

# Deep Dive, a 201 AI policy course

*Prepared by Rafael Andersson Lipcsey & Kambar Orazbekov*

**Course Description:** *Deep Dive* is an advanced curriculum for those who have a solid foundation in AI governance (e.g. completion of an introductory course such as BlueDot AI Governance 101 or equivalent).

This course welcomes participants from all backgrounds – various ages, professions, and cultures. Its aim is to deepen understanding of AI governance by exploring cutting-edge frameworks, historical analogies, and best practices from related policy domains.

By the end, participants will have found their footing in the vast AI governance landscape, having engaged with legal, domestic, international, economic and other critical dimensions of AI policy. Each session combines targeted readings with discussion prompts or exercises to maximize learning of key content while optimizing reading time.

## Objectives

By the conclusion of this 9(+3)-week course, participants will:

1. **Evaluate** at least two distinct global governance frameworks for AI policy by comparing their strengths and weaknesses using provided case studies.
2. **Develop** a policy brief or comparative case study (8–20 pages) articulating a clear, well-supported position on an AI governance issue, demonstrating synthesis of economic, technical, geopolitical, and legal perspectives.
3. **Identify and analyze** three historical policy analogies (e.g., nuclear arms control, internet regulation) to derive applicable lessons for contemporary AI governance, providing concrete examples in discussions and written exercises.
4. **Critically appraise** proposed AI regulations, clearly articulating at least two potential institutional challenges and two viable multilateral or domestic solutions in class discussions

or written assignments.

5. **Demonstrate proficiency** in articulating and defending positions on controversial AI governance issues (e.g., AI alignment, power concentration, regulatory trade-offs) through structured debates or discussions in class.

## Session 1: Introduction – AI Governance Landscape and Key Themes

*Learning Objective:* Set the stage for the course by reviewing foundational concepts from AI Governance 101 and introducing the major themes and challenges that will be explored. Participants will also get to know each other and articulate their concerns about AI through an interactive icebreaker.

- **Introduction & Icebreaker:** Welcome and introductions. Ask the learners about the origin of their personal interest in AI & AI Safety. Each participant shares one potentially catastrophic impact of AI that worries them and explains why.
- **Review of AI Governance 101:** A rapid refresher on key concepts:
  - *Discussion:* What social or political factors might impede AI safety efforts?
  - *Discussion:* Compare, at a high level, the current state of AI regulation in the EU, China, and the US.
  - *Discussion:* Should society prioritize slowing AI development, mitigating its impacts, or some other goal altogether?
- **Course Overview & Key Themes:** Outline of the course structure and a teaser of upcoming topics:
  - **Institutional Destabilization:** Preview reading “*How AI Threatens Democracy*” (4.2.1) – **read abstract** (<https://www.journalofdemocracy.org/articles/how-ai-threatens-democracy/>). This article discusses how AI could undermine democratic systems. This sets the stage for examining AI's impact on political institutions.
  - **AI in Warfare and National Security:** Preview reading “*Reducing the Risks of Artificial Intelligence for Military Decision Advantage*” (5.1.2) – **read executive summary** (<https://cset.georgetown.edu/publication/reducing-the-risks-of-artificial-intelligence-for>

- [-military-decision-advantage/](#)). This report directly addresses military applications of AI and associated risks, introducing a critical security dimension.
- **AI Policy Responses:** Preview the presented perspectives (Regulatory sandboxes, Risk-weighted regulation, etc):
    - *“The AI regulations that aren’t being talked about” (7.2.1) – read excerpt* (from “Two roads diverged in a digital world” up to “Regulators aren’t alone...”)  
(<https://www2.deloitte.com/us/en/insights/industry/public-sector/ai-regulations-around-the-world.html>). Highlights overlooked areas of AI regulation, giving a glimpse into the breadth of regulatory approaches.
    - *“Global AI Governance: Barriers and Pathways” (8.1.1) - skim the abstract and introduction*  
(<https://academic.oup.com/ia/article/100/3/1275/7641064?login=false>). Introduces the challenges and possibilities of international cooperation on AI, a core theme of the later sessions.
  - **Legal Aspects:** Preview reading *“AI, Governance Displacement, and the (De)Fragmentation of International Law” (9.2.3) – read abstract*  
(<https://matthijsmaas.com/uploads/Maas%20-%202021%20-%20AI.%20Governance%20Displacement,%20and%20the%20%28De%29Fragmenta.pdf>). Explores how AI may disrupt international legal order, a crucial legal consideration.
  - **Economic Impacts:** Preview reading *“Navigating AI’s Impact on Labor: Challenges, Scenarios, and Policy Pathways” (10.1.2) – read executive summary*  
([https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=5032882](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5032882)). Investigates AI’s effect on employment, giving a preview of the economic policy discussions.

## Session 2: Technical Alignment Challenges – The AI Control Problem

*Learning Objective:* Understand why advanced AI systems might be difficult to control or align with human values. We explore technical reasons for loss of control, the concept of AI “power-seeking” behavior, and emerging issues in AI alignment.

- **2.1 Technical Reasons for Lack of Control:** Foundational technical issues that could lead to loss of control over AI systems.

- **2.1.1 X-Risk Analysis for AI Research – read pp. 1–9**  
(<https://arxiv.org/pdf/2206.05862.pdf>). This paper categorizes existential risks from AI.
  - *Discussion:* Which categories of AI-related risk identified in the paper do you find most concerning, and why? Are any potential risks overlooked or underestimated?
- **2.1.2 Concrete Problems in AI Safety, Revisited - read all**  
(<https://arxiv.org/pdf/2401.10899>). An updated look at known AI safety problems.
  - *Discussion:* Considering recent AI advancements, which of the problems discussed have become more urgent? Are any less pressing now than they seemed previously?
- **2.2 Power-Seeking and Enfeeblement:** Scenarios where AI might seek power or cause human enfeeblement (dependency).
  - **2.2.1 “Is Power-Seeking AI an Existential Threat?” - watch the video**  
(<https://www.youtube.com/watch?v=UbruBnv3pZU>), *skim the paper*  
(<https://arxiv.org/abs/2206.13353>) – A presentation examining whether AI could intentionally seek power.
    - *Discussion:* Are the arguments for power-seeking AI as an existential threat convincing? Why or why not? What counterpoints or mitigating factors would you consider?
  - **2.2.2 Scheming AIs – read pp. 5–22** (<https://arxiv.org/abs/2311.08379>). Explores the concept of AI systems behaving deceptively to achieve goals.
    - *Discussion:* What do the authors define as “scheming” in AI, and do you agree this is a plausible near-term threat rather than a far-fetched scenario?
- **2.3 Alignment** (Ensuring AI systems act in accordance with human intentions and values):
  - **2.3.1 Alignment Faking in Large Language Models**  
(<https://www.anthropic.com/news/alignment-faking>). Describes instances where AI appears aligned but isn't truly.
    - *Discussion:* In what ways could “alignment faking” be even more dangerous in the long term than an AI that is openly unaligned? Consider trust and over-reliance.

## Session 3: AI and Information Warfare – Disinformation & Infosecurity

*Learning Objective:* Examine how AI can amplify misinformation and propaganda, and how it factors into cyberwarfare. Participants will learn about AI-generated disinformation, its societal effects, and the security implications of AI in cyber operations and terrorism.

- **3.1 AI-Driven Propaganda – Background:** How AI changes the landscape of misinformation and influence operations.
  - **3.1.1 Weapons of the Weak: Russia and AI-Driven Asymmetric Warfare** (<https://www.brookings.edu/research/weapons-of-the-weak-russia-and-ai-driven-asymmetric-warfare/>). A case study of how a nation with relatively fewer resources leverages AI for influence.
    - *Discussion:* How might AI reshape the balance of power between nations with differing tech capabilities? Do you agree with the paper’s assessment of Russia’s strategy, and what could it mean for global security?
- **3.2 Disinformation Tactics:** The role of AI in generating and spreading false information.
  - **3.2.1 Generative Language Models and Automated Influence Operations: Emerging Threats and Potential Mitigations** (<https://arxiv.org/pdf/2301.04246.pdf>). **Read pp. 1–4, 22–29, 38–65.** This study details how GPT-like models could be used at scale for propaganda.
    - *Discussion:* How do generative AI models alter traditional influence operations?
    - *Discussion:* Consider the “liar’s dividend” effect (dismissing real content as AI-generated). Could this be more dangerous than blatant AI falsehoods?
  - **3.2.2 How Persuasive is AI-Generated Propaganda?** (<https://cset.georgetown.edu/publication/how-persuasive-is-ai-generated-propaganda/>). An empirical look at the effectiveness of AI-created propaganda.
    - *Discussion:* Were you surprised by how persuadable people are by AI-generated content? Why might people trust or believe AI-crafted propaganda at such rates?
- **3.3 AI in Cybersecurity and Terrorism:** How AI tools can be misused for cyber attacks or by terrorist organizations.
  - **3.3.1 Algorithms and Terrorism: The Malicious Use of AI for Terrorist Purposes** (<https://www.un.org/counterterrorism/sites/www.un.org.counterterrorism/files/malicious-use-of-ai-uncct-unicri-report-hd.pdf>). **Skim pp. 10–25; read pp. 26–48.** Focuses on possible malicious AI applications by non-state actors.

- *In-class Exercise:* Based on the report, brainstorm a plausible scenario of an AI-enabled cyberattack (choose a target, method, and outcome). Share and discuss scenarios in groups.
- *Discussion:* Which potential terrorist use of AI did you find most alarming? What proactive measures can governments and international bodies take to prevent or mitigate such threats? Do you think an international agreement to limit malicious AI use is achievable?
- **3.3.2 State of Cyberwarfare 2024**  
(<https://media.armis.com/pdfs/rp-state-of-cyberwarfare-2024-en.pdf>). Overview of current trends in cyberwarfare with AI in the mix.
  - *Discussion:* How has AI integration transformed nation-state cyber attackers' strategies and capabilities?
  - *Discussion:* What are the pros and cons of using AI for cyber defense against these evolving threats?
- **3.3.3 Securing AI Model Weights**  
([https://www.rand.org/pubs/research\\_reports/RRA2849-1.html](https://www.rand.org/pubs/research_reports/RRA2849-1.html)). **Read the Summary, Chapter 2, Chapter 5, and Conclusion.** RAND report on preventing theft and misuse of frontier models.
  - *Discussion:* How should policymakers balance promoting open AI research and development with the imperative to secure AI model weights against adversaries?

## Session 4: AI's Impact on Political Systems – Institutional Destabilization

*Learning Objective:* Understand how AI might destabilize political institutions and alter power dynamics between autocracies and democracies. This session blends historical perspective with current analyses to discuss how emerging technologies influence governance models.

- **4.1 Emerging Technology & Institutions:** How new tech (like AI) challenges existing political institutions.
  - **4.1.1 AI and the Future of Democracy – A European Perspective**  
(<https://www.europarl.europa.eu/cmsdata/237002/Working%20Paper%20on%20AI%20and%20the%20Future%20of%20Democracy.pdf>). Reviews political party positions on AI in the EU.

- *Discussion:* Which political group's stance on AI governance resonated with you the most, and why? (Consider that you need not share a party's general ideology to agree with aspects of its AI stance.)
  - **4.1.2 *The Impact of the Internet on Institutions in the Future*** (<https://www.pewresearch.org/internet/2010/03/31/the-impact-of-the-internet-on-institutions-in-the-future-2/>). An earlier survey on how the internet could change institutions.
    - *Exercise:* As a class, conduct a mini "Pew-style" survey for AI. Predict how AI might impact key institutions in the next 10–20 years. Then, critically evaluate each other's predictions in light of what we know today.
- **4.2 Autocracy vs. Democracy in the AI Era:** Will AI favor authoritarian control or empower open societies?
  - **4.2.1 *How AI Threatens Democracy*** (<https://www.journalofdemocracy.org/articles/how-ai-threatens-democracy/>). Argues that AI can undermine democratic norms.
    - *Discussion:* In your opinion, how should education systems adapt to help people identify AI-generated misinformation?
    - *Discussion:* What responsibility do tech companies have in promoting digital literacy to combat AI-driven misinformation?
  - **4.2.2 *AI-tocracy*** (<https://academic.oup.com/qje/article/138/3/1349/7076890>). **Read pp. 1349–1356 & 1368–1380;** skim other sections. Introduces the idea that AI could strengthen autocracies ("AI-tocracy").
    - *Discussion:* What insight do the findings give about how political systems (authoritarian vs. democratic) shape and are shaped by AI adoption?
    - *Discussion:* How does AI innovation interplay with political control differently in autocracies compared to democracies?
  - **4.2.3 *AI and the Transition Paradox*** (<https://blog.practicaethics.ox.ac.uk/2022/04/ai-and-the-transition-paradox/>). Discusses how AI-induced economic disruption (e.g. via automation) might affect political engagement.
    - *Discussion:* If large numbers of people lose jobs to AI and rely on UBI (universal basic income), will they use their extra time to become *more* politically engaged, potentially counteracting the erosion of democracy? Or could the opposite happen?

## Session 5: AI in Warfare and National Security

*Learning Objective:* Explore how AI is transforming national security, military strategy, and the risks of autonomous weapons. This session covers the dual-use nature of AI in warfare — both offering strategic advantages and introducing new threats.

- **5.1 AI and National Security Policy:** How governments view AI in the context of defense and security.
  - **5.1.1 Artificial Intelligence and National Security**  
(<https://crsreports.congress.gov/product/pdf/R/R45178/10>). **Read pp. 16–20 & 35–39.** Provides a broad overview of AI applications in national security.
    - *Discussion:* Summarize key strategic implications of AI for national security identified in the report. How might AI change international competition or conflict dynamics (e.g., arms races, deterrence)?
  - **5.1.2 Reducing the Risks of AI for Military Decision Advantage**  
(<https://cset.georgetown.edu/publication/reducing-the-risks-of-artificial-intelligence-for-military-decision-advantage/>). **Read pp. 16–27.** Focuses on mitigating AI risks in military use-cases.
    - *Discussion:* How could failures in AI systems during military operations lead to unintended escalation of conflicts? (E.g., misclassification of targets.)
    - *Discussion:* What changes in training or military culture could help prevent or manage AI-related failures?
- **5.2 Autonomous Weapons Systems:** The challenges of regulating and controlling AI-driven weaponry.
  - **5.2.1 Challenges to International Regulation of AI in Military Applications**  
(<https://drive.google.com/file/d/1NnW8zLpnpR8cEfs7hIleZEaC7zEywGum/view?usp=sharing>). **Read pp. 3–8.** Reviews obstacles to global agreements on AI weapons.
    - *Discussion:* What are the biggest obstacles to international regulation of AI-enabled weapons? Consider trust issues, verification, great-power politics, etc.
    - *Discussion:* Can you think of an incentive structure that might persuade major powers to agree to AI arms control?
  - **5.2.2 AI Warfare is Already Here**  
(<https://www.bloomberg.com/features/2024-ai-warfare-project-maven/>). *Begin at “How Maven Works”.* Describes a real AI battlefield program.



- *Discussion:* How does the reality of current AI warfare programs (like Project Maven) compare to how AI in war is portrayed in science fiction?
  - *Discussion:* What ethical or legal concerns are raised by using AI for selecting or engaging targets in conflict?
- **5.3 AI and Other Strategic Weapons:** AI's impact on nuclear stability and other high-stakes domains.
  - **5.3.1 How Might AI Affect the Risk of Nuclear War?**  
 (<https://www.rand.org/pubs/perspectives/PE296.html>). Considers AI in nuclear command, control, and targeting.
    - *Discussion:* Could improvements in AI-driven targeting undermine the credibility of a second-strike (and thus destabilize nuclear deterrence)?
    - *Discussion:* Might AI-enabled surveillance make mobile nuclear launchers easier to find and destroy, thereby increasing the temptation for a preemptive strike? What would that mean for global stability?

## Session 6: Governing AI – Lessons from History and Emerging Frameworks

*Learning Objective:* Learn from historical attempts to govern disruptive technologies and examine theoretical frameworks for AI governance. This session sets the foundation for regulatory approaches by looking at past regulatory successes/failures and current governance proposals.

- **6.1 Lessons from History:** What history teaches us about regulating transformative technologies.
  - **6.1.1 History's Message About Regulating AI**  
 (<https://www.brookings.edu/articles/historys-message-about-regulating-ai/>). Draws parallels between AI and past tech (industrial revolution, nuclear energy, etc.).
    - *Discussion:* The author argues against succumbing to fear of new tech and instead addressing its societal effects (like job disruption). Is this a wise approach for AI?
    - *Discussion:* Have we missed the window to halt or significantly slow AI development? Or is there still an opportunity through policy action?
- **6.2 Theory and Practice of AI Regulation:** Current frameworks and strategies for AI governance, split into theoretical proposals and practical implementations.

- **6.2.1 (Theory) Understanding and Avoiding AI Failures: A Practical Guide** (<https://www.mdpi.com/2409-9287/6/3/53>). Introduces a framework for AI safety failures. *Start from the “Classification Schema for AI Systems” section.*
  - *Discussion:* What limitations of traditional safety engineering for complex systems does the paper identify in the context of AI? How does its framework propose to overcome these?
  - *Discussion:* Should any classic safety engineering practices be integrated into AI development?
- **6.2.2 (Theory) AI Is Testing the Limits of Corporate Governance** (<https://hbr.org/2023/12/ai-is-testing-the-limits-of-corporate-governance>). Discusses the limitations of corporate governance of AI in post OpenAI crisis context.
  - *Discussion:* Do you agree with the judgement delivered by this article upon corporate governance of AI?
  - *Discussion:* Is government regulation required to ensure companies act in the public interest regarding AI? If so, is such regulation feasible today?
- **6.2.3 (Theory) Artificial Canaries: Early Warning Signs for Democratic Governance of AI** ([https://www.ijimai.org/journal/sites/default/files/2021-02/ijimai\\_6\\_5\\_10.pdf](https://www.ijimai.org/journal/sites/default/files/2021-02/ijimai_6_5_10.pdf)). Proposes indicators (“canaries”) to detect early risks. *Read Abstract, Introduction, Sections III–V, Conclusion.*
  - *Discussion:* Might focusing on expert-identified early warning signs narrow public discourse on AI risks? Or could it help broaden engagement by highlighting concrete issues for debate?
- **6.2.4 (Practice) EU AI Act – Article 57: Regulatory Sandboxes** (<https://artificialintelligenceact.eu/article/57/>). Details how the EU plans to foster AI sandboxes. *Read together with 6.2.5 for context.*
  - *Discussion:* The sandbox provision offers liability protection (“safe harbor” from fines) if participants comply with sandbox plans, while keeping them liable for harm. Does this balance innovation and accountability appropriately?
  - *Discussion:* What positive or negative side effects might arise from this approach to liability in sandboxes?
- **6.2.5 (Practice) Regulatory Sandboxes in Artificial Intelligence** (<https://www.oecd.org/sti/regulatory-sandboxes-in-artificial-intelligence-8f80a0e6-en.htm>). Global survey of AI sandboxes. *Read Executive Summary, pp. 12–28.*
  - *Discussion:* The report calls for international coordination among AI sandboxes to ensure interoperability. What do you see as the biggest hurdle

to creating interoperable AI governance sandboxes globally? Propose one concrete step to address this hurdle.

## Session 7: Comparative AI Policy – Global Approaches

Note: We are actively looking for pieces dealing with the US policy horizon that would not get outdated within a week or two. Please do feel free to suggest materials that meet this daunting criteria.

*Learning Objective:* Compare how different governments and regions approach AI governance. This session covers the EU's strategy (including the concept of "AI sovereignty"), the US approach, and how these policies might (or might not) set global standards. We also introduce perspectives from China and others for contrast.

- **7.1 AI Policy Across the World:** Key regional strategies and their implications.
  - **7.1.1 EU AI Sovereignty:** *For whom, to what end, and to whose benefit?* (<https://www.tandfonline.com/doi/full/10.1080/13501763.2024.2318475>). An analysis of the EU's pursuit of "AI sovereignty."
    - *Discussion:* How does the EU's notion of AI sovereignty differ from that of the US or China? (Think in terms of jurisdictional control, citizen rights, competitive vs. protective approaches.)
  - **7.1.2 The EU AI Act's Global Impact (and its limits)** (<https://www.brookings.edu/articles/the-eu-ai-act-will-have-global-impact-but-a-limited-brussels-effect/>). Argues the EU AI Act will influence global norms but with a limited "Brussels effect."
    - *Discussion:* The article lists factors that limit the AI Act's international influence. Which factor do you believe most hinders the Act from becoming a de facto global standard, and why?
  - **7.1.3 "CERN for AI": The EU's Seat at the Table** ([https://icfg.eu/wp-content/uploads/2024/09/CERN\\_for\\_AI\\_FINAL\\_REPORT.pdf](https://icfg.eu/wp-content/uploads/2024/09/CERN_for_AI_FINAL_REPORT.pdf)). Proposes a large-scale European AI research hub (analogous to CERN).
    - *Discussion:* Would creating a "CERN for AI" help the EU close the AI innovation gap with the US (and China)? What advantages or challenges do you foresee with this approach?

- **7.1.4 United States Approach to AI**  
([https://www.europarl.europa.eu/RegData/etudes/ATAG/2024/757605/EPRS\\_ATA\(2024\)757605\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/ATAG/2024/757605/EPRS_ATA(2024)757605_EN.pdf)). Overview of recent US AI policy initiatives. (*Content updated after changes in administration.*)
  - *Note:* The US approach has evolved rapidly (e.g., new executive orders, agency directives). Keep this in mind as a moving target rather than a static policy.
- **7.2 Contrasting Policies – US, EU, and China:** Broader comparisons and under-the-radar policies.
  - **7.2.1 The AI Regulations That Aren't Being Talked About**  
(<https://www2.deloitte.com/us/en/insights/industry/public-sector/ai-regulations-around-the-world.html>). Highlights lesser-known AI regulations worldwide.
    - *Reading focus:* Excerpt covering different regulatory approaches that often escape mainstream discussion. This provides context on varied global regulatory efforts beyond the usual US/EU narrative.
  - **7.2.2 Comparing EU, China, and US AI Policy Landscapes**  
(<https://www.plurusstrategies.com/insights/2024/5/6/a-cross-sectional-comparison-of-eu-china-and-us-artificial-intelligence-policy-landscapes-1>). A cross-sectional analysis of how the three superpowers approach AI governance.
    - *Discussion:* After reviewing 7.1 and 7.2 readings, consider how cultural, political, and economic differences shape each region's AI policy. Which approach seems most comprehensive? Most flexible? Most likely to influence others?

## Session 8: International Coordination and Strategic Governance

*Learning Objective:* Investigate efforts toward international governance of AI and address strategic issues like compute (hardware) control. Participants will learn about challenges and initiatives in global coordination, including the role of international organizations and treaties, as well as the geopolitics of compute power.

- **8.1 International Coordination efforts:** How nations and international bodies attempt to collaborate (or clash) on AI governance.

- **8.1.1 Global AI Governance: Barriers and Pathways**  
(<https://academic.oup.com/ia/article/100/3/1275/7641064?login=false>). Discusses diplomatic and structural challenges to global AI governance.
  - *Discussion:* What do you identify as the greatest barrier to effective global AI governance? (e.g., great power rivalry, lack of trust, divergent values) How might that barrier be overcome?
- **8.1.2 India's Advance on AI Regulation.** A perspective on AI policy from India's viewpoint  
(<https://carnegieendowment.org/research/2024/11/indias-advance-on-ai-regulation?lang=en>).
  - *Discussion:* How does India's approach to AI governance compare to the US/EU approaches discussed earlier? In what ways could emerging economies influence global AI norms?
- **8.1.3 Global AI Cooperation on the Ground: AI R&D on a Global Scale**  
(<https://www.brookings.edu/wp-content/uploads/2022/11/FCAI-October-2022.pdf>).  
Case studies of international AI research collaborations. **Read pp. 12–16.**
  - *Discussion:* What are the benefits of international research cooperation in AI? Identify one successful model of global collaboration (from the reading or elsewhere) that could be replicated or expanded in AI governance.
- **8.2 Compute Governance and Geopolitics:** Regulating the hardware and infrastructure that power AI.
  - **8.2.1 Computing Power and the Governance of AI**  
([https://cdn.governance.ai/Computing\\_Power\\_and\\_the\\_Governance\\_of\\_AI.pdf](https://cdn.governance.ai/Computing_Power_and_the_Governance_of_AI.pdf)).  
Discusses control of high-end compute (chips, data centers) as a governance lever. **Read pp. 2–7, 24–33, 60–72.**
    - *Discussion:* Should access to advanced AI compute resources be internationally regulated (similar to how nuclear material is)? What might an effective compute governance regime look like?
  - **8.2.2 China Goes on the Offensive in the Chip War** (<https://archive.ph/pxybU>).  
Analysis of China's semiconductor strategy.
    - *Discussion:* How do export controls and national investments in AI chips factor into AI governance? Discuss the potential risks and rewards of the ongoing “chip war” for global AI safety.

## Session 9: Capstone Projects

*Note:* Please refer to the Capstone Project section at the end of this document for the description of the activities carried out on Session 9 and beyond.

*Reminder:* Beyond Session 9, the Capstone Project writing process runs concurrently with the optional sessions, should you elect to attend them.

## Session 10: Special Topic – AI and the Law (Optional Module)

*Learning Objective:* Delve into legal challenges posed by AI, including freedom of speech, liability of AI outputs, international law disruptions, and intellectual property. This module is optional for those who want a deeper dive into legal implications of AI.

- **10.1 AI and Speech Law:** How AI-generated content tests legal boundaries.
  - **10.1.1 Freedom of Speech and AI Output**  
(<https://www.journaloffreespeechlaw.org/volokhlemleyhenderson.pdf>). Examines whether AI-generated speech is protected and how it might be regulated.
  - **10.1.2 Section 230 Won't Protect ChatGPT**  
(<https://www.lawfaremedia.org/article/section-230-wont-protect-chatgpt>). Argues that current internet liability shields (like Section 230 in the US) may not apply to AI systems. *Discussion:* (For 9.1.1 & 9.1.2) How should the law treat AI-generated content? Who is liable when AI outputs cause harm — the user, the developer, or the AI itself?
- **10.2 AI and International Law:** AI's impact on global legal structures.
  - **10.2.1 A View from 40,000 Feet: International Law and the Invisible Hand of Technology** (<https://papers.ssrn.com/abstract=987524>). Early scholarship on technology's challenge to international law.
  - **10.2.2 International Law Does Not Compute: AI and the Global Legal Order** ([https://law.unimelb.edu.au/\\_\\_data/assets/pdf\\_file/0005/3144308/Maas.pdf](https://law.unimelb.edu.au/__data/assets/pdf_file/0005/3144308/Maas.pdf)). Explores how AI could fragment or transform the international legal order.
  - **10.2.3 AI, Governance Displacement, and (De)Fragmentation of International Law** (<https://matthijsmaas.com/uploads/Maas%20-%202021%20-%20AI,%20Governance%20Displacement,%20and%20the%20%28De%29Fragmenta.pdf>). Discusses how

AI governance might bypass traditional international law mechanisms. *Discussion:* (For 9.2.\*) In what ways might AI development outpace or undermine existing international law? Can international law adapt, or will new governance structures be needed (a “parallel” legal order for AI)?

- **10.3 AI and Intellectual Property:** Copyright and creativity in the age of AI.
  - **10.3.1 Copyright Policy Options for Generative AI**  
(<https://cepr.org/voxeu/columns/copyright-policy-options-generative-artificial-intelligence>). Policy proposals for copyright in the era of AI-generated content.
  - **10.3.2 AI and Artists' IP: Copyright in Andersen v. Stability AI**  
(<https://itsartlaw.org/2024/02/26/artificial-intelligence-and-artists-intellectual-property-unpacking-copyright-infringement-allegations-in-andersen-v-stability-ai-ltd/>). An ongoing legal battle over AI-generated art and copyright infringement.
  - **10.3.3 The Work of Copyright in the Age of Machine Production**  
([https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4581738](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4581738)). Theoretical perspective on how AI challenges the fundamental concepts of IP law.
  - *Discussion:* How should copyright law evolve to handle AI-generated works? Should AI outputs be eligible for copyright at all, and if so, who is the author? What balance would incentivize innovation without stifling human creators?

## Session 11: Special Topic – AI and the Economy (Optional Module)

*Learning Objective:* Analyze the economic implications of AI, including labor market impacts, effects on developing countries, and policy mechanisms like safety nets to handle AI-driven economic transformation.

- **11.1 Economic Trajectories of AI:** How AI might reshape global economic patterns.
  - **11.1.1 The Transformative Effects of AI on International Economics**  
(<https://www.policyjournal.net/the-transformative-effects-of-ai-on-international-economics.html>). Discusses AI's potential to change global trade and economic power balances.
  - **11.1.2 Navigating AI's Impact on Labor: Challenges, Scenarios, and Policy Pathways**  
([https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=5032882](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5032882)). **Read executive summary.** Examines how AI automation will affect jobs and what policies can address these changes.

- *Discussion:* What sectors and workers are most vulnerable to AI-driven disruption? Which policy pathways (e.g., retraining programs, UBI, job sharing) seem most promising to mitigate labor impacts?
- **11.2 AI in the Global South:** Opportunities and risks of AI for developing economies.
  - **11.2.1 AI Diffusion to Low- and Middle-Income Countries: A Blessing or a Curse?** (<https://arxiv.org/abs/2405.20399>). Considers how AI could either help or hurt developing nations (through productivity vs. inequality).
    - *Discussion:* How might AI widen or narrow the development gap between high-income and low-income countries? What policies could help the Global South benefit from AI while minimizing downsides?
- **11.3 Economic Safety Nets for AI Disruption:** Policies to ensure broad sharing of AI benefits.
  - **11.3.1 The Windfall Clause: Distributing the Benefits of AI** (<https://www.fhi.ox.ac.uk/windfallclause/>). Suggests a mechanism where extremely profitable AI developments (“windfalls”) are shared with society.
    - *Discussion:* Is a “windfall clause” a viable idea to address potential extreme inequality from AI breakthroughs? What might be the challenges in implementing such a clause internationally?

## Capstone Project

*The capstone project is a chance to synthesize course learnings into practical output. Participants will choose a policy topic related to AI governance and develop a brief or case study. (Participants may work in pairs for the project.)*

- **Project Options:**
  - **Option 1 – Policy Brief:** Analyze how a current or proposed AI policy initiative might conflict with existing laws or practices in another sector. Propose recommendations to harmonize these policies.
  - **Option 2 – Comparative Case Study:** Select a particular AI policy (actual or hypothetical) and compare it to a historical policy in a related domain (e.g., data privacy, biotech, aviation). Derive insights for AI governance from this comparison.
  - **Option 3 - Custom Project:** If you have a very clear idea of what you’d like to do instead of the two proposed options, you can discuss it with your facilitator.
- **Deliverables:**



- A written report of **2,000–5,000 words** (approx. 8–20 pages) detailing your analysis and recommendations.
- A **3–5 minute presentation** summarizing key findings of your project, to be shared in the final session.
- **Resources for Project Development:**
  - *How to write case studies and policy papers:*
    - **Case Writing Guide** – Schreyer Institute, Penn State  
(<https://www.schreyerinstitute.psu.edu/pdf/CaseWritingGuide.pdf>)
    - **Case Study Research: In-Depth Understanding in Context** – (Academic paper on case study methodology)  
(<https://drive.google.com/file/d/1JDHhgB5FumVZZAkdsTC5Dm78Jxj38r44/view?usp=sharing>)
    - **10 Tips for Developing Effective Policy Case Studies** – Stanford Law guidelines  
([https://law.stanford.edu/wp-content/uploads/2015/03/10\\_Tips\\_Case\\_Studies.pdf](https://law.stanford.edu/wp-content/uploads/2015/03/10_Tips_Case_Studies.pdf))
    - **Tips for Writing Policy Papers** – Stanford Law (White Paper guidelines)  
(<https://law.stanford.edu/wp-content/uploads/2015/03/Tips-for-Writing-Policy-Papers.pdf>)
    - **Policy Briefs** – UNC Writing Center guide  
(<https://writingcenter.unc.edu/tips-and-tools/policy-briefs/>)
- **Timeline:** (*Rough estimate, may be subject to change*)
  - **Capstone Kickoff (Week 9):** Week of September 15-21 – Come up with project ideas and an outline. In this, as well as in the Week 10 session, we'll discuss each participant / team's plan and provide feedback.
  - **Brainstorm Deadline (Week 10):** Week of September 22-28 – A check-in session
  - **Draft Deadline (Week 11):** Week of September 29 - October 5 – Submit a rough draft of your report (it can be partial). We will review and return comments.
  - **Final Deadline (Week 12):** Week of October 6-12 – Submit the final report. Presentations will be delivered this week in a mini symposium.

We recognize that some Capstone Projects might be more difficult and/or ambitious than others. We also would not want to force an incomplete work out for the sake of uniform pace alone. As such, if provided with sound reasoning, we are open to stretching the timeline.