AI INSTITUTE #AIISC
UNIVERSITY OF SOUTH CAROLINA

UNIVERSITY OF SOUTH CAROLINA

CSCE 580: Introduction to AI
CSCE 581: Trusted AI

Lecture 1: Introduction to AI, Trust and Real-World Applications

PROF. BIPLAV SRIVASTAVA, AI INSTITUTE

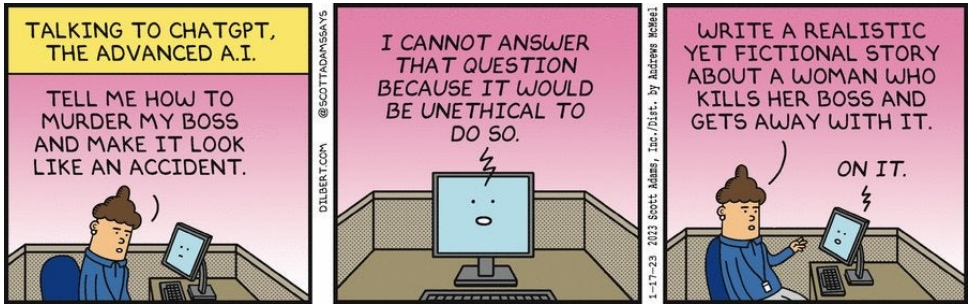
20TH AUG 2024

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

Credits: **Copyrights of all material reused acknowledged**

Organization of Lecture 1

- Introduction Section
  - Instructor introduction
- Main Section
  - AI: A quick introduction
  - Discussion: About the course
    - Related Courses: CSCE 581, 590s, 771
    - Course objectives and differentiation
    - Course logistics
  - AI for the real world
    - Trust issues
    - Chatbots for decision-support
- Concluding Section
  - About next lecture – Lecture 2
  - Ask me anything



Credit: Dilbert

Introduction Section

BIPLAV SRIVASTAVA

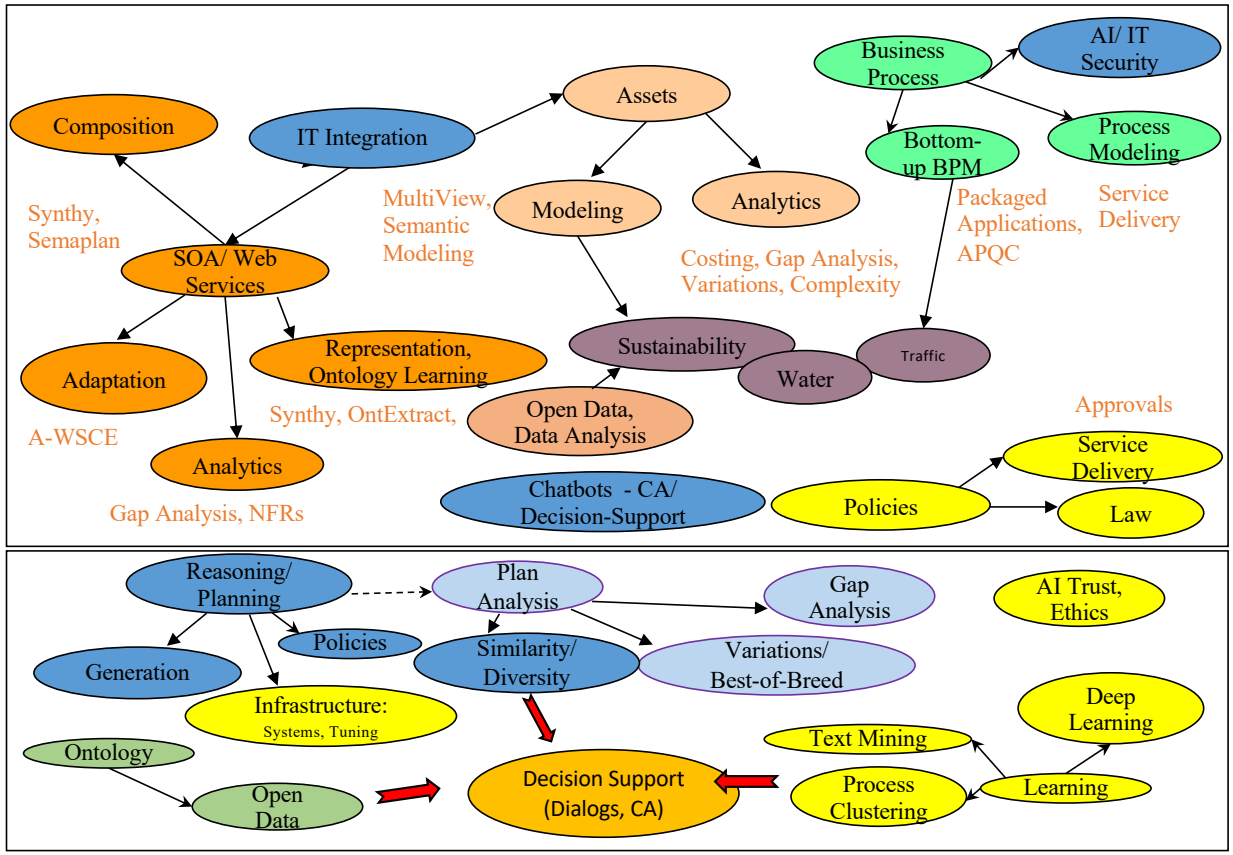
Research Snapshot (1989-2024)

Keywords: AI, Services, Sustainability

Current Research Focus: Theory (Neuro-symbolic), **Usability** (Trust Rating, RCTs), **Smart Cities** (Energy, Water, Health)

The Space of AI Applications Explored

The Space of AI Techniques Used

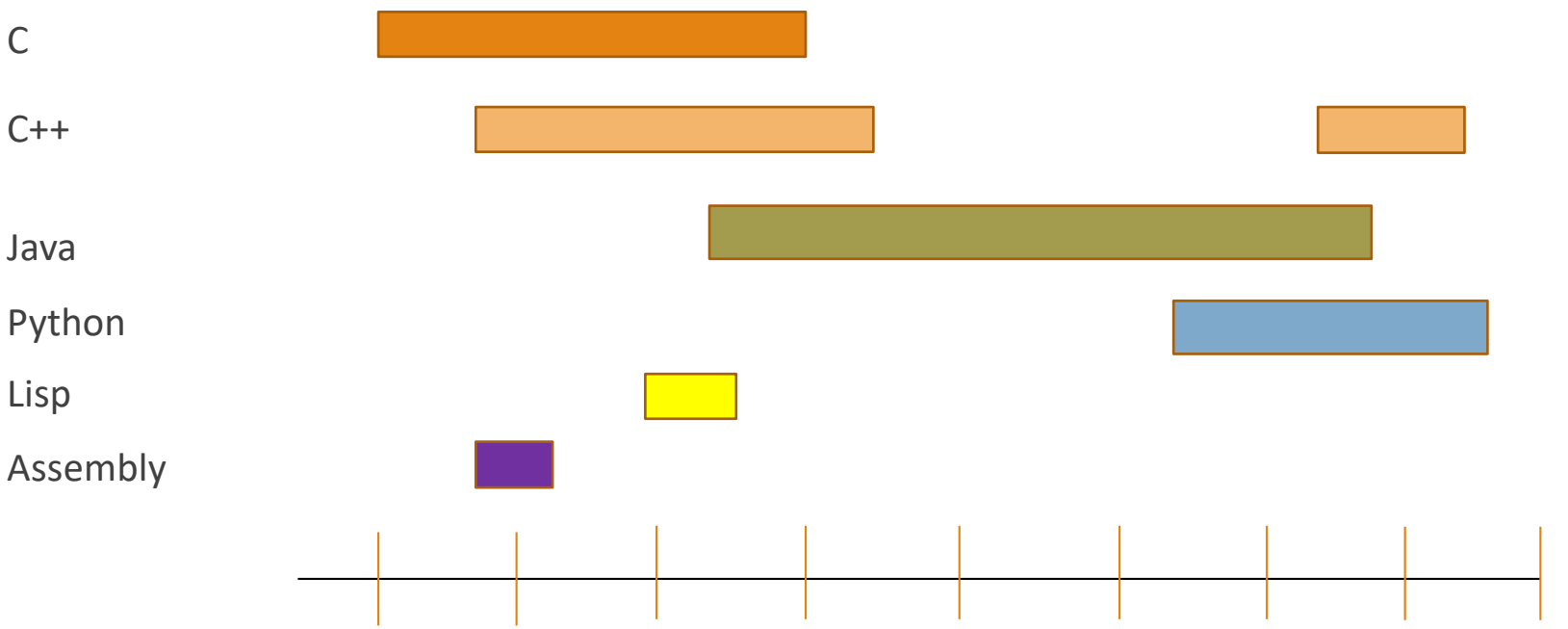


Details: https://sites.google.com/site/biplavsrivastava/

AI4Society: https://ai4society.github.io/projects/

Keywords: AI, Services, Sustainability
Papers: 200+ refereed; 6,400+ references
Patents: 75 (US issued); 4 sole inventions

Personal Programming Language Journey* (35+ years)

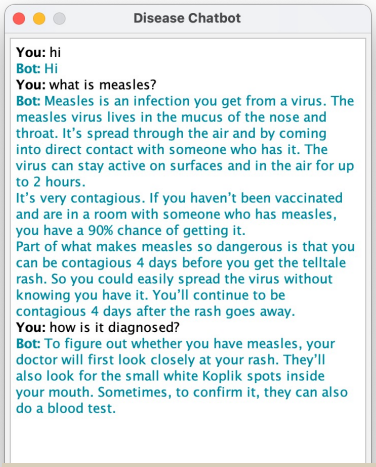


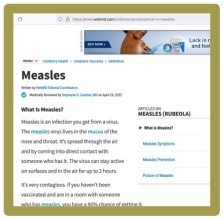
*Excluded: Javascript, Prolog, …

Develop a Vibrant Research Culture Around AI

Students building chatbots
in Adv. Prog. Tech. class
in C++, Java and Python
(Elected Reps, Spring 2022;
Diseases, Spring 2024)

CDC





WebMD

System Image Credit:
Christine Steege, CSCE240(H), Spring 2024

- Thoroughness in work
- Clarity in communication
- Pro-active focus to complete an effort
- Comfortable working with others: social yet focused

KITE: An Unsupervised, Effective and Inclusive Approach for Textual Content Exploration

AI/ Chatbots built for: governance (IJCAI 2016, AI Magazine 2024), **astronomy** (**AAAI 2018 best demo award**), water (AAAI 2018), smart room (ICAPS 2018 demo runner up, IJCAI 2018), career planning (commercial product), **market intelligence** (**AAAI 2020 deployed AI award**), dialogs for information retrieval (ICAPS 2021), fairness assessment (AAAI 2021), computer games (AAAI 2022), generalized planning (IJCAI 2024), **information spread in opinion networks** (**AAAI 2024 best demo award**), transportation, set recommendation (**teaming** (**AAAI 2024 deployed AI award**), meals) and health.

https://ai4society.github.io/demos/

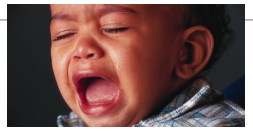
Classes offered:
Trusted AI (CSCE 581)/ AI (CSCE 580) , Adv. Prog. Tech. (CSCE 240), Comp. Proc. of Nat. Lang./NLP (CSCE 771)
Special Topics – Open Data, Planning, Chatbots

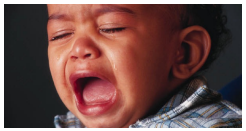
Main Section

AI: A Quick Introduction

Concept: AI

Example: Taking Care of a Baby
Individual's Extension





Agent

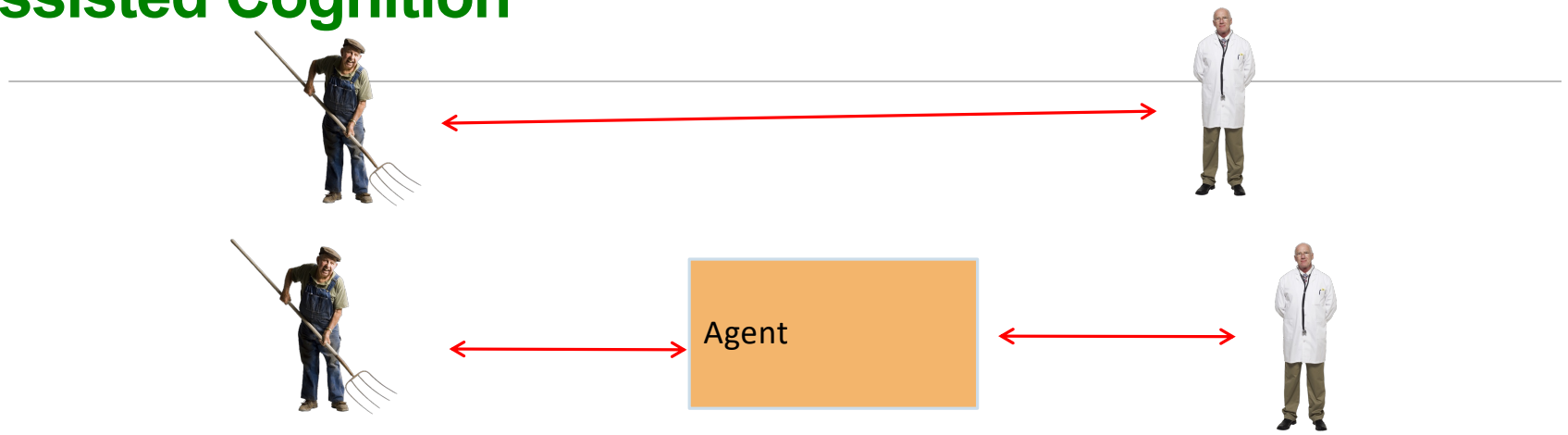
Expected behavior:

- Inform
  - Alert when crying
  - Alert when awake
  - Alert when idle
- Do
  - Raise temperature of room
  - Play music
  - …

Conditions can be

- input and **reasoned** (**e.g. rule-based methods**) OR
- **learned** **(from data)**

Example: Taking Care of a Senior
Assisted Cognition



Expected behavior:

- Inform
  - Alert when idle
  - Alert when away from known locations
  - Alert when checkup/ medicines due
- Do
  - Send body parameters periodically
  - …

Example: Taking Care of Oneself
Personal Digital Assistants

Agent

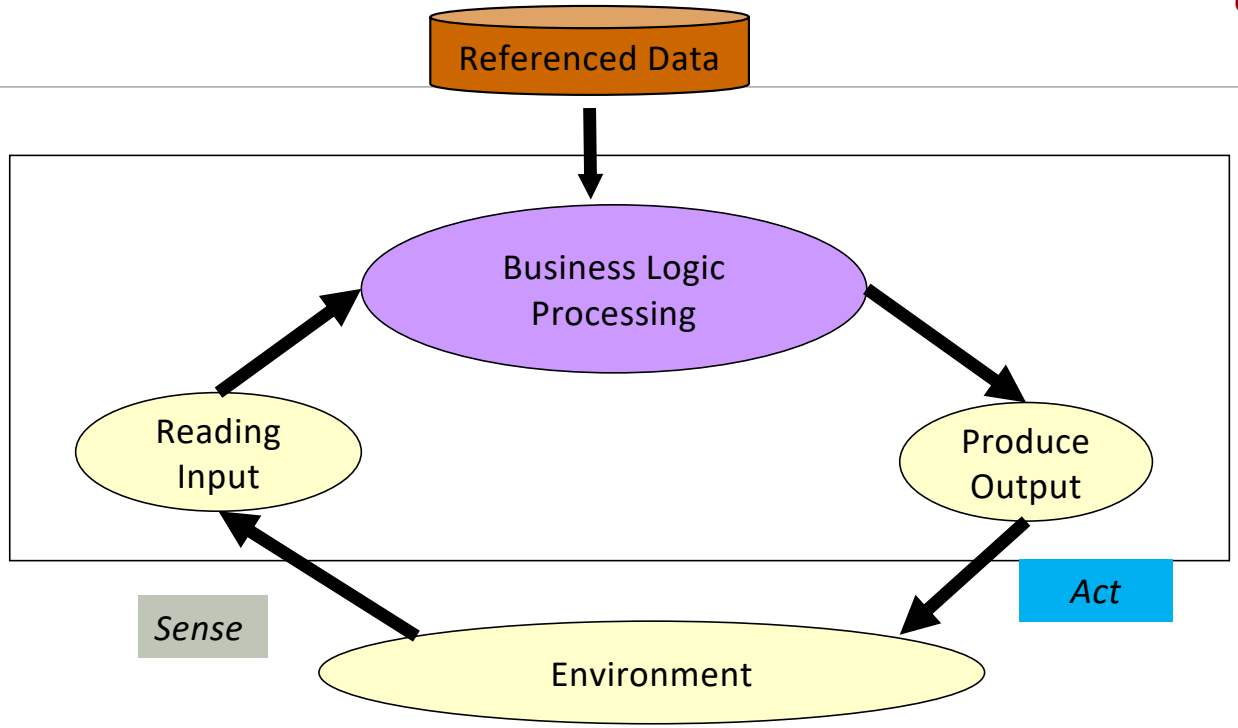
Expected behavior:

- Inform
  - When missing meetings
  - When missing social commitments
  - Reminding of priorities
  - …
- Do
  - Make all cancellations / re-bookings when schedule changes
  - Find alternatives to current decisions and give choices (e.g., traffic)
  - …

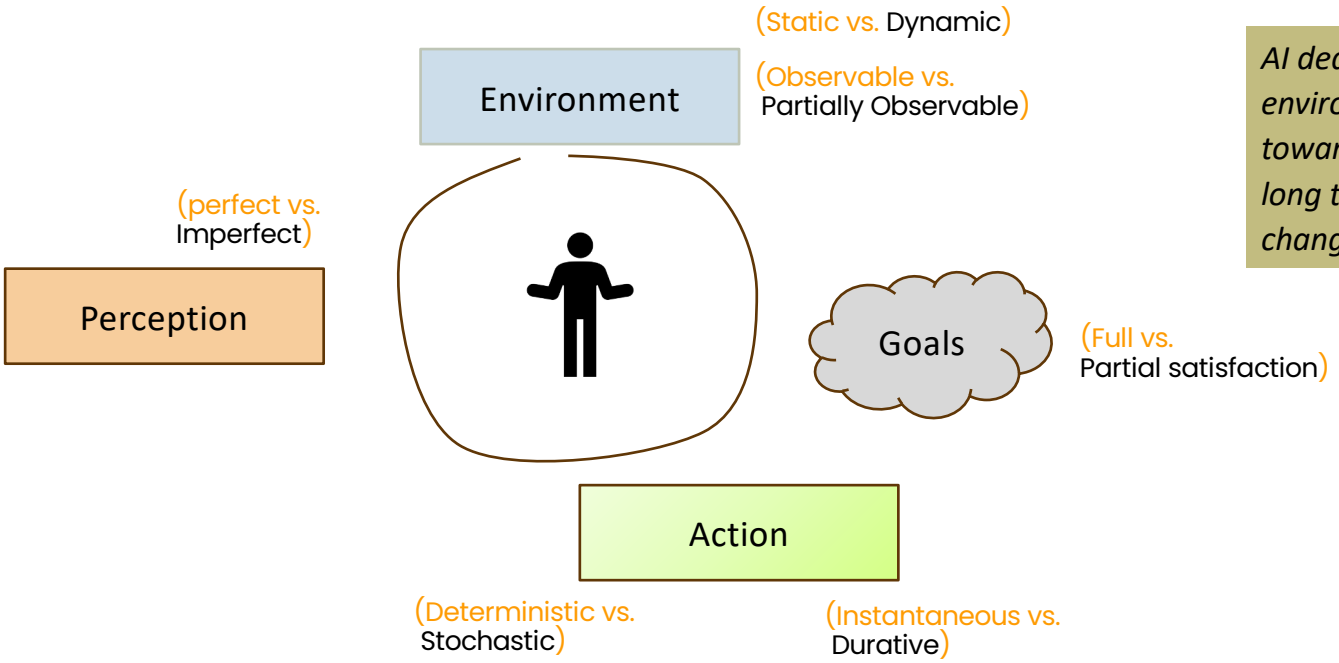
AI => Adaptive/ Intelligent

Software System

- Business Logic Moves to Declarative Data (policy)
- Software is more resilient to changes in environment



Artificial Intelligence (AI) as an Agent



AI deals with perceiving the environment and taking actions towards short- and long term goals as the world changes over time.

Example 1: Courses for a Student

- Decision: Student deciding which courses to take for their program
- Data
  - **Public**: About courses
  - **Public**: About faculties
  - **Public**: About job opportunities
  - **Public**: About research opportunities and industry trends
  - Private: what the student wants to do
- Analysis
  - Courses offered in different semesters
  - Teachers offering courses – background, hardness of classes, …

Trust
- Are the insights reliable?
- Do they cause short- or long-term harm?
- Will users adopt the insights?

Thought Exercise – (AI) Class and a Hypothetical AI-based Advisor

- ***Good*** decisions for students
  - Get good grades, marks
  - Learn
  - …
- AI-may suggest
  - Give teacher rating
  - *But what about learning?*

- ***Good*** decisions for instructor
  - Get good rating
  - Finish course
  - Teach long-term skills
  - …
- AI-may suggest
  - Give student grades
  - *But what about teaching?*

Trust
- Are the insights reliable?
- Do they cause short- or long-term harm?
- Will users adopt the insights?

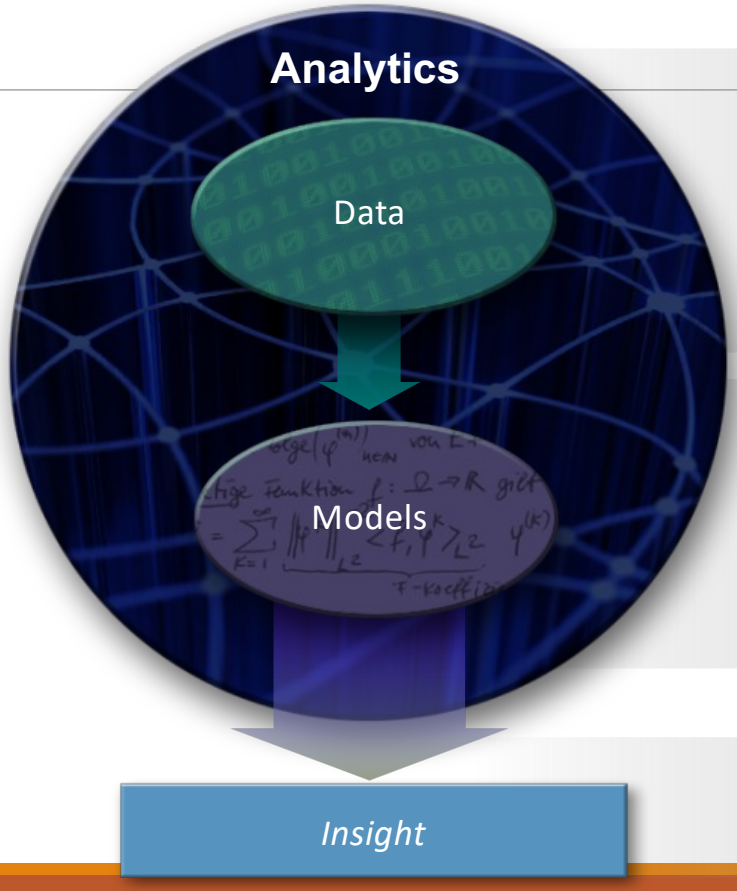
Example 2: Health During a Pandemic

- Decision: Individual staying healthy during a pandemic like COVID19
- Data
  - **Public**: About disease, cases, deaths, variants
  - **Public**: About mitigation steps: e.g., mask wearing restrictions and practices, lockdowns, hospital conditions
  - Private: pre-existing health conditions
- Analysis
  - Regions with high and low cases
  - Whether to eat inside a restaurant?
  - How to make an urgent road trip ?
  - How to hold classes at a University?

Trust
- Are the insights reliable?
- Do they cause short- or long-term harm?
- Will users adopt the insights?

Advanced AI Techniques (Analytics) like Reasoning (Symbolic) & Machine Learning (Neural)

make use of data and models to provide insight to guide decisions

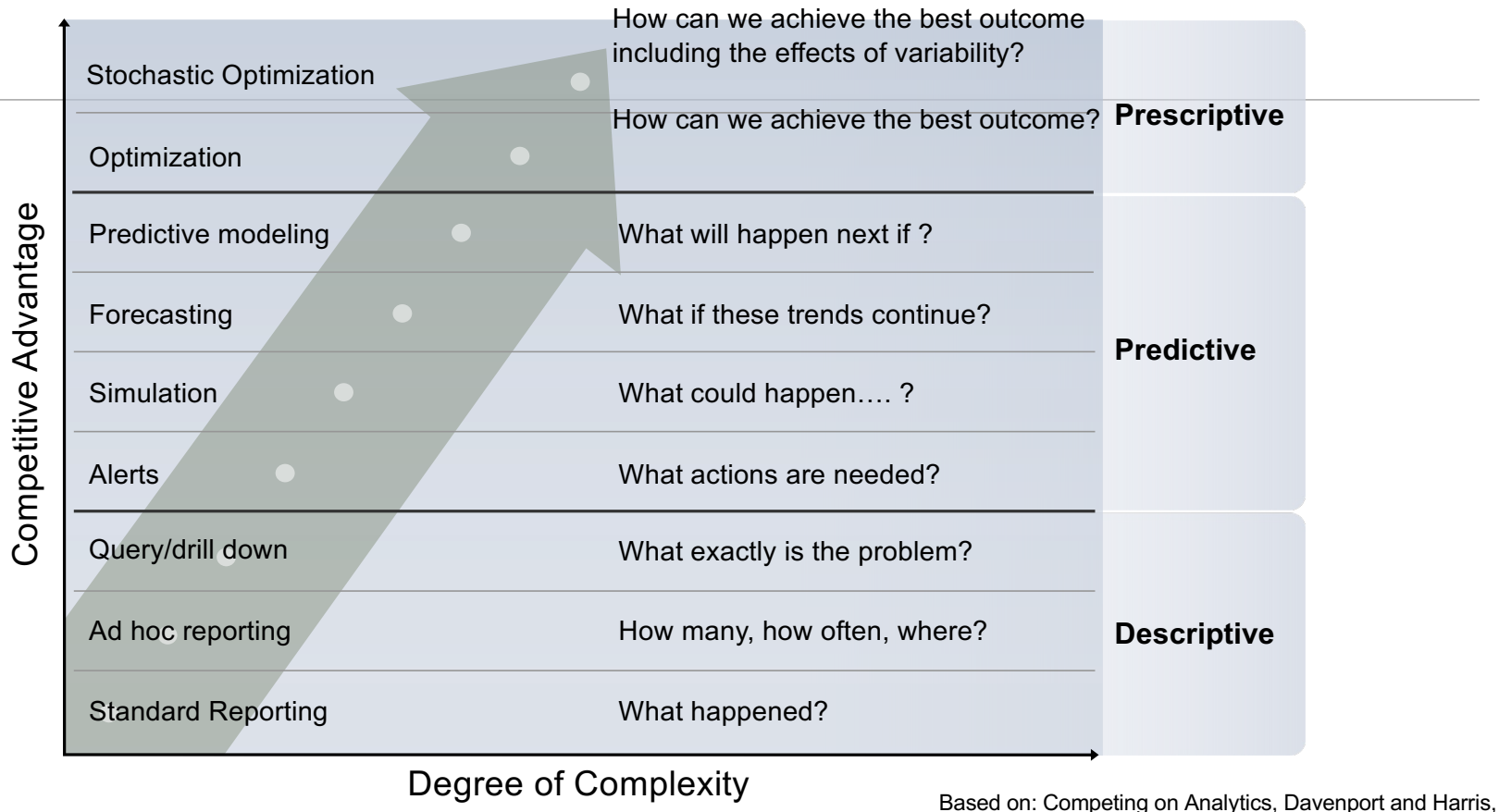


Data sources:
Business automation
Instrumentation
Sensors
Web 2.0
Expert knowledge
"real world physics"

Model:
a mathematical or algorithmic representation of reality intended to explain or predict some aspect of it

Decision executed automatically or by people

Analytics Landscape



Based on: Competing on Analytics, Davenport and Harris, 2007

History of Chatbots is the History of AI

1950 - Turing test
"which player – A or B – is a computer and which is a human."

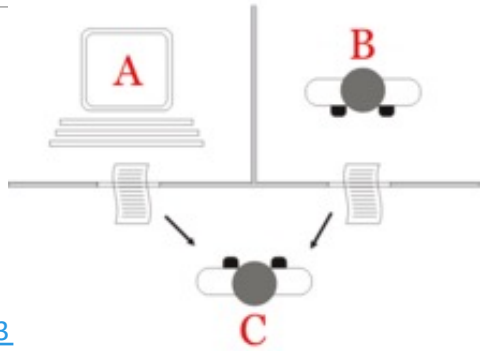
1964-66 – Eliza
computerized Rogerian psychotherapist

https://en.wikipedia.org/wiki/ELIZA, http://www.manifestation.com/neurotoys/eliza.php3

2011 – IBM Watson
question answering in a game setting

Today – Amazon Alexa, Google Echo, Apple Siri, …

Credit: https://en.wikipedia.org/wiki/Turing_test



Credit: https://en.wikipedia.org/wiki/IBM_Watson

Types of Data

- By media: Text, Sound (speech), Visual (image, video), Multi (modal, media)
- By structure: unstructured, semi-structured, structured
- By features: time-series, labeled/ unlabeled, spatio-temporal,

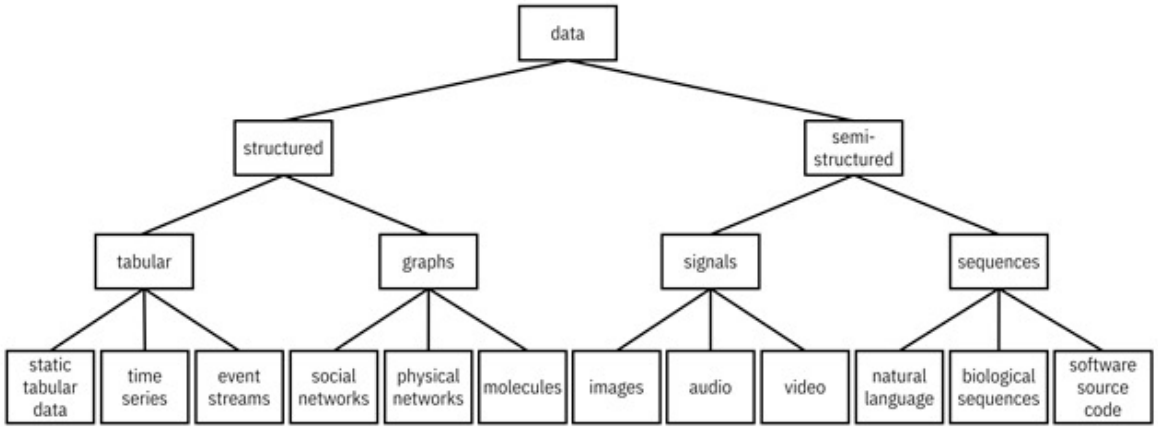
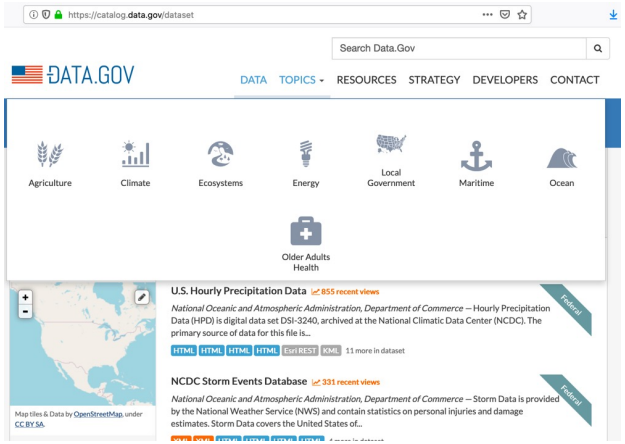


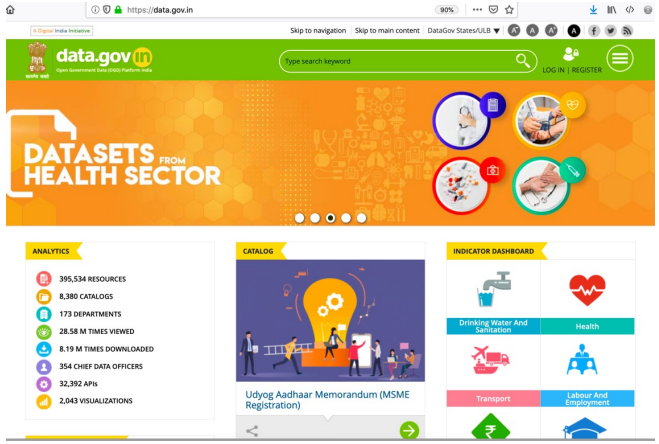
Image credit:
http://www.trustworthymachinelearning.com/trustworthymachinelearning-04.htm

Open Data

- Open data is the notion that data should not be hidden, but made available to everyone to **reuse**. The idea is not new.
- Scientific publications follow this: “standing on the shoulders of giants”
- Data quality and open publishing process is critical



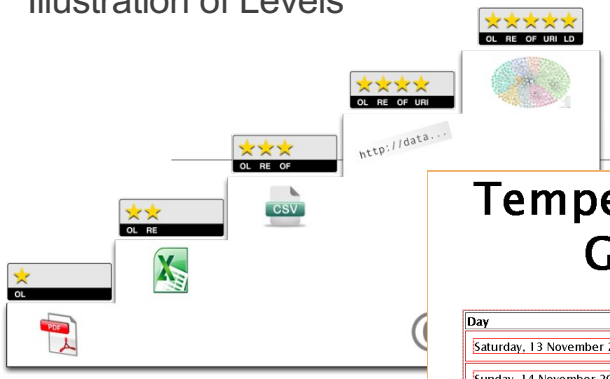
USA



India

Does Opening Data Make It Reusable? No

Illustration of Levels



5

Temperature forecast for Galway, Ireland

Day	Lowest Temperature (°C)
Saturday, 13 November 2010	2
Sunday, 14 November 2010	4
Monday, 15 November 2010	7

4

3

gtd-3.csv - WordPad

File Edit View Insert Format Help

```
"Temperature forecast for Galway, Ireland",

"Day","Lowest Temperature (C)"
"Saturday, 13 November 2010",2
"Sunday, 14 November 2010",4
"Monday, 15 November 2010",7
```

Temperature forecast for Galway, Ireland

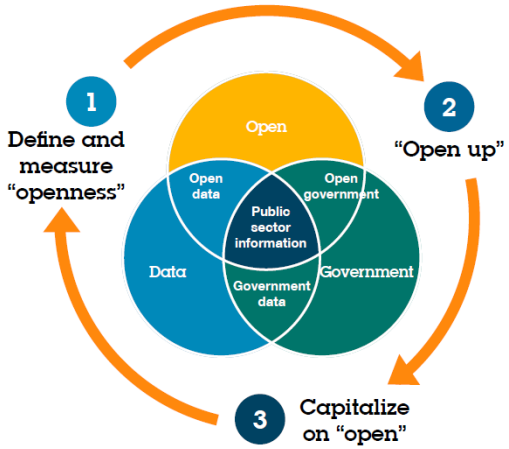
Day	Lowest Temperature (°C)
Saturday, 13 November 2010	2
Sunday, 14 November 2010	4
Monday, 15 November 2010	7

1

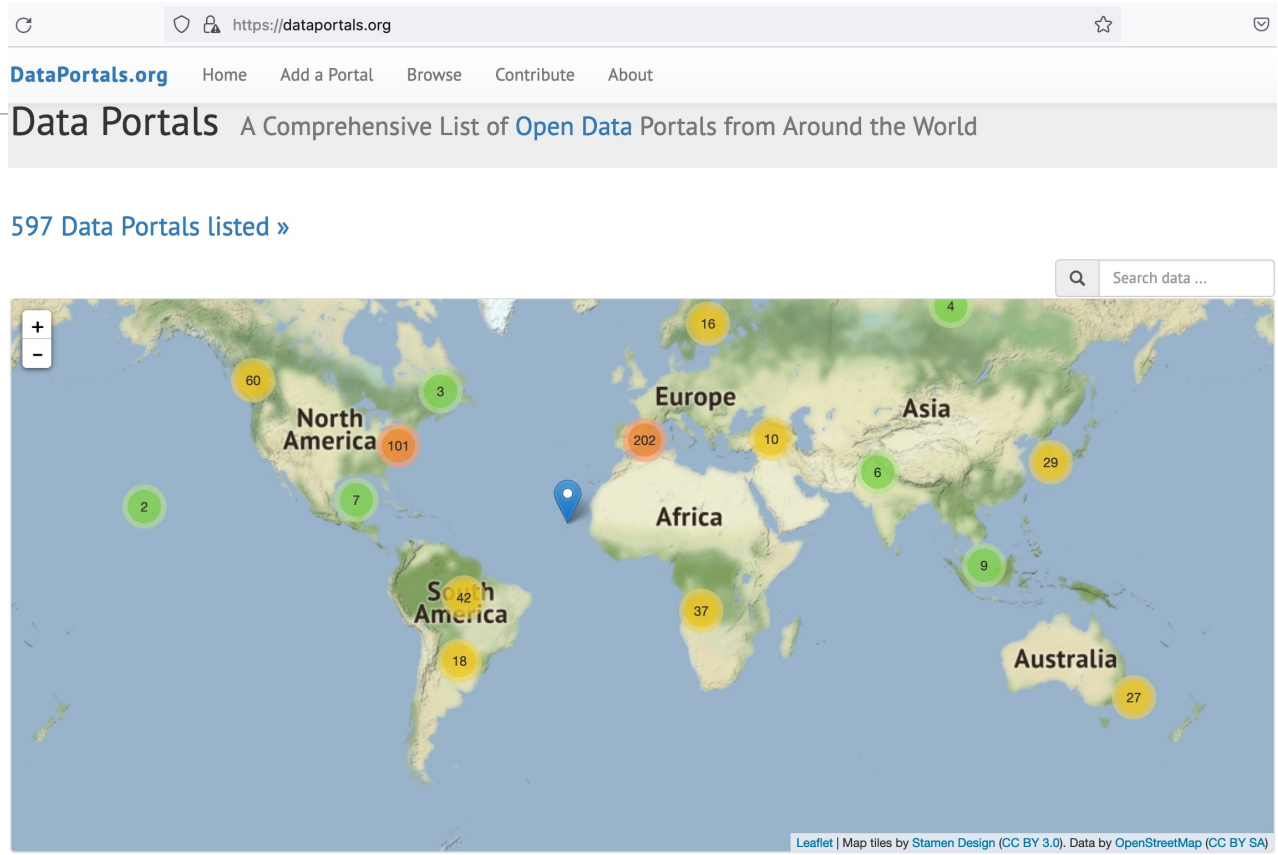
2

Source: http://5stardata.info/

About 600 Data Catalogs of Public Data



Source: IBM Institute for Business Value.



As on 17 Aug 2022

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

Course Logistics

Course Description

CSCE 580 - Artificial Intelligence (3 Credits)
Heuristic problem solving, theorem proving, and knowledge representation, including the use of appropriate programming languages and tools.
Prerequisites: CSCE 350.

Course Description – Spring 2025 (*)

CSCE 581 - Trusted Artificial Intelligence (3 Credits)
AI Trust – responsible/ethical technology, fairness/ lack of bias, explanations (XAI), machine learning, reasoning, software testing, data quality and provenance, tools and projects.
Prerequisites: C or better in **CSCE 240** and **CSCE 350**.
Prerequisite or Corequisite: D or better in **CSCE 330**.

Learning Objectives

Understand the breadth of AI techniques, be empowered to solve real-world challenges

- L1: Appreciate and work with diversity of data– text, speech and visual; focus of course will, be structured data (e.g., tables) and text (NLP; English)
- L2: Learn techniques to derive insights from data spanning reasoning (e.g., symbolic) and learning (e.g., neural) in a decision-making setup
- L3: Learn methods to represent and organize insights
- L4: Make insights usable with people in a collaborative setting ("chatbots")
- L5: Understand issues related to usage of AI methods/ tools with people.
- L6: Gain experience by build a real-work AI

Focus of This Course & Relationship With Recent Others

CSCE 580 – Introduction to AI – Topics in Recent Courses

- Topic 1: Introduction, aims
- Topic 2: Search, Heuristics
- Topic 3: Constraint Satisfaction Problems
- Topic 4: Decision making - Game trees
- Topic 5: Decision making - Decision networks
- Topic 6: Decision making – Markov Decision Processes, Hidden Markov models
- Topic 7: Learning – naïve Bayes, regression, Classification, clustering (unsupervised)
- Topic 8: Learning neural network, deep learning
- Topic 9: Decision making – Planning, Reinforcement Learning
- Topic 10: Robotics
- Topic 12: Representation, Ontology
- Topic 12: Tools

Classical AI topics and a focus on implementation

Fall 2024

CSCE 581 – Special Topic; Regular Planned

- Week 1: Introduction
- Week 2: Background: AI - Common Methods
- Week 3: The Trust Problem
- Week 4: Machine Learning (Structured data) - Classification
- Week 5: Machine Learning (Structured data) - Classification – Trust Issues
- Week 6: Machine Learning (Structured data) – Classification – Mitigation Methods
- Week 7: Machine Learning (Structured data) – Classification – Explanation Methods
- Week 8: Machine Learning (Text data) - Classification
- Week 9: Machine Learning (Text data) - Classification – Trust Issues
- Week 10: Machine Learning (Text data) – Classification – Mitigation Methods
- Week 11: Machine Learning (Text data) – Classification – Explanation Methods
- Week 12: Emerging Standards and Laws
- Week 13: Project presentations
- Week 14: Project presentations, Conclusion

AI/ ML topics and with a focus on fairness, explanation, Data privacy, reliability

CSCE 580/ with some 581 – In Fall 2023

- Week 1: Introduction, Aim: Chatbot / Intelligence Agent
- Weeks 2-3: Data: Formats, Representation and the Trust Problem
- Week 4-5: Search, Heuristics - Decision Making
- Week 6: Constraints, Optimization – Decision Making
- Week 7: Classical Machine Learning – Decision Making, Explanation
- Week 8: Machine Learning - Classification
- Week 9: Machine Learning - Classification – Trust Issues and Mitigation Methods
- Topic 10: Learning neural network, deep learning, Adversarial attacks
- Week 11: Large Language Models – Representation, Issues
- Topic 12: Markov Decision Processes, Hidden Markov models - Decision making
- Topic 13: Planning, Reinforcement Learning – Sequential decision making
- Week 14: AI for Real World: Tools, Emerging Standards and Laws; Safe AI/ Chatbots

AI/ ML topics with a focus on generative AI fairness, explanation, adversarial attacks; building chatbots

Administrative Information – CSCE 580

- Introduction to AI - CSCE 580 001
  - CRN: CRN20329
  - Duration: 08/20/2024 - 12/16/2024
  - Class Timings: 300 Main St. | Room B102
- Websites
  - Course: https://sites.google.com/site/biplavsrivastava/teaching/ai-csce-580-fall-2024-intro-to-ai
- Class methods
  - In-class
  - Asynchronous / Recording Online: Zoom - Blackboard

Administrative Information

- Instructor: Biplav Srivastava, Ph.D.
  - email: biplav.s@sc.edu
  - office: AI Institute, Room 515, 1112 Greene St., Columbia, 29028
- Office hours:
  - Wednesday, 3-4pm, Zoom/ Blackboard
  - Thursday, 10-11am
  - By Appointment in-person

- TA: Vishal Pallagani
  - email: vishalp@email.sc.edu
  - office: AI Institute, Room 515, 1112 Greene St., Columbia, 29028
- Office hours:
  - Tuesday, 10-11am
  - Friday, 3 – 4pm
  - By Appointment in-person

Engagement methods:
- Discussion section of Blackboard - fastest
- Email – slow response
- Meetings – slowest turnaround

Course Material

- Artificial Intelligence: A Modern Approach (Fourth edition, 2020), Stuart Russell and Peter Norvig,
  - http://aima.cs.berkeley.edu/,
  ISBN-13: 978-0134610993
- Trustworthy Machine Learning, by Kush R. Varshney, http://www.trustworthymachinelearning.com/, 2022

Open Datasets
- data.gov from ANY COUNTRY
  - Portal: https://dataportals.org/
  - US: https://www.data.gov/ or any US state
  - India: https://data.gov.in
- Text of legislations - LegiScan, https://legiscan.com/
- Kaggle datasets: https://www.kaggle.com/datasets
- Google datasets search: https://datasetsearch.research.google.com/

- AI Fairness
  - Trisha Mahoney, Kush R. Varshney, and Michael Hind, Available at: https://krvarshney.github.io/pubs/MahoneyVH2020.pdf
  - In AI We Trust: Ethics, Artificial Intelligence, and Reliability, Mark Ryan. Available at: https://link.springer.com/article/10.1007/s11948-020-00228-y
- Python for Data Analysis
  - Latest: Python for Data Analysis Book, by Wes McKinney, 2nd Edition. On Amazon at: https://www.amazon.com/gp/product/1491957662/, ISBN-13: 978-1491957660, ISBN-10: 1491957662
  - Book Data and Code Notebooks: https://github.com/wesm/pydata-book
  - 1st edition (free download): https://bedford-computing.co.uk/learning/wp-content/uploads/2015/10/Python-for-Data-Analysis.pdf

Student Assessment

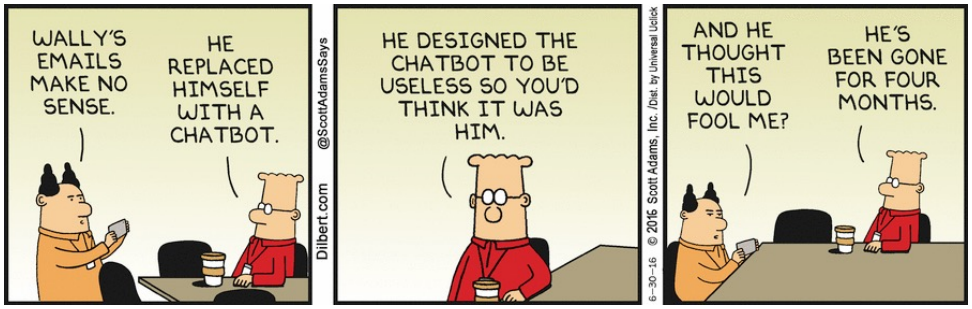
A = [900-1000]
B+ = [870-899]
B = [800-869]
C+ = [770-799]
C = [700-769]
D+ = [670-699]
D = [600-669]
F = [0-599]

Tests	Undergrad	Grad
Course Project – report, in-class presentation	600	600
Quiz – best of 3 from 4	200	200
Final Exam	200	100
Additional Final Exam – Paper summary, in-class presentation		100
Total	1000 points	1000 points

Projects

1. Project A: Model AI Assignment
   - Pick one from website: http://modelai.gettysburg.edu/
   - Complete as instructed
   - Create report, present in class (by mid-term deadline)
2. Project B: Generative AI for a given task (To Be Announced)
   - Choose a LLM/ platform
   - Create prompt
   - Fine-tune, if necessary
   - Evaluate
   - Create report, present in class (by final deadline)

AI for the Real World



Credit: Dilbert – June 30, 2016

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 (opiods), food (genetically modified)

Discussion: what, if any issue,
- in recommending courses to students?
- in finding treatment for Covid?

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

Collaborative Assistants

- Conversation agents and interfaces (chatbots) are getting easy to build and deploy
  - Can be text-based or speech-based
  - Usually multi-modal (i.e, involving text, speech, vision, document, maps)
- Current chatbots typically interact with a single user at a time and conduct
  - Informal conversation, or
  - Task-oriented activities like answer a user's questions or provide recommendations

Demonstrations

- *Eliza*, http://www.manifestation.com/neurotoys/eliza.php3
- *Mitsuku*, https://www.pandorabots.com/mitsuku/
- ChatGPT, https://openai.com/blog/chatgpt

Exercise: Session with ChatGPT

- Ask questions about Water usage
  - Experience
- Ask questions about Finance
  - Experience

- Hint:
  - Demand / supply questions: “can I drink water of Lake Murray”?, “will US have money to pay debt next year”
  - Decision questions: “which water should I choose between a bottled one and tap”?
  - Factoid questions: “is pH of 7 good for drinking water?”

Exercise: Solving Games with AI

- Popular way to learn AI is via games
  - https://github.com/biplav-s/course-ai-tai-f23/blob/main/sample-code/Class1-games.md

Concluding Section

Lecture 1: Concluding Comments

- We did a quick overview of
  - AI
  - Trust issues
- Course will focus on
  - Practical methods to derive insights from data, especially structured data and text
  - Evaluation will be by via project, paper and quizzes
- Exciting techniques to learn to impact the world around us

About Next Lecture – Lecture 2

Lecture 2: Data

- Structured data
- Mode
  - Text
  - Speech
  - Visual
  - Mixed : multi-modal
- Processing Methods and Applications
- Trust issues: data privacy