

TIRTHO ROY

☎ +1 515-509-9498

🌐 Tirtho-Roy

🔗 Tirtho-Roy

📧 Tirtho-Roy

🔄 tirtho149

✉ tirtho@iastate.edu

SUMMARY

Software and machine-learning engineer with a research-driven approach (B.S. Data Science and M.S. Mechanical Engineering, Iowa State University). Builds end-to-end systems—from TypeScript/React web applications and data-processing pipelines to large-scale machine learning, computer vision, and vision-language models—and is committed to turning research prototypes into reliable, well-tested, production-minded software. Co-first-authored work accepted at an IEEE/CVF Conference on Computer Vision and Pattern Recognition 2026 workshop.

TECHNICAL SKILLS

Languages: Python, TypeScript, JavaScript, SQL

Web & Full-Stack: React, web application development, REST APIs, responsive UI/UX

Machine Learning & AI: machine learning, deep learning (PyTorch), computer vision, vision-language models, large language models, multi-agent systems

Data & Infrastructure: data processing & pipelines, large-scale datasets, databases/SQL, cloud (AWS/GCP), Git, developer tooling

EDUCATION

Bachelor of Science in Data Science

Aug 2023 – Present

Iowa State University, Ames, IA

Master's of Science in Mechanical Engineering

Jan 2025 – Present

Iowa State University, Ames, IA

EXPERIENCE

Graduate Research Assistant

Jan 2025 – Present

Translational AI Center, Iowa State University

Ames, IA

- Lead M.S. thesis research on **agentic vision-language systems for cyber-agricultural applications** under Prof. Soumik Sarkar; co-first-authored work accepted at an **IEEE/CVF Conference on Computer Vision and Pattern Recognition 2026** workshop.
- Co-engineered **SAGE**, an autonomous crop-disease-diagnosis agent powered by the largest plant-disease image-symptom dataset to date (335 crops, 1,251 classes, ~839K images); symptom grounding raised accuracy by **+16.2** points and generalizes to unseen crops with *no retraining*.
- Build large-scale LLM/VLM evaluation pipelines and reproducible benchmarks spanning agriculture, healthcare, and finance, turning research prototypes into deployable, rigorously tested systems.
- **Skills:** Python, PyTorch, multi-agent LLM/VLM orchestration, large-scale data engineering, retrieval & grounding, benchmark and metric design.

Undergraduate Research Assistant

Apr 2024 – Present

Translational AI Center, Iowa State University

Ames, IA

- Built a GPT-4o pipeline analyzing 430 Shark Tank pitches to quantify how market orientation and brand storytelling drive investor decisions; findings published in the *Journal of Business & Industrial Marketing*.
- **Skills:** Python, LLM prompting, human-AI annotation, statistical validation.

Undergraduate Software Developer

June 2024 – December 2024

Dept. of Agricultural & Biosystems Engineering, Iowa State University

Ames, IA

- Developed a chatbot web frontend for an agricultural research tool with Prof. Ryan McGhee, improving usability through iterative, user-tested design.
- **Skills:** web frontend development, responsive UI/UX design, user-centered design.

- Ran controlled probing experiments on how large language models resolve knowledge conflicts (conflicting context vs. parametric memory) under Prof. Qi Li.
- **Skills:** experimental design, LLM evaluation, data analysis.

PROJECTS

AgAdvisor — Agentic RAG for Pesticide-Label Guidance

[🔗 Live Demo](#)

Translational AI Center / AIIRA, Iowa State University

(Iowa Soybean Association beta)

- Built and deployed a multi-agent RAG assistant that answers pesticide-label questions grounded in **CDMS** labels, live as a public demo on Hugging Face Spaces; stack: FastAPI + LangGraph multi-agent orchestration, OpenAI embeddings + Qdrant vector search, GPT-4o, and a Streamlit/Gradio UI.
- **Industry connection:** developed in the Translational AI Center (AIIRA) and **beta-tested with the Iowa Soybean Association**, iterating directly on grower and agronomist feedback (wrong-herbicide bias, answer verbosity, mobile confidence badge, cold-start latency).
- Engineered **faithful, abstaining retrieval**—a product-disambiguating catalog plus abstention so the agent refuses rather than answer from the wrong herbicide label, with page-level citations: a concrete reliability guarantee for regulatory agrochemical guidance.
- Productionized with account authentication and per-user persistent history, per-user daily quotas to protect shared API keys, response caching, and a concurrency-safe shared vector store; a 62-test suite runs in CI.

PUBLICATIONS & PRESENTATIONS

SAGE: Scalable Agentic Grounded Evaluation for Crop Disease Diagnosis

IEEE/CVF Conference on Computer Vision and Pattern Recognition 2026 (CORE A*), Workshop on Emerging Directions in Data for Multimodal Foundation Models, Accepted (*Co-first author*)

- Co-built the largest plant-disease image-symptom dataset to date: 335 crops, 1,251 disease classes, and ~839K images with source-grounded, expert-validated symptom descriptions.
- Engineered an autonomous vision-language agent with stepwise diagnostic reasoning (anatomical context → symptom-guided narrowing → reference-image comparison → explainable trace); symptom grounding raised accuracy by +16.2 points on average and extends to new crops without retraining.
- Core component of M.S. thesis research.

SAGE: Scalable Agentic Grounded Evaluation for Crop Disease Diagnosis

AI Scientist Summer Workshop, Microsoft Research New England, Cambridge, MA, Poster (*Selected for presentation; co-first author*)

- Selected for poster presentation at the AI Scientist Summer Workshop hosted by Microsoft Research New England (August 4, 2026), presenting the SAGE crop-disease-diagnosis agent to the AI-for-science community.

Counterfactual Dynamic Embodiment for Field Deployable Agricultural Robots

IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) 2026 (CORE A), Pittsburgh, PA — Workshop on From Innovation to Adoption: Hardware, AI, Policy, and Funding Challenges in Agricultural Robotics, Oct. 2026 (*Accepted; poster*)

- Proposed AgriMorph, which treats a field robot's morphology as a runtime decision rather than a fixed hardware property, using a learned world model to project candidate futures under each available body before terrain is entered.
- Designed a dual latent encoding that keeps terrain perception separate from internal kinematic state, a counterfactual projection over five candidate morphologies that nets projected utility against transformation cost, and an ensemble-disagreement confidence gate that blocks shape changes made on weak sensing.

- Invited to submit an extended version to the Springer Special Issue on Agricultural Robotics.

bioMoR: Biology-Guided Mixture-of-Recursions for Effective Genomic Learning

Manuscript under review, 2026

(Under review; second author)

- Co-developed the first Mixture-of-Recursions framework for gene- and pathway-level omics learning, compressing high-dimensional expression inputs into marker-gene or Reactome pathway tokens and routing each token through an adaptive number of recursive steps in one weight-shared Transformer block.
- Injected biological structure at three sites — graph-based embedding smoothing, a structural attention bias over co-expression and pathway–pathway relations, and a zero-initialized graph-aware router that sets per-token recursion depth.
- Across eight single-cell and multi-omics benchmarks under a unified five-fold protocol, improved macro-F1 by 8.2 points and balanced accuracy by 7.1 points over the strongest biology-agnostic MoR baseline while using 75% fewer parameters and up to 58% fewer FLOPs than a non-recursive Transformer.

Towards Trustworthy Biological Alignment in TabPFN-Probed Pathology Foundation Models

Manuscript in preparation, 2026

(In preparation; fourth author)

- Built a training-free framework that audits whether frozen pathology foundation models encode biologically meaningful signal, pairing histology with spatial transcriptomics (HEST-1k; 240 samples across breast, skin, and brain) and using TabPFN as a standardized probe against Ridge, kNN, and MLP baselines.
- Benchmarked five encoders (UNI, Phikon, Phikon-v2, Lunit DINO ViT, CTransPath-style SimCLR ResNet); UNI led pathway decodability at 0.303, and pathway-level activity recovered more reliably than individual genes.
- Quantified generalization failure modes: organ-level shift cost 62% of the pathway signal and section identity remained linearly decodable up to 214× chance, while benign augmentations (rotation, flip) retained $\geq 95\%$ performance.

When Do Engineered Features Beat Raw Prices? A Walk-Forward Benchmark of Classical, Neural, and World-Model Predictors on a Global Equity Dataset

World Models @ Booth (WM@Booth) 2026, University of Chicago Booth School of Business, Aug.–Sep. 2026
(Accepted; poster)

- Built a walk-forward backtesting framework benchmarking classical, neural, and world-model predictors on a global equity dataset.
- Isolated the conditions under which engineered features outperform raw price inputs across model families.

TRAPS: Therapeutic Response Analysis via Pathway-informed Stratification

arXiv preprint, 2026

- Built the first unified benchmark for pathway-guided cancer therapy-response modeling, evaluating three biologically informed architectures (BINN, GraphPath, PATH) under identical training and evaluation conditions.
- Engineered a multi-task pipeline over 2,622 TCGA patients across 5 cohorts (Reactome pathway-activity features), jointly predicting targeted therapy, radiation therapy, and 6-month survival; best model reached 0.92 AUROC on prostate targeted-therapy prediction at 11% class prevalence.

Leveraging Temporal Distribution Shifts for Rice Disease Detection

3rd IEEE International Conference on Computing, Applications and Systems 2026, IEEE Xplore, Oct. 2026
(Accepted; corresponding author)

- Designed a chronological evaluation protocol for rice-leaf disease recognition on Bangladeshi field imagery: ten public sources merged into a 12,683-image four-class corpus, trained on 2021–2025 and tested only on a held-out 2026 partition that never entered training, tuning, or model selection.

- Compared four backbones (ResNet50, DenseNet121, EfficientNet-B0, DINOv2 ViT-S/14) under full and partial fine-tuning across 10 seeds (80 runs), with claims fixed before testing and assessed by seed-matched paired tests, assumption checks, Holm correction, effect sizes, and confidence intervals.
- Future-year macro-F1 fell 43.91 points below same-period validation in all eight configurations (95% CI: 39.42–48.39; retention 48.28%), and the better fine-tuning strategy proved architecture-dependent (Friedman $\chi^2 = 28.08$, $p = 3.49 \times 10^{-6}$) — DINOv2 with partial fine-tuning generalized best despite ranking only fifth of eight on validation.

Improving the Safety and Trustworthiness of Medical AI via Multi-Agent Evaluation Loops

arXiv preprint, 2026

- Built a multi-agent iterative-refinement system pairing generator models (DeepSeek R1, Med-PaLM) with evaluator agents (LLaMA 3.1, Phi-4) to align medical LLM outputs to AMA ethics principles and a five-tier Safety Risk Assessment (SRA-5).
- Cut ethical violations by 89% and achieved a 92% risk-downgrade rate across 900 clinical queries spanning nine ethical domains.

An Agentic Framework for Mechanistic Therapeutic Reasoning in Oncology via Pathway-Grounded Tree-of-Thought Inference

North East AI Agents Day 2026, Accepted; ACM Conference on Bioinformatics, Computational Biology, and Health Informatics 2026 (CORE B), Posters

- Designed an agentic, pathway-grounded tree-of-thought inference pipeline that structures LLM reasoning over biological pathway knowledge for mechanistic therapeutic reasoning in oncology.

Intelligent Health Intervention System for Syndemic Management Among University Students: An AI-Driven Approach Using Reinforcement Learning

Association for the Advancement of Artificial Intelligence 2026 Spring Symposium Series, San Francisco, CA, Oral Presentation

- Co-developed a reinforcement-learning intervention engine for syndemic health management, modeling intervention timing and selection as a sequential decision problem.

Interpreting Transformers through the Lens of Physics and Dynamical Systems

American Physical Society Global Physics Summit 2026, Presentation

(First author)

- First-authored an analysis interpreting transformer architectures through physics and dynamical-systems theory to explain their internal computation dynamics.

Leveraging Vision Language Models for Specialized Agricultural Tasks (AgEval Benchmark)

Proceedings of the 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) (CORE A)

- Co-developed AgEval, a 12-task plant-stress phenotyping benchmark measuring zero- and few-shot in-context learning of state-of-the-art VLMs (Claude, GPT, Gemini, LLaVA).
- Demonstrated rapid domain adaptation: best-model F1 rose from 46.24% (zero-shot) to 73.37% (8-shot); a coefficient-of-variation metric exposed per-class gaps of 26–58%, and strategic example selection added +15.38% F1.

The Impact of Market Orientation and Brand Storytelling on Shark Tank Evaluations

Journal of Business & Industrial Marketing, 40(1), 265–280, 2025

- Built a GPT-4o analysis pipeline over 430 Shark Tank pitches to quantify how market orientation and brand storytelling drive investor decisions.
- Combined expert annotation with statistical validation to test and calibrate the LLM-derived predictions for a peer-reviewed B2B marketing study.

Enhancing Medical AI Safety Through Multi-Agent Evaluation and Iterative Refinement

The 16th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics 2025 (CORE B),

Pennsylvania, USA, Poster

- Prototyped the multi-agent medical-LLM safety framework (generators DeepSeek R1 / Med-PaLM, evaluators LLaMA 3.1 / Phi-4) aligned to AMA principles and the SRA-5 protocol; later extended into the arXiv study above.

Federated In-Context Prompt Selection for Multi-Modal 3D Dental Imaging: A Theoretical Framework with Privacy-Preserving Guarantees

Oral and Dental Image Analysis Workshop 2025, held with the 29th International Conference on Medical Image Computing and Computer Assisted Intervention 2025 (CORE A), Poster

- Designed a privacy-preserving federated-learning framework for multi-modal 2D/3D dental imaging, featuring cross-modal prompt alignment, hierarchical multi-modal optimization, and Byzantine-resilient aggregation with differential privacy.

Emissai: LLM-Powered Analysis of Carbon Footprint

American Institute of Chemical Engineers Annual Meeting 2025, Presentation

- Built an LLM-powered tool to estimate and analyze carbon footprint from operational data.

Benchmarking Bioemu-1: Establishing a Baseline for Protein Structural Ensemble Prediction

American Institute of Chemical Engineers Annual Meeting 2025, Boston, USA, Poster

- Established baseline benchmarks for Bioemu-1 on protein structural-ensemble prediction, defining reproducible metrics for computational structural biology.

LlamaHealth: Assessing LLaMA’s Performance in Multi-Class Healthcare Test Result Prediction

International Conference on Engineering Research, Innovation and Education 2025, Shahjalal University of Science and Technology, Sylhet, Bangladesh, Poster

- Benchmarked LLaMA on multi-class healthcare test-result prediction, assessing reliability for responsible medical-AI deployment.

LLM-based Insights into How Market Orientation and Brand Storytelling Affect Investors’ Evaluation of Startup Pitches

Research Experiences for Undergraduates Summer Symposium 2025, Poster; Translational AI Center Day 2025, Poster & Flash Talk

- Presented the Shark Tank investor-evaluation study (GPT-4o over Shark Tank data) at the Research Experiences for Undergraduates Summer Symposium and as a Translational AI Center Day 2025 flash talk; mentored a peer researcher within the Translational AI Center.

INTERNSHIPS

Undergraduate Summer Research Intern – Translational AI Center	May 2024 - July 2024
--	----------------------

HONORS & AWARDS

Dean’s High Impact Award for Undergraduate Research	Fall 2025
Dean’s List	Fall 2024
Best-In-Show, George Washington Carver Poster Challenge	2025
Borlaug Poster Competition (1st Place, Undergraduate Division) (<i>certificate</i>)	2025
Bangladesh–Sweden Trust Fund Travel Grant	2025
Iowa State University College of Liberal Arts and Sciences Travel Grant	2025

ACADEMIC SERVICE

Member, [Institute of Electrical and Electronics Engineers \(IEEE\)](#) — Graduate Student Member (through Dec. 2026); [IEEE Computer Society](#), [IEEE Antennas and Propagation Society](#), and [IEEE Young Professionals](#). **Member**, [Association for Computing Machinery \(ACM\)](#) — Student Member (since 2026)

Program Committee, Workshop on Secure and Trustworthy Quantum Machine Learning (SaTQuML) @ Conference on Neural Information Processing Systems 2026 (CORE A*)

Program Committee, [Trustworthy Agentic Systems \(TAS 2026\)](#) @ Association for the Advancement of Artificial Intelligence 2026 Fall Symposium Series, Arlington, VA (Nov. 5–7, 2026)

Ethics Reviewer, Conference on Neural Information Processing Systems 2026 (CORE A*) (Main Conference and Evaluations & Datasets Tracks)

Reviewer, [Medical Reasoning with Vision Language Foundation Models Workshop \(Med-Reasoner\)](#) @ Conference on Neural Information Processing Systems 2026 (CORE A*)

Reviewer, [AI for Drug Discovery: Bridging the Translation Gap Workshop \(AI4DD\)](#) @ Conference on Neural Information Processing Systems 2026 (CORE A*)

Reviewer, [AI for Meta-Science Workshop \(AI4MetaScience\)](#) @ Conference on Neural Information Processing Systems 2026 (CORE A*)

Reviewer, [Diffusion Language Models Workshop \(DiffuLM\)](#) @ Conference on Neural Information Processing Systems 2026 (CORE A*)

Reviewer, [New In ML Workshop \(NewInML\)](#) @ Conference on Neural Information Processing Systems 2026 (CORE A*)

Reviewer, [AI for Science Workshop \(AI4Science\)](#) @ Conference on Neural Information Processing Systems 2026 (CORE A*)

Reviewer, [Economics for Machine Learning Workshop \(EconML\)](#) @ Conference on Neural Information Processing Systems 2026 (CORE A*)

Reviewer, World Models @ Booth 2026

Reviewer, Context Beyond the Window Workshop @ Conference on Language Modeling (COLM) 2026 (top-tier)

Reviewer, Social Simulation Workshop (Social Sim '26) @ Conference on Language Modeling (COLM) 2026 (top-tier)

Ethics Reviewer, Conference on Neural Information Processing Systems 2025 (CORE A*) (Main Conference and Datasets & Benchmarks Tracks)

Reviewer, Rashomon Effect Workshop @ Association for the Advancement of Artificial Intelligence 2026 (CORE A*)

Reviewer, AI for Good Workshop @ International Conference on Machine Learning 2026 (CORE A*)

Reviewer, Computer Vision for Animals Workshop @ IEEE/CVF Conference on Computer Vision and Pattern Recognition 2026 (CORE A*)

Reviewer, AI for Critical National Infrastructure Workshop @ International Conference on Autonomous Agents and Multi-Agent Systems 2026 (CORE A*)

Reviewer, LatinX in AI Workshop @ Conference on Neural Information Processing Systems 2025 (CORE A*)

Reviewer, LatinX in Computer Vision Workshop @ IEEE/CVF Conference on Computer Vision and Pattern Recognition 2025 (CORE A*)

Reviewer, Computer Vision for Animals Workshop @ IEEE/CVF Conference on Computer Vision and Pattern Recognition 2025 (CORE A*)

Reviewer, Second Workshop on Bangla Language Processing @ International Joint Conference on Natural Language Processing and Asia-Pacific Chapter of the Association for Computational Linguistics 2025

Reviewer, Indian Conference on Computer Vision, Graphics & Image Processing ([ICVGIP 2026](#))

Reviewer, 4th International Conference on Computing Advancements (ICCA 2026) ([certificate](#))

Reviewer, 3rd International Conference on Machine Intelligence and Emerging Technologies (MIET 2026) ([certificate](#))