Agentic Artificial Intelligence and the Law
Harvard Law School
Spring 2025

Instructor: Jonathan Zittrain | Jordi Weinstock | Joshua Joseph

Teaching Assistants: Brigitte Fink

COURSE OVERVIEW

The recent success of large language models has led to a surge in the development of agentic artificial intelligence systems. Agentic AI – autonomous, goal-directed systems capable of planning and executing long sequences of actions on behalf of a user – are being built for various applications and are being granted increasing discretion. Some experts believe that Agentic AI systems that can pursue goals autonomously and grasp the complexities of adaptive decision-making will be integrated into daily experiences within the next two years.

Depending on how they are designed and implemented, autonomous systems can present challenges to established structures in our legal system that are designed to protect individuals from harm caused by other actors and their human and corporate agents. This course will explore the boundaries of that conflict. Students will develop concrete policies designed to ensure that society is able to take advantage of the benefits of agentic AI while mitigating when its behavior and effects inevitably go awry.

Session 1: Intros & Overview (01/27)

No Guest

- What are AI agents?, by Melissa Heikkila, MIT Technology Review (2024)
- We Need to Control AI Agents Now, by Jonathan Zittrain, The Atlantic (2024)
- Through the Chat Window and Into the Real World: Preparing for AI Agents, by Helen Toner et al., Center for Security and Technology, CSET (2024) **[Read: executive summary]**

Session 2: What are AI Agents? (01/28)

No Guest

- "Chapter 2: The Basics and Eccentricities of Machine Learning" Gaining Power, Losing Control, Our Choices as We Confront the Next Technology Revolution, by Jonathan Zittrain, upcoming publication (2024) **[NOTE: This reading is only available in print copy, please collect from Lewis 319 prior to class]**
- Artificial Intelligence: A Modern Approach, by Stuart Russell and Peter Norvig (2010) **[Read: Chapter 1.3, "The History of Artificial Intelligence," 16-28]**
- Is it an agent, or just a program?: A taxonomy for autonomous agents, by Stan Franklin and Art Graesser, Intelligent Agents III Agent Theories, Architectures, and Languages (1996) **[Read: Sections 1-3]**
- Large Language Models explained briefly, by 3Blue1Brown (2024) **[7-minute explainer video on Large Language Models]**
    - **Optional***:* How large language models work, a visual intro to transformers | Chapter 5, Deep Learning
- Building effective agents, by Anthropic(2024)

Session 3: Agency (02/03)

No Guest

- The Routledge Handbook of Philosophy of Agency, edited by Luca Ferrero (2022) **[Read: sections 3 and 5 of the introduction]**
- Adopting the intentional stance toward natural and artificial agents, by Jairo Perez-Osorio & Agnieszka Wykowska, Philosophical Psychology Vol.33 (2020) **[Read: sections 1-3, and 7]**
- Agency Law Primer, by Holger Spamann & Jens Frankenreiter, Corporations (2023)
- The Law of AI is the Law of Risky Agents without Intentions, by Ian Ayres & Jack Balkin, Yale Law School, Public Law Research Paper (2024)

Session 4: AI Agents for Healthcare (02/04)

Guest Speaker: Maulik Shah

Hippocratic AI, General Counsel

Maulik Shah is an engineer turned lawyer with deep expertise in copyright, intellectual property, and data privacy. Advisor to large and small technology companies, including surgical robotics. Former litigator for Fortune 50 companies. He received a B.S. in Mechanical Engineering from Yale University, an M.S. in Aerospace Engineering from Stanford University, and a J.D. from Harvard Law School.

- Hippocratic AI, by Hippocratic AI (2024) **[Please explore the following links]**
  - Hippocratic AI's homepage
  - Read page on Hippocratic's Foundation Model
  - Watch video here
- 'I will never go back': Ontario family doctor says new AI notetaking saved her job, by Colin D'Mello & Isaac Callan, Global News (2024)
- *"Section 1, Chapter 3: The Suicide Meter" Gaining Power, Losing Control, Our Choices as We Confront the Next Technology Revolution*, by Jonathan Zittrain, upcoming publication (2024) **[NOTE: This reading is only available in print copy, please collect from Lewis 319 prior to class]**
- Quest to develop fully autonomous surgical robot attracts award up to $12 million from ARPA-H, by School of Engineering, Vanderbilt University (2024) **[Read: post and watch video]**

Session 5: AI Agents for Law (02/10)

Guest Speaker: Scott Stevenson

Spellbook AI, CEO

Scott Stevenson is the CEO and Co-Founder of Spellbook. With a background in software engineering, he has been building technology products for over 15 years. Scott was inspired to found Spellbook after being unsatisfied with the legal work's cost and turnaround time in a previous venture.

- Spellbook Associate, by Scott Stevenson, Spellbook.legal (2024) **[Read: blog post and watch demo video]**
- AI, by Lawrence Lessig (2024) **[NOTE: only available on Canvas, read p. 6-12, "HLS Innovator's Dilemma"]**
- Promises and pitfalls of artificial intelligence for legal applications, by Sayash Kapoor, Arvind Narayanan, and Peter Henderson, Center for Information Technology Policy, Princeton University (2024)
- AI on Trial: Legal Models Hallucinate in 1 out of 6 (or More) Benchmarking Queries, by Varun Magesh, Faiz Surani, Matthew Dahl, Mirac Suzgun, Christopher D. Manning, Daniel E. Ho, Human-Centered Artificial Intelligence, Stanford University (2024)
- **Please set up CoCounsel AI-Assisted Research within your personal Westlaw account prior to class.**

Session 6: Personal AI Agents (02/11)

Guest Speaker: Omar Shaya
Please, Founder
Omar Shaya is the founder of Please (previously MultiOn), an AI company that develops autonomous agents. Prior to this, he led product management for Core Product Ranking at Meta and worked on product management in Microsoft's Search, Assistant, and Intelligence (MSAI) division. Additionally, Omar has worked as a product manager at Crealytics and LiquidM, two successful startups where he helped incubate and develop AI-based marketing automation platforms.

- About, Please, by Omar Shaya, Please(2024) **[Explore website and watch video compilation]**
- Claude | Computer use for automating operations, by Anthropic (2024) **[video]**
- System Prompts, by Anthropic (2024) **[Read: Claude 3.5 Sonnet's system prompt, November 22nd version and July 12th version; Claude 3 Haiku's July 12th system prompt]**
- I Took a 'Decision Holiday' and Put A.I. in Charge of My Life, by Kashmir Hill, New York Times (2024)
- Can a Chatbot Named Daenerys Targaryen Be Blamed for a Teen's Suicide?, By Kevin Roose, New York Times (2024)
- The ethics of advanced AI assistants, Iason Gabriel & Arianna Manzini, Google Deepmind (2024) **[Read: 1.2 "Key Questions" and 20.1 "Key Themes and Insights"]**

Session 7: Scaling and Self-improvement (02/24)

No Guest

- 4 Charts That Show Why AI Progress Is Unlikely to Slow Down, by Will Henshall, TIME (2023)
- The "most important century" blog post series, by Holden Karnofsky, cold-takes (2021)
- The way we measure progress in AI is terrible, by Scott J Mulligan, MIT Technology Review (2024)
- Is AI progress slowing down?, by Arvind Narayanan & Sayash Kapoor, AI Snake Oil (2024)
- Levels of AGI for Operationalizing Progress on the Path to AGI by Meredith Ringel Morris, et al., Arxiv (2023) **[Read: Chapters 2, 4, and 6]**
- Automated Design of Agentic Systems**,** by Shengran Hu, Cong Lu & Jeff Clune, Arxiv (2024) **[Read: Sections 1-3]**

- The Rogue Replication Threat Model, by Josh Clymer, Hjalmar Wijk, & Beth Barnes, METR (2024)

Session 8: Torts and Products Liability (02/25)

No Guest

- Tort Law and Frontier AI Governance, by Matthew van der Merwe, Ketan Ramakrishnan, & Markus Anderljung, Lawfare (2024)
- Tort Law Should Be the Centerpiece of AI Governance, by Gabriel Weil, Lawfare (2024)
- It's Time for Courts to Tackle AI Harms with Product Liability – EPIC, by Grant Fergusson & Maria Villegas Bravo, Electronic Privacy Information Center (2024)
- U.S. Tort Liability for Large-Scale Artificial Intelligence Damages: A Primer for Developers and Policymakers, by Ketan Ramakrishnan, Gregory Smith & Conor Downey, RAND (2024) **[Read: summary page, skim paper]**
- Innovating Liability: The Virtuous Cycle of Torts, Technology and Liability Insurance, by Anat Lior, 25 Yale J.L. & TECH. 448 (2023)

Session 9: Play w/ Agents & Paper Brainstorming (03/03)

No Guest

This day will incorporate hands-on interactions with AI. There will be more information for this session given during the semester.

Session 10: Open Source Models and Agents (03/04)

No Guest

- The Gradient of Generative AI Release: Methods and Considerations, by Irene Solaiman, arxiv (2023) **[Read: Abstract, Sections 3, 4, 5 (plus the three figures in the paper)]**
- Legal primer on open genAI models, by Ken D. Kumayama & Pramode Chiruvolu, Reuters (2024)
- There Was Never Such a Thing as 'Open' AI, by Matteo Wong, The Atlantic (2024)
- Open Source AI Is the Path Forward, by Mark Zuckerberg, Meta (2024)
- Mark Zuckerberg gave Meta's Llama team the OK to train on copyrighted works, filing claims, by Kyle Wiggers, TechCrunch (2025)

Session 11: SB1047 - California Regulation (03/10)

Guest Speaker: Irene Solaiman

Hugging Face, Head of Global Policy

Irene Solaiman is an AI safety and policy expert. She is Head of Global Policy at Hugging Face, where she is conducting social impact research and leading public policy. Irene serves on the Partnership on AI's Policy Steering Committee and the Center for Democracy and Technology's AI Governance Lab Advisory Committee. Irene advises responsible AI initiatives at the OECD and IEEE. Her research includes AI value alignment, responsible releases, and combating misuse and malicious use. Irene was named MIT Tech Review's 35 Innovators Under 35 for her research.

- When the tech boys start asking for new regulations, you know something's up, by John Naughton, The Guardian (2023)
- SB 1047: Safe and Secure Innovation for Frontier Artificial Intelligence Models Act, Bill Author Scott Wiener (D-D11), Digital Democracy Cal Matters (2024)
- **Letters of response/commentary on SB1047:**
    - Letter to CA state leadership from Professors Bengio, Hinton, Lessig, & Russell, by Professors Bengio, Hinton, Lessig, & Russell, Safe & Secure AI Innovation (2024)
    - A Right to Warn about Advanced Artificial Intelligence, by current and former employees at frontier AI companies, Right To Warn AI (2024)
    - Letter to Governor Newsom RE: SB 1047 (Wiener) from Dario Amodei, by Dario Amodei, Anthropic (2024)
    - Big Tech Is Very Afraid of a Very Modest AI Safety Bill, by Larry Lessig, The Nation (2024)
    - Letter to Senator Wiener RE: SB 1047 from Andreessen Horowitz "a16z", by Jaikumar Ramaswamy, Andreessen Horowitz (2024)
    - RE: Response to inaccurate, inflammatory statements by Y Combinator & a16z regarding Senate Bill 1047, by Senator Wiener, CA Gov (2024)
        - **Optional**: SB 1047 August 15 Author Amendments Overview by Nathan Calvin, Center for AI Safety Action Fund (2024)
- SB1047 Veto Message by Governor Gavin Newsom, CA Gov (2024)

Session 12: AI Agent Safety and Societal-scale Risks (03/11)

Guest Speaker: Dan Hendrycks

Safe.ai, Director

Dan Hendrycks received his PhD from UC Berkeley where he was advised by Dawn Song and Jacob Steinhardt. He is now the director of the Center for AI Safety. He is interested in AI Safety. His research is supported by the NSF GRFP and the Open Philanthropy AI Fellowship. He helped contribute the GELU activation function (the most-used activation in state-of-the-art models including BERT, GPT, Vision Transformers, etc.), the out-of-distribution detection baseline, and distribution shift benchmarks.

w

- When A.I. Passes This Test, Look Out, by Kevin Roose, New York Times (2025)
- Managing extreme AI risks amid rapid progress, by Bengio, Hinton et al., Science (2024)
- An Overview of Catastrophic AI Risks, by Center for AI Safety (2024) **[Live blog post recommended]**
- Safetywashing: Do AI Safety Benchmarks Actually Measure Safety Progress?**,** by Ren, Basart, et al., Arxiv (2024) **[Read: Chapters 1, 4-8]**
- Safety Cases: How to Justify the Safety of Advanced AI Systems**,** by Joshua Clymer, et al., Arxiv (2024) **[Read: Executive Summary]**

Session 13: International-Level Regulation (03/24)

Guest Speaker: Gabriele Mazzini

MIT Media Lab

Gabriele Mazzini is the architect and lead author of the EU Artificial Intelligence Act, Gabriele Mazzini is a pioneer and world-renowned expert in AI governance and regulation. In his capacity as Team Leader at the European Commission until July 2024, he designed and led the drafting of the Commission EU AI Act and acted as the principal advisor during the legislative negotiations with the Parliament and the Council. He also shaped earlier policy work of the Commission on the European approach to AI since 2017, including the White Paper on the ecosystem of excellence and trust for AI and the work on liability for emerging technologies.

- High-level summary of the EU AI Act, by Future of Life Institute, EU Artificial Intelligence Act (2024)
- The Proposal for the Artificial Intelligence Act: Considerations around Some Key Concepts, by Gabriele Mazzini, in Camardi (a cura di), La via europea per l'Intelligenza artificiale (2023)

- Ensuring AI Accountability Through Product Liability: The EU Approach and Why American Businesses Should Care, by Jana S. Farmer & Thomas M. DeMicco, The National Law Review (2024)
- Long awaited EU AI Act becomes law after publication in the EU's Official Journal, by Tim Hickman, Dr. Sylvia Lorenz, Dr. Constantin Teetzmann, Aishwarya Jha, White & Case LLP (2024) **[Read: implementation timeline]**
- Four things to know about China's new AI rules in 2024, by Zeyi Yangarchive, MIT Technology Review (2024) **[Skim]**
- One Year Later, How Has the AI Executive Order Delivered on its Promises?, by Aaron Klein, Cameron F. Kerry, Courtney C. Radsch, Mark MacCarthy, Sorelle Friedler, and Nicol Turner Lee, Brookings (2024) **[Skim]**

Session 14: Personhood & Remedies (03/25)

Guest Speaker: James Boyle
Duke Law School, Professor
James Boyle is William Neal Reynolds Professor of Law at Duke Law School and founder of the Center for the Study of the Public Domain. Boyle's new book, The Line: AI and the Future of Personhood, will be published by MIT Press on Oct 22 2024 under a Creative Commons License. Per Boyle: "The book is a labor of (mainly) love — together with the familiar accompanying authorial side-dishes: excited discovery, frustration, writing block, self-loathing, epiphany, and massive societal change that means you have to rewrite everything. So just the usual stuff. It is not a run-of-the-mill academic discussion, though."

Boyle spends as much time on art and constitutional law as he does on ethics, treats movies and books and the heated debates about corporate personality as seriously as he does the abstract philosophy of personhood. These are the cultural materials with which we will build our new conceptions of personhood, elaborate our fears and our empathy, stress our commonalities and our differences.

- The Line: AI and the Future of Personhood, by James Boyle, MIT Press (2024) **[Introduction 1-32 (32 pages) and conclusion 235-275 (40 pages)]**
- Is AI on Its Way to Gaining Rights?, by James Boyle, MIT Press (2024)

Session 15: Standards Bodies (03/31)

Guest Speaker: Joseph Lorenzo Hall
ISOC, Distinguished Technologist

At the Internet Society, Joseph Lorenzo Hall leads the programs focused on online trust and safety and an open and trustworthy Internet, which aim to defend the Internet against decisions that weaken online security, and advocate for policy, technology, and commercial decisions that put people's safety, security, and privacy first. Prior to joining the Internet Society in 2019, he was the chief technologist and director of the Internet architecture project at the Center for Democracy & Technology. Before that, he was an academic, completing postdoctoral research with Helen Nissenbaum at New York University, Ed Felten at Princeton University, and Deirdre Mulligan at University of California, Berkeley. Hall received hisPh.D. in information systems from the UC Berkeley School of Information in 2008 with a thesis on electronic voting as a critical case study in digital government transparency.

- Global Standard Setting in Internet Governance, by Alison Harcourt et al., (2020) **[Read: Introduction, 1-14]**
- Enabling AI governance and innovation through standards, by Christopher Thomas & Dr Florian Ostmann, UNESCO (2024)
- NIST's Responsibilities Under the October 30, 2023 Executive Order, by National Institute of Standards and Technology(2024) **[Skim website]**

Session 16: Robotic Agents & Military Technologies (04/01)

Guest Speaker: Adam Bry

Skydio, Co-Founder & CEO

Adam Bry is co-founder and CEO at Skydio, the leading consumer and commercial U.S. drone company, and the world leader in autonomous flight. He has over two decades of experience with small UAS, starting at 16 when he was a national champion R/C airplane aerobatics pilot. As a grad student at MIT, he helped lead an award-winning research program that pioneered autonomous flight for drones, transferring much of what he learned as an R/C pilot into software that enables drones to fly themselves. After graduate school, Adam was part of Google[x]'s Project Wing's original team. He has co-authored numerous technical papers and patents, and was recognized on MIT's TR35 list for young innovators. In 2021, Adam was appointed to the FAA's Drone Advisory Committee.

- Patrol Overwatch For Every Officer with Autonomous Drones, by Skydio (2024) **[Watch short video]**
- 75% Faster Bridge Inspections with Autonomous Drones, by Skydio (2024) **[Watch short video]**
- FAA Issues Revolutionary Approval to NYPD to conduct Drone as First Responder operations with no Visual Observers, by Jakee Stoltz, Skydio (2024)
- Skydio is pivoting to enterprise — its consumer drones are dead, by Emma Roth & Sean Hollister, The Verge (2023)

- Many Americans have come to rely on Chinese-made drones. Now lawmakers want to ban them, by Didi Tang, AP News (2024)
- China's Sanctions on Skydio, by Adam Bry, Skydio (2024)
- U.S. Weighs Ban on Chinese Drones, Citing National Security Concerns, by Ana Swanson, New York Times (2025)

Session 17: Corporate Governance & Federal Oversight, Part 1 (04/07)

Guest Speaker: Helen Toner

CSET, Director of Strategy and Foundational Research Grants

Helen Toner is the Director of Strategy and Foundational Research Grants at Georgetown's Center for Security and Emerging Technology (CSET). She previously worked as a Senior Research Analyst at Open Philanthropy, where she advised policymakers and grantmakers on AI policy and strategy. Between working at Open Philanthropy and joining CSET, Toner lived in Beijing, studying the Chinese AI ecosystem as a Research Affiliate of Oxford University's Center for the Governance of AI.

Toner has written for Foreign Affairs and other outlets on the national security implications of AI and machine learning for China and the United States, as well as testifying before the U.S.-China Economic and Security Review Commission. Helen holds an MA in Security Studies from Georgetown, as well as a BSc in Chemical Engineering and a Diploma in Languages from the University of Melbourne.

- AI Is Testing the Limits of Corporate Governance, by Roberto Tallarita, Harvard Business Review (2023)
- Senate hearing: Oversight of AI: Insiders' Perspectives, by Helen Toner, U.S. Senate Committee on the Judiciary, Subcommittee on Privacy, Technology, and the Law (2024)
- Inside OpenAI's Crisis Over the Future of Artificial Intelligence, by Tripp Mickle, Cade Metz, Mike Isaac, and Karen Weise, New York Times (2023)

Session 18: Corporate Governance & Federal Oversight, Part 2 (04/08)

Guest Speaker: Natasha Crampton

Microsoft, Chief Responsible AI Officer

Natasha Crampton leads Microsoft's Office of Responsible AI, as the company's first Chief Responsible AI Officer. The Office of Responsible AI puts Microsoft's AI principles into practice by defining, enabling, and governing the company's approach to responsible AI. The Office of Responsible AI also collaborates with stakeholders within and outside the company to shape new laws, norms, and standards to help ensure that the promise of AI technology is realized for the benefit of all.

Prior to this role, Crampton served as lead counsel to the Aether Committee, Microsoft's advisory committee on responsible AI. Crampton also spent seven years in Microsoft's Australian and New Zealand subsidiaries helping Microsoft's highly regulated customers move to the cloud. Prior to Microsoft, Crampton worked in law firms in Australia and New Zealand, specializing in copyright, privacy, and internet safety and security issues. Crampton graduated from the University of Auckland in New Zealand with a Bachelor of Laws (Honours) and a Bachelor of Commerce majoring in Information Systems.

- Responsible AI Standard, by Microsoft (2022) **[Skim]**
- Progress with our AI commitments: an update ahead of the UK AI Safety Summit, by Microsoft (2023)
- Anthropic's Responsible Scaling Policy, by Anthropic (2023)