DIKSHA SHRIVASTAVA

dikshashraigarh57@gmail.com · *diksha-shrivastava13.github.io* · *Research & Blog*

RESEARCH INTERESTS

Causal Discovery for Safe ASI · Formal Verification & Scalable Oversight · Natural Abstractions · Continual Learning in Complex World Models

EDUCATION

2021-2025 Bachelor of Technology in Computer Science

Bennett University (The Times Group), India

CGPA: 9.12/10 · **Specialization in Artificial Intelligence**

Coursework · Statistical Machine Learning, Artificial Intelligence, Intelligent Model Design Thinking, Natural Language Processing, Special Topics in AI, Undergraduate Research in CS.

Research Work · Investigated frameworks for continual reasoning in world models. Developed a framework for Divergent Problem Generation with modifications to GRPO. Developed pipelines for the automation of complete ML cycles from a high-level description.

RESEARCH EXPERIENCE

2025-2026 Visiting Researcher

Lossfunk, India

Testing the Scientist AI agenda: How can truth-seeking agents exploit natural abstractions to inherently understand any system & build a causal world model under safety mechanisms?

Residency: 6-months work to train truth-seeking agents to discover the causal structures underlying any environment and build a bayesian causal world model.
Initial Research Proposal

Jun–Jul, 2025 FAEB (ARENA) Scholar

Finnish Center for Safe AI, Tutke

Can the agents develop a causal world model by engaging in a scientific debate?

Training: 6-Week Finnish Alignment Engineering Bootcamp on Technical AI Safety based on the ARENA curriculum, spanning neural network fundamentals, mechanistic interpretability, reinforcement learning, and LLM evaluations.

Capstone: Developed a proof-of-concept for Causal Discovery using Debate protocols as a proxy in the five days duration. Connected debater agents to an external world model, with formal verification layers for the scientific discovery process.
Presentation Deck

Jan–May, 2025 Research Intern, AI Reasoning

School of CSET, Bennett University

How can agents discover unseen dependencies in structured world representations?

Continual Reasoning: Developed a memory-integrated framework for iterative self-correction in complex inference tasks.

Rearrangement Sampling: Proposed a sampling technique that converts divergent solutions into new problem formulations, enhancing generalization across reasoning tasks.

Execution-Guided Generation: Implemented a feedback-driven decoding pipeline where execution traces refine model-generated hypotheses, reducing error propagation.

Automated ML Pipelines: Developed a system that dynamically generates and executes end-to-end ML pipelines from high-level problem specifications using a decision graph.
Technical Blog

Sept–Dec, 2024 Independent Researcher, Causal Discovery

Independent Research, Remote

Can language models formulate ML problems from deep, interacting subsystems?

Observation · LLMs recognize surface correlations but fail to uncover deep causal structures governing the complex, evolving world models.

Reasoning in Holistic World Models: Designed experiments to test agent learning, adaptation, and generalization in dynamically interwoven systems represented by hybrid vector-graphs system.

Beyond Static Models: Stress-tested frameworks for causal discovery and formal verification from abstract data of complex world models, including transduction & induction reasoning methods, symbolic regression, open-endedness and automated theorem-proving.

Continual Learning with Dynamic Database: Designed a self-updating framework for hypothesis-driven link prediction and structured learning in evolving datasets.
Technical Blog

May–Aug, 2022 Student Researcher, AI & Psychophysics

Nvidia-Bennett Center for AI, Bennett University

How do cognitive disorders affect neural music perception?

Neural Pattern Analysis: Applied SPM12 and PRoNTO V3.0 in MATLAB to analyze fMRI data, isolating superior temporal gyrus (STG) activity for genre-based classification.

Machine Learning for Cognition: Designed an SVM-based classifier to distinguish neural responses to music genres, leveraging voxel-based feature extraction.

Conference Acceptance: Selected to present at Fechner Day 2022, Sweden, showcasing ML-driven insights into music cognition and mental health applications. *(Withdrawn for Grant Reasons)*
Abstract | Website

research engineering & product development

Jun–Sept, 2024 AI Engineer, Founding Team

Digital Product School, Munich with German Federal Ministry, BMZ

Can AI reason across complex world policy decisions spanning decades for maximal gain?

Product: Designed and piloted an AI-driven Decision-Making System for policy officers in 60+ countries, modeling hierarchical government initiatives as a 5-level structured world model to support strategic policy decisions.

Pipeline: Developed multi-layered agentic reasoning pipelines (54+ iterations over 200–2000 entities from unstructured reports) to track causal shifts in policy evolution.

Tools: Built 7+ AI tools—situational similarity models, graph-based retrieval, and AI-driven action plans—to surface risk factors and rank interventions by structural importance.

Inference: Explored and benchmarked reasoning methods (agentic workflows, multi-hop reasoning, few-shot planning, Monte Carlo Tree Search, graphrag, etc.) to capture implicit relationships over time.

Handover: Delivered the system to BMZ's DataLab with AI-driven recommendations, strategic planning insights, and roadmaps for SLM training on structured decision-making tasks.
Technical Blog

Feb–May, 2024 AI Engineer, Founding Team

Digital Product School with SAP, Munich

How can an AI system continually learn from feedback to refine information retrieval?

Product: Prototyped ai-SAP, an LLM-powered search and retrieval system for 100,000 SAP employees, unifying access to internal documentation, GitHub, and Slack.

Retrieval & Reasoning: Designed a multi-step retrieval pipeline with 15+ data readers and 13+ LLM calls, integrating filtering, recursion, and intent classification.

Optimization & Efficiency: Integrated a CI/CD pipeline on Google Cloud and enhanced LlamaIndex with custom chunking and extraction strategies, improving recall, MRR, and reducing debugging time from 14+ hours to 5 seconds.

Continuous Learning: Designed a generative feedback loop that dynamically updates answerable question metadata based on user feedback, aligning the knowledge base with evolving user needs.

Investor Presentations: Presented the product to investors at Meta, IBM, UnternehmerTUM, MTZ, AWS, and United Internet Media GmBH, demonstrating the system's capabilities.
Onsite Pitch Video

fellowships & open-source

Jul–Oct, 2023 Core Contributor - ML

Unify.ai (YC W23), London

Built unified backend APIs (TensorFlow, PyTorch, JAX, MindSpore, PaddlePaddle) for cross-framework compatibility. Designed universal loss functions, neural network ops, and convolution layers.

2022-2023 Google KaggleX Fellow

KaggleX Fellowship

How can AI understand and generate emotions in music through symbolic representation? Explored symbolic music generation with Music Transformers, built MIDI/audio models (Librosa, Music21, LSTM), analyzed structural patterns.

2022–2024 Google Women Engineer Scholar

TalentSprint · Google, India

Completed 2-year ML and software training, showcased projects at IIIT Hyderabad bootcamp. Received mentorship in communication, strategic planning, and design thinking.

Research Directions

Causal Discovery for Critical AI Safety

2025 · Causal Discovery in Evolving Curricula for Safe Open-Ended Foundation Models

2025 · Externalizing Latent Reasoning to an Interpretable Dynamic World Model

2025 · The Problem of Perspectives: How Perspective Shapes Causality & Alignment

2025 · Evolving Debate in Agents: Question-Asking as Alignment in Scientific Inquiry

AI for Formal Logic & Decision-Making

2025 · Modeling Wayfinding: A Hybrid Neurosymbolic and RL Approach to Dynamic Decision-Making in Quest-Driven Narrative Worlds

2025 · AI for Astrophysics: Automating Domain Specific Tasks with LLMs

2024 · Automated ML Cycles: From High-Level Description to Research & Analysis for Deployed APIs with Minimal Supervision

Reinforcement Learning

2025 · Decisive-Agents: Leveraging Graph of Decisions for Intermediate Reward Modeling

2025 · Execution-Guided Continuous Code Generation for Adaptive Agent Reasoning

2023 · ChessGAN: Designing Agents to Lose at Chess 50% of the Time & The Turing Test

Publications

Reinforcement Learning

2025 · Reasoning Beyond Correctness: Problem Generation through Divergent Rearrangement Sampling. Diksha Shrivastava, Mann Acharya, Dr. Tapas Badal. *Communicated to AAAI 2026.*

Cognitive Sciences

2022 · Analysis of Neural Correlates of Different Music Genres using Machine Learning. Diksha Shrivastava, Dr. Anuj Bharti. *Accepted to Fechner Day 2022, by International Society for Psychophysics (withdrawn due to grant reasons).*

Miscellaneous

Awards & Recognition

2024 · **UVC Partners' Summer BBQ:** Personally invited to UVC Partners' highly exclusive, invite-only Summer BBQ to network with top VCs, angel investors, and startups.

2024 · **Investor Presentations:** Presented products to investors at Meta, IBM, MTZ, AWS, receiving funding offers for a white-label version.

2022 · **Google KaggleX Grantee:** Selected as one of the Top 152 globally among AI researchers and engineers, awarded a $1,000 research grant and $1,000 in GCP credits.

2022 · **Google TalentSprint WE Scholar:** Chosen as one of the Top 250 from 30,000+ applicants, awarded a 100% Scholarship for training by Google and TalentSprint experts.

Relevant Skills

Technical Skills · Python, Java, C++. TensorFlow, PyTorch, Keras, HuggingFace. LLM, Agentic & Custom Frameworks. Vector, Graph, Hybrid Databases. GCP, Azure, AWS. LLMOps. FastAPI, Redis, Docker, Git.

Product Development · AI Product Management, Low and Hi-Fi Prototyping, Proof of Concepts, Lean Start-Up, Risk Validation & Deployment, Iterative MVP Development.

August 18, 2025