

# Introduction to Scientific Computing and Data Science

## Scientific Computing

Stefan Abi-Karam

Summer 2023

# Table of Contents

1. Course Overview
2. About Your Teacher
3. Scientific Computing Overview
4. Closing Advice

## 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

# Table of Contents

1. Course Overview
2. About Your Teacher
3. Scientific Computing Overview
4. Closing Advice

# About Your Teacher

- ▶ 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



# Table of Contents

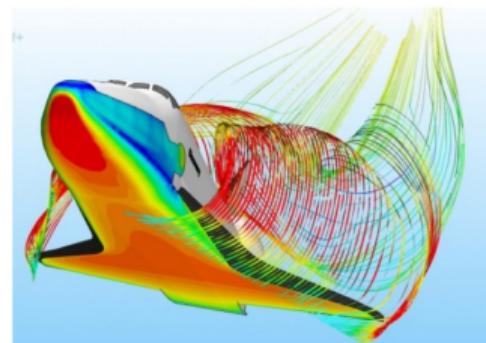
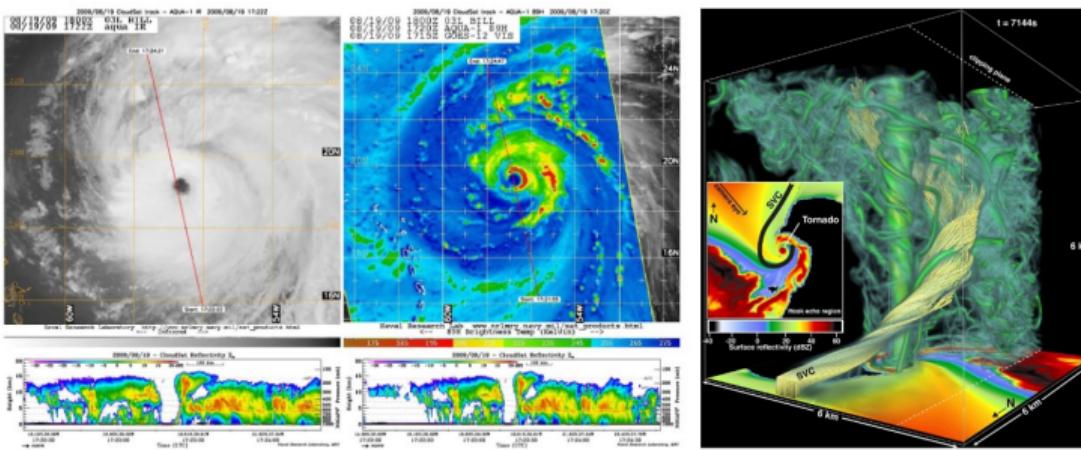
1. Course Overview
2. About Your Teacher
3. Scientific Computing Overview
4. Closing Advice

# What is Computing?

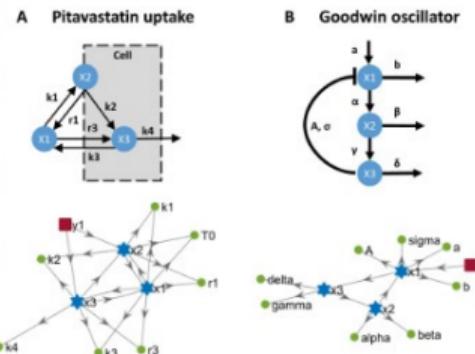
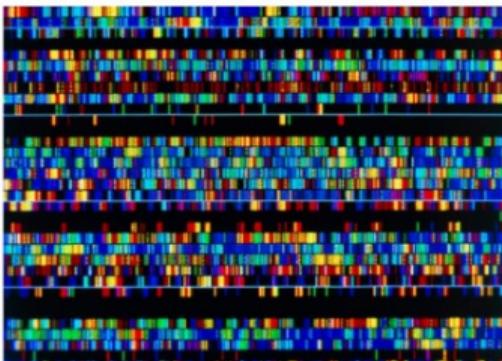
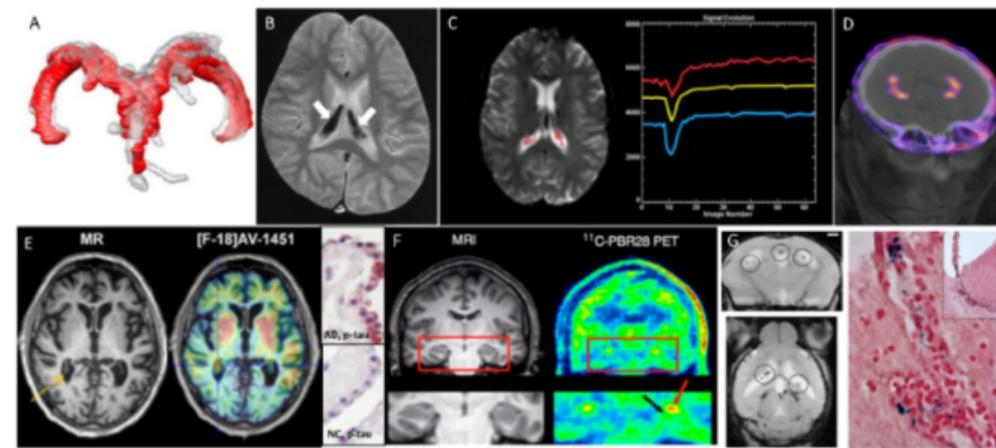
*Computer Science is no more about computers than astronomy is about telescopes.*

—Edsger W. Dijkstra

# Scientific Computing - Earth Science and Engineering



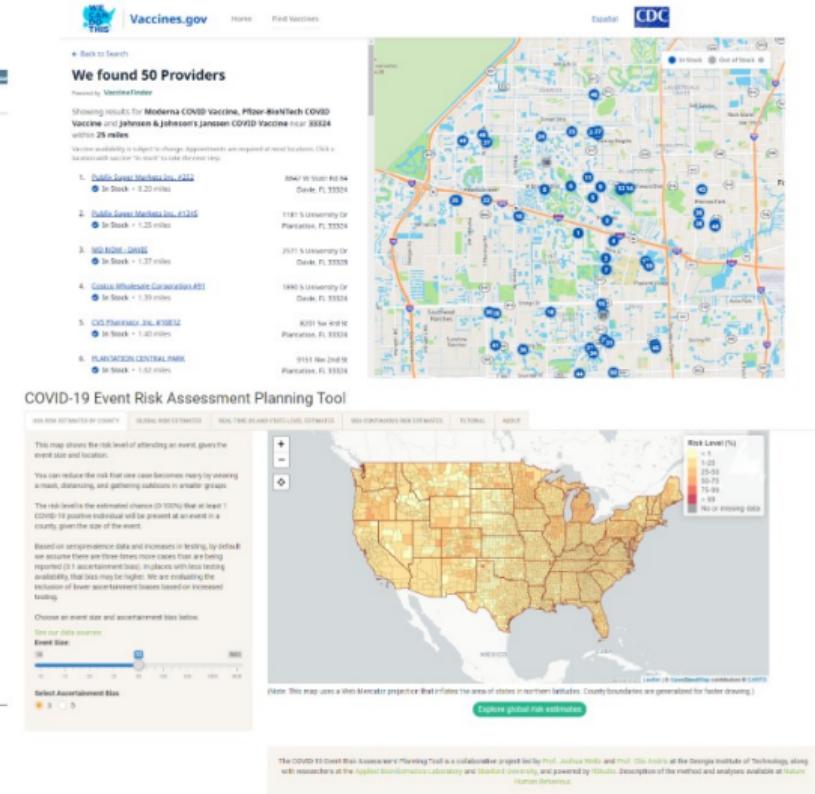
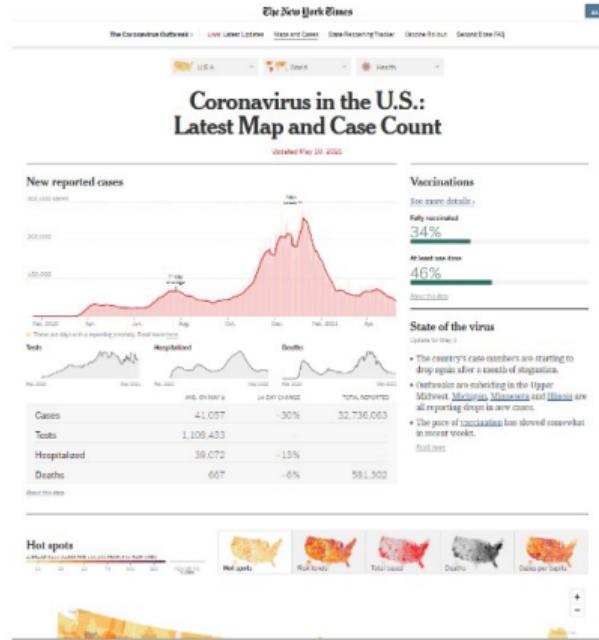
# Scientific Computing - Biology and Medicine



# 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



**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

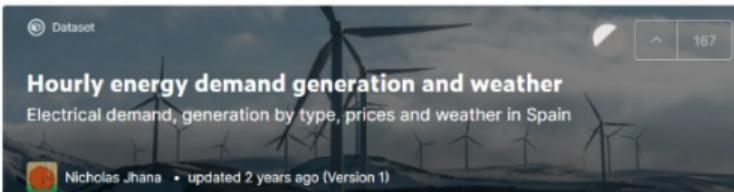
Usability 8.8 License Attribution 4.0 International (CC BY 4.0) Tags transportation, covid19, intermediate, aviation



**Heart Attack Analysis & Prediction Dataset**  
A dataset for heart attack classification  
Rashik Rahman • updated 2 months ago (Version 2)

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

Usability 10.0 License CC0: Public Domain Tags health, health conditions, classification, heart conditions, binary classification



**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 (12) 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?

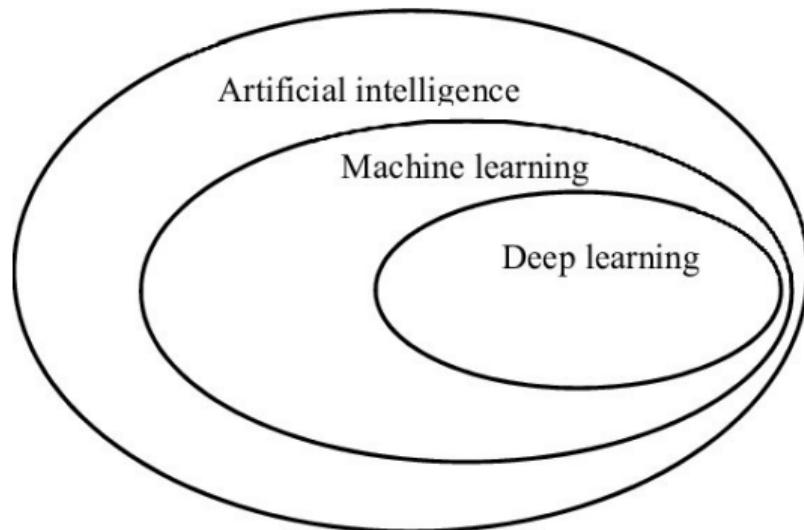
*AI is the new electricity.*

—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



DreamStudio  
beta

MENU

-  Dream
-  History
-  Prompt Guide
-  Social
-  FAQ
-  Support

DreamStudio Lite



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

Dream



1.0 credits / image

The width of the generated image.

Height 512

The height of the generated image.

Cfg Scale 7

Cfg scale adjusts how much the image will be like your prompt. Higher values keep your image closer to your prompt.

Steps 50

How many steps to spend generating (diffusing) your image.

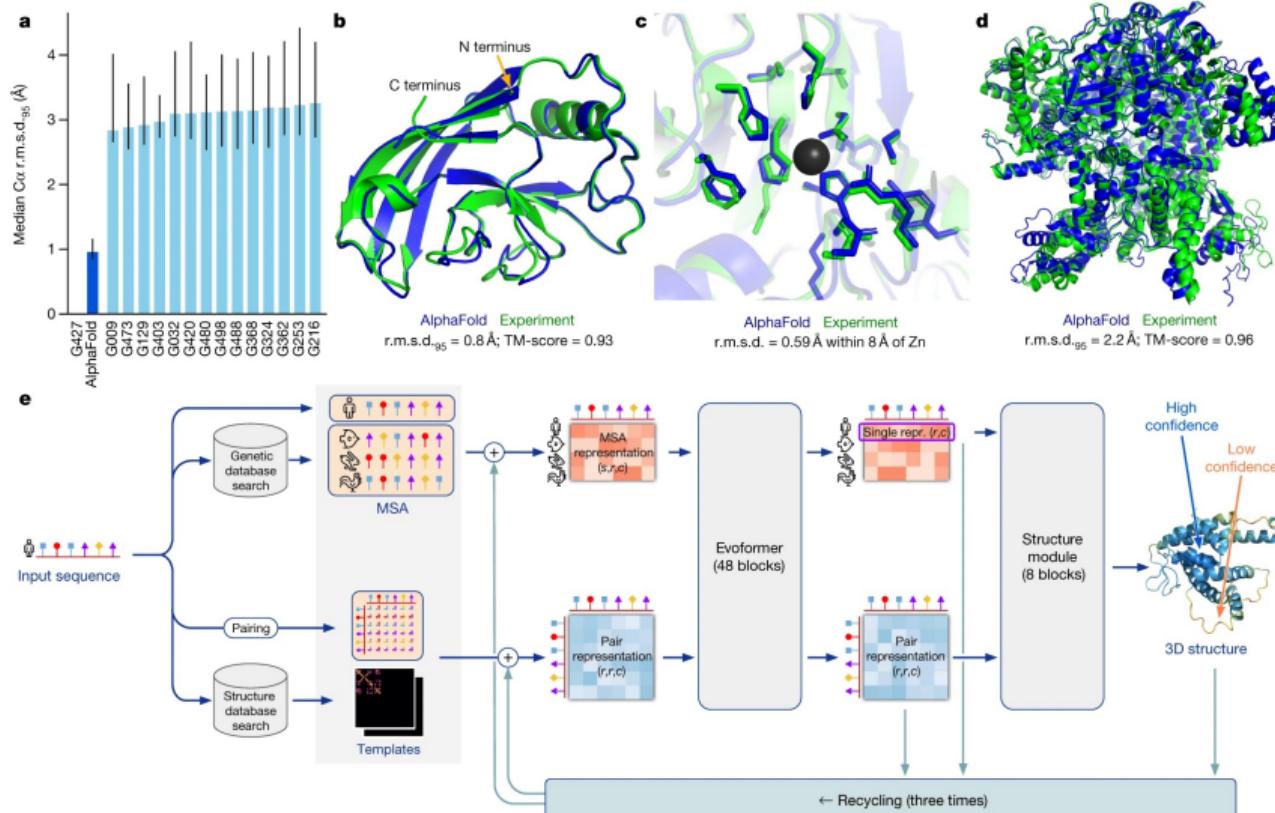
Number of Images 4

To generate multiple images from one prompt.

Sampler

The diffusion sampling method. Cannot be modified if CLIP Guidance is enabled.

# 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 blue navigation bar with links for EMBL-EBI, Services, Research, Training, About us, and a logo for EMBL-EBI. The main title "AlphaFold Protein Structure Database" is displayed in white on the left, and a navigation menu with "Home", "About", "FAQs", and "Downloads" is on the right. The background features a large, stylized blue protein structure model. The central title "AlphaFold Protein Structure Database" is written in large, bold, white letters. Below it, the text "Developed by DeepMind and EMBL-EBI" is also in white. At the bottom, there is a search bar with the placeholder "Search for protein, gene, UniProt accession or organism", a "BETA" button, and a "Search" button. Below the search bar, examples of search terms are listed: "Free fatty acid receptor 2", "At1g58602", "Q5VSL9", "E. coli", "Help:", and "AlphaFold DB search help". There is also a link for "Feedback on structure: Contact DeepMind".

# Table of Contents

1. Course Overview
2. About Your Teacher
3. Scientific Computing Overview
4. Closing Advice

## 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.