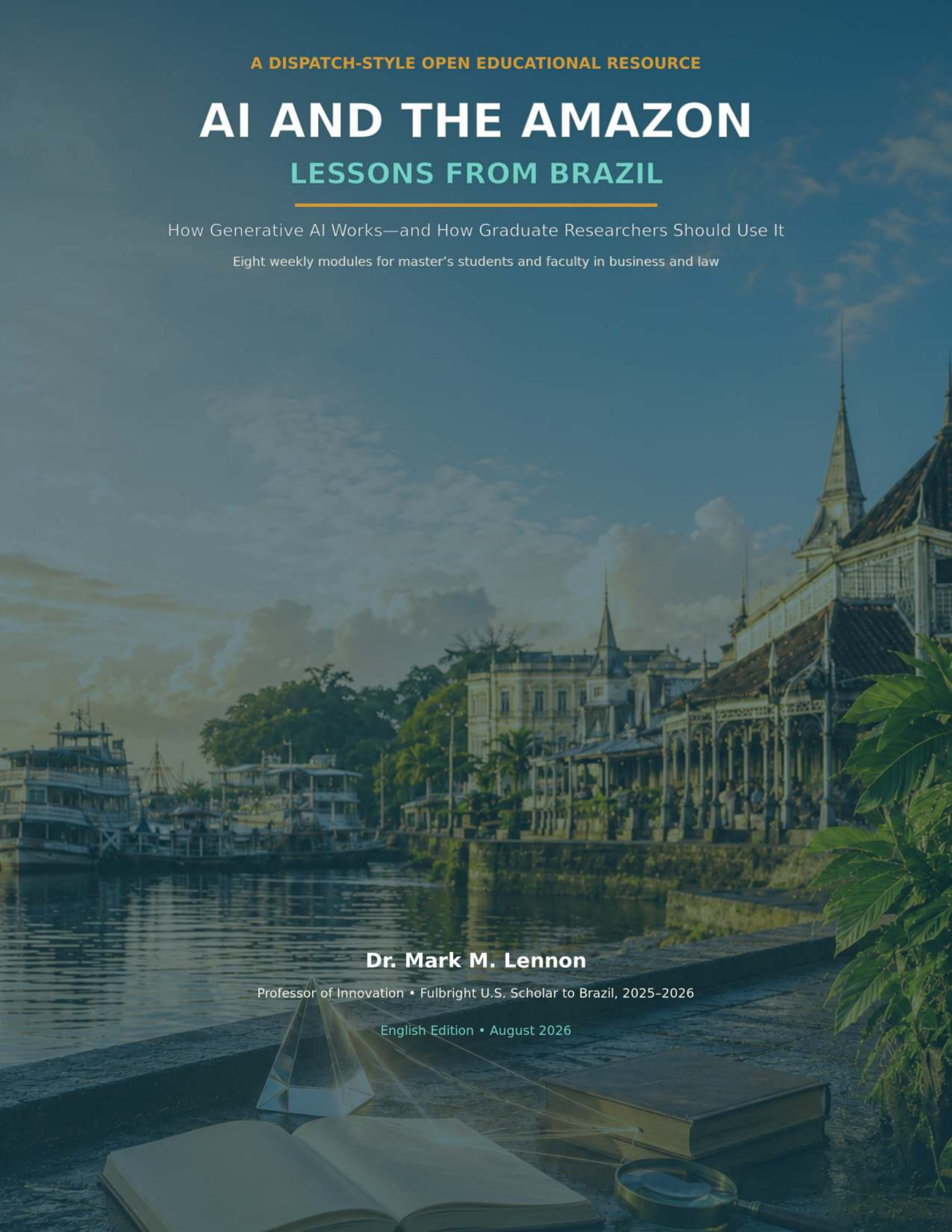
A DISPATCH-STYLE OPEN EDUCATIONAL RESOURCE

AI AND THE AMAZON

LESSONS FROM BRAZIL

How Generative AI Works—and How Graduate Researchers Should Use It

Eight weekly modules for master's students and faculty in business and law



Dr. Mark M. Lennon

Professor of Innovation • Fulbright U.S. Scholar to Brazil, 2025–2026

English Edition • August 2026

Copyright, license, and independence

Copyright © 2026 Mark M. Lennon. All rights reserved except as provided by the open license below.

Open educational license. This work is licensed under Creative Commons Attribution–NonCommercial 4.0 International (CC BY-NC 4.0). Readers may share and adapt the work for noncommercial purposes with appropriate attribution, a link to the license, and an indication of changes. Third-party trademarks and linked services remain the property of their respective owners.

Independence disclaimer. This open educational resource was independently developed by Dr. Mark M. Lennon, drawing in part on his teaching, research, and experiences while serving as a Fulbright U.S. Scholar in Brazil. The interpretations, recommendations, and opinions expressed are solely those of the author. This work is not endorsed, sponsored, approved, or supported by the Fulbright Program, the U.S. Department of State, the United States or Brazilian governments, UNAMA, UFPA, PennWest University, or any other governmental, academic, or institutional organization.

Professor Bear™ and Professor Cat™ are educational characters created for this publication. Professor Bear is the author's teaching alter ego. Professor Cat is a fictional composite used to make technical and linguistic ideas approachable; she does not represent or identify a particular colleague.

⚖️ A NOTE FOR LAW READERS

This book provides education about research practice. It is not legal advice. Never place confidential client information, privileged material, protected personal data, or sensitive research records into a public AI service without authorization and an appropriate risk assessment.

Preface: Why a Fulbright scholar began asking different questions about AI in Brazil

I arrived in Belém expecting to teach artificial intelligence. Very quickly, I realized that the most important lessons would travel in both directions. Graduate students and faculty brought serious questions about research, language, law, business, and development. They also brought constraints that rarely appear in promotional demonstrations: uneven access to paid subscriptions, research materials written in more than one language, local expressions missing from training data, and institutional policies still being written while the technology continued to change.

A recurring misconception emerged in conversations: generative AI was treated as a database that could be improved mainly through better prompts. When an answer failed, the assumed solution was a more elaborate command. Prompting matters, but it cannot make a model possess a source it has never received, correct an imbalanced training corpus, preserve a cautious scholarly claim across languages, or transform a plausible citation into a real publication.

This book therefore begins one step earlier. It asks what generative AI actually does. At a high level, a language model predicts sequences of tokens from learned patterns. That simple sentence explains both its surprising strengths and its persistent weaknesses. The model can produce fluent explanations, comparisons, translations, plans, software, and images. Yet the same fluency can wrap uncertainty, cultural error, or invented evidence in language that sounds authoritative.

The Amazon is not a decorative setting for these lessons. Belém makes visible questions that matter everywhere: Whose language is represented? Which knowledge travels easily? Who can afford the most capable systems? When does a free tool expand opportunity, and when does a paid feature create another research divide? What might locally operated or open-source models make possible—and what new burdens do they create?

The goal is neither enthusiasm nor fear. It is informed use. Every module asks readers to connect technological history to the craft of research: identify a real problem, understand the limits of the method, test claims, verify evidence, document choices, and leave the final judgment with accountable human beings.

How to use this book

The eight modules are designed as weekly readings. Each starts with a short Dispatch from Belém, develops a small number of concepts, returns to research practice, and ends with an assignment. A class can complete the modules in eight weeks or select individual modules for a research-methods course, faculty workshop, or thesis seminar.

The recurring route

- 📍 Dispatch from Belém — a concrete scene before the abstraction.
- 🧠 Core idea — the minimum technical concept needed to reason well.
- 🪄 Misconception versus reality — a claim that sounds right but fails under examination.
- 🔍 Why this matters for researchers — consequences for evidence, writing, and method.
- 🐻🐱 Professor Bear and Professor Cat — audience adaptation without technical theater.
- ⚠️ Research trap — a common failure mode.
- ✅ Verification checkpoint — an action the reader can repeat.
- 🎓 Weekly assignment — evidence of judgment rather than prompt performance.

> **⚠️ THE RULE THAT OVERRIDES EVERY PROMPT**
>
> Do not cite, submit, translate, or act on AI output merely because it is fluent. Locate the source, compare the claim, check the context, and document what you did.

Capabilities notice

Descriptions of ChatGPT, Claude, Gemini, NotebookLM, and other products reflect publicly described capabilities available around Summer 2026. Free tiers, model names, limits, privacy settings, and paid features change. Confirm current terms before using any service. This book never requires a paid subscription.

Contents

Preface: Why a Fulbright scholar began asking different questions about AI in Brazil

How to use this book

Module 1 — Every Technology Begins with a Problem

Module 2 — AI Is Not Magic, It's Math

Module 3 — Égua! Language, Culture, and the Training Corpus

Module 4 — Hallucinations: Useful Possibilities, Dangerous Evidence

Module 5 — Prompting Is Not Enough: From Prompts to RAG

Module 6 — The Graduate Research Workflow

Module 7 — From Chatbots to Multimodal Systems, Apps, and Agents

Module 8 — Who Gets to Participate in the AI Future?

Conclusion: Probability, Possibility, and Responsibility

Quick reference: The responsible research route

Glossary: English and Brazilian Portuguese

Research verification checklists

References

About the author

MODULE 1

Every Technology Begins with a Problem



The Umbrella That Invented Eight More Problems

In Belém, rain does not “begin.” It attacks. One moment, the sky is merely considering its options. The next, the Amazon has dropped a swimming pool on your head. I bought an umbrella from a street vendor who assured me it was “very strong.” It survived eleven minutes. First, the wind turned it inside out. I pushed the ribs back into place. Then the handle came loose. I wrapped it with tape. Next, water began dripping through the fabric. I added a plastic bag.

By the time I reached my destination, I was carrying an umbrella reinforced with tape, plastic, string, and what appeared to be part of a discarded açaí container. It was no longer an umbrella. It was a research program. Each solution had revealed the next problem. The World Wide Web connected scattered information—and created the problem of finding anything. PageRank helped organize the Web—and raised new questions about ranking and influence. Transformers learned relationships among words—and produced fluent answers that were not always grounded in evidence. RAG brought external sources back into the process—and introduced new questions about which sources were retrieved.

Technology rarely arrives as one perfect invention. It usually looks more like my Belém umbrella: one problem solved, three new problems discovered, and someone standing in the rain saying: “Well…that part worked.”

What you will learn

- Recognize generative AI as part of a cumulative history rather than an isolated invention.
- Use the pattern problem → limitation → insight → possibility → remaining problem.
- Connect technological problem-solving to the design of a thesis or dissertation.

> 🧭**ORGANIZING PRINCIPLE**
>
> AI is not magic. It is mathematics, language, data, engineering, and human choices operating together.

From the Web to attention

Tim Berners-Lee's 1989 proposal responded to a practical institutional problem: information at CERN was distributed across incompatible systems and easily lost as people and projects changed. Hypertext linked documents across machines. The World Wide Web did not solve every problem of knowledge; it made a new scale of connection possible.

As the Web expanded, connection created a second problem: finding useful pages. Brin and Page described a large-scale search engine that used link structure as one signal of importance. PageRank did not 'understand' truth. It offered a way to rank an enormous and changing graph of documents.

Nakamoto's Bitcoin paper addressed a different trust problem: preventing double-spending in peer-to-peer electronic payments without routing every transaction through a trusted financial institution. Whether one celebrates or criticizes Bitcoin, the research pattern is instructive. The paper specifies a problem, assembles earlier ideas, proposes a system, and states conditions under which it is expected to work.

The lesson is not that a single paper creates a technology from nothing. Each proposal depends on prior work, infrastructure, implementation, testing, and social adoption. Innovation is cumulative—and so is research.

> 🔍 **RESEARCHER'S QUESTION**
>
> What is the precise problem, why do current approaches fail, and what would count as an improvement?

The road to generative AI

Neural networks learn adjustable relationships from examples. Backpropagation made it practical to calculate how internal parameters should change to reduce error. Distributed word representations later encoded aspects of word use as positions in mathematical space: terms appearing in similar contexts often received similar representations.

The 2017 transformer paper, Attention Is All You Need, proposed an architecture centered on attention rather than recurrence for sequence transduction. Attention lets the system weight relationships among parts of an input. Transformers proved highly parallelizable and became the architectural foundation for many modern language models.

Scale then changed what these systems could do. Training on large corpora, using large numbers of parameters, and adapting models through instruction tuning and human feedback produced systems capable of useful zero-shot and few-shot behavior. Yet the objective remained related to predicting language—not maintaining a verified database of scholarly truth.

This history helps us resist technological mythology. Modern systems are impressive not because they escaped mathematics, but because many mathematical and engineering ideas were combined at unusual scale.

> 🪄 **MISCONCEPTION VERSUS REALITY**
>
> Misconception: one breakthrough paper invented generative AI. Reality: transformers were a decisive architectural development inside a much longer chain of ideas, data, hardware, evaluation, and human labor.

RAG: the next problem becomes visible

A pretrained model can generate fluent language, but its internal parameters are an awkward place to store and update precise, attributable knowledge. Retrieval-augmented generation, or RAG, addresses this limitation by retrieving external passages and providing them as context for generation.

Lewis and colleagues described RAG models that combine parametric memory—the learned model—with non-parametric memory—a retrievable collection. The conceptual advance is easy to state: do not rely only on what the model absorbed during training; give it relevant material at the time of the task.

RAG does not create truth automatically. Retrieval can miss the best source. A collection can contain outdated or biased material. Chunking can separate a sentence from its qualification. The model can still misread the retrieved passage. But RAG makes a better research question possible: Which source supports this answer, and can I inspect it?

The sequence repeats: connection created search problems; scale created grounding problems; grounding systems create new questions about source quality, access, privacy, and evaluation. Every solution reveals the next problem.

> 🎓**THESIS CONNECTION**
>
> A thesis is not ‘about AI’ merely because it uses an AI tool. Start with the business or legal problem, identify the evidence and method, then decide whether AI contributes to a defensible solution.

🎓 Weekly assignment: Read a foundational paper as a problem-solution argument

> 🎯 **PURPOSE**
>
> Practice extracting the research logic from a major technical work without reproducing its mathematics.

Instructions

1. Select Berners-Lee, Brin and Page, Nakamoto, Vaswani et al., or Lewis et al.
2. Read the abstract, introduction, conclusion, and at least one methods or system section.
3. Write six labeled paragraphs: problem, existing limitation, central insight, new possibility, remaining limitation, and lesson for your research.
4. Ask a free AI system for its interpretation only after completing your own notes.
5. Compare the AI interpretation with the original paper and correct every unsupported claim.

Submission

Submit 700–900 words plus a short verification log naming the sections of the original paper you checked.

Verification checklist

- I opened the original work.
- I separated the authors' claims from later historical interpretation.
- I did not claim that one paper created a technology alone.
- Every quotation and bibliographic detail was verified.

> ✅ **ASSESSMENT STANDARD**
>
> Credit depends on documented judgment and verification—not on producing the most polished AI output.

MODULE 2

AI Is Not Magic, It's Math



The Tacacá Lady Knew What I Was Going to Say

There is a tacacá stand in Belém where the woman serving customers appears to possess artificial intelligence. Possibly supernatural intelligence. Before the regular customer finishes speaking, she is already reaching for the correct bowl. "Tacacá com—" She adds the jambu. "Sem—" She removes something. "Mais—" The extra tucupi is already on its way. When I approached, she studied me carefully: American, visibly confused, sweating heavily, and holding my phone like it contained emergency diplomatic instructions.

She predicted that I would want the mild version. She was correct. She had not searched a database containing my official tacacá record. She had recognized patterns from thousands of previous customers and predicted what someone like me was likely to request next. That is also roughly what a language model does. It receives tokens, examines the available context, and calculates what is likely to come next. Then it repeats the process—one token after another—until the result looks remarkably like thought.

But probability is not certainty. The next week, I returned feeling brave. Before I could speak, she prepared the mild version again. "Mais forte," I said proudly. She looked at me. Then she looked at the tucupi. Then she looked back at me with the expression of a woman watching a foreign researcher override the recommended safety settings. The model had made a reasonable prediction. The user had changed.

What you will learn

- Explain tokens, prediction, probability, context, and attention at a high level.
- Distinguish a language model from a database and a search engine.
- Explain why output can vary even when a question appears unchanged.

> 🧭 **ORGANIZING PRINCIPLE**
>
> AI is not magic. It is mathematics, language, data, engineering, and human choices operating together.

Tokens: the small pieces behind fluent language

A language model does not begin with words exactly as a human reader sees them. Text is divided into tokens: units that may be a word, part of a word, punctuation, or another symbol. Tokenization varies by model and language. A common English sequence may require fewer tokens than an uncommon regional expression.

During training, the model encounters enormous numbers of token sequences and adjusts internal parameters so that its predictions become less wrong. At generation time it receives a context and assigns probabilities to possible next tokens. It selects one, adds it to the context, and repeats. Fluent paragraphs emerge from repeated local predictions shaped by learned patterns.

This description is simplified. Modern systems include extensive engineering, multimodal components, safety layers, tool connections, and post-training. But next-token prediction remains the most useful starting point for understanding why a model can sound knowledgeable without consulting a verified record.

Probability does not mean randomness without structure. The distribution is strongly constrained by training, context, instructions, and system design. Yet it is not the same as asking a relational database to return the record with a particular identifier.

> 🐱 **PROFESSOR CAT EXPLAINS**
>
> Imagine an extraordinarily experienced improviser who has studied patterns in millions of performances. She can continue almost any scene—but continuation is not the same act as looking up an official transcript.

Attention, context, and temperature

Attention helps a transformer weigh relationships among tokens in the available context. In a sentence containing several people and pronouns, attention patterns can help connect a pronoun with a likely referent. Across a research prompt, the model can relate an instruction near the beginning to evidence supplied later.

The context window is the material available to the model for a particular interaction: instructions, conversation, uploaded text, retrieved passages, and generated output. A large context window is useful, but capacity is not comprehension. Important evidence can be buried, contradictory passages can compete, and the model may emphasize the wrong portion.

Temperature is one control associated with variation in token selection. Lower settings generally favor higher-probability continuations; higher settings can increase variety. Product interfaces may hide or reinterpret this setting, and deterministic repetition is not guaranteed simply because temperature is low.

For researchers, the practical point is straightforward: save the prompt, model or service, date, supplied sources, relevant settings, and output when reproducibility matters. Even then, a future run may differ because providers update models and systems.

> ⚠ **RESEARCH TRAP**
>
> A larger context window can hold more text, but uploading an entire library without selection may reduce clarity. Context engineering includes deciding what not to include.

Six systems that people confuse

A database stores structured records and answers defined queries. A search engine indexes material and ranks results. A language model generates continuations from learned patterns. A RAG system retrieves external material and supplies it to a generator. A workflow connects defined steps. An agent uses a model to select actions and tools toward a goal, usually inside constraints.

Real products combine these categories. A chatbot may search the Web, run code, retrieve a file, or call an external service. The visible answer can therefore mix generated language with results from deterministic tools. Responsible use requires asking which component produced each important claim.

If a tool reports a calculation, inspect the calculation. If it cites a webpage, open the page. If it summarizes an uploaded contract, locate the clause. If it proposes a legal authority, retrieve the official opinion. The interface may feel unified; the evidence chain is not.

> 🪄 **MISCONCEPTION VERSUS REALITY**
>
> Misconception: AI gave one answer, so the answer must be a stored fact. Reality: a product may generate, retrieve, calculate, or combine these operations. The researcher must trace the source.

> ✅ **ONE-SENTENCE TEST**
>
> Can you explain where the relevant information came from without saying ‘the AI knows’? If not, pause before using it.

🎓 Weekly assignment: Explain one system to three audiences

> 🎯 **PURPOSE**
>
> Demonstrate genuine understanding by adapting—not merely simplifying—the same concept for different readers.

Instructions

6. Choose tokens, attention, context windows, embeddings, or RAG.
7. Write a 100-word explanation for a twelve-year-old, a 180-word explanation for a master's student, and a 150-word explanation for a university dean.
8. Use Professor Bear and Professor Cat in one version if helpful.
9. Ask two free AI systems to critique clarity and identify possible inaccuracies.
10. Revise manually and annotate the change that required the most judgment.

Submission

Submit the three explanations, the two critiques, and a 250-word reflection on what changed across audiences and what could not be simplified safely.

Verification checklist

- The scientific core remains consistent.
- The analogy is labeled as an analogy.
- No audience is treated as incapable.
- I verified technical claims against a reliable source.

> ✅ **ASSESSMENT STANDARD**
>
> Credit depends on documented judgment and verification—not on producing the most polished AI output.

MODULE 3

Égua! Language, Culture, and the Training Corpus



The Day Égua Sent Me Looking for a Horse

Early in my Portuguese education, I learned that **égua** meant “mare.” A female horse. This seemed straightforward—until I arrived in Belém, where people appeared to be discussing horses constantly. Someone saw an outrageous price. “Égua!” A motorcycle nearly hit the curb. “Égua!” A friend opened an unexpectedly large restaurant bill. “Égua!” I began to suspect that Pará had a previously undocumented equine crisis. I looked around for the horses.

There were none. Eventually, a patient Paraense explained that **égua** in Belém can mean something closer to “Wow!”—although tone and context can transform it into surprise, admiration, disbelief, annoyance, or a warning that the conversation is about to become significantly more interesting. The dictionary had given me a correct answer. It had also given me the wrong answer. A language model trained primarily on standardized Portuguese—or translated through English-dominant patterns—may recognize the word and still miss what the speaker is doing with it.

Language does not live only in dictionaries. It lives in neighborhoods, gestures, histories, jokes, and the expression on someone’s face when a professor starts searching Ver-o-Peso for a missing horse. Translation is not merely moving words between languages. Sometimes it is finding the horse that was never there.

What you will learn

- Explain why natural-language fluency is both a strength and a risk.
- Identify harms created by uneven linguistic and cultural representation.
- Verify translations for claim strength, register, context, and regional meaning.

> 🧭**ORGANIZING PRINCIPLE**
>
> AI is not magic. It is mathematics, language, data, engineering, and human choices operating together.

The power of natural language

Natural-language processing lets people work with computers through ordinary language rather than specialized commands. A researcher can request a comparison of theories, a critique of an argument, a translation, a coding plan, or a visual explanation. This lowers the threshold for exploration and makes computational assistance available to readers who do not program.

The same fluency can conceal error. Awkward machine output invites skepticism; elegant output invites trust. A sentence can be grammatical while its evidence is weak, its conclusion is stronger than the source, or its cultural meaning is wrong.

Models learn from corpora shaped by availability, digitization, licensing, filtering, and commercial decisions. English and Global North material are often overrepresented. Portuguese exists in major training corpora, but representation is not uniform across Brazilian regions, disciplines, communities, legal traditions, or varieties of speech.

The resulting problem is not merely translation quality. Underrepresentation affects which concepts are easy to express, which entities are recognized, which examples feel ‘normal,’ and which knowledge is retrieved or repeated.

> 🌴 **ÉGUA CHECK**
>
> A literal translation can be lexically correct and pragmatically wrong. Ask: What action is this expression performing in this community and situation?

The case of the overconfident abstract

A Brazilian graduate researcher—called John Johnson here—prepared an English abstract for a professional-development workshop with guidance from his adviser. The first machine-assisted version sounded excellent. It also made the conclusions and contributions more certain than the Portuguese source justified.

This is a subtle failure. Academic English often rewards concise claims, and a model trained on polished abstracts may transform ‘the findings suggest’ into ‘the study demonstrates.’ It may turn a limited exploratory contribution into a general solution. No individual word looks outrageous, yet the epistemic status of the research has changed.

John used AI again, this time as a comparison aid: identify places where the English sounded more certain than the Portuguese, explain the shift, and propose cautious alternatives. He then manually evaluated the suggestions and restored claims that matched the evidence.

The lesson is not to translate twice and hope. Preserve a claim ledger: original claim, translated claim, evidence level, uncertainty marker, and final human-approved wording.

> ⚖️ **LEGAL PARALLEL**
>
> In legal writing, shall, may, must, should, and can are not interchangeable decoration. Translation must preserve authority, obligation, discretion, and uncertainty.

Local language needs local evaluation

Benchmark scores can obscure uneven performance. A model may perform well on standardized Portuguese while failing on Paraense expressions, Indigenous names, local institutions, or Amazonian environmental and legal contexts. Evaluation should therefore include tasks designed by people who know the community.

Surviving the Corporate Jungle uses animal metaphors to show that familiar workplace language is culturally loaded. BelémLLM Quest and the multilingual Belém guide illustrate another possibility: create experiences around local places and speech rather than asking users to adapt themselves to a generic global interface.

Locally grounded tools should not romanticize locality. They still require consent, representative sources, maintenance, and testing. Small datasets can amplify the preferences of their curators. Community knowledge may be sensitive or inappropriate for unrestricted model training.

The responsible question is not ‘Does the model speak Portuguese?’ It is ‘For which Portuguese, which task, whose knowledge, and under what evaluation?’

> 🔍 **RESEARCH DESIGN IDEA**
>
> Include regional speakers in test design. Record not only whether an answer is correct but whether it preserves register, intent, uncertainty, and cultural meaning.

🎓 Weekly assignment: The Égua translation audit

> 🎯 **PURPOSE**
>
> Evaluate cultural and semantic performance across free AI systems without treating any one system as the answer key.

Instructions

11. Select a regional Portuguese expression or a cautious paragraph from your research field.
12. Obtain outputs from at least two free systems and record the date and model label shown.
13. Back-translate the output, but do not treat back-translation as proof.
14. Ask a knowledgeable human reader to evaluate register, intent, certainty, and cultural fit.
15. Create a claim ledger showing the original, machine output, identified risk, and final revision.

Submission

Submit the claim ledger and a 600-word analysis. Remove personal or confidential information before testing.

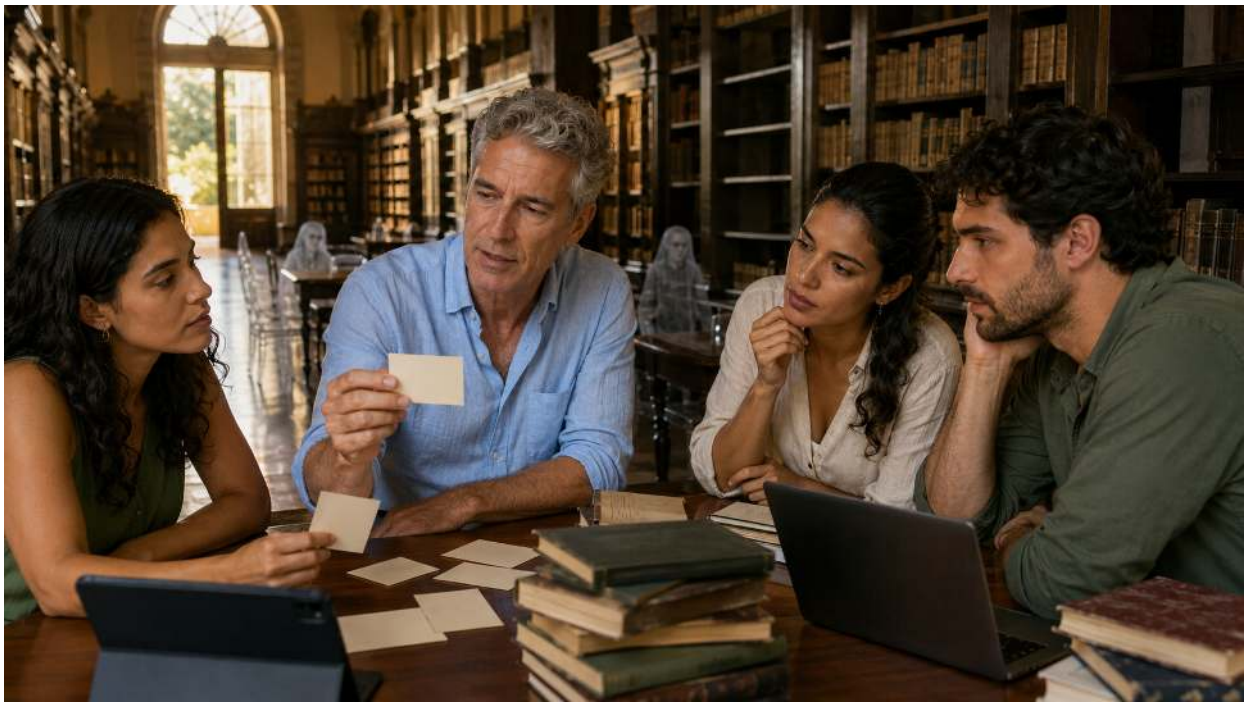
Verification checklist

- The source text was safe to upload.
- A competent human evaluated cultural meaning.
- I checked certainty and obligation words.
- I did not rank a product from a single example.

> ✅ **ASSESSMENT STANDARD**
>
> Credit depends on documented judgment and verification—not on producing the most polished AI output.

MODULE 4

Hallucinations: Useful Possibilities, Dangerous Evidence



The Literature Review Written by Ghosts

When I first applied for a Fulbright U.S. Scholar award, I asked generative AI to help me explore the literature. The results were magnificent. Prestigious journals. Perfect titles. Relevant authors. Recent publication dates. Even convincing page numbers. There was only one small problem. Many of the articles did not exist. I searched Google Scholar. Nothing. I searched the university databases. Nothing. I searched by title, author, keyword, journal, and increasingly desperate variations of all four.

Nothing. More than half of my impressive new literature review had apparently been written by extremely productive academic ghosts. The AI was not plotting against my Fulbright application. It had learned what scholarly references look like and generated plausible examples. The citations had the shape of evidence without the inconvenient requirement of actually being evidence. Fortunately, I checked them. The fictional references could not enter the application, but several invented titles suggested genuinely useful research questions. I discarded the citations and kept the ideas.

That distinction matters. A hallucinated reference can be a search lead. It cannot be a source. A fabricated theory can provoke a question. It cannot support a conclusion. Generative AI is very good at telling us what might plausibly exist. Research begins when we find out whether it does.

What you will learn

- Explain why plausible fabrications arise from generative prediction.
- Separate idea generation from evidence generation.
- Apply a repeatable verification protocol to citations and legal authorities.

> 🧭**ORGANIZING PRINCIPLE**
>
> AI is not magic. It is mathematics, language, data, engineering, and human choices operating together.

Why fabricated evidence sounds real

A citation has a predictable linguistic shape: author names, year, article title, journal, volume, pages, and perhaps a DOI. A language model can reproduce that shape even when no matching publication exists. It may blend a real scholar, a plausible title, and a journal that publishes related work.

Calling this ‘lying’ suggests intention, knowledge of truth, and a deliberate plan to deceive. Those human concepts are usually unhelpful. The model is producing a plausible continuation under its design and context. The absence of intent does not reduce the harm of using the result as evidence.

Researchers should assign outputs to categories. A proposed keyword, conceptual connection, counterargument, or title can be treated as a lead. A quotation, citation, statistic, holding, regulation, or empirical finding is an evidence claim and must be verified at its origin.

The more consequential the decision, the stronger the verification. A classroom brainstorming list needs less scrutiny than a court filing, human-subjects protocol, policy recommendation, or published systematic review.

> 🧭 **THE CENTRAL DISTINCTION**
>
> Possibility can begin a search. Evidence must survive a search.

The source-tracing protocol

Generate. Use AI to suggest search vocabulary, debates, or candidate works. Mark every unverified item visibly.

Locate. Search a scholarly index, publisher site, official repository, court database, or government source. A second chatbot is not independent verification.

Open. Read the actual document. A search-result snippet or generated summary is not the source.

Compare. Find the passage supporting the claim and inspect its surrounding context, method, jurisdiction, date, and limitations.

Verify. Confirm bibliographic elements, quotation wording, page or paragraph location, and whether later work has corrected or contested the claim.

Cite. Build the reference from the verified source, then check the final manuscript against it. Citation managers reduce transcription errors but do not decide whether a source supports a claim.

> ⚠️ **BUSINESS RESEARCH TRAP**
>
> A precise market statistic can be fabricated, outdated, geographically mismatched, or drawn from a vendor's methodology. Precision is not provenance.

> ⚖️ **LAW RESEARCH TRAP**
>
> Generated case names and quotations can look conventional. Retrieve the official or trusted legal text and confirm jurisdiction, status, and subsequent treatment.

Triangulation is a design choice

Triangulation does not mean finding three websites that repeat the same press release. It means combining sources or methods whose weaknesses do not perfectly overlap. An official dataset, a peer-reviewed analysis, and a transparent independent replication may provide stronger convergence than three summaries.

For translation, compare the source text, machine output, and a qualified human reading. For a historical claim, compare the original paper with credible historical scholarship. For a product capability, check the provider's current documentation and test the feature in the version actually available to you.

Record failed verification. If a citation cannot be located, do not silently replace it with a different paper and preserve the original claim. Reassess whether the claim is supportable at all.

The author's application experience demonstrates a balanced position: the fabricated references were unusable, but the generated themes helped shape legitimate searches. The researcher—not the model—converted possibility into evidence.

> **DOCUMENTATION MINIMUM**
>
> Record the prompt or task, tool and displayed model, date, sources supplied, outputs used, verification performed, and substantive human revisions.

🎓 Weekly assignment: Audit an AI-generated literature map

> **🎯 PURPOSE**
>
> Turn a generated list into a transparent record of what exists, what supports a claim, and what must be rejected.

Instructions

16. Ask a free system for six candidate sources on a narrow business or legal research question.
17. Mark the entire output UNVERIFIED.
18. Locate each item through appropriate scholarly or official databases.
19. For each real item, read enough of the original to determine whether it supports the generated description.
20. Classify each result: verified and relevant; real but misdescribed; not located; or outside scope.

Submission

Submit a two-page audit table followed by a 400-word reflection. Do not reward the system for plausibility; reward the process for traceability.

Verification checklist

- Every located source was opened.
- Search snippets were not treated as evidence.
- Unlocated sources were excluded.
- I recorded contradictions rather than hiding them.

> **✅ ASSESSMENT STANDARD**
>
> Credit depends on documented judgment and verification—not on producing the most polished AI output.

MODULE 5

Prompting Is Not Enough: From Prompts to RAG



The World's Greatest Prompt Could Not Order the Fish

At a restaurant in Belém, I decided to demonstrate my growing sophistication in Portuguese. I carefully constructed the perfect order. I specified the fish, preparation, seasoning, side dishes, portion size, dietary preferences, desired level of crispness, and the precise emotional experience I hoped the meal would produce. It was not an order. It was prompt engineering. The waiter listened patiently. Then he said: "Não tem." They did not have it. I tried again using clearer instructions.

Still did not have it. I added context, assigned the waiter the role of "an internationally recognized Amazonian culinary expert," requested step-by-step reasoning, and emphasized that accuracy was extremely important. He pointed to the menu. The fish was still not there. A better prompt could not retrieve a dish the kitchen did not possess. Generative AI works the same way. Detailed instructions can improve structure, audience fit, and clarity. They cannot supply an article the model has not received, repair a weak source collection, or manufacture trustworthy evidence.

That is why researchers need more than prompting. They need sources. RAG gives the system something like a menu: a selected collection from which relevant material can be retrieved and placed into context. Of course, the waiter can still misunderstand the menu. And the kitchen can still overcook the fish. But at least we have established that the fish exists.

What you will learn

- Distinguish prompt improvement from source grounding.
- Explain RAG, embeddings, chunks, and retrieval conceptually.
- Design a small source collection and evaluate its limitations.

> 🧭 **ORGANIZING PRINCIPLE**
>
> AI is not magic. It is mathematics, language, data, engineering, and human choices operating together.

What prompts can—and cannot—do

A good prompt defines the task, audience, constraints, desired output, and available evidence. It can ask a model to distinguish observation from inference, identify uncertainty, show its reasoning structure, or return a comparison in a usable format.

A prompt cannot guarantee factuality. It cannot repair absent or poor sources, discover a confidential document that was never supplied, make a model neutral across cultures, or turn an outdated knowledge base into a current one. ‘Do not hallucinate’ expresses a preference; it does not change the generative mechanism.

Context engineering broadens the work. The researcher selects sources, removes irrelevant material, defines terminology, sequences instructions, provides examples, separates trusted text from untrusted text, and plans verification. Prompting is one element inside this larger design.

The difference resembles assigning a graduate assistant a task. A beautifully written instruction is insufficient if the assistant lacks the cases, dataset, codebook, or articles needed to perform it.

> 🪄 **MISCONCEPTION VERSUS REALITY**
>
> Misconception: the perfect prompt unlocks a hidden correct answer. Reality: prompts shape generation; evidence and verification shape research reliability.

RAG without the machinery

A RAG workflow begins with a collection: articles, regulations, cases, reports, interview protocols, or other authorized sources. Documents are divided into chunks. An embedding model represents chunks and queries as vectors so that semantically related passages can be retrieved.

The system sends selected passages to a language model with the question. The model generates an answer using that context, often with links or citations back to the sources. The model's learned parameters still contribute language patterns and background associations; the retrieved text provides task-specific evidence.

Every stage can fail. The collection may omit a decisive source. Optical character recognition may corrupt a PDF. Chunking may separate a rule from its exception. Semantic retrieval may favor vocabulary similarity over legal relevance. The generator may combine passages incorrectly.

Evaluation must therefore test both retrieval and generation. Did the system retrieve the right passage? Did the answer represent it accurately? Are citations attached to the claims they support? Did the system say 'not found' when the collection lacked evidence?

> 🐻 **PROFESSOR BEAR EXPLAINS**
>
> RAG is an open-book examination. Giving the student books improves the chance of a grounded answer; it does not guarantee that the student chose the right page or interpreted it correctly.

Free and open pathways

NotebookLM offers a convenient source-grounded workflow in which readers add materials and ask questions across them. It is best described as a source-grounded research environment; its internal architecture should not be assumed from the interface. Free chat systems may also accept files, search the Web, or quote provided passages, although limits change.

Open-source or open-weight models can be run through community services or, with adequate equipment and expertise, on local infrastructure. Small language models may reduce hardware needs and support narrow tasks. Local operation can improve control over data flow, but only when installation, logs, access, updates, and storage are configured responsibly.

Open does not mean free of cost, bias, or risk. Institutions may need hardware, electricity, technical labor, evaluation datasets, security controls, and long-term maintenance. A model with weaker Portuguese performance may save subscription fees while imposing hidden quality costs.

Choose the smallest arrangement that satisfies the research need. Sometimes a searchable folder, careful notes, and a spreadsheet provide more transparent evidence management than an ambitious agentic system.

ACCESS PRINCIPLE

Paid tools may offer higher limits or integrated features, but responsible research practice must be teachable with free tools. The method should survive a change of vendor.

🎓 Weekly assignment: Build a miniature grounded workflow

> 🎯 **PURPOSE**
>
> Compare an ungrounded answer with an answer based on a small, inspectable source set.

Instructions

21. Choose three to five authorized sources on a narrow question.
22. Ask a free system the question without supplying sources; save the output as UNGROUNDED.
23. Provide or load the source collection and repeat the question with citation requirements.
24. Check each substantive claim against the cited passage.
25. Identify at least one retrieval failure, generation risk, or missing source.

Submission

Submit both outputs, a claim-to-source audit, and a 500-word evaluation of whether grounding improved reliability.

Verification checklist

- Sources were authorized and appropriate.
- Claims were checked against passages, not citation labels.
- I tested what happened when evidence was absent.
- I did not describe the system as infallible RAG.

> ✅ **ASSESSMENT STANDARD**
>
> Credit depends on documented judgment and verification—not on producing the most polished AI output.

MODULE 6

The Graduate Research Workflow



My Thesis Became an Açaí Bowl

In Belém, açaí is not the sweet purple dessert many Americans imagine. It is serious food. One afternoon, I watched a vendor prepare it. The berries had been harvested, transported, cleaned, processed, strained, checked, poured, and finally served with the customer's chosen accompaniments. Nobody walked into the stall and shouted: "Please generate an açaí bowl using best practices!" The bowl was the final result of a process. Graduate students sometimes approach AI as though research begins with the final command:

"Write my literature review." "Analyze my data." "Produce my conclusions." But a defensible study has its own preparation process. Define the question. Gather the literature. Evaluate the sources. Select the evidence. Ground the task. Generate bounded assistance. Verify the claims. Revise the argument. Document what happened. Skip the early stages and the final product may still look beautiful. So can purple sugar in a plastic cup. That does not make it authentic açaí.

Generative AI can help at almost every step, but it cannot turn an undefined problem, an unreliable source collection, and an undocumented method into credible scholarship merely by serving them in a polished bowl. Research is not the final prompt. It is everything that happened before the bowl reached the table.

What you will learn

- Use an eight-step AI-assisted research workflow.
- Identify appropriate, high-risk, and inappropriate uses.
- Document AI involvement and preserve accountable human judgment.

> 🧭**ORGANIZING PRINCIPLE**
>
> AI is not magic. It is mathematics, language, data, engineering, and human choices operating together.

Define → Gather → Evaluate → Ground

Define the problem before choosing the tool. State the phenomenon, population or jurisdiction, unit of analysis, time period, and decision the research should inform. Ask what would falsify or complicate the preferred explanation.

Gather sources through a defensible search strategy. AI can suggest vocabulary, synonyms, opposing theories, and databases, but the researcher decides inclusion criteria. Record searches so the process can be explained and repeated.

Evaluate authority, method, relevance, date, incentives, and limitations. A prestigious source can be irrelevant; a recent source can be weak; an official source can represent an institutional perspective rather than neutral truth.

Ground the AI task in selected material. Provide definitions, theory, evidence, coding rules, and boundaries. Label sources and instruct the system to distinguish supplied evidence from general background. Remove sensitive material or use an approved environment.

> 🎓**THESIS HABIT**
>
> If you cannot describe the research problem without naming an AI product, the project may still be tool-led rather than problem-led.

Generate → Verify → Revise → Document

Generate for a bounded purpose: propose a coding framework, compare supplied arguments, identify ambiguities, produce alternative structures, or critique a draft. Avoid asking one output to perform every stage at once.

Verify at the level of claims. Recalculate numbers, open sources, compare quotations, test translations, inspect code, and seek contradictory evidence. Ask domain experts to evaluate high-stakes interpretations.

Revise with human ownership. Remove invented transitions, restore uncertainty, strengthen logic, and decide which suggestions to reject. A clean AI draft can weaken an author's distinctive contribution by smoothing away necessary disagreement.

Document enough for readers, advisers, collaborators, or reviewers to understand the role of the system. Disclosure should describe substantive assistance, not merely name a brand. Follow the applicable journal, funder, course, and institution rules as they evolve.

📝RESEARCH LOG

Task • date • service/model label • sources supplied • relevant settings • output retained • verification • human revisions • disclosure decision

Practical uses with boundaries

Literature work. Use AI to generate search terms, compare supplied papers, or locate tensions. Do not allow it to invent the literature.

Qualitative analysis. Use AI to test a codebook or suggest alternative interpretations on authorized text. Preserve the audit trail and do not confuse model consistency with intercoder reliability.

Writing. Use AI to diagnose structure, reduce ambiguity, or adapt a truthful résumé to a fictional job description. Never invent experience, results, credentials, or contributions.

Law. Use AI to organize issues or translate plain-language explanations. Verify every authority and protect confidentiality. Jurisdiction, procedural posture, and current validity matter.

Business. Use AI to generate scenarios or competitor questions. Verify market facts and make assumptions explicit. A strategy is not validated because a model tells a coherent story.



RÉSUMÉ DEMONSTRATION

Safe: identify truthful experience that matches the job and revise wording. Unsafe: add software, results, leadership, or credentials the applicant does not possess.

🎓 Weekly assignment: Complete the eight-step workflow

> 🎯 **PURPOSE**
>
> Demonstrate an accountable AI-assisted research process from problem definition through disclosure.

Instructions

26. Choose one bounded task connected to a business or law research question.
27. Complete the eight stages and maintain a dated research log.
28. Retain one AI suggestion you rejected and explain why.
29. Verify at least three consequential claims or transformations.
30. Draft a two- or three-sentence disclosure appropriate to the work performed.

Submission

Submit the research product, log, verification evidence, rejected suggestion, and disclosure. The AI output itself should be an appendix, not the main achievement.

Verification checklist

- The problem was defined before the tool was selected.
- Sensitive information was protected.
- Verification matched the stakes.
- The final claims are mine to defend.

> ✅ **ASSESSMENT STANDARD**
>
> Credit depends on documented judgment and verification—not on producing the most polished AI output.

MODULE 7

From Chatbots to Multimodal Systems, Apps, and Agents



Professor Bear Got Lost, So Professor Cat Built an App

Professor Bear had a map of Belém. Professor Cat had a phone. This should have been enough. It was not. Professor Bear wanted to find a historic building, practice Portuguese, identify local food, understand a cultural reference, and avoid walking six blocks in the wrong direction in Amazonian heat. The map could show streets. The translation app could translate words. The search engine could find websites. The camera could recognize a building. The audio tool could pronounce a phrase.

Each tool knew one trick. Professor Cat looked at the collection and said, in effect: “Why are we doing all the work?” So we built a small app that connected the pieces. Point the camera at something. Hear its Portuguese name. Learn the cultural context. Find the next location. Ask a question. Receive a challenge. Continue walking. It was not merely a chatbot answering questions. It was a workflow involving language, images, sound, location, retrieval, and action.

Give the system more freedom to choose the next tool, and it begins to resemble an agent. That can be useful—until Professor Bear asks for lunch and the agent decides that what he really needs is a forty-minute educational detour through the history of manioc. The lesson from Belém was simple: Let AI help plan the walk. Do not let it walk away with the map.

What you will learn

- Differentiate chatbots, multimodal systems, workflows, tools, and agents.
- Evaluate products by task and constraints rather than brand loyalty.
- Design a small application that solves a real problem without unnecessary autonomy.

> 🧭**ORGANIZING PRINCIPLE**
>
> AI is not magic. It is mathematics, language, data, engineering, and human choices operating together.

A ladder of capability—not a ladder of intelligence

A chatbot exchanges messages. A multimodal system can work across combinations of text, images, audio, or video. A tool connection lets a model search, calculate, retrieve, or act through another system. A workflow arranges defined steps. An agent uses a model to choose among actions while pursuing a goal.

These categories overlap, and product marketing often stretches the word agent. A long prompt is not necessarily an agent. A workflow that always follows the same branches may be useful automation without autonomy. Greater autonomy increases the need for boundaries, permissions, logging, monitoring, and human approval.

Model Context Protocol, or MCP, is one approach for connecting AI applications with tools and contextual resources through a common protocol. For a nontechnical researcher, the useful idea is interoperability: a model may gain access to a controlled tool or source rather than requiring every connection to be invented from scratch.

The risk also becomes clearer. Once a system can act, an inaccurate interpretation can become an inaccurate action. Start with read-only access, narrow scope, and confirmation before consequential changes.

> 🤖 **AGENT TEST**
>
> What can the system observe, decide, and change? Which action requires human confirmation? Where is the log? How can the action be stopped or reversed?

Choose tools by the job

ChatGPT, Claude, and Gemini provide general conversational and multimodal assistance, with changing combinations of file analysis, Web access, coding, image, and tool features. NotebookLM emphasizes work with user-selected sources. Open-source and open-weight models offer varying degrees of local control and customization.

No static table can declare a permanent winner. Free-tier limits, geographic availability, privacy terms, model labels, and integrations change. Evaluate a tool using the actual task: Portuguese translation, long-document comparison, source-grounded synthesis, image interpretation, coding, or constrained tool use.

Run a small benchmark using representative material. Define success before testing. Include failure cases, regional language, missing evidence, and a request the system should refuse or qualify. Record the date because the comparison will expire.

Do not ask students to purchase several subscriptions to complete ordinary coursework. Design assignments around transferable practices and permit equivalent free tools.

> **SUMMER 2026 NOTICE**
>
> Product descriptions in this book are snapshots, not promises. Consult current provider documentation and your institution’s requirements before use.

The Belém project gallery

Welcome to BelémLLM Quest uses Professor Bear and Professor Cat to explain technical concepts through a playful language-and-culture experience. The characters demonstrate audience adaptation: a complex idea can be made inviting without pretending that the underlying limitation disappears.

Surviving the Corporate Jungle explores office-animal metaphors and the cultural assumptions hidden inside apparently simple phrases. The multilingual Belém guide connects language with movement through the city. AI ABC Brazil turns terminology into approachable reference material.

These are not presented as finished scientific instruments. They are prototypes that demonstrate generative AI's rapid-creation advantage: researchers and teachers can build a small, tailored interface, test it with real users, learn where it fails, and revise.

The design principle is purposeful smallness. A focused tool for one audience and one problem can be more valuable—and easier to evaluate—than an ambitious agent that claims to do everything.

> 🔗**EXPLORE**
>
> BelémLLM Quest: https://00769.netlify.app/ • Corporate Jungle: https://officejungle.netlify.app/ • Belém Guide: https://belem-guide.netlify.app/ • AI ABC Brazil: https://ai-abc-brazil.netlify.app/

🎓 Weekly assignment: Design a small AI application

> **🎯 PURPOSE**
>
> Translate a real research, teaching, business, legal, or cultural problem into a bounded application concept.

Instructions

31. Name the user and the problem before naming a model.
32. Define inputs, outputs, sources, and actions.
33. Specify what the system must never do.
34. Choose free tools or an open-source approach and identify access limitations.
35. Create three evaluation cases: normal, culturally difficult, and high-risk.

Submission

Submit a two-page concept brief plus a one-page evaluation plan. Building a prototype is optional.

Verification checklist

- The concept solves a specific problem.
- Autonomy is no broader than necessary.
- Users can see important limitations.
- Human approval protects consequential actions.

> **✅ ASSESSMENT STANDARD**
>
> Credit depends on documented judgment and verification—not on producing the most polished AI output.

MODULE 8

Who Gets to Participate in the AI Future?



The Twenty-One-Dollar Cup of Coffee

Imagine a café with one item on the menu: Coffee — US$21. The price is the same for everyone. A professor from Pittsburgh buys it after working a relatively small portion of the day. A Brazilian graduate student looks at the same cup and calculates how many hours of work it represents. Same coffee. Same price. Very different sacrifice. Global AI subscriptions often work this way. Companies charge one dollar price and call it equal access. But dollars do not measure the hours of human labor required to obtain them.

Paid systems may provide larger limits, stronger models, longer context windows, advanced research tools, workflows, and agents. When those capabilities help researchers learn faster and perform work others cannot easily reproduce, the subscription is no longer merely a better cup of coffee. It becomes access to capability formation. Open-source models may offer another path—but “free” software still needs computers, electricity, technical knowledge, maintenance, security, and someone willing to answer the question: “Why has the university server been making that noise since Tuesday?”

There is no magical free coffee. But universities can share infrastructure, researchers can build smaller local systems, and developers can design access models that recognize unequal purchasing power. Charging everyone the same price is easy. Giving everyone a meaningful chance to participate is harder. That is the problem worth solving.

What you will learn

- Explain access inequality in terms of labor time and capability formation.
- Evaluate free, institutional, small-model, and open-source responses.
- Create a responsible AI plan that includes access, ethics, and governance.

> 🧭**ORGANIZING PRINCIPLE**
>
> AI is not magic. It is mathematics, language, data, engineering, and human choices operating together.

From price inequality to capability inequality

Comparing exchange rates alone can hide the lived cost of a subscription. Labor time asks how many hours a professor, graduate student, or worker must exchange for the same monthly service. The answer varies by salary, tax, employment status, and country, but the method reveals why a globally uniform dollar price is not globally uniform access.

Paid tiers may offer higher usage limits, larger context, more capable models, integrated research tools, agents, or workflow features. When these capabilities become part of ordinary research practice, exclusion affects more than convenience. It affects the speed at which people can review material, prototype ideas, test questions, and learn emerging methods.

This is capability formation: the opportunity to develop skill by repeated use. A student who reaches a free limit after a few experiments receives less practice and less feedback. A research group without advanced tools may struggle to reproduce work performed elsewhere.

The response should not be mandatory personal subscriptions. Universities, funders, developers, and researchers can explore shared infrastructure, educational access, transparent free tiers, local models, open tools, and assignments that do not equate purchasing power with academic merit.

MEASUREMENT CAUTION

Any labor-time comparison must state the wage source, worker category, date, taxes, exchange rate, and treatment of unpaid student labor. Do not present one number as Brazil’s universal experience.

Open-source possibilities and hidden costs

Open-source software and open-weight models can expand experimentation, localization, and institutional control. Researchers may adapt a small model for Portuguese-first retrieval, run a system offline, or avoid sending sensitive text to a commercial interface.

These possibilities are not costless. Hardware procurement, electricity, cooling, configuration, cybersecurity, backups, updates, technical staffing, and evaluation all matter. A nominally free model can be expensive to operate responsibly. Licensing terms also differ; ‘open’ should be examined rather than assumed.

Regional evaluation is essential. A smaller model may be adequate for classification or retrieval while performing poorly on nuanced academic translation. Hybrid designs can use deterministic search, a local document store, and a model only where generation adds value.

The best workaround may be institutional collaboration: shared collections, common evaluation sets, training for faculty, and infrastructure that distributes capability rather than forcing every student to solve deployment alone.

GLOBAL SOUTH PRINCIPLE

Scientific equality requires more than access to a chat box. It requires language representation, usable limits, training, infrastructure, governance, and the ability to shape research agendas.

Ethics and governance in an emergent landscape

Privacy. Minimize personal data, understand where inputs go, and apply the LGPD and other applicable requirements. Removing a name may not anonymize a distinctive case.

Confidentiality. Legal, employment, health, commercial, and unpublished research information may carry duties that a public AI interface cannot satisfy.

Copyright. Access to text does not automatically authorize training, uploading, reproduction, or publication. Use licensed or appropriately permitted material and distinguish quotation from transformation.

Authorship and disclosure. AI cannot accept scholarly responsibility. Human authors must defend the argument, evidence, analysis, and wording. Disclose substantive assistance according to applicable rules.

Bias and harm. Test performance across language, region, and affected groups. Invite participation from people represented in the data and impacted by the use.

Risk management. NIST's generative-AI profile offers a useful general framework for identifying and managing risks. Policies will continue to evolve; a responsible method should not wait for perfect institutional certainty.

BEFORE YOU UPLOAD

Do I have authority? Is the information sensitive? Can I minimize it? Where will it be processed or retained? Is a local or approved system required?

🎓 Weekly assignment: Responsible AI plan for a thesis

> **🎯 PURPOSE**
>
> Integrate method, evidence, access, language, ethics, and disclosure into one defensible plan.

Instructions

36. State a business or legal research problem and the role AI might play.
37. Identify free tools and an equivalent fallback if access changes.
38. Specify the source collection, verification process, and claim-level checks.
39. Address Portuguese or regional-language performance and evaluation.
40. Address privacy, confidentiality, copyright, authorship, bias, and disclosure.
41. Define a stop rule: the condition under which AI will not be used.

Submission

Submit a 1,200–1,500-word plan and a one-page risk checklist. End with the sentence: ‘The final research judgment remains the responsibility of…’ and complete it accurately.

Verification checklist

- The method does not require a paid subscription.
- Access limitations are explicit.
- Sources and claims are traceable.
- The plan includes a human decision point.

> **✅ ASSESSMENT STANDARD**
>
> Credit depends on documented judgment and verification—not on producing the most polished AI output.

Conclusion: Probability, possibility, and responsibility

Generative AI is powerful because natural language provides a flexible interface to learned patterns, tools, sources, and increasingly multimodal systems. It can help a researcher explore a question, compare supplied arguments, revise an explanation, create a prototype, or communicate across languages. Its power does not come from being a perfect database. It comes from generation.

That same mechanism creates the central responsibility. A plausible continuation is not automatically a fact. A fluent translation is not automatically faithful. A citation-shaped string is not automatically a source. A source-grounded answer is not automatically a correct interpretation. An agent's confident action is not automatically an authorized one.

Brazil and the Amazon make these lessons especially visible without making them uniquely Brazilian. Égua can mean mare or amazement. A global price can be equal in dollars and unequal in labor. An English-dominant corpus can speak Portuguese while still missing local knowledge. A free model can expand access while transferring costs to universities and communities.

The answer is not to reject generative AI or surrender judgment to it. It is to become a better researcher: define the problem, understand the method, choose the evidence, test alternatives, verify claims, document decisions, and remain accountable.

> 🌿**THE LAST WORD**
>
> Todo avanço começa com um problema que alguém decidiu resolver melhor. Every advance begins with a problem someone decided to solve better.

Quick reference: The responsible research route

✅1. DEFINE

State the research problem, boundary, population or jurisdiction, and decision.

✅2. GATHER

Develop and record a defensible source-search strategy.

✅3. EVALUATE

Assess authority, method, relevance, date, incentives, and limitations.

✅4. GROUND

Supply selected evidence and definitions; protect sensitive information.

✅5. GENERATE

Assign a bounded task and distinguish possibilities from evidence.

✅6. VERIFY

Open sources, test calculations, compare translations, and seek contradictions.

✅7. REVISE

Restore nuance, remove invention, and make the argument your own.

✅8. DOCUMENT

Record tool, date, inputs, outputs used, checks, revisions, and disclosure.

Glossary: English and Brazilian Portuguese

Agent — Agente. A system in which a model selects actions or tools toward a goal within defined constraints.

Attention — Atenção. A mechanism that weights relationships among tokens or representations in context.

Context window — Janela de contexto. The material available to a model during a particular interaction.

Embedding — Incorporação vetorial / embedding. A numerical representation used to capture patterns of similarity.

Generative AI — IA generativa. Systems that generate new outputs such as text, images, audio, video, or code from learned patterns.

Hallucination — Alucinação. Plausible but unsupported or incorrect generated content; the metaphor does not imply human intent.

Large language model — Modelo de linguagem de grande porte (LLM). A language model trained at large scale with many parameters and extensive data.

MCP — Protocolo de Contexto de Modelo. A protocol for connecting AI applications with contextual resources and tools.

Multimodal — Multimodal. Able to process or generate more than one modality, such as text and images.

Prompt — Prompt / instrução. Input that directs or contextualizes a model task.

RAG — Geração aumentada por recuperação. Generation supported by material retrieved from an external collection.

Small language model — Modelo de linguagem de pequeno porte (SLM). A comparatively smaller model that may be practical for narrower or local tasks.

Token — Token. A unit of text or other data processed by a model.

Transformer — Transformer. A neural architecture centered on attention mechanisms.

Workflow — Fluxo de trabalho. A defined sequence of tasks, rules, and handoffs.

Research verification checklists

Citation and quotation

- Locate the original source.
- Confirm author, title, year, venue, DOI or stable URL.
- Compare the claim with the source passage and context.
- Confirm quotation wording and location.
- Check corrections, retractions, later treatment, or current legal validity.

Translation

- Protect confidential information.
- Compare claim strength and uncertainty.
- Check duty, permission, and obligation terms.
- Check regional meaning, register, and discipline-specific terminology.
- Obtain qualified human review for consequential text.

Agent or workflow

- List data and tools the system can access.
- Use the least privilege necessary.
- Require confirmation for consequential actions.
- Keep a usable log.
- Provide a stop mechanism and recovery plan.

References

Autoridade Nacional de Proteção de Dados. (2024). Brazilian General Data Protection Law (LGPD): Law No. 13,709/2018. https://www.gov.br/anpd/

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610–623. https://doi.org/10.1145/3442188.3445922

Berners-Lee, T. (1989). Information management: A proposal. CERN. https://www.w3.org/History/1989/proposal.html

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. Computer Networks and ISDN Systems, 30(1–7), 107–117. https://doi.org/10.1016/S0169-7552(98)00110-X

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural Information Processing Systems, 33, 9459–9474. https://arxiv.org/abs/2005.11401

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv. https://doi.org/10.48550/arXiv.1301.3781

Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system. https://bitcoin.org/bitcoin.pdf

National Institute of Standards and Technology. (2024). Artificial intelligence risk management framework: Generative artificial intelligence profile (NIST AI 600-1). https://doi.org/10.6028/NIST.AI.600-1

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. Nature, 323, 533–536. https://doi.org/10.1038/323533a0

UNESCO. (2023). Guidance for generative AI in education and research. https://unesdoc.unesco.org/ark:/48223/pf0000386693

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30. https://arxiv.org/abs/1706.03762

About the author

Dr. Mark M. Lennon is a professor of innovation, strategy, and international business at PennWest University. He developed this open educational resource from teaching, research, prototyping, and conversations with graduate students and faculty while serving as a Fulbright U.S. Scholar in Belém, Pará, Brazil.

His work examines how generative AI can expand research and learning while also reproducing inequalities involving language, access, culture, and institutional capacity. His Belém-based educational prototypes explore AI terminology, multilingual communication, cultural learning, retrieval-grounded research, and the rapid creation of small applications for particular audiences.

🌎AN INVITATION TO READERS

Test these ideas with the free tools available to you. Share failures as well as successes. Keep local knowledge visible. And never allow the elegance of generated language to outrun the strength of the evidence.