2023-09-28 AIM TF Meeting Notes

Meeting Date & Time

- Next meeting 26 Oct 2023
  - **09:00-10:00 PT / 16:00-17:00 UTC**

Zoom Meeting Links / Recordings

Meeting: https://zoom.us/j/98931559152?pwd=d0ZwM1JHQ3d5cXRqVTh4NlRHeVJvQT09

Recording: No recording

Attendees

- Wenjing Chu X
- Anita Rao X
- Jacob Yunger
- Neil Thomson
- @Alex Khachaturian X
- Mary Lacity
- Judith Fleenor X
- Steven Milstein
- Savita Farooqui
- Daniel Bachenheimer

Main Goal of this Meeting

This is the AIM TF's #28 meeting.

One of our main goals is to have individual member presentations on what problems/challenges they see in AI & Metaverse related to trust.

Starting in the new year (2023), we plan to start drafting white papers or other types of deliverables of the task force.

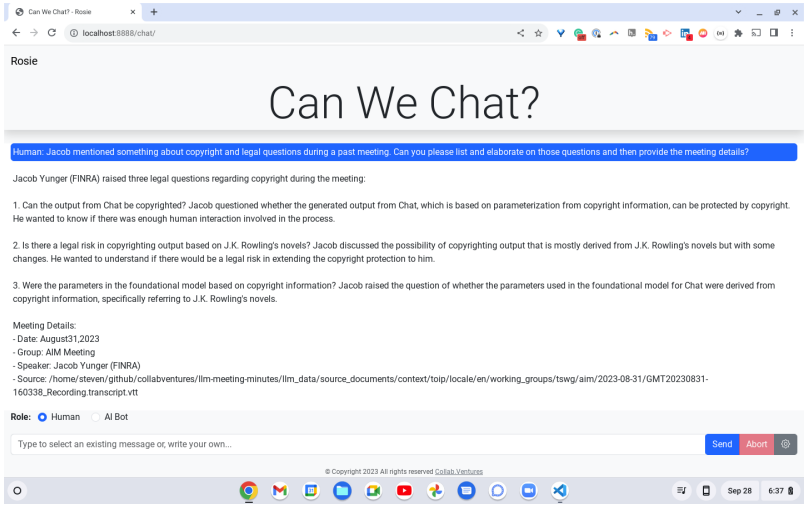
Agenda Items and Notes (including all relevant links)

Time	Agenda Item	Lead	Notes
5 min	• Start recording • Welcome & antitrust notice • Introduction of new members • Agenda review	Chairs	• **Antitrust Policy Notice:** *Attendees are reminded to adhere to the meeting agenda and not participate in activities prohibited under antitrust and competition laws.* • ***ToIP Policy**: Only members of ToIP who have signed the necessary agreements are permitted to participate in this activity beyond an observer role.* • ***ToIP TSWG IPR Policy:** see TF wiki page. AI & Metaverse Technology Task Force*
10 mins	• Introduction of new members • Any general announcements, news, that could be of interest to the TF	All	• Copyright lawsuit • ChatGPT new release

	US Copyright Office NoI	All	**Planned:**
			
			• Topic: draft a response to the US Copyright Office's Notice of Inquiry RFC
			  ◦ The US Copyright Office Notice of Inquiry (August 30, 2023) (Also: https://www.copyright.gov/ai/)
			• The DRAFT is here: https://docs.google.com/document/d/1TjYoR4ICVusnLYOKUeJThKSNiHB1hAv6Lu57UY_LqxA/edit
			• Discussions
			  ◦ Select a few questions to answer - focus on where we intersect with our direct concerns and/or expertise.
			  ◦ Review draft responses
			
			In Meeting: There was minimal discussion on extending existing comments in the response draft, and no one offered additional written contributions beyond the comments in the draft. General consensus is that there was insufficient understanding of the issues and insufficient time (lots of conferences and other commitments) to produce a comprehensive set of comments by the Oct 18 deadline.
			
			However, there was agreement that these are important issues to understand, most of which are beyond just the copyright aspects of AI.
			
			It is suggested that:
			
			1. Package comments to date and submit for the Oct 18 deadline.
			2. Put issues raised in the document (and expand on it) for future discussions on copyright and other AI topics and potentially write a separate paper (or equivalent) as input to the US Copyright Office at some later date.

	General Comments on AI, Copyright	All	The following topics/issues came up in discussions:
			
			• Defining data to be used in answering a given prompt/query (whether at training time or providing post-training supplemental data) is currently inadequate.
			  ◦ Controls on the training process may/are unable to avoid “hallucinating” while training w data.
			  ◦ Building a reliable question environment (for a specific topic) is not deterministic/reliable as the data scoping and combination rules are weak.
			• Attribution/citing of sources is unreliable and (due to the tech) may be inherently lost in the training model process (of data/sources). This undermines the concept of Authenticity (authentic data), which undermines trust.
			  ◦ Only authentic data sources (ACDC type provenance) will be reliable.
			  ◦ Limiting data used in a prompt/query to specific sources and provenance/traceability to data (sources) explicitly used in the answer are required for verifiability.
			• Reasoning (workflow, processing) for an answer is highly desirable for trust/verifiability.
			• Controls on the accuracy of answers are inadequate (“don’t hallucinate”, “be truthful”, “tell me if you don’t have an answer,” as explicit instructions are crude)
			• The question of copyright at the current state of GenAI is likely moot as answers may not be independently verifiable (or repeatable), and determining copy/unfair use may not be knowable beyond straight tests for text copying.
			
			Details
			
			Mary Lacity's experience with ChatGPT, including asking for sources:
			
			• Does not remember any corrections to sources
			• Invents fictitious sources which cannot be found with alternate search tools
			• The Current tool does not remember past sessions without turning on history options
			
			Mary Lacity's experience with the Copyright Office: if you want to influence them, then you need to have direct contact with staff within the Copyright Office (vs submissions)
			
			Savita Farooqui’s experience (screenshots) is working with organizations (Savita details?); they have found that ChatGPT is not reliable for accessing and interpreting data, which has driven the need for best practices on both providing data to LLM models and on “crafting” of prompts/queries.
			
			Steven Milstein - being able to control the processing of data into Vector DBs, including adding/extending metadata, is highly desirable (to control data use/training better). See the screenshot below as an example. The context (e.g., who, what, when, where, why, how) for the data is an example of metadata. It does have an advantage over (non-AI) Google searches in that you have some control over directing the use of data and how to derive answers, which includes being able to direct ChatGPT to self-check answers (including recursively).
			
			Discussion on attribution (which applies to humans as well) is that attribution is pragmatic on the main influences for derived work (top 1-5) vs. every book, paper, article, etc., that was used, read or reviewed.
			
			It was also suggested that any LLM should have a set of test suites, including repeatability tests for continually updated datasets. Test results would have to be verifiable by separate approaches, which are essential to verifiability and trust with respect to the data and resulting LLM model.
			
			Steven and Savita discussed self-checking and verification
			
			Jacob Yunger - how do you direct (what is best practice) to ensure a useful result
			
			Neil Thomson - I’m unclear why this technology does not have a clear body of knowledge, provided by the developers, as to how to prepare and control data input/training, plus the “syntax”, structure and configuration controls that are compatible with the internal design. I find it odd that people expect success from using trial/error to reverse engineer how to get useful, trustable answers.
			
			At least one lawsuit claims a violation of “fair use” (including Game of Thrones author RR. Martin). Question: would the objection(s) be dropped if GenAI attributed the source of derivative work to those author’s works?
			
			General consensus - to be useful/trustable, GenAI needs to emulate academic researchers in providing citations to data and answers (and the data they are based on) to support verification through alternate data sources/processing models.
			
			LLM training means that, in most cases, the data used is not current. ChatGPT models currently offered include data training from Sep 2021
			
			Dan Backenheimer - With the exception of direct quotes (which are permitted copying with attribution), if a GenAI model/engine output is to be copyrightable, then it has to be demonstrably a derived work of one/multiple sources
			
			What rights does copyright provide?

5 mins	• Review decisions /action items • Planning for next meeting • AOB	Chairs	• Topic #2 (Probably next time): Worldcoin review of this white paper. Not addressed in this meeting



Screenshot example of trained AI searching across multiple meetings and providing metadata with its response.