

DOC	JD-HMT-B02
TYPE	Internship position description
TRACK	AI & Digital Health
LEVEL	Student / early career

Data Science Intern, Dataset Curation & Model Evaluation

Dataset survey and curation, annotation schema design, evaluation framework construction, and bias and reliability analysis for health-related AI research

KEY TO TERMS USED

ANNOTATION	Human-assigned labels on data. The definition matters more than the tool.
AGREEMENT	How often two independent annotators give the same label to the same item.
SPLIT	Keeping training and test data separate so that results are not self-congratulatory.
SUBGROUP	Performance reported per group, rather than one average that hides a failure.
LICENSE	The terms under which a dataset may legally be used, including for commercial work.
THRESHOLD	The success bar set before results exist, and not moved afterwards.

INTENDED ROLE

Before a health AI model can be trusted, someone has to define what the data means and what counts as good enough. This role does that upstream work. You will survey available datasets and their licenses, design annotation schemas, and build the evaluation framework the project will be judged against. The emphasis is on defining measurement honestly before results exist, not on chasing a headline accuracy number.

PRINCIPAL DUTIES

- Survey publicly available image and health datasets, documenting content, licensing terms, and stated limitations.
- Assess dataset representativeness across demographics, capture conditions, and label quality, and state gaps plainly.
- Design annotation schemas and written labeling guidelines that two independent annotators could apply consistently.
- Measure annotation agreement, and revise guidelines where disagreement shows the definition is ambiguous.
- Define the evaluation framework, including metrics, splits, baselines, and success thresholds, before any comparison.
- Run subgroup and bias analyses so that reported performance is not an average hiding an unacceptable failure.
- Use approved AI tools for exploration and drafting, then verify all dataset claims against the original documentation.

EXAMPLE ASSIGNMENTS

Assignments depend on active studies, authorized data access, and the internship timeline. Examples may include:

- A dataset landscape review with licensing, representativeness, and suitability recommendations
- An annotation schema and labeling guideline with an inter-annotator agreement study
- An evaluation framework defining metrics, baselines, splits, and subgroup reporting requirements

YOUR FIRST TWO WEEKS

- Survey candidate datasets and record content, license terms, and stated limitations.
- Draft the first annotation guideline and label a pilot set yourself.
- Run a two-annotator pilot and measure how often the labels agree.
- Propose the metrics, splits, and success thresholds before any model is compared.

ABOUT HOPKINS MEDTECH

Hopkins MedTech Group LLC is a U.S. company working in medical technology and in vitro diagnostics. Its work covers clinical performance studies, protocol development and site coordination, data review and quality assurance, and regulatory support for United States and European pathways. The group is also exploring AI and digital health through an early-stage research initiative in facial and skin analysis.

REQUIRED QUALIFICATIONS

- Currently pursuing, or recently completed, a degree in data science, statistics, computer science, public health, or a related field, or able to show equivalent self-study or project work.
- Working Python or R skills for data handling, analysis, and visualization from coursework or personal projects.
- Understanding of evaluation fundamentals such as train and test separation, baselines, and the difference between accuracy and usefulness.
- The rigor to define success criteria before seeing results, and to leave them unchanged afterward.
- Careful reading of licensing and consent terms, and the discipline to respect authorized data boundaries.
- No prior data science, research lab, or professional internship experience is required.

PREFERRED QUALIFICATIONS

Not screening criteria. Coursework, labs, personal projects, and self-study all count as evidence.

- Coursework or self-study in machine learning, experimental design, epidemiology, or measurement theory
- Experience labeling or cleaning a dataset and discovering how ambiguous a definition can be
- Awareness of algorithmic fairness literature, especially in health and dermatology contexts

WHAT YOU WILL GAIN

- Learn the part of AI work that decides whether any later result means anything.
- Practice writing definitions precise enough that two people reach the same label.
- Understand how subgroup performance and dataset bias become clinical and regulatory concerns.
- Leave with an evaluation framework and agreement study that demonstrate real methodological maturity.

DATA HANDLING, PRIVACY, AND RESPONSIBLE AI

Interns work only with data and documents they have been explicitly authorized to access. Patient-identifiable information, study data, client-confidential files, and proprietary materials must never be uploaded to public or unapproved AI services, shared outside the project team, or stored on personal accounts or devices. AI tools may be used only

DELIVERABLES

- A dataset review with licensing and representativeness findings
- An annotation guideline plus an inter-annotator agreement report
- A written evaluation framework with metrics, baselines, and subgroup requirements

SUPERVISION AND EVALUATION

Interns report to a named project lead, meet on a set weekly cadence, and receive written feedback on each deliverable. Work is evaluated on accuracy and traceability of sources, documentation quality, responsiveness to feedback, judgment about what to escalate, and disciplined use of approved AI tools. We expect steady learning and iteration rather than polished perfection on the first day.

WHAT WE DO NOT EXPECT

- Prior internship experience, or a name-brand employer on your CV.
- Knowing the regulations, tools, or vocabulary before you start; the terms box above is where most people begin.
- Certainty. Saying what the evidence does not support is treated as a strength, not a gap.

APPLICATION AND SELECTION

An application is a CV plus a short note naming the track you want and one project from this document you would like to work on, and why. First-round interviews are with the project lead who would supervise the work, not with a screening panel.

where the project lead has approved them for that task, and all AI-assisted work must be independently verified before it enters any deliverable. The intern and the project lead remain accountable for the final result.

Hopkins MedTech Group LLC provides equal internship and employment opportunities without regard to race, color, religion, sex, sexual orientation, gender identity, national origin, age, disability, veteran status, or any other legally protected characteristic. Internships are educational in nature; interns do not provide medical advice, make regulatory determinations, or represent the company to any health authority, client, or study site. The AI and digital health initiative described here is early-stage research, not a cleared, approved, or clinically validated product. Assignments, data access, duration, and supervision are confirmed in the final project description.