Research proposal aims page

De-identified format sample · Specific aims

THE GAP

Scientists who study complex materials read three kinds of evidence together: images, spectra, and the written record of how a sample was made. Current AI systems read each kind separately. A model that describes a micrograph accurately may still contradict the spectrum taken from the same sample, and it cannot say which source it trusted.

OBJECTIVE

We will build and evaluate a multi-modal reasoning system that answers questions about a sample by drawing on images, spectra, and text at once, and that reports the evidence behind each answer.

CENTRAL HYPOTHESIS

A model trained to reconcile disagreeing sources will make fewer unsupported claims than a model trained on each source alone, without losing accuracy on single-source questions.

AIM 1. ASSEMBLE A RECONCILED BENCHMARK

Collect 2,000 samples that each carry an image, a spectrum, and a preparation record, with expert annotations marking where the sources agree and where they conflict. **Outcome:** a public benchmark and scoring method.

AIM 2. TRAIN A MODEL THAT CITES ITS EVIDENCE

Extend an open multi-modal model so that each statement in an answer points to the region of the image, the band of the spectrum, or the sentence of the record that supports it. **Outcome:** a model and an evidence-tracing interface.

AIM 3. TEST THE MODEL WITH WORKING SCIENTISTS

Run a blinded study in which twelve researchers answer questions with and without the system. Measure accuracy, time to answer, and how often researchers catch a planted error. **Outcome:** evidence on whether cited answers improve human judgment.

EXPECTED IMPACT

If the hypothesis holds, laboratories gain an assistant whose answers can be checked in seconds. If it fails, the benchmark still gives the field its first shared measure of cross-source consistency.