Reflection on Learning: The Journey Through Research Methods and Tools

Charles Watson Ndethi Kibaki

2025-04-10

Abstract

This reflection documents my transformative journey through the Research Methods and Tools course, chronicling the evolution from viewing research as an intimidating, mathematically-focused endeavor to understanding it as a structured process of inquiry accessible to all dedicated learners. I discuss my development of critical research skills including literature review proficiency, methodological literacy, and ethical awareness, particularly as applied to my interest in NLP for low-resource African languages. The document explores challenges faced in navigating methodological complexity and time management, and outlines how these lessons will inform my future academic work. Through this metacognitive process, I've gained confidence in my ability to conduct meaningful research that aligns with my personal interests and has real-world impact for underrepresented communities.

1 Introduction

When I began the Research Methods and Tools course, I approached it with a mixture of anticipation and apprehension. My previous encounters with research methodologies during my undergraduate studies, particularly struggling through Calculus classes, had left me with the impression that research was a domain reserved for mathematical prodigies and statistical virtuosos. This perception was further reinforced when I was advised against pursuing a data science graduate course due to concerns about my quantitative abilities. Entering this course, I carried these doubts with me, unsure if I possessed the necessary aptitude to engage meaningfully with formal research methods.

What I discovered instead was a transformative journey that fundamentally changed how I approach knowledge acquisition, analysis, and contribution to academic discourse. I learned that while statistics and quantitative methods are important components of research, they are merely tools within a broader framework of inquiry. The course revealed the multifaceted nature of research methodology—from choosing research questions to selecting appropriate instruments and applying existing analytical frameworks. In today's era of artificial intelligence, many technical aspects that once seemed intimidating can be abstracted away, allowing researchers to focus on the conceptual and design elements that drive meaningful inquiry.

This reflection documents this journey, from theoretical understanding to practical application, highlighting key learning moments, challenges overcome, and the evolution of my perspective as a researcher. Through this process of metacognition—thinking about my own thinking and learning—I hope to consolidate my growth and establish a foundation for my continuing development as a researcher in the field of responsible AI and low-resource natural language processing (NLP).

2 Evolution of Research Understanding

2.1 From Implementation to Investigation

My background in technology implementation had conditioned me to approach problems with a solution-oriented mindset: identify a technical issue, research existing solutions, implement and test. This approach, while effective for development work, lacks the rigor and systematic inquiry that characterizes academic research. Through this course, I've transitioned from viewing research as merely "finding information" to understanding it as a structured process of inquiry, evidence-gathering, and critical analysis.

A pivotal moment in this evolution came during Week 2's literature review sessions, when I realized that effective research isn't just about finding answers but about asking the right questions. The structured approach to formulating research questions using criteria like specificity, measurability, and relevance transformed how I conceptualized my research interests. Rather than thinking broadly about "improving NLP for African languages," I learned to formulate specific, answerable questions like "What patterns of linguistic diversity exist within the CORD-19 dataset, and what representation gaps can be identified for low-resource languages?"

This shift from implementation-focused thinking to investigation-focused thinking represents perhaps the most fundamental change in my approach to academic work. I now understand research not as a means to an immediate practical end, but as a rigorous process of building and contributing to knowledge that ultimately serves broader aims.

2.2 Developing Methodological Literacy

Prior to this course, my understanding of research methodologies was limited and largely intuitive rather than formal. Terms like "qualitative analysis," "experimental design," and "correlation research" were familiar but not clearly differentiated in my mind. The course provided a comprehensive framework for understanding different methodological approaches and their appropriate applications.

Week 4's exploration of qualitative versus quantitative research was particularly enlightening. I had previously assumed that quantitative approaches were inherently more "scientific" or rigorous, reflecting a bias common in technical fields. Learning about the strengths of qualitative methods for exploring complex phenomena and generating rich insights helped me appreciate the complementary nature of different methodological traditions.

This methodological literacy proved invaluable when designing my research proposal on linguistic diversity in the CORD-19 dataset. I was able to make informed decisions about mixed-methods approaches, combining quantitative analysis of language distribution with qualitative assessment of content characteristics. Rather than defaulting to purely computational methods, I developed a more nuanced approach that could address both the "what" and "why" questions about language representation gaps.

3 Development of Critical Research Skills

3.1 Literature Review Proficiency

The development of literature review skills represented one of my most significant areas of growth. Prior to this course, my approach to literature was haphazard and often focused on confirming existing ideas rather than critically engaging with diverse perspectives. The CRAAP framework (Currency, Relevance, Authority, Accuracy, Purpose) introduced in Week 3 transformed how I evaluate research sources.

Applying this framework to my research on NLP for low-resource languages led me to discover critical perspectives I had previously overlooked. For example, examining the authority and purpose of sources led me to Joshi et al.'s (2020) taxonomy of language resource levels and Nekoto et al.'s (2020) participatory research approach for African languages. These sources fundamentally shaped my understanding of the challenges facing low-resource languages and potential methodological approaches to address them.

Perhaps most importantly, I learned to approach literature not as a collection of facts to be compiled, but as an ongoing scholarly conversation to which I could contribute. This shift from passive consumption to active engagement with literature enabled me to position my own research questions within existing knowledge gaps, making my work more relevant and impactful.

3.2 Data Collection and Analysis Techniques

The course provided comprehensive exposure to both qualitative and quantitative data collection methods. Week 6's focus on questionnaire design principles proved unexpectedly valuable for my CORD-19 analysis project. Although I was working with existing data rather than collecting new information, the principles of clarity, specificity, and unbiased question formulation directly informed how I structured my content analysis categories.

My data analysis skills developed substantially through both theoretical understanding and practical application. The statistical concepts covered in Week 7 helped me interpret patterns in the CORD-19 language distribution with greater sophistication. Rather than simply reporting frequency counts, I learned to apply correlation analyses and representation indices that provided more meaningful insights into linguistic diversity patterns.

A particular breakthrough came when applying text complexity analysis across different languages in the dataset. Initially, I struggled with comparing readability metrics across languages that use different scripts and structural patterns. The course's emphasis on methodology adaptation for different research contexts encouraged me to develop language-agnostic complexity metrics based on sentence structure and terminology density, resulting in more valid cross-linguistic comparisons.

3.3 Ethical Research Practices

Perhaps the most profound development in my research approach has been a heightened awareness of ethical considerations. Week 10's focus on plagiarism and intellectual property expanded my understanding of ethical research beyond simply avoiding direct copying to encompass proper attribution, respect for others' ideas, and transparent reporting of findings.

More broadly, the course heightened my awareness of the ethical implications of research choices throughout the entire research process. When analyzing the CORD-19 dataset, this awareness led me to consider not just technical questions about language identification, but broader questions about what representation gaps might mean for information equity and global health outcomes.

This ethical sensitivity is particularly important for my interest in African language NLP, where issues of data sovereignty, community involvement, and potential unintended consequences of technology deployment are critical considerations. I now approach these questions not as tangential concerns but as central to responsible research practice.

4 Methodological Challenges and Solutions

4.1 Navigating Methodological Complexity

One of the most significant challenges I encountered was navigating the complexity of methodological choices. The range of potential approaches, each with its strengths and limitations, sometimes felt overwhelming. This was particularly evident when designing my research proposal on linguistic diversity in the CORD-19 dataset. I initially struggled to determine whether a primarily quantitative, qualitative, or mixed-methods approach would be most appropriate.

I overcame this challenge through iterative refinement of my research questions and careful consideration of which methodologies would best address each aspect of my inquiry. The course's emphasis on aligning methodology with research questions rather than forcing questions to fit preferred methods was invaluable

in this process. By focusing first on what I wanted to know and then selecting appropriate methods, I developed a more coherent and effective research design.

This experience taught me that methodological complexity is not a barrier to be avoided but a resource to be leveraged. The diversity of research approaches provides flexibility to address different types of questions and triangulate findings through multiple methods. Rather than seeing methodological decisions as binary choices, I now approach them as strategic selections from a continuum of options based on specific research needs.

4.2 Balancing Depth and Breadth

Another significant challenge was balancing depth and breadth in my research. When analyzing the CORD-19 dataset, I was initially tempted to explore every possible aspect of linguistic diversity, from language distribution to content characteristics, text complexity, and named entity patterns. While comprehensive in scope, this approach risked superficial treatment of each aspect.

The solution came through Week 5's discussion of experimental design principles, particularly the importance of clearly defined variables and controlled scope. I realized that effective research often requires strategic focusing rather than exhaustive coverage. By prioritizing key aspects of linguistic diversity that most directly addressed representation gaps for low-resource languages, I was able to conduct more meaningful analysis within the available time and resources.

This lesson in strategic focus has important implications for my capstone project on African language NLP. Rather than attempting to address all aspects of low-resource language processing, I now understand the value of identifying specific, well-defined research questions that can be investigated with appropriate depth. This approach is likely to yield more substantial contributions than a broader but shallower investigation.

4.3 Computational Challenges in NLP Research

A significant challenge I encountered was confronting the computational demands of natural language processing tasks. My work with the CORD-19 dataset revealed how resource-intensive NLP research can be, particularly when analyzing documents across multiple languages. Even basic tasks like multilingual text preprocessing and language identification required substantial computation time and occasionally exceeded available memory resources.

These limitations became particularly evident when implementing cross-lingual analysis techniques. What appeared conceptually straightforward in research papers often proved computationally intractable without specialized hardware. I found myself constantly balancing methodological rigor against practical constraints, making strategic decisions about sampling approaches to work within available resources.

This experience has important implications for my future research in low-resource African languages, where computational efficiency becomes even more critical. It also highlighted broader equity issues in NLP research, where access to computational resources often determines which languages receive attention. Moving forward, I plan to explore more efficient methodologies while seeking collaborative arrangements that can provide access to necessary computing infrastructure for work with underrepresented languages.

4.4 The Role of AI in Research

Throughout this course, I've developed a nuanced understanding of artificial intelligence's role in the research process. While AI tools can significantly assist with data analysis, literature review, and even writing tasks, they cannot replace human judgment in defining research objectives and interpreting results within meaningful contexts.

This insight became particularly clear during my CORD-19 analysis project, where I used AI-assisted tools for language identification and content classification. While these tools enhanced efficiency, the critical decisions about what questions to ask, which methodologies to apply, and how to interpret the findings in

relation to my research questions remained fundamentally human tasks. AI functioned as an amplifier of my research capabilities rather than a substitute for my cognitive engagement with the material.

Professor Mouzonni's guidance on AI use was especially valuable—he emphasized using AI as a thought partner after initial human ideation rather than as a primary idea generator. This advice proved remarkably insightful as I discovered through experience that current AI systems, built on transformer architectures, generally struggle to produce genuine novelty. As Marcus and Davis (2020) argue, these systems excel at pattern recognition and recombination of existing knowledge but lack the conceptual understanding and creative leaps that characterize human innovation. I found that AI functions better as a "regurgitation machine" that can elaborate on and refine human-generated ideas rather than as a source of truly original concepts. This reinforced my understanding that authentic creativity and innovative research questions remain primarily the domain of human researchers, while AI can serve as a powerful amplifier and refinement tool for these human-originated ideas.

I've learned that effective AI integration in research requires clearly defined objectives and careful oversight. The quality of AI-generated outputs depends heavily on the quality of human inputs and direction. This realization has helped me view AI as a powerful collaborator rather than either a threat to research integrity or a magical solution to all research challenges.

5 Research Planning and Time Management

5.1 The Importance of Structured Planning

Effective research planning emerged as a crucial skill throughout this course. Prior to these studies, my approach to project planning was often reactive and ad hoc, responding to immediate needs rather than strategically organizing the entire research process. The course's emphasis on structured research design and timeline development in Week 5 transformed my approach to planning.

For my CORD-19 analysis project, I implemented a phased research plan with clearly defined milestones for literature review, data acquisition, preliminary analysis, in-depth investigation, and synthesis of findings. This structured approach allowed me to track progress and adjust strategies when certain phases, particularly data preprocessing, required more time than initially anticipated.

I've learned that effective research planning is not about creating rigid schedules but about developing flexible frameworks that accommodate discovery and refinement while maintaining overall direction. This balance between structure and adaptability will be crucial for my future research endeavors, including my capstone project.

5.2 Time Management Challenges

Time management emerged as a practical challenge throughout my research projects. The iterative nature of research, with its cycles of literature review, methodology refinement, data analysis, and interpretation, requires careful planning to ensure completion within defined timeframes.

I initially underestimated the time required for certain research phases, particularly data preprocessing and preliminary analysis. For the CORD-19 project, language identification and text cleaning took nearly twice as long as anticipated, creating pressure on subsequent analysis phases. This experience taught me the importance of building buffer time into research schedules and periodically reassessing timelines based on emerging insights.

The course's emphasis on research planning provided valuable strategies for more effective time management. I've learned to view research planning not as a fixed linear process but as an adaptive framework that accommodates discovery and refinement. For future projects, including my capstone, I plan to implement more explicit milestone tracking and regular progress assessment to ensure timely completion without sacrificing research quality.

I will take special regard for Professor Mouzonni's caution about time management and put into practice his particularly insightful advice that follows a logarithmic curve—investing more rigor and grit upfront in the research process, then easing into a more relaxed pace toward thesis defense season. This approach acknowledges that the early phases of research, including literature review, methodology design, and data collection, require intensive effort and meticulous attention to establish a solid foundation. With this groundwork properly laid, the later stages of analysis, writing, and revision can proceed more smoothly and confidently. This logarithmic distribution of effort will likely lead to higher quality outcomes while reducing stress during the final stages of project completion.

6 Application to Future Research

6.1 Direct Methodological Applications

The methodologies and techniques learned throughout this course have direct applications to my capstone project on NLP for low-resource African languages. The literature review framework will enable me to systematically map existing work in this field, identifying key gaps and opportunities for contribution. The mixed-methods approach demonstrated in my CORD-19 analysis provides a template for combining computational techniques with qualitative assessment of language representation and content characteristics.

Particularly valuable will be the experimental design principles covered in Week 5, which will help structure evaluations of transfer learning approaches for cross-lingual information retrieval between high-resource and low-resource languages. The emphasis on clearly defined variables, appropriate controls, and replicable procedures will strengthen the validity of these evaluations.

The data collection and analysis techniques from Weeks 6-8 will inform both corpus development for African languages and subsequent analysis of language patterns. The understanding of sampling strategies will be particularly valuable for creating representative datasets within the constraints of limited available content for low-resource languages.

6.2 Research with Purpose and Context

Perhaps the most significant insight I'll carry forward is the importance of researching topics that deeply resonate with my personal interests and cultural context. My passion for improving NLP resources for underrepresented African languages stems from my own experiences and cultural background. This connection creates an intrinsic motivation that makes the research process more meaningful and sustainable, even when facing challenges.

As Professor Mouzonni emphasized throughout the course, research is not merely an academic exercise but a foundational approach to creating impactful work. By focusing on questions that have real-world significance for communities I care about, I can ensure that my research contributes meaningfully to both academic knowledge and practical applications. This purpose-driven approach will guide my selection of research topics and methodologies in future academic and professional endeavors.

6.3 Continuing Education and Collaboration

I recognize that my development as a researcher is an ongoing journey rather than a destination reached at the conclusion of this course. I plan to revisit the course materials and practice exercises as I embark on my capstone project, using them as reference points for methodological decisions and research planning.

Additionally, I value the importance of maintaining open communication channels with mentors like Professor Mouzonni, whose patient and approachable teaching style helped demystify research methodology and make it accessible. This experience has shown me that research is not reserved for an elite few with exceptional mathematical abilities, but is a disciplined approach to inquiry that can be learned and applied by anyone with dedication and proper guidance.

7 Recommendations for Course Improvement

While the course was generally well-structured and informative, I believe it could be enhanced in two key areas. First, more personalized feedback on assignments would significantly benefit students' development as researchers. I understand the challenge this presents given the number of students and the depth of analysis required, but even brief, targeted comments on methodology choices or analytical approaches would provide valuable guidance. Perhaps a rotating system where each assignment receives more detailed feedback, or leveraging AI tools to provide initial assessment followed by human refinement, could make this more manageable. Second, incorporating mandatory presentation sessions where students must explain their research designs and preliminary findings would strengthen their ability to communicate research effectively. As preparation for thesis defense and professional research communication, these opportunities to present work—even in brief 5-10 minute sessions—would build confidence and clarity in articulating complex methodological choices. These sessions could be structured as peer-review exercises to maximize learning while managing time constraints, with students providing constructive feedback to one another under faculty guidance. These enhancements would further develop two critical research skills that complement the excellent methodological foundation already provided: receiving and incorporating targeted feedback, and effectively communicating complex research designs to diverse audiences

8 Conclusion

My journey through the Research Methods and Tools course has transformed my understanding of research from an intimidating, statistically-focused endeavor to a multifaceted process of structured inquiry that can be approached from various methodological perspectives. I've developed critical research skills that will serve as a foundation for my academic and professional work, particularly in my areas of interest related to NLP for low-resource African languages.

The challenges I've faced—from navigating methodological complexity to managing research timelines—have provided valuable learning experiences that have strengthened my capabilities as a researcher. By finding effective solutions to these challenges, I've gained confidence in my ability to design and execute meaningful research projects.

Looking ahead, I will apply these insights to my capstone project and beyond, approaching research not merely as an academic requirement but as a fundamental tool for creating knowledge with purpose and impact. I am particularly committed to continuing research that supports linguistic diversity and representation in technology, ensuring that digital advances benefit all language communities equitably.

This course has opened doors to new ways of thinking about and engaging with complex problems in my field. Rather than viewing research methodology as a restrictive set of rules, I now see it as an enabling framework that enhances the rigor, credibility, and impact of my work. With these tools at my disposal, I am better equipped to make meaningful contributions to my field while addressing real-world challenges that matter to the communities I hope to serve.

References

Joshi, P., Santy, S., Budhiraja, A., Bali, K., & Choudhury, M. (2020). The state and fate of linguistic diversity and inclusion in the NLP world. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 6282–6293).

Marcus, G., & Davis, E. (2020). *Rebooting AI: Building artificial intelligence we can trust.* Vintage.

Nekoto, W., Marivate, V., Matsila, T., Fasubaa, T., Kolawole, T., Fagbohungbe, T., … & Bashir, A. (2020). Participatory research for low-resourced machine translation: A case study in African languages. In *Findings of the Association for Computational Linguistics: EMNLP 2020* (pp. 2144–2160).