

# COGS138: Neural Data Science

## **Lecture 1**

C. Alex Simpkins, PhD

UCSD Dept. of Cognitive Science, Spring 2023

[http://casimpkinsjr.radiantdolphinpress.com/pages/cogs138\\_sp23](http://casimpkinsjr.radiantdolphinpress.com/pages/cogs138_sp23)

[rdrobotics@gmail.com](mailto:rdrobotics@gmail.com) | [csimpkinsjr@ucsd.edu](mailto:csimpkinsjr@ucsd.edu)

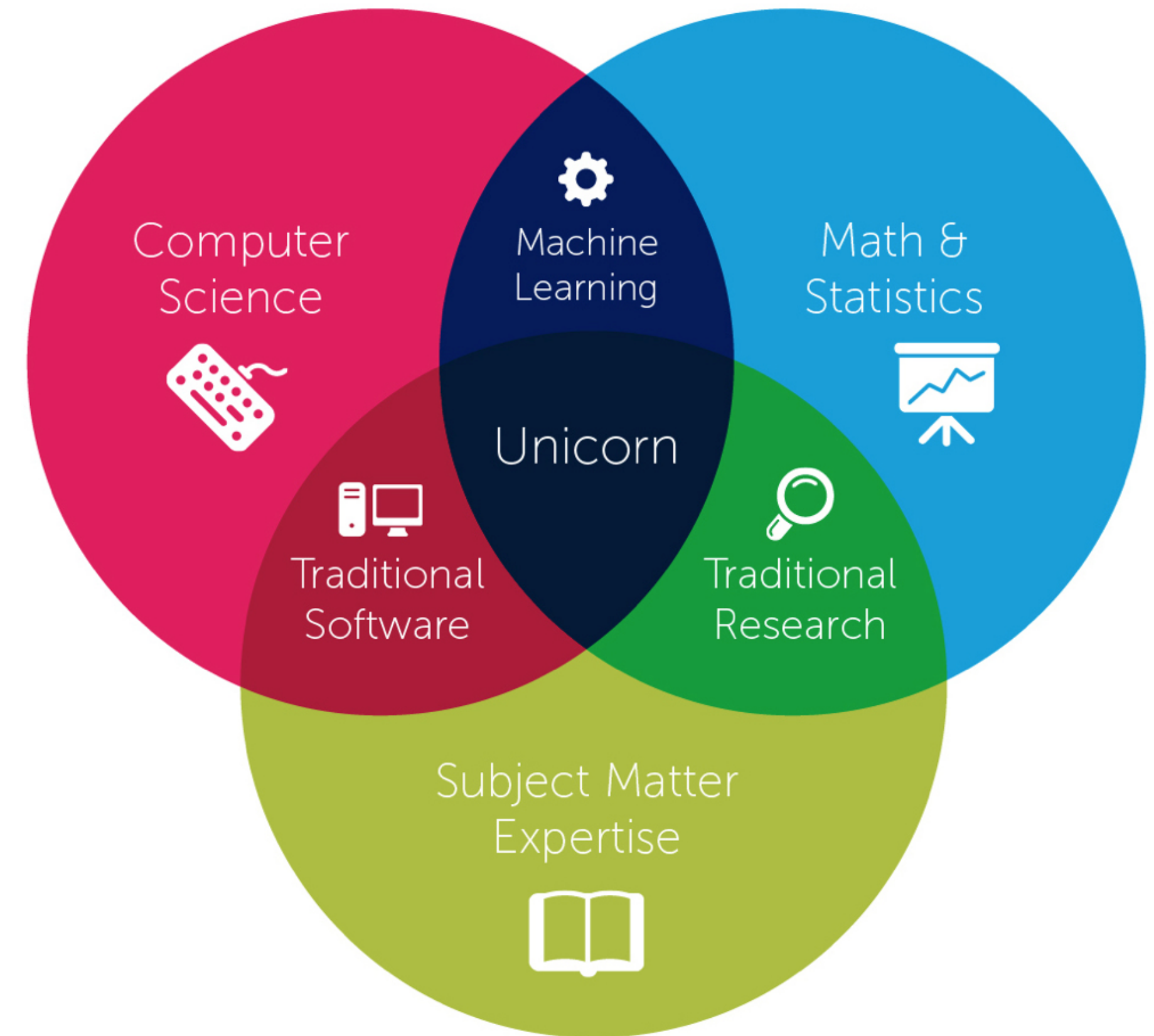
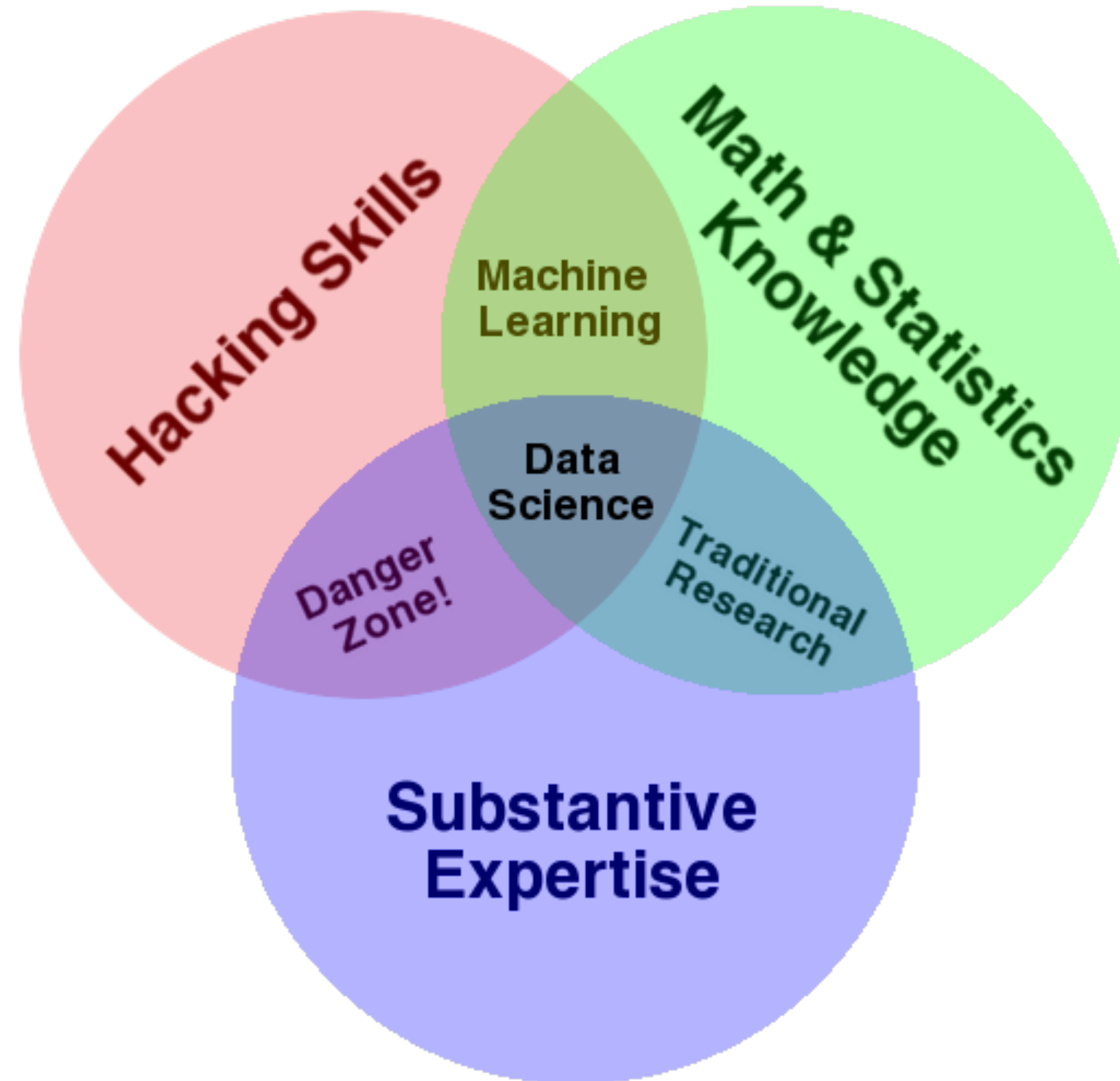
# Plan for today

- Introduction
- Course outline and structure
- What is data science?
- What is neural data science?
- What will we be doing?
- Part I review of datahub, jupyter

# Course structure

- Course structure
- Lectures (weekly) - at times one lecture one workshop day,
  - Location: SOLIS 109
  - Time: TuTh 12:30-1:50
- Readings + reading quizzes (5-10)
- Assignments (~5, datahub)
- Project (1, github)
- Lecture quizzes (~5, canvas)
- No final
- Piazza

# What is data science?



Copyright © 2014 by Steven Geringer Raleigh, NC.  
Permission is granted to use, distribute, or modify this image,  
provided that this copyright notice remains intact.

# Defining Data Science

*a "concept to unify statistics, data analysis, machine learning and their related methods" in order to "understand and analyze actual phenomena" with data.<sup>[3]</sup> It employs techniques and theories drawn from many fields within the context of mathematics, statistics, information science, and computer science. -Wikipedia*

*"This coupling of scientific discovery and practice involves the collection, management, processing, analysis, visualization, and interpretation of vast amounts of heterogeneous data associated with a diverse array of scientific, translational, and interdisciplinary actions." -David Donoho ("50 years of Data Science)*

*"an emerging discipline that draws upon knowledge in statistical methodology and computer science to create impactful predictions and insights for a wide range of traditional scholarly fields" - from a panel Rafael Irizarry moderated, shared on SimplyStatistics ("The role of academia in data science education")*

*"an umbrella term used by organizations to describe the processes used to extract value from data" -Rafael Irizarry's personal definition in "The role of academia in data science education"*

*"The study of how the quantification of observable phenomena can lead to human understanding of the processes giving rise to those phenomena—or even the ability to predict future outcomes absent human understanding—and why certain phenomena require more or less data to lead to human understanding and/or prediction accuracy". -Brad Voytek's definition*

**"The scientific process of extracting value from data"**

Data scientists ask  
interesting questions  
& answer them with  
data

What is NEURAL Data Science?

# What will we be doing?

Learn how to:

- Think from a “data first” perspective: what data *would* you need to answer your scientific questions of interest?
- Develop hypotheses specific to big data environments in neuroscience.
- Work with many different neuroscience data types that might include data on behavior, brain structure and connectivity, single-unit spiking, field potential, gene expression, and even text-mining of the peer-reviewed neuroscientific literature.
- Read and analyze data stored in standard formats (*e.g.*, Neurodata Without Borders and Brain Imaging Data Structure).
- Integrate multiple heterogeneous datasets in scientifically meaningful ways.
- Choose statistical model(s) informed by the underlying data.
- Design a big data experiment and integrate data from multiple open data sources.
- Consider alternative hypotheses and assess for spurious correlations and results.



# Data Science as the Extended Mind

# Introduction to DataHub

- datahub.ucsd.edu
- Logging in
- Navigating
  - Intro to file structures and how they relate to your computer
- Upload
- Download
- Rename files
- Make folders
- Delete
- Submitting assignments, fetching assignments
- Validating!

The screenshot shows the homepage of the UC San Diego Jupyterhub (Data Science) Platform. At the top, there is a navigation bar with the text "DATA SCIENCE / MACHINE LEARNING PLATFORM" on the left and the "UC San Diego" logo on the right. Below this is a blue banner with the text "Information Technology Services - Academic Technology Services" and links for "Help" and "FAQ". The main content area features a large image of a Jupyterhub logo and a photograph of a large, colorful sculpture of a bear. A yellow "Log In" button is prominently displayed, with the text "Registered Users" and "username@ucsd.edu" below it. Below the image, the title "UC San Diego Jupyterhub (Data Science) Platform" is displayed in blue. A note reads: "If you are unable to log in: Please try opening a private/incognito window in your browser | FAQ". The page is divided into two columns: "Student Resources" and "Instructor Resources".

DATA SCIENCE / MACHINE LEARNING PLATFORM

UC San Diego

Information Technology Services - Academic Technology Services

Help ▾ FAQ

Log In

Registered Users  
"username@ucsd.edu"

## UC San Diego Jupyterhub (Data Science) Platform

*If you are unable to log in: Please try opening a private/incognito window in your browser | [FAQ](#)*

### Student Resources

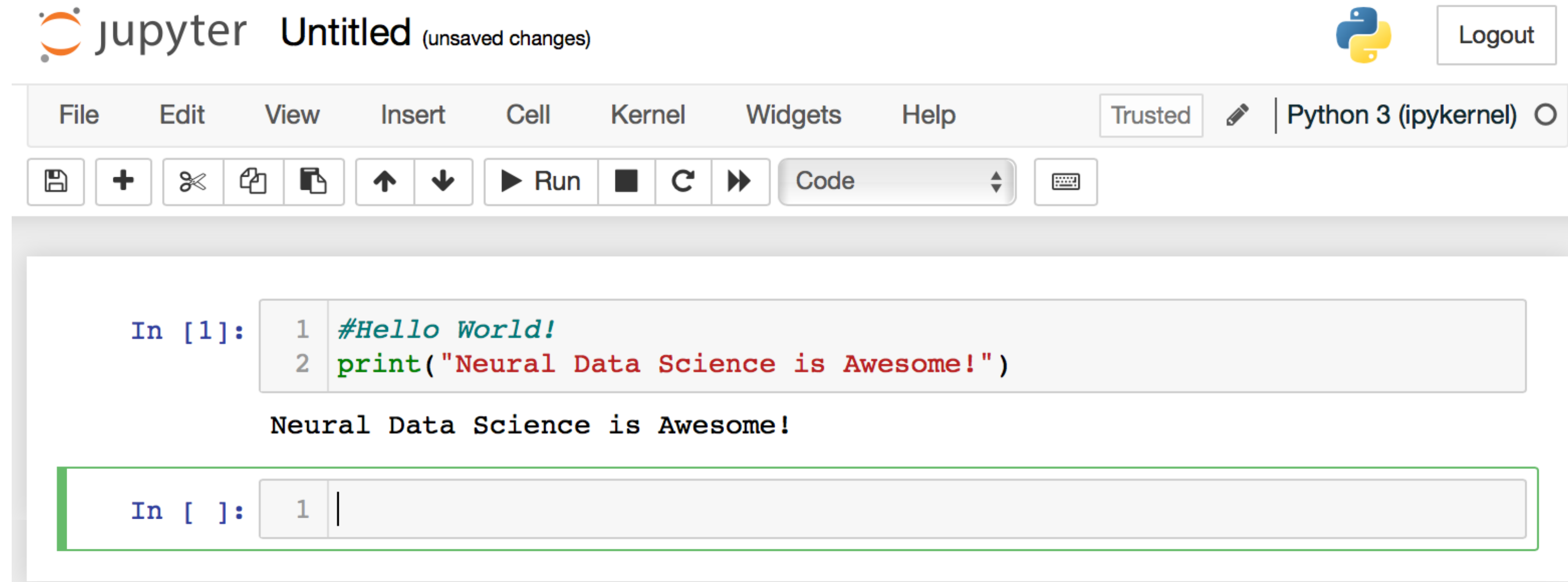
- [Datahub/DSMLP Cluster Status](#)
- [Independent Study Access Request](#)
- [Data Science Resources](#)
- [Datahub/DSMLP Knowledge Base](#)
  - [Launching Containers from the Command Line](#)
  - [Configuring Your Container Launch](#)

### Instructor Resources

- [Request Datahub/DSMLP - Instructional Technology Request \(CINFO\)](#)
- [Instructor Guidance for Datahub/DSMLP](#)
- [Educational Technology Services Instructional Github](#)
- [Blink Documentation](#)
- [Datahub Grading Tools](#)

# Jupyter notebooks review

- Installing anaconda
- <https://github.com/COGS108/Tutorials>
- <https://github.com/NeuralDataScience/Tutorials>
- Correcting common issues
- Up to students to correct and resubmit so grading can be timely



The screenshot displays the Jupyter Notebook interface. At the top, the title bar shows the Jupyter logo, the text "jupyter Untitled (unsaved changes)", the Python logo, and a "Logout" button. Below the title bar is a menu bar with options: File, Edit, View, Insert, Cell, Kernel, Widgets, Help, Trusted, and Python 3 (ipykernel). A toolbar below the menu bar contains icons for saving, adding, deleting, copying, pasting, undo, redo, and running code. The main area shows a code cell with the following content:

```
In [1]: 1 #Hello World!
        2 print("Neural Data Science is Awesome!")
```

The output of the code cell is displayed below the code:

```
Neural Data Science is Awesome!
```

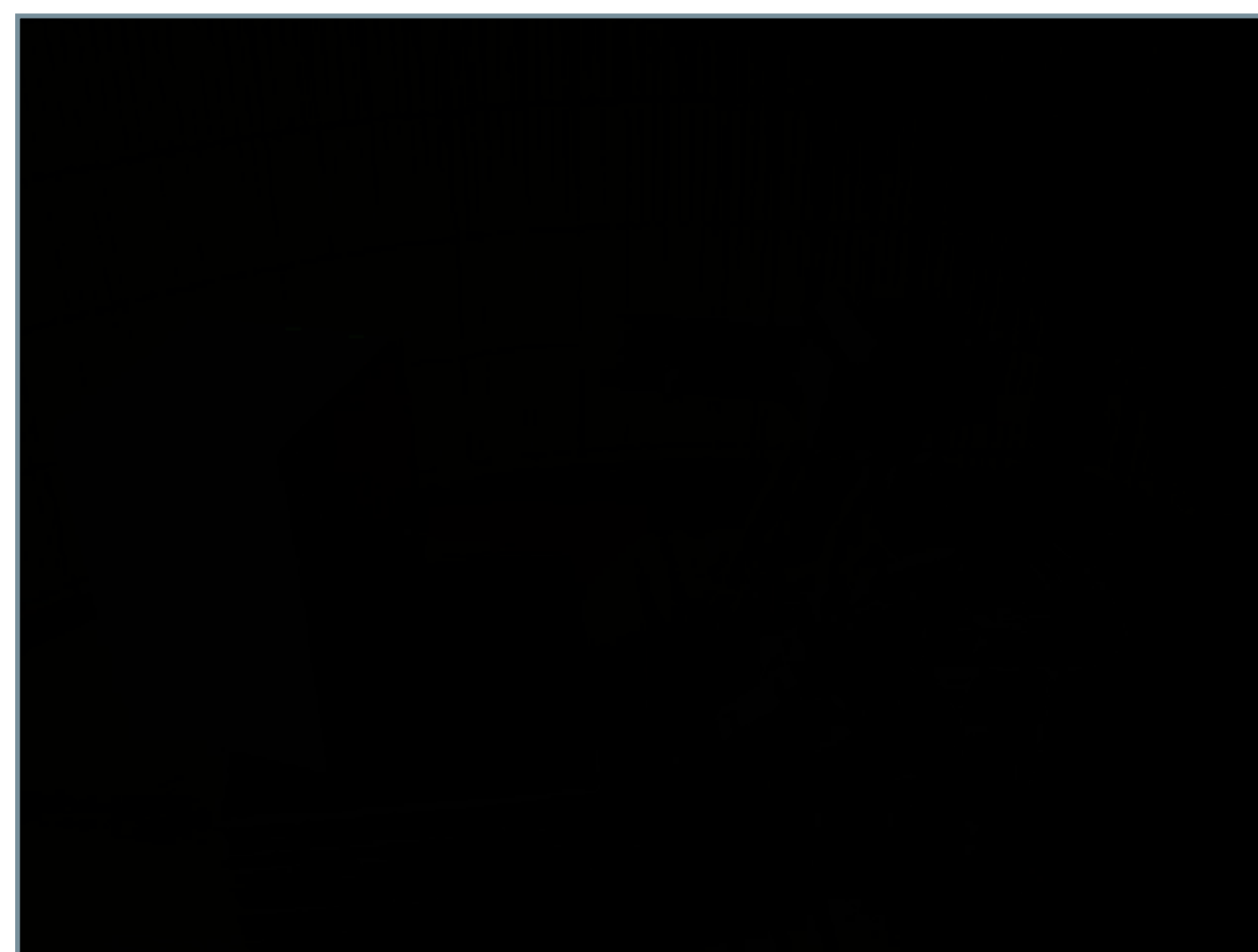
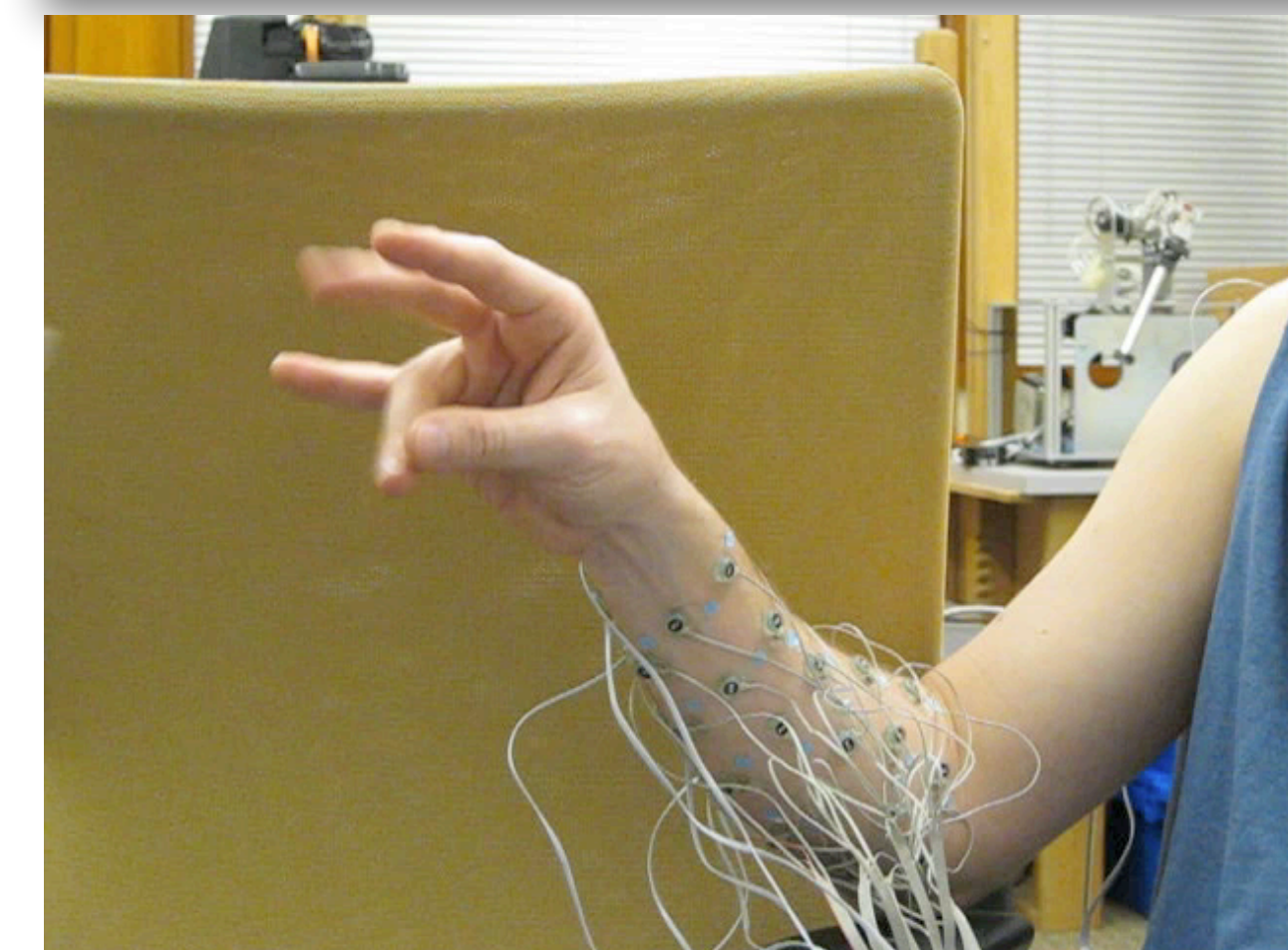
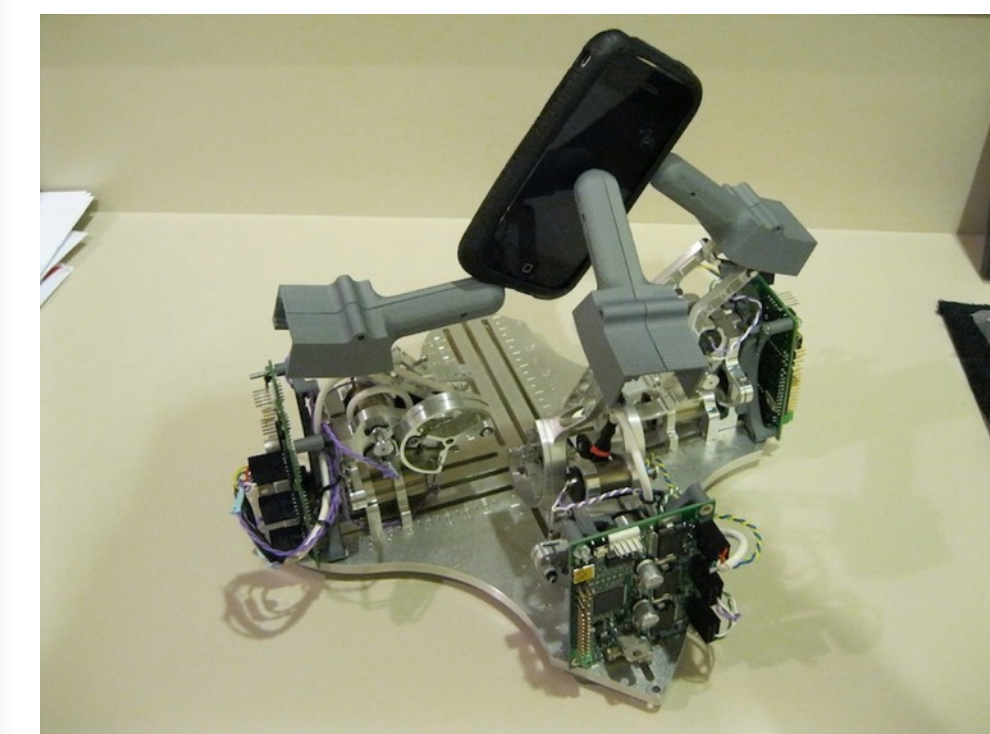
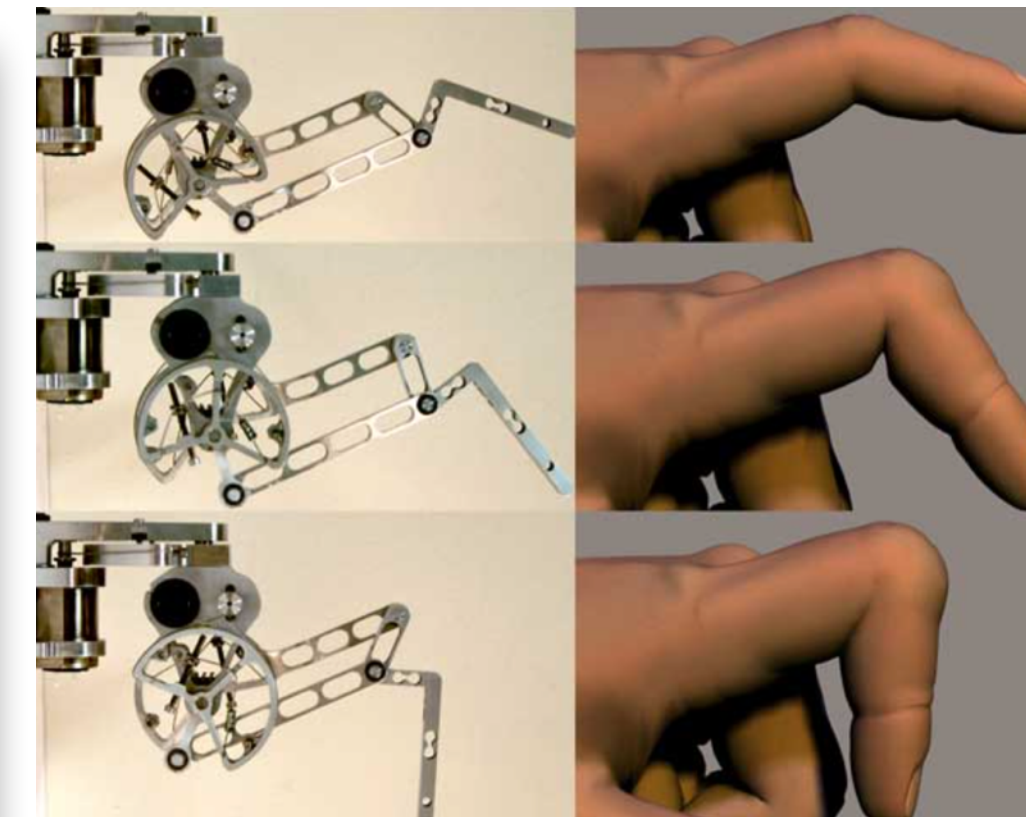
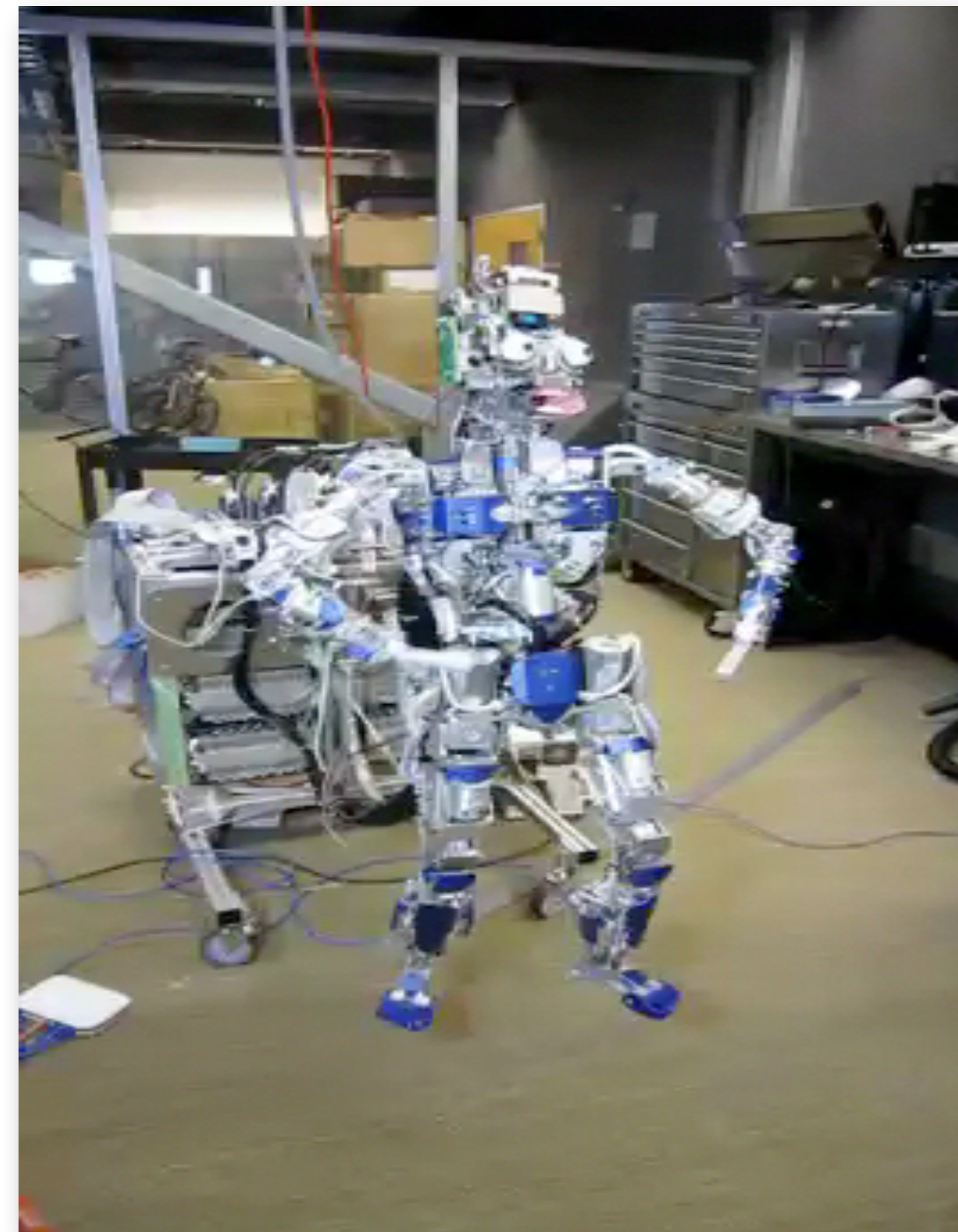
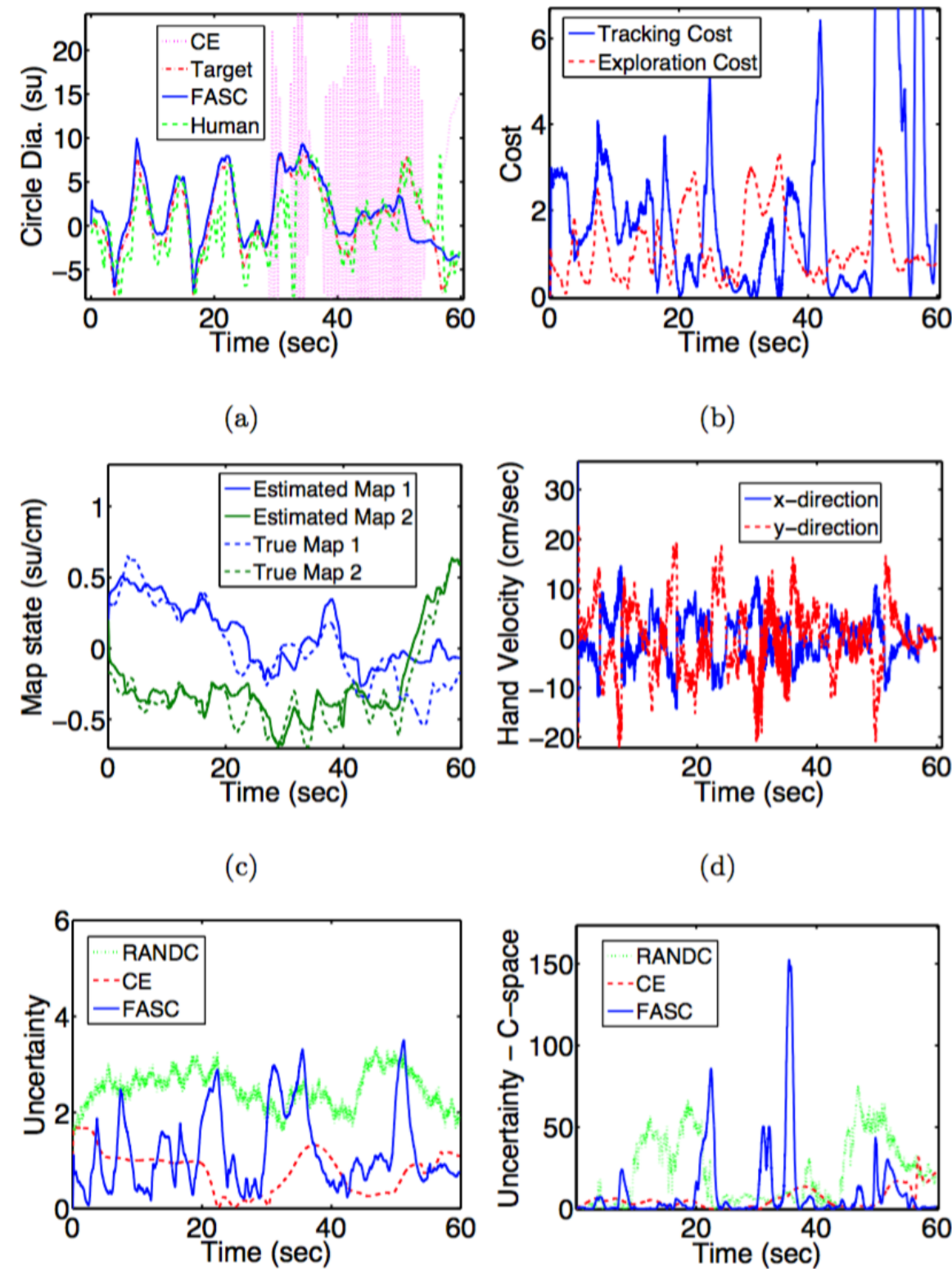
Below the output, there is an empty code cell with the prompt "In [ ]:" and a cursor.

# Who am I?

- C. Alex Simpkins Jr. Ph.D.
- BS/BS/MS/PhD UCSD Psyc, AMES, MAE, MAE, 2 postdocs UCSD Cogs and UW CSE
- Taught as TA ~20 times as a student, taught COGS109 as a grad student, taught at SDSU for a year in ME in Design, came back to UCSD last quarter - COGS100 and 108
- Been involved in teaching for over 30 years teaching martial arts

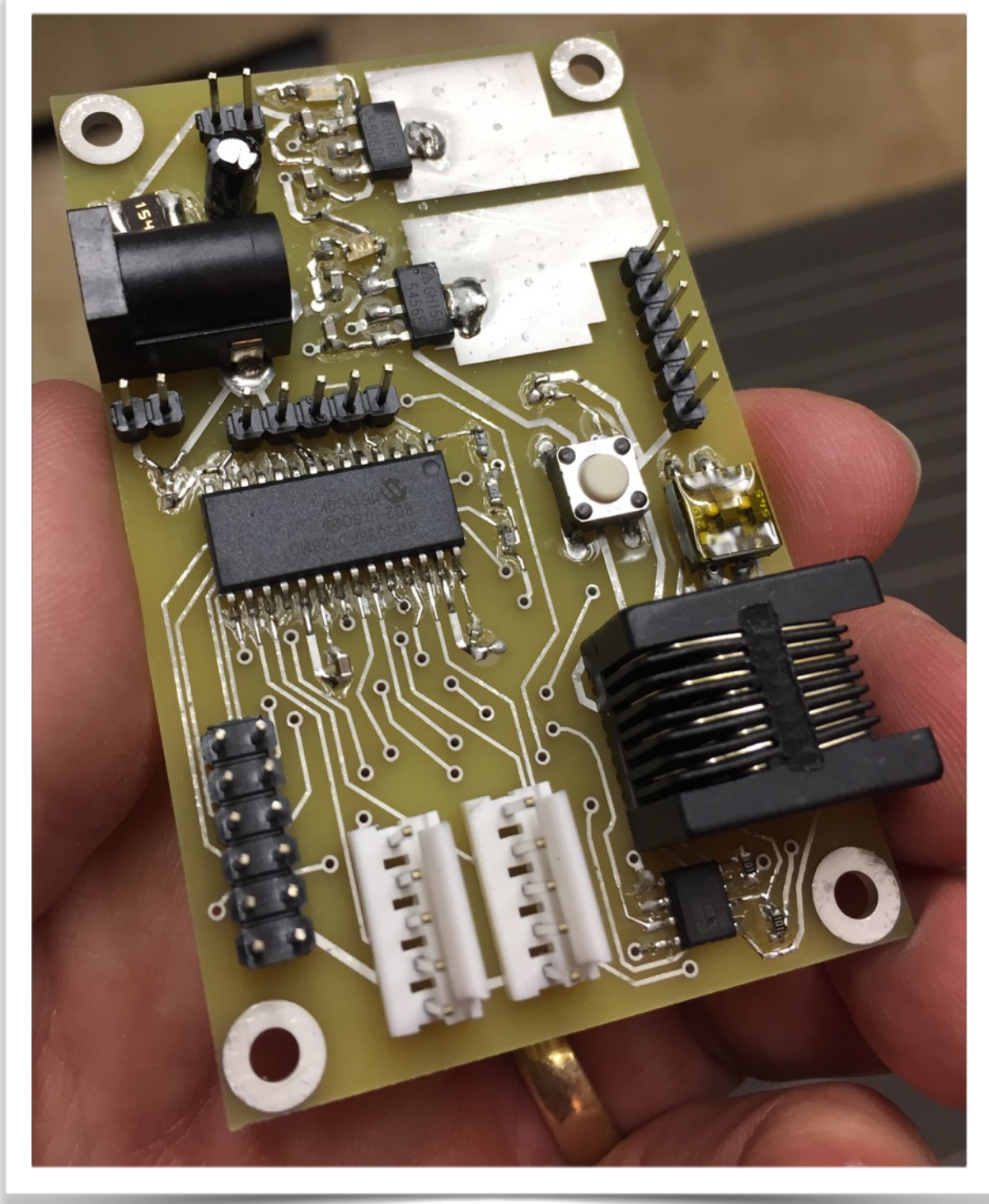
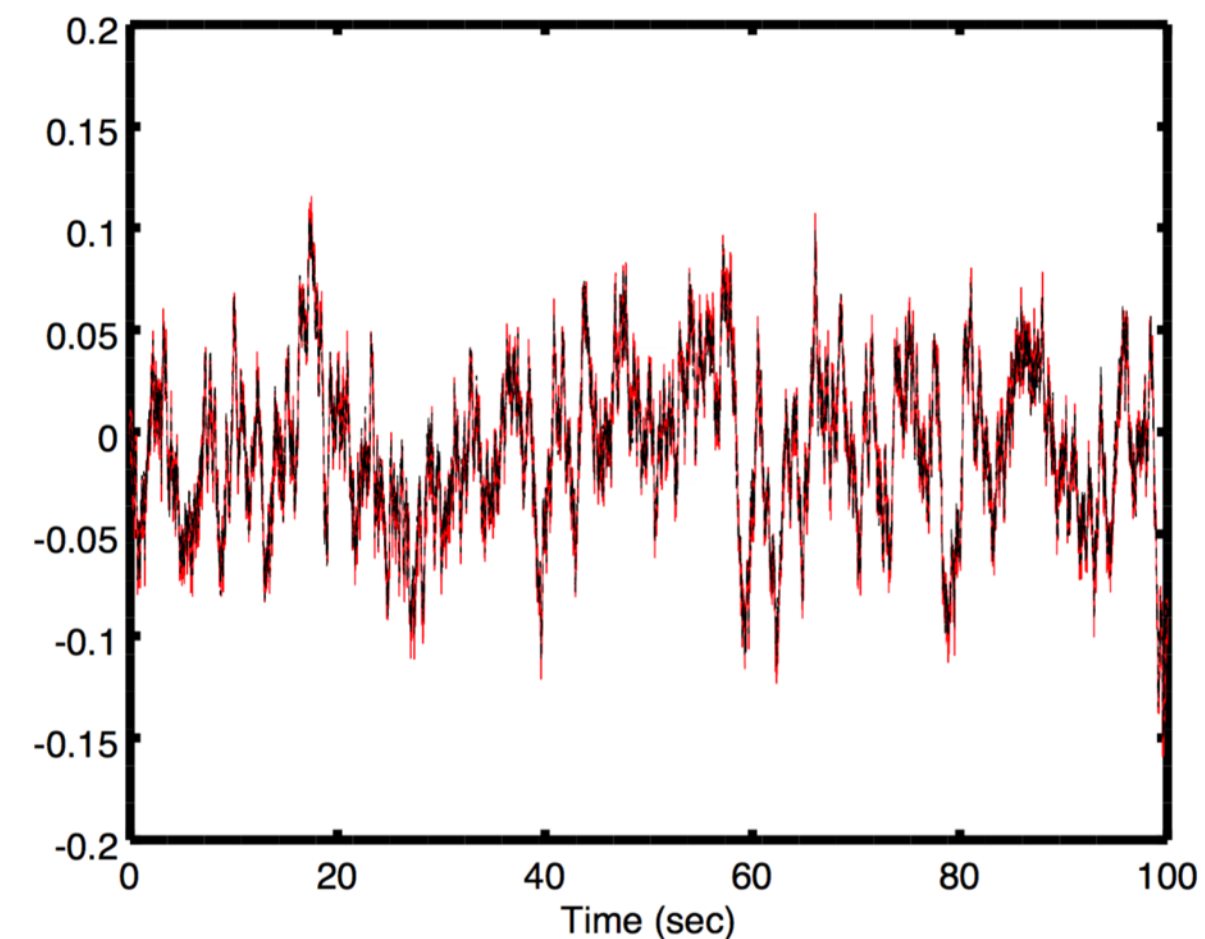
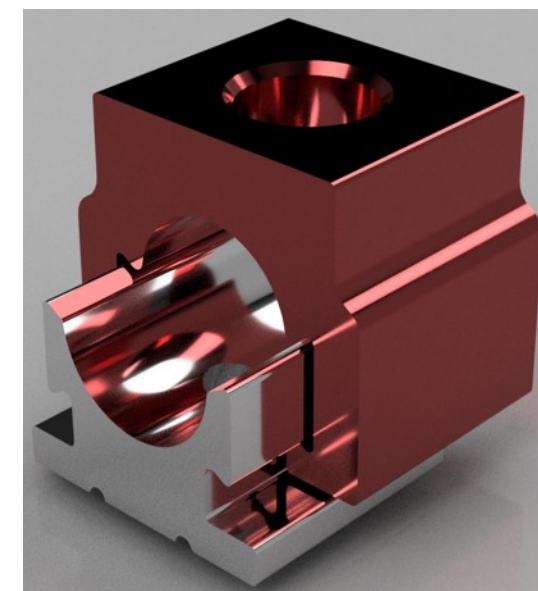
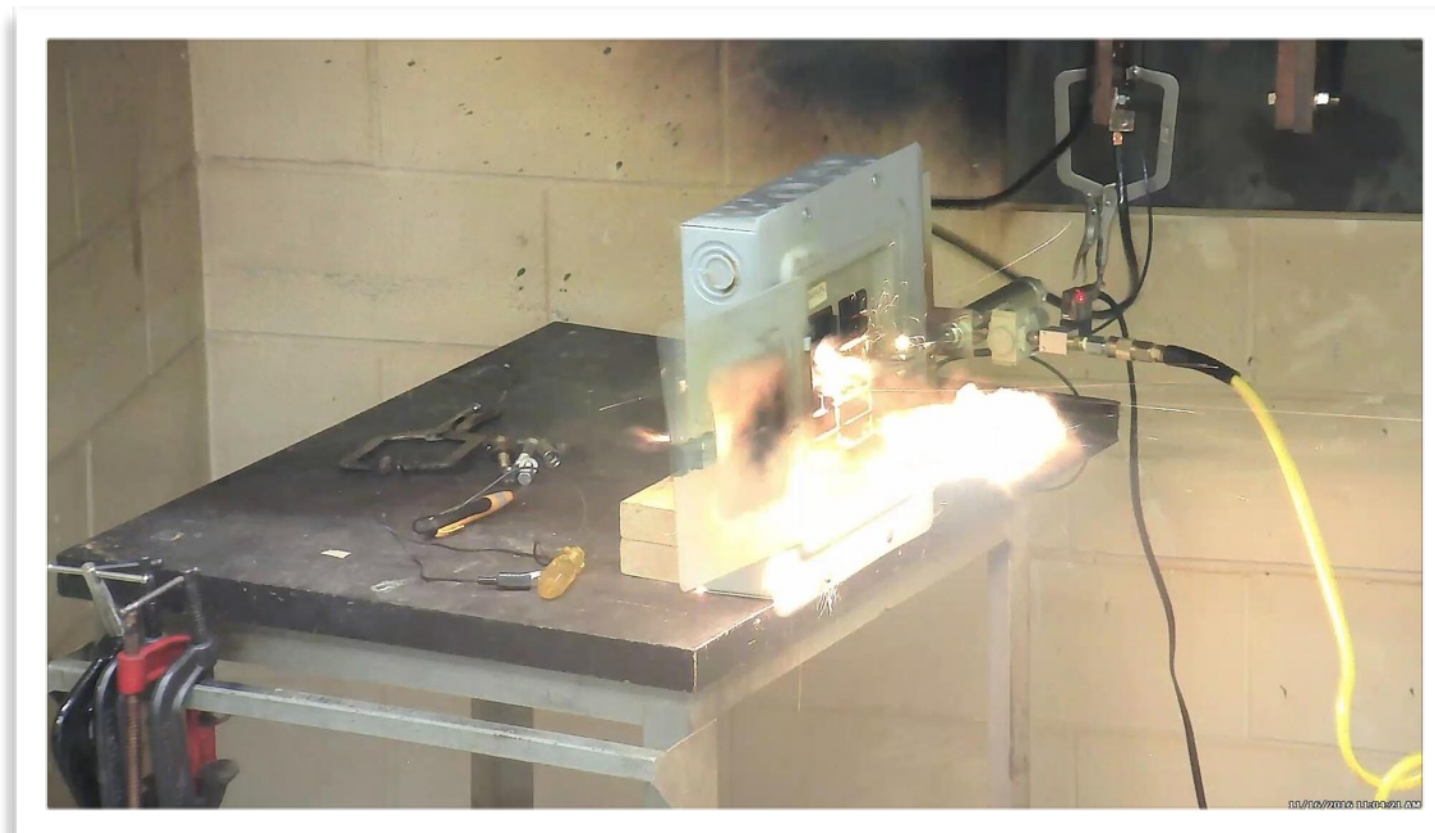


- PhD in Control Theory (and Design) is based on cybernetics
- Robotics



# Who am I?

- Many industrial projects - robotics, AI, drones
- Consulting
- Entrepreneurial startup work
- Work with larger companies
- Research - COVID, robotics, AI, control



# What to expect in this class

- What is my (our) role?
  - I am (we are) here to help you to learn and succeed, to open the door
  - NOT here to weed anybody out
  - NOT here to compete with you
  - Mutual respect

# What to expect in this class II

- What is your role?
- Learn! Open your mind
- Put in the effort - you must walk through the door
  - Watch/attend lectures, do the readings, complete assignments and tests, and think about it all
- Treat each other well, help each other to succeed (but do your own work of course)