HRI Ethics Integration

Pancho Cabrera & Lynn Kirabo

Course Overview

(16-467) Human Robot Interaction with Prof. Henny Admoni

This course provides an introductory-level overview of the field of HRI.

It is primarily lecture-based, with in-class participatory mini-projects, homework assignments, a final exam, and a group term project that will enable students to put theory to practice.

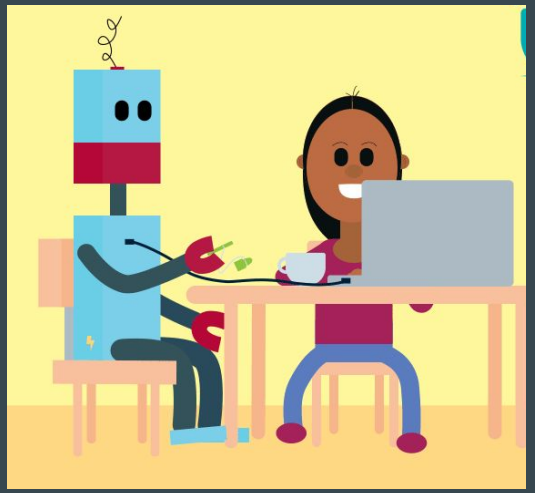


Image Credit: Jennifer Perdomo, Travels with Robots

Course Constraints

- **Course structure** i.e., the first third focusing on foundations and the last two thirds on the integration of those foundations.

- Undergraduate students from **both technical and non-technical** backgrounds. Class size ~ **60 students** enrolled.

- **TA Grading Workload** (~25:1 student-to-TA ratio)

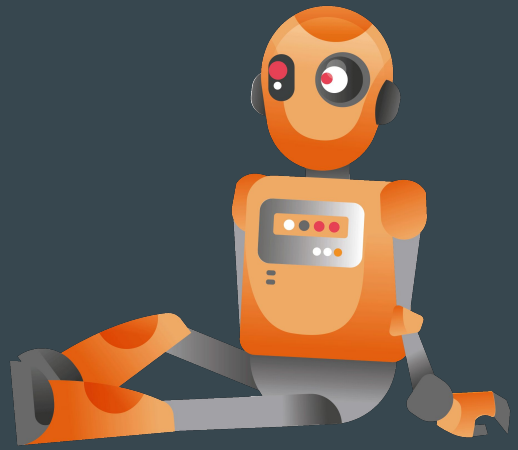


Image Credit: Giphy, Pixup Media

Learning Objectives

- Goal 1: **Identify** any **ethical topics** embedded in HRI concepts, applications, or conversations.
- Goal 2: Develop **vocabulary to express ethical concerns** they observe.
- Goal 3: Develop the ability to **identify comparisons between ethical events in current news and their future implications**.

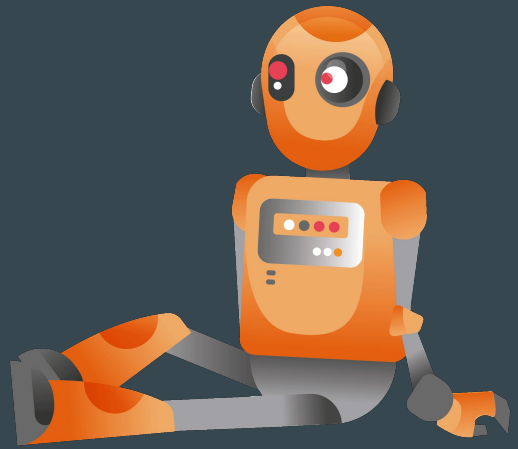


Image Credit: Giphy, Pixup Media

Our strategy: **introduce ethics discussion** that **builds up to** the main ethics lecture at the end of the semester

News Articles

Videos

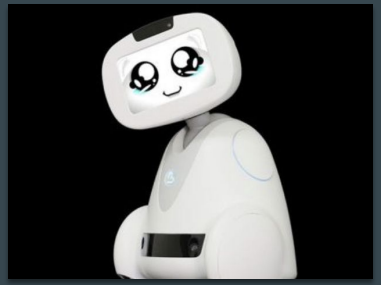
Online Discussion Posts

Quiz

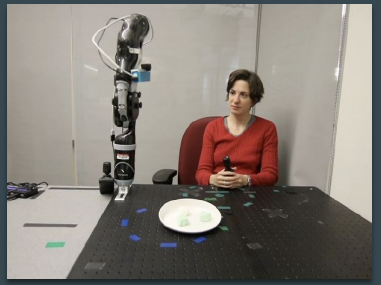
Individual reflections

Debate & Class Discussion

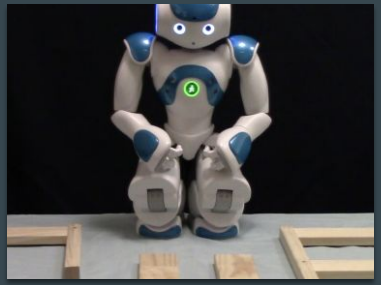
7 Lectures Selected + Final Project



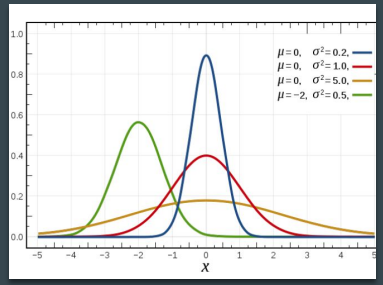
Introduction



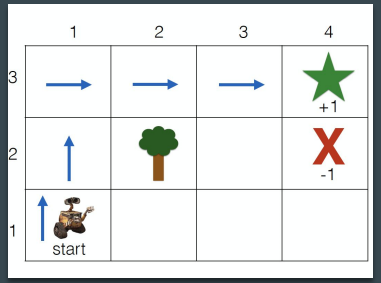
Autonomy



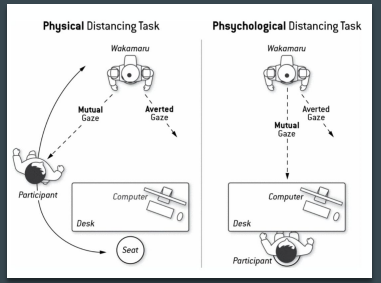
Verbal & Emotions



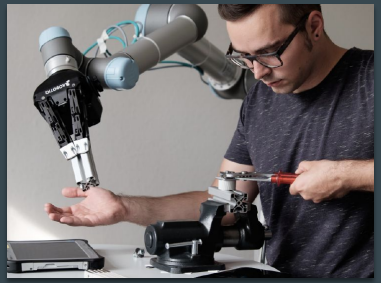
Data Analysis



MDPs



Social Navigation



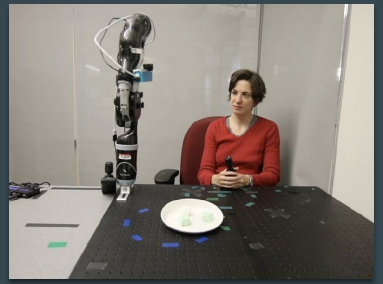
Collaboration

Extra Credit

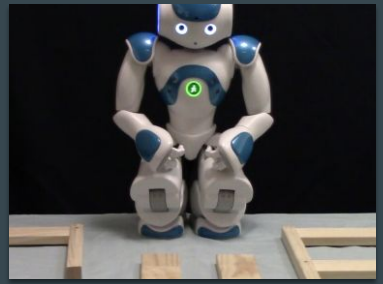
Image Credit: Prof. Admoni S'20 Slides & Illustration by Freepik Storyset

7 Lectures Selected

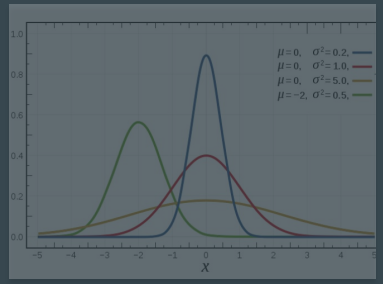
Introduction



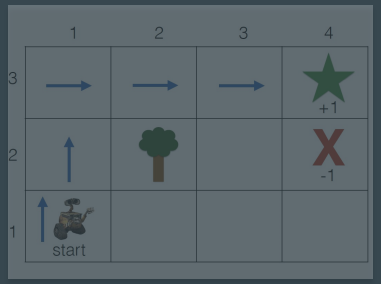
Autonomy



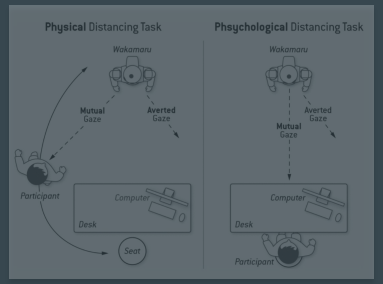
Verbal & Emotions



Data Analysis



MDPs



Social Navigation

Collaboration

Final Project

Image Credit: Prof. Admoni S'20 Slides & Illustration by Freepik Storyset

Robots of the Future

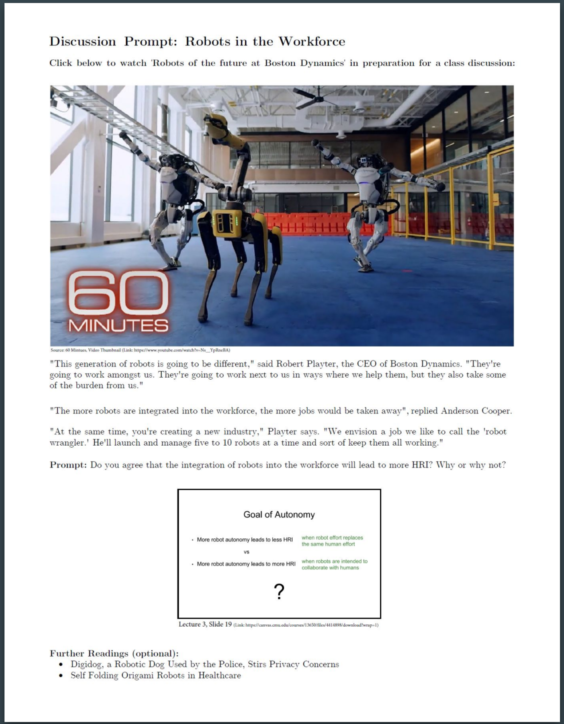
Course relevance: Lecture 3 - Autonomy

Learning objectives: Goal 1 + Goal 3

Implementation:

- Video Clip
- Discussion Post
- In-class Large Group Discussion

In-class Timing: 5 mins



Discussion Prompt: Robots in the Workforce

Click below to watch 'Robots of the future at Boston Dynamics' in preparation for a class discussion:

Source: 60 Mintues, Video Thumbnail (Link: https://www.youtube.com/watch?v=Ns__YpRneBA)

"This generation of robots is going to be different," said Robert Playter, the CEO of Boston Dynamics. "They're going to work amongst us. They're going to work next to us in ways where we help them, but they also take some of the burden from us."

"The more robots are integrated into the workforce, the more jobs would be taken away", replied Anderson Cooper.

"At the same time, you're creating a new industry," Playter says. "We envision a job we like to call the 'robot wrangler.' He'll launch and manage five to 10 robots at a time and sort of keep them all working."

Prompt: Do you agree that the integration of robots into the workforce will lead to more HRI? Why or why not?

Lecture 3, Slide 19 (Link: https://canvas.cmu.edu/courses/13650/files/4414898/download?wrap=1)

Further Readings (optional):

- Digidog, a Robotic Dog Used by the Police, Stirs Privacy Concerns
- Self Folding Origami Robots in Healthcare

Racial Divide in Speech-Recognition Systems

Course relevance: Lecture 8 - Emotions & Verbal Communication

Learning objectives: Goal 2

Implementation:

- Article Reading
- Individual Reflection
- In-class Large Group Discussion

In-class Timing: 5 - 10 mins

Reflection: Racial Divide in Speech-Recognition Systems

Read the article 'There Is a Racial Divide in Speech-Recognition Systems, Researchers Say'.

There Is a Racial Divide in Speech-Recognition Systems, Researchers Say

Technology from Amazon, Apple, Google, IBM and Microsoft misidentified 35 percent of words from people who were black. White people fared much better.

Amazon's Echo device is one of many similar gadgets on the market. Researchers say there is a racial divide in the usefulness of speech recognition systems. Grant Hindsley for The New York Times

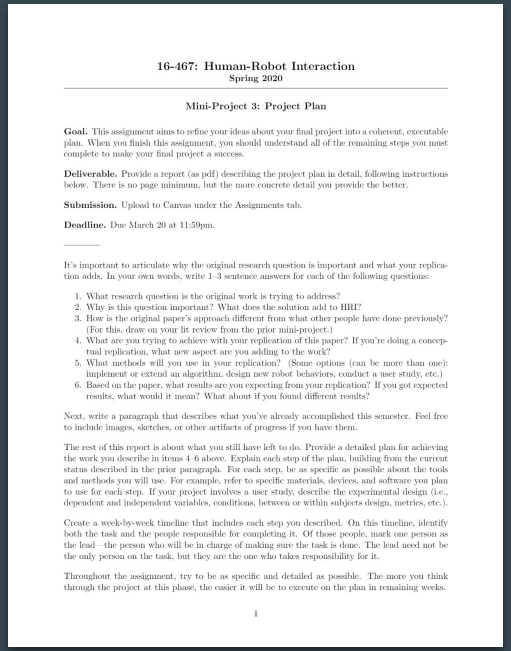
Source: New York Times(Link: https://www.nytimes.com/2020/03/23/technology/speech-recognition-bias-apple-amazon-google.html)

Eric Schmidt is quoted as saying, 'We know the data has bias in it. You don't need to yell that as a new fact. Humans have bias in them, our systems have bias in them. The question is: What do we do about it?'

Prompt: Do you agree that the challenge of bias is an application problem? Why/Why not?

Final Project

Mini-Project 3: Project Plan



16-467: Human-Robot Interaction
Spring 2020

Mini-Project 3: Project Plan

Goal. This assignment aims to refine your ideas about your final project into a coherent, executable plan. When you finish this assignment, you should understand all of the remaining steps you must complete to make your final project a success.

Deliverable. Provide a report (as pdf) describing the project plan in detail, following instructions below. There is no page minimum, but the more concrete detail you provide the better.

Submission. Upload to Canvas under the Assignments tab.

Deadline. Due March 20 at 11:59pm.

It's important to articulate why the original research question is important and what your replication adds. In your own words, write 1–3 sentence answers for each of the following questions:

1. What research question is the original work is trying to address?
2. Why is this question important? What does the solution add to HRI?
3. How is the original paper's approach different from what other people have done previously? (For this, draw on your lit review from the prior mini-project.)
4. What are you trying to achieve with your replication of this paper? If you're doing a conceptual replication, what new aspect are you adding to the work?
5. What methods will you use in your replication? (Some options (can be more than one): implement or extend an algorithm, design new robot behaviors, conduct a user study, etc.)
6. Based on the paper, what results are you expecting from your replication? If you got expected results, what would it mean? What about if you found different results?

Next, write a paragraph that describes what you've already accomplished this semester. Feel free to include images, sketches, or other artifacts of progress if you have them.

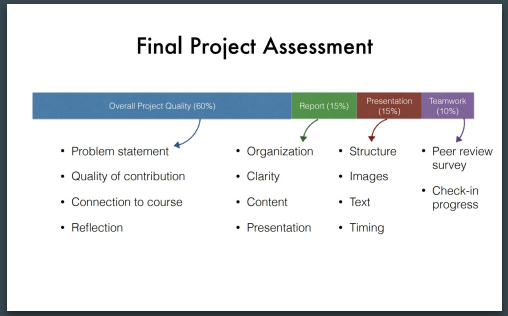
The rest of this report is about what you still have left to do. Provide a detailed plan for achieving the work you describe in items 4–6 above. Explain each step of the plan, building from the current status described in the prior paragraph. For each step, be as specific as possible about the tools and methods you will use. For example, refer to specific materials, devices, and software you plan to use for each step. If your project involves a user study, describe the experimental design (i.e., dependent and independent variables, conditions, between or within subjects design, metrics, etc.).

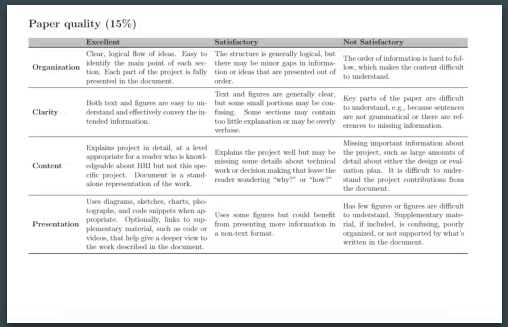
Create a week-by-week timeline that includes each step you described. On this timeline, identify both the task and the people responsible for completing it. Of those people, mark one person as the lead—the person who will be in charge of making sure the task is done. The lead need not be the only person on the task, but they are the one who takes responsibility for it.

Throughout the assignment, try to be as specific and detailed as possible. The more you think through the project at this phase, the easier it will be to execute on the plan in remaining weeks.

1

Final Project Report





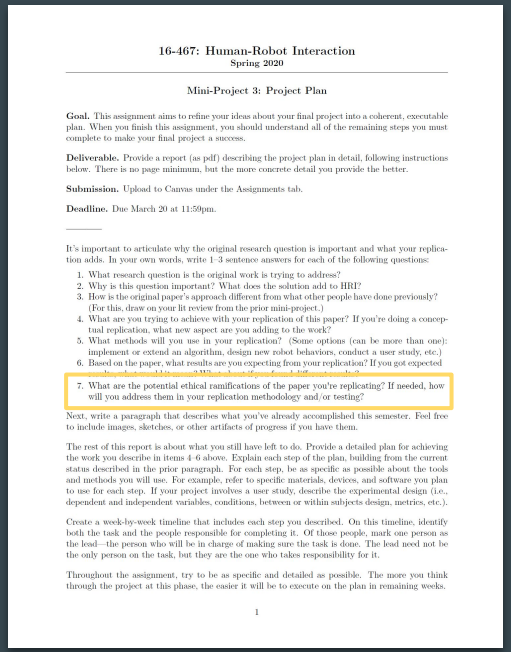
Paper quality (15%)

	Excellent	Satisfactory	Not Satisfactory
Organization	Clear, logical flow of ideas. Easy to identify the main point of each section. Each part of the project is fully presented in the document.	The structure is generally logical, but there may be minor gaps in information or ideas that are presented out of order.	The order of information is hard to follow, which makes the content difficult to understand.
Clarity	Both text and figures are easy to understand and effectively convey the intended information.	Text and figures are generally clear, but some small portions may be confusing. Some sections may contain too little explanation or may be overly verbose.	Key parts of the paper are difficult to understand, e.g., because sentences are not grammatical or there are references to missing information.
Content	Explains project in detail, at a level appropriate for a reader who is knowledgeable about HRI but not this specific project. Document is a standalone representation of the work.	Explains the project well but may be missing some details about technical work or decision making that leave the reader wondering "why?" or "how?"	Missing important information about the project, such as large amounts of detail about either the design or evaluation plan. It is difficult to understand the project contributions from the document.
Presentation	Uses diagrams, sketches, charts, photographs, and code snippets when appropriate. Optionally, links to supplementary material, such as code or videos, that help give a deeper view to the work described in the document.	Uses some figures but could benefit from presenting more information in a non-text format.	Has few figures or figures are difficult to understand. Supplementary material, if included, is confusing, poorly organized, or not supported by what's written in the document.

Image Credit: Prof. Admoni S'20

Final Project

Mini-Project 3: Project Plan



16-467: Human-Robot Interaction
Spring 2020

Mini-Project 3: Project Plan

Goal. This assignment aims to refine your ideas about your final project into a coherent, executable plan. When you finish this assignment, you should understand all of the remaining steps you must complete to make your final project a success.

Deliverable. Provide a report (as pdf) describing the project plan in detail, following instructions below. There is no page minimum, but the more concrete detail you provide the better.

Submission. Upload to Canvas under the Assignments tab.

Deadline. Due March 20 at 11:59pm.

It's important to articulate why the original research question is important and what your replication adds. In your own words, write 1–3 sentence answers for each of the following questions:

1. What research question is the original work is trying to address?
2. Why is this question important? What does the solution add to HRI?
3. How is the original paper's approach different from what other people have done previously? (For this, draw on your lit review from the prior mini-project.)
4. What are you trying to achieve with your replication of this paper? If you're doing a conceptual replication, what new aspect are you adding to the work?
5. What methods will you use in your replication? (Some options (can be more than one): implement or extend an algorithm, design new robot behaviors, conduct a user study, etc.)
6. Based on the paper, what results are you expecting from your replication? If you got expected
7. What are the potential ethical ramifications of the paper you're replicating? If needed, how will you address them in your replication methodology and/or testing?

Next, write a paragraph that describes what you've already accomplished this semester. Feel free to include images, sketches, or other artifacts of progress if you have them.

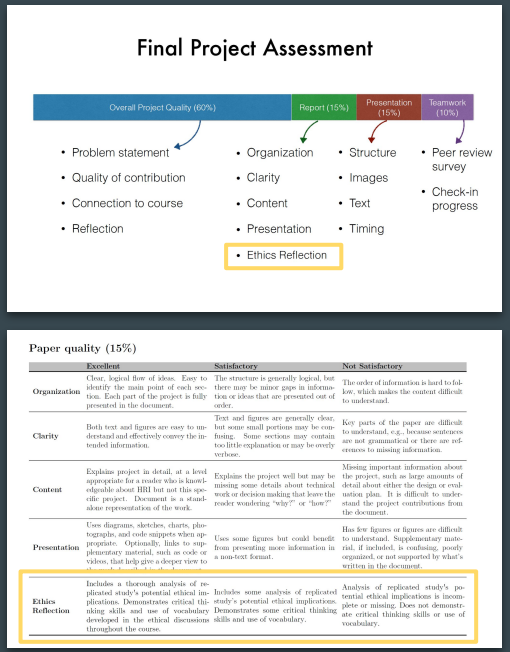
The rest of this report is about what you still have left to do. Provide a detailed plan for achieving the work you describe in items 4–6 above. Explain each step of the plan, building from the current status described in the prior paragraph. For each step, be as specific as possible about the tools and methods you will use. For example, refer to specific materials, devices, and software you plan to use for each step. If your project involves a user study, describe the experimental design (i.e., dependent and independent variables, conditions, between or within subjects design, metrics, etc.).

Create a week-by-week timeline that includes each step you described. On this timeline, identify both the task and the people responsible for completing it. Of those people, mark one person as the lead—the person who will be in charge of making sure the task is done. The lead need not be the only person on the task, but they are the one who takes responsibility for it.

Throughout the assignment, try to be as specific and detailed as possible. The more you think through the project at this phase, the easier it will be to execute on the plan in remaining weeks.

1

Final Project Report



Final Project Assessment

Paper quality (15%)

	Excellent	Satisfactory	Not Satisfactory
Organization	Clear, logical flow of ideas. Easy to identify the main point of each section. Each part of the project is fully presented in the document.	The structure is generally logical, but there may be minor gaps in information or ideas that are presented out of order.	The order of information is hard to follow, which makes the content difficult to understand.
Clarity	Both text and figures are easy to understand and effectively convey the intended information.	Text and figures are generally clear, but some small portions may be confusing. Some sections may contain too little explanation or may be overly verbose.	Key parts of the paper are difficult to understand, e.g., because sentences are not grammatical or there are references to missing information.
Content	Explains project in detail, at a level appropriate for a reader who is knowledgeable about HRI but not this specific project. Document is a standalone representation of the work.	Explains the project well but may be missing some details about technical work or decision making that leave the reader wondering "why?" or "how?"	Missing important information about the project, such as large amounts of detail about either the design or evaluation plan. It is difficult to understand the project contributions from the document.
Presentation	Uses diagrams, sketches, charts, photographs, and code snippets when appropriate. Optionally, links to supplementary material, such as code or videos, that help give a deeper view to	Uses some figures but could benefit from presenting more information in a non-text format.	Has few figures or figures are difficult to understand. Supplementary material, if included, is confusing, poorly organized, or not supported by what's written in the document.
Ethics Reflection	Includes a thorough analysis of replicated study's potential ethical implications. Demonstrates critical thinking skills and use of vocabulary developed in the ethical discussions throughout the course.	Includes some analysis of replicated study's potential ethical implications. Demonstrates some critical thinking skills and use of vocabulary.	Analysis of replicated study's potential ethical implications is incomplete or missing. Does not demonstrate critical thinking skills or use of vocabulary.

Image Credit: Prof. Admoni S'20

Extra Credit

Robot Rights & Citizenship

Social Justice in Tech & Climate Impact of ML

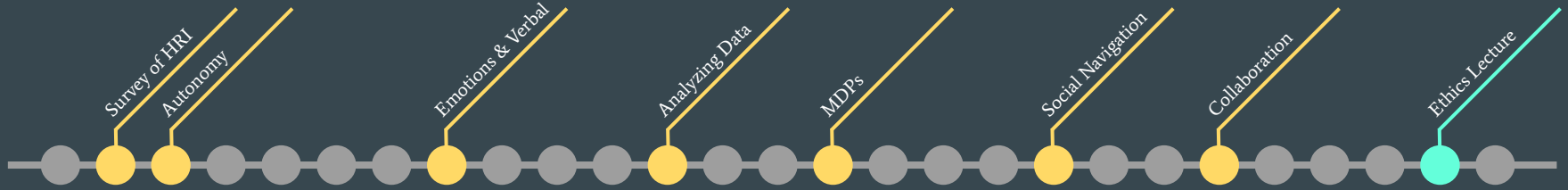
Hypothetical: Trust in Robots

Algorithmic Bias

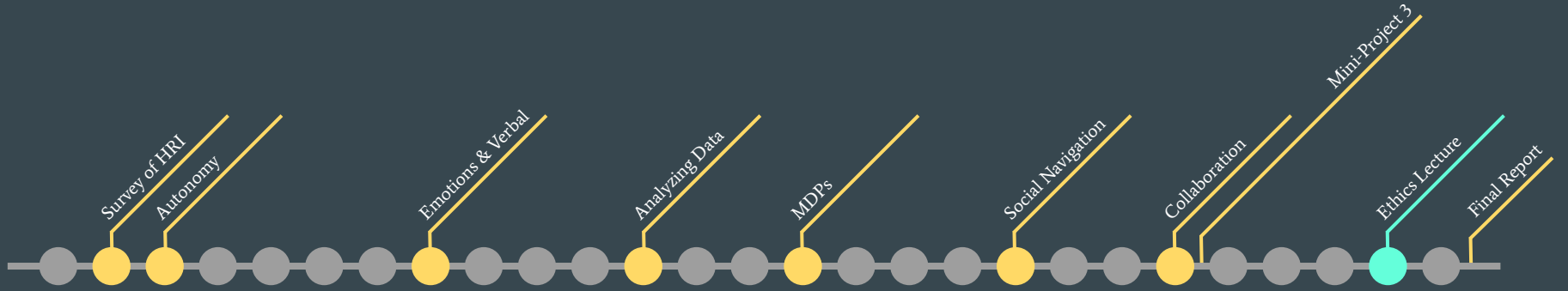
Image Credit: Tony Robbins, Wikipedia, Google Images, New York Times

Overview

Overview


Survey of HRI
Autonomy
Emotions & Verbal
Analyzing Data
MDPs
Social Navigation
Collaboration
Ethics Lecture

Overview


Survey of HRI
Autonomy
Emotions & Verbal
Analyzing Data
MDPs
Social Navigation
Collaboration
Mini-Project 3
Ethics Lecture
Final Report

Overview


Survey of HRI
Autonomy
Extra Credit: Sophia
Emotions & Verbal
Analyzing Data
MDPs
Extra Credit: Policy
Social Navigation
Collaboration
Mini-Project 3
Ethics Lecture
Extra Credit: Gebru
Extra Credit: Bias
Final Report

Questions?

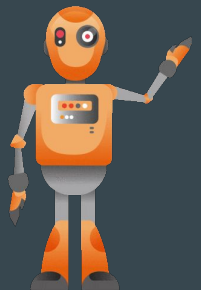


Image Credit: Giphy, Pixup Media