

Cost Recovery Model for AI and Machine Learning Support

Consultation

Initial consultation is provided at no charge to discuss study design, data requirements, feasibility, what a given data type can and cannot show, and potential collaboration opportunities.

Grant Proposal Development

Support is available for grant proposal development, including study design, analysis plans, data management plans, and budget preparation.

Investigators are expected to include appropriate percent effort for computational analysis in grant budgets when such support will be required during the project period. Where large datasets or model training are involved, compute and storage should appear as their own budget lines.

Funded Research Projects

For funded projects, computational support should be provided through percent effort on the grant whenever feasible.

Support may include:

- Study design, and review of data collection plans before samples are processed
- Data quality control, and batch effect assessment
- Pipeline development: extraction of features from images, signals, text or records
- Representation learning, clustering and dimensionality reduction
- Supervised model development, training and validation
- Statistical analysis and interpretation
- Manuscript and abstract preparation
- Data management, and delivery of reproducible analysis code

What We Work On

The methods are shared across data types, so the group is organized by what the work is rather than by which instrument produced the data.

- Imaging: digital pathology and whole-slide images, radiology, microscopy and fluorescence, and other volumetric or tiled image data

- Signals and time series: physiological recordings, monitoring data, wearables
- Text and structured records: clinical notes, registries, electronic health records
- Molecular data where the question needs a model rather than a standard workflow
- Multimodal work combining any of the above with each other or with outcome data

Estimating a New Project

What drives cost is not which of those a project belongs to. It is whether a pipeline for that data already exists. The first project on a given data type and acquisition setup carries the cost of building and validating one; later projects on comparable data reuse that work. On a first project, build and validation typically account for about half the effort, so a later project on comparable data costs well under half the original.

Engagement	Example	Indicative effort
Established pipeline, new data	Same data type, acquisition protocol and instrument as a previous project	0.2 to 0.3 person-months
Established pipeline, adapted	Same data type, but a changed protocol, instrument or population	0.3 to 0.5 person-months
New modality	No pipeline exists for this data type yet	0.5 to 1.0 person-months
Unsupervised analysis	Clustering on learned or measured features, embeddings, dimensionality reduction, outlier detection	0.2 to 0.4 person-months
Supervised model development	Annotation, training, benchmarking and validation	Quoted per project
Ongoing scientific involvement	Repeated rounds, interpretation, figures, manuscripts	Percent effort

Those figures are for the analysis itself and assume the data are usable. An estimate should say which row it sits in, because the difference between the first and the third is two to five times.

What the Data Has to Carry

Some conditions determine whether a result can be interpreted at all and cannot be recovered afterwards: control material processed identically to the experimental material, randomization of samples across processing batches and acquisition runs, and per-sample outcome data. Where supervised models are planned, expert annotation on a subset is also required, with an estimate of agreement between annotators.

Unsupervised work does not escape that requirement so much as move it. No labels are needed to produce clusters, but someone has to say what the clusters are, and a cluster nobody can name is not a result. Plan for expert time at the end of an unsupervised project in the way a supervised one plans for it at the start.

These are best settled at consultation.

Unfunded, Pilot, and Small Projects

For projects without external funding, support may be provided based on project scope, available resources, and team capacity.

Internal rates:

- Faculty / Scientific Director: \$150/hour
- Senior AI or computational scientist: \$110/hour
- AI or computational scientist: \$85/hour

These rates include GPU time, working storage and routine data transfer for the duration of the project. Training and inference on large data consume enough machine time that an open compute line is hard for an investigator to budget against, so it is folded into the rate and the estimate is a single number. Long-term storage beyond the project period, commercial software licenses and dedicated computational environments are quoted separately.

Projects requiring substantial effort may be encouraged to identify departmental, divisional, institutional, or external funding mechanisms to support ongoing collaboration.

Trainee Projects

Research projects led by fellows, residents, students, and other trainees will be evaluated based on educational value, project scope, and available team resources. Reduced costs or subsidized support may be available at the discretion of the group and subject to resource availability.