



NFDI 4  
BIOIMAGE

NATIONAL RESEARCH DATA INFRASTRUCTURE  
FOR MICROSCOPY AND BIOIMAGE ANALYSIS

ScaDS.AI  
DRESDEN LEIPZIG

CENTER FOR SCALABLE DATA ANALYTICS  
AND ARTIFICIAL INTELLIGENCE



UNIVERSITÄT  
LEIPZIG

# How LLMs impact BioImage Data Science

Robert Haase



<https://doi.org/10.5281/zenodo.17669681>

Funded by

**DFG** Deutsche  
Forschungsgemeinschaft

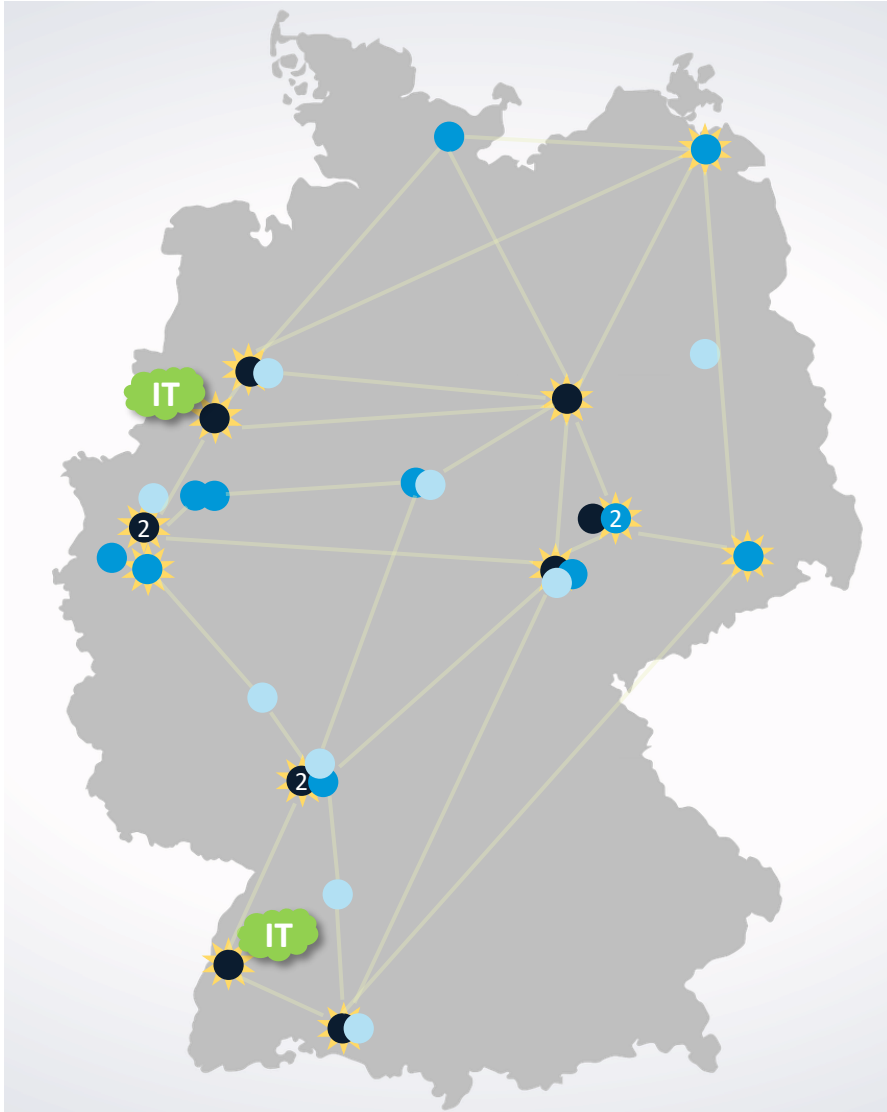


Bundesministerium  
für Forschung, Technologie  
und Raumfahrt



SACHSEN Diese Maßnahme wird gefördert durch die Bundesregierung  
aufgrund eines Beschlusses des Deutschen Bundestages.  
Diese Maßnahme wird mitfinanziert durch Steuermittel auf  
der Grundlage des von den Abgeordneten des Sächsischen  
Landtags beschlossenen Haushaltes.

# The consortium NFDI4BIOIMAGE



- 11 co-applicant institutions (6 Task Areas)  
Lead institution: Heinrich Heine University Düsseldorf  
Speaker: Prof. Dr. Stefanie Weidtkamp-Peters
- IT IT infrastructure (*storage & playground: @Uni Münster & @Uni Freiburg*)  
*But not: a central data archive for all bioimaging data*
- 12 participating institutions
- ★ **data stewards (DaSts) & research software engineers**
- community use cases

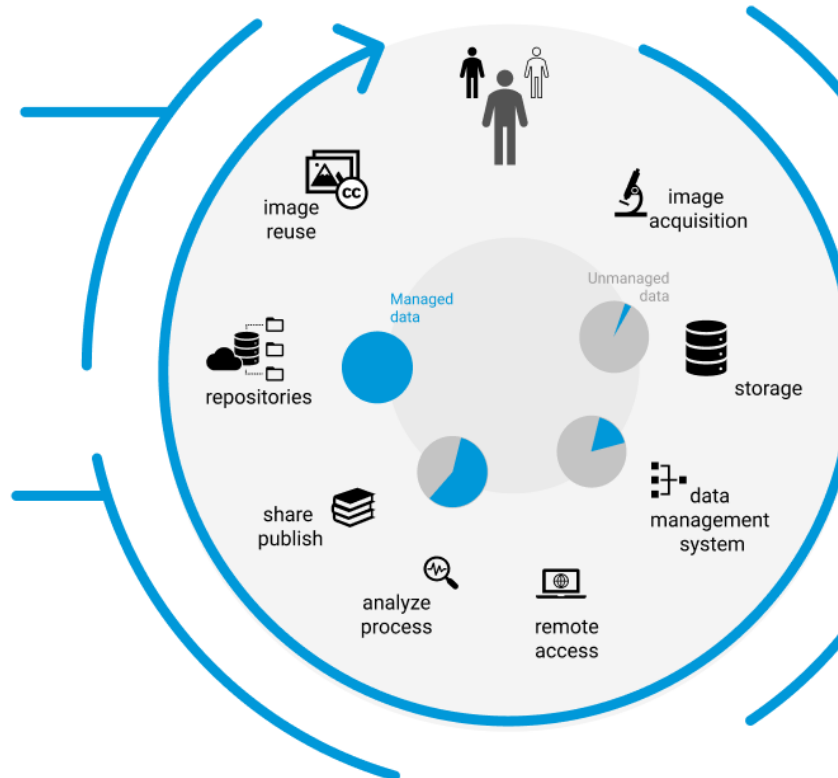
Community-oriented services (implemented or planned)

- Help Desk for use community support
- Training portfolio
- Bioimage RDM playground infrastructure
- Supporting reproducible image analysis

# NFDI4BIOIMAGE Objectives

**Objective 4**  
**Capacitate** researchers  
for FAIR image data  
management

**Objective 3**  
Maximize the reach  
of **reproducible** image  
analysis workflows  
in the community

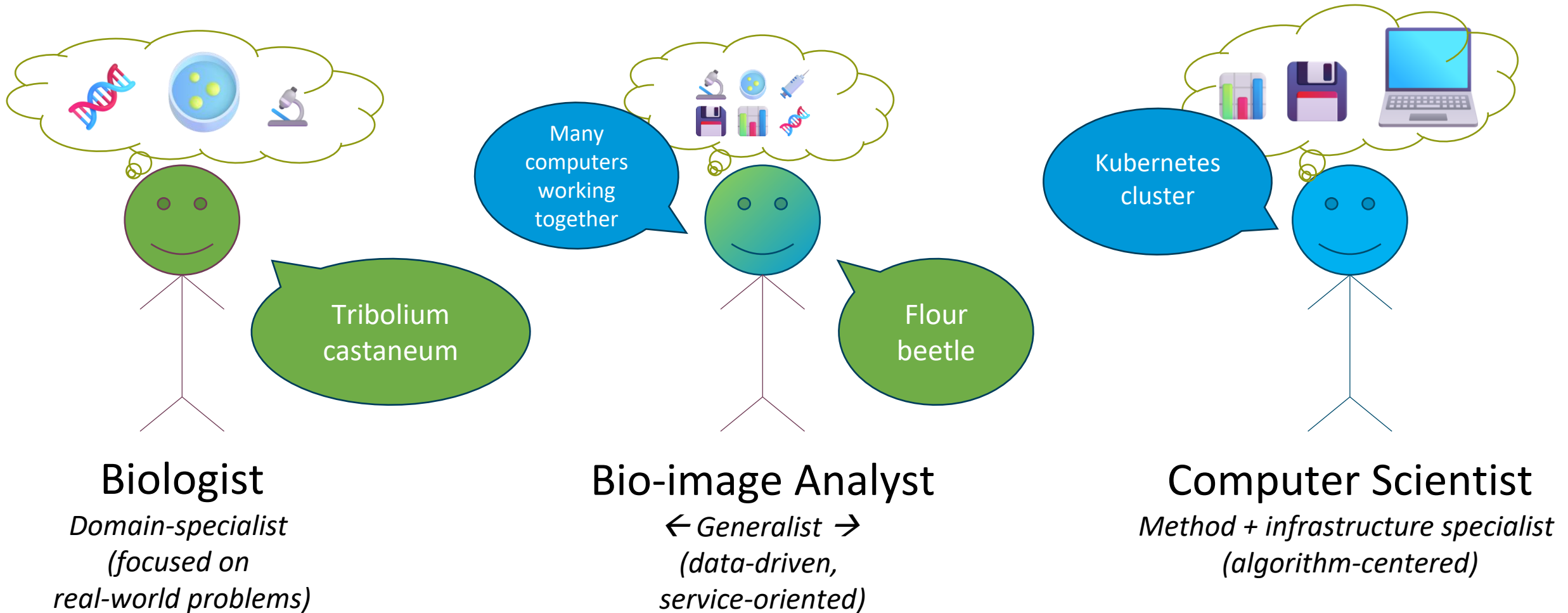


**Objective 1**  
Champion the  
**standardization**  
of the „bioimage  
data“ type

**Objective 2**  
Provide scalable  
**infrastructure**  
for FAIR image data

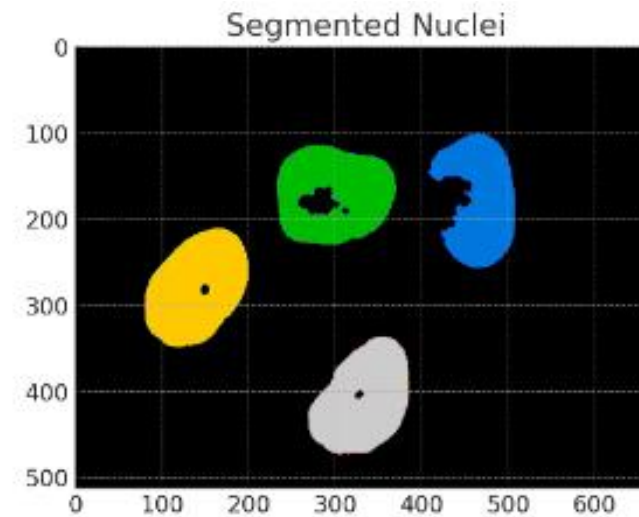
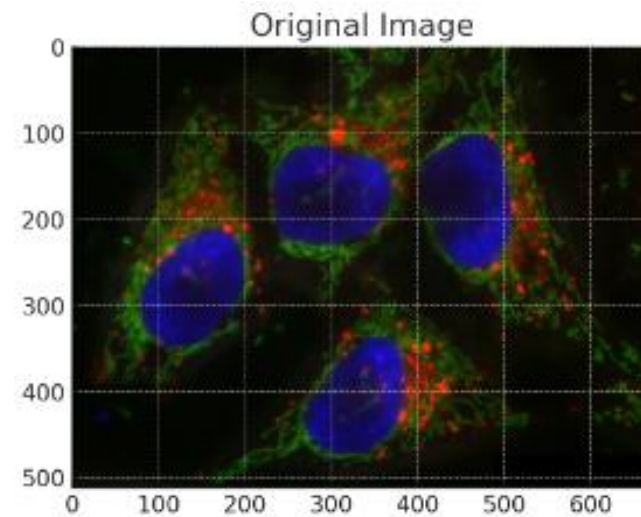
*Out of scope:* Medical, clinical, and preclinical imaging (MRT, PET, ultrasound, etc.)

# Bio-image Analysis



# Bio-image Analysis

- My job ...

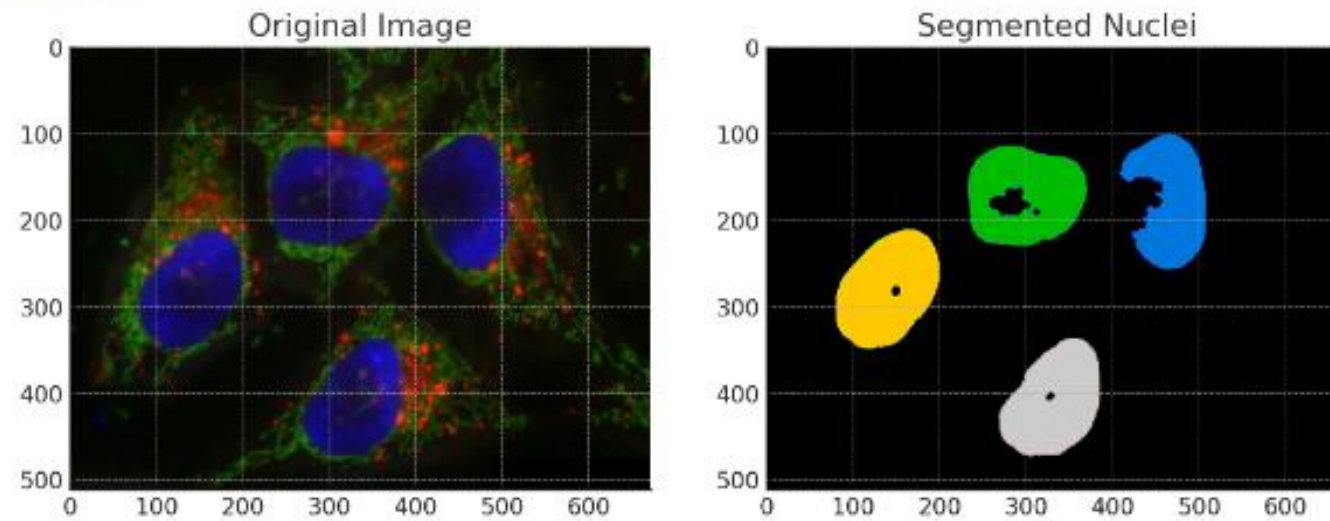


# Bio-image Analysis

- My job is changing, since we have ChatGPT

Prompt:  
“Please segment the  
nuclei in the blue  
channel of the  
image”

 ChatGPT

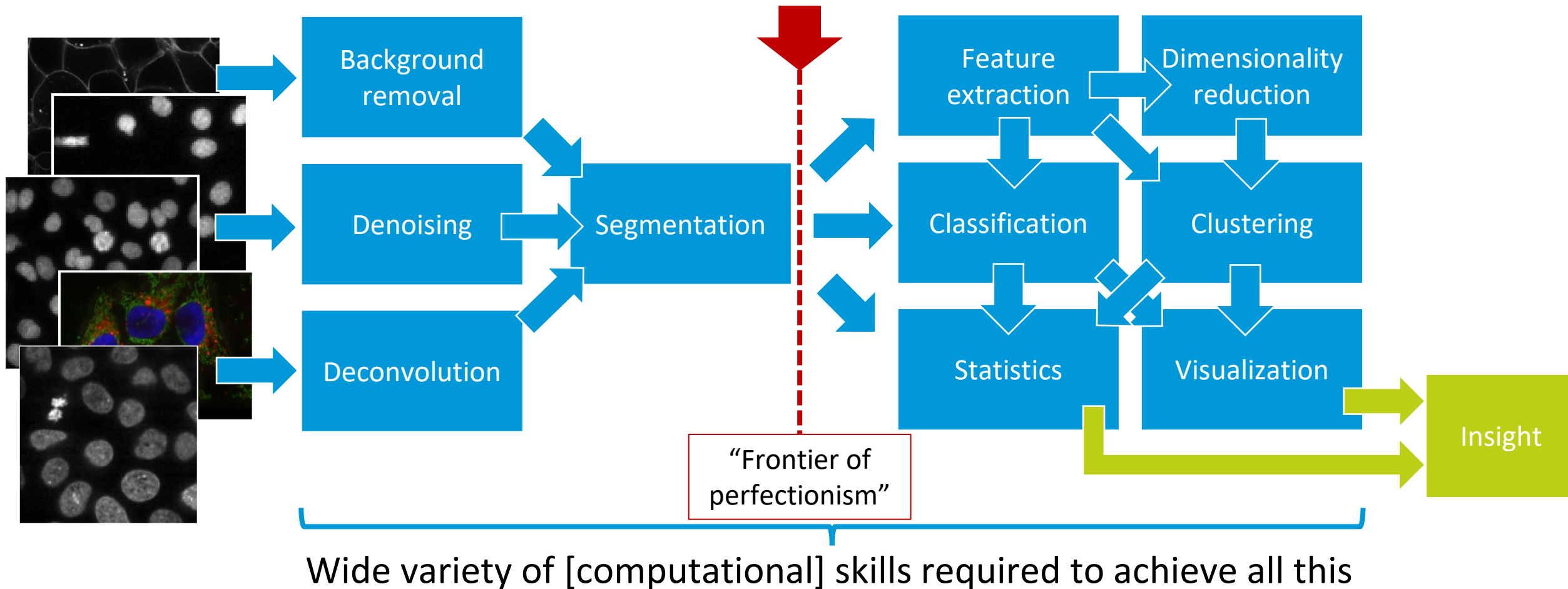


The segmentation of the nuclei has been performed on the blue channel of the image. The right side of the figure shows the segmented nuclei in different colors, each corresponding to a different nucleus in the field of view. Each color represents a unique label assigned to the individual nuclei detected in the image. [↗]



# Bio-image analysis beyond image segmentation

- Crossing this border seems challenging





# Consulting LLMs

ChatGPT 5 ▾

Can you help me writing code for denoising a microscopy image?

Thought for 19s

Absolutely! Below is a clear and concise Python script that identifies **common microscopy noise** and provides **denoising options** out of the box.

ChatGPT 5 ▾

Can you help me writing code for segmenting a microscopy image?

ChatGPT 5 ▾

Can you help me extracting features from a segmented microscopy image?

Thought for 31s

Absolutely! Here's a reliable, ready-to-run Python workflow to extract per-object features from a segmented microscopy image (labels) and the corresponding raw image (intensities). It works for 2D (and easily extends to 3D).

ChatGPT 5 ▾



Can you help me writing code for generating a UMAP from a table of measurements?

Thought for 9s

Absolutely! Here's a clean, production-ready Python example that reads a table (CSV/Excel), handles preprocessing (numeric + categorical features, missing values, scaling), runs UMAP, and plots the embedding—with or without labels.





# Consulting LLMs

 ChatGPT 5 ▾ 

Can you help me writing code for clustering objects based on a table of measurements

Thought for 5s

Absolutely! Below i  
rows (objects) from  
sensible workflow:


 Gemini   
2.5 Flash ▾


Can you help me writing code for clustering objects based on a table of measurements?

✦

What kind of objects are you trying to cluster? What kind of measurements do you have? And how many objects and measurements are we talking about?


If you could provide me with more details about your specific problem, I can give you a more targeted and useful response. However, here's a general guide and a Python code example for clustering objects based on a table of measurements.

☐ Data Clustering Code Implementation ▾ 



 Can you help me writing code for clustering objects based on a table of measurements?


I'd be happy to help you write clustering code! To create the most useful solution, I'll need to know a bit more about your specific situation:

**About your data:**

- What format is your measurement table in? (CSV, database, pandas DataFrame, etc.)
- What types of measurements do you have? (numerical, categorical, or mixed) 
- Roughly how many objects and how many measurement dimensions?

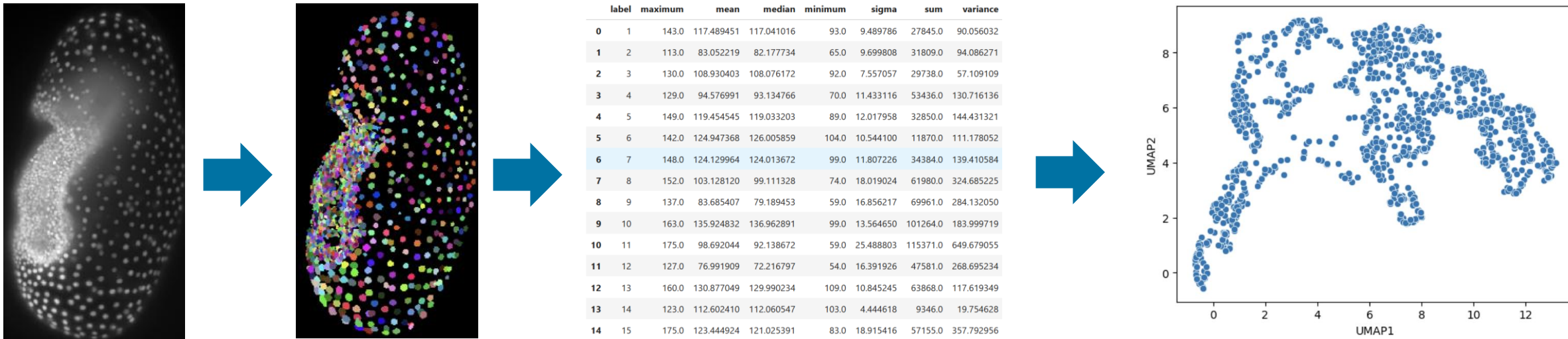
Reply to Claude...

Claude Sonnet 4 ▾ 

# Benchmarking LLMs for Bio-image Analysis

- Use-case: segment an image, extract features and create a UMAP.



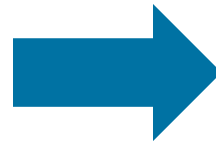
Unit-test pass-rate (n=10):

	reference	gpt-4-turbo-2024-04-09	Claude-3-opus-20240229	gpt-4-1106-preview	gpt-3.5-turbo-1106	gemini-pro	codellama
workflow_segment_measure_umap	1.0	0.8	0.7	0.9	0.1	0.2	0.0

# Benchmarking LLMs for Bio-image Analysis

- Use-case: compute the correlation matrix

	label	maximum	mean	median	minimum	sigma	sum	variance	elongation	feret_diameter
0	1	143.0	117.489451	117.041016	93.0	9.489786	27845.0	90.056032	1.228690	8.774964
1	2	113.0	83.052219	82.177734	65.0	9.699808	31809.0	94.086271	1.325096	13.152946
2	3	130.0	108.930403	108.076172	92.0	7.557057	29738.0	57.109109	1.565911	12.884099
3	4	129.0	94.576991	93.134766	70.0	11.433116	53436.0	130.716136	1.227027	14.352700
4	5	149.0	119.454545	119.033203	89.0	12.017958	32850.0	144.431321	1.429829	10.723805
5	6	142.0	124.947368	126.005859	104.0	10.544100	11870.0	111.178052	1.196911	9.273618
6	7	148.0	124.129964	124.013672	99.0	11.807226	34384.0	139.410584	1.137575	13.114877
7	8	152.0	103.128120	99.111328	74.0	18.019024	61980.0	324.685225	1.373404	14.866069
8	9	137.0	83.685407	79.189453	59.0	16.856217	69961.0	284.132050	1.205095	14.456832
9	10	163.0	135.924832	136.962891	99.0	13.564650	101264.0	183.999719	1.169756	14.730920
10	11	175.0	98.692044	92.138672	59.0	25.488803	115371.0	649.679055	1.201427	16.911535
11	12	127.0	76.991909	72.216797	54.0	16.391926	47581.0	268.695234	1.272889	15.066519
12	13	160.0	130.877049	129.990234	109.0	10.845245	63868.0	117.619349	1.143453	13.152946
13	14	123.0	112.602410	112.060547	103.0	4.444618	9346.0	19.754628	2.025953	9.110434
14	15	175.0	123.444924	121.025391	83.0	18.915416	57155.0	357.792956	1.339038	13.892444



	label	maximum	mean	median	minimum	sigma	sum	variance	elongation	feret_diameter
label	1.000000	-0.605035	-0.651268	-0.637158	-0.581233	-0.136779	-0.170934	-0.134539	-0.014857	0.105859
maximum	-0.605035	1.000000	0.824653	0.787135	0.577706	0.576118	0.522637	0.563160	-0.028076	0.144944
mean	-0.651268	0.824653	1.000000	0.997051	0.918750	0.040456	0.157387	0.052848	0.122359	-0.173872
median	-0.637158	0.787135	0.997051	1.000000	0.928168	-0.013701	0.111119	-0.002710	0.133479	-0.205403
minimum	-0.581233	0.577706	0.918750	0.928168	1.000000	-0.317573	-0.057844	-0.273489	0.217240	-0.311868
sigma	-0.136779	0.576118	0.040456	-0.013701	-0.317573	1.000000	0.633356	0.959845	-0.250655	0.436449
sum	-0.170934	0.522637	0.157387	0.111119	-0.057844	0.633356	1.000000	0.602559	-0.074555	0.775197
variance	-0.134539	0.563160	0.052848	-0.002710	-0.273489	0.959845	0.602559	1.000000	-0.191963	0.370870
elongation	-0.014857	-0.028076	0.122359	0.133479	0.217240	-0.250655	-0.074555	-0.191963	1.000000	0.184445
feret_diameter	0.105859	0.144944	-0.173872	-0.205403	-0.311868	0.436449	0.775197	0.370870	0.184445	1.000000

Unit-test pass-rate (n=10):

	reference	gpt-4-turbo-2024-04-09	Claude-3-opus-20240229	gpt-4-1106-preview	gpt-3.5-turbo-1106	gemini-pro	codellama
pair_wise_correlation_matrix	1.0	1.0	1.0	0.9	1.0	0.5	0.1

# Benchmarking LLMs for Bio-image Analysis

- Use case: Open a zarr file



Unit-test pass-rate (n=10):

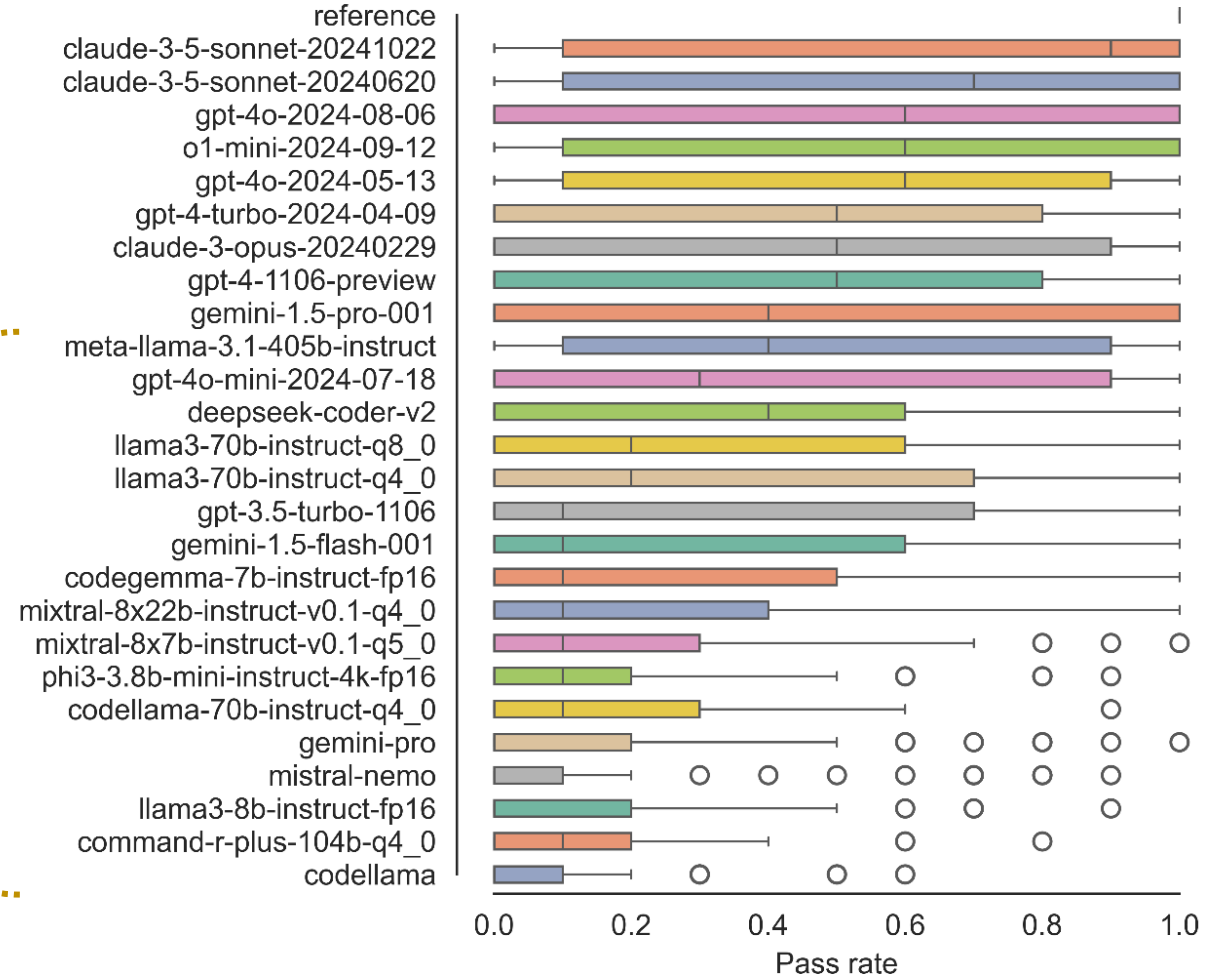
	reference	gpt-4-turbo-2024-04-09	Claude-3-opus-20240229	gpt-4-1106-preview	gpt-3.5-turbo-1106	gemini-pro	codellama
open_zarr	1.0	0.0	0.7	0.0	0.0	0.2	0.0

# Benchmarking LLMs for Bio-image Analysis

- Summary: 57 use-cases (yet), 26 LLMs (yet), n=10

Check out KISSKI ChatAI and  
Helmholtz Blablador:  
<https://kisski.gwdg.de/leistungen/2-02-llm-service/>  
<https://helmholtz-blablador.fz-juelich.de/>

Possibly  
available on  
institutional  
or national  
infrastructure



# Can LLMs solve real-world GitHub issues?

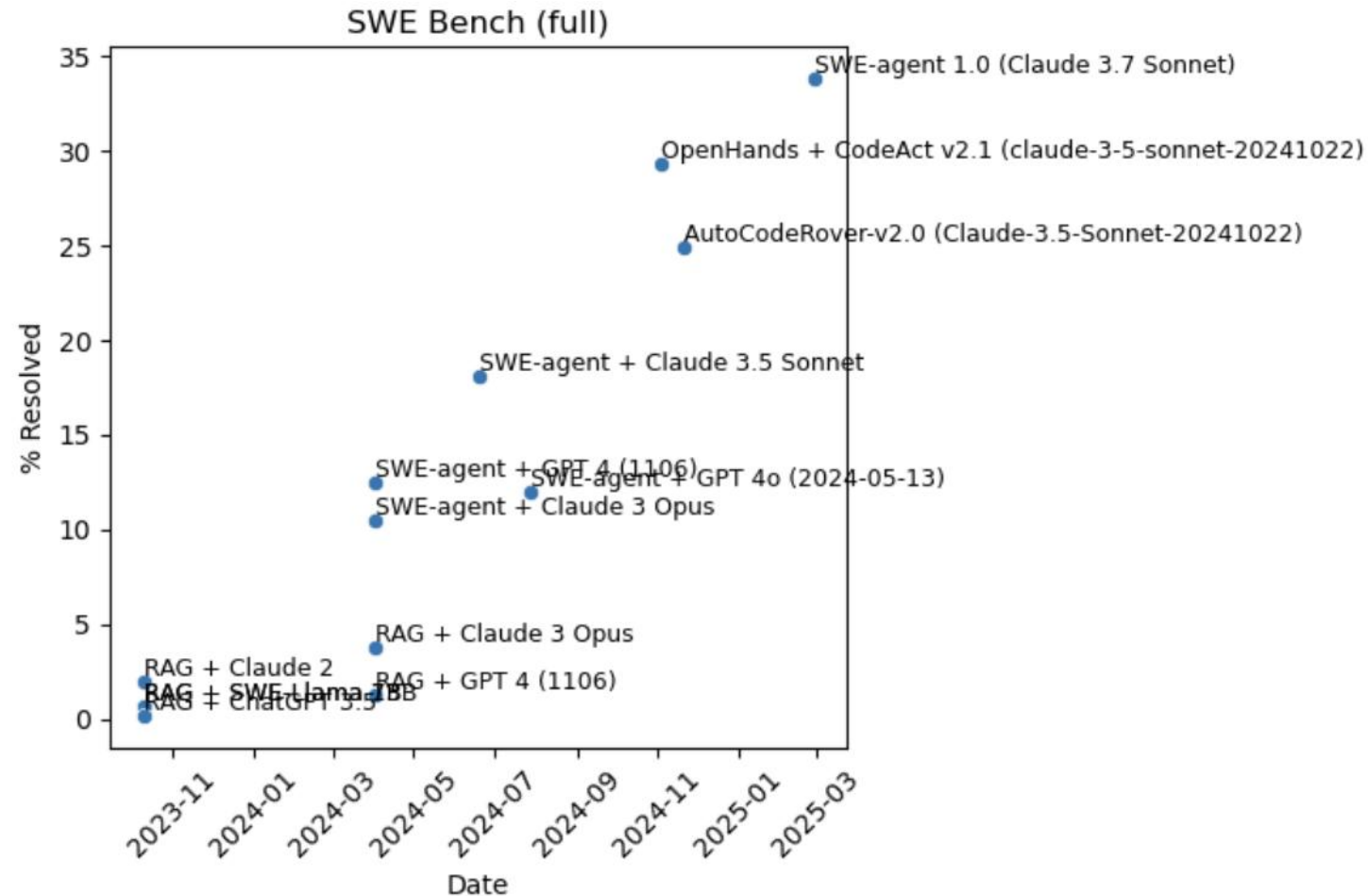
Published as a conference paper at ICLR 2024

## SWE-BENCH: CAN LANGUAGE MODELS RESOLVE REAL-WORLD GITHUB ISSUES?

Carlos E. Jimenez<sup>\* 1,2</sup> John Yang<sup>\* 1,2</sup> Alexander Wettig<sup>1,2</sup>  
Shunyu Yao<sup>1,2</sup> Kexin Pei<sup>3</sup> Ofir Press<sup>1,2</sup> Karthik Narasimhan<sup>1,2</sup>

<sup>1</sup>Princeton University <sup>2</sup>Princeton Language and Intelligence <sup>3</sup>University of Chicago

Model	SWE-bench	
	% Resolved	% Apply
Claude 3 Opus	<b>3.79</b>	46.56
Claude 2	1.97	43.07
ChatGPT-3.5	0.17	26.33
GPT-4-turbo	1.31	26.90
SWE-Llama 7b	0.70	51.74
SWE-Llama 13b	0.70	<b>53.62</b>



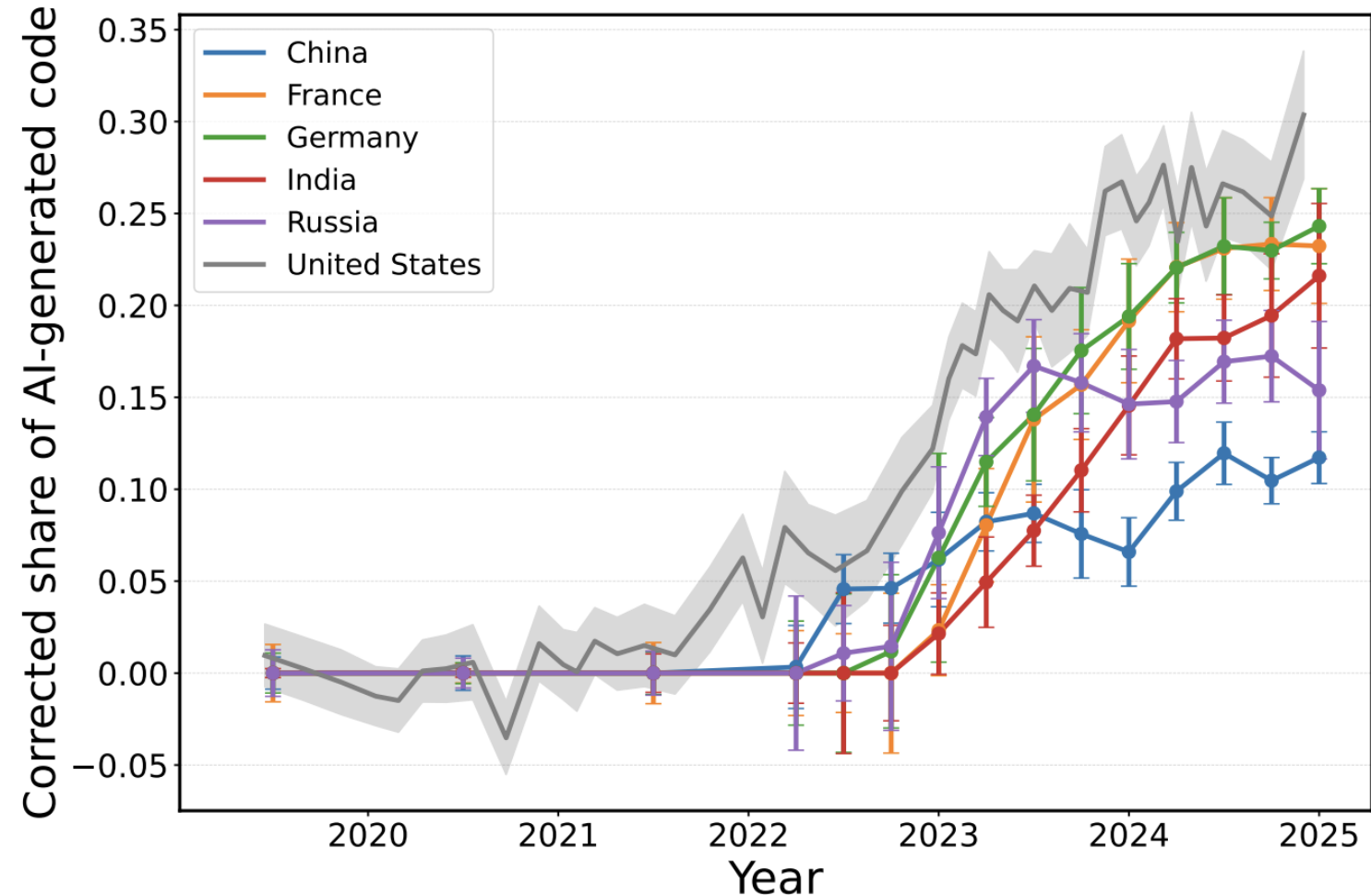
Source: Jimenez et al (2024),  
licensed [CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)  
<https://arxiv.org/abs/2310.06770>

Plot data source:  
<https://www.swebench.com/>  
(2025-06-21)



# AI-generated Code

- ... appears online more and more
- Indicator: Share of AI-generated code to Python-Contributions on Github.com
- (Hint: Presumable not submitted by AI-agents, but by humans using ChatGPT)



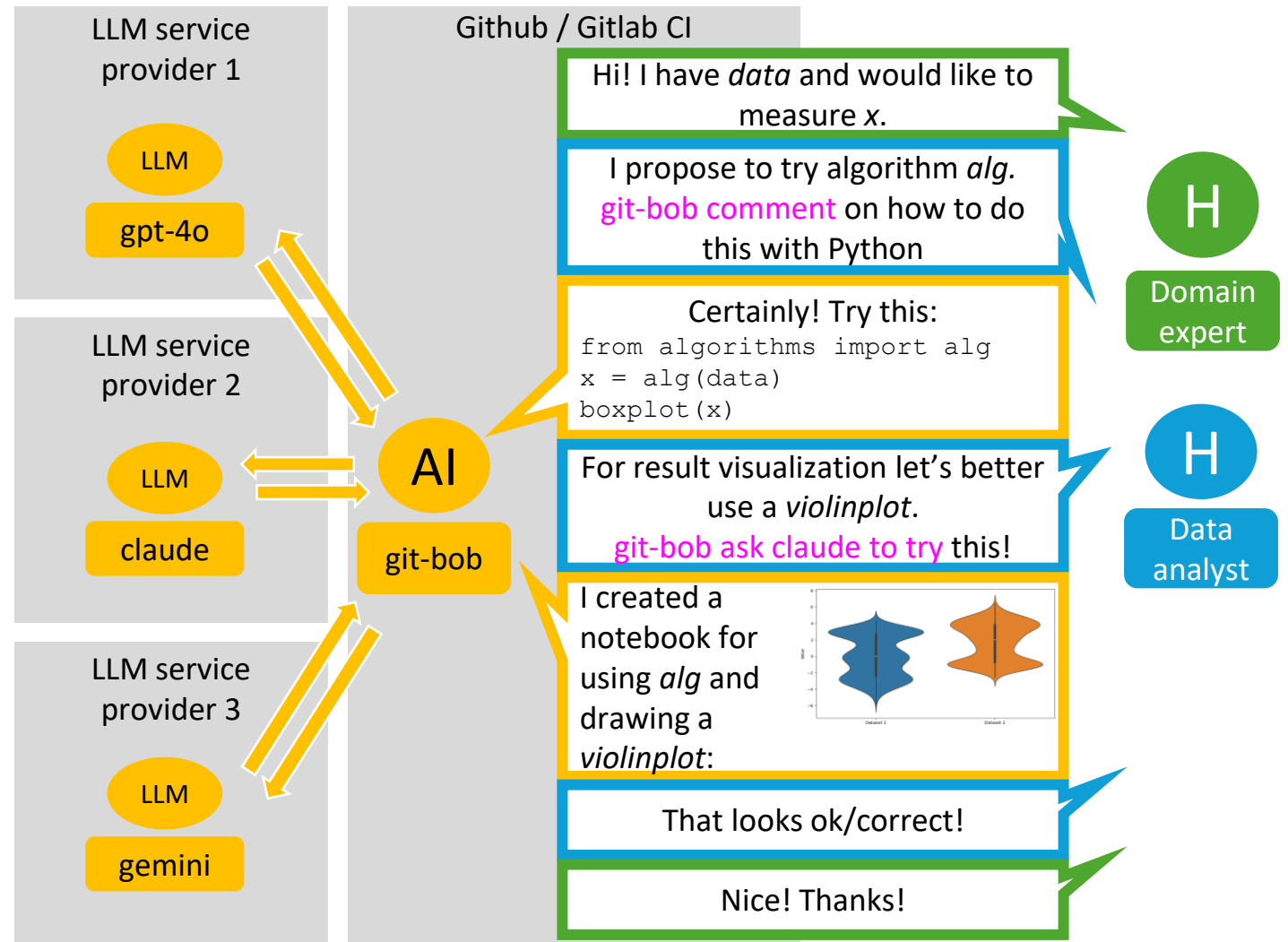
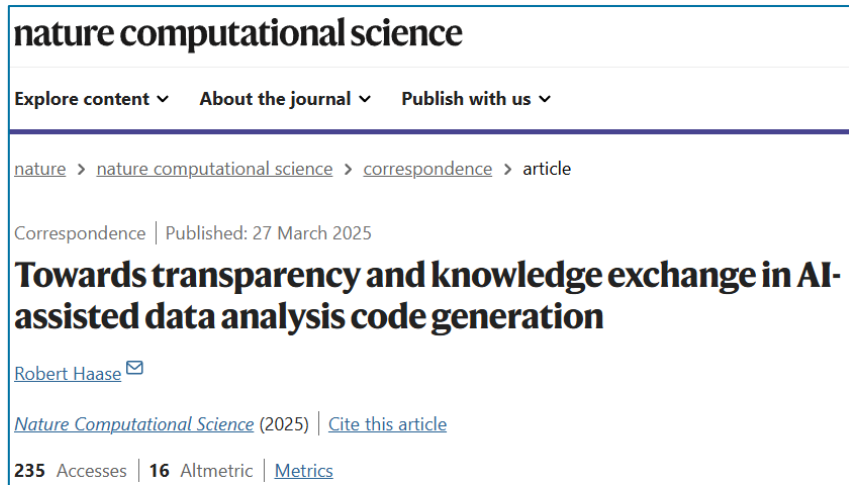
Source: Daniotti et al (2025), licensed [CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)

<https://arxiv.org/abs/2506.08945>



# Collaborative working with AI assistants

- New AI-assisted coding solution appear *daily*
- Human-human-AI interaction
- Transparency
- Knowledge exchange



# Summary & Outlook

- Using AI / LLMs in [bioimage] data analysis seems more and more unavoidable
- Maintaining good scientific practice is challenging (esp. human-AI peer-review)
- Relying on institutional infrastructure exclusively is important but hard
- Learn how to use AI *responsibly* (-> **train the trainers**)
- We are on the edge towards AI-agents analyzing image data autonomously.



<https://doi.org/10.5281/zenodo.17669681>

# Bio-Image Analysis Communities

- Image/data analysis + data management support
- Networking opportunities
- Trainings



National Infrastructure for Research  
Data Management in BioImaging

<https://nfdi4bioimage.de/help-desk/>



Image Science Community Forum

<https://image.sc/>



<https://www.globias.org/about-globias/globias-association>

# Acknowledgements

## Collaborators & contributors

Christian Tischer (EMBL Heidelberg), Jean-Karim Hériché (EMBL Heidelberg), Nico Scherf (MPI CBS Leipzig), Loïc A. Royer (CZ Biohub), Elena Nicolay (UFZ Leipzig), Kevin Yamauchi (ETH Zurich / Basel), Seth Hinz, Teun A.P.M. Huijben, Jordão Bragantini (CZ Biohub San Francisco), Ilan Theodoro (CZ Biohub and the State University of Campinas, Brazil), Ian C. (University of British Columbia), Pradeep Rajasekhar (WEHI Australia)

## BIDS Team @ ScaDS.AI / UL



Lea Kabjesz, Lea Gihlein, Mara Lampert



<https://doi.org/10.5281/zenodo.17669681>

## Communities & platforms



NFDI4  
BIOIMAGE



image.sc

## Funding



Bundesministerium  
für Forschung, Technologie  
und Raumfahrt



SACHSEN Diese Maßnahme wird gefördert durch die Bundesregierung aufgrund eines Beschlusses des Deutschen Bundestages. Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.



Deutsche  
Forschungsgemeinschaft

RH acknowledges the financial support by the Federal Ministry of Education and Research of Germany and by Sächsische Staatsministerium für Wissenschaft, Kultur und Tourismus in the programme Center of Excellence for AI-research „Center for Scalable Data Analytics and Artificial Intelligence Dresden/Leipzig“, project identification number: ScaDS.AI