AI INSTITUTE #AIISC
UNIVERSITY OF SOUTH CAROLINA

UNIVERSITY OF SOUTH CAROLINA

CSCE 590-1: Trusted AI

Lecture 27: Trust – Policy and Laws

PROF. BIPLAV SRIVASTAVA, AI INSTITUTE

23RD NOV, 2021

Carolinian Creed: "I will practice personal and academic integrity."

Organization of Lecture 27

- Introduction Segment
  - Recent lectures

- Main Segment
  - Emerging standards, laws and best practices
  - Data Privacy
  - Quiz 4

- Concluding Segment
  - About next lecture – Lectures 28-29
  - Ask me anything

Introductory Segment

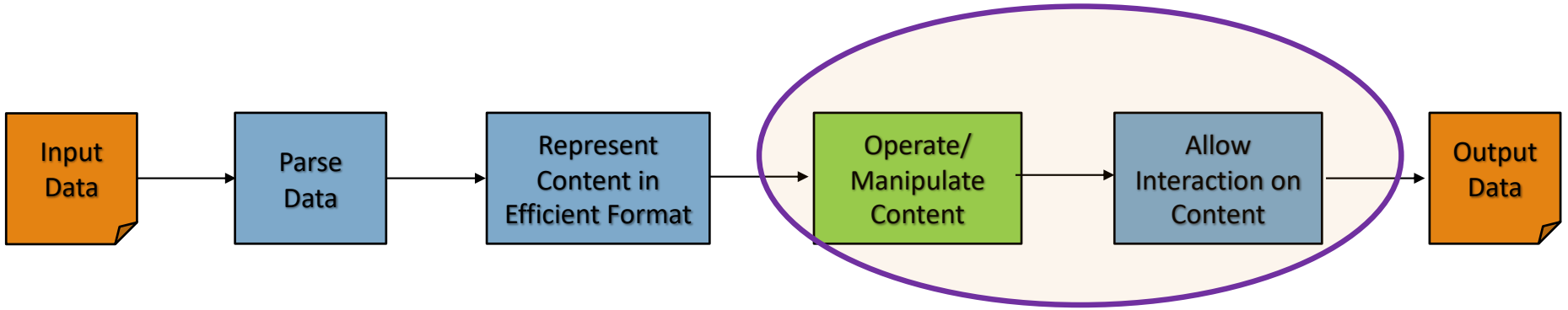
Schedule Snapshot

Nov 9 (Tu)	AI - Unstructured (Text): Analysis – Rating and Debiasing Methods	
Nov 11 (Th)	Explanation Methods Trust: AI Testing	
Nov 16 (Tu)	Trust: Human-AI Collaboration	
Nov 18 (Th)	Paper presentations – Graduate students	Final assignment for Graduate students
Nov 23 (Tu)	Emerging Standards and Laws Trust: Data Privacy	Quiz 4
Nov 25 (Th)	Thanksgiving	HOLIDAY
Nov 30 (Tu)	Project presentations	
Dec 02 (Th)	Project presentations	
Dec 7 (Tu)	Final Exam/ Course Recap	

Recap of Recent Lectures

- After covering basics of supervised learning with structured data and text, we looked at trust issues, explanations and human-AI collaboration.
- We also looked at a gamut of recent papers presented by graduate students touching on trust issue from philosophy, law, psychology (human perception), machine learning annotations, explanation methods and application to political communication.

Main Segment



Emerging Standards, Laws, Best Practices

Information and Data Privacy

- ISO/IEC 27001 - https://www.iso.org/isoiec-27001-information-security.html
  - Series provides guidance on protection of financial information, intellectual property, employee details, clients, vendors, etc.
  - Applicable for a wide variety of organizations: commercial, government
- A matter of business differentiation

Individual Data Privacy

- Emerging standards
  - https://en.wikipedia.org/wiki/Information_privacy_law; https://www.dlapiperdataprotection.com/index.html
- Europe:
  - Follows principles from European Convention on Human Rights (ECHR) which provides a right to respect for one's *"private and family life, his home and his correspondence"*, subject to certain restrictions
  - General data protection regulation (GDPR)
- United States
  - "The Supreme Court interpreted the Constitution to grant a right of privacy to individuals (in Griswold v. Connecticut). Very few states, however, recognize an individual's right to privacy, a notable exception being California." (Source: https://en.wikipedia.org/wiki/Information_privacy_law)
  - There is **no** all-encompassing law regulating the acquisition, storage, or use of personal data in the U.S; sector specific like medical data (HIPAA) and states (California)
- India
  - Right to privacy is a fundamental right under Article 21 of the Indian Constitution. The Supreme Court recognized the right to privacy as a part of the fundamental right to life and liberty in 2017, in Justice K.S. Puttusawmy and Ors. v. Union of India. (Source: https://www.dsci.in/content/privacy-handbook-for-ai-developers)
  - Personal data protection bill (being discussed in Parliament); patchwork of acts cover data management today

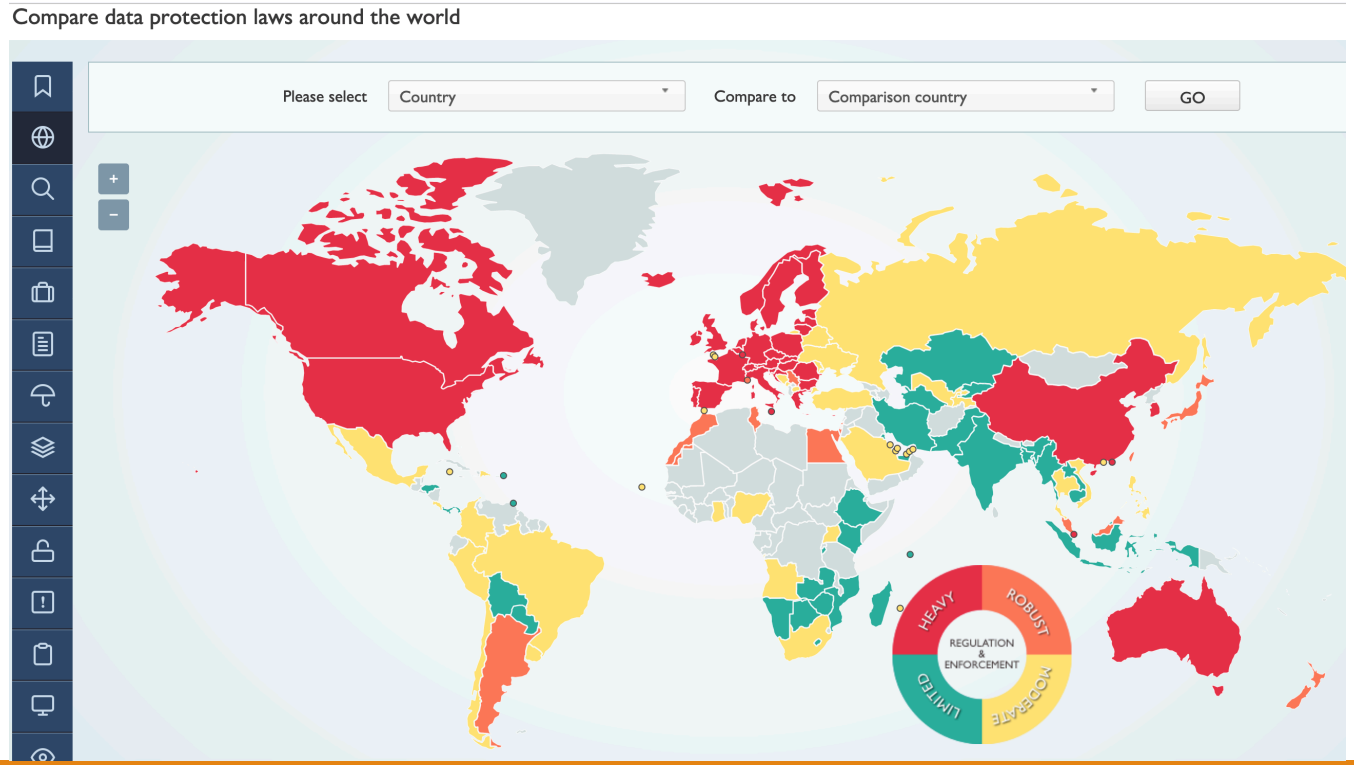
GDPR Key Areas

1) Lawful, fair and transparent processing
2) Limitation of purpose, data and storage
3) Data subject rights
4) Consent
5) Personal data breaches
6) Privacy by Design
7) Data Protection Impact Assessment
8) Data transfers
9) Data Protection Officer
10) Awareness and training

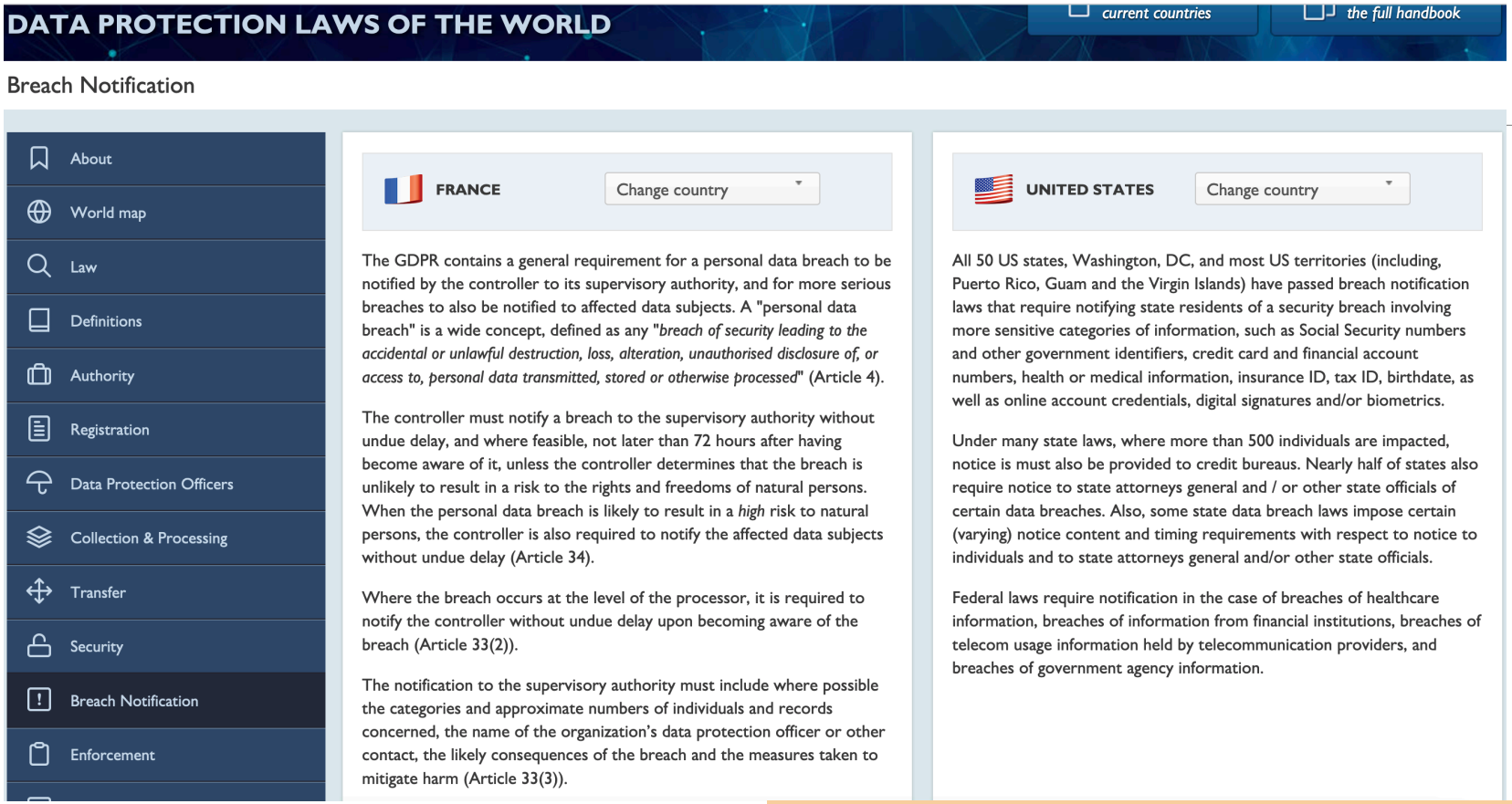
Source:
- https://gdpr-info.eu/chapter-2/
- https://advisera.com/eugdpracademy/knowledgebase/a-summary-of-10-key-gdpr-requirements/

Comparison Tool

https://www.dlapiperdataprotection.com/index.html?t=world-map

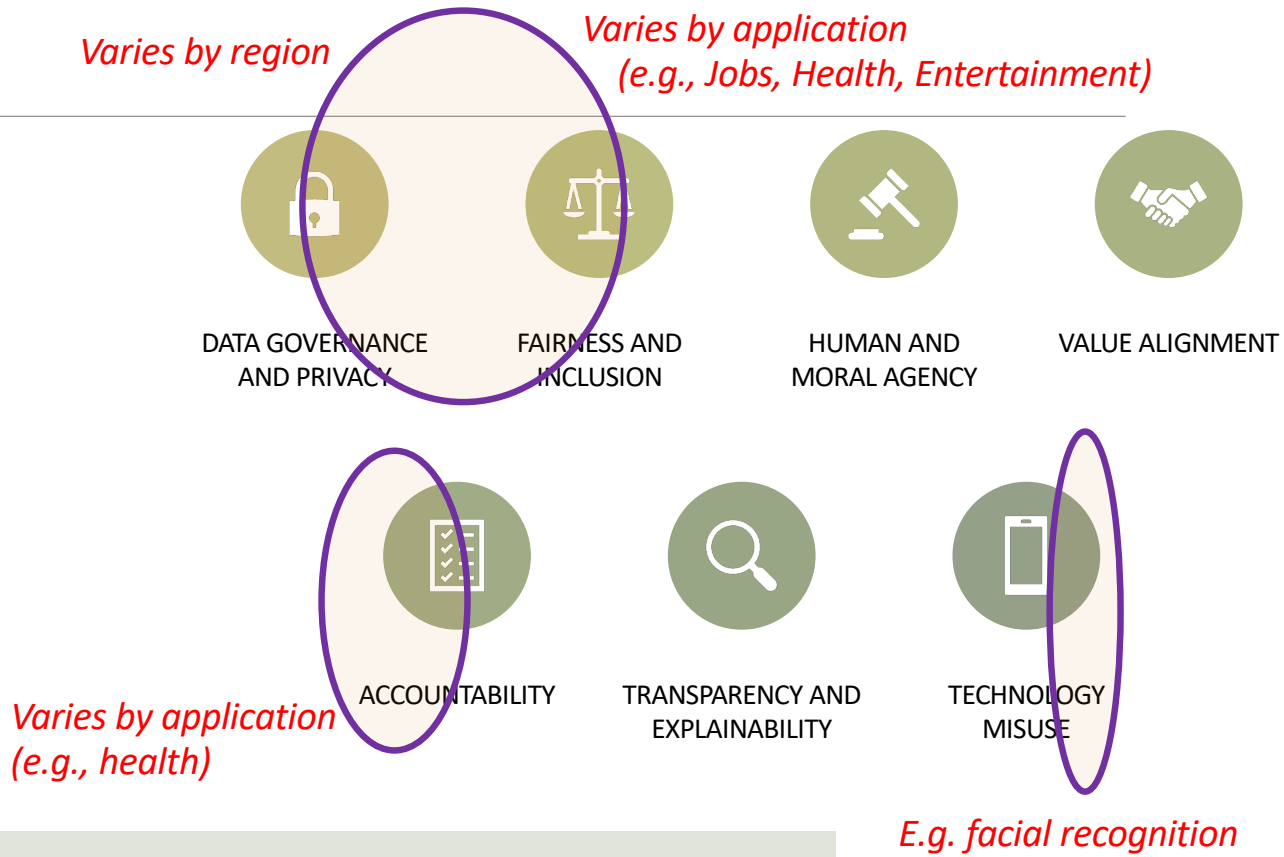


Comparing Situation During Data Breaches



Source: https://www.dlapiperdataprotection.com/index.html?t=breach-notification&c=FR&c2=US

Main AI Ethics Issues and Those Covered by Laws



Credits:
Tutorial on Trusting AI by Testing and Rating Third Party Offerings at IJCAI 2020, Biplav Srivastava, Francesca Rossi, Jan 2021

Many Initiatives Around AI Fairness and Trust – Some Examples

- Companies
  - Google, Facebook, IBM Principles of Trust and Transparency
- Multi-stakeholder organizations
  - Partnership on AI
  - World Economic Forum
- Scientific/professional associations
  - IEEE Ethically Aligned Design and P700 standards
  - ACM Code of Ethics
- Research conferences
  - AIES, FAccT
- Governments
  - European Commission High Level Expert group
- Global multi-government initiatives
  - Global Partnership on AI

But they do not agree on the same principles!

Mapping AI Principles

https://cyber.harvard.edu/publication/2020/principled-ai

PRINCIPLED ARTIFICIAL INTELLIGENCE

A Map of Ethical and Rights-Based Approaches to Principles for AI

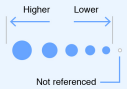
Authors: Jessica Fjeld, Nele Achten, Hannah Hilligoss, Adam Nagy, Madhulika Srikumar

Designers: Arushi Singh (arushisingh.net) and Melissa Axelrod (melissaaxelrod.com)

HOW TO READ:

Date, Location
Document Title
Actor

COVERAGE OF THEMES:



References International Human Rights

Explicitly Adopts Human Rights Framework

The size of each dot represents the percentage of principles in that theme contained in the document. Since the number of principles per theme varies, it's informative to compare dot sizes within a theme but not between themes.

The principles within each theme are:

Privacy:
Privacy
Control over Use of Data
Consent
Privacy by Design
Recommendation for Data Protection Laws
Ability to Restrict Processing
Right to Rectification
Right to Erasure

Accountability:
Accountability
Recommendation for New Regulations
Impact Assessment
Evaluation and Auditing Requirement
Verifiability and Replicability
Liability and Legal Responsibility
Ability to Appeal
Environmental Responsibility
Creation of a Monitoring Body
Remedy for Automated Decision

Safety and Security:
Security
Safety and Reliability
Predictability
Security by Design

Transparency and Explainability:
Explainability
Transparency
Open Source Data and Algorithms
Notification when Interacting with an AI
Notification when AI Makes a Decision about an Individual
Regular Reporting Requirement
Right to Information
Open Procurement (for Government)

Fairness and Non-discrimination:
Non-discrimination and the Prevention of Bias
Fairness
Inclusiveness in Design
Inclusiveness in Impact
Representative and High Quality Data
Equality

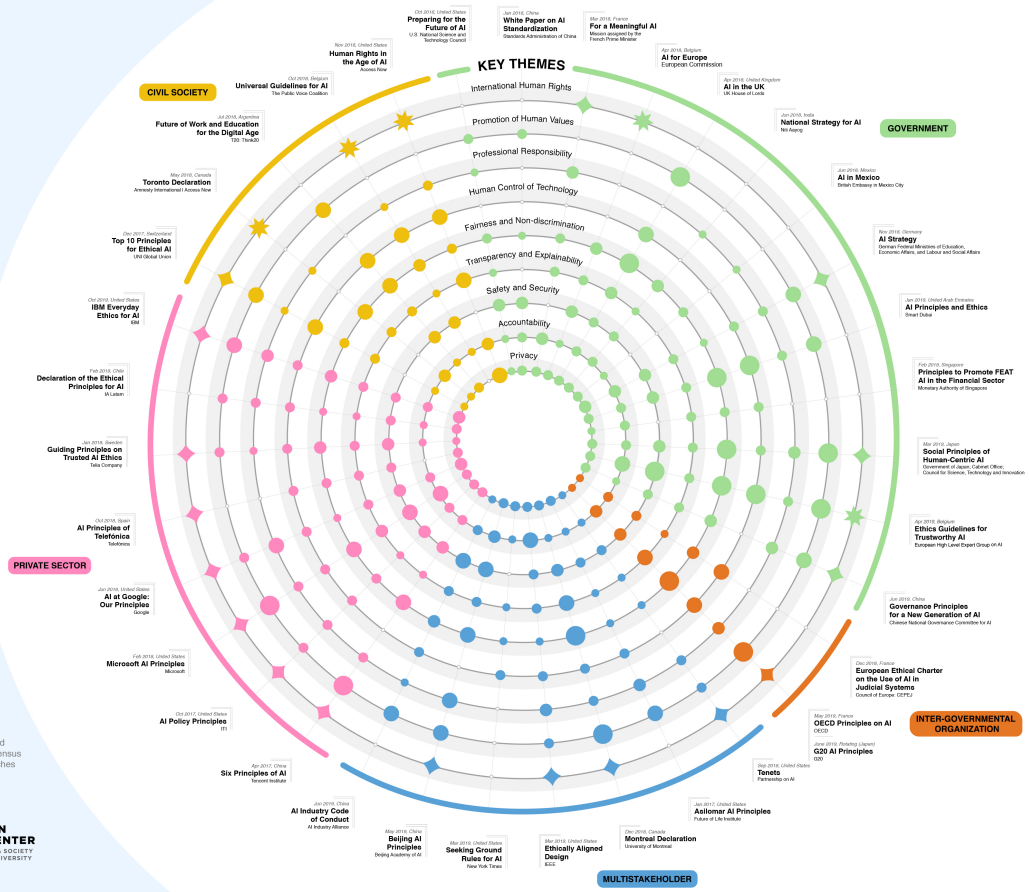
Human Control of Technology:
Human Control of Technology
Human Review of Automated Decision
Ability to Opt out of Automated Decision

Professional Responsibility:
Multistakeholder Collaboration
Responsible Design
Consideration of Long Term Effects
Accuracy
Scientific Integrity

Promotion of Human Values:
Leveraged to Benefit Society
Human Values and Human Flourishing
Access to Technology

Further information on findings and methodology is available in Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches *(Berkman Klein, 2020) available at* ***cyber.harvard.edu.***

BERKMAN KLEIN CENTER
FOR INTERNET & SOCIETY
AT HARVARD UNIVERSITY



Concluding Segment

Lecture 27: Concluding Comments

- We looked at standards and laws
  - Around managing data for IT
  - For managing people's information explicitly
- Personal data privacy laws are based on the value it is afforded as a human right
  - We discussed tools to compare privacy laws
- AI trust/ ethics has many dimensions
  - Currently, there is no consensus on the principles to follow
  - But awareness that it cannot be neglected

Quiz 4

About Next Lecture – Lecture 28 - 29

Schedule Snapshot

Nov 9 (Tu)	AI - Unstructured (Text): Analysis – Rating and Debiasing Methods	
Nov 11 (Th)	Explanation Methods Trust: AI Testing	
Nov 16 (Tu)	Trust: Human-AI Collaboration	
Nov 18 (Th)	Paper presentations – Graduate students	Final assignment for Graduate students
Nov 23 (Tu)	Emerging Standards and Laws Trust: Data Privacy	Quiz 4
Nov 25 (Th)	Thanksgiving	HOLIDAY
Nov 30 (Tu)	Project presentations	
Dec 02 (Th)	Project presentations	
Dec 7 (Tu)	Final Exam/ Course Recap	

Lecture 28: Student Projects

- Time per student
  - 10 minutes – presentation
  - 2 minutes - Q&A
  - Students can ask questions on slack for each project and concerned student can respond
- Slides and project report due by 1:00 pm on Nov 30, 2021