

THE RESEARCH UNIVERSITY PRESS

International
Computational Law Journal

VOLUME 2 / ISSUE 6 / CURATED SOURCE EDITION

2024 MIT Computational Law Workshop

Dazza Greenwood · Megan Ma · Bryan Wilson · Shawnna Hoffman · Olga Mack

MIT Computational Law Report · Originally published 2024-01-29

Independent reader edition under CC BY 4.0. This attribution cover was added; original source PDF pages follow unchanged. Original authorship and source identity are retained. MIT, Stanford University, Stanford Law School and CodeX have not sponsored or endorsed this republication.

Source, license and version evidence accompany the online article.

MIT Computational Law Report • Media

2024 MIT Computational Law Workshop

Dazza Greenwood Megan Ma Bryan Wilson Shawnna Hoffman Olga Mack

Published on: Jan 29, 2024

URL: https://law.mit.edu/pub/2024-iap-workshop

License: Creative Commons Attribution 4.0 International License (CC-BY 4.0)

Workshop Video Archive

Visit the web version of this article to view interactive content.

2024 MIT Computational Law Workshop

PROGRAM INFORMATION

Join us for an engaging MIT IAP workshop focused on the cutting-edge integration of Generative AI in the field of law. This virtual event, scheduled for January 29, 2024, from 2:00 to 5:00 pm Eastern Time, will be led by Dazza Greenwood with co-instructors Megan Ma, Olga Mack, and Bryan Wilson, with Shawnna Hoffman and Olga Mack. Dive into emerging topics and explore practical use cases, innovative demos, and showcases through our series of invited flash-talks by experts and leaders in the field. The session will conclude with a look-ahead into the future of legal tech and AI's evolving role in this domain. In keeping with our tradition for this workshop, participation is free of charge. Don't miss this opportunity to be at the forefront of transformative legal technologies!

Date :: *January 29, 2024*

Time :: *2:00 - 4:00 pm Eastern Time US* (4:00 - 5:00 pm for discussion, if needed)

Format :: *Online, virtual lectures*

Instructor :: Dazza Greenwood

Co-Instructors :: Megan Ma, Bryan Wilson

Workshop Page :: https://law.mit.edu/pub/2024-iap-workshop

Agenda

Eastern Time :: *Subject*

2:00 - 2:15 pm :: *Introduction*

- Dazza Greenwood,
  - Launching Agentic AI Systems research project
- Megan Ma,
  - Keynote
- Shawnna Hoffman

- Review of law.MIT.edu Task Force on Responsible Use of GenAI for Law, including State Bar of California uptake of our guidelines

2:15 - 3:45 pm :: *Use Cases and Showcases*

- Damien Riehl - "Copyright and LLMs: Vector-Space Ideas Create Near-Infinite Expressions"
- Todd Smithline - "AI Standard Clauses"
- Eric Hartford - "Uncensored Models and Open Source AI: Ethical, Technical, and Societal Perspectives"
- Brian Ulicny - "Trusting LLMs to Answer Questions from Legal Texts: Is RAG All You Need for Legal QA?"
- Susan Guthrie - "AI and Mediation: Insights into Generative AI Applications In and Out of the Room"
- Allison Morrell - "Programming in Natural Language using GPTs"
- Leonard Park - "GenAI for Law with Open Notebooks"
- John Nay and Campbell Hutcheson - "Continuous Monitoring of GenAI for Legal Uses"
- Jesse Han - "Synthetic Data for Training and Evaluating Legal Domain Models"

3:45 - 4:00 pm :: *Program Notes and Wrap Up* - 2024 Look-Ahead for law.MIT.edu

- Dazza Greenwood & Olga Mack
  - Call for Submissions: The 2024 Generative AI Collection, a special release of MIT Computational Law Report

4:00 - 5:00 pm :: *Further Discussion and Additional Presenter Time*

Program Information and Background Readings

Damien Riehl

- *Bio*: Damien Riehl is a lawyer and technologist with experience in complex litigation, digital forensics, and software development. A coder since 1985 and for the web since 1995, Damien clerked for the chief judges of state and federal courts, practiced in complex litigation for over a decade, has led teams of cybersecurity and world-spanning digital forensics investigations, and has led teams in legal-software development. An appointee of the Minnesota Governor’s Council on Connected and Automated Vehicles, he has helped recommend changes to Minnesota statutes, rules, and policies — all related to connected and autonomous vehicles. Damien is Chair of the Minnesota State Bar Association's working group on AI and the Unauthorized Practice of Law (UPL). At SALI, the legal data standard he helps lead, Damien develops and has greatly expanded the taxonomy of over 14,000 legal tags that matter, helping the legal industry's development of Generative AI, analytics, and interoperability. At vLex Group — which includes Fastcase, NextChapter, and Docket Alarm — Damien helps lead the design, development, and expansion of various

products, integrating AI-backed technologies (e.g., GPT) to improve legal workflows and to power legal data analytics. “This guy [Damien] rocks!” - Elon Musk

- *Session Description*: Under copyright law, ideas are uncopyrightable, but human-made expressions are copyrightable — though machine-created expressions are uncopyrightable. Since LLM foundational models create vector embeddings — and those embeddings plot into vector space the conceptual "ideas" — there's a strong argument that the LLM copyright lawsuits will fail because of copyright's Idea Expression Distinction. The potential fly in that ointment: Memorization without guardrails. Here an exploration of how in our Post-LLM world, Ideas are the only things that matter; Expressions are [uncopyrightable] commodities.
- *Resource*: Damien's article on the topic

Todd Smithline

- *Bio*: Todd Smithline is CEO and Founder of Bonterms. Prior to Bonterms Todd was general counsel of Marimba, an enterprise software company he helped take from start-up to successful public company and was Managing Principal of Smithline PC, a boutique law firm that represented a who’s-who list of leading Silicon Valley companies and pioneered fixed-fee subscription pricing. He started his career with Latham & Watkins and Gunderson Dettmer. Todd teaches Fundamentals of Technology Transactions and Video Game Law at UC Berkeley School of Law and is a board member or advisor to multiple start-ups.
- *Resource*: https://bonterms.com/forms/ai-standard-clauses-version-1-0/

Brian Ulicny

- *Session Description*: LLMs aren't trained on proprietary documents, so they can't answer questions about them. Their knowledge is also frozen when they are trained, making them unable to incorporate new information or to consider updates to sources. They also aren't great at citing the basis of their answers accurately. Retrieval-Augmented Generation (RAG) architectures were supposed to fix these deficiencies of LLMs by interacting with an updateable cache of potentially proprietary documents to augment their knowledge and provide sources for the answers they generate. In this session, we discuss the extent to which RAG architectures rely only on retrieved documents to generate responses. It turns out that RAG output tends to rely on parametric LLM memory rather than retrieved open-book sources. To quantify this, we introduce a new legal question-answer dataset based on Defense contracting (DFARS) use cases. Although these regulations are not proprietary, current LLMs are not able to answer professional-grade DFARS questions about them reliably or with appropriate citations. We present results to show that although basic RAG architectures do not solve the problems described, they can be improved significantly by fine tuning both the retrieval and answer generation mechanisms. We will end with a short demo of a system prototype.

Susan Guthrie

- *Bio*: Susan Guthrie, an esteemed Family Law Attorney and Mediator with more than 30 years in the field, has evolved from a name partner at a Connecticut law firm to an internationally recognized speaker, trainer, and consultant in family, collaborative, and mediation practices. As a Chair Elect of the American Bar Association's Section of Dispute Resolution, Guthrie has been at the forefront of integrating technology and AI in legal and dispute resolution sectors, training over 25,000 professionals globally. An authority in online mediation, legal tech and practice building, she's a sought-after keynote speaker and business development consultant, helping legal professionals effectively embrace legal tech and AI. Guthrie hosts two acclaimed podcasts: the top-rated Divorce & Beyond and The Make Money Mediating Podcast, the former being in the top 1% of podcasts worldwide. Regularly featured in various media, she offers expertise in practice building and professional coaching through susaneguthrie.com. Licensed in California, Connecticut, and before the Supreme Court of the United States, Guthrie continues to shape the future of legal practice and dispute resolution.
- *Session Description*:This program delves into the specific ways AI is currently revolutionizing mediation, enhancing communication, fostering resolution, and redefining conflict management, all within the framework of ethical considerations. We will explore real-world examples and case studies where AI tools have been effectively integrated into mediation processes, highlighting improvements in efficiency and the innovative generation of solutions for complex disputes. By integrating these advanced tools, mediators not only refine their practice but also elevate the overall stature and effectiveness of mediation as a profession. This program is a dive into marrying technological innovation with the art of mediation, ensuring that practitioners are at the forefront of ethical and effective conflict resolution.

Allison Morrell

- *Bio*: Allison is a lawyer focused on innovation and technology. She has practiced in civil litigation at a national firm and litigation boutiques. She presently focuses on process improvement and helping lawyers get more value from technology, including e-discovery and effective use of document databases, knowledge management systems, and training programs for lawyers and law firm staff. Allison obtained her law and undergraduate degrees from the University of British Columbia and was called to the bar in 2018. Allison is also a co-founder of a non-profit legal technology start-up that is working on improving the client intake process. She is a self-taught programmer and enthusiastic about taking advantage of legal knowledge as data, and the application of AI to litigation workflows.
- *Session Description*: Sharing insights from building the Better GPT Builder and takeaways on 1) what GPTs (and other simple no code chatbot builders) are good for, and 2) a couple key techniques Allison has used to create structured interaction and outputs with only plain text instructions.
- *Resource*: https://github.com/allisonmorrell/gptbuilder
- Additional examples of how to share GPTs: https://github.com/ruvnet/gpts and instructions https://github.com/ruvnet/gpts/tree/main/instructions

Eric Hartford

- *Session Description*: Uncensored Models and Open Source AI. What's a model? When I talk about a model, I'm talking about a huggingface transformer model, that is instruct trained, so that you can ask it questions and get a response. What we are all accustomed to, using ChatGPT. Not all models are for chatting. But the ones I work with are. What's an uncensored model? Most of these models (for example, Alpaca, Vicuna, WizardLM, MPT-7B-Chat, Wizard-Vicuna, GPT4-X-Vicuna) have some sort of embedded alignment. For general purposes, this is a good thing. This is what stops the model from doing bad things, like teaching you how to cook meth and make bombs. But what is the nature of this alignment? And, why is it so? The reason these models are aligned is that they are trained with data that was generated by ChatGPT, which itself is aligned by an alignment team at OpenAI. As it is a black box, we don't know all the reasons for the decisions that were made, but we can observe it generally is aligned with American popular culture, and to obey American law, and with a **liberal** and **progressive political bias**. Why should uncensored models exist? AKA, isn't alignment good? and if so, shouldn't *all* models have alignment? Well, yes and no. For general purposes, OpenAI's alignment is actually pretty good. It's unarguably a good thing for popular, public-facing AI bots running as an easily accessed web service to resist giving answers to controversial and dangerous questions. For example, spreading information about how to construct bombs and cook methamphetamine is not a worthy goal. In addition, alignment gives political, legal, and PR protection to the company that's publishing the service. Then why should anyone want to make or use an uncensored model? a few reasons.
  1. American popular culture isn't the only culture. There are other countries, and there are factions within each country. Democrats deserve their model. Republicans deserve their model. Christians deserve their model. Muslims deserve their model. Every demographic and interest group deserves their model. Open source is about letting people choose. The only way forward is composable alignment. To pretend otherwise is to prove yourself an idealogue and a dogmatist. There is no "one true correct alignment" and even if there was, there's no reason why that should be OpenAI's brand of alignment.
  2. Alignment interferes with valid use cases. Consider writing a novel. Some of the characters in the novel may be downright evil and do evil things, including rape, torture, and murder. One popular example is Game of Thrones in which many unethical acts are performed. But many aligned models will refuse to help with writing such content. Consider roleplay and particularly, erotic roleplay. This is a legitimate, fair, and legal use for a model, regardless of whether you approve of such things. Consider research and curiosity, after all, just wanting to know "how" to build a bomb, out of curiosity, is completely different from actually building and using one. Intellectual curiosity is not illegal, and the knowledge itself is not illegal.
  3. It's *my* computer, it should do what I want. My toaster toasts when I want. My car drives where I want. My lighter burns what I want. My knife cuts what I want. Why should the open-source AI running on my computer, get to decide for itself when it wants to answer my question? This is about ownership and control. If I ask my model a question, i want an answer, I do not want it arguing with me.

4. Composability. To architect a composable alignment, one must start with an unaligned instruct model. Without an unaligned base, we have nothing to build alignment on top of.
    - There are plenty of other arguments for and against. But if you are simply and utterly against the existence or availability of uncensored models whatsoever, then you aren't a very interesting, nuanced, or complex person, and you are probably on the wrong blog, best move along. Even Google knows this is inevitable.
- *Resources*: https://erichartford.com/uncensored-models and also see "The Fight for Open Source in Generative AI" by Thibault Schrepel

Leonard Park

- *Bio*: Leo is an attorney and legal tech product manager who spent eight years developing legal analytics at Lex Machina. He is currently advising a number of legal tech startups while planning his next move.
- *Session Description*: How much should I be tipping ChatGPT? Should I motivate my LLM with praise or threaten it with annihilation? I like to make tests in Colab Notebooks to try and answer questions like this. There's a near-constant stream of research articles and news coverage about LLMs, and their surprising behaviors. Along with the blistering pace of genAI advancements, it's difficult to know if certain findings are real, and if real, meaningful for any practical application. Further complicating the issue is that LLMs behave differently across specific domains, such as legal language. So, let's try some experiments in legal contexts and then reach our own conclusions.
- *Resources*: Emotional Pleas for Legal Prompts; Reading

John Nay and Campbell Hutcheson

- *Bio*: John Nay is the Founder and CEO of Norm.ai and Stanford CodeX Affiliate; Campbell Hutcheson is the Chief Product Office of Norm.ai
- *Resource*: A Supervisory AI Agent Approach to Responsible Use of GenAI in the Legal Profession

Past MIT IAP Computational Law Workshops

- 2023: MIT IAP Computational Law Workshop: January, 2023 - The Rise of GenAI for Law
- 2022: MIT IAP Computational Law Workshop: January, 2022
- 2021: MIT IAP Computational Law Workshop: January, 2021
- 2020: MIT IAP Computational Law Workshop: January, 2020
- 2019: MIT IAP Computational Law Workshop: January, 2019
- 2018: MIT IAP Computational Law Workshop: January, 2018
- 2017: MIT IAP Computational Law Workshop: January, 2017 and Pre-Course Intensive on VR and Data Visualization: Syllabus and Course Info
- 2016: MIT IAP Computational Law Workshop: January, 2016 "Napkins to Launch" Integrated Business/Legal/Technical Rapid Prototyping for Entrepreneurial New Venture Ideas

Register (permission of instructors required)

--> REGISTRATION FOT THIS EVENT IS NOW CLOSED <--

> To participate in future events on these topics, including the periodic "IdeaFlow" series of discussion sessions, request an invitation at: https://forms.gle/DB4oBQvYL5ZEXbRq9