



CSE 591 Human-Aware AI

Fall 2018 ASU

Course Schedule

For the **required** papers, students will need to post reflections by Thursday night preceding the class.

1. 8/17: Intro class.
 1. [Syllabus](#)
 1. (optional) : [AI Challenges in Human-Robot Cognitive Teaming](#)
 2. (optional) [Explainable Agency for Intelligent Autonomous Systems](#)
 2. [Slides used in the class](#); [videos of the class](#)
2. 8/24: **Topic: Explicability/Predictability** [[Rao's slides for the class](#)][[Rao's audio](#)] [[Anagha's slides](#)] [Class notes [@21](#)]
 1. [Required] [Plan Explicability and Predictability for Robot Task Planning](#).
[Reflections [@13](#)]
 2. [Required] [Generating Plans that Predict themselves](#) [Reflections [@14](#)]
 1. (Optional) [Legibility and Predictability in Robot Motion Planning](#)
(This one is an [overview paper of that group's work](#))
 2. (Optional) [Action selection for transparent planning](#)
3. 8/31: **Topic: Explanations** [[Rao's background slides](#)]
 1. [Required] [Explainable AI: Beware of Inmates Running the Asylum](#) [Reflections [@28](#)] [[Sarath's slides on the paper](#)]

2. [Required] [Plan Explanations as Model Reconciliation: Moving Beyond Explanation as Soliloquy](#) [Reflections @29][Tathagata's slides]
 1. (Optional) [Much longer version of Tim Miller's paper on Explanations in AI & Social Science](#)
 2. (Optional--but easy read) [Towards A Rigorous Science of Interpretable Machine Learning](#)
4. 9/7 **Topic: Explanations--other perspectives**
 1. [Required] [Verbalization: Narration of Autonomous Robot Experience](#) [Reflections @35; Rao's comments] [Daniel's presentation]
 1. (Optional) [Interpretability of a Service Robot: Enabling User Questions and Checkable Answers.](#)
 2. [Required] [The Challenge of Crafting Intelligible Intelligence](#) . [Reflections @36; Rao's comments] [Sriram's presentation]
 1. (Optional) [The Mythos of Model Interpretability](#)
 2. (Optional) [Recent Algorithms for Interpretable Machine Learning](#) (Cynthia Rudin's ICML-18 talk that I mentioned in class)
 3. (Optional) [Inherent Trade-Offs in the Fair Determination of Risk Scores](#) (Kleinberg paper mentioned in class)
5. 9/14: **[Topic: Providing assistance & Getting out of the way]**
 1. [Required] [A Decision-theoretic Model of Assistance](#) [Reflections @44; Rao's comments] [Slides]
 1. (Optional) [Longer version of Decision-theoretic Model of Assistance](#)
 2. (Optional) [The "Serendipity" paper that got mentioned many times during the discussion](#)
 2. [Required] [Grandpa Hates Robots — Interaction Constraints for Planning in Inhabited Environments](#) [Reflections @45; Rao's comments]
 1. (Optional) [Toward Human-Aware Robot Task Planning](#)
 2. (Optional) [Human-Aware Motion Planning](#) (notice the similarity between human-aware task vs. motion planning)
6. 9/21:
 1. **[Topic: Explanations in MDPs]** (Required) [Minimal Sufficient Explanations for Factored Markov Decision Processes](#) (Alberto)[Reflections @49; Rao's comments]
 1. (Optional) [MDP Explanations in Course Scheduling domain](#) (2013 paper that follows up on the previous paper; by different authors)
 2. (Optional) [Explainable Planning](#) (Similar to the one above, but geared to factored deterministic planning)

2. **[Topic: Providing assistance]** (Required) [Planning for Serendipity](#)
[Reflections [@50](#); [Rao's comments](#)] [[slides from tathagata](#)]
7. 9/28 **[Topic: Human Factors]** *Class Project Proposals Due*
 1. [Required] [Human-Automation Interaction](#) [Shaojia] [Reflections [@60](#)]
 2. [Required] [RADAR - A Proactive Decision Support System for Human-in-the-Loop Planning](#) [Sailik][Reflections [@61](#)]
8. 10/5: **[Special class] 12:30pm: Dan Weld's Distinguished Lecture talk on Intelligible AI in BYAC110; 2-3pm: Discussion in MI-09 [Lecture video]**
9. 10/12[Topic: Affective computing/Mixed-Initiative Interfaces]
 1. [A domain-independent framework for modeling emotion](#)
(Sarah)[[Slides](#)][Reflections [@67](#)]
 2. [Principles of Mixed Initiative Interfaces](#) (Sunny's [Slides](#)) [Reflections [@68](#)]
 1. [Optional] [Jeff Heer's Agency & Automation \[Slide Deck\]](#)
10. 10/15: Make up class for 10/19 **NOTE SPECIAL ROOM AND TIME: MONDAY 1pm BY 420**
 1. [Rao will "lecture"] [[Slides](#)] [[Audio recording](#)]
11. 10/29: <Make up class not held>
12. 11/2: David Smith in town..
 1. [Topic: Mixed-initiative planning] [Planning as an iterative process](#) [Aditya]
 2. [Topic: Interpretable RL] [Interpretable Reinforcement Learning](#) {Manish}[[Slides](#)]
13. 11/9: **Note: we meet at 12 noon in BYENG 210**
 1. [Topic: Preference Learning] [Cooperative Inverse Reinforcement Learning](#) [Utkarsh Soni] [[Slides](#)]
 1. Optional: [Off-Switch Game](#) (they talk about making the robot has built-in uncertainty about the reward that it learned from the human)
 2. (Optional: [Impossibility of Preference Learning](#))
 2. [Topic: Learning Mental Models] [Machine Theory of Mind](#) (Kathryn Wifvat) [[Slides](#)]
 1. (Optional) [Plan Recognition with Bayesian Theory of Mind](#)
14. 11/16: **Normal Class Time/venue**
 1. [Towards Minimal Explanations of Unsynthesizability for High-Level Robot Behaviors](#) [Zahra]
 2. [Socially intelligent robots: dimensions of human–robot interaction](#) [Jay Shah]
15. 11/30: (No Class; Instead we meet on 12/2 Sunday; and 12/5 Wednesday for presentations)

Student Projects

Sunday

- Slot 1 - Xiangyu Guo [[PDF_Version](#)] [[PPTX_Version](#)] [[Report](#)]
Slot 2 - Chetan [[presentation.pdf](#)] [[ChetanEEGReport.pdf](#)]
Slot 3 - Kuldeep/Utkarsh [[slides](#)] [[Report.pdf](#)]
Slot 4 - Sarath/Aditya/Alberto [[Report](#)] [[Slides](#)]  [Accepted to IJCAI 2019](#) 
Slot 5 - Anagha [[Report](#)] [[Slides](#)]
Slot 6 -
Slot 7 - Manish/Jay/Shaojia [[Report](#)] [[Slides](#)] [[Google Slides](#)]
Slot 8 - Sriram [[Report](#)][[Slides](#)] [[slides google drive](#)]
Slot 9 - Sunny [[HAAI_PROJECT_REPORT_1.pdf](#)]
Slot 10 -
Slot 11 - Kathryn [[slides.pdf](#)] [[Report.pdf](#)]

Wednesday

- Slot 1 - Sailik/Zahra [[report](#)][[slides](#)]  [Accepted to Trust in Agent Societies Workshop/AAMAS 2019](#) 
Slot 2 - Hari/Daniel [[Presentation_Slides.pdf](#)][[Project_Report.pdf](#)]
Slot 3 - Venkatesh/Samarth [[Slides.pptx](#)] [[Report.pdf](#)]
Slot 4 - Brandon/Sarah [[CSE_591_Final_Report.pdf](#)][[slides](#)]
Slot 5 - Yantian [[HAAI_Slides_1.pptx](#)][[report_HAAI.pdf](#)]
Slot 6 - Sachin [[Report](#)][[Slides](#)] [[pdf](#)] [[ppt](#)]