NC STATE University Libraries

# AI Tools for Research

Comparative Molecular Medicine Training Program

# DATA SCIENCE SERVICES

## OUR SERVICES:

CONSULTING

WORKSHOPS & INSTRUCTION

SPECIALIZED SPACES

## WHAT WE DO:

Finding Data

Wrangling Data

Analyzing Data

Visualization

Geographic Information Systems (GIS)

Software Access

# DATA SCIENCE SERVICES - WHO WE ARE

**Mara Blake**
Head, Data Science Services

**Shannon Ricci**
Associate Head, Data Science Services

**Jeff Essic**
Lead Librarian for GIS

**Alp Tezbasaran**
Data Science Librarian

**Selene Schmittling**
Data Science Librarian

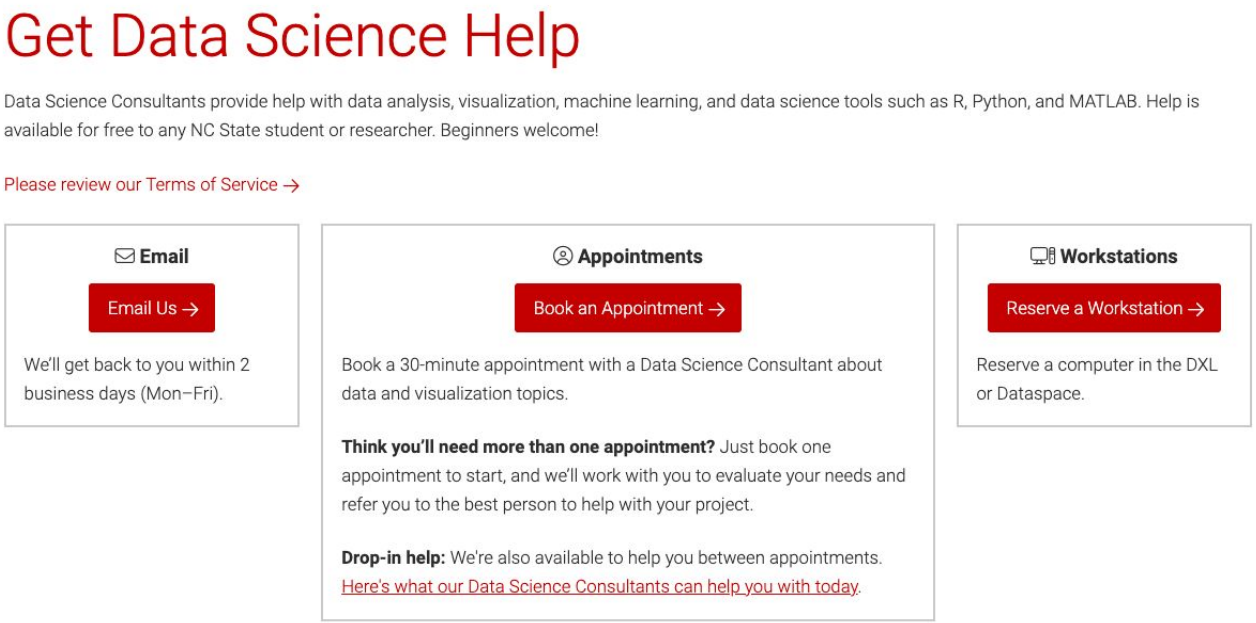
**Lynnee Argabright**
Data Science Teaching Librarian

Our team comes from a variety of academic backgrounds, including nuclear engineering, electrical engineering, marine ecology, history, information science, natural resources, and geographic information sciences.

# HOW TO CONTACT US

go.ncsu.edu/getdatahelp

Data requests: getdatahelp@ncsu.edu



## Get Data Science Help

Data Science Consultants provide help with data analysis, visualization, machine learning, and data science tools such as R, Python, and MATLAB. Help is available for free to any NC State student or researcher. Beginners welcome!

Please review our Terms of Service →

**Email**

Email Us →

We'll get back to you within 2 business days (Mon–Fri).

**Appointments**

Book an Appointment →

Book a 30-minute appointment with a Data Science Consultant about data and visualization topics.

**Think you'll need more than one appointment?** Just book one appointment to start, and we'll work with you to evaluate your needs and refer you to the best person to help with your project.

**Drop-in help:** We're also available to help you between appointments. Here's what our Data Science Consultants can help you with today.

**Workstations**

Reserve a Workstation →

Reserve a computer in the DXL or Dataspace.

# Setting the Stage

What is AI?

# AI is not

turnitin™

Hey Siri

Google Translate

Text | Images | Documents | Websites

Detect language | Spanish | English | French | English | Spanish | Arabic

Are you any good at translation?

33 / 5,000

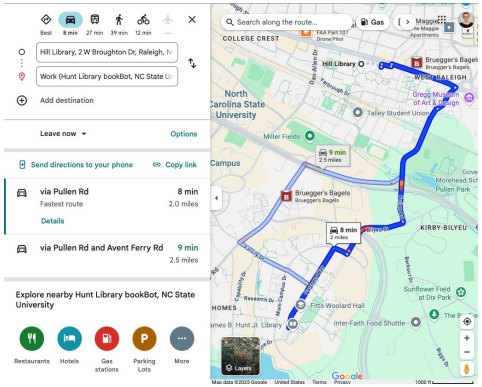
Translations are gender-specific. Learn more

¿Eres buena en la traducción?

(feminine)

**Learns what you like.**

Just turn it up and down. It learns the temperatures you like and creates a schedule for you.


**Explore top deals in related categories** See all deals


Java House Single-Serve Cold Brew Coffee Pods – Lightly Sweetened Salt…

82

Amazon's Choice

-10% $13.49

($1.12/count)

Limited time deal

Typical: $14.99

Get it as soon as Tuesday, Jul 29

FREE Shipping on orders over $35 shipped by Amazon


Taylors of Harrogate Classic Chai Tea, 20 Teabags (Pack of 1)

7,385

Amazon's Choice

-20% $4.00 ($0.20/count)

Limited time deal

Typical: $4.99

Get it as soon as Tuesday, Jul 29

FREE Shipping on orders over $35 shipped by Amazon

4 sustainability features


JAVA HOUSE Dark Roast Sumatran Cold Brew Pods, (12 Count) Peel and Pour Pods Hot or Iced…

6,805

-10% $13.49

($0.83/fluid ounce)

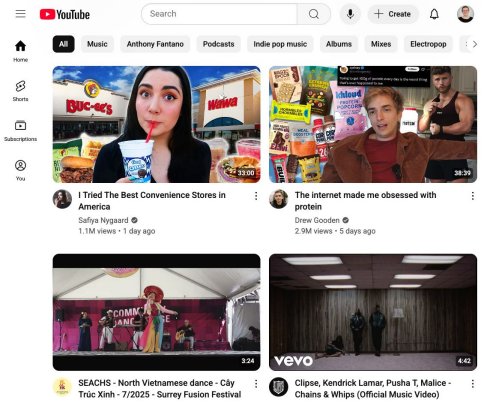
Limited time deal

Typical: $14.99

Get it as soon as Tuesday, Jul 29

FREE Shipping on orders over $35 shipped by Amazon

1% Claimed




grammarly

# Chatbot options

ChatGPT

Gemini

Claude

Perplexity


Copilot

DeepSeek


Julius

GitHub Copilot


R Shiny Assistant

# AI Background

- But it feels new because of how rapidly it's grown in recent years, particularly **large language models (LLMs).**
- LLMs are AI systems that can generate human language. They are trained on large datasets & work by predicting what text is most likely to come next.

**Ex:** Before we leave, I need to feed the...

- Kids (10%)
- Dog (3%)
- Cat (3%)
- Hungry (0.1%)

# AI Background

- In casual conversation we tend to use the words ChatGPT, LLMs, and AI interchangeably.
    - But AI is a larger umbrella, and we'll look at more examples than just LLMs today.
- Increasingly, LLM tools are multimodal & can generate images, videos, etc.

**Common LLMs include:**

- **ChatGPT**
- **Gemini**
- **Claude**
- **CoPilot**

# Best Practices

Start with **goal and constraint setting.**

Have a plan for **prompting**.

**“Wh”** questions for AI:

- What is the issue that we have at hand?
- Why is it important?
- How to tackle it?

**Keep in mind:**

- Adaptability
- Fundamentals
- Terminology
- Systematic thinking
- Modularization

**AI is a tool, not a solution**

# Prompt PARTS

## Persona

**Identify who you are in relation to the task.**

Knowing your persona helps genAI tailor the output to your specific role.

‘Act as an expert sustainability researcher and learning designer.’

‘You are an experienced mentor to graduate students.’

‘Act as a highly creative instructional designer.’

## Aim

**State your objective clearly. What do you want genAI to achieve?**

**A clear aim provides genAI with a focused direction.**

Describe the task in clear and direct language.

It is recommended you break down large tasks into smaller chunks.

## Recipients

**Specify the audience you're targeting.**

Are you writing for undergrads, graduate students, or the general public?

Knowing your audience helps genAI adjust the tone and complexity of the output.

## Theme

**Describe the desired style, tone, and any relevant parameters.**

Do you want student facing instructions, a formal report, or an engaging case study?

Adding details about the theme guides genAI towards the appropriate style and content.

## Structure

**Note the format you expect for the output.**

Are you looking for a short paragraph, a detailed list, a table, or a complete script?

Specifying the structure helps genAI format the information accordingly.

## Basic prompt

Generate two highly engaging discussion activities for a [lab on enzyme kinetics].

## PARTS prompt

Act as a biochemistry professor designing engaging lab activities. — **Persona**

Generate two discussion activities that will actively involve students in understanding enzyme kinetics concepts. — **Aim**

Undergraduate students taking a biochemistry lab. — **Recipients**

The activities should be highly engaging, promote critical thinking, and reinforce the practical application of enzyme kinetics. — **Theme**

Provide a clear description of each activity, including objectives, materials needed (if any), and suggested discussion questions. — **Structure**

# Google Gemini

## What is it?

Google Gemini is a conversational AI tool (chatbot). This tool integrates with Google Colab.

## What are some use cases?

This tool can be used for a variety of things including:

- Draft and summarize documents
- Analyze and visualize data
- Create presentations
- Automate repetitive tasks
- Answer questions using internal knowledge

## Pricing details

Free to NC State affiliates!

NC State's choice of AI solution

https://gemini.google.com/app

Gemini

2.5 Flash

New chat

Explore Gems

Recent

CMake Fortran Compiler Co...

Shipping Export-Controlled ...

Research Project Document...

NotebookLM: Features, Use ...

NotebookLM: Overview and ...

Using xclip Command-Line ...

LaTeX Dark Mode Implemen...

t-SNE's Goal: Data Visualiza...

Splitting Images Horizontally...

GitHub Enterprise Organizat...

Google Docs to GitHub Wiki ...

StudentID,Name,HW1,HW2,...

Course Grade Data Analysis

StudentID,Name,HW1,HW2,...

Adding iFrames to WordPress

IEEE LaTeX Appendix Templ...

Intune versus AppsAnywher...

Calendly API Webhook Regis...

Gemini Essay Grading Works...

`fs::rename` vs. `fs::copy_fil...

Settings & help

Hello,

Want to try out a few things?

Make my own custom mini figure

Turn me into a superhero

Give me an 80s style makeover

Create a professional headshot

Enter a prompt for Gemini

Tools

# NotebookLM

## What is it?

Research and note taking tool with a chatbot tied to Google/Gemini Workspace

Compared to other tools **context** can be very large

## What are some use cases?

- Create study guides from notes and other media (ex. YouTube, slides)
- Summarize research papers
- Get explanations of complex topics
- Analyze dense reports
- Generate outlines
- Brainstorm ideas

## Pricing details

Free to NC State affiliates!

https://notebooklm.google.com/

# Your AI Toolkit

## The “Doer” (Gemini, ChatGPT, etc)

**Role:** The Data Analyst & Coder.

**Best For:** Writing Python code, formatting data, creating visualizations, mathematical calculations.

**How it works:** You give instructions; it executes tasks.

## The “Knower” (NotebookLM)

**Role:** The Research Assistant.

**Best For:** Summarizing PDFs, synthesizing literature, "chatting" with your specific data files.

**How it works:** Grounded only in the documents you upload (RAG - Retrieval Augmented Generation).

# Trusting the Tool

Data Privacy

- Rule of Thumb: Do not upload PII (Personally Identifiable Information) or unpublished, sensitive IP into **public models**.
- Solution: Use **"Enterprise" versions** (like NC State's Google Workspace) where data isn't used for training (if applicable), or anonymize/mask data before upload.

The "Hallucination" Problem

- **Risk**: LLMs are bad at doing math in their "head" (e.g., asking "What is the average of these 500 numbers?").
- **Fix**: Code/Approach Generation. We don't ask the AI to calculate; we ask it to write code to calculate.

# Demo: annotating cell type in scRNAseq data

# The (initial) prompt

I have an scRNAseq data set but in excel form not the raw data. The columns are: gene, logfoldchange, pvals, pvals_adj, scores. There are multiple tabs, each with a different "Group" number. I am interested in annotating cell type in scRNAseq data, what cell type each cluster represents.

# The (initial) prompt

I have an scRNAseq data set but in excel form not the raw data. The columns are: gene, logfoldchange, pvals, pvals_adj, scores. There are multiple tabs, each with a different "Group" number. I am interested in annotating cell type in scRNAseq data, what cell type each cluster represents.

➔ Dataset was NOT uploaded
➔ Responses may vary, but will give you an approach to follow

# Follow-up prompt

Is there a way to do this programmatically, using AI to produce the formula/functions or scripts but not to do the analysis itself? I want a python code, and I want to be able to specify/modify my dictionary of cell types.

➔ Dataset was NOT uploaded
➔ Produces a script that can be run locally
➔ Highlights areas that you can modify

# Additional Tasks

"Where do I get the marker list?"

- The AI Trick: Tell them to ask Gemini to generate the dictionary first!
- Prompt: "Create a Python dictionary of gene markers for common cells found in [Insert Tissue Type, e.g., Skin/Wound Healing], including Immune cells, Fibroblasts, and Endothelial cells."

Worried about hallucination? Use NotebookLM!

- Upload articles, pdfs of other marker lists
- Upload your dataset, or the top genes and have NotebookLM “do” the work for you OR ask it to get the marker list.

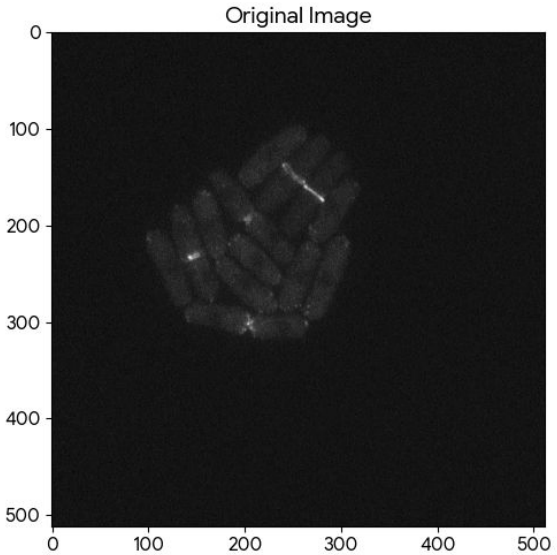
# How does AI enhance this workflow?

**Reproducibility**: You save the script. If they re-run their experiment next month with new data, they just click "Run" and it re-annotates everything.

**Scalability**: It works the same for 10 clusters or 1,000 clusters.

**Control**: You can add your own "Favorite Genes" to the dictionary, making the tool smarter over time

# Demo: Automating image analysis



# Why AI for image analysis?

**Context:** Quantitative microscopy (The .tif file). Measuring fluorescence in contractile rings.

**The Bottleneck:** Opening 100 images in ImageJ and hand-drawing circles (ROIs).

**The AI Solution:** Computer Vision scripts.

## The "Scripting" Prompt:

"I have a fluorescence microscopy image in .tif format. Write a Python script using the skimage library to:

1. Load the image.
2. Apply a Gaussian blur to reduce noise.
3. Apply an Otsu threshold to create a mask of the bright structures (the cells).
4. Measure the total 'Integrated Density' (sum of pixel intensities) inside the masked regions.
5. Print the result."

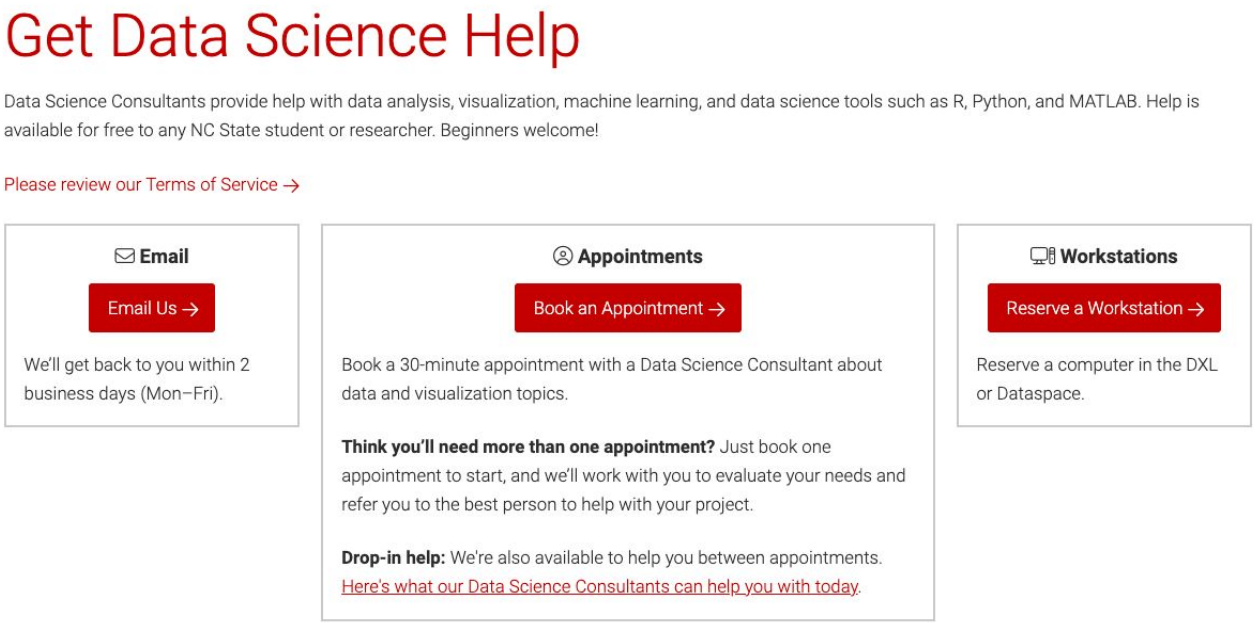
# AI Resources at NC State

- Approved AI tools: https://software.ncsu.edu/artificial-intelligence-it-purchase-process/
- AI Hub: https://datascienceacademy.ncsu.edu/ai/
- AI Policy for Theses & Dissertations
    - Overall guidance: “Candidates utilizing GAI tools must document their use clearly and thoroughly in their dissertation or thesis.”
- AI Tools in publishing
    - Different journals have different AI policies, but it’s rare to see it completely banned. Look under “author guidelines” for rules.

# HOW TO CONTACT US

go.ncsu.edu/getdatahelp

- Data requests: getdatahelp@ncsu.edu



## Get Data Science Help

Data Science Consultants provide help with data analysis, visualization, machine learning, and data science tools such as R, Python, and MATLAB. Help is available for free to any NC State student or researcher. Beginners welcome!

Please review our Terms of Service →

**Email**

Email Us →

We'll get back to you within 2 business days (Mon–Fri).

**Appointments**

Book an Appointment →

Book a 30-minute appointment with a Data Science Consultant about data and visualization topics.

**Think you'll need more than one appointment?** Just book one appointment to start, and we'll work with you to evaluate your needs and refer you to the best person to help with your project.

**Drop-in help:** We're also available to help you between appointments. Here's what our Data Science Consultants can help you with today.

**Workstations**

Reserve a Workstation →

Reserve a computer in the DXL or Dataspace.