



RESEARCH METHODS MODULE

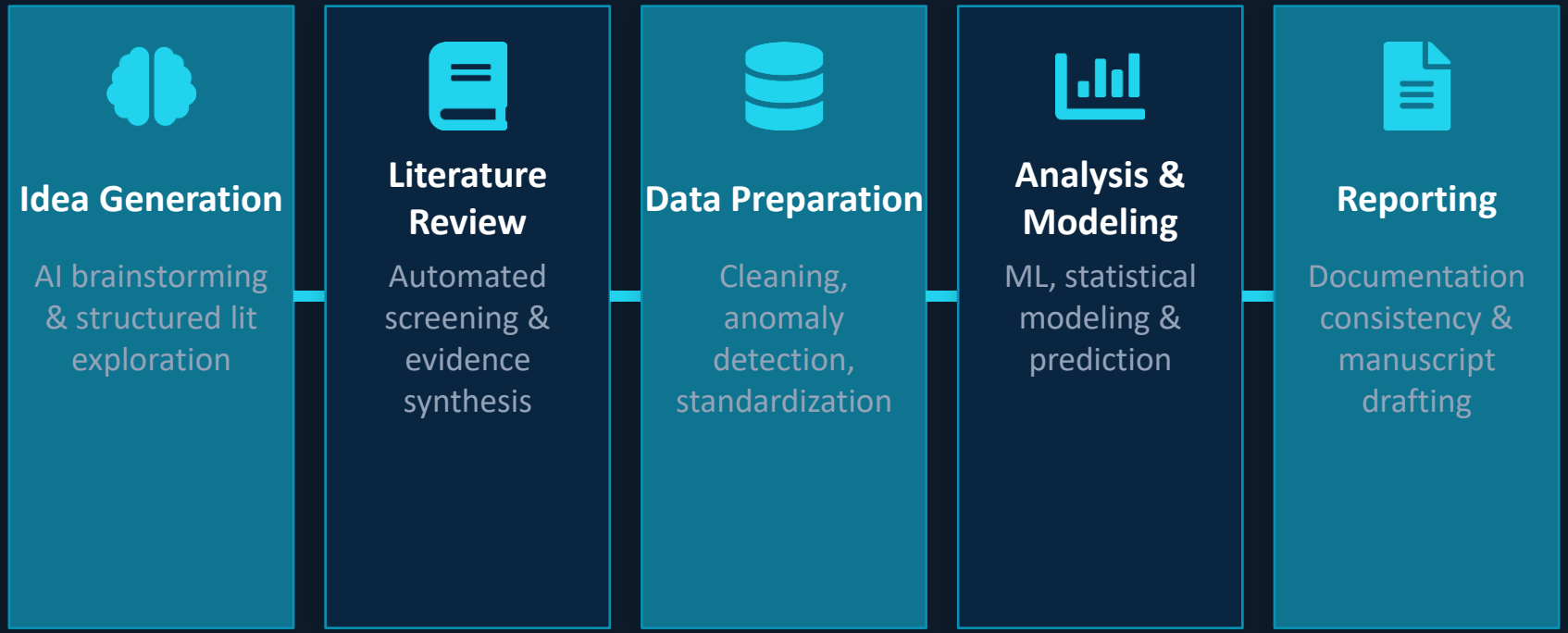
# Leveraging AI for Transparency & Reproducibility

Practical and ethical applications of AI in modern research

By Damian Louis

# AI Across the Research Lifecycle

From ideation to publication — AI enhances every phase



# AI & Research Transparency

- Document all tools, parameters, and workflows used in your research
- AI outputs must remain verifiable and reproducible by independent researchers
- Clearly disclose AI assistance in all publications and reports
- Avoid 'black-box' systems — hidden algorithms obscure decision-making
- Open documentation reduces uncertainty and strengthens public trust

# Reproducibility Challenges in AI Research

## Randomness

AI models rely on randomized processes — fix random seeds for consistent results.

## Software Versions

Different software versions may produce varying results. Always document versions.

## Proprietary Systems

Closed algorithms limit access to underlying structures, hindering independent verification.

## Uncontrolled Parameters

Uncontrolled parameters introduce unintended analytical bias into findings.

✓ **Solution: Standardized documentation + controlled settings + version tracking = reproducible AI research**

# AI for Literature Reviews

- Rapidly screen and categorize thousands of academic publications
- Automated extraction significantly reduces manual coding errors
- NLP identifies thematic patterns across large bodies of literature
- Always independently validate AI-generated summaries for accuracy
- Transparent, documented search strategies are essential for systematic reviews

# AI in Data Cleaning & Preparation

- Automatically detect missing values and inconsistent data entries
- Standardize variables across complex, multi-source datasets
- Pattern recognition accurately identifies potential outliers
- All preprocessing steps must be fully documented for auditability
- Save scripts and use version tracking to ensure reproducible cleaning

# Machine Learning & Statistical Modeling

## Nonlinear Patterns

ML discovers complex relationships invisible to traditional statistics

## Unstructured Data

Neural networks efficiently process text, images, and large datasets

## Hyperparameter Reporting

Full transparency requires documenting all model configurations

## Overfitting Risks

Over-complex models fail to generalize — validate on held-out data

## Interpretability

Tools like SHAP explain algorithmic decisions to stakeholders

# Version Control & Reproducible Workflows

- Version control tracks every incremental change in your research code
- Platforms like GitHub support transparent, collaborative research practices
- Tagged releases preserve stable analytical environments over time
- Shared repositories enhance accountability and enable peer verification
- Proper README documentation ensures long-term reproducibility of analyses

# Ethical Considerations in AI Research

## Embedded Bias

AI systems may reflect and amplify existing societal biases in training data.

## Fairness Audits

Researchers must assess model fairness across demographic groups before deployment.

## Data Privacy

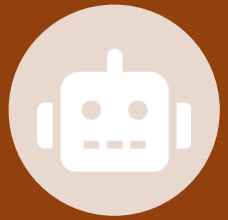
Sensitive data requires strict confidentiality protections and compliance with privacy laws.

## Human Oversight

Automated decision-making must be supervised — humans remain accountable.

## Ethics Committees

AI evaluation components should be integrated into institutional review processes.



# Automation Bias & Human Responsibility

- Automation bias: the tendency to over-trust algorithmic outputs without scrutiny
- Independent verification of AI results is essential for scientific accountability
- AI should augment critical thinking and judgment — not replace it
- Researchers retain full responsibility for all published findings
- Transparency in methods strengthens accountability in AI-supported research

# AI for Research Reporting

## ✓ BENEFITS

- Assists drafting structured manuscripts efficiently
- Automated grammar tools improve clarity and coherence
- Speeds up literature summary and methods sections

## ⚠ RISKS & RULES

- Always manually verify AI-generated citations
- Hallucinated references threaten academic integrity
- Disclose AI assistance in all publications

# AI Policy & Institutional Guidelines

- Universities must establish clear AI disclosure requirements for all research
- Policies should define both acceptable and prohibited AI uses explicitly
- Reproducibility statements must include full AI workflow documentation
- Training programs should promote responsible AI literacy among researchers
- Institutional leadership drives sustainable and ethical AI governance frameworks

# AI Applications in Kenyan Universities

- AI can significantly strengthen systematic review capacity at local institutions
- Automated analysis reduces resource constraints for under-funded research teams
- Infrastructure limitations require strategic, phased technology investments
- Capacity building programs improve AI literacy among researchers and faculty
- Institutional collaboration accelerates responsible and sustainable AI integration

# Best Practices Checklist

Responsible AI in Research

- ✓ Document AI tools, versions, and configurations thoroughly
- ✓ Fix random seeds to ensure reproducible and consistent outputs
- ✓ Archive all datasets and code repositories securely
- ✓ Disclose AI usage transparently within research manuscripts
- ✓ Conduct bias audits across demographic groups before publication

# Acknowledgements

This project was funded by the Berkeley Initiative for Transparency in the Social Sciences (BITSS), under the Center for Effective Global Action (CEGA).