

Introduction to Scientific Computing and Data Science

Course Goals

- ▶ Overview of topics in **scientific computing, data science, and machine learning**
- ▶ Apply these topics to your own research
- ▶ Familiarize yourself with the tools and software

Course Survey

Topics

- ▶ Programming Foundations
- ▶ Computer Science and Math Foundations
- ▶ Overview of Scientific Computing
- ▶ Exploratory Data Science and Visualization
- ▶ Machine Learning
- ▶ Deep Learning

Tools

- ▶ Python
- ▶ NumPy, SciPy, Pandas
- ▶ Matplotlib, Seaborn
- ▶ Scikit-Learn, Scikit-Image
- ▶ PyTorch

Course Logistics

Refer to the current year's course syllabus for more details.

We will typically split the class into the following components:

- ▶ Lectures: Twice a week
- ▶ Homework: Once a week
- ▶ Office Hours: Typically once a week with demand

Grading is typically split as follows:

- ▶ Homework: 80%
- ▶ Participation + Other: 20%

Communication

- ▶ Email: stefan.abi-karam@ahschool.com
 - ▶ Absence, Personal Concerns, Project Questions and Feedback
 - ▶ Please use your Heritage email address as well
- ▶ Google Classroom
 - ▶ **Homework Assignments**
 - ▶ Class meeting link
 - ▶ Class announcements and changes
- ▶ Campuswire
 - ▶ **Questions about Homework**
- ▶ Office Hours
 - ▶ **Questions about Homework**
 - ▶ **Questions about Projects**
- ▶ Slack
 - ▶ Project Questions and Feedback, Quick Updates, Meeting Links and Reminders

About Your Teacher

Timeline

- ▶ Graduated from American Heritage in 2018
- ▶ B.S. in Electrical Engineering in 2021
- ▶ M.S. in Electrical Engineering in 2022
- ▶ Ph.D. + Research Faculty 2023 and beyond

Research

My research is focused on hardware acceleration of machine learning algorithms and applied machine learning for hardware design tools and emerging technologies.



About Your Teacher

More Details

- ▶ SHARC Lab, ECE, Georgia Tech
- ▶ CIPHER Lab, Georgia Tech Research Institute
- ▶ Georgia Tech Undergraduate Admissions

Hobbies

- ▶ Meteorology
- ▶ Music Production and Analog Music Synthesis
- ▶ Sailing
- ▶ Hobby Electronics and Programming
- ▶ Reading

About Your Teacher

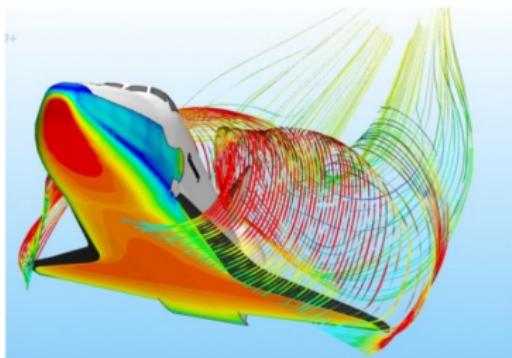
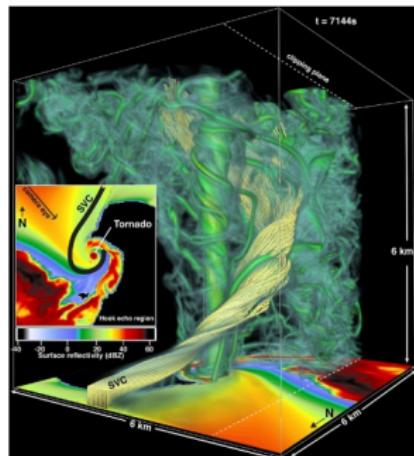
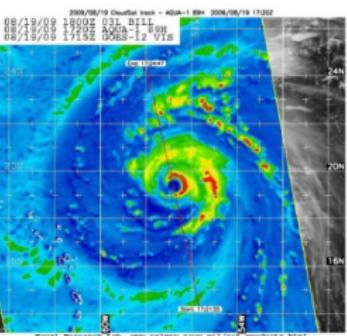
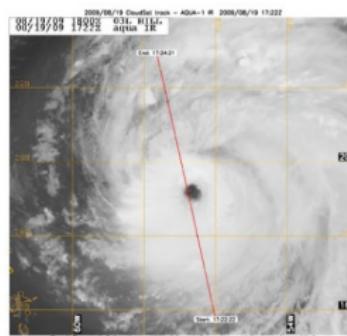


What is Computing?

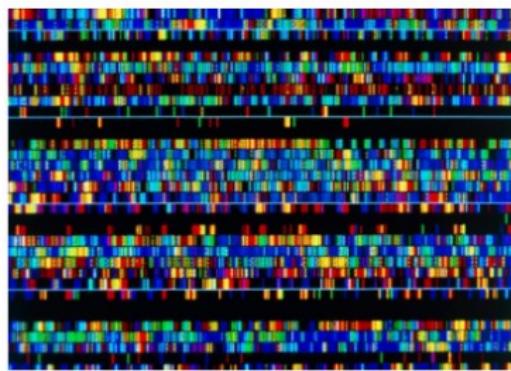
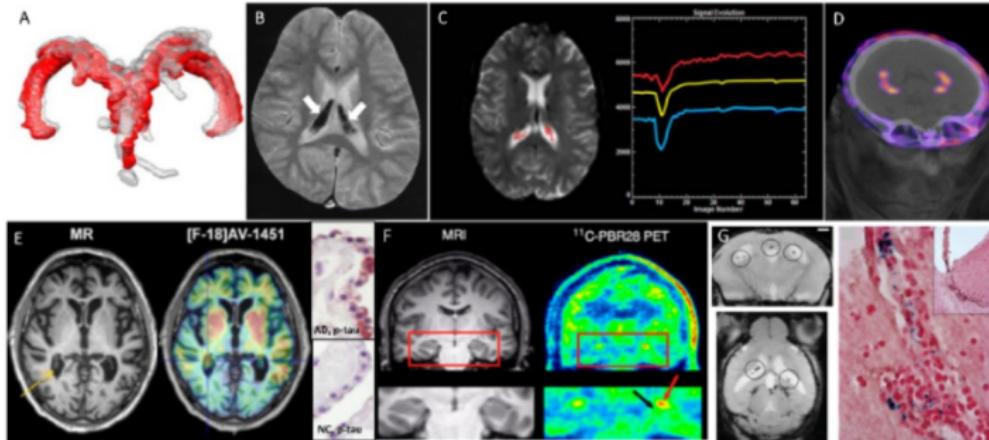
Computer Science is no more about computers than astronomy is about telescopes.

—Edsger W. Dijkstra

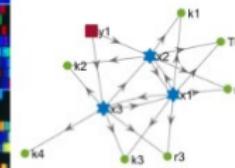
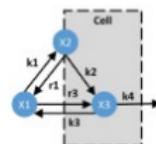
Scientific Computing - Earth Science and Engineering



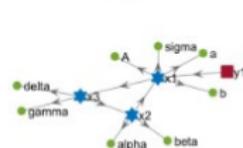
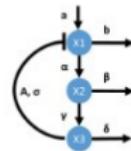
Scientific Computing - Biology and Medicine



A Pitavastatin uptake



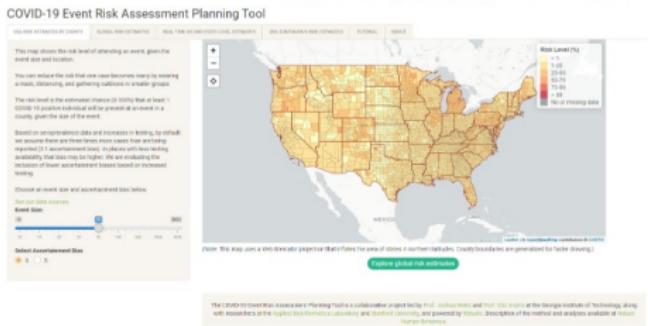
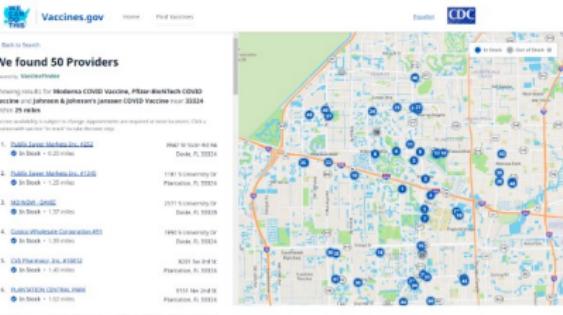
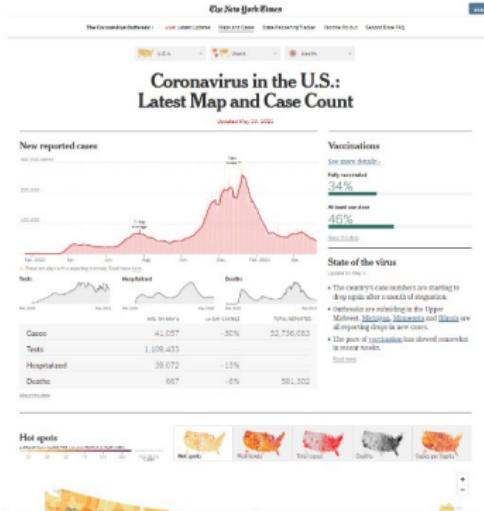
B Goodwin oscillator



What is Data Science?

Data is the new oil.

Data Science - COVID-19



Data Science - General Applications



Loan Default Prediction
Beginners data set for financial analytics.

Kamal Das • updated a month ago (Version 2)

Data Tasks (1) Code (6) Discussion (1) Activity Download (269 KB) New Notebook

Usability 10.0 License Data files © Original Authors Tags business, finance, tabular data, banking, beginner



Heart Attack Analysis & Prediction Dataset
A dataset for heart attack classification

Rashid Rahman • updated 2 months ago (Version 2)

Data Tasks (1) Code (203) Discussion (12) Activity Download (32 KB) New Notebook



Covid-19 period air-traffic dataset
air traffic data from The OpenSky Network 2020

IshaDS • updated 4 days ago (Version 2)

Data Tasks (1) Code (1) Discussion Activity Download (3 GB) New Notebook

Usability 8.8 License Attribution 4.0 International (CC BY 4.0) Tags transportation, covid19, intermediate, aviation



Hourly energy demand generation and weather
Electrical demand, generation by type, prices and weather in Spain

Nicholas Jhana • updated 2 years ago (Version 1)

Data Tasks (1) Code (62) Discussion (2) Activity Download (25 MB) New Notebook

Usability 10.0 License CC0: Public Domain Tags business, news, energy, weather and climate, renewable energy

What is Machine Learning?

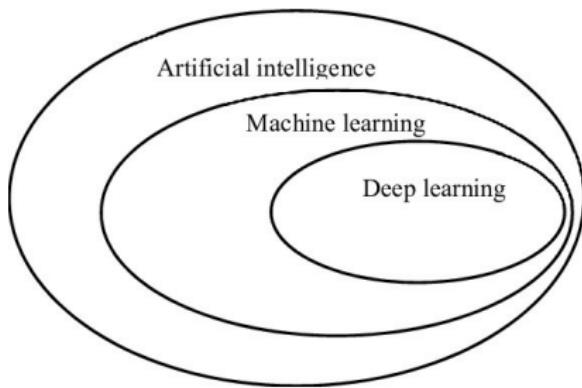
AI is the new electricity.

—Andrew Ng

In the same way that electricity transformed our world, AI will transform our world and will be incorporated into every aspect of our lives.

AI vs. Machine Learning vs. Deep Learning

- ▶ **AI:** The study of how to make computers intelligent.
- ▶ **Machine Learning:** A subfield of computer science that gives computers the ability to learn without being explicitly programmed.
- ▶ **Deep Learning:** A subfield of machine learning that uses neural-network-like structures and large amounts of data to learn.



Machine Learning - Self-Driving Cars



Machine Learning - Generative Models

The screenshot shows the DreamStudio Lite interface. On the left is a sidebar with a logo, navigation links (MENU, Dream, History, Prompt Guide, Social, FAQ, Support), and a back arrow. The main area displays four generated images of cats playing banjos in different styles. Below the images is a text input field containing a prompt: "Painting of a cat playing the banjo, abstracted observed, minimal indication, thick outlines, contented peaceful, medium saturation with trichromatic similar hues". To the right are several configuration sliders: Credits per image (1.0), Height (512), Cfg Scale (7), Steps (50), Number of Images (4), and Sampler (set to EMA). A teal "Dream" button is at the bottom right.

DreamStudio Lite

1.0 credits / image

Height 512

Cfg Scale 7

Steps 50

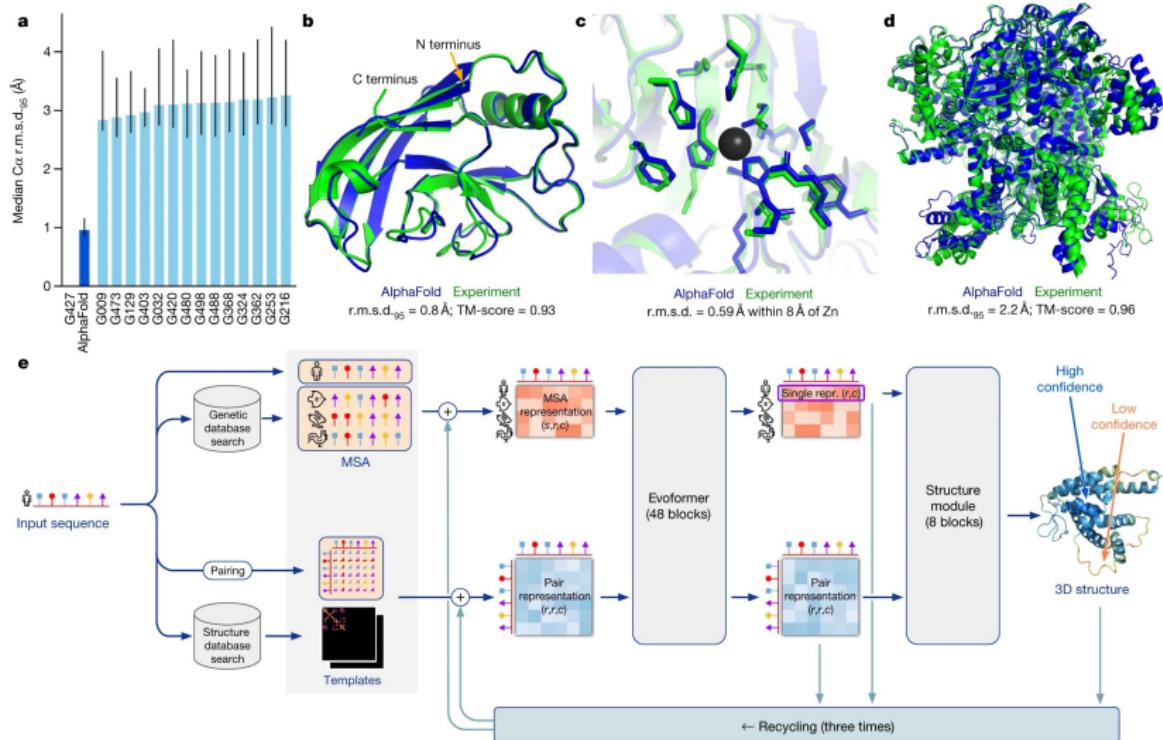
Number of Images 4

Sampler

Dream

Painting of a cat playing the banjo, abstracted observed, minimal indication, thick outlines, contented peaceful, medium saturation with trichromatic similar hues

Machine Learning - Protein Folding



Machine Learning - Protein Folding

The screenshot shows the homepage of the AlphaFold Protein Structure Database. At the top, there is a green header bar with the text "Highly accurate protein structure prediction with AlphaFold" and a link to "PubPeer". Below this is a dark grey navigation bar with links for EMBL-EBI, Services, Research, Training, About us, and EMBL-EBI logo. The main title "AlphaFold Protein Structure Database" is on the left, and a navigation menu with Home, About, FAQs, and Downloads is on the right. The background features a blue gradient with a faint, stylized protein structure graphic. The central title "AlphaFold Protein Structure Database" is displayed in large, bold, white font. Below it, the text "Developed by DeepMind and EMBL-EBI" is shown in a smaller white font. A search bar at the bottom left contains the placeholder "Search for protein, gene, UniProt accession or organism" and a "BETA" button. To the right of the search bar is a blue "Search" button. Below the search bar, there is a row of example queries: "Free fatty acid receptor 2", "At1g58602", "Q5VSL9", "E. coli", "Help", and "AlphaFold DB search help". At the very bottom, there is a footer with links for "Feedback on structure" and "Contact DeepMind", along with standard browser navigation icons.

"Highly accurate protein structure prediction with AlphaFold" has comments on PubPeer

EMBL-EBI Services Research Training About us EMBL-EBI

AlphaFold Protein Structure Database

Home About FAQs Downloads

AlphaFold Protein Structure Database

Developed by DeepMind and EMBL-EBI

Search for protein, gene, UniProt accession or organism BETA Search

Examples: Free fatty acid receptor 2 At1g58602 Q5VSL9 E. coli Help AlphaFold DB search help

Feedback on structure: Contact DeepMind

Closing Advice

- ▶ Find connections between your interest and what we will cover and explore those connections; always keep that in the back of your mind.
- ▶ There is no such thing as a dumb question; if you have a question or a gap in your understanding, ask about it as soon as possible.
- ▶ I am flexible about a lot of things if you come to me early and talk to me directly.
- ▶ I have a good sense of humor, so any class or research memes are encouraged.