

Undergraduate Education at UCSD on the NRP: Present and Future

Sam Lau

Assistant Teaching Professor, HDSI

7NRP Workshop

May 6, 2026



Early versions of DataHub @ Berkeley as an undergrad TA



2015-2018

2018-2023

2023+

 note 

388 views

Actions ▾

RETIRED: Official 500/503 Error Reporting Thread

THIS THREAD IS RETIRED (because it got too long).

503 errors are being reported all over the place, and it's getting hard to keep track of them all. We're not sure of the cause of the errors yet, but in the meantime please post a followup on this thread if you're getting a 503 error, and one of the TAs will start up your server.

Early versions of DataHub @ Berkeley as an undergrad TA

Research on data scientists + notebooks as a PhD student

2015–2018

2018–2023

2023+

Pivot tables: The pandas `.pivot_table()` function aggregates data into a pivot table, an operation inspired by spreadsheets.

Here is how Pandas Tutor visualizes an example pivot table call where the kids row index pivots into being a column index and the values in the longevity column aggregate into a 2x2 cross-tab:

```
dogs.pivot_table(index='size', columns='kids', values='longevity')
```

	size	kids	longevity
0	medium	low	8
1	small	high	12
2	small	high	13
3	medium	high	14
4	medium	high	10

	size	high	low
medium		12	8
small		12.50	NaN

Similar to reshapings, pivot tables can be hard to understand since passing in different values can result in very different outputs.

The screenshot shows a browser window with a Data Theater interface. On the left, Python code is displayed: `from random import choices, seed; seed(42); avgs = []; dice = 2; for _ in range(100): rolls = choices([1, 2, 3, 4, 5, 6], avg = sum(rolls) / len(rolls)); avgs.append(avg)`. A slider labeled 'Step: 87' is below the code. On the right, a declarative JSON spec is shown, mapping Python values to visual glyphs. The visualization on the far right shows a scatter plot of averages, with a slider at the bottom.

Figure 1. Using Data Theater to create an explorable that rolls two dice, calculates the average of the two rolls, and displays the history of all averages. (A) Instructors implement the core logic using Python code. (B) Instructors visualize the code through a declarative JSON spec that maps Python values to visual glyphs. (C) The visualization live-updates as either the Python or JSON change. (D) Explorables authored in Data Theater contain a single interactive element: a slider that steps forward and backward through execution steps and causes the visualization to animate as Python values change.

Early versions of DataHub @ Berkeley as an undergrad TA

Research on data scientists + notebooks as a PhD student

Teaching + research on AI in CS/DS education as teaching faculty

2015–2018

2018–2023

2023+

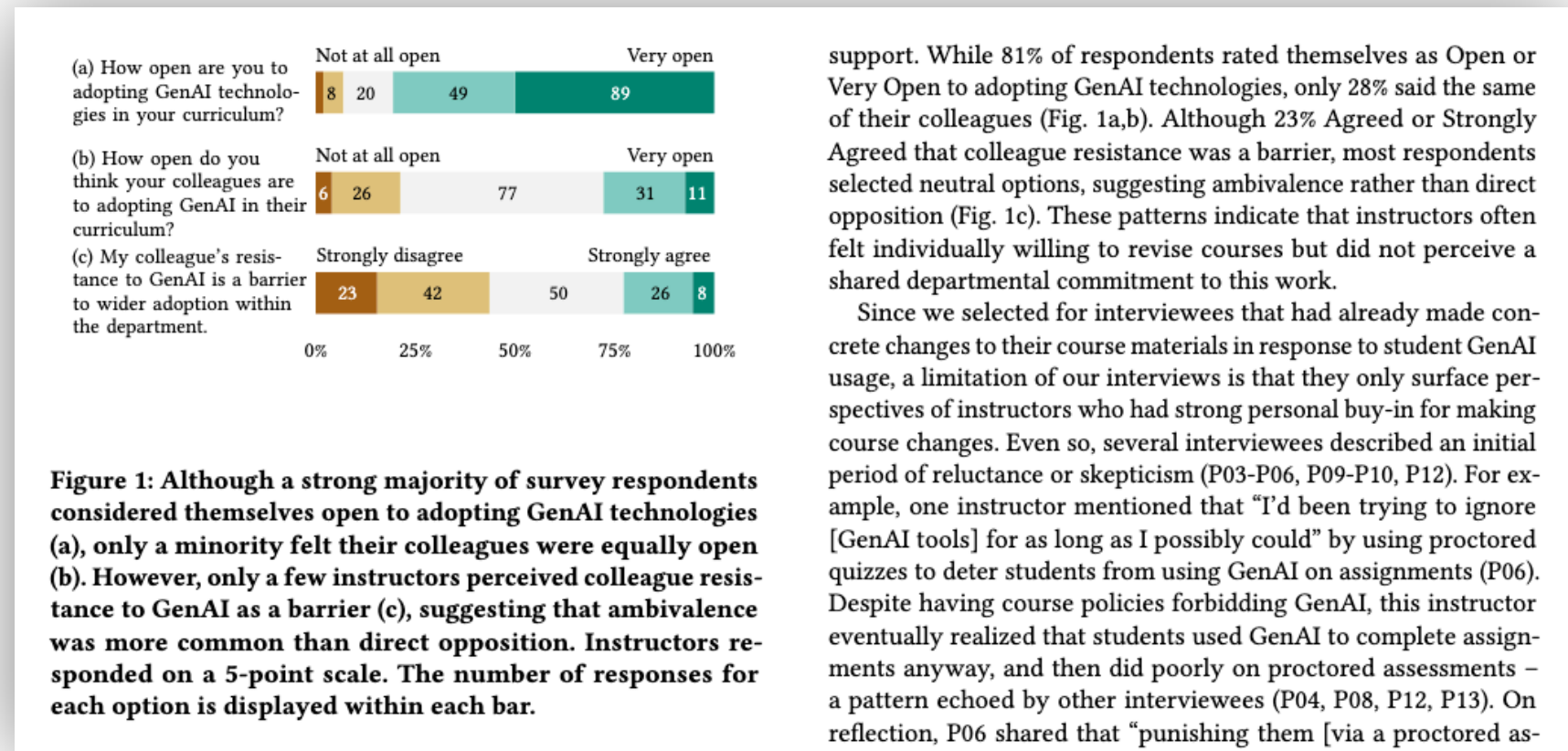
From "Ban It Till We Understand It" to "Resistance is Futile": How University Programming Instructors Plan to Adapt as More Students Use AI Code Generation and Explanation Tools such as ChatGPT and GitHub Copilot

Sam Lau
UC San Diego
La Jolla, California, USA
lau@ucsd.edu

Philip J. Guo
UC San Diego
La Jolla, California, USA
pg@ucsd.edu

ABSTRACT
Over the past year (2022–2023), recently-released AI tools such as ChatGPT and GitHub Copilot have gained significant attention from computing educators. Both researchers and practitioners have discovered that these tools can generate correct solutions to a variety of introductory programming assignments and accurately explain the contents of code. Given their current capabilities and likely advances in the coming years, how do university instructors

ACM Reference Format:
Sam Lau and Philip J. Guo. 2023. From "Ban It Till We Understand It" to "Resistance is Futile": How University Programming Instructors Plan to Adapt as More Students Use AI Code Generation and Explanation Tools such as ChatGPT and GitHub Copilot. In *Proceedings of the 2023 ACM Conference on International Computing Education Research V.1 (ICER '23 V1)*, August 7–11, 2023, Chicago, IL, USA. ACM, New York, NY, USA, 16 pages. <https://doi.org/10.1145/3568813.3600138>



A class with programming, pre-2017

Environment Setup



2 years ago in **General**



Hello,

1. I followed the instructions but the environment is 3.9 and I'm not sure if the correct python version is happening?

2 years ago in **General** PRIVATE 35 VIEWS

<https://github.com/conda-forge/miniforge>

[README](#) [License](#)

Optional support for PyPy in place of standard Python interpreter (aka "CPython") is provided in the installers with `-pypy3-` in their filename.

Download

Miniforge installers are available here: <https://github.com/conda-forge/miniforge/releases>

Miniforge3

Latest installers with Python 3.12 (*) in the base environment:

OS	Architecture	Minimum Version	Download
Linux	x86_64 (amd64)	glibc >= 2.17	Miniforge3-Linux-x86_64
Linux	aarch64 (arm64) (**)	glibc >= 2.17	Miniforge3-Linux-aarch64
Linux	ppc64le (POWER8/9)	glibc >= 2.17	Miniforge3-Linux-ppc64le
macOS	x86_64	macOS >= 10.13	Miniforge3-MacOSX-x86_64
macOS	arm64 (Apple Silicon) (***)	macOS >= 11.0	Miniforge3-MacOSX-arm64
Windows	x86_64	Windows >= 7	Miniforge3-Windows-x86_64

Miniforge3-Windows-x86_64 (1).exe isn't a recognized file type. It may be a virus or malware. Make sure you trust Miniforge3-Windows-x86_64 (1).exe before running it.

It won't let me run the executable to set up the mamba application

PRIVATE 90 VIEWS

```
event.<locals>.event_logger(wrapped, self, args, kwargs)
event.<locals>.event_logger(wrapped, self, args, kwargs)
apped return value
tebook.__init__(self, nb_path, tests_dir, tests_url_prefix, colab, jup
tadataConfig.from_notebook(cls, nb)
tadataConfig.__init__(self, user_config, documentation_mode, require_v
bool = False, require_valid_keys: bool = False) -> None:
ts")
```

What would happen if we made compute
free for instructors?

– *circa 2017*

Today: Fill out 1 form, then students can:

JupyterHub-UCSD

UC San Diego

DATAHUB & RESEARCH CLUSTER POWERED BY DSMLP

Information Technology Services - Academic Technology Services

Help - FAQ

JupyterHub

UC San Diego Data Research Platform

If you are unable to log in: Please try...

JupyterHub Home Token Cluster Status DSMLP

final_projec... (3) - JupyterLab

lec01.ipynb lab2.ipynb final_project.ipynb

File Edit View Run Kernel Git Nbgrader LaTeX Tabs Settings Help

Markdown git Validate Open in... Python 3 (ipykernel)

Question 1.1

While the 'latitude' and 'longitude' tell us the precise location of a meteorite, it would be helpful to have a coarser description of where a meteorite was recorded. The file 'data/continents.csv' is a CSV with two columns: 'id' (the ID of a meteorite) and 'continent' (the continent on which it was recorded).

Assign the variable `df` to a DataFrame that contains all of the columns already in `raw_df` with one additional column, 'continent', which includes the continent in which each meteorite was recorded. Set the index of `df` to 'id'.

Hint: The file 'data/continents.csv' does not have the same number of rows as `raw_df`, which means you will have to merge!

```
[4]: continents = bpd.read_csv('./data/continents.csv')
df = raw_df.merge(continents, on='id').set_index('id')

df
```

	id	name	reclass	seen_falling	mass	year	decade	latitude	longitude	conti
--	----	------	---------	--------------	------	------	--------	----------	-----------	-------

Simple 0 \$ 3 mai Python 3 (ipykern... Mem: 637.28 / 819... Mode: Co... Ln 1, ... final_projec... 1

Syncs with course roster

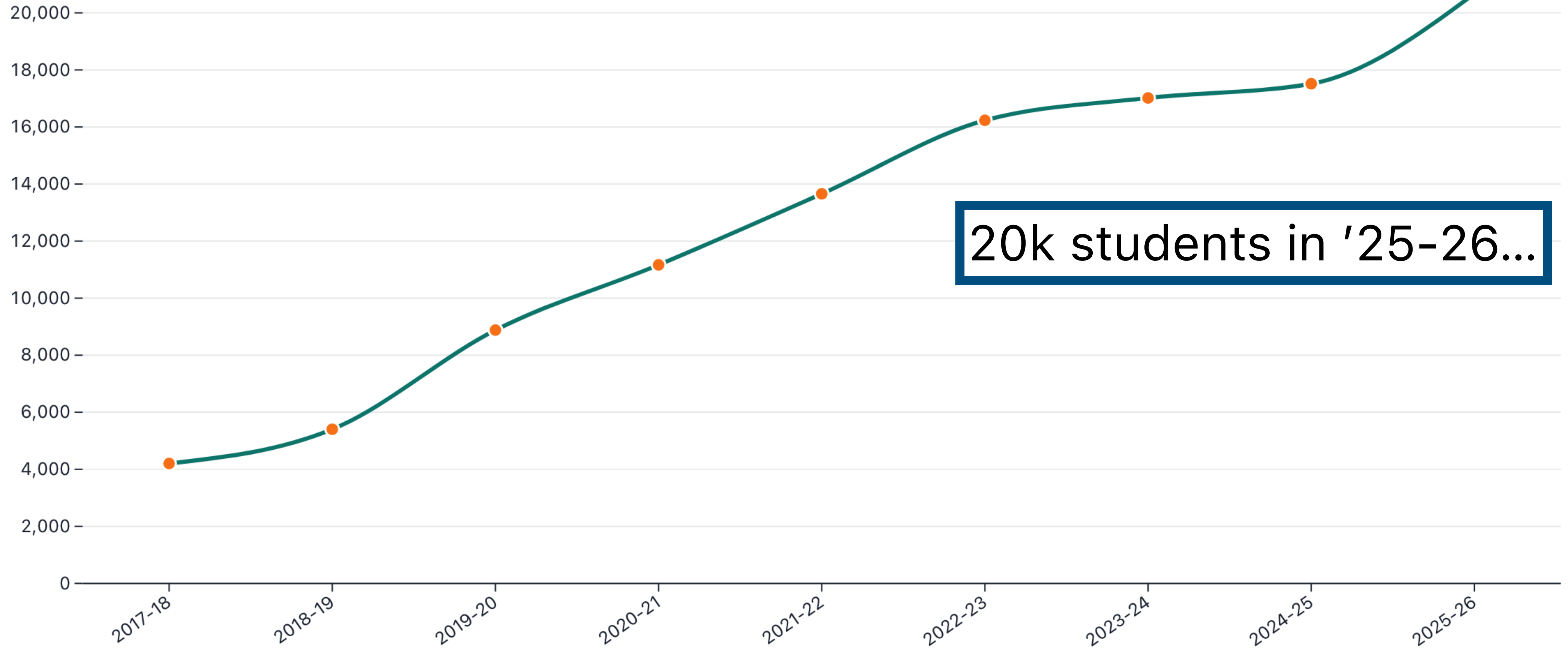
Select Your Notebook Environment

DSC10_WI26_A00 - Samuel Lau [WI26]
ghcr.io/ucsd-ets/dsc10-notebook:wi26 (2 CPU, 8G RAM)

Launch Environment

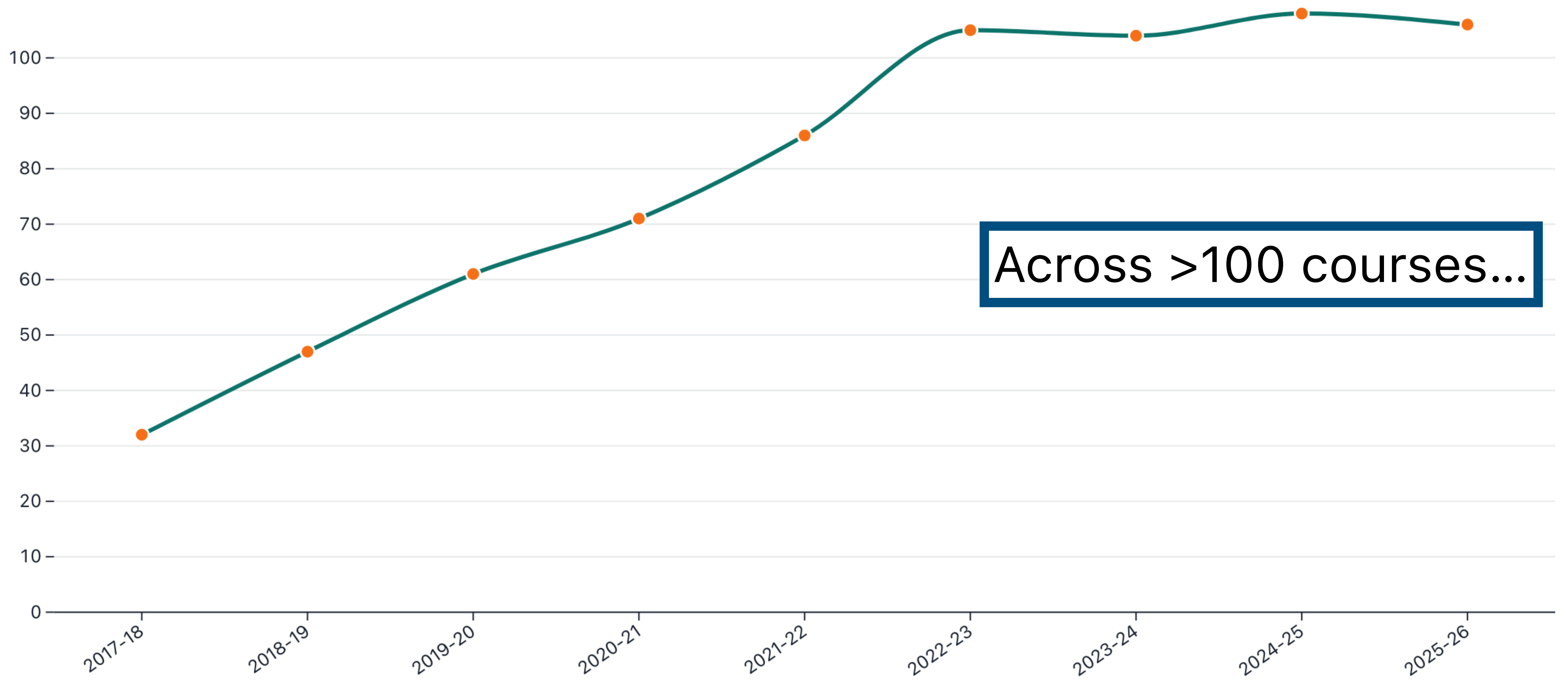
Students using DataHub at UCSD

↑ Students



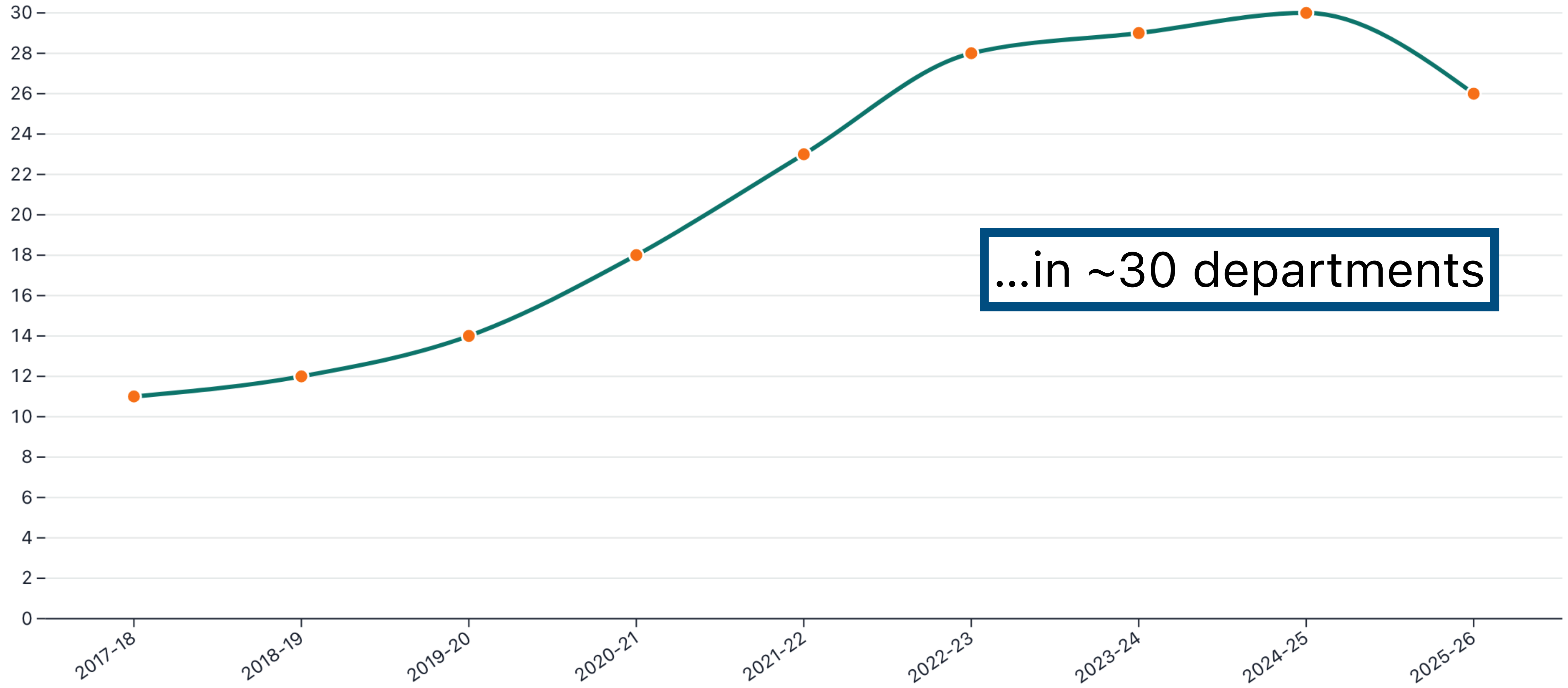
Unique courses using DataHub by academic year

↑ Courses



Unique departments using DataHub by academic year

↑ Departments



Ranging from a small visual arts seminar...

Deep Hysteria (2023) – #! Amy Alexander

● ○ ○ ○ ○ ○ ○ ○ ○ ○

The people in these artworks are "AI"-generated twins that vary in gender presentation — titled as another "AI," trained on human perceptions, identifies the emotion on their faces.

The faces are generated using deep learning algorithms trained on still frames of

Night" at panke.gallery, Berlin, 23 May
Talk/panel March 21: SIGGGRAPH DAC SPARKS !

Amy in the Social Networksphere

Facebook
Instagram
Bluesky

Subscribe to Amy's Sporadic Email
Updated

...to a required class for 6,000 biology majors
(just launched in 2024!)

The screenshot displays an RStudio environment within a Zoom meeting. The RStudio window is titled 'grader-bild5-03' and shows a coding lab with the following content:

Type the statement "Hello, World" (in quotes) on the line that begins with `q1 <-`, replacing the placeholder text "your answer here". This will assign the text "Hello, World" to an object called `q1`.

Then, insert a new line underneath and type `print(q1)`. Running this line of code will display the value of `q1` on the screen.

```
{r}
q1 <- "your answer here"
```

Question 2 (3 points)

Examine the course schedule. Which topic are you the most excited about? Why?

```
{r}
q2 <- "your answer here"
# Type your answer in the quotes above, replacing the placeholder text.
```

Question 3 (3 points)

Examine the course schedule. Which topic do you think will be the hardest or most confusing for you? Why?

```
{r}
q3 <- "your answer here"
# Type your answer in the quotes above, replacing the placeholder text.
```

The console at the bottom shows the following R code being executed:

```
R - R 4.3.3 ~ /
> # Run this code chunk to import the dataset and save it to the object 'admissions'.
> admissions <- read.csv('admissions_data.csv')
> admissions
>
```

The file explorer on the right shows a directory structure with folders for Coding Lab 01 through 05. A context menu is open over the file explorer, showing options like 'Copy...', 'Copy To...', 'Move...', and 'Copy Folder Path to Clipboard'.

The Zoom meeting interface is visible at the top and bottom of the screenshot, showing the meeting name 'Mingyu Yang's screen' and participants 'Mingyu Yang' and 'Sam Lau'.

Domino effect: since 1st years take this course, now upper-divs can start thinking about using data science / computation too

In Data Science (my home department)

Installing software is hard, just use Datahub

You need to learn how to manage your own environment!

Installing software is hard, just use Datahub

Year 1

Years 2-3

Year 4+

GPU drivers, distributed computing, TB+ datasets, etc.

But also in research + industry too

Installing software is hard, just use Datahub

Year 1

You need to learn how to manage your own environment!

Years 2-3

Installing software is hard, just use {**Databricks, SageMaker, cloud compute, etc.**}

Year 4+

GPU drivers, distributed computing, TB+ datasets, etc.

What would happen if we made compute
free for instructors?

– *circa 2017*

AI

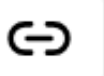
What would happen if we made ~~compute~~
free for instructors?

– *circa* ~~2017~~

2026

Progress in Open Weights vs. Proprietary Intelligence

2 of 3 license types

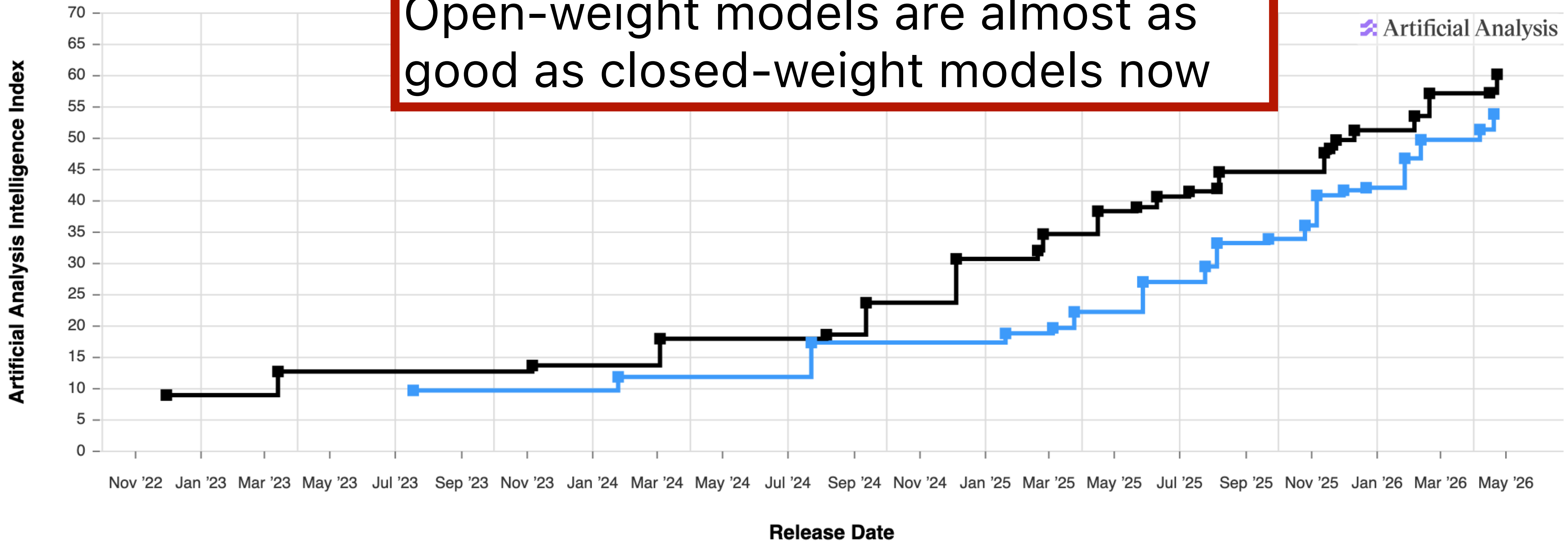


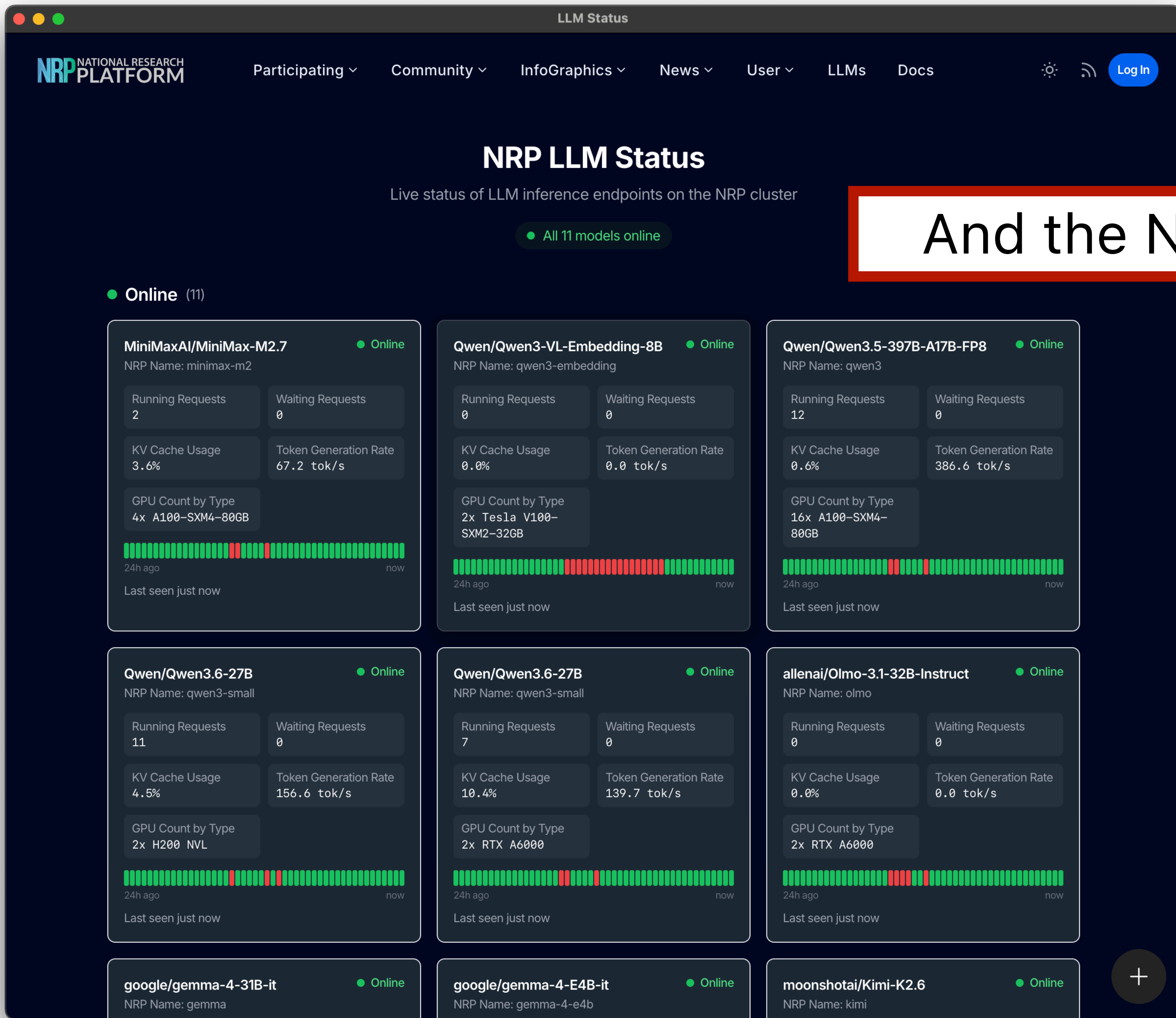
Artificial Analysis Intelligence Index v4.0 incorporates 10 evaluations: GDPval-AA, τ^2 -Bench Telecom, Terminal-Bench Hard, SciCode, AA-LCR, AA-Omniscience, IFBench, Humanity's Last Exam, GPQA Diamond, CritPt

■ Open Weights ■ Proprietary

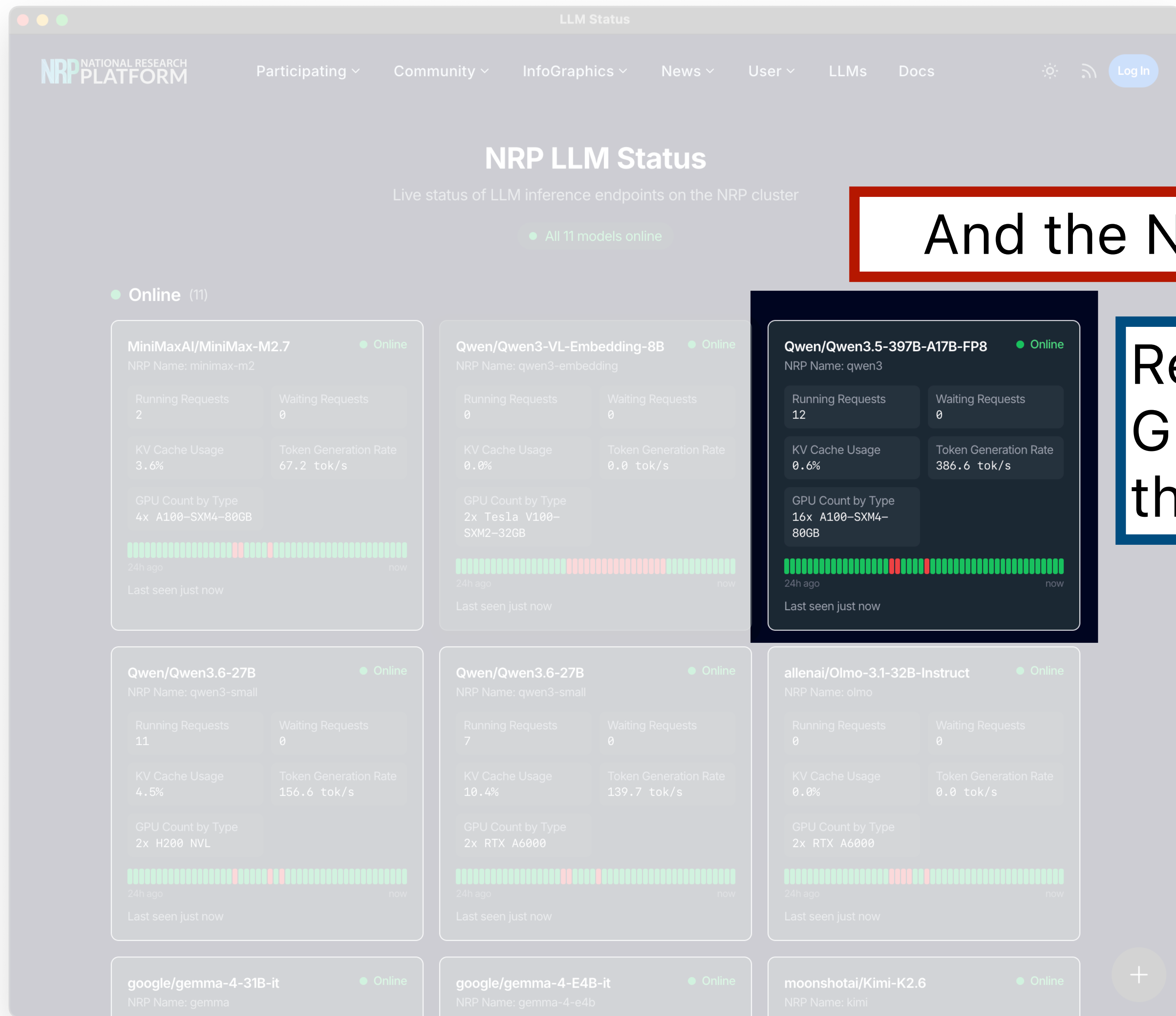
Open-weight models are almost as good as closed-weight models now

Artificial Analysis





And the NRP hosts them!

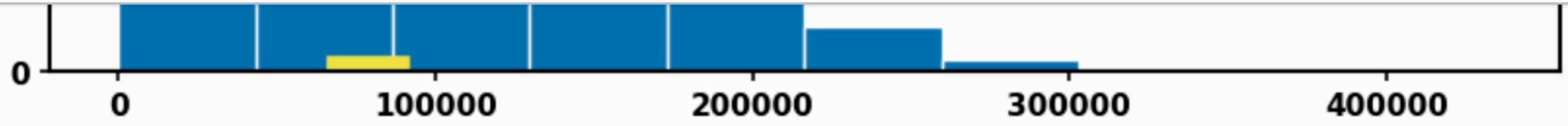


And the NRP hosts them!

Retail price of these GPUs is \$320k, by the way...

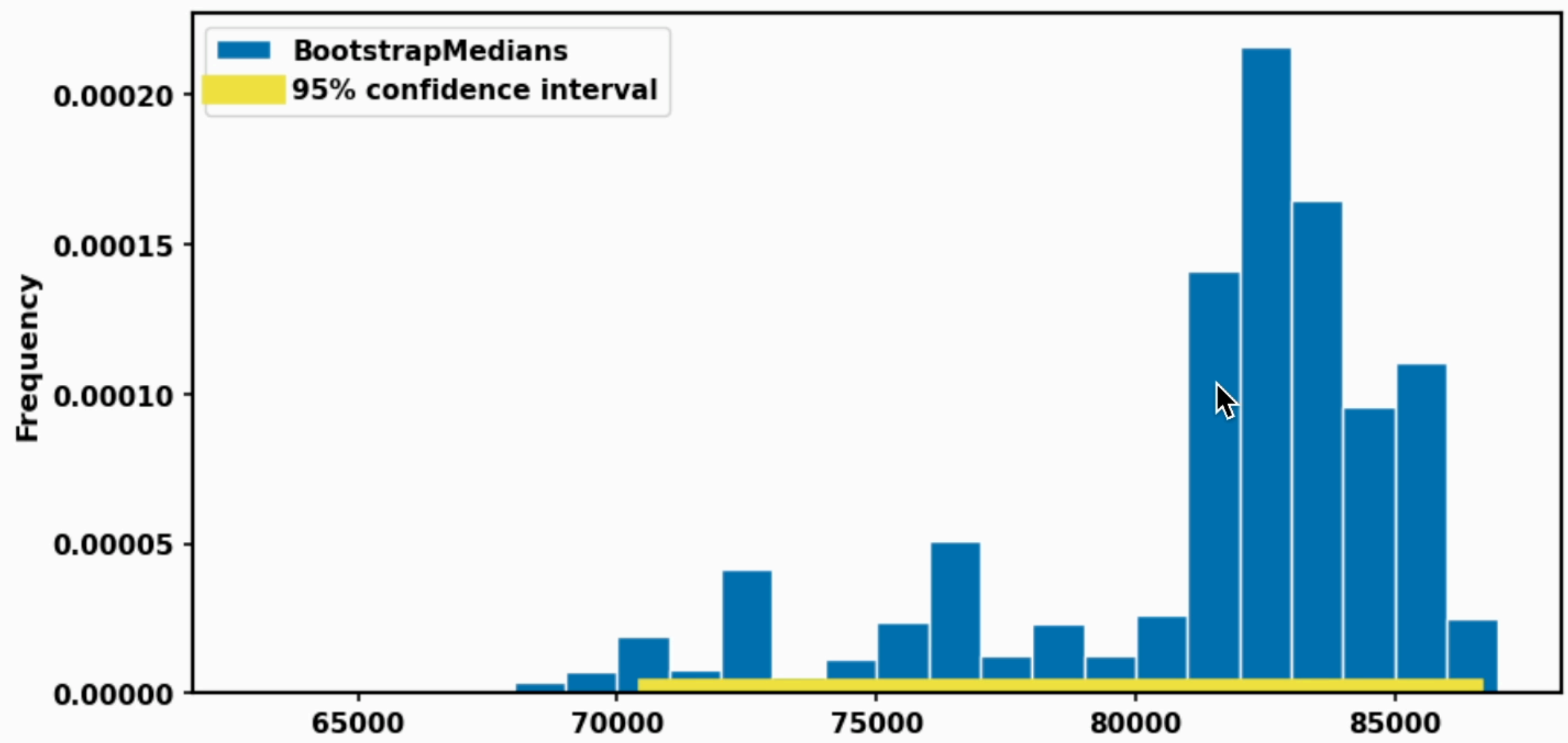
Launcher x lec15-live.ipynb +

+ ✂ 📄 📄 ▶ ⏪ ⏩ Code ⌚ git Validate ▾ Open in... 🗨️ 🛑 Python 3 (ipykernel) ☰



However, this interval does contain 95% of all bootstrapped median salaries.

```
[15]: bpd.DataFrame().assign(BootstrapMedians=boot_medians).plot(kind='hist', density=True, bins=
plt.plot([left, right], [0, 0], color=gold, linewidth=12, label='95% confidence interval');
plt.legend();
```



```
[16]: # Our interval:
[left, right]
```

[16]: [70671.5, 86405.0]

Is there is a 95% chance that this interval contains the population parameter? No! ❌

DSC 10 AI Tutor

New Conversation Tutor ChatGPT

Notebook: lec15-live.ipynb 112 cells loaded. I'm ready to help!

What does this code do?



Send

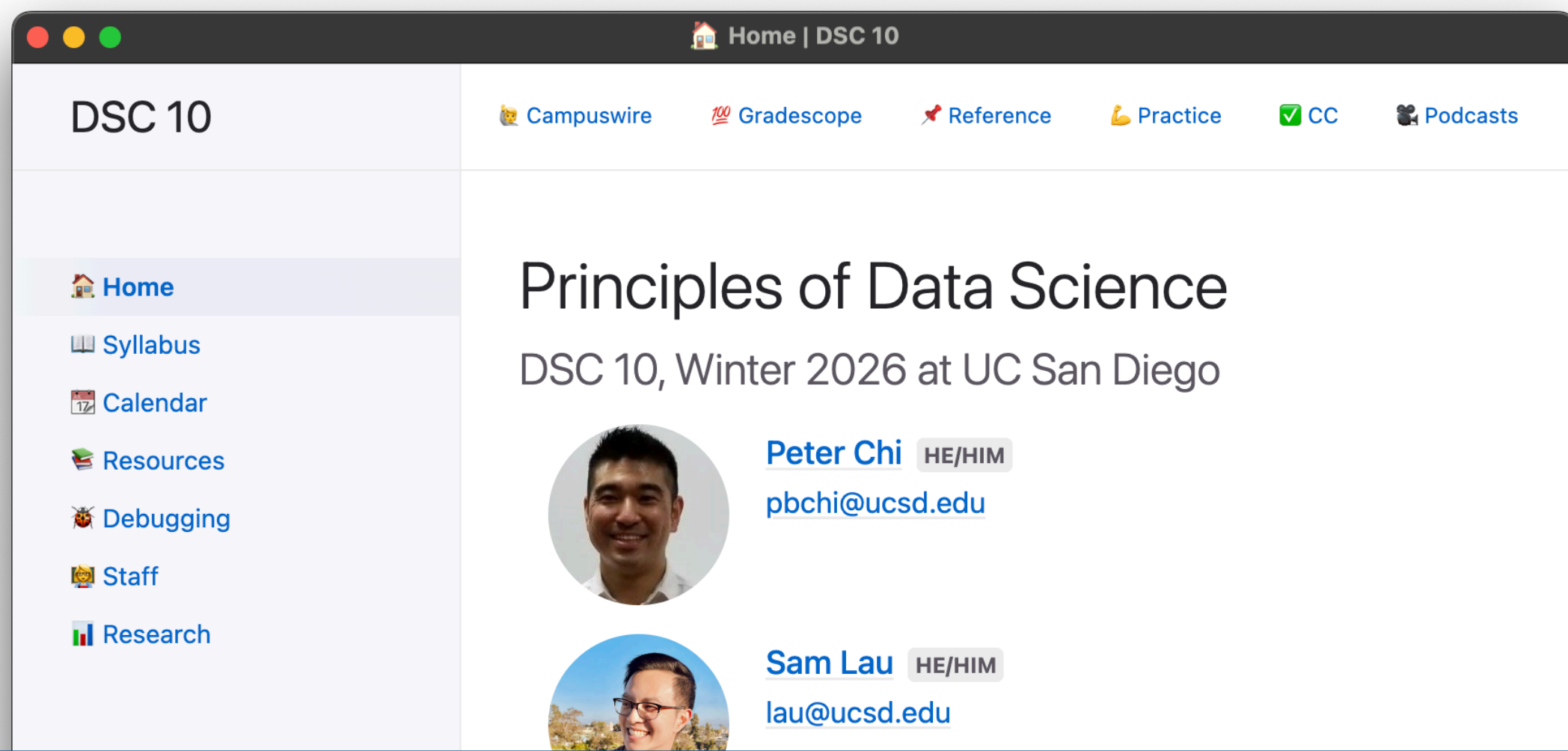
⚙️

🕒

AI Tutor

Create Assignment

🛑



Winter '26:
200 students sent 3k
messages per week

I set up my own logging (not available by default on NRP yet)

```
Mar 8, 03:37:21 PM tutor_query
is it allowed to use in babypandas?

Mar 8, 03:37:27 PM tutor_response
Yes, absolutely! The `~` (tilde) symbol is the standard way to flip (or negate) a sequence of Booleans in `babypandas`. It is completely valid and expected for you to use it in this course when doing Boolean filtering.

To put it all together, we know that:
1. `probabilities.index.isin(continent)` gives us `True` for the continents that are in the list.
2. Putting `~` in front of it flips those Booleans, giving us `True` for the continents that are NOT in the list.

Since we want to extract these remaining continents from `probabilities.index`, how would you finish the skeleton code provided in the notebook?

```python
some_continents = probabilities.index[
```

Give it a try and let me know what you come up with!
```

Students don't intuitively use AI tools well, but now I can monitor student usage and make suggestions about how to use AI better

“Once you graduate from UCSD, you’ll be able to do ~50% of what LLMs can do.”

– *circa 2026*

“Your skills + LLMs > LLMs alone.”

– *circa 2030?*

Domain knowledge
Technical knowledge
Mindset
AI knowledge
Wise delegation
etc.

+

NRP enables equitable LLM access at scale for instructors / students.

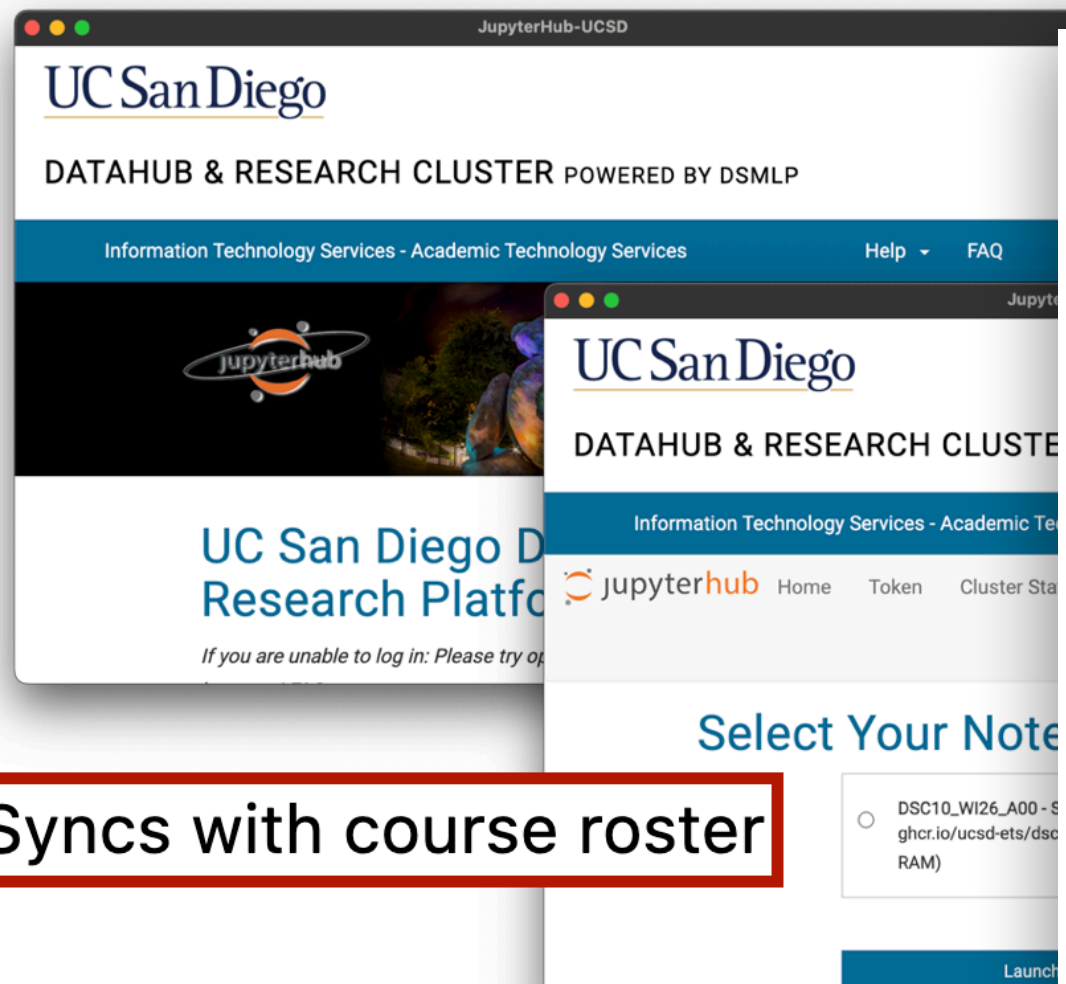
Lots of practice working with modern LLM tools on hard problems

“Your skills + LLMs > LLMs alone.”
– *circa 2030?*

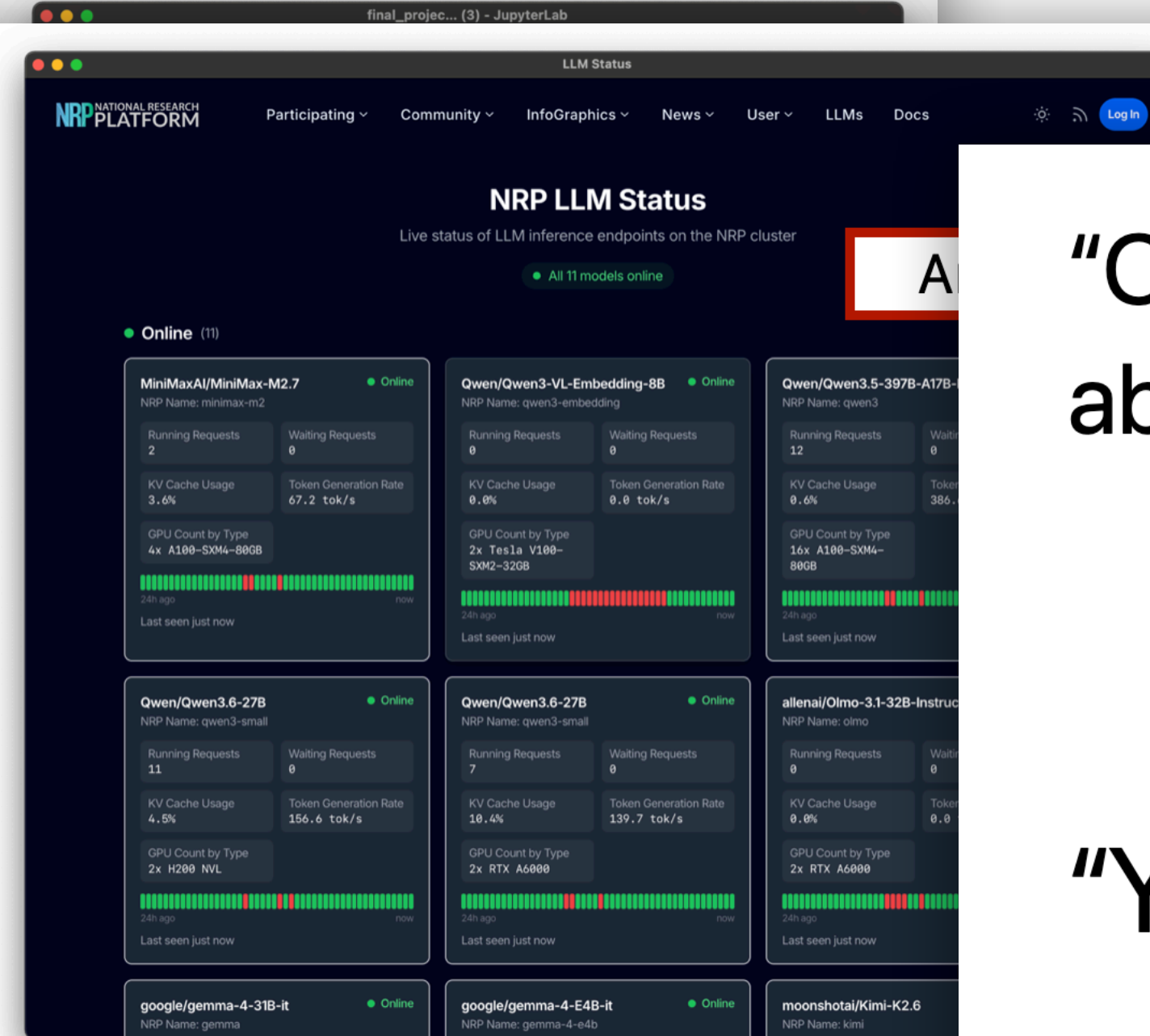
Thanks!

Email: lau@ucsd.edu

Today: Fill out 1 form, then students can:



Syncs with course roster



“Once you graduate from UCSD, you’ll be able to do ~50% of what LLMs can do.”
– circa 2026

“Your skills + LLMs > LLMs alone.”
– circa 2030?