

BRIDGE

National Center to Democratize Multi-Disciplinary Data Science and AI



Prof. Chen Li
UC Irvine

Rui Chen, Steve Parker, Ardalan Amiri Sani, Naomi Sugie
Kristin Turney, Wei Wang, Xiangmin Xu

Team members



Chen Li
Lead PI | UCI
Data systems • AI/ML



Ardalan Amiri Sani
Co-PI | UCI
Privacy • Security



Rui Chen
Co-PI | UCI
Single-cell genomics



Kristin Turney
Co-PI | UCI
Social science



Wei Wang
Co-PI | UCLA
AI • Machine learning



Xiangmin Xu
Sr. Personnel | UCI
Neurobiology



Naomi Sugie
Sr. Personnel | UCLA
Sociology



Stephen Parker
Sr. Personnel | UMich
Bioinformatics

Data science and AI in research communities



1 Python analysis

```
analysis.py

import pandas as pd

df =
pd.read_csv("data.csv")
clean = wrangle(df)
model = train(clean)
plot_results(model)
```

2 Push code

```
$ git add analysis.py
README.md
$ git commit -m "share
analysis"
$ git push origin main
```

3 GitHub repository

```
research-analysis/

• data/
• results/
  analysis.py
  README.md
  requirements.txt
```

README.md

```
# Reproducible analysis
How to run the code
and reproduce figures.
```

Public code + README

Needs and Challenges

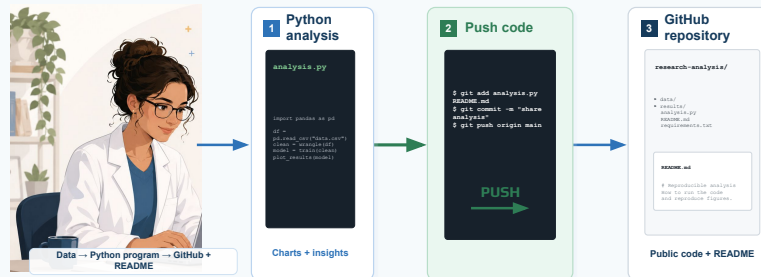
Coding requirements

Barriers to use AI/ML techniques

Reusability and reproducibility

Need of collaboration and sharing

Limited computing resources



Why now?

Recent surge in AI technologies

- More researchers want to adopt AI
- Face barriers in skills and computing resources

Latest technological advances

- Faster networks
- Browser-based solutions (e.g., CRDT)
- Open-source parallel solutions (e.g., Akka/Pekko)
- More mature cloud infrastructure

Higher demand on collaboration

- Expect tools similar to Google Docs, Overleaf, Draw.io, etc.

Our solution: Apache Texera (Incubating) system



Started in 2016

GUI workflows

Real-time collaboration

Supporting Python, R, Java, Scala, ...

Sharing data, workflows, and computing units

Cloud friendly

The screenshot displays the Apache Texera web interface. On the left, a sidebar menu is visible with sections for 'Community', 'Your Work', and 'Admin'. The 'Your Work' section is expanded, showing 'Projects', 'Workflows', and 'Datasets'. A red box highlights the 'Machine Learning' category under 'Workflows', which includes 'Sklearn', 'Passive Aggressive', 'Multinomial Naive Bayes', and 'Extra Trees'. A red arrow points from the 'Public and private resources' text to the 'Workflows' section. The main area shows a workflow titled 'Jail Data Analysis' with steps: 'Coded Transcripts', 'Projection', 'Remove stopwords', 'Training/Testing Split', 'SVM Model', 'LR Model', 'Hash Join', 'Projection', and 'Testing Model'. A red box highlights the 'Share' button in the top right. Below the workflow, a 'Result Panel: Testing Model' shows a table with accuracy and F1 scores. A red box highlights the 'Jiadong Bai' user profile. Text overlays include 'Easy access to AI/ML' and 'Real-time collaboration between researchers'.

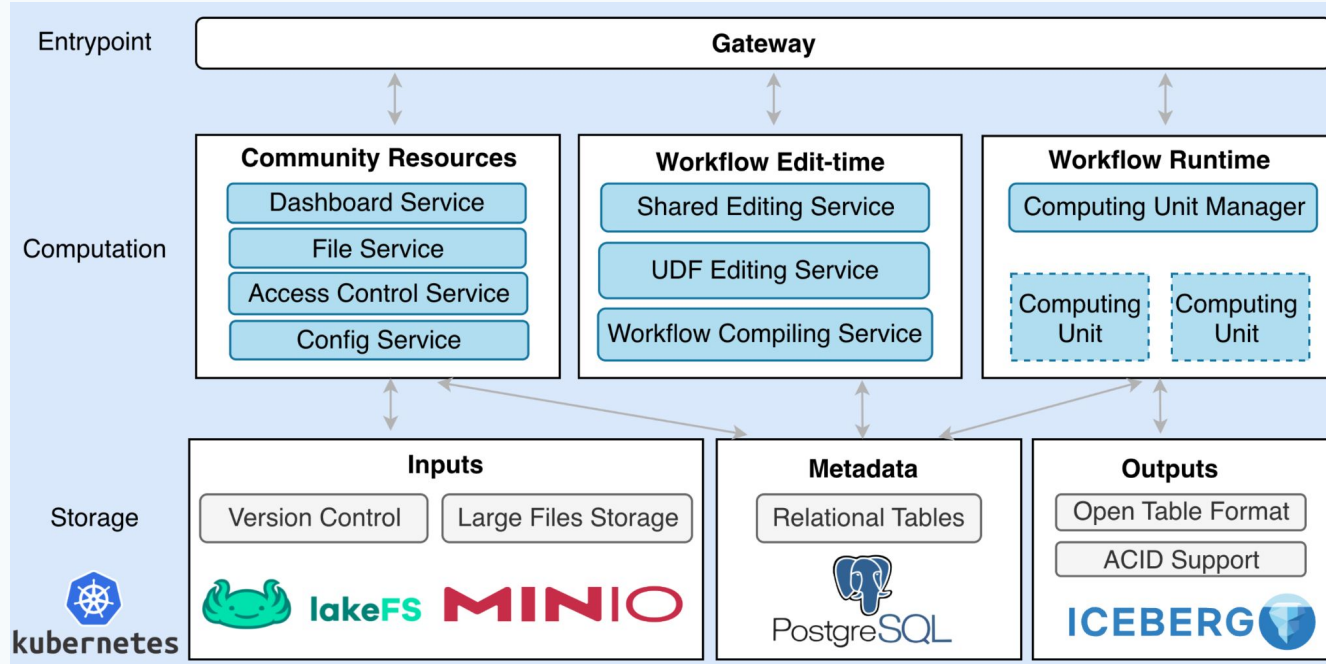
Public and private resources

Easy access to AI/ML

Real-time collaboration between researchers

name		Accuracy		F1		Precisi
IsUrgent	100.00%	Min	0.95	Min	0.86	Min
	0.00%	Average	0.95	Average	0.86	Average
Other	0.00%	Max	0.95	Max	0.86	Max
IsUrgent		0.9523809523809523		0.8571428571428571		0.75

Texera system architecture



Micro-services

Extensible to
new applications

Deployable on
public clouds
and on-prem

Parallel engine
for big data

Adoptions



> 800

users

> 4,000

workflows

> 300,000

workflow versions

> 62,000

executions

20 interdisciplinary projects

Public health, neuroscience,
criminology, and more.

Educational programs

Used in 10 courses

Active open-source community



16

PPMC & committers

226

GitHub contributors

6,856

pull requests

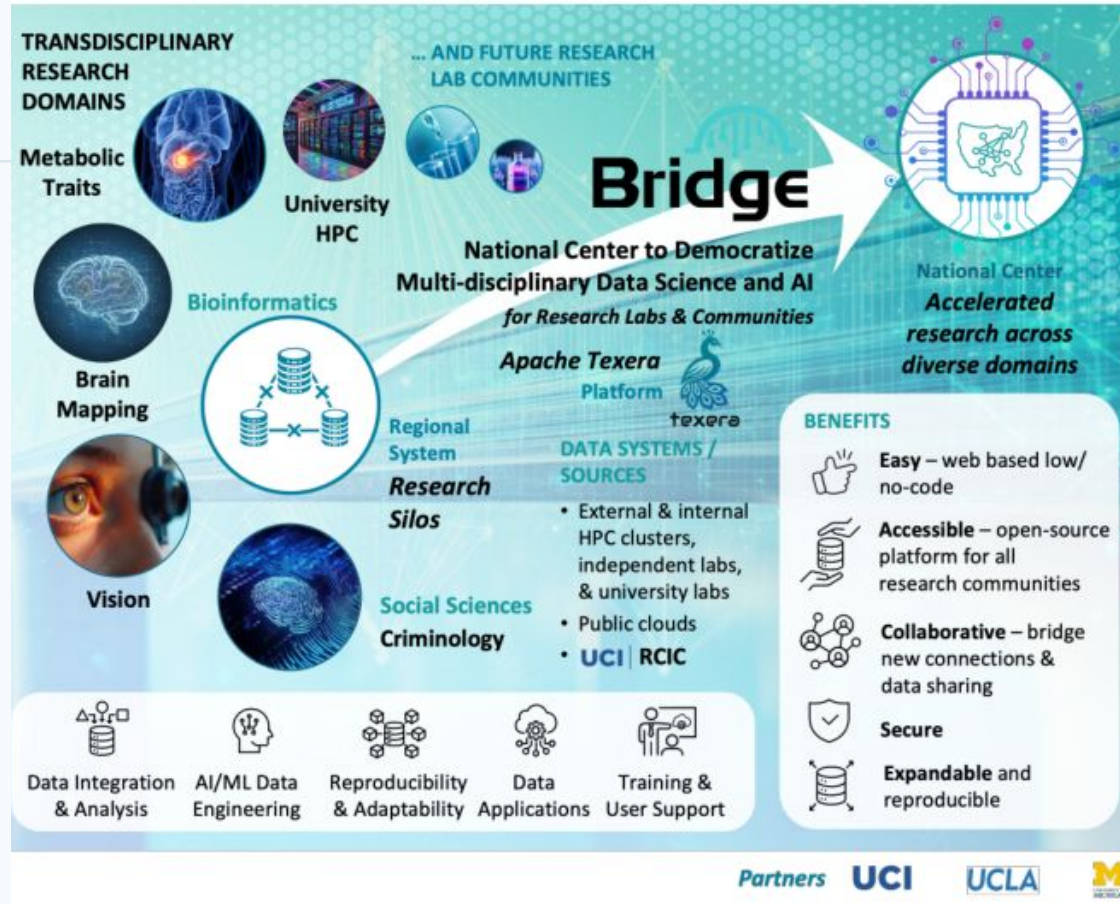
The screenshot shows the Apache Texera GitHub repository page. The repository is public and has 19 branches and 13 tags. The main branch is selected. The repository has 8,197 commits, 263 stars, 170 forks, and 24 watchers. The repository is described as "Human-AI Collaborative Data Science Using Visual Workflows". The repository is licensed under the Apache-2.0 license. The repository has a code of conduct, contributing guidelines, security policy, and activity feed. The repository has 13 releases, with the latest being 1.1.0-incubating, released 2 months ago.

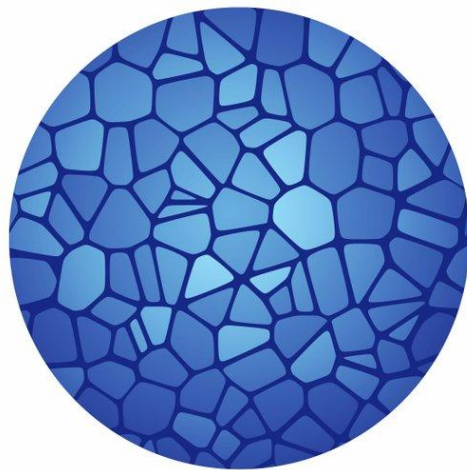
File	Commit Message	Time Ago
.github	revert: route all Codecov flags through the deferred workf...	18 hours ago
.run	feat(build): bump Scala/Java services to Java 17 LTS (#49...	2 months ago
access-control-service	chore(licensing): refresh LICENSE-binary for transitive drif...	last week
agent-service	fix(deps, agent-service): bump the agent-service-patch gr...	3 days ago
amber	test(workflow-execution-service): cover ResultExportServi...	2 days ago
bin	fix(k8s): add missing AUTH_JWT_SECRET to values-devel...	4 days ago
common	test(workflow-core): add storage unit test coverage (Docu...	10 hours ago
computing-unit-managing-service	test(computing-unit): cover ComputingUnitAccessResour...	3 days ago
config-service	feat(config-service): move site_settings API from texera-w...	last week
docs	chore(ci): drop Python 3.10 support, raise floor to 3.11 (#6...	2 weeks ago
file-service	test(file-service): cover DatasetResource path validation (...)	3 days ago
frontend	test(frontend): cover WorkflowUtilService untested helper...	8 hours ago
licenses-3rd-party-code	docs(license): bundle full third-party MIT license texts (#4...	2 months ago
licenses	chore: add LICENSE-binary, NOTICE-binary and dependen...	3 months ago
notebook-migration-service	chore(licensing): refresh LICENSE-binary for transitive drif...	last week

BRIDGE: a national center for scientific communities

Help labs and communities deploy their own Texera-based computing platforms.

Start from 5 domains and support more labs and communities





HUMAN CELL ATLAS

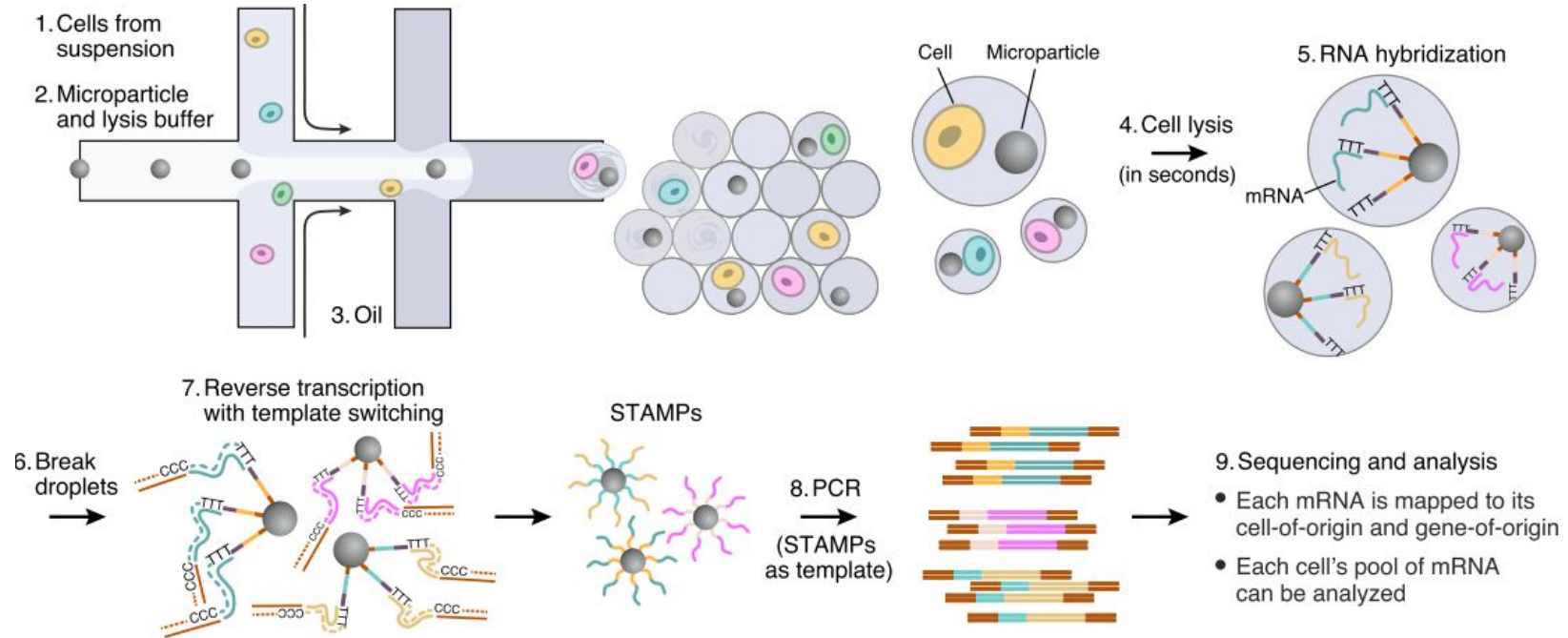
To create comprehensive reference maps of all human cells—the fundamental units of life—as a basis for both understanding human health and diagnosing, monitoring, and treating disease.

4100 members

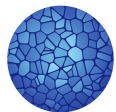
1987 institutions

103 countries

Droplet Single Cell RNA-Seq (Drop-Seq)



Macosko et al, Cell 2015



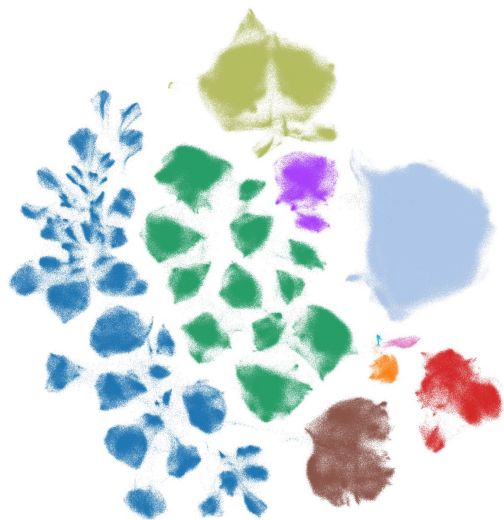
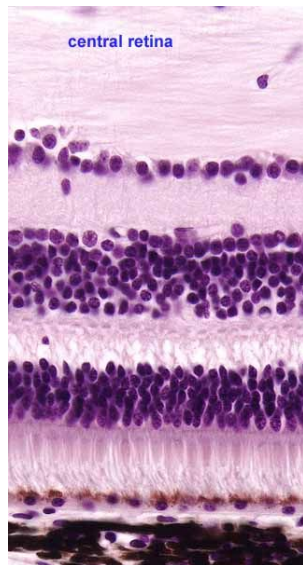
HUMAN
CELL
ATLAS

nature genetics

Article

<https://doi.org/10.1038/s41588-025-02454-1>

Single-cell atlas of the transcriptome and chromatin accessibility in the human retina

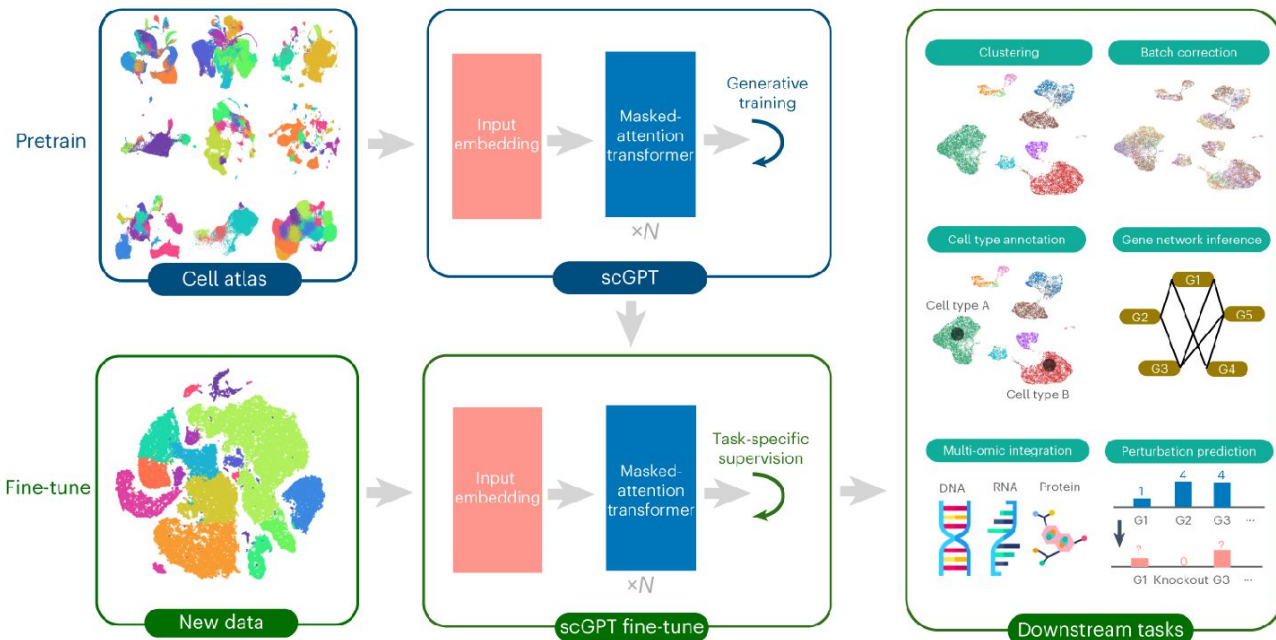


- AC
- Astrocyte
- BC
- Cone
- HC
- MG
- Microglia
- RGC
- RPE
- Rod

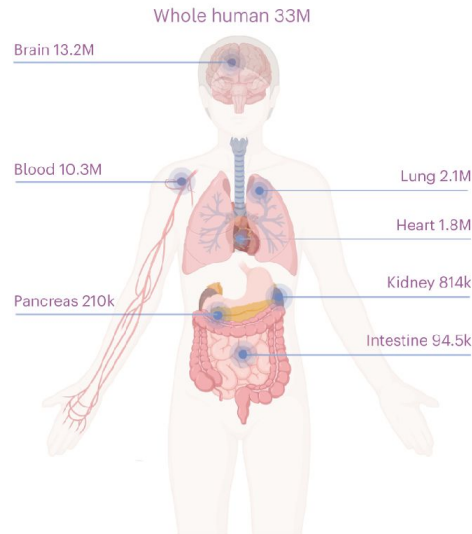
*3.9M total number of single cell/nuclei from 125 donors;
Over 130 cell types identified.*

Single-cell GPT (scGPT)

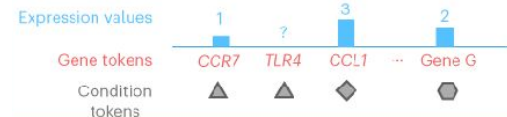
- Pre-training: over 33 million normal human cells
- Fine-tuning: for downstream tasks



Cell numbers and origin tissues included in the pretraining



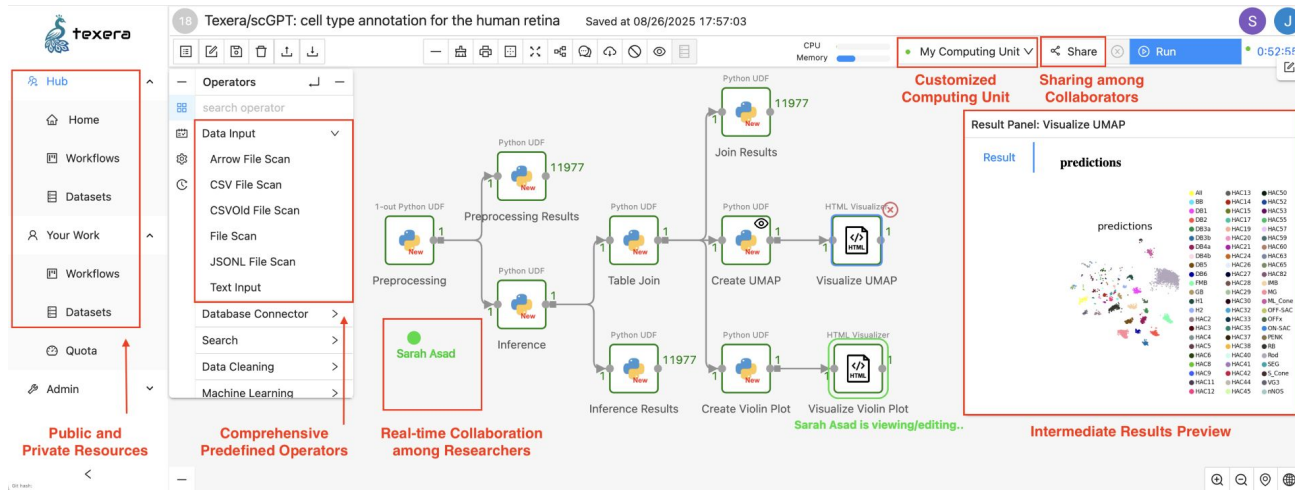
51 organs/tissues



Bridge the Gap of Data Analysis



**Wet Lab
Researcher**



- Further establish additional workflows based on community needs.
- Make the server accessible by the broad community.

Services provided by the BRIDGE Center



Deployment

Deploy Texera for labs and communities on private clusters or public clouds.



Documentation

Maintain guides, tutorials, examples, and reusable workflow templates.



Training

Run onboarding, workshops, office hours, and domain-specific demos.



Support

Full-time staff handle bug reports, tickets, releases, and community needs.



Outreach

Recruit users, collect feedback, grow platforms, and build sustainable communities.



Research

Improve usability, performance, security, AI assistance, and cloud elasticity.

New research opportunities

**Workflow-Script
Migration**



**Agent-based Data
Science**



**Federated Sharing
and Learning**



Flexible Billing



**One Codebase for
Multiple Domains**



Uniqueness and Strengths



No-code and accessible

- No-code experience using GUI-based workflows
- Lowers the barrier to data science and AI for non-coders and domain experts
- AI agents assist with natural language



Flexible deployment

- Deployable across domains and labs
- Runs locally or scales to the cloud
- Meets data-privacy needs
- Elastic cloud compute for heavy workloads

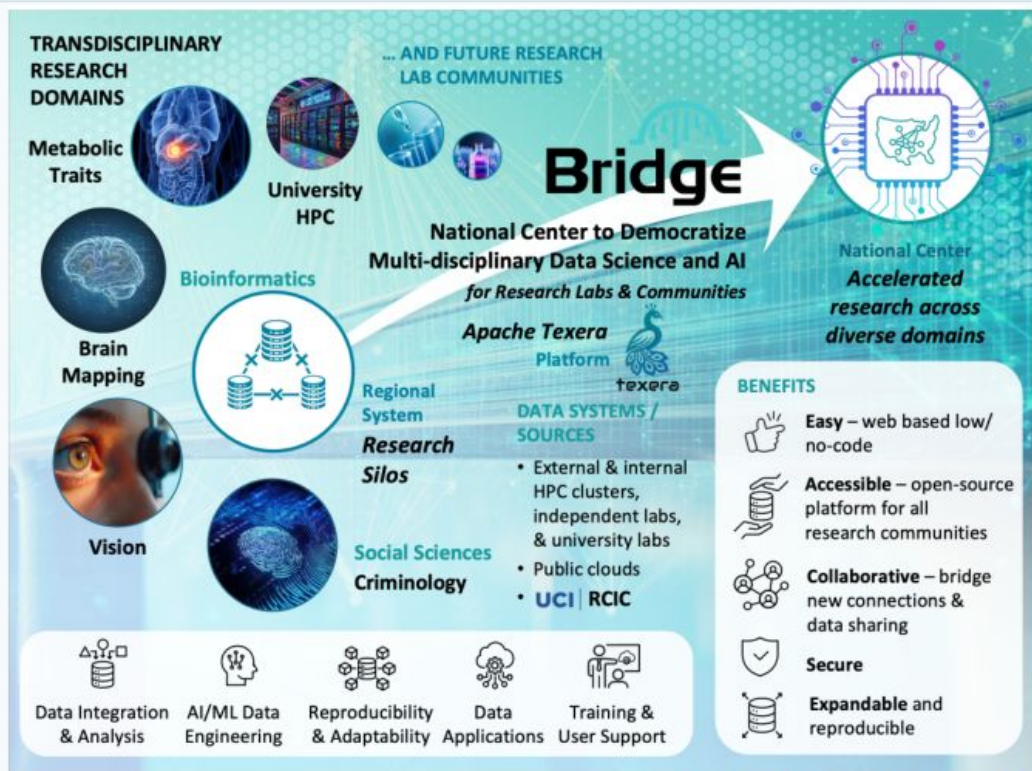


Built for collaboration

- Shared editing, execution, and debugging in real time
- Built-in version control
- Shared resources
 - Datasets
 - Workflows
 - ML models
 - Compute units

BRIDGE

National Center to Democratize Multi-Disciplinary Data Science and AI



BRIDGE

National Center to Democratize Multi-Disciplinary Data Science and AI



Prof. Chen Li
UC Irvine

Rui Chen, Steve Parker, Ardalan Amiri Sani, Naomi Sugie
Kristin Turney, Wei Wang, Xiangmin Xu