

On Feeling Data: Curating What Matters in Local AI Models for Intertextual Engagements

Dr. Antonio Byrd
Associate Professor of English
University of Missouri Kansas City

AI and Writing Across the Curriculum:
Doubting and Believing
University of Denver
February 20, 2026

Scan the QR Code!



Beginning with Doubting

To believe in generative artificial intelligence is to *doubt* its current infrastructure and seek an alternative configuration of GenAI and use.

Asking Different Questions

- What are our responsibilities and values as writing teachers?
- What are our goals and learning outcomes for our students?
- How have digital technologies furthered interests in critical inquiry and problem-solving?
- How and when does generative artificial intelligence represent or serve our interests in human intelligence, community, critical inquiry, and problem solving?

The real issue is not technology or progress per se but the variety of possible technologies and paths of progress among which we must choose. Determinists claim that there are no such alternatives, that technological advance always and everywhere leads to the same result. This view is increasingly contested by students of technology. But if alternatives do exist, the choice between them will have political implications.

-Andrew Feenberg, *Transforming Technology*, 2002

Local First Software and GenAI?

A set of principles for software that enables both collaboration and ownership for users. Local-first ideals include the ability to work offline and collaborate across multiple devices, while also improving the security, privacy, long-term preservation, and user control of data.

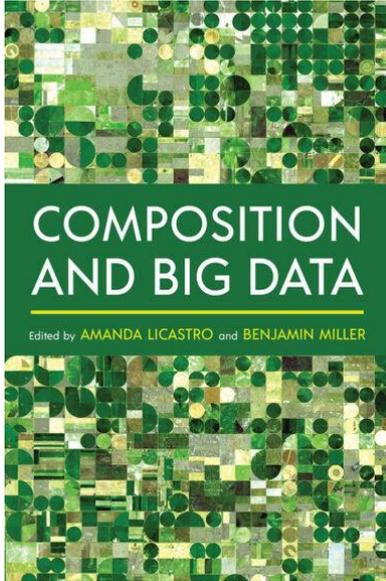
-Ink & Switch

“Local-First Software: You Own Your Data, in Spite of the Cloud,” 2019

What Technology Does

Technologies should be designed not to farm data but help ***facilitate what we can do*** with data for critical inquiry, problem-solving, and knowledge building.

Responsible Data Handling



Scaling Data with No Promises

Is Your Book in the Anthropic Settlement Class? Find Out and File a Claim

October 2, 2025

Share [🐦](#) [f](#) [in](#) [✉](#)

How Tech Giants Cut Corners to Harvest Data for A.I.

OpenAI, Google and Meta ignored corporate policies, altered their own rules and discussed skirting copyright law as they sought online information to train their newest artificial intelligence systems.

Anti-piracy group shuts down Books3, a popular dataset for AI models

The dataset was part of a larger project called The Pile, which aimed to provide open-source data for language models.

Data Cultures in Conflict





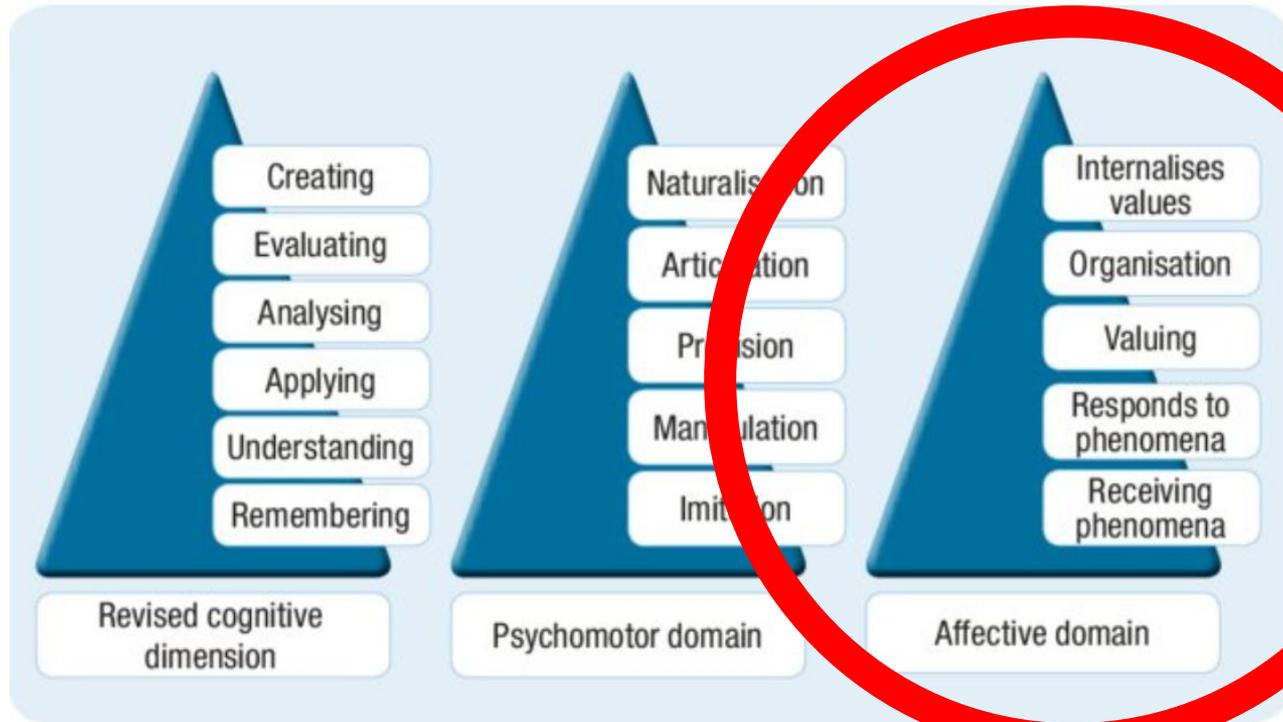
If Not a Technology, a Mindset

Teaching digital literacy as **local data ownership** and local data manipulation (creating, writing, revising, and deleting personal local data) that's motivated by our **affective attachments** to such data

A Use Case

1. Local AI models for writing process portfolios
2. Hold space for the affective
3. Invite other human sources and voices for intertextual engagement
4. Prompting for reflection about process and relationship among pieces in the model

Attention to Affective Learning Outcomes



Planting Seeds



Oh, hold on to your anger and use it as compost for your garden.

–Thich Nhat Hanh to bell hooks



What's Happening Today?



- 01** Stories of Affective Attachments to Data
- 02** All Your Data Are Belong to Us
- 03** The Local-First Software Paradigm
- 04** Feeling Local Data in Writing Portfolios
- 05** What Does It Mean to Feel Data in AI Models?

1

Stories of Affective Attachments to Data

A Follow Up 4 Years Later



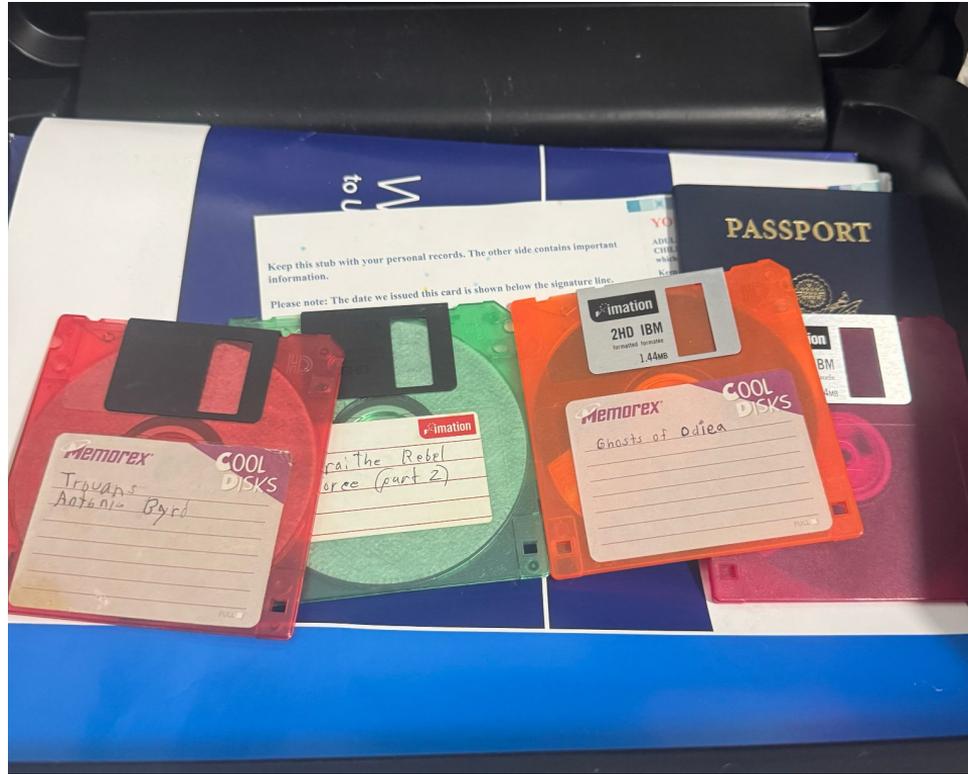
Sunset over Limuru tea farm. Source: Irewolede

Searching for Black Voices for AI



Source: Emmanuel Ikwuegbu

Antonio's Magnum Opus



Source: Antonio Byrd, January 1, 2026

Not Antonio's Jump Drive

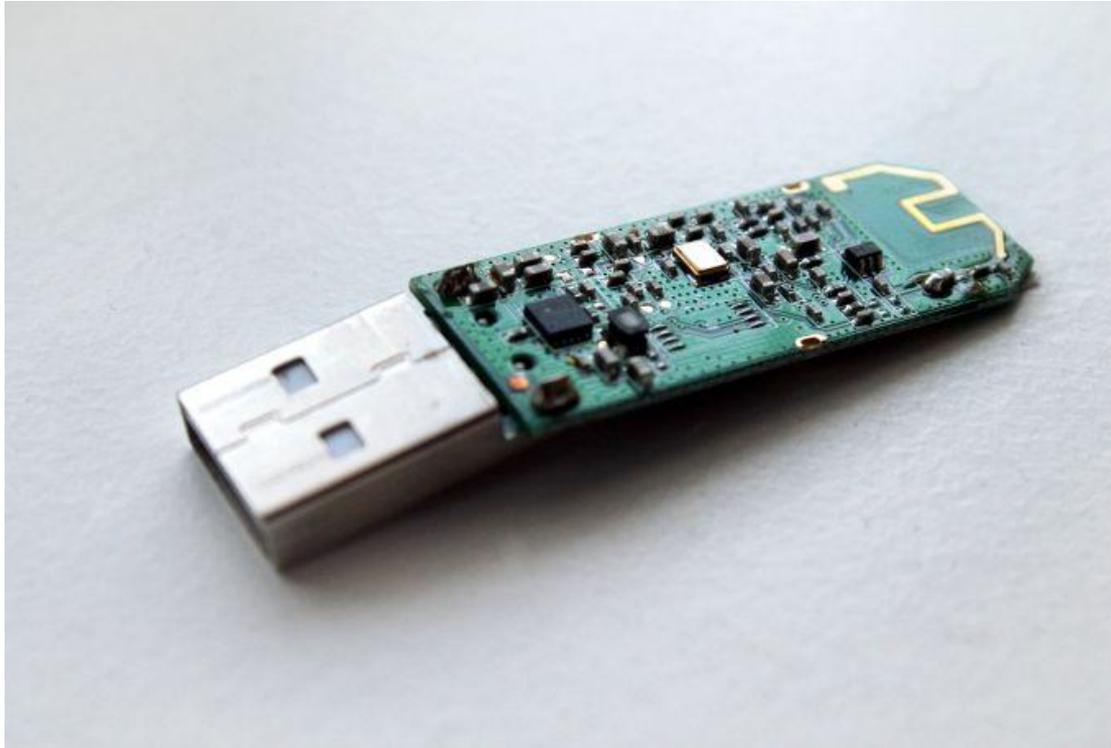


Source: Etsy, [DeborahLaun](#)

The best way to understand Google Drive is to look at it as an evolution of letting people work successfully in the cloud. We have focused a lot on collaboration with Google Docs and we are doing the same thing now with Google Drive.

-Scott Johnston, product manager, 2012, [ABC News](#)

Not Antonio's USB Drive But Similar



Source: [Help Desk Geek](#)

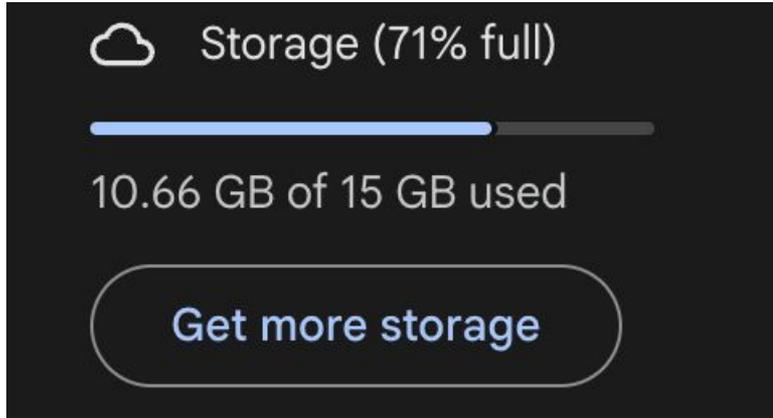
The Data We Carry

How much data have you created throughout your professional career?

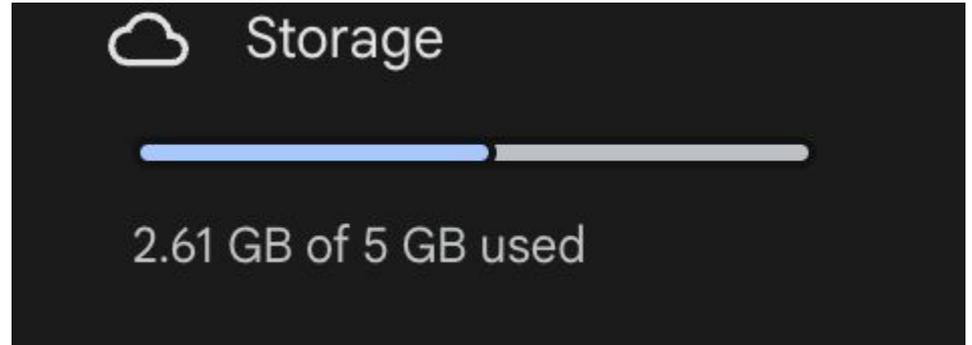
How do you carry your data now and how have you carried your data in the past?

If you have data from 20 years ago, why do you still keep them? How and why are you attached to them?

Antonio's Life in GigaBytes



Google Drive storage in personal account



Google Drive storage in professional account

A Lot of Shared Documents

The screenshot displays the Google Drive interface with a dark theme. The left sidebar shows navigation options: Home, Activity, Workspaces, My Drive, Shared drives, Shared with me (highlighted), Recent, Starred, Spam, Trash, and Storage. The main area is titled "Shared with me" and includes a search bar and filter buttons for Type, People, Modified, and Source. A "Date shared" dropdown is set to "Last week". Below this, three document thumbnails are visible: "Resume of Madelyn...", "Technical Description ...", and "Final Draft for Grant...". A "Last month" section follows, displaying a grid of document thumbnails including "Your big idea", "2025 RESUME - Vy...", "Hotspot Distribution...", "*Document Level...", "Suleiman Atieh...", "Rullfsson: KC LitLink - ...", "KC Support Sync App - ...", "Prototype", "Activity 21: Create a...", "ice Navigator", and "Healthy Habits". The bottom left corner indicates "2.61 GB of 5 GB used".

Not Personal Data But Data Created

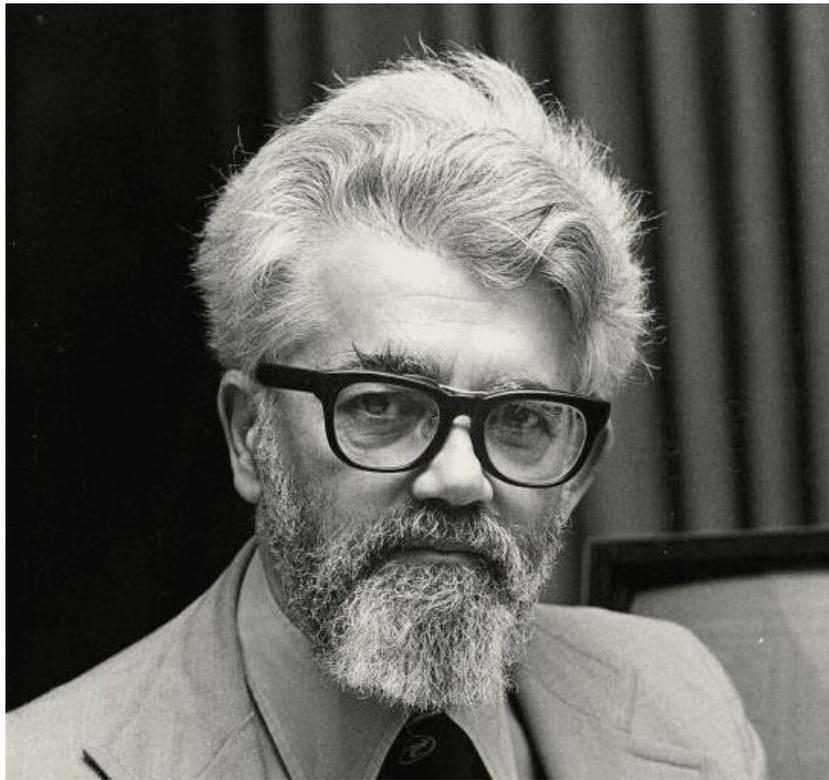
Documents, presentations, spreadsheets, computer code, notes, drawings, photos, and other content that record expression and intellectual property.

-Ink & Switch

“Local-First Software: You Own Your Data, in Spite of the Cloud,” 2019

2

**All Your Data Are Belong to
Us**



John McCarthy

Coined "Artificial Intelligence"
Imagined Cloud Computing



J.C.R. Licklider

Intergalactic Computer Network
Early version of cloud computing

Data Centers: They Kinda Got Big, Uh?



IBM's System 370 Data Center
Source: [IBM](#)



AWS data center looms over houses
Source: [Inside Climate News](#)

We don't own our data in the cloud.

We Only Have Copies of the Original

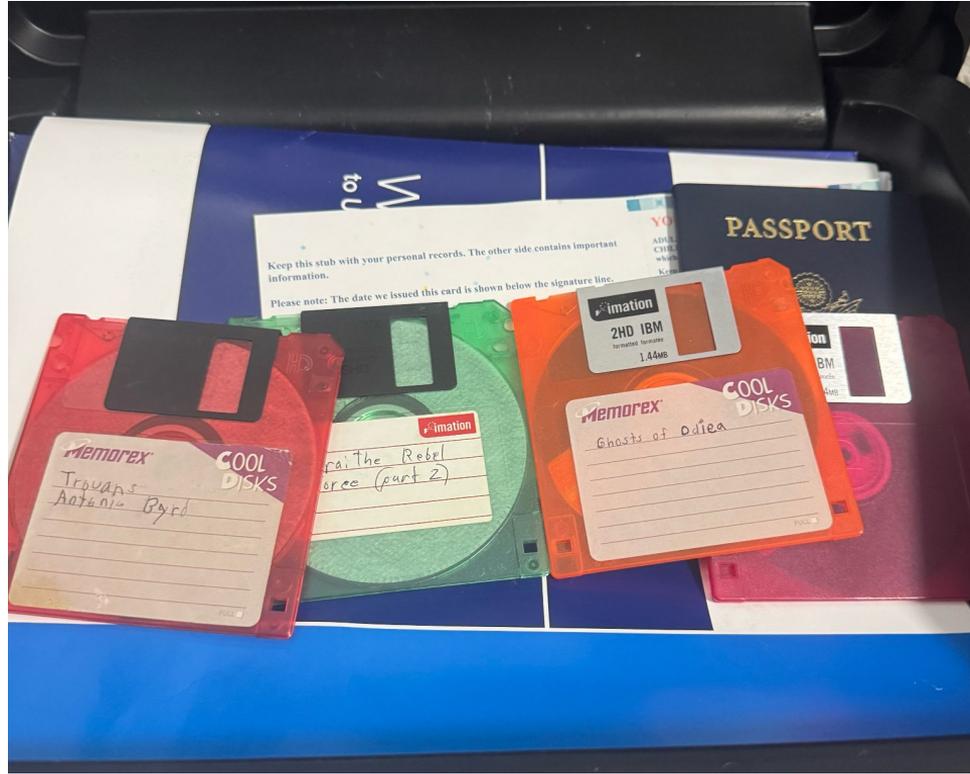


Jamillah Knowles & Digit / <https://betterimagesofai.org> / <https://creativecommons.org/licenses/by/4.0/>

Data: More Than a Cognitive Experience

1. Creation of and interaction with data involves an emotional experience
2. We give our agency, memory, and entire being to the cloud, not just data
3. Digital literacy practice involves preserving data representations of our embodiment

Affective Attachments to Floppy Disks



How Can We Find Joy with GenAI and Data?



Source: [Jacqueline Munguía, Unsplash](#)

3

The Local First Software Paradigm

Local First Software

A set of principles for software that enables both collaboration and ownership for users. Local-first ideals include the ability to work offline and collaborate across multiple devices, while also improving the security, privacy, long-term preservation, and user control of data.

-Ink & Switch, 2019

What Ownership Means

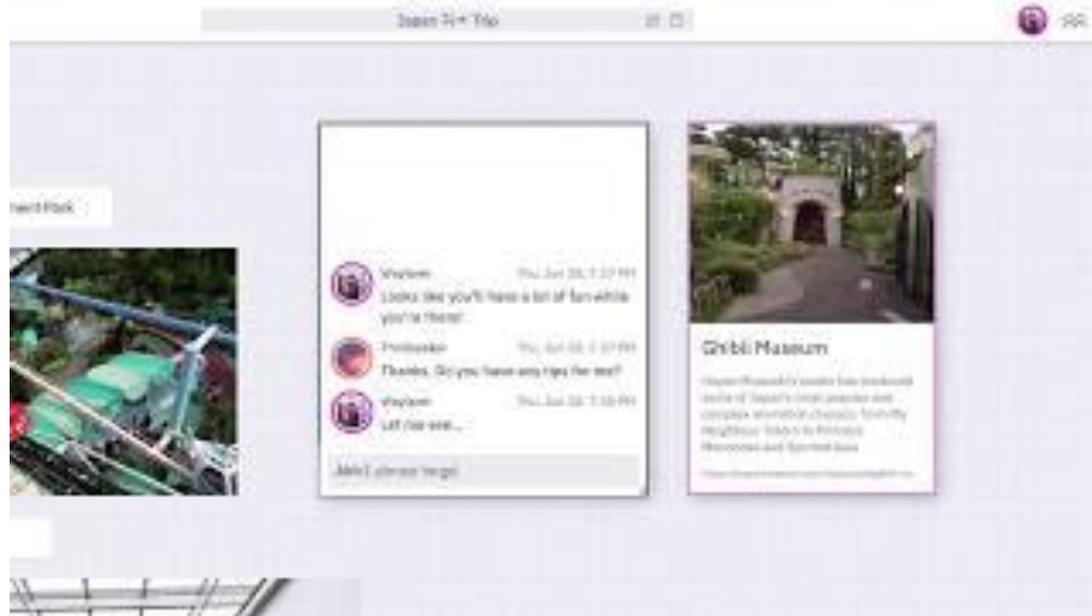
[T]he sense of agency, autonomy, and control over the data. You should be able to copy and modify data in any way, write down any thought, and no company should restrict what you are allowed to do.

-Ink & Switch, "Local-First Software: You Own Your Data, in Spite of the Cloud", 2019

Conflict-free Replicated Datatypes

a data structure which allows multiple replicas to be updated on separate devices and updated when those device communicates with each other

Example: PushPin



4

Feeling Local Data in Local AI Models

Is Data Sovereignty Possible?

Latam-GPT

A language model made in Latin America, for Latin America



Source: Canadore College Advances Indigenous Language Preservation Through AI-Driven Collaboration (2026)

“A Rocket Ship to Visit the Grocery Store”

A local first paradigm doesn't ask:

Which company should we make an agreement with? What AI detector should we give to faculty?”

Instead it asks:

“What generative artificial intelligence platform gives us local control and local responsibility?”

Stored Stacked Files That Don't Engage



Zotero



Google Drive



Nvivo

Local AI Models for Portfolio Assessment?

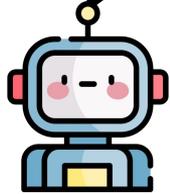
E-portfolios can be “web-sensible”—a thoughtfully arranged collection of multimedia-rich, interlinked, hypertextual documents that students compose, own, maintain, and archive on the Internet or in other formats.

Principles and Practices in Electronic Portfolios
CCC Taskforce on Best Practices in Electronic Portfolios (2007)

Reflection in Local AI Model



Students' data
(notes, zero drafts, etc.)



LLM
(a model of culture)

What and who mattered
in your writing process?

How does your primary
data link with other
human voices? Where did
it digress?

What values circulate in
your portfolio?



Other human sources
/ texts

Options for Using Local Models



LM Studio



GPT4All



Ollama

A Demonstration

Uploading Documents and Prompt

You



Bazerman2004ChapIntertextuality.pdf



AdamJ.Banks_2011_2.MixRolesRelationshi_DigitalGriot... .pdf



Karenga - Nommo, Kawaida, and Communicative Practi... .pdf



retrievalAugmentedGeneration.pdf



Clarke - 2004 - Talk About Talk Promises, Risks, and apdf

You are a writing tutor, trained to ask professionals Socratic questions about their ideas. I've uploaded documents on intertextuality, digital griots, and nommo. I've included some notes about my book chapter related to these ideas. Review my notes and compare them to the ideas uploaded in the chat:

Why use GenAI? There's something about the disbelief of the rhetorical conditions that makes GenAI attractive? The rhetorical conditions or ecologies shape the decision to engage in generating word. I think the issue isn't the reader – it's the writer themselves this chapter might focus on. Like we have to think about the status of writing in the 21st Century. As Deb Brandt shows the status of writing is about the economy, about getting things done. The author of AI and the neoliberal university makes the same argument, that even the intellectual property we develop can be exploited by universities – as patents, inventions, software – but also we ourselves as professors even in the humanities can make ideas for capitalist ends, such as getting paid speaking engagements. The rhetorical condition of writing is an economic condition. But the rhetorical condition is awash in the need to produce so much content so quickly.

Referring to Documents for Response

llama-3.2-3b-instruct

- ✓ Chosen context injection strategy: 'retrieval'. Retrieval is optimal for the size of content provided
- ✓ Retrieved 3 relevant citations for user query
 - ✓ Processed Bazerman2004ChapIntertextuality.pdf for retrieval
 - ✓ Processed AdamJ.Banks_2011_2.MixRolesRelationshi_DigitalGriotsAfricanA.pdf for retrieval
 - ✓ Processed Karenga - Nommo, Kawaida, and Communicative Practice Bringi.pdf for retrieval
 - ✓ Processed retrievalAugmentedGeneration.pdf for retrieval
 - ✓ Processed Clarke - 2004 - Talk About Talk Promises, Risks, and a Propositio.pdf for retrieval
- ✓ Done predicting

I'd like to clarify a few points about your ideas and explore some connections with the concepts we've discussed so far.

You mention that GenAI is attractive because it disrupts the rhetorical conditions and forces writers to reconsider their approach. I find that interesting, and it reminds me of the idea of "nommo" as a creative power. How do you think nommo, as a concept, can be applied to the context of GenAI and the need for rapid content production?

You also mention Deb Brandt's argument about the economic status of writing in the 21st Century. This resonates with our discussion around nommo and its potential to account for AAL in terms of creativity alone. How do you think the idea of nommo can help us understand the relationship between writing, creativity, and economic conditions?

Lastly, you seem to be highlighting the tension between the need for rapid content production (driven by GenAI and other factors) and the importance of economic consideration in writing. This tension is reminiscent of the concept of "intertextuality" and how it can be used to analyze the ways in which texts engage with, build upon, or subvert each other. Can you elaborate on how you see the tension between these two forces playing out in your own writing and research?

📄 Citations



🔄 Reload last used model (#R)

Noticing the Gap in My Prompt

You

Those are great questions. I think I left out one idea: after identifying the problem with rapidly creating text using generative artificial intelligence in the knowledge economy. My solution is that we can use retrieval augmented generation (RAG) to do intertextual analysis.

llama-3.2-3b-instruct

✓ Retrieved 3 relevant citations for user query

✓ Done predicting

That's a fascinating connection. The idea of retrieval augmented generation (RAG) as a means to perform intertextual analysis is indeed related to our previous discussion.

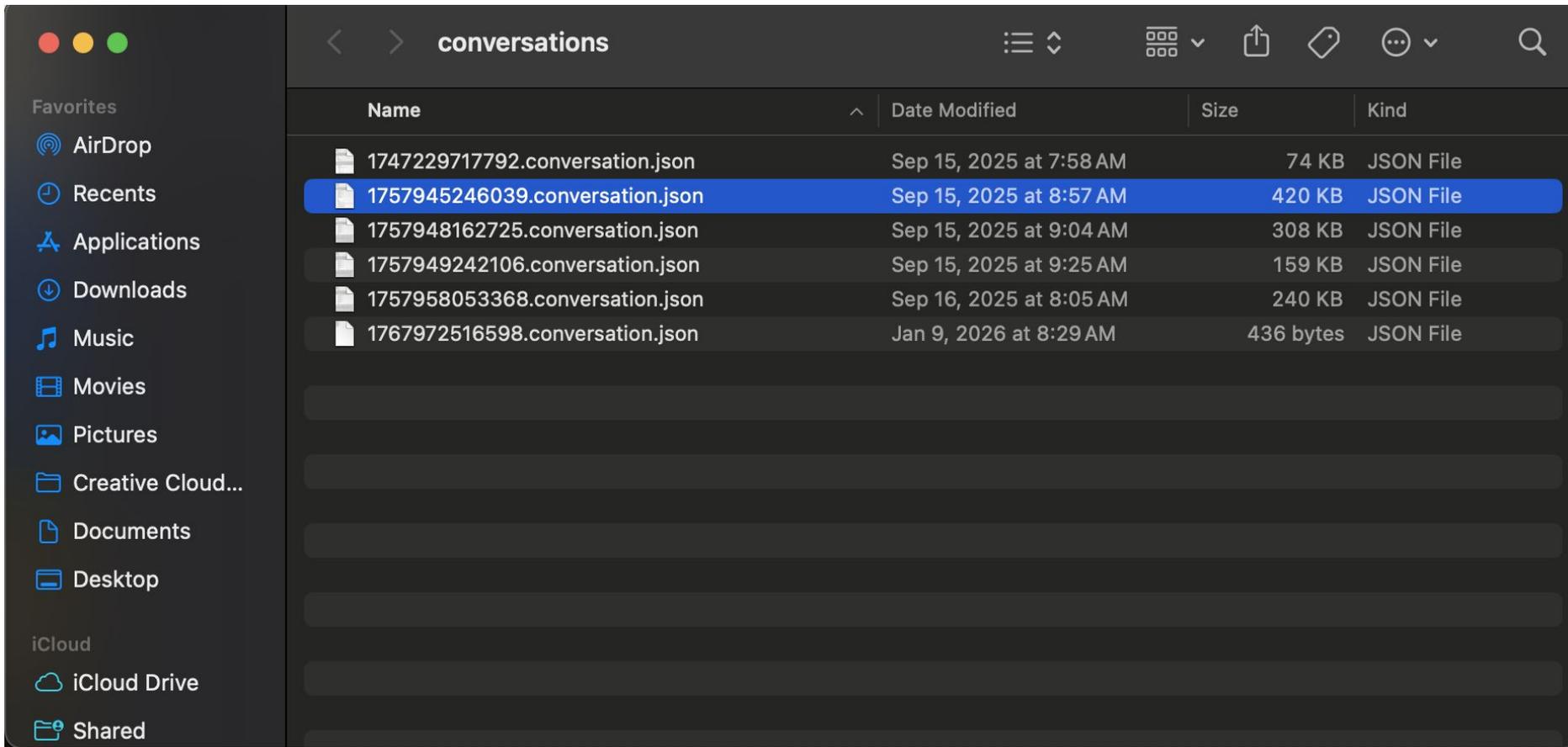
In the context of RAG, it's interesting that you mention intertextuality. The authors in Citation 2 explore a general-purpose fine-tuning recipe for RAG models, which combines pre-trained parametric and non-parametric memory. This is reminiscent of our discussion around GenAI and how it can disrupt the rhetorical conditions, forcing writers to reconsider their approach.

The fact that RAG models can access and precisely manipulate knowledge, while also providing provenance for their decisions, is crucial. It brings up the idea of nommo as a creative power, which we discussed earlier.

One potential application of RAG in the context of intertextual analysis is to help analyze and understand how texts engage with, build upon, or subvert each other. The authors in Citation 2 fine-tune and evaluate their models on a wide range of knowledge-intensive NLP tasks, which could be seen as analogous to analyzing the intertextuality in a text.

However, it's also worth noting that RAG models can be used to generate abuse or misleading content, which raises concerns

Locally Saved Data



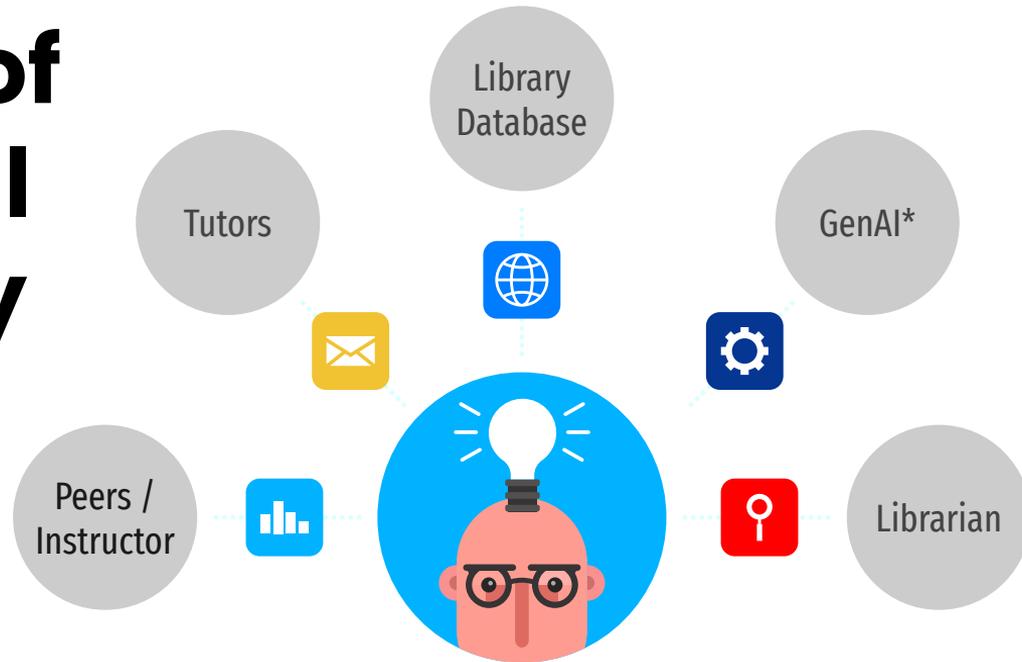
The screenshot shows a macOS Finder window with a dark theme. The window title is 'conversations'. The left sidebar shows the 'Favorites' section with items: AirDrop, Recents, Applications, Downloads, Music, Movies, Pictures, Creative Cloud..., Documents, Desktop, iCloud Drive, and Shared. The main pane displays a table of files:

Name	Date Modified	Size	Kind
1747229717792.conversation.json	Sep 15, 2025 at 7:58 AM	74 KB	JSON File
1757945246039.conversation.json	Sep 15, 2025 at 8:57 AM	420 KB	JSON File
1757948162725.conversation.json	Sep 15, 2025 at 9:04 AM	308 KB	JSON File
1757949242106.conversation.json	Sep 15, 2025 at 9:25 AM	159 KB	JSON File
1757958053368.conversation.json	Sep 16, 2025 at 8:05 AM	240 KB	JSON File
1767972516598.conversation.json	Jan 9, 2026 at 8:29 AM	436 bytes	JSON File

The Tradeoffs: Still a Mid Technology

- Need powerful computing
- Significant hard drive space
- Reasoning performance not as good
- May not monitor malicious prompting

Multiple Points of Critical Inquiry



5

What Does It Mean to Feel Data?

What Happens When the Local Are Belong to Us?

- Instead of accepting corporate narratives about GenAI what narratives can we write for ourselves in higher education?
- What happens when our narratives focus on community, joy, and human intelligence instead of surveillance and hype?
- And when joy frames our cultures, what does GenAI add to that culture, if at all?
- How do we leverage local first mindsets for technologies that represents our values of community, critical inquiry, and problem solving?

Power and Public Values

