by automation. Should it be government or the companies causing the displacement? Why?
“Governments should intervene, at a minimum, *when private action has negative public consequences*; when shortsighted actions threaten to cause long-term harm; when failure to intervene undermines significant constitutional values and important individual rights; when a form of life emerges that may threaten values we believe to be fundamental; and when we can see that failing to intervene on the side of right will simply strengthen the interventions on the side of wrong.”

-- Larry Lessig, *Code 2.0*
Society in the Loop

- Society must resolve the tradeoffs between the different values that AI systems can strive towards
- Society must agree on which stakeholders would reap which benefits and pay which costs
Forging the Social Contract

- Evolutionary pressures favored the emergence of complex social institutions to coordinate people’s behaviors
- Centralized government in the form of a “Leviathan”
- Evolution of the social contract to ensure sovereign power implements the general will
What does it mean to govern new technologies? It requires:

- Measuring or anticipating their impacts
- Identifying benefits, costs, and trade-offs
- Surfacing conflicting preferences, values, goals
- Understanding how interests are organized
- Aggregating preferences through some legitimate decision-making process
- Articulating policies, rules, standards, and/or norms
- Enforcing policies, rules, standards, and/or norms
- Adapting policies, rules, standards, and/or norms
One Result: A Risk-Based Approach

- **Unacceptable risk**
  - e.g. social scoring
  - Prohibited

- **High risk**
  - e.g. recruitment, medical devices
  - *Not mutually exclusive*
  - AI with specific transparency obligations
    - ‘Impersonation’ (bots)
  - Permitted subject to compliance with AI requirements and ex-ante conformity assessment

- **Minimal or no risk**
  - Permitted with no restrictions
Today’s Agenda

1. Models for Human-AI interaction
   • Human in the loop
   • Human on the loop
   • Human out of the loop
2. Human relationship with AI
   • Do we have a right to know we’re interacting with AI
3. Bringing society in the loop – the role of government
4. Approaches to addressing economic consequences of AI
5. Value Pluralism
How Big a Threat?

JOBS THREATENED BY AUTOMATION
PROJECTED PERCENTAGE OF US JOBS AT HIGH RISK OF AUTOMATION BY 2030

- FREY & OSBORNE (OXFORD): 47%
- PRICewaterHOUSE COOPERS (PWC): 38%
- ORGANIZATION FOR ECONOMIC COOPERATION & DEVELOPMENT (OECD): 9%

Source: PwC, Frey and Osborne, OECD
©2017 Grayline Group
The Critical Role of Policy

Rather than regulating AI per se, it may make more sense to explore the role of policies in mitigating the costs of adjustment.

For example, addressing the potential of AI and automation to increase inequality is one valid societal concern.
What Are Our Obligations?

When people experience technological unemployment, what are our societal responsibilities, if any? What should be done?

One answer: people should be guaranteed a right to a job.
India’s Approach

“The Mahatma Gandhi National Rural Employment Guarantee Act aims at enhancing the livelihood security of people in rural areas by guaranteeing hundred days of wage-employment in a financial year to a rural household whose adult members volunteer to do unskilled manual work.”

- One of the world’s largest social welfare schemes
- Covers over 800 million people
What Are Our Obligations?

When people experience technological unemployment, what are our societal responsibilities, if any? What should be done?

One answer: people should be guaranteed a right to a job.

An alternative: As a society, we want a government that cushions people against adverse economic shocks (unemployment, disability, trade, etc.) through policies and programs.
The Great Depression & CCC

[Image of two men holding signs reading "WANTED: A DECENT JOB"
by a decent man age 37 - family war veteran paying on home college trained native Chicagoan.

[Image of a newspaper clipping with headlines:
"ROOSEVELT ASKS CONGRESS TO MAKE WORK FOR 250,000; FARM BILL FOES UNAPPEASED"
" quick job action sought"
"President asks power to begin recruiting idle in 2 weeks.

$200,000,000 IS AVAILABLE"

"Points of First Roosevelt bill aimed at unemployment relief"
"critics of both parties assail farm bill, but will vote for it."
"Passage is set for today"
"Snell rails at drastic rule, barring changes and limiting debate."

"House backs president"
"Weather - fair and slightly colder today; tomorrow fair"
In the event that robots and computers are capable of doing many human jobs, would you support or oppose the following policies?

- Machines are limited to doing dangerous or unhealthy jobs
- Government offers all Americans a guaranteed income that would meet their basic needs
- Government creates a national service program that would pay people to perform tasks
From Stanford to Stockton

Basic Income In Cities

A guide to city experiments and pilot projects
Nearly 80% of you support a universal basic income as a response to automation.

- On what grounds do you support a UBI?
- At what level should a UBI be set? Should it be sufficient to keep people above the poverty line, or simply to complement other income sources?
- Why is UBI preferable to alternative policy responses?
Three Rationales

1. A UBI would provide an income sufficient to help people in the event they are not employed

   Depends on how much it is!

2. A UBI would provide essential resources to enable people to make investments (e.g. savings, education, etc.) that help them to weather economic shocks

   Depends on how people spend the UBI and use their time!

3. A UBI would be a less paternalistic way of delivering assistance

   How much do we weigh autonomy over efficacy?
Who Benefits from UBI?

• Our current tax and transfer system is largely targeted towards those in the lower half of the income distribution—which means that it works to reduce poverty and income inequality.

• Shifting to a universal cash grant would mean that less of the system would be targeted toward those at the bottom, leading to an increase in income inequality.

• Unless we are willing to raise taxes substantially, it would be difficult to provide a common amount to everybody while also making sure the amount is sufficient to cover the needs of the poorest households.

(Furman 2016)
What Does UBI Cost?

Domestic Spending Compared with UBI

- Social Security: $2.6 Trillion
  - Veteran's Benefits
  - Income Security
  - Other Health
  - Medicaid
  - Medicare

- Other Domestic Programs: $540.3 Billion
  - Other
  - Federal Retirement
  - Justice
  - Transportation
  - Education

- Universal Basic Income: $2.4 Trillion
  - $2 Trillion
  - $1 Trillion

Does UBI Work?

• Despite its long history, there is relatively little evidence on the impacts of a guaranteed income: on how much people work, on their economic wellbeing, etc.

• Manitoba (1974-79): guaranteed income to 10k+ residents of Dauphin; 20 years later results suggested hugely beneficial effects on health and wellbeing, no negative employment effects (but no control group!)

• Give Directly (present): unconditional cash transfers to rural villagers in Uganda and Kenya; early evidence suggests increases in consumption and psychological wellbeing, though results do not seem to be sustained over time
Preliminary Results from Finland

- Finland ran the world’s largest UBI experiment, providing $640 per month to 2000 randomly selected unemployed Finns.
- There were no significant effects on employment and earnings (either positive or negative).
- Access to the UBI grant had significant effects on psychological health and wellbeing.
Findings in Stockton

STOCKTON GUARANTEED INCOME PROGRAM

- Less anxious and depressed
- Improvements in overall well-being
- More easily handled unexpected expenses
- Obtained full-time employment at more than twice the rate of non-recipients

Source: University of Tennessee at Knoxville, University of Pennsylvania
A Moment of Experimentation
What Are the Alternatives?

- Means-tested social benefits (e.g. unemployment insurance, food stamps, etc.)
- Minimum wage
- Significant investments in education and retraining
- Earned income tax credit
- Public employment schemes
- Strengthening unions
Today’s Agenda

1. Models for Human-AI interaction
   • Human in the loop
   • Human on the loop
   • Human out of the loop
2. Human relationship with AI
   • Do we have a right to know we’re interacting with AI
3. Bringing society in the loop – the role of government
4. Approaches to addressing economic consequences of AI
5. Value Pluralism
WHAT TO DO IN THE FACE OF VALUE TENSIONS AND TRADE-OFFS?

Maximize/optimize an objective function?
But this runs into problems:
Not every problem has a technical solution!
And even when technical approaches are available, beware the problems of optimizing:
1. Optimizing is a means, not an end
2. Proxies can become the end
3. Successfully optimizing for one end can upset a broader social balance of multiple values
What to do about Value Tensions?

Algorithmic decision-making: predictive accuracy vs. fairness, privacy, explainability/auditability

Data privacy: privacy vs. national security, personal safety, innovation, convenience

AI and autonomous systems: productivity vs. material welfare and human agency
Utilitarianism

A general theory of what is valuable and of right action.

1. A Theory of the Good?

A Theory of what is valuable: happiness (and only happiness) is intrinsically good.

2. A Theory of the Right

The right action is that action, of all available alternatives, that produces the greatest net balance of happiness.
When confronting a value tension, decide by engaging in cost-benefit analysis.
Take the action that creates the greatest net benefits.
Domestic justice, not global justice
Society is a cooperative venture for mutual advantage
A theory of distributive justice
Provide an alternative to utilitarianism (which “fails to take seriously the distinction between persons”)
Rights as Constraints: Some things should not be part of moral calculus

An **interest**? (one consideration among many; can be defeated by other values or rights?)

A **right**? (strong presumption against infringement)

A **constitutional right**? (ordinary lawmaking can’t remove; sits above majoritarian decisionmaking)

An **inalienable constitutional right**? (ordinary lawmaking can’t remove and right holder can’t renounce – no notice and consent!)

A **fundamental human right**? (no government can remove/abridge)
The Trolley Problem
Man on the Bridge
Suppose that Jones has suffered an accident in the transmitter room of a television station. Electrical equipment has fallen on his arm, and we cannot rescue him without turning off the transmitter for fifteen minutes. A World Cup match [the Super Bowl] is in progress, watched by many people, and it will not be over for an hour. Jones’s injury will not get any worse if we wait, but his hand has been mashed and he is receiving extremely painful electrical shocks.
Thomas Scanlon’s
*What We Owe Each Other*

Should we rescue him now or wait until the match is over?

Does the right thing to do depend on how many people are watching — whether it is one hundred, one million or five million, or a hundred million?"

Implication: the numbers don’t matter. No moral calculus here.
The ideas of economists and political philosophers, both when they are right and when they are wrong, are more powerful than is commonly understood. Indeed the world is ruled by little else. Practical men, who believe themselves to be quite exempt from any intellectual influence, are usually the slaves of some defunct economist. Madmen in authority, who hear voices in the air, are distilling their frenzy from some academic scribbler of a few years back.

John Maynard Keynes.
At some point I realised that what all these views had in common was a Platonic ideal: in the first place, that, as in the sciences, all genuine questions must have one true answer and one only, all the rest being necessarily errors; in the second place, that there must be a dependable path toward the discovery of these truths; in the third place, that the true answers when found must necessarily be compatible with one another and form a single whole, for one truth cannot be incompatible with another. This kind of omniscience was the solution of the cosmic jigsaw puzzle...”

Isaiah Berlin (5-6)
The notion of the perfect whole, the ultimate solution, in which all good things co-exist, seems to me to be not merely unattainable – that is a truism – but conceptually incoherent; I do not know what is meant by a harmony of this kind. Some among the Great Goods cannot live together. That is a conceptual truth.

We are doomed to choose, and every choice may entail an irreparable loss. Happy are those who live under a discipline which they accept without question, who freely obey the orders of leaders, spiritual or temporal...I can only say that those who rest on such comfortable beds of dogma are victims of self-induced myopia, blinkers that may make for contentment, but not for understanding of what it is to be human. Berlin (13-14)
Value Pluralism

• There are **objective/universal values** (eg., suffering is bad; truth-telling is good)

• **Values are plural** (and internally complex – admit of cultural variation)

• **Values conflict** (cannot simultaneously maximize all values)

• Values may be **incommensurable** (unable to compare and rank order; no common metric)

• Value pluralism is true within individual lives, within single societies, and across societies
Implications of Value Pluralism

• The really terrible choices in life are between rival good things.
• Tragedy is not the inevitable unfolding of some evil.
• Tragedy is something different: the terrible choice between rival good things.
What could this mean?

OPTIMAL LIVING MICRO CLASS

HOW TO OPTIMIZE YOUR LIFE
Do I contradict myself?
Very well then I contradict myself
(I am large, I contain multitudes)
Berlin’s Conclusions?

1. Avoid suffering (17)
2. Embrace uncertainty and ambiguity; be willing to trade-off and compromise
3. Humility

A flat, uninspiring answer?

“The concrete situation is almost everything. There is no escape: we must decide as we decide; moral risk cannot, at times, be avoided. All we can ask for is that none of the relevant factors be ignored, that the purposes we seek to realize should be seen as elements in a total form of life, which can be enhanced or damaged by decisions.”

Isaiah Berlin (18-19)