



## CSCE 590-1: Trusted AI

### Lecture 2: First Look at Trusted Decisions

---

PROF. BIPLAV SRIVASTAVA, AI INSTITUTE

24<sup>TH</sup> AUG 2021

*Carolinian Creed: “I will practice personal and academic integrity.”*

# Organization of Lecture 2

---

- Introductory Segment
  - Recap from Lecture 1
- Main Segment
  - Trust in technology
  - About AI Ethics
  - Quality of decisions
  - Case Study: Towards trusted decisions during COVID-19
- Concluding Segment
  - A framework to think of course project
  - About next lecture – Lecture 3
  - Ask me anything

# Introductory Segment

---

# Recap of Lecture 1

---

- We looked at course expectations
- Introduced AI
  - Adaptive Software System
  - Intelligent Assistant
  - Data science is a subset; types of analysis – descriptive, predictive and prescriptive
- To complete discussion on the topics of Trust and Ethics

# Main Segment

---

# Trust – In Technology

---

# AI Results – Are They Representative ?

No



Breakfast items searched on Google



Employees



Marriage



Waiting Customers

Many perspectives  
missing by default,  
and the system  
does not even  
inform about it



Hospital Patients

# Example: Regional Perspectives Matters for Trust



*Breakfast items searched on Google*



*Waiting Customers*

**How does this impact ?**

- Training data
- Preferences and constraints
- Inductive bias (i.e., implicit defaults in the algorithms)
- Integration of solution to wider ecosystem

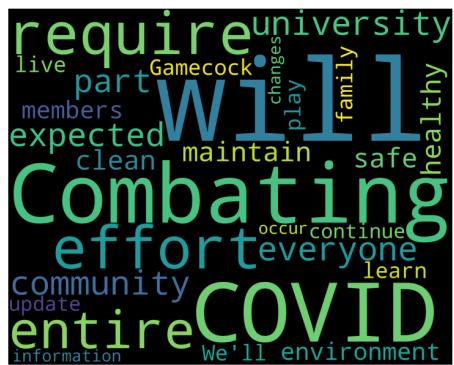
**An AI System that is Not Relevant to User Cannot Become Trustable**

# Insights from Text

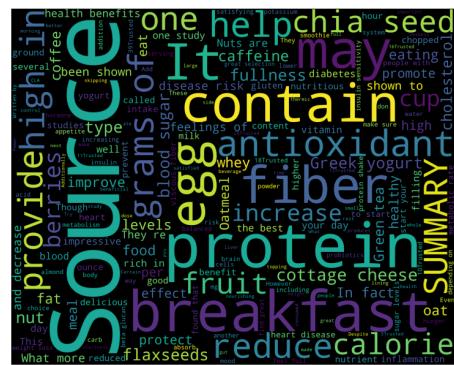
# Recap

<https://github.com/biplav-s/course-tai/blob/main/sample-code/l1-intro/Introduction%20Text%20Issues.ipynb>

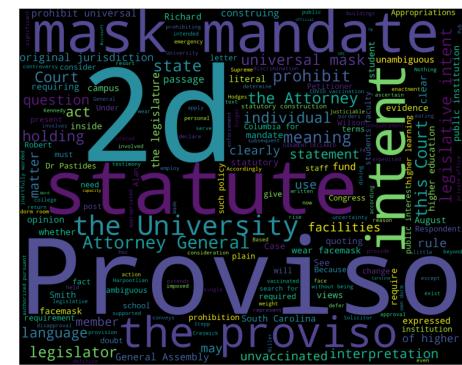
**Key Message:** AI useful but one needs to be cautious before generalizing



Covid from UoSC



## 12 Healthy breakfast dishes



## Mask ruling by SC Supreme Court

# What are the Components of Trust (Technology)

---

1. Competent – does what it is supposed to do
2. Reliable – including, well tested
3. Upholds human values, social good
  1. Fairly and ethically used
  2. Adequate data management & preserves privacy
4. Allows human-technology interaction
  1. Explainable, transparent
  2. How does the system give its result?

# Components of Trust - Illustration

1. Competent – does what it is supposed to do
2. Reliable – including, well tested
3. Upholds human values
  1. Fairly and ethically used
  2. Adequate data management & preserves privacy
4. Allows human-technology interaction
  1. Explainable, transparent
  2. How does the system give its result?

	Car – cruise control	Nuclear Energy
Competent	X	X
Reliable	X	X
Upholds human values	-	?
Allows human interaction	X	-

x: yes; -: not applicable; ?: questionable

# Components of Trust for AI

1. Competent – does what it is supposed to do
2. Reliable – including, well tested
3. Upholds human values
  1. Fairly and ethically used
  2. Adequate data management & preserves privacy
4. Allows human-technology interaction
  1. Explainable, transparent
  2. How does the system give its result?

	AI – Word Tag Cloud	AI – Image Search	AI – Self-driving Car	AI-powered Chatbot: Medical Guide
Competent	x	x	?	x
Reliable	x	?	?	?
Upholds human values	?	?	?	?
Allows human interaction	x	x	?	?

x: yes; -: not applicable; ?: questionable

# Group Discussion (5 mins):

## What are the Components of Trust for People

---

1. Examples
  1. Competence – do they get work done?
  2. Reliable – are they consistent?
  3. Uphold human values
  4. ...
2. Are the components for people different from those for technology?

# Review: AI During COVID-19 - A Crisis of Trust

---

- **Machine Learning algorithms** were not reliable to be deployed
  - Hundreds of AI tools have been built to catch covid. None of them helped, <https://www.technologyreview.com/2021/07/30/1030329/machine-learning-ai-failed-covid-hospital-diagnosis-pandemic/>, Aug 2021
  - Turing Institute report, [https://www.turing.ac.uk/sites/default/files/2021-06/data-science-and-ai-in-the-age-of-covid\\_full-report\\_2.pdf](https://www.turing.ac.uk/sites/default/files/2021-06/data-science-and-ai-in-the-age-of-covid_full-report_2.pdf)
  - Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal, BMJ 2020; 369 doi: <https://doi.org/10.1136/bmj.m1328> (Published 07 April 2020)
- **Chatbots** could not be used at scale
  - Biplav Srivastava, Did Chatbots Miss Their 'Apollo Moment'? A Survey of the Potential, Gaps and Lessons from Using Collaboration Assistants During COVID-19, [Cell PATTERNS](#), Aug 2021
- **Image processing** for COVID-19 detection
  - Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans, <https://www.nature.com/articles/s42256-021-00307-0> , March 2021

# AI Ethics

---

# Why is Ethics Even an Issue?

---

- When a technology works with humans and relates to inter-personal issues, the question of ethics comes into picture
- Examples: medicine (opioids), food (genetically modified)

# AI Ethics

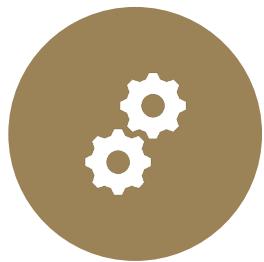
---



Multidisciplinary field of study



How to optimize AI's beneficial impact while reducing risks and adverse outcomes



How to design and build AI systems that are aware of the values and principles to be followed in the deployment scenarios



To identify, study, and propose technical and nontechnical solutions for ethics issues arising from the pervasive use of AI in life and society

**Credits:**

Tutorial on [Trusting AI by Testing and Rating Third Party Offerings at IJCAI 2020](#), Biplav Srivastava, Francesca Rossi, Jan 2021

# What is Specific to AI?

- AI needs **data**
  - Data privacy and governance
- AI is often a **black box**
  - Explainability and transparency
- AI can make **decisions/recommendations**
  - Fairness and value alignment
- AI is based on statistics and has always a small percentage of **error**
  - Who is accountable if mistakes happen?
- AI can infer our preferences and **manipulate them**
  - Human and moral agency
- AI is very **pervasive and dynamic**
  - Larger negative impacts for tech misuse
  - Fast transformation of jobs and society

## Credits:

Tutorial on [Trusting AI by Testing and Rating Third Party Offerings at IJCAI 2020](#), Biplav Srivastava, Francesca Rossi, Jan 2021

# Main AI Ethics Issues



DATA GOVERNANCE  
AND PRIVACY



FAIRNESS AND  
INCLUSION



HUMAN AND  
MORAL AGENCY



VALUE ALIGNMENT



ACCOUNTABILITY



TRANSPARENCY AND  
EXPLAINABILITY



TECHNOLOGY  
MISUSE

**Credits:**

Tutorial on [Trusting AI by Testing and Rating Third Party Offerings at IJCAI 2020](#), Biplav Srivastava, Francesca Rossi, Jan 2021

# Decisions – AI as a Decision-Support Tool

---

# The Quality of Everyday Decisions



Source: <https://www.umassd.edu/fycm/decision-making/process/>

Major variability due to:

- Emotions
- Biases
- Increasing data volume
- Cognitive ability to process
  - Decreases under stress and constraints
  - Decreases with age\*

\* **Source:** A Review of Decision-Making Processes: Weighing the Risks and Benefits of Aging, Mara Mather,  
<https://www.ncbi.nlm.nih.gov/books/NBK83778/>

# Evidence #1: Poor Medical Adherence

## Taking medicines

- 20 -30 % of medication prescriptions are never filled
  - ~50 % of medications for chronic disease are not taken as prescribed

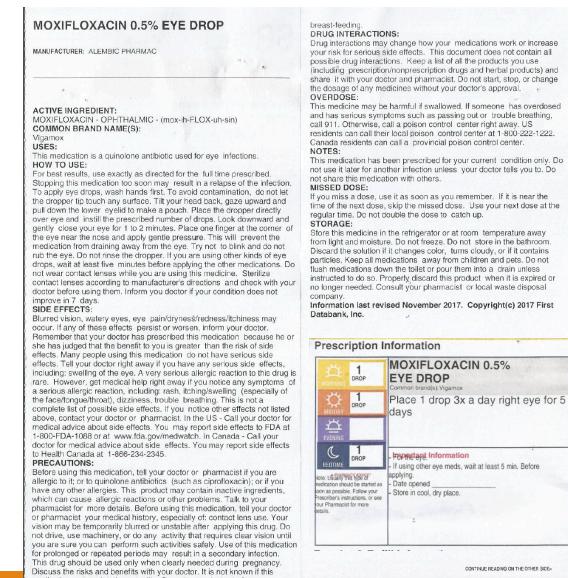
## Impact

- causes 125,000 deaths, at least 10 percent of hospitalizations
  - Costs the American health care system between \$100 billion and \$289 billion a year.

Finding relevant guidance is hard,  
one reason for non-adherence and high  
costs in health

## Sources:

- Medication Nonadherence, A Diagnosable and Treatable Medical Condition, Zachary A. Marcum, Mary Ann Sevick, Steven M. Handler, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3976600/>, 2013.
  - <https://www.nytimes.com/2017/04/17/well/the-cost-of-not-taking-your-medicine.html>



# Evidence #2: Matching Demand to Supply of Jobs is Inadequate Demand-Supply Gap in Jobs Market<sup>[1]</sup> and Yet, Low Work Satisfaction/ Engagement<sup>[2]</sup>

The screenshot shows the Indeed job search interface. The search bar at the top has 'human resources' entered. Below the search bar, there's a tip: 'Tip: Enter your city or zip code in the "where" box to show results in your area.' The main results page displays several job listings for 'Human Resources Manager' in Cortland, NY. Each listing includes the employer name (Byrne Dairy), review rating (★★★☆☆), and location (Cortland, NY 13045). A detailed description of the job responsibilities follows each listing. On the left sidebar, there are filters for salary estimates (e.g., \$30,000+, \$35,000+) and job types (e.g., Full-time, Part-time). At the bottom of the page, there's a section titled 'Job search at a portal'.

- Finding jobs was generally hard around the world (Dec 2019), except for in tight labor markets like US (3.5% unemployment)
- Workforce satisfaction/ engagement was generally low around the world – people did not find jobs they were match for [1,2]
- COVID-19 impact [3]:
  - *Nearly half of global workforce at risk of losing livelihoods in informal sector*
  - *9-12% job loss in the formal sector around the world*
  - *14.7% unemployment in US by end of April 2020* [4]

1. **Source:** Global Skills Trends, Training Needs and Lifelong Learning Strategies for the Future of Work, ILO & OECD Report 2018, [http://www.g20.utoronto.ca/2018/g20\\_global\\_skills\\_trends\\_and\\_lll\\_oecd-ilo.pdf](http://www.g20.utoronto.ca/2018/g20_global_skills_trends_and_lll_oecd-ilo.pdf)
2. **Source:** For 2016, job satisfaction: US – 32%, Global – 13%, <https://www.gallup.com/workplace/236495/worldwide-employee-engagement-crisis.aspx>
3. [https://www.ilo.org/global/about-the-ilo/newsroom/news/WCMS\\_743036/lang--en/index.htm](https://www.ilo.org/global/about-the-ilo/newsroom/news/WCMS_743036/lang--en/index.htm)
4. <https://www.bls.gov/news.release/empsit.nr0.htm>

# Case Study: Towards Trusted Decisions During COVID-19

---

# Decision Imperative: Corona Virus Pandemic

---

## Emerging Scenario Around the World\*

- Millions of cases, hundreds of thousands of deaths
- Businesses disrupted, millions going out of business
- Millions of jobs lost

\* Numbers changing continuously; see reference for details

## Decisions Need to be Made

- About disease
  - Understand disease
  - Tackle disease
- Understand impact to society: economy, supply chain
- Advise on actions to take
  - Individual
  - Group
  - Societal policy

**Resource:** <https://github.com/biplav-s/covid19-info/wiki/Important-Information-About-COVID19>

# AI-Based Decision-Support for COVID-19

1. Understanding the disease
  - (a) Disease spread and simulation models
  - (b) Insights by visualization
2. Understanding impact on society
  - (a) Understanding mental depression from social posts
  - (b) Assessing economic impact – job loss, industrial decline
  - (c) Effect on supply chain
  - (d) Assess risks
3. Observing disease in people
  - (a) Fever detection via images
  - (b) Tracking people's movement
1. Guidance for individual actions
  - (a) Screening/ triage tools
  - (b) Guidance about government benefits
  - (c) Vaccine appointments and scheduling
2. Guidance for group-level actions
  - (a) Models for when to open economy
  - (b) Contact tracing following an incident
  - (c) Matching producers and consumers to meet demand, reduce loss: food, medical supplies
3. Insights for policy actions
  - (a) Understanding impact of policy choices (e.g. lockdowns, travel restrictions)
  - (b) Design of economic interventions
  - (c) Fighting fake news

Resource: <https://github.com/biplav-s/covid19-info/wiki/Important-Information-About-COVID19>

**Source:** Biplav Srivastava, Did Chatbots Miss Their 'Apollo Moment'? A Survey of the Potential, Gaps and Lessons from Using Collaboration Assistants During COVID-19, [Cell PATTERNS](#), Aug 2021

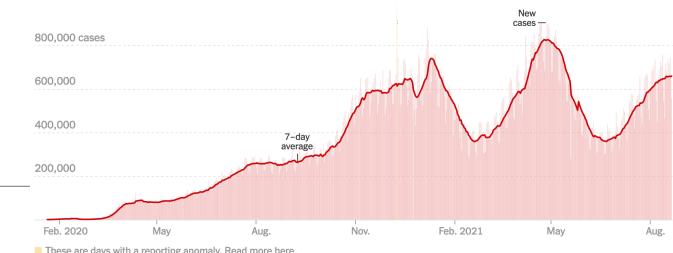
# Case Study: COVID-19

<https://www.nytimes.com/interactive/2020/world/coronavirus-maps.html>

Q: How good has a country done and is doing?

	CASES DAILY AVG.	PER 100,000	14-DAY CHANGE	DEATHS ▾ DAILY AVG.		PER 100,000	FULLY VACCINATED
				DAILY AVG.	PER 100,000		
Indonesia	19,072	7	-42%	1,282.3	0.47		11%
United States ›	151,227	46	+39%	1,007.0	0.30		51%
Russia	20,443	14	-8%	780.3	0.54		23%
Brazil ›	29,478	14	-12%	774.4	0.37		25%
Mexico ›	17,921	14	+8%	680.0	0.53		24%
Iran	35,944	43	+1%	592.3	0.71		5.3%
India ›	33,094	2	-17%	448.9	0.03		9.3%
South Africa	12,111	21	+12%	340.3	0.58		8.2%
Vietnam	10,178	11	+28%	300.4	0.31		1.7%
Thailand	20,715	30	+4%	249.1	0.36		8.2%

## New reported cases



	DAILY AVG. ON AUG. 21	14-DAY CHANGE	TOTAL REPORTED
Cases	659,135	+5%	211,484,780
Deaths	9,879	+4%	4,423,743

# Case Study: COVID-19

Updated Aug. 22, 2021

Q: How good has a US state done and is doing?

	CASES DAILY AVG.	PER 100,000	14-DAY CHANGE	HOSPITALIZED DAILY AVG.	PER 100,000	14-DAY CHANGE	DEATHS DAILY AVG.	PER 100,000	FULLY VACCINATED
United States	151,227	46	+39%	88,653	27	+43%	1,007.0	0.30	51%
Florida ›	23,314	109	+21%	16,728	78	+29%	212.3	0.99	51%
Texas ›	18,892	65	+57%	12,552	43	+44%	135.6	0.47	46%
California ›	13,673	35	+23%	7,776	20	+43%	64.6	0.16	55%
Louisiana ›	5,020	108	+9%	2,693	58	+27%	57.0	1.23	39%
Alabama ›	4,549	93	+47%	2,813	57	+48%	33.6	0.68	36%
North Carolina ›	5,374	51	+62%	2,493	24	+59%	32.9	0.31	45%
Mississippi ›	3,586	120	+71%	1,682	57	+39%	32.9	1.10	37%
Georgia ›	7,596	71	+71%	5,146	48	+61%	30.3	0.28	40%
Tennessee ›	4,989	73	+77%	2,186	32	+45%	28.7	0.42	41%
Arkansas ›	2,248	74	-4%	1,420	47	+16%	28.6	0.95	39%
Missouri ›	2,497	41	-13%	2,411	39	+11%	26.3	0.43	44%
South Carolina ›	3,621	70	+48%	1,676	33	+72%	24.9	0.48	42%

New reported cases



<https://www.nytimes.com/interactive/2020/us/coronavirus-us-cases.html>

# Pressing Issue: Distribution of Vaccines

---

- **Problem:** Limited supply, larger demand, many technologies, recipient hesitancy;  
How do distribute equitably, fairly and efficiently
- Possible (automated) solutions
  - Random: pick receiver based on random choice
    - **Benefit:** Easy to implement
    - **Problems:** Equitable but not fair, receiver may not be at risk or not want it, others wanting it may not get it
    - **Question:** assumes we can give vaccine quickly to the selected person
  - Prioritized random: make a prioritized list of groups, assign randomly in each group
    - **Benefit:** identifies affected groups
    - **Problems:** receiver may not want the vaccine
    - **Question:** who comes up with groups?, is it rewarding groups who have not been taking precautions ? Assumes we can give vaccine quickly to the selected person
    - ...
  - Benefit-cost: based on contribution to economy
    - **Benefit:** efficient

# Guideline: Human Impact of AI

---

- We study technology (AI) but it works with data
- Data, when from people or about people, can have issues like bias
  - **Example:** data reveals a view which is influenced by data collection practices
  - **Difference:** **World as it is**, world according to data and **world as it should be**
- The course and instructor believes in
  - Not promoting bias of any kind
  - Respecting everyone regardless of background

# Lecture 2: Concluding Comments

---

- We did a quick overview of the course
- Looked at AI, Data and Decisions
- Course will focus on
  - Practical methods to derive insights from open data
  - Evaluation will be by via project, paper and quizzes
  - **Bring your ideas to your project**
- Exciting techniques to learn to impact the world around us

# Concluding Segment

---

# Concluding Discussion: Course Project

---

- **Start thinking of your ideas for the project**

- **A Framework**

1. (Problem) Think of a problem whose solution may benefit people (e.g., health, water, air, traffic, safety)
2. (User) Consider how the primary user (e.g., patient, traveler) may be solving the problem today
3. (AI Method) Think of what the solution will do to help the primary user
4. (Data) Explore the data for a solution to work
5. (Reliability:Testing) Think of the evaluation metric we should employ to establish that the solution will work? (e.g., 20% reduction in patient deaths)
6. (Holding Human Values) Discuss if there are fairness/bias, privacy issues?
7. (Human-AI) Finally, elaborate how you will explain the primary user that your solution is trustable to be used by them

# About Next Lecture – Lecture 3

---

# Lecture 3: Data Prep, Knowledge Representation/ Graph

---

- Types
  - By format: Structured, Semi-structured, Unstructured,
  - By media: text, audio, video, multi-media;
  - By source: enterprise, open data, social, private
- Open Data
- Representation, Knowledge Graphs