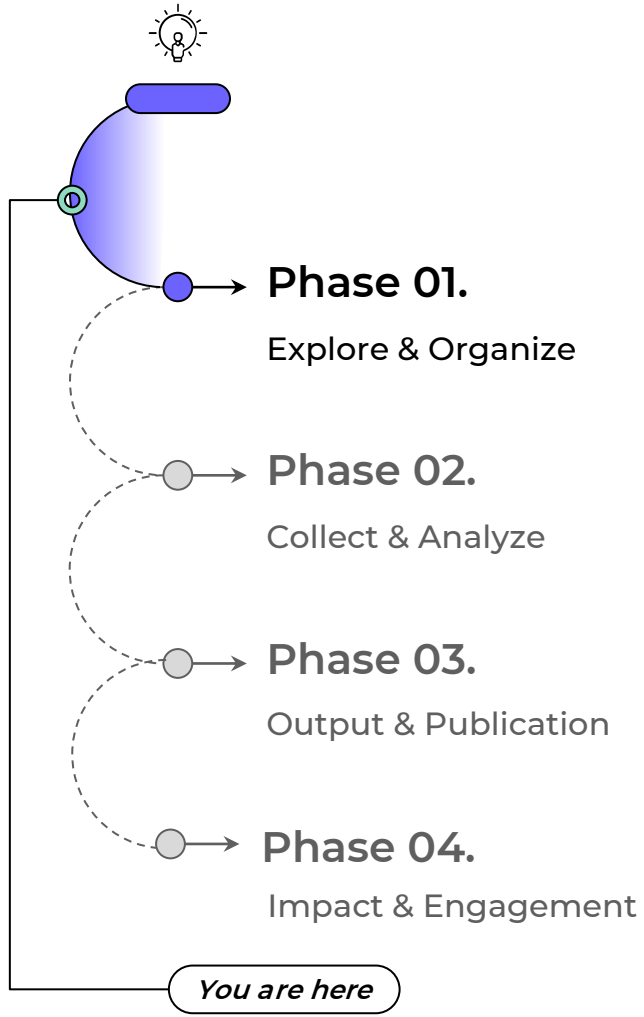
The Research Lifecycle

From Idea to Paper



Presentation by Mansura Naznine | 25.10.2025

Mansura Naznine

Research Assistant

Email: naznine31@gmail.com
linkedIn-mansura-naznine ;
googleScholar-mansura-naznine

About Me

AI/ML researcher focused on computer vision, biomedical imaging, and data-driven healthcare. Published, project-driven; blends research with competitive programming rigor and real-world automation.

Education

Master of Science, Electrical Electronic & System Engineering
National University of Malaysia (UKM)
Mar 2025 - Present; Research (Computer Vision)

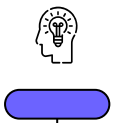
Bachelor's Degree, Computer Science & Engineering
Rajshahi University of Engineering & Technology (RUET)
Jan 2018 - Sep 2023; Research (Video Compression)

Experience

Research Assistant, Qatar University
National University of Malaysia (UKM)
Mar 2024 – Present

Lecturer, Computer Science & Engineering
Lalon University of Science and Arts, Kushtia
Aug 2014 - Apr 2025

Outcomes — What You'll Leave With



1. Prerequisites
2. Research Workflow
3. How to Read a Paper
4. Literature Review
5. Data Preparation
6. Experiment Phase
7. Writing a Paper

Prerequisites for AI/ML Research

01.

Python Foundation

syntax, functions, data types;
libs: NumPy, Pandas, Matplotlib;
scikit-learn; PyTorch/Keras

02.

Statistics

data types; mean/median;
variance/std; CIs;
distributions

03.

Linear algebra

vectors, matrices, eigen-
intuition; where it shows
up (attention, PCA)

04.

Algorithms & DS

complexity,
arrays/lists/dicts/queues;
when optimizations matter

05.

Domain knowledge

Knowing the field = Asking the right questions

You don’t need mastery before starting—treat these as parallel tracks while you work on your first research.

- https://www.youtube.com/@programmingwithmosh
- https://www.programiz.com/
- https://www.youtube.com/c/3blue1brown
- https://www.youtube.com/@statquest

End-to-End AI/ML Research Flow

01 **Define problem & feasibility**: novelty × impact × resources

02 **Literature → find the gap** — read papers, fill a Gap Table (Claim/Evidence/Gap/Your idea), pick angle. “

03 **Data & ethics** — choose dataset, license/consent, clear splits, leakage checks, risks/IRB if needed

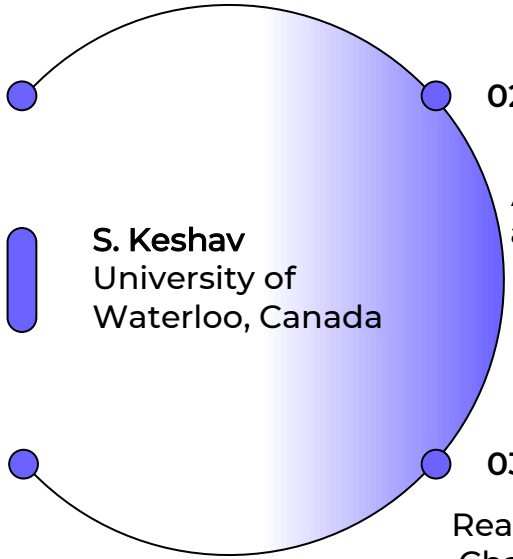
04 **Metrics, baselines, plan** — Preprocessing, strong baselines, compute budget & stop rules

05 **Build reproducible stack → Training** — repo layout, code, configs, fixed seeds, logging, data versioning

06 **Run experiments & validate** — ablations, limited sweeps, val/test, error/fairness/robustness; post-processing.

07 **Write, publish, iterate** — figs + table first, draft paper, publish

How to Read a Paper: THE THREE-PASS APPROACH



First Pass

01.

Read title → abstract → intro → conclusion
Scan section/subsection headings
Answer: Category · Context · Correctness · Contributions · Clarity
Decide: stop or go deeper

S. Keshav
University of Waterloo, Canada

Second Pass

02.

Read carefully: skip details for now
Annotate margins; jot key claims & assumptions, Inspect figures (axes, units, error bars, statistical validity)
Outcome: you can explain the paper's main idea + evidence

Continue

*.

.....

Third Pass

03.

Read implementation details, Method, Challenge line by line, Note strengths, weaknesses, missing citations
Capture: ideas for future work

Where to Find and Download Research Papers?

- 🔷 Google Scholar – The go-to place for free access to papers!
- 🔷 IEEE Xplore – For cutting-edge tech and engineering papers.
- 🔷 ArXiv – Free preprints for the latest in AI, ML, and more!
- 🔷 ACM Digital Library – A treasure trove for CSE research.
- 🔷 ResearchGate – Connect with researchers and request papers directly.
- 🔷 PubMed, Scopus, Web of Science.
- 🔷 Your University Library – Access to journals, paid papers, and more!

💡 Pro Tip: Keep an eye on references in papers – they'll lead you to more gems!

- ❑ Open Access Journals
- ❑ ResearchGate
- ❑ Sci-Hub: The Boss of Accessing Papers!
- ❑ ArXiv
- ❑ University Library Access
- ❑ From Your Network

Literature Review

- Choose Papers from Different Categories
- Summarize Key Points: Focus on Title, Abstract, Problem Domain, and Techniques Used.
- Analyze Dataset & Methods: Understand the data and visualize the methodology with flowcharts.
- Identify Limitations & Gaps: Spot common gaps, limitations, and areas for further research.
- Look for New Research Questions: Use identified gaps to guide your own research direction.

Publisher	Journal Title	Publication Date	Impact factor/ (Quartile)	Article Title	Dataset	Method/ Architecture	Results	Code availability

Gather Domain Knowledge: Before you dive in, know your stuff!”

Data — Source → Prep → Explore → Split

Data is your best friend...
until it's not

01.

Find / Permission: Pick or build a dataset (Kaggle, HF, Roboflow).

02.

Analyze: Shape, missingness, class balance, outliers, histograms, boxplots, scatter

Clean & Prepare: Fix types, deduplicate, impute missing, handle outliers, Scale, normalize; encode categories; balance classes (stratify, weights/SMOTE).

03.

Text: Tokenize, normalize (case/punct), strip HTML/URLs, language detect, Augment

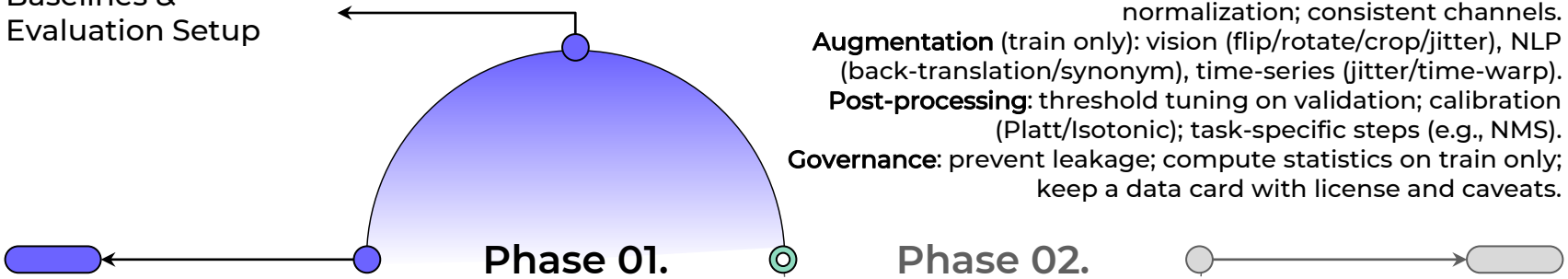
Image methods: Resize/pad, RGB normalize, denoise/CLAHE, Augment

Tabular / Time-Series methods: One-hot/target encode; Standard/MinMax scale; bin/log transforms; Temporal resample, windowing, per-series z-norm; jitter/time-warp (TS aug)

Experiment Phase

Cleaning: types, duplicates, missingness; outlier handling (cap/log/winsorize) with justification.
Balancing: stratified sampling, class weights, focal loss; SMOTE (tabular) when appropriate.
Preprocessing: scaling/normalization; tokenization; image normalization; consistent channels.
Augmentation (train only): vision (flip/rotate/crop/jitter), NLP (back-translation/synonym), time-series (jitter/time-warp).
Post-processing: threshold tuning on validation; calibration (Platt/Isotonic); task-specific steps (e.g., NMS).
Governance: prevent leakage; compute statistics on train only; keep a data card with license and caveats.

Baselines &
Evaluation Setup

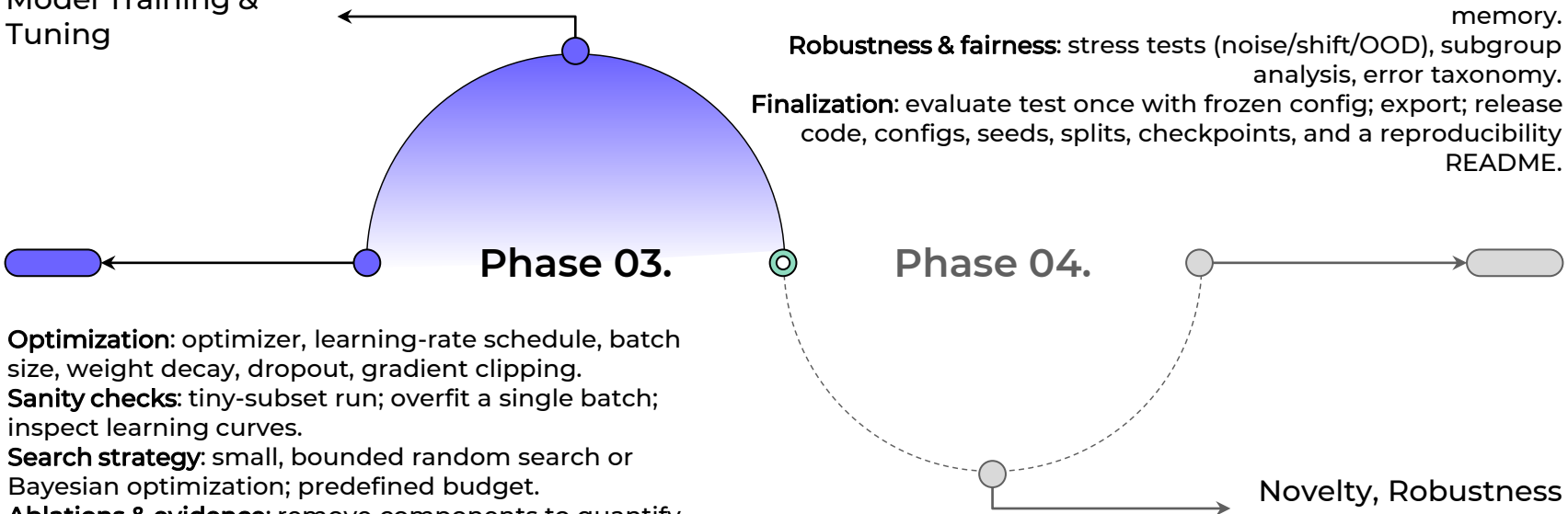


Task & objective: define loss, primary metric, and success criterion (effect size, CI method).
Data protocol: fixed train/val/test splits; stratification or time-aware as appropriate; seeded for reproducibility.
Baselines: trivial → classical → strongest prior you can run; replicate published baseline if feasible.
Reporting standard: median (or mean) ± confidence interval over multiple seeds; document compute, versions, and splits.

Data-Centric
Improvements

Experiment Phase (Cont.)

Contribution: clearly state the novel element (module/loss/pipeline/eval protocol) and the hypothesis it tests.
Controlled comparisons: evaluate against the strongest baseline under identical settings.
Cost–benefit: report accuracy vs. parameters/FLOPs/latency and memory.
Robustness & fairness: stress tests (noise/shift/OOD), subgroup analysis, error taxonomy.
Finalization: evaluate test once with frozen config; export; release code, configs, seeds, splits, checkpoints, and a reproducibility README.



Optimization: optimizer, learning-rate schedule, batch size, weight decay, dropout, gradient clipping.
Sanity checks: tiny-subset run; overfit a single batch; inspect learning curves.
Search strategy: small, bounded random search or Bayesian optimization; predefined budget.
Ablations & evidence: remove components to quantify contribution; keep runs config-driven; log metrics and artifacts; checkpoint best-on-validation.

Writing

Figure-First Workflow

- **Decide first**: 4 plots + 1 table

Main results table with Cis; Training curve (loss/metric vs steps); Ablation bar chart; Error analysis (confusion/failure grid)

- **Storyboard each figure**: question → data → takeaway → caption
- **Table schema**: rows = methods; cols = primary metric (+ CI), key seconds metrics, params/FLOPs
- **Style**: consistent notation, units, axis labels; readable font sizes; highlight primary metric

Paper Structure & Paragraph

Paragraph: Claim → Evidence → Takeaway

- **Abstract**: Problem → Gap → Method → Results (with numbers) → Impact
- **Introduction**: Importance → Gap/limits → Contributions (measurable bullets)
- **Literature Review**
- **Method**: assumptions, model/data pipeline, training details reproducible by others, datasets, baselines, metrics + CI method, protocol, ablations
- **Results**: table/plots first, then analysis + failure cases
- **Limitations & Ethics, Related Work, Conclusion**

Conference vs Journal: Writing Differences

❑ Scope & Length

Conference: concise, sharp contribution; strict page limits; appendices for extras
Journal: comprehensive, extended experiments/theory; more space for proofs & surveys

❑ Novelty & Rigor

Conference: clear delta vs prior work; tight ablations, strong baseline parity
Journal: deeper rigor (theory/derivations), broader datasets, sensitivity studies, reproducibility audit

❑ Artifacts & Transparency

Conference: code/checkpoints encouraged; minimal dataset card acceptable
Journal: full reproducibility package, data cards, ethics statements, extended error/bias analysis

❑ Timeline & Tone

Conference: faster cycles; punchy, result-driven narrative
Journal: slower cycles; exhaustive background, complete method proofs, expanded limitations

Thank you

“TODAY IS THE BEST DAY TO START”