

Spring 2026, 7 Week Curriculum

Week 1: Introduction to machine learning and AI

1. [But what is a neural network?](#) (3Blue1Brown, 2017). *Watch first 12.5 min.*
2. [The AI Triad and what it means for national security strategy](#) [Exec Summary Only] (Buchanan, 2020) *5 min.*
3. [4 charts that show why AI progress is unlikely to slow down](#) (Henshall, 2023). *8 min.*
4. [Can AI Scaling Continue Through 2030?](#) [Introduction, Chip Manufacturing Constraints (only the summary at the top, up to “Current Production and Projections”, then Figure 4), Data Scarcity, What Constraint is Most Limiting?, Will Labs Attempt to Scale These to New Heights?] (Sevilla et al., 2024) [20 m]

Additional readings:

- [But what is a GPT? Visual intro to transformers | Chapter 5, Deep Learning.](#) *Watch until 15:13.*
- [**What is a foundation model?](#) (Jones, 2023)
- [What will GPT-2030 look like?](#) (Steinhardt, 2023).
- [Literature review of transformative AI timelines](#) (Wynroe et al., 2023).
- [Visualizing the deep learning revolution](#) (Ngo, 2023). *15 min.*
- [Tracking Large-Scale AI Models](#) (Rahman et al. 2024)

Week 2: Overview of risks from advanced AI systems

1. [International Scientific Report on the Safety of Advanced AI](#) (Bengio et al. 2024)
 - a. Executive Summary (pp. 9-14)
 - b. § 4.1 Malicious use risks (pp. 41-47)
 - c. § 4.2.3 Loss of control (pp. 51-53)
 - d. § 4.3.1-4.3.1 Labour market risks, Global AI Divide, & Market concentration risks and single points of failure (pp. 54-59)
 - e. § 4.4.2 Cross-cutting societal risk factors (pp. 66-67)
2. [Harms from increasingly agentic algorithmic systems](#) (Chan et al., 2023). *Read pgs. 2-6 (§ 1 & 2). 10 min.*
3. [An overview of catastrophic AI risks](#) (Hendrycks et al., 2023). *Read pgs. 38–40 (§ 5.3, half of 5.4). 10 min.*
4. [Why AI alignment could be hard with modern deep learning](#) ^{**}(Cotra, 2021). *20 min.*

Additional readings:

- [Without specific countermeasures, the easiest path to transformative AI likely leads to AI takeover](#) (Cotra, 2022).
- [The alignment problem from a deep learning perspective](#) (Ngo et al., 2022).

Week 3: Safety standards and regulations

1. [From Principles to Rules: A Regulatory Approach for Frontier AI](#) (Schuett et al., 2024) [Parts I, IIB, IIIA, IVB]
2. [Managing extreme AI risks amid rapid progress](#) (Bengio et al. 2024)
3. [Breaking Down the Biden AI EO: Ensuring Safe and Secure AI](#) (Baker, 2023)
4. [Existing authorities for oversight of frontier AI models](#) (Bullock et al., 2024) *Read pgs. 1-16.*

Additional readings:

- [**Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence**](#) (The White House, 2023). *Read §4. 15 min.*
- [Primer on safety standards and regulations for industrial-scale AI development](#) (BlueDotImpact, 2023).
- [**Frontier AI regulation blueprint**](#) (Avin, 2023).
- [Biden Artificial Intelligence Executive Order Action Tracker](#) (Mayer Brown, 2024).
- [Frontier AI regulation: Managing emerging risks to public safety](#) (Anderljung et al., 2023).

Week 4: Self- and Corporate Governance

1. [Common Elements of Frontier AI Safety Policies](#) (METR, 2024). *Skip the section "Covered Threat Models". 15 pages.*
2. [If-Then Commitments for AI Risk Reduction](#) (Karnofsky, 2024). 15 pages.
3. [**AI is Testing the Limits of Corporate Governance**](#) (Tallarita, 2023) 15 min.

Additional readings:

- [Auditing large language models: A three-layered approach](#) (Mökander et al., 2023).
- [Evaluating language-model agents on realistic autonomous tasks](#) (Kinniment et al., 2023).
- [**Scaling AI Safely: Can Preparedness Frameworks Pull Their Weight?**](#) (Titus, 2024) 15 min.
- [Responsible scaling policy](#) (Anthropic, 2023). 25 min.
- [**The Long-Term Benefit Trust**](#) (Anthropic, 2023) 10 min.
- [When can we trust model evaluations?](#) (Hubinger, 2023).
- OpenAI Preparedness Framework ([blog summary](#), [complete](#)).

Week 5: Compute governance

1. [Computing Power and the Governance of Artificial Intelligence](#) (Sastry et al., 2024). *Read pgs. 2-6, 11-18, 34-59. (§ 1, § 2.C, § 4). 45 min.*
2. [Legal Challenges to Compute Governance](#) (Bernabei, Baker, and Fabrizio, 2024) [6 pages]

Additional readings:

- [**AI chips: What they are and why they matter**](#) (Khan & Mann, 2020).
- [Securing semiconductor supply chains](#) (Khan, 2021).
- [Mapping the Chip Smuggling Pipeline and Improving Export Control Compliance](#) (Harithas 2024)
- [Oversight for Frontier AI through a Know-Your-Customer Scheme for Compute Providers](#) (Egan & Heim 2023)
- [**U.S. Export Controls of AI Models**](#) (Bloomfield, 2024)

Week 6: Law and AI

1. **Torts:** [U.S. Tort Liability for Large-Scale Artificial Intelligence Damages: A Primer for Developers and Policymakers](#), RAND (Ramakrishnan, Smith, & Downey 2024).
 - For law students: Chapter 3 (12 pgs), Chapter 4 (9 pgs), Chapter 5 (3 pgs)
 - For non-law students: Summary (5 pgs), Chapter 2 (7 pgs), Chapter 3 (12 pgs)
2. **First Amendment:**
 - [Freedom of Speech and AI Output](#), Journal of Free Speech Law (Volokh, Lemley, & Henderson 2023) [10 pgs]
 - [AI Outputs Are Not Protected Speech](#), Washington University Law Review (Salib 2024), Part III.A only [pp. 29–45, 16 pgs]
3. **Section 230:**
 - [Section 230 Won't Protect ChatGPT](#), Journal of Free Speech Law (Perault 2023) [12 pgs]

Additional readings:

- [AI & Antitrust](#), Yale Journal of Law & Technology (Hua & Belfield 2021)
- [Open Questions in Law and AI Safety: An Emerging Research Agenda](#), Lawfare (Arbel et al. 2024)

Week 7: International governance, next steps, and careers

1. [High-level summary of the AI Act](#) (EU AIA, 2024). *10 min.*
2. [AISIs' Roles in Domestic and International Governance](#) (Ziosi et al., 2024). *15 min.*
3. [**China's Views on AI Safety Are Changing—Quickly**](#) (Sheehan, 2024). *5 min*
4. [Career resources on AI governance, policy, and strategy](#) (AISST/80K, 2023). *15 min.*
5. [Feedback form](#), scheduling 1:1s, and next steps.** *30 min*

Additional readings:

- [Winning the tech talent competition](#) (Zwetsloot, 2021).
- [Structured access: An emerging paradigm for safe AI deployment](#) (Shevlane, 2022).
- [The Windfall Clause: Distributing the benefits of AI for the common good](#) (O'Keefe et al., 2020).
- [International Institutions for Advanced AI](#) (Ho et al. 2023). *Read executive summary; § 3. 20 min.*

- [**AI Policy Resources**](#) (Horizon Institute for Public Service, 2024)
- [12 tentative ideas for U.S. AI policy](#) (Muehlhauser, 2023). 5 *min.*
- [China's AI regulations and how they get made](#) (Sheehan, 2023). *Read until "The underlying structure..."*. 20 *min.*
- [Open Questions in Law and AI Safety: An Emerging Research Agenda](#), Lawfare (Arbel et al. 2024)