

# **AI for the Sceptical Scholar: Practical Strategies for Using LLMs in Research**

Teal Emery

2025-07-15

# Table of contents

<b>Preface</b>	<b>4</b>
Learning Objectives . . . . .	4
Who This Workshop Is For . . . . .	4
Workshop Scope . . . . .	5
What We Will Cover . . . . .	5
What We Will Not Cover . . . . .	5
Understanding LLMs in Research Context . . . . .	5
About Your Instructor . . . . .	6
Our Two-Hour Journey . . . . .	6
Part 1: Foundations (20 minutes) . . . . .	6
Part 2: Practical Applications (70 minutes) . . . . .	6
Part 3: Advanced Possibilities (15 minutes) . . . . .	6
Part 4: Q&A and Discussion (15 minutes) . . . . .	6
Approaching This Material . . . . .	7
Preparing for the Workshop . . . . .	7
How to Use This Book . . . . .	7
<b>1 About Your Instructor</b>	<b>8</b>
1.1 How I Got Here . . . . .	8
1.2 What I Do Now . . . . .	9
1.3 Other Relevant Experience . . . . .	9
1.4 Why This Workshop? . . . . .	10
<b>I Foundations</b>	<b>11</b>
<b>2 Understanding LLMs as Collaborative Research Assistants</b>	<b>12</b>
2.1 Learning Objectives . . . . .	12
2.2 Why This Matters for Your Research . . . . .	12
2.3 A New Way to Think About AI Collaboration . . . . .	13
2.4 The Jagged Frontier: Understanding AI's Uneven Capabilities . . . . .	13
2.5 What LLMs Do Well vs. What They Don't . . . . .	15
2.5.1 AI Strengths . . . . .	15
2.5.2 AI Limitations . . . . .	15
2.6 The Collaborative Model: You Provide Expertise, AI Provides Scale . . . . .	16

2.7	Triaging Tasks Along the Frontier . . . . .	17
2.7.1	Human-Only Tasks . . . . .	17
2.7.2	Collaborative Tasks (Near the Frontier) . . . . .	17
2.7.3	AI-Assisted Tasks (Inside the Frontier) . . . . .	17
2.8	Finding Your Own Jagged Frontier . . . . .	17
2.9	Key Principles for Success . . . . .	18

# Preface

This book accompanies the workshop **AI for the Skeptical Scholar: Practical Strategies for Using LLMs in Research** for SOAS College of Social Sciences. In two hours, we'll explore how this new technology—despite its limitations—can enhance your research capabilities by handling routine tasks while you focus on critical analysis and theoretical contributions.

## Learning Objectives

By the end of this workshop, you will be able to:

- Evaluate and select appropriate LLM tools for your research needs
- Design effective prompts that leverage your domain expertise
- Use LLMs to enhance literature reviews and cross-disciplinary understanding
- Apply LLMs for coding assistance and data analysis support
- Understand validation approaches for LLM-generated content
- Recognize both the transformative potential and important limitations of these tools

## Who This Workshop Is For

This workshop is designed for experienced researchers who want to explore how LLMs might enhance their work. We assume you have:

- Deep expertise in your research domain
- Healthy skepticism about new technologies and their promises
- No prior knowledge of LLMs or coding experience
- Interest in practical tools that could streamline routine research tasks

Your skepticism is justified—LLMs have real limitations we'll address directly. This workshop provides a realistic assessment of both capabilities and constraints.

## Workshop Scope

Artificial Intelligence is a vast and rapidly evolving field. In two hours, we can only cover a small portion of this landscape. This workshop aims to provide you with three core concepts that will equip you with immediately useful tools and a framework for continued learning:

1. **A mental model** for understanding when and how to use AI effectively
2. **Practical techniques** for common research tasks
3. **Validation strategies** to maintain research integrity

## What We Will Cover

- Consumer-friendly LLM interfaces you can use immediately
- Hands-on practice with real research applications
- Introduction to programmatic possibilities for larger projects
- Case studies demonstrating successful academic use

## What We Will Not Cover

- Comprehensive discussion of AI's social implications (though we acknowledge them)
- Detailed API programming instruction
- Exhaustive review of AI startup tools
- Solutions to replace human critical thinking

## Understanding LLMs in Research Context

Large Language Models represent a new category of research tool. Like any emerging technology, they come with significant limitations: training data biases, lack of contextual understanding, tendency to generate plausible-sounding but incorrect information, and important ethical considerations around consent and knowledge production.

However, when used strategically and with appropriate validation, these tools can transform research workflows. By automating time-consuming routine tasks—initial literature categorization, draft translations, basic coding—LLMs free researchers to dedicate more time to what humans do best: critical analysis, theoretical development, contextual interpretation, and ethical judgment.

## About Your Instructor

## Our Two-Hour Journey

### Part 1: Foundations (20 minutes)

#### Understanding LLMs as Research Tools

- The “Jagged Frontier”: Where AI excels versus where humans remain essential
- Key concepts: model capabilities, cost structures, context windows
- Why Google Gemini for academic work (citations, extended context, NotebookLM)

### Part 2: Practical Applications (70 minutes)

#### Hands-On Tools and Techniques

- Prompt engineering fundamentals with practice exercises
- Creating reusable “Gems” for common research tasks
- Enhancing literature reviews across languages and disciplines
- Getting coding assistance without programming expertise
- Brief exploration of complementary tools (Perplexity, ChatGPT, Claude)

### Part 3: Advanced Possibilities (15 minutes)

#### Scaling Your Research

- Case study: How I classified 18,000 Chinese overseas lending projects in 15 hours (versus 1,500 hours manually).
- Validation strategy: achieving 91.8% agreement with human raters
- Enabling policy-relevant analysis: quantifying green lending patterns across the Belt and Road Initiative
- Introduction to programmatic approaches for large-scale research
- When and how to consider API-based workflows

### Part 4: Q&A and Discussion (15 minutes)

#### Your Questions and Next Steps

## Approaching This Material

This workshop takes a pragmatic stance. We neither dismiss AI’s real limitations nor accept inflated claims about its capabilities. Instead, we focus on practical applications where LLMs demonstrably save time and enhance research capacity while maintaining academic standards.

Throughout, we’ll use clear language and define technical terms as they arise. When we discuss “context windows,” we’ll explain this means how much text an AI can process at once. When we mention “hallucinations,” we’ll clarify this refers to AI’s tendency to generate false but plausible information.

## Preparing for the Workshop

You’ll need:

- A free Google Gemini account (setup instructions in Appendix A)
- A research question or paper you’re currently working on
- Willingness to experiment while maintaining healthy skepticism

## How to Use This Book

Each chapter provides:

- Clear explanation of concepts without unnecessary jargon
- Step-by-step instructions with visual guides
- Hands-on exercises using real research scenarios
- Common pitfalls and how to avoid them
- Validation strategies specific to each application

This book serves as both a workshop companion and a reference for future exploration. The goal is not to make you an AI expert but to provide practical tools that enhance your existing research practice.

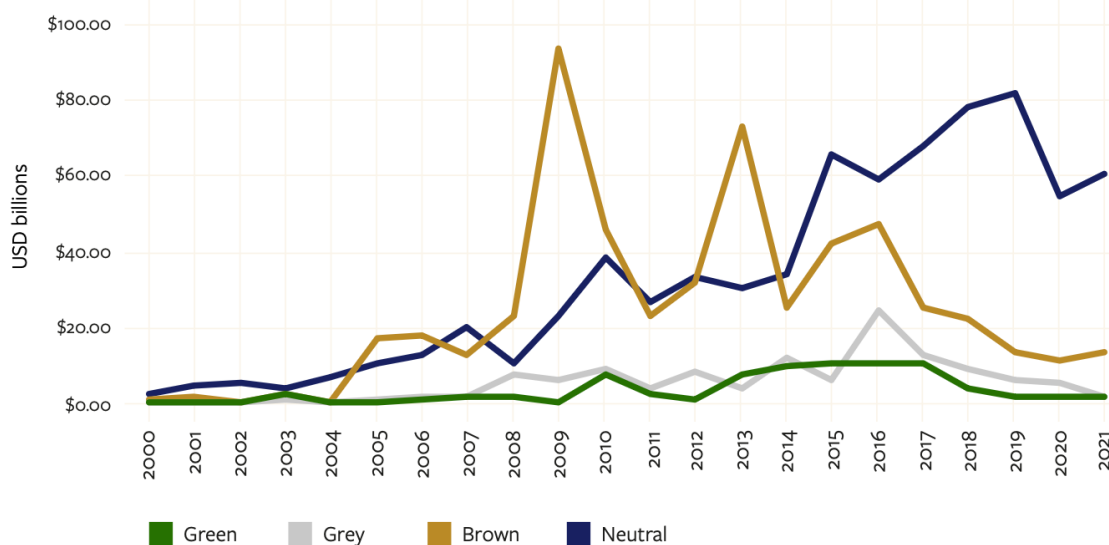
Let’s begin exploring how these tools can support your important work.

# 1 About Your Instructor

## 1.1 How I Got Here

Carlos Oya reached out after seeing a recent ODI Global paper I co-authored with Yunnan Chen called [Greener on the other side? Mapping China’s overseas co-financing and financial innovation](#). We used a novel LLM-based approach to classify “green” Chinese lending projects—something that would have taken a large research team months to do manually.

**Figure 15** Trends in project type lending (2000–21)



Source: Authors’ chart, authors’ categorisation and calculations based on AidData GCDF v3.0

Figure 1.1: Source: Chen & Emery 2025

When we did this work, I looked for “best practices” for validating LLM findings. There weren’t many. So we developed our own validation method and published both a [methodological](#)



[appendix](#) and our [GitHub repository](#). It's not perfect, but it's something for others to build upon.

#### **i** Key Numbers from Our Chinese Lending Project

- **18,000 projects** classified in 15 hours using Deepseek v3
- **\$1.58 total cost** vs. estimated \$22,500 for manual classification
- **91.8% agreement** with human raters on validation sample
- **First comprehensive analysis** of China's green overseas lending portfolio

This experience showed me how LLMs can enhance what's possible for policy-relevant research. Two policy researchers on a tight budget accomplished what traditionally required large, grant-funded research teams. There's a long way to go to establish best practices for the use of LLMs in policy research, so I'm trying to do my best to move the conversation forward.

## 1.2 What I Do Now

**Day job:** Running Teal Insights, where we help Global South finance ministries navigate complex debt sustainability and climate investment challenges. We're philanthropically funded with a mandate to build open-source tools—including LLM tools—so countries don't have to pay exorbitant fees to financial advisors.

**Our approach:** Small team (US, Nigeria, Kenya) using AI tools heavily to amplify our impact in research and code development.

## 1.3 Other Relevant Experience

- **EM sovereign debt research analyst**, Morgan Stanley Investment Management
- **Adjunct Lecturer**, Johns Hopkins SAIS (teaching students to do real-world data analysis on financial and sustainability data)
- **Thought leadership** on sovereign debt + sustainability, World Bank
- **Chinese debt restructuring & flows** research, AidData
- **Big nerd**

## 1.4 Why This Workshop?

### ! A Note on Expertise

This technology is very new. Nobody is really an “expert” yet. But since we’re using these tools extensively, we’ve learned hard lessons about how to use them well—and badly.

When Carlos asked me to teach this, I figured it was a great excuse to organize my thoughts on something I discuss with skeptical, curious researchers all the time.

This is my first attempt at articulating practical guidance for academics who want to use AI responsibly. I hope it’s useful, and I invite all feedback on how to make it better.

# **Part I**

## **Foundations**

## 2 Understanding LLMs as Collaborative Research Assistants

### 2.1 Learning Objectives

By the end of this section, you will be able to:

- **Develop a mental model** for understanding AI as a collaborative research tool rather than a replacement for human expertise
- **Understand the “jagged frontier” concept** and use it to predict where AI will excel versus where it will struggle
- **Triage research tasks responsibly** by categorizing them as human-only, collaborative, or AI-assisted based on the frontier
- **Recognize AI’s key limitations** (hallucination, bias, missing context) and plan accordingly
- **Begin exploring your own jagged frontier** through systematic experimentation with low-stakes tasks

### 2.2 Why This Matters for Your Research

Before diving into the technical details, let’s be clear about why you might want to learn to work with AI: these tools can dramatically expand what’s possible for individual researchers and small teams. When used effectively, AI can handle routine tasks that typically consume enormous amounts of time—literature searches, initial coding, translation, summarization—freeing you to focus on what only you can do: critical analysis, theoretical development, fieldwork insights, and interpretation.

The researchers I know who’ve learned to work well with AI aren’t replacing their expertise; they’re amplifying it. They’re tackling more ambitious projects, exploring research questions they previously couldn’t afford the time to pursue, and spending more of their energy on the intellectually rewarding aspects of research rather than the drudgery.

## 2.3 A New Way to Think About AI Collaboration

Think of Large Language Models not as magical oracles or human replacements, but as sophisticated research assistants with a unique set of strengths and blind spots. [Ethan Mollick](#) suggests a particularly useful analogy:

**treat AI like an infinitely patient new coworker who forgets everything you tell them each new conversation, one that comes highly recommended but whose actual abilities are not that clear.**

This analogy helps us understand how to work with AI effectively:

### Human-like aspects:

- **New on the job:** Needs clear instructions and guidance, may not understand your specific context
- **Coworker relationship:** Works best through collaboration and back-and-forth dialogue

### Non-human aspects:

- **Infinite patience:** Never gets frustrated with repetitive requests or extensive revisions
- **Complete forgetfulness:** Starts fresh in each conversation with no memory of previous interactions

Unlike traditional software that follows predictable rules, LLMs work more like collaborating with a capable but quirky colleague who can be creative and insightful, but may also confidently present plausible-sounding information that's completely wrong.

### **i** Building on Ethan Mollick's Work

This chapter builds heavily on the work of Ethan Mollick, particularly his concept of the “jagged frontier” and his research on human-AI collaboration. I’ve found his insights invaluable in my own journey learning to work with AI. I highly recommend reading his book [Co-Intelligence](#) and following his Substack “[One Useful Thing](#)” for deeper insights into working effectively with AI.

## 2.4 The Jagged Frontier: Understanding AI's Uneven Capabilities

The most important concept for working with LLMs is what Mollick (& esteemed co-authors) calls the “**jagged frontier**” of AI capabilities. Imagine a fortress wall with towers and battlements jutting out at irregular points. Some parts of the wall extend far into the countryside, while others fold back toward the center. This wall represents AI's capabilities—everything

inside the wall represents tasks AI can handle well, while everything outside represents tasks where AI struggles.

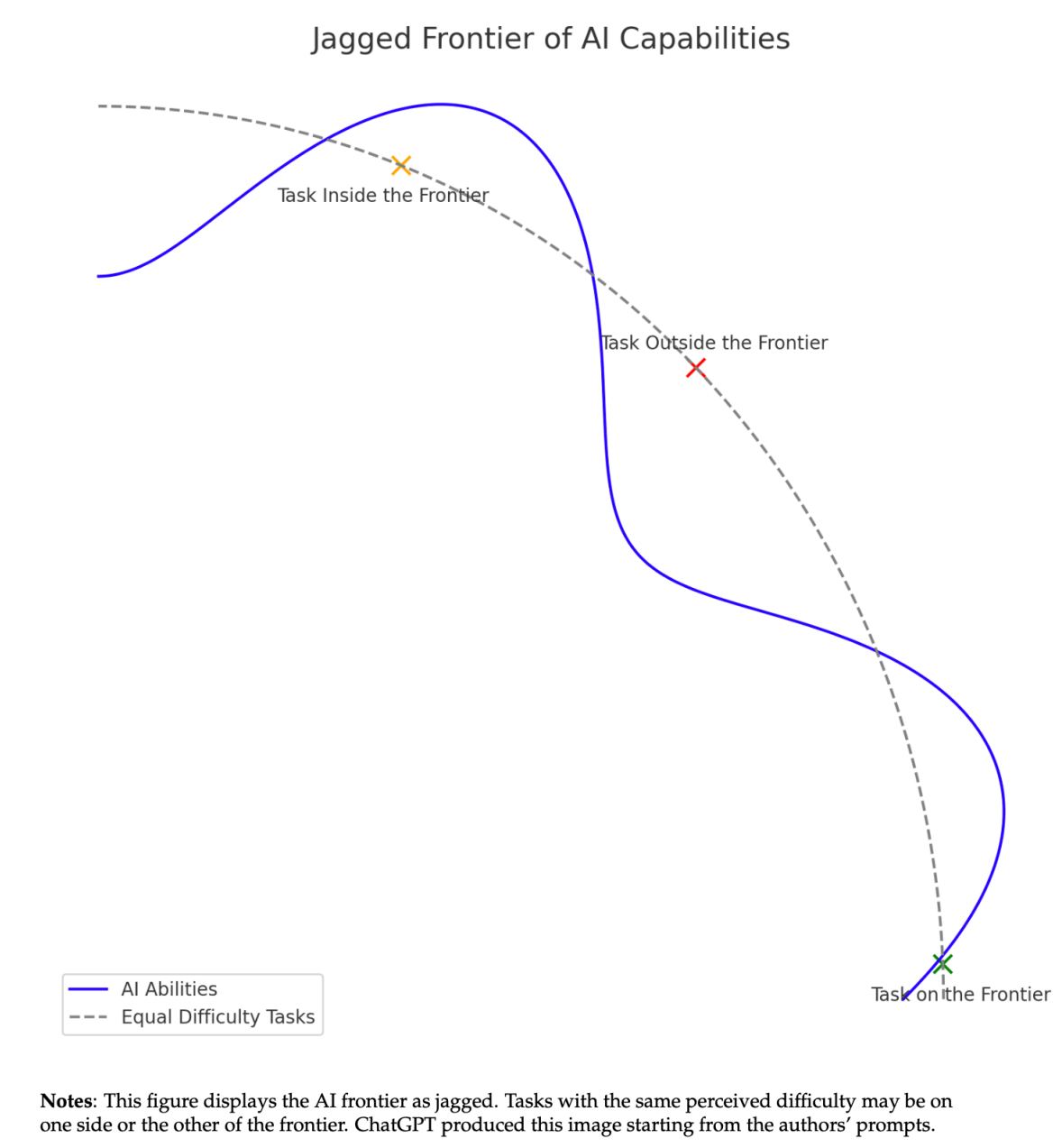


Figure 2.1: Jagged Frontier of AI Capabilities

The challenge is that this wall is invisible. Tasks that seem equally difficult to humans often

fall on opposite sides of the frontier. For example:

**Inside the Frontier (AI excels):**

- Summarizing academic papers and identifying key themes
- Creating first drafts of literature reviews
- Translating research documents between major languages
- Generating research questions and hypotheses to explore
- Coding assistance in most major languages (R, Python, STATA, etc..)
- Writing and formatting citations and bibliographies

**Outside the Frontier (AI struggles):**

- Grasping context that isn't explicitly stated
- Making ethical judgments about research implications
- Humor. Seriously, try it. It's all dad joke vibes

This unpredictability means you cannot assume that because AI handles one complex task well, it will handle a seemingly simpler related task with equal competence.

## 2.5 What LLMs Do Well vs. What They Don't

### 2.5.1 AI Strengths

**Scale and Speed:** LLMs can process vast amounts of text in seconds. Need to identify key themes across 50 research papers? AI can help you get started in minutes rather than weeks.

**Pattern Recognition:** AI excels at identifying patterns across large datasets of text, finding connections you might miss, and synthesizing information from multiple sources.

**First-Draft Generation:** Whether it's grant applications, literature reviews, or research summaries, AI can create useful first drafts that you can then refine with your expertise.

**Language Tasks:** Translation, summarization, and style adaptation are genuine AI strengths that can save researchers enormous amounts of time.

### 2.5.2 AI Limitations

**Hallucination:** LLMs confidently generate plausible-sounding but false information. They might cite papers that don't exist, create realistic-sounding statistics, or confidently state "facts" they've essentially made up.

### What is Hallucination?

“Hallucination” refers to when AI generates plausible-sounding but factually incorrect information. This isn’t a bug—it’s how these models work. They predict what text should come next based on patterns, not facts. A hallucinated research paper might have a realistic title, believable authors, and a publication year that makes sense, but the paper simply doesn’t exist.

**Cultural and Geographic Bias:** LLMs are trained predominantly on text from wealthy countries in the Global North, often in English. They reflect the biases in that data and may default to Western-centric perspectives on development, governance, or social issues.

**Missing Context:** AI only knows what’s explicitly written down. It misses the unspoken context that you understand from fieldwork—the power dynamics in a room, historical tensions between communities, or the significance of what isn’t being said.

**Lack of True Understanding:** When I read IMF documents, boring bureaucratic language often hides spicy geopolitical tensions that you can detect if you understand the context. AI reads the words but misses the subtext entirely.

## 2.6 The Collaborative Model: You Provide Expertise, AI Provides Scale

The most effective approach treats AI as a collaborator rather than a replacement. Here’s how to think about the division of labor:

### **Your Unique Value:**

- Domain expertise and contextual understanding
- Critical analysis and theoretical frameworks
- Ethical judgment and interpretation
- Understanding of implicit meanings and power dynamics
- Ability to validate and verify AI outputs

### **AI’s Unique Value:**

- Processing large volumes of text quickly
- Identifying patterns across many documents
- Generating first drafts and creative alternatives
- Handling routine, time-consuming tasks
- Providing different perspectives to consider



## 2.7 Triaging Tasks Along the Frontier

Use the jagged frontier concept to categorize your research tasks:

### 2.7.1 Human-Only Tasks

Tasks where AI is unreliable or where human judgment is essential:

- Final interpretation of sensitive field data
- Ethical analysis of research implications
- Understanding implicit cultural dynamics
- Making final decisions about research direction

### 2.7.2 Collaborative Tasks (Near the Frontier)

Tasks where AI can help but requires careful human oversight:

- Literature reviews (AI helps find patterns, you verify and interpret)
- Data analysis (AI helps with initial coding, you validate themes)
- Cross-language work (AI provides translations, you check accuracy)
- Grant writing (AI creates drafts, you ensure accuracy and voice)

### 2.7.3 AI-Assisted Tasks (Inside the Frontier)

Tasks you can safely delegate with light oversight:

- First-pass summarization of documents
- Formatting and citation cleanup
- Translation of straightforward technical content
- Creating multiple versions of the same content for different audiences

## 2.8 Finding Your Own Jagged Frontier

The jagged frontier varies between individuals, research domains, and even specific projects. You need to discover it yourself through experimentation. Here's how:

**Start with your own work:** Begin by testing AI on your own papers and research. You'll quickly spot when it gets things wrong because you know the material intimately.

**Begin with low-stakes tasks:** Try AI first on tasks where errors won't matter much—reformatting text, creating bullet point summaries, or generating initial ideas.

**Test systematically:** When you find a task AI handles well, try similar but slightly different tasks to map the boundaries of its capabilities.

**Stay updated:** The frontier is expanding rapidly but unevenly. AI that was terrible at math six months ago may now be excellent due to integrated calculation tools. Assume the AI you're working with today is the worst AI you'll ever use.

## 2.9 Key Principles for Success

1. **Always verify:** Never trust AI output without checking, especially for facts, citations, or quantitative claims.
2. **Use your expertise:** Work with AI on topics where you have deep knowledge so you can catch errors and guide the process effectively.
3. **Embrace iteration:** AI works best through conversation and refinement, not one-shot requests.
4. **Maintain critical thinking:** AI should amplify your analytical capabilities, not replace them.
5. **Document your discoveries:** Keep track of what works and what doesn't for your specific research context.

The goal isn't to become an AI expert—it's to become more effective at research by understanding how to collaborate with these powerful but imperfect tools. In the next section, we'll explore the practical considerations of choosing and using specific AI systems for academic work.