

CSCE 590-1: Trusted AI

Lecture 26: Graduate Paper Presentations

PROF. BIPLAV SRIVASTAVA, AI INSTITUTE

18TH NOV, 2021

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

Organization of Lecture 26

- Introduction Segment
 - Recap from last lecture
 - Paper selection by graduate students
- Main Segment
 - Paper selection by graduate students
- Concluding Segment
 - About next lecture – Lecture 27
 - Ask me anything

Introductory Segment

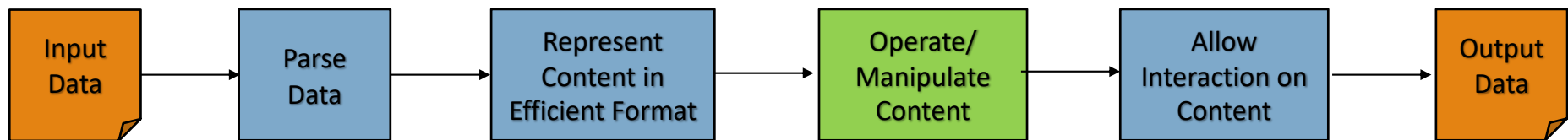
Schedule Snapshot

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 Lecture 25

- We looked at what it means for AI to collaborate with Humans
 - Experience with other technologies shows that human has to control technology (AI)
- Collaborative assistant is a promising avenue to deliver decision support
 - But can sources of bias
- Rating of chatbots can be a promising direction to control behavior
- Looked at initial material on adversarial learning

Main Segment



Lecture 26:

Paper Presentations – Graduate Students

- Select a paper appearing at a top-AI or data conference (AAAI, IJCAI, NeurIPS, ICML, ACL, CVPR, CIKM, WWW ... or discuss with instructor) during 2019-2021
- Present in class for 9 + 3 minutes of Q/A
- Things to cover
 - **Summary:** problem, solution, related work, contributions
 - **Opinion:** What you liked or did not like
- Put paper details in Google sheet
- Dates
 - In-class presentation on Nov 18, 2021 (Thursday)
 - 1-page written report on Nov 23, 2021 (Tuesday)

Papers List

1. From ethics washing to ethics bashing: a view on tech ethics from within moral philosophy;
<https://dl.acm.org/doi/abs/10.1145/3351095.3372860>;
<https://facctconference.org/2020/acceptedpapers.html>
2. CheXbert: Combining Automatic Labelers and Expert Annotations for Accurate Radiology Report Labeling Using BERT, <https://aclanthology.org/2020.emnlp-main.117.pdf>
3. GNNExplainer: Generating Explanations for Graph Neural Networks <https://arxiv.org/pdf/1903.03894v4.pdf>
4. Mitigating Political Bias in Language Models Through Reinforced Calibration <https://arxiv.org/abs/2104.14795>
5. Opportunities and Challenges in Explainable Artificial Intelligence (XAI): A Survey
(<https://arxiv.org/abs/2006.11371>)
6. Ethically Compliant Sequential Decision Making <https://www.aaai.org/AAAI21Papers/AAAI-3534.SvegliatoJ.pdf>
7. The Role of Domain Expertise in User Trust and the Impact of First Impressions with Intelligent Systems
<https://arxiv.org/pdf/2008.09100.pdf>

Concluding Segment

Lecture 26: Concluding Comments

- A wide spectrum of papers

About Next Lecture – Lecture 27

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 27: Socio-Technical Environment

- Emerging Standards and Laws
- Data Privacy considerations
- Quiz 4 (in-class)