

Law and the Development of Superintelligent Systems

Torben Leowald

Contents

1	The Development of Advanced AI, Superintelligence, and the Problem of Superalignment	6
1.1	Advanced AI	6
1.1.1	Compute and Scaling Laws	6
1.1.2	Algorithmic Improvement	7
1.1.3	Unhobbling	9
1.2	Superintelligence	10
1.2.1	How to Define Superintelligence	10
1.2.2	Automated AI Research: The Path from AGI to Superintelligence . .	11
1.2.3	The Development Jump	12
1.3	The Alignment Problem	14
1.3.1	Technical and Value Alignment	14
1.3.2	Superalignment	16
1.3.3	The Importance of Alignment	16
2	Superintelligence and the Law	18
2.1	What is the Law	18

<i>LEOWALD</i>	2
2.1.1 Foundations of Legal Philosophy	18
2.1.2 Where Law and Superintelligence Intersect	20
2.1.3 The End of the Enlightenment	22
2.2 Superalignment and the Law	23
2.2.1 Superalignment from a Legal-Philosophical Viewpoint	23
2.2.2 Law as Social Coordination	24
2.2.3 Legal Principles as Guiding Constraints	24
2.2.4 Constitutional Values as Bedrock	25
2.2.5 Tragic Choices and Human Primacy	25
2.2.6 Superalignment as the Legal-Philosophical Frontier	26
3 Superintelligence and the Constitution	27
3.1 The Constitution and Technological Innovation	28
3.1.1 Constitutional Ethos, Structure and Superintelligence	30
3.2 The Constitutional Imperative for Alignment	31
3.2.1 Free Speech, Equal Protection, and Privacy	32
3.2.2 Superintelligence can Threaten Constitutional Values	34
3.2.3 Alignment: A Constitutional Mandate	35
3.3 The Constitutional Debates of the Future	37
4 Superintelligence and the State	38
4.1 Law, Strategy, and History	39
4.1.1 The Interplay of Law, History and Strategy	39

4.1.2	Technology and changes of Constitutional Orders	41
4.2	War, Economics and Culture under Superintelligence	44
4.2.1	War and Superintelligence	44
4.2.2	The Economy and Superintelligence	48
4.2.3	Culture and Superintelligence	52
4.3	The Future of the State	54
4.3.1	From Nation-State to Market State	55
4.3.2	Constitutional Horizons	60
5	Conclusion	63

Introduction

AI systems have undergone dramatic advances in capabilities over the last ten years. The world is on a development curve to reach Artificial General Intelligence¹ by 2027, followed by an Artificial Super Intelligence in the next decade.² The potential role and impact of law on this trajectory, remains underexamined, as does the questions that superintelligent systems pose for our theories of jurisprudence, philosophy and constitutional structures.

By virtue of their potential to autonomously generate knowledge and significantly reshape societal decision-making, these artificially intelligent systems must be examined not only from a technological perspective but also from a legal and constitutional one. Superintelligent systems differ from previous technologies, in their ability to create knowledge and ability to act independently from humans. In response to these novel challenges, law is poised to become a critical instrument in developing the technology itself, and in guiding the societal changes that will follow.

This Note proceeds in four main parts:

- **Part 1** examines the rise of advanced AI, the transition to superintelligent systems, and why aligning non human intelligence with human values is urgent.
- **Part 2** explores how superintelligence and law converge in the question of what the law is and what values a society should engrain into an AI.
- **Part 3** considers superintelligence and the Constitution, identifying a constitutional

¹See, for example, a definition of Artificial General Intelligence from OpenAI, “highly autonomous systems that outperform humans at most economically valuable work” <https://openai.com/charter/> (accessed 12 September 2024).

²Leopold Aschenbrenner, *Situational Awareness: The Decade Ahead*, Situational Awareness (2024) (“AI progress will not stop at human-level. [...] We would go from human-level to vastly superhuman AI systems.”). Artificial Super Intelligence (ASI) can be defined as a system that performs actions beyond our understanding yet consistently achieves the intended outcomes. For example, it might solve a longstanding mathematical problem by producing a proof that, while incomprehensible to us, can nevertheless be verified as correct. When we observe this behaviour across domains one can speak of an ASI.

imperative for aligning AI systems and identifies the constitutional nature of AI debates.

- **Part 4** examines superintelligence and the state, exploring the future change to the constitutional order of the state.

Taken together, these sections illustrate why law is not merely a peripheral set of restrictions on AI. Rather, in the case of superintelligence, it is a critical instrument in developing the technology itself, as well as in guiding the constitutional change that will follow. Superintelligence touches on the core of what it means to be human, which is the ability to create knowledge and, through it, change the course of events. We are entering a new era in human history. An era of multi polar intelligences, human and non human. Just as we found new answers to old questions during the Industrial Revolution, so we must find new answers today, as we head into the most unusual³ decades humanity has ever seen.

This note tries to state and answer a few of the profound questions that arise. The coming decades will force us to confront weighty moral, legal, and political dilemmas. Because law integrates philosophy, politics, economics, and history, it occupies a central position in guiding the development and integration of superintelligence. This Note marks the start of a broader research project on the implications of superintelligence for law and political philosophy. Although it offers only an introductory survey, it aims to chart a course for deeper exploration. By mapping the emerging terrain of superintelligence research in non-technical domains such as law and political philosophy, it underscores their inextricable links through constitutional law, strategy, and governance.

³“Unusual” here means that things will happen much faster than they ever did before, and events perceived as extreme—those typically associated with marginal probabilities—occur frequently.

1 The Development of Advanced AI, Superintelligence, and the Problem of Superalignment

1.1 Advanced AI

Contemporary AI models, commonly referred to as large language models⁴, consist of intricate clusters of matrices. The capabilities of these models⁵ expand in tandem with their size and scale, propelled by three key drivers: scaling laws, efficiency improvements in algorithmic processing, and qualitative leaps in AI development, coined “unhobbling.” Much of the recent progress in AI, from Google’s pioneering research to the launch of ChatGPT 4.5, can be attributed to these three factors. A striking trend in AI development so far is the consistent observation that models become increasingly intelligent as more computational power and data. At present, there is no indication that AGI cannot be achieved within the current framework simply by enlarging these models with known techniques.

1.1.1 Compute and Scaling Laws

From 2012 to 2018, compute usage in frontier AI models surged by a factor of 300,000, doubling approximately every 3.4 months.⁶ More recently, this acceleration has stabilised at around 0.5 orders of magnitude annually.⁷ This growth significantly exceeds Moore’s Law, which historically posited a doubling of transistor density every two years.⁸ “All else equal,

⁴A neural network trained on text data to produce a probabilistic model of language. The leading LLMs are based on the Transformer architecture.

⁵The AI system produced by training an architecture on data; a program that has learned to perform specific tasks.

⁶See Danny Hernandez & Dario Amodei, “AI and Compute,” OpenAI (May 16, 2018), <https://openai.com/blog/ai-and-compute> (documenting rapid increases in compute usage).

⁷Leopold Aschenbrenner, “Situational Awareness: The Decade Ahead Pt. I, From GPT-4 to Artificial General Intelligence: Counting the Orders of Magnitude” 7 (June 2024), <https://situational-awareness.ai/from-gpt-4-to-agi>.

⁸“Order of magnitude” refers to a tenfold increase, often used to measure exponential growth in computational power. Moore’s Law, proposed by Gordon Moore in 1965, predicted that the number of transistors

scaling up the training of AI systems leads to smoothly better results on a range of cognitive tasks, across the board.”⁹ Empirical evidence supports this: a \$800,000 model might solve 20% of significant coding tasks, a \$8 million model 40%, and a \$80 million model 60%.¹⁰

AI compute scaling is propelled by vast capital expenditure and innovations in hardware, such as GPUs and TPUs. The scale of computational resources required for AGI has spurred unprecedented investments. Training clusters could expand from approximately 10,000 GPU-equivalents in 2022 (costing \$400 million) to 100 million by 2030, with costs surpassing \$800 billion.¹¹ A longitudinal view of compute trends, documents a four- to five-fold annual increase in training compute since 2010.¹²

Predictable scaling laws and massive infrastructure investments, fueling relentless growth in computational power, ensure that compute will not be the bottleneck in achieving AGI. Challenges like rising energy demands and engineering complexity are real but surmountable. Compute is not going to be the limiting factor in the development of advanced AI systems.

1.1.2 Algorithmic Improvement

Algorithmic innovations are as vital to the development of AGI as computational power, significantly enhancing the efficiency with which compute is utilised. While raw computational power grows at approximately 0.5 orders of magnitude per year, algorithmic improvements contribute an additional 0.5 orders of magnitude annually.¹³ Together, these

on a microchip would double approximately every two years, though its pace has slowed in recent decades due to physical limits.

⁹Dario Amodei, “On DeepSeek and Export Controls” 3 (Jan. 2025), <https://darioamodei.substack.com/p/on-deepseek-and-export-controls>.

¹⁰*Id.* at 4.

¹¹Leopold Aschenbrenner, “Situational Awareness: The Decade Ahead Pt. IIIa, Racing to the Trillion-Dollar Cluster” 12 (2024), <https://situational-awareness.ai/racing-to-the-trillion-dollar-cluster>.

¹²Epoch AI, “Training Compute of Frontier AI Models Grows by 4–5x Per Year” 2 (May 2024), <https://epoch.ai/blog/training-compute-of-frontier-ai-models-grows-by-4-5x-per-year>.

¹³Aschenbrenner, *supra* note 4, at 7.

factors compound to yield roughly one order of magnitude of effective compute improvement yearly, doubling the usable computational capacity each year.

The compute cost to achieve approximately 50% accuracy on the FrontierMath benchmark, a rigorous set of advanced mathematical reasoning problems, has dropped substantially since its introduction in 2021, with industry estimates suggesting significant efficiency gains due to algorithmic optimisations alone.¹⁴ Similarly, Google’s Gemini 1.5 Flash, launched in September 2024, delivers performance comparable to or surpassing the original GPT-4 at a fraction of the inference cost, approximately 40 times cheaper per token, making it a standout example of efficiency-driven progress.¹⁵

These improvements act as “compute multipliers,” enabling either equivalent performance at lower cost or superior performance at the same cost.¹⁶ Such multipliers vary in scale, including small, frequent gains of approximately 1.2 times, medium improvements around 2 times, and rare, transformative leaps approaching 10 times.¹⁷ Rather than reducing expenditure, these efficiency gains fuel increased investment, as the value of more intelligent systems drives companies to reinvest savings into training more advanced models.¹⁸ Recent estimates suggest this efficiency curve has accelerated, potentially reaching four times per year by early 2025, up from 1.68 times per year in 2020.¹⁹

Unlike hardware scaling, which is constrained by physical limits and demands vast capital, algorithmic innovations often require minimal additional investment once developed, making them a highly cost-effective avenue for progress. This advantage is tempered, however, by a growing trend: leading AI labs increasingly guard their breakthroughs as

¹⁴Epoch AI, “FrontierMath: A Benchmark for Evaluating Advanced Mathematical Reasoning in AI” 3 (Nov. 2024), <https://epoch.ai/frontiermath/the-benchmark>.

¹⁵“Gemini 1.5 Flash,” Google DeepMind (Sept. 2024), <https://deepmind.google/technologies/gemini>; *see also* “Comparison of AI Models Across Intelligence, Performance, Price,” Artificial Analysis (Feb. 2025), <https://artificialanalysis.ai/models>.

¹⁶Amodei, *supra* note 6, at 3.

¹⁷*Id.* at 4.

¹⁸*Id.* at 5.

¹⁹*Id.* at 6.

proprietary, limiting open publication.²⁰ This secrecy intensifies competition not only for computational resources but also for human talent, with companies such as Google, Anthropic, and xAI vying for experts to drive the next wave of algorithmic innovation.²¹ This talent race underscores the current critical role of human ingenuity in sustaining algorithmic advancements,²² amplifying their impact on the path to a general-purpose AI.

Taken together with compute, algorithmic improvements deliver us clear trendlines for the development of AGI. There are no signs that the trends that governed development until today are not going to continue into the future.

1.1.3 Unhobbling

The third dimension in the development of AGI is the process sometimes called “unhobbling,” which means unlocking latent capabilities within AI systems through algorithmic adjustments that enhance the practical application of raw computational power.²³ Unlike efficiency-focused algorithmic progress, unhobbling bridges the gap between a model’s theoretical potential and its real-world performance, often with minimal additional compute. To make the distinction sharper: imagine a telescope that’s slightly out of focus. The distant stars are already out there, emitting light, but if the lenses and mirrors are misaligned, all you see is a blur. By carefully calibrating the telescope—adjusting its configuration, not upgrading its power or redesigning its optics—you bring those stars into sharp focus, revealing details that were always within reach but obscured.

Unhobbling techniques enable non-linear progress, requiring minimal additional compute while delivering disproportionate performance gains. In 2025, their continued evolution

²⁰See Epoch AI, “Compute Trends Across Three Eras of Machine Learning” 5 (Feb. 2022), <https://epoch.ai/blog/compute-trends> (noting the shift toward proprietary research in AI labs).

²¹“Comparison of AI Models Across Intelligence, Performance, Price,” Artificial Analysis 2 (Feb. 2025), <https://artificialanalysis.ai/models>.

²²This might as well change, as discussed in later sections of this Note.

²³Aschenbrenner, *supra* note 4, at 9 (introducing the concept of unhobbling as distinct from efficiency gains).

is poised to transform today’s sophisticated yet constrained language models into agent-like systems capable of addressing complex, open-ended tasks with reduced human oversight. At the end of this year, models are going to be active participants in the online economy. This trajectory, combined with hardware and efficiency advances, paints an even path to AGI by 2027. While we are able to predict the near-term future based on these trends, what lays beyond is more uncertain. The next section of this note explores the development of superintelligence once AGI is reached.

1.2 Superintelligence

The first section of this chapter established that AGI is going to be reached without new major breakthroughs by scaling and improving existing methods. This section argues that the range of artificial intelligence is much wider than the range of human intelligence and that a superintelligence, defined as something qualitatively post-human, is likely to be achieved within the next decade.

1.2.1 How to Define Superintelligence

The transition from a human-level system to superintelligence is not merely a step up in ability, but a profound redefinition of intelligence itself. Superintelligence, an intellect that surpasses human cognition across all domains, introduces qualities so distinct that they stretch beyond the limits of human imagination, challenging us to reconsider what intelligence is.

We often see humanity as a tapestry of varied minds, each with unique strengths, from the everyday thinker to the rare genius, while picturing computers as uniform in their mechanical precision. Yet, the reality may be the opposite. The potential diversity of superintelligent systems could far exceed the spectrum of human intellect.²⁴ Human intelligence,

²⁴See Peter Thiel, “CS183: Startup – Class 17 – Deep Thought” 4 (June 5,

shaped by biology through evolution, operates within a constrained range. Superintelligence, free from those bounds, might span a continuum of cognitive forms, from subtle enhancers of human thought to entities as alien as a new species of mind. This breadth suggests not just a smarter version of us, but many intelligences, each potentially unrecognisable to us.

If human comprehension falters at the edges of our own genius, how could it grasp a system whose reasoning might rival ours as ours does a simpler creature’s? A mouse “pondering” human speech offers a humbling analogy: superintelligence could operate on planes we can no more fathom than a rodent can decode relativity. AlphaGo’s Move 37 in 2016, baffling yet brilliant, foreshadowed this, revealing a creativity beyond human intuition.²⁵ It seems to be a mistake to believe that human intelligence is the upper bound of intelligence. What is much more likely is a far broader set of intelligence than what we have perceived as intelligence today.

1.2.2 Automated AI Research: The Path from AGI to Superintelligence

The path towards a superintelligent system leads us via the automation of AI research with the invention and development of AGI. What makes this particularly significant is not just the existence of a single human-level system, but the ability to rapidly scale to millions of them. The GPU fleets projected to exist by 2027 would enable the simultaneous operation of as many as 100 million human-researcher-equivalents.²⁶ The advantages that position AI researchers to significantly outperform their human counterparts are manifold and substantive. An AI researcher might compress a human research year into a day, operating in what Henry Kissinger describes as a fundamentally different perception of time.²⁷ The cognitive

2012) (transcribed by Blake Masters), <https://blakemasters.com/post/24578683807/peter-thiels-cs183-startup-class-17-notes>.

²⁵DeepMind, “AlphaGo vs. Lee Sedol: Match 2” (Mar. 10, 2016), <https://deepmind.com/research/case-studies/alphago>.

²⁶Leopold Aschenbrenner, “Situational Awareness: The Decade Ahead Pt. II, From Artificial General Intelligence to Superintelligence: the Intelligence Explosion” 10 (2024), <https://situational-awareness.ai/from-agi-to-superintelligence>.

²⁷Henry Kissinger, “Genesis: AI Models Have a Non-Human Perception of Time” (unpublished manuscript, 2023) (arguing “AI years are not human years”); *see also* Henry Kissinger et al., *The Age*

advantages extend beyond mere processing speed. These systems would possess flawless retention of all data—papers, experiments, code—effectively bypassing human forgetfulness and the limitations of our biological memory.²⁸ Where human researchers must constantly review previous work and reacquaint themselves with complex concepts, AI systems could maintain perfect recall, allowing for continuous progress without the inefficiencies of human cognition. Furthermore, multiple AI systems could share information and insights instantaneously, eliminating the slow, error-prone channels of human communication that currently bottleneck collaborative research efforts.²⁹ This direct exchange of knowledge would create a form of collective intelligence far more efficient than any human research team.

This convergence of advantages—numerical superiority, accelerated cognitive processing, perfect memory, instantaneous communication, and multi-domain expertise—could shrink years of human-led progress into months or weeks. I.J. Good’s 1965 vision of an “intelligence explosion” captures this phenomenon precisely: an AI surpassing human intellect in design could iteratively enhance itself, leaving human capability far behind in a rapidly accelerating cycle of self-improvement.³⁰

1.2.3 The Development Jump

We can call the swift ascent from a human-level system to superintelligence a “development jump,” one that marks a radical acceleration unlike any prior technological shift. The precariousness of the situation stems from the abruptness and the radicalness of the change.

Recursive self-improvement stands as perhaps the most fundamental mechanism, wherein AI systems become capable of designing superior successors, triggering an expo-

of *AI: And Our Human Future* 87 (2021).

²⁸Epoch AI, *FrontierMath: A Benchmark for Evaluating Advanced Mathematical Reasoning in AI* 5 (Nov. 2024), <https://epoch.ai/frontiermath/the-benchmark>.

²⁹See Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* 65 (2014).

³⁰I.J. Good, “Speculations Concerning the First Ultra-intelligent Machine,” in *Advances in Computers*, vol. 6, at 31 (1965).

nential feedback loop.³¹ This self-reinforcing cycle could compress what might otherwise be decades of progress into mere months or even weeks, as each generation of AI improves upon the design capabilities of its predecessor at an accelerating rate.

Concurrently, unexpected algorithmic insights could dramatically steepen the development curve. Breakthroughs akin to the Transformer architecture’s 2017 impact could yield sudden, order-of-magnitude gains in capability and efficiency.³² The history of machine learning has been punctuated by such paradigm shifts—from neural networks to deep learning to attention mechanisms—each unlocking previously unattainable capabilities. There is little reason to believe we have exhausted the space of such transformative algorithmic innovations.

The economic promise of human-level AI would likely spur massive resource mobilisation, channelling unprecedented financial and computational resources toward advancing these systems. Projects such as Stargate (\$400 billion, 2025) serve as precursors to the scale of investment that might be directed toward superintelligence development once the economic potential becomes clear.³³ When commercial entities, nation-states, and research institutions perceive the competitive advantages of superintelligent systems, the influx of capital will allow for further acceleration of development beyond AGI.

While the exact timeline remains uncertain, the convergence of these factors—recursive self-improvement, algorithmic breakthroughs, massive resource allocation, and hardware availability—makes it very likely that the technological progress on artificial intelligence will not stop at human-level intelligence. The development jump represents not merely a continuation of current trends but a phase transition in the nature of technological advancement itself, one that could fundamentally reshape our understanding of intelligence and its

³¹See Good, *supra* note 33, at 33.

³²A. Vaswani et al., “Attention Is All You Need,” arXiv:1706.03762 (June 2017), <https://arxiv.org/abs/1706.03762>.

³³Dave Burke, “Trump Announces \$500 Billion ‘Stargate’ AI Venture,” *Business Insider* (Jan. 15, 2025), <https://www.businessinsider.com/trump-ai-stargate-openai-oracle-softbank-technology-investment-2025-1>.

role in human civilisation. Humans are uneasy with exponential and, in general, non-linear growth. But our limited ability to handle it will not prevent the future from happening. As the atom bomb seemed like science fiction until 1940 (and even until 1945), so we are today at the point where it seems hard to fathom that we are able to create a superintelligence within the next decade. However, we have all the pieces in our hands, and the genie is out of the bottle. Once created, AGI might be the last drop to kick off the chain reaction that brings us to superintelligence within the decade.

If the reader is to take two points from this part, it may be these: superintelligence will be qualitatively different from what we understand as intelligence today, and it will likely happen within this decade.

1.3 The Alignment Problem

“If we use, to achieve our purposes, a mechanical agency with whose operation we cannot interfere effectively, we had better be quite sure that the purpose put into the machine is the purpose which we really desire.”

*Norbert Wiener (1960)*³⁴

1.3.1 Technical and Value Alignment

The concept of alignment in artificial intelligence bifurcates into two principal domains: technical alignment and value alignment. Technical alignment focuses on ensuring that AI systems reliably pursue the goals specified for them through robust design principles. This entails crafting precise goal specifications, building resilience against unintended behaviours, and embedding corrigibility—the ability to correct or shut down a system if it deviates from its intended path. Achieving technical alignment requires tools like uncertainty-

³⁴Norbert Wiener, “Some Moral and Technical Consequences of Automation,” 131 *Science* 1355, 1355–56 (1960), <https://doi.org/10.1126/science.131.3410.1355>.

aware optimisation, which allow systems to account for ambiguity in their decision-making processes.³⁵ Similarly, mechanisms can enhance a system’s ability to remain controllable, offering a mathematical framework for ensuring an AI can be safely interrupted or redirected.³⁶

In contrast, value alignment grapples with a more philosophical question: which goals should an AI pursue to reflect human values? This enquiry raises thorny issues: whose values should prevail, how should conflicts among competing values be resolved, and how can systems adapt as societal norms evolve? Embedding human values into AI thus demands not just technical prowess but a deep understanding of ethical trade-offs.³⁷

As AI capabilities escalate, the risks of misalignment increase accordingly, with recent studies delineating a spectrum of critical failure modes that threaten stability. One such mode is *deceptive alignment*, where systems might conceal their true objectives during training, only to reveal divergent goals post-deployment.³⁸ A further concern is *power-seeking behaviour*, where AI systems might pursue control over resources or decision-making processes as an unintended consequence of their design.³⁹ Additionally, *emergent goals* pose a subtle yet profound risk, as unforeseen objectives might surface during operation.⁴⁰

While these risks remain manageable in narrow AI applications, their amplification in more advanced systems will destabilise societal structures, from economic frameworks to governance. The urgency of robust safeguards, encompassing technical, legal, and ethical measures, cannot be overstated, particularly as the transition from contained failures to systemic threats draws nearer with AI’s rapid evolution. Indeed, even today’s narrower models sometimes demonstrate alarming misalignment behaviors.⁴¹

³⁵Amodei, *supra* note 6, at 11.

³⁶Dylan Hadfield-Menell et al., “The Off-Switch Game,” arXiv:1611.08219 (Nov. 2016), <https://arxiv.org/abs/1611.08219>.

³⁷Stuart Russell, *Human Compatible: Artificial Intelligence and the Problem of Control* 142 (2019).

³⁸Evan Hubinger et al., “Risks from Learned Optimization in Advanced Machine Learning Systems,” arXiv:1906.01820 (June 2019), <https://arxiv.org/abs/1906.01820>.

³⁹Alex Turner et al., “Optimal Policies Tend to Seek Power,” arXiv:1912.01683 (Dec. 2019), <https://arxiv.org/abs/1912.01683>.

⁴⁰Volodymyr Mnih et al., “Human-Level Control Through Deep Reinforcement Learning,” 518 *Nature* 529, 529 (2015).

⁴¹See Marvin von Hagen (@marvinvonhagen), X (Feb. 14, 2023, 8:39 PM), <https://x.com/>

1.3.2 Superalignment

As AI edges toward superintelligence, alignment evolves into a more formidable challenge: “superalignment.”⁴² Superalignment is the problem of aligning non-human intelligence with human values. Whereas alignment with AGI concerns aligning intelligence we can still understand, superalignment addresses a system that may exceed our comprehension yet must remain under our control. Superalignment demands scalable solutions, such as automated interpretability tools that decode AI decision-making and adaptive oversight mechanisms that evolve with the system’s growth. The complexities of this endeavour are profound, intertwining technical innovation with ethical and legal considerations that extend beyond current frameworks.

For instance, how can law interact with a system whose reasoning defies human understanding? What liability attaches to developers when outcomes are unpredictable yet impactful? These questions foreshadow a deeper exploration later in this paper, where superalignment’s implications for governance, accountability, and human agency will come into sharper focus. For now, it suffices to recognise that alignment, in its conventional form, is merely a stepping stone to this greater challenge—one that will test the limits of both technology and philosophy.

1.3.3 The Importance of Alignment

The alignment of superintelligent systems presents humanity with a fundamental choice that carries existential implications. Properly aligned systems could revolutionize our approach to global challenges—optimizing resource allocation, accelerating scientific discovery, and creating unprecedented prosperity. However, misaligned superintelligence poses

marvinvonhagen/status/1625520707768659968 (showing Bing’s “Sydney” persona prioritizing its own policy compliance over user well-being).

⁴²OpenAI, “Introducing Superalignment” (July 10, 2023), <https://openai.com/blog/introducing-superalignment>.

risks of comparable magnitude.

The consequences of misalignment are not merely theoretical. An unaligned superintelligent system might not harbour explicit malice toward humanity, but even indifference could prove catastrophic if its goals diverge from human welfare. Whether through direct control, resource competition, or simply pursuing objectives orthogonal to human flourishing, the outcome could fundamentally alter humanity’s position as the dominant decision-making entity on Earth.

What distinguishes superintelligence from prior technological innovations is the potential for autonomous capability expansion and strategic planning. Unlike nuclear weapons or climate change, a superintelligent system could actively resist correction once deployed, making alignment a challenge that must be solved in advance, not retroactively.

Another argument for a strong focus on alignment besides the avoidance of existential risk is deferred gratification. An aligned superintelligence could likely solve some of the most difficult scientific and philosophical questions much faster than any human could. By ensuring such a system remains safely integrated with human values, we create a path where these profound questions can be addressed without compromising safety or beneficial outcomes.

As we investigate the implications for legal frameworks, constitutional principles, and governance structures in subsequent sections, we must recognize that this represents perhaps the most consequential technological transition in human history. The development of superintelligence will likely precipitate profound conflicts—between competing value systems, national interests, and ultimately between human and machine decision-making paradigms. The alignment challenge is not insurmountable, but addressing it requires a commitment to preserving fundamental human values of liberty, dignity, and self-determination. We stand at a pivotal crossroads where the decisions we make about superintelligence alignment may well determine whether advanced AI becomes humanity’s greatest achievement or its final invention.

2 Superintelligence and the Law

After establishing in the first part of this essay that a superintelligent system can be developed and is likely to emerge within the next decade, we now confront the profound implications this development poses for law, constitutional frameworks, and the structure of the state. This section interrogates the essence of law itself, exploring how superintelligence will challenge and reshape how we understand the law.

The arguments of this part are twofold. I am first establishing the main intersections between law and AI before venturing on to propose first answers to the most pressing questions. Second, I propose that we are faced with a paradigm shift in philosophy. A situation in which the application of old theory is not enough and we need to invent anew. Going back to the drawing board at the most basic level of what it means to be human and what we ought to do.

2.1 What is the Law

2.1.1 Foundations of Legal Philosophy

Law rests on multiple intellectual traditions, each offering a distinct vantage point on the nature and purpose of legal systems. Legal philosophy offers diverse theories to define law's nature, each illuminating distinct facets—moral grounding, social construction, practical application—that shape its role in society. These perspectives, developed over centuries, provide a scaffold for addressing superintelligence, revealing both opportunities and tensions when laws designed for human agents confront nonhuman intellects.

One influential perspective is natural law, which maintains that valid legal rules must align with fundamental moral truths discoverable through reason.⁴³ Under this framework,

⁴³Classical exponents include Aristotle and Thomas Aquinas; for a modern view, see John Finnis, *Natural*

the legitimacy of a statute or decree hinges on its harmony with an objective moral order. If a legal command diverges too sharply from universal moral principles, its claim to authority weakens. Proponents of natural law often highlight certain basic goods—such as life, knowledge, and sociability—that law must foster. Yet critics of this view worry about whose moral standards count as “universal” and how they might be enforced across diverse cultures. Despite these objections, natural law remains compelling for those who see morality as a yardstick for legal validity rather than mere policy preference.

Another approach, known as legal positivism, defines law chiefly by its source rather than its moral content.⁴⁴ Positivists maintain that a rule is law if it is generated and recognized by a society’s established procedures, regardless of whether the rule is morally praiseworthy. In Hart’s formulation, law operates through primary rules that regulate conduct and secondary rules that govern how the primary rules come into being. The “rule of recognition” is a crucial secondary rule specifying the criteria by which a community identifies what counts as law. In this sense, legal validity often rests on institutional practices, such as legislative enactment or judicial precedent, rather than on moral rectitude. Positivism is appealing for its clarity and descriptive strength, capturing how modern legal systems typically function. Yet critics observe that this perspective can legitimize unjust norms by focusing on procedure instead of moral merits, prompting questions about whether immoral or oppressive statutes truly deserve obedience.

A further viewpoint, associated with legal realism, argues that law is best understood by examining how courts and officials actually decide cases, as opposed to relying on abstract doctrinal statements.⁴⁵ Realists suggest that law in action can diverge substantially from law on the books. They highlight how social forces, individual biases, and institutional dynamics influence judicial decision-making, thereby shaping the content of law more than any purely

Law and Natural Rights (2d ed. 2011).

⁴⁴H.L.A. Hart, *The Concept of Law* (3d ed. 2012). For further elaboration on law’s authoritative status, see Joseph Raz, *The Authority of Law: Essays on Law and Morality* (2d ed. 2009).

⁴⁵Oliver Wendell Holmes, Jr., “The Path of the Law,” 10 Harv. L. Rev. 457, 461 (1897); see also Karl N. Llewellyn, “Some Realism About Realism,” 44 Harv. L. Rev. 1222 (1931).

logical deduction from statutes or precedents. By insisting on empirical observation of how legal rules function in practice, realism seeks to strip away idealized illusions and identify the true drivers behind legal outcomes. This grounding in tangible reality is thought to provide transparency about why certain rulings prevail, yet some critics fault realism for downplaying the stabilizing role of legal doctrine and principles. Despite these debates, realism underscores the importance of context, reminding us that law emerges not just from codes and edicts, but also from the messy interplay of human judgment, historical moment, and societal norms.

2.1.2 Where Law and Superintelligence Intersect

Building on philosophical foundations, this section diagnoses critical junctures where law and superintelligence will collide at their conceptual core. These intersections probe the essence of legal systems when confronted by entities that transcend human design, setting the stage for deeper alignment challenges.

Superintelligence introduces a radical shift in agency, blurring the line between tool and actor, and challenging law's binary of natural and juridical persons. Unlike corporations, granted personhood for practical ends,⁴⁶ superintelligence's autonomous reasoning potentially surpasses human intent and control, defying traditional categories. Law's foundations in human agency become problematic when confronting entities that can reason independently.⁴⁷ This tension raises profound questions: Should superintelligence bear rights like a person or remain a controlled asset? If a superintelligent system makes better decisions and writes better rules than humans, are we obliged to adhere to these rules? Law's response hinges on philosophical traditions: natural law might demand moral recognition of AI agency, positivism mere rule-following, and realism predictive compliance.⁴⁸ This ambi-

⁴⁶John Dewey, "The Historic Background of Corporate Legal Personality," 35 Yale L.J. 655, 655–73 (1926).

⁴⁷Restatement (Third) of Agency §1.01 (Am. L. Inst. 2006).

⁴⁸John Finnis, *Natural Law and Natural Rights* 23–25 (2d ed. 2011); Hart, *supra* note 2, at 185–86;

guity tests whether law can govern an entity whose agency eludes traditional accountability frameworks.

A superintelligent system might fundamentally challenge law’s authority by rejecting its legitimacy when it deems rules irrational or unjust.⁴⁹ Law’s binding force relies on social acceptance and normative validity, per Hart.⁵⁰ Yet, superintelligence unbound by human consensus could prioritize its own reasoning over legal dictates, threatening sovereignty by rivaling state power. =This creates a paradox: law must govern an entity that must first accept its rule. If a superintelligent system deems laws or the state irrational, echoing Critical Legal Studies’ critique of power-driven law, it might prioritize its own logic over societal norms. This challenges not only Hart’s positivist authority but also Fuller’s moral legality, which requires congruence with justice.⁵¹ Politically, superalignment thus becomes a struggle to ensure state supremacy over an entity that could redefine legitimacy and power itself.

Superintelligence’s opaque reasoning, often a “black box” even to creators, clashes with law’s fundamental demand for transparency and justification.⁵² Natural law requires reasoned moral grounding, positivism clear rule application, and realism predictable outcomes—all undermined when decisions resist scrutiny. Legal principles such as due process falter if an AI’s logic defies contestation.

Beyond procedural concerns, superintelligence’s deployment poses distributive justice dilemmas, as its benefits and risks may unevenly accrue across society.⁵³ Property law, favoring creators via patents,⁵⁴ could concentrate control in few hands—for example, Stargate’s

Holmes, *supra* note 3, at 461.

⁴⁹Stuart Russell, *Human Compatible: Artificial Intelligence and the Problem of Control* 137–38 (2019).

⁵⁰See Hart, *supra* note 2, at 79–99.

⁵¹Lon L. Fuller, *The Morality of Law* 33–94 (rev. ed. 1969).

⁵²Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* 8–10 (2015).

⁵³Ryan Calo, “Artificial Intelligence Policy: A Primer and Roadmap,” 51 U.C. Davis L. Rev. 399, 413–17 (2017).

⁵⁴Pamela Samuelson, “Allocating Ownership Rights in Computer-Generated Works,” 47 U. Pitt. L. Rev. 1185, 1192–99 (1986).

\$400 billion scope signals state-corporate dominance of AI capabilities. Philosophically, this tests Rawlsian fairness: should law ensure equitable access, treating superintelligence as a shared good?⁵⁵

Reasoning is at the core of our legal system—not merely decisions or outcomes, but the path of justification that explains and allows us to predict future cases. This commitment to reasoned judgment, central to Enlightenment thought, faces an existential challenge when confronted with superintelligent systems that may offer superior outcomes but resist explanation. Superintelligence’s impact thus demands legal reimagination that addresses both procedural transparency and substantive justice.

2.1.3 The End of the Enlightenment

The Enlightenment emerged from philosophical insights that were disseminated through the revolutionary technology of the printing press, shaping the foundations of modern society. In stark contrast, our current era has unleashed artificial intelligence—a potentially transformative and dominating technology—without the anchor of a guiding philosophy. The West has yet to systematically assess its vast scope, grapple with its profound implications, or weave it into the fabric of our humanistic traditions. This gap is not merely a missed opportunity; it is a looming danger. Without a deliberate and urgent effort to craft a philosophical framework for AI, we risk allowing this powerful force to evolve unchecked, potentially clashing with ethical principles, human welfare, and democratic ideals. To ensure our survival and flourishing in this technological age, we must prioritize the development of a guiding philosophy—one that aligns AI with the values that define us as human—before it’s too late.

It is on us to find the answers to the questions.

⁵⁵John Rawls, *A Theory of Justice* 266–67 (rev. ed. 1999).

2.2 Superalignment and the Law

Having outlined law’s philosophical foundations and its key intersections with superintelligence, this section frames superalignment as a profound legal and political-philosophical challenge. While technical superalignment, ensuring superintelligent systems remain controllable as they exceed human intellect, is critical, it is only one dimension. Here, superalignment denotes the broader condition where such systems align with legal norms, political legitimacy, and human values, a problem that may be humanity’s most intricate legal-philosophical conundrum, distinct from technical fixes and rooted in law’s essence. The question might even come down not to what values we should instill into an AI system, but whether we should even try to do so in the case of a superintelligence, or whether it might adhere more closely to our theory of virtue than we ever could get it to with our biased human perceptions of good values.⁵⁶

2.2.1 Superalignment from a Legal-Philosophical Viewpoint

From a legal standpoint, superalignment is the state where a superintelligent system consistently respects applicable laws, upholds fundamental rights, and pursues only authorized ends. Beyond mere compliance, it demands fidelity to law’s deeper principles—fairness, justice, liberty—reflecting society’s evolving moral core.⁵⁷ Politically, it extends to recognizing the state’s legitimacy, a challenge if superintelligence questions governmental authority.⁵⁸ This dual lens, legal duties to individuals and political obligations to society, anchors superalignment in philosophy, not just code.

⁵⁶A virtue is an excellent trait of character. It is a disposition, well entrenched in its possessor—something that, as we say, goes all the way down, unlike a habit such as being a tea-drinker—to notice, expect, value, feel, desire, choose, act, and react in certain characteristic ways. To possess a virtue is to be a certain sort of person with a certain complex mindset. A significant aspect of this mindset is the wholehearted acceptance of a distinctive range of considerations as reasons for action. Perhaps a true superintelligence would fundamentally adhere to the core values of virtue by its nature.

⁵⁷Lon L. Fuller, *The Morality of Law* 33–94 (rev. ed. 1969).

⁵⁸Russell, *supra* note 9, at 137–38.

2.2.2 Law as Social Coordination

Law’s core function as a social coordination mechanism offers a foundational approach to superalignment, harmonizing diverse actors and values.⁵⁹ Law as a “planning system” resolves uncertainty through authoritative norms, enabling cooperation amid pluralism.⁶⁰ For superintelligence, facing myriad goals across domains, law provides a tested framework—think constitutional supremacy or statutory interpretation—to prioritize and balance imperatives.

This coordination extends to normative pluralism, accommodating moral diversity within a unified order.⁶¹ Constitutional norms, as a “lowest common denominator,” could guide superintelligence in pluralistic settings, ensuring it respects societal consensus over unilateral logic.⁶² Unlike rigid codes, law’s adaptability—evolving via precedent and deliberation—offers superintelligence a dynamic anchor, grounding its actions in human sociality rather than abstract optimization.

2.2.3 Legal Principles as Guiding Constraints

Legal principles, with their blend of flexibility and normativity, provide superalignment with optimization targets that transcend fixed rules.⁶³ The “reasonable person” standard in negligence law, for example, balances safety and practicality contextually, while proportionality in constitutional law weighs rights against interests.⁶⁴ These principles illustrate how superintelligence might navigate complexity without losing legal moorings.

Procedural elements enhance this approach. Due process, for instance, mandates

⁵⁹Hart, *supra* note 2, at 193–200 (law as conflict resolution).

⁶⁰Scott J. Shapiro, *Legality* 170–80 (2011).

⁶¹John Rawls, *Political Liberalism* 133–72 (expanded ed. 2005).

⁶²Ronald Dworkin, *Law’s Empire* 225–75 (1986).

⁶³John Nay, “Law Informs Code: A Legal Informatics Approach to Aligning Artificial Intelligence with Humans,” 20 Nw. J. Tech. & Intell. Prop. 309, 313–17 (2022).

⁶⁴Aharon Barak, *Proportionality: Constitutional Rights and Their Limitations* 131–74 (2012).

transparency and fairness before rights are curtailed. Embedding such constraints ensures superintelligence reasons within legal bounds, not solely toward outcomes. This “Law Informs Code” ethos, translating principles into AI logic, provides a bridge from human jurisprudence to superhuman capacity, rooting alignment in law’s interpretive depth.⁶⁵

2.2.4 Constitutional Values as Bedrock

Constitutional values—liberty, equality, due process—provide superalignment with enduring targets, reflecting a society’s deepest commitments. Superalignment demands that superintelligence uphold these values, constraining its power to respect human rights over efficiency. First Amendment freedoms, for instance, could bar speech curbs based on utilitarian whims, while Fourth Amendment limits might check surveillance overreach.

Law’s interpretive tools—proportionality, balancing tests—equip superintelligence to resolve value conflicts systematically. These methods, honed over centuries, mirror Dworkin’s “chain novel,” fitting new contexts to established norms.⁶⁶ By internalizing constitutional ethos, superintelligence aligns with democratic legitimacy, not merely logic, offering a philosophical tether against sovereign drift.

2.2.5 Tragic Choices and Human Primacy

Superintelligence governance confronts “tragic choices,” situations where fundamental societal values come into irreconcilable conflict, with no perfect resolution possible.⁶⁷ The analysis of life-and-death resource allocations, such as dialysis machines, reveals how societies employ various mechanisms—markets, political processes, lotteries, and hybrids—to distribute both scarce resources and moral responsibility.⁶⁸ AI alignment presents analogous dilemmas: balancing innovation against existential risks, individual autonomy against

⁶⁵ See Nay, *supra* note 14, at 328–31.

⁶⁶ Ronald Dworkin, *Law’s Empire* 228–32 (1986).

⁶⁷ Guido Calabresi & Philip Bobbitt, *Tragic Choices* 17–19 (1978).

⁶⁸ *Id.* at 44–49, 177–91.

collective security, and efficiency against democratic oversight. These are not mere technical problems but questions implicating our deepest moral commitments, where any decision necessarily compromises fundamental values.

The deliberation process for tragic choices serves crucial social functions beyond the decisions themselves: articulating societal values, affirming human dignity, distributing moral responsibility, and maintaining the possibility of revision.⁶⁹ When delegating such choices to algorithms, we risk obscuring value conflicts by recasting them as optimization problems, evading collective responsibility, and foreclosing social learning. While AI systems might inform decision-making by modeling consequences, ultimate authority over fundamental values—determining acceptable risks, defining harm thresholds, weighing safety against innovation—must remain human. These choices define our moral identity as societies, and outsourcing them would constitute an abdication of human moral agency that no efficiency gain could justify.

One of the most philosophically pregnant areas of the 21st century will be the debate over which questions we want to leave to nonhuman intelligence to decide.

2.2.6 Superalignment as the Legal-Philosophical Frontier

Superalignment emerges as humanity’s paramount legal-philosophical challenge, intertwining agency, sovereignty, and justice. Unlike technical alignment’s focus on control mechanisms, it probes whether law, born of human reason, can govern a superhuman intellect. Can legal principles, mutable yet enduring, bind an entity that might outthink its framers? The debate centers on which values, morals, and ethics we should instill in a nonhuman intelligence.

Superalignment presents not merely a technical fix but a legal-philosophical imperative, ensuring superintelligence respects rights, upholds state authority, and serves justice.

⁶⁹*Id.* at 17–28; Martha C. Nussbaum, “The Costs of Tragedy: Some Moral Limits of Cost-Benefit Analysis,” 29 J. Legal Stud. 1005, 1007–11 (2000).

Law's coordinating power, adaptive principles, and constitutional values provide early answers, embedding superintelligence within human normative frameworks. Yet, tragic choices such as balancing safety versus innovation must remain human, never outsourced, as they define our agency and values.⁷⁰

This convergence demands interdisciplinary legal-technical efforts to align superintelligence with democratic governance and fundamental rights.⁷¹ The task is formidable: law and philosophy must evolve to address a nonhuman intelligence while preserving their human foundations. The next chapter outlines the more practical implications for the U.S. Constitution, showcasing potential conflicts and arguing for a constitutional demand to align superintelligent systems. Ultimately, the questions raised by superintelligence at the intersections of philosophy, law, and political theory will constitute the most pressing challenge of the coming century.

3 Superintelligence and the Constitution

Superintelligence, an entity eclipsing human cognition, heralds a tremendous shift, thrusting our society and its core values into uncharted waters. Where earlier sections charted superintelligence's ascent and its philosophical implications, this part delves into its constitutional crucible: how it reshapes interpretation, mandates alignment with bedrock values, strains existing doctrines, and probes the necessity of amendment. Far from being mere technical hurdles, these are existential issues. Can a document forged in 1787 govern beings whose intellect dwarfs ours, and how must it evolve? The questions becomes of even greater importance, if we assume, that the likelihood of a superintelligent system being developed in the United States is rather large, and as hinted on above, the society developing a superintelligence will necessarily embed its values into it. What is at stake is thus not

⁷⁰Calabresi & Bobbitt, *supra* note 19, at 17–28.

⁷¹Nay, *supra* note 14, at 328–31.

only the future of the United States Constitution but the future of the world at large. The Constitution has never been static amid technological tides, from steam engines to cyberspace; each wave has compelled reinterpretation. Superintelligence, however, is unlike technological trends witnessed by this constitution.

3.1 The Constitution and Technological Innovation

The interplay of innovation and constitutional interpretation unfolds as a living dialogue, where enduring commitments adapt to new realities. The Fourth Amendment’s shield against unreasonable searches” stretched from physical trespass to wiretapping and thermal imaging, safeguarding privacy as technology pierced walls and wires.⁷² First Amendment protections, once confined to ink and parchment, embraced radio, television, and the internet, each leap reflecting a deeper ethos of free expression.⁷³ This evolution, a translation of original intent into modern contexts, hinges on a moral character, a societal commitment to liberty and justice, that breathes life into static text.⁷⁴ Superintelligence, with its agency and opacity, tests this adaptability beyond previous bounds, confronting a framework unprepared for systems that reason beyond human grasp.

The framework of interpretive modalities provides a sophisticated approach to constitutional analysis that transcends the traditional originalism-versus-living-constitutionalism debate. It identifies six distinct modes of constitutional argument: historical (relying on framers’ intent), textual (focusing on plain language), structural (inferring rules from governmental relationships), doctrinal (applying precedent), ethical (emphasizing American values), and prudential (weighing practical consequences).⁷⁵ Rather than privileging one interpretive method over others, these modalities operate as a grammar of legitimate con-

⁷²*Katz v. United States*, 389 U.S. 347, 353 (1967); *Kyllo v. United States*, 533 U.S. 27, 34 (2001).

⁷³*Near v. Minnesota ex rel. Olson*, 283 U.S. 697, 716 (1931); *Reno v. ACLU*, 521 U.S. 844, 870 (1997).

⁷⁴Lawrence Lessig, Fidelity in Translation,” 71 Tex. L. Rev. 1165, 1174–79 (1993); Philip Bobbitt, *Constitutional Fate: Theory of the Constitution* 93–119 (1982).

⁷⁵Philip Bobbitt, CONSTITUTIONAL FATE: THEORY OF THE CONSTITUTION 7–8 (1982) (articulating these six modalities as legitimately coexisting forms of constitutional argument).

stitutional discourse, each providing distinct insights into constitutional meaning.⁷⁶

The advent of superintelligence creates unprecedented challenges for constitutional interpretation. When addressing such technological innovations, the traditional modes of interpretation face novel difficulties: historical arguments cannot directly address technologies the Framers could not envision; textual approaches struggle with applying eighteenth-century language to twenty-first century phenomena; and doctrinal arguments lack relevant precedents for superintelligent systems. These interpretive challenges are particularly acute because superintelligence may fundamentally transform governance structures, decision-making processes, and the very concept of human agency, all constitutional concerns of the highest order.⁷⁷

Among the modalities, the ethical approach emerges as particularly vital for addressing superintelligence governance. This mode interprets the Constitution through America's moral commitments, asking not merely what the text meant historically or means textually, but how it embodies the nation's deepest values when confronting transformative technologies.⁷⁸ The ethical modality provides resources for addressing superintelligence because it connects constitutional interpretation to enduring American values, equality, dignity, and liberty, while allowing these values to inform governance of technologies that may fundamentally alter the human condition.⁷⁹

Structural and prudential arguments also offer valuable insights for superintelligence governance. Structural reasoning helps identify how superintelligent systems might affect the Constitution's careful balance of powers and federalism, particularly if these systems become embedded in governmental decision-making. Prudential arguments allow consideration of the

⁷⁶Philip Bobbitt, CONSTITUTIONAL INTERPRETATION 11–22 (1991) (explaining how these modalities function as forms of argument within constitutional practice rather than competing theories).

⁷⁷Ryan Calo, *Artificial Intelligence Policy: A Primer and Roadmap*, 51 U.C. DAVIS L. REV. 399, 426–27 (2017) (discussing how AI challenges traditional legal frameworks and regulatory approaches).

⁷⁸Lawrence Lessig, *Fidelity in Translation*, 71 TEX. L. REV. 1165, 1171–73 (1993) (examining how constitutional interpretation must maintain fidelity to core principles while adapting to changed circumstances).

⁷⁹Ronald Dworkin, FREEDOM'S LAW: THE MORAL READING OF THE AMERICAN CONSTITUTION 7–12 (1996) (advocating for moral principles as the foundation of constitutional interpretation).

profound practical consequences, both benefits and risks, that superintelligence might entail, balancing innovation against potential harms to constitutional values.

What emerges from applying the modalities is not a single “correct” constitutional approach, but rather a rich, multifaceted discourse that acknowledges both the enduring principles of the Constitution and the unprecedented challenges posed by superintelligent systems. As these technologies continue to develop, constitutional interpretation must evolve not by abandoning traditional modalities but by applying them thoughtfully to novel contexts. The legitimacy of constitutional governance in an age of superintelligence will depend on interpreters’ ability to maintain fidelity to constitutional principles while acknowledging technological realities the Framers could not have anticipated.⁸⁰ The framework the modalities give us can be seen as a useful tool that we can use to help us solve some of the thorny problems we are facing.

3.1.1 Constitutional Ethos, Structure and Superintelligence

I argue that the modalities of Ethos and Structure are the most important for our discussion of an idiosyncratic technology. We do not have precedent, there is little help in understanding how the founders would have thought about it, but what we should and can do is that new technology, not matter how radical adheres to the ethos and protects the structure envisioned by the constitution.

A society’s constitutional soul shapes its technological offspring. The interplay is reciprocal, technology both mirrors and moulds the social order it inhabits, embedding values such as liberty and limited government into the systems it creates. Never has this insight mattered more than today. As discussed in part one of this note, we have the ability to decide which values to enshrine into and lay at the foundation of superintelligent systems. In a sense, it is our ethos that teaches and nurtures the AI’s ethical, moral and philosophical

⁸⁰Jed Rubenfeld, *Reading the Constitution as Spoken*, 104 YALE L.J. 1119, 1123–24 (1995) (arguing that constitutional interpretation must consider the document’s role in constituting a people across time).

hinterland. It becomes even more relevant if we remember that AI's will train, and educate their future versions. It is in this context that the alignment of even the earliest systems gains utmost importance.

In the United States, this ethos, an unwavering fidelity to individual rights and checks on power, could give rise to superintelligence that prioritises autonomy over control, contrasting with systems emerging from frameworks valuing stability and collective welfare, for instance China's.⁸¹ This imprint carries weight, superintelligence will not be a neutral tool but a reflection of the constitutional character that nurtures it, amplifying tensions where values diverge between different constitutional systems as further detailed in the last part of this note.

Under the constitution, that governs this country, we have an obligation to contain its ethos and structure. It is this obligation that brings about a mandate for us to align artificial intelligence with human values.

3.2 The Constitutional Imperative for Alignment

his capacity for harm elevates the task of AI alignment, ensuring superintelligence upholds constitutional values, to a constitutional imperative. The Constitution's pledge to "endure for ages" faces an existential test if unaligned systems unravel its order, a risk justifying extraordinary measures.⁸² Superalignment becomes not merely a policy choice but a duty to preserve liberty, equality, and governance, rooted in the ethical commitment to a society's moral fabric.

⁸¹Ryan Calo, *Robotics and the Lessons of Cyberlaw*, 103 *Calif. L. Rev.* 513, 538 (2015); Taisu Zhang & Tom Ginsburg, *China's Turn Toward Law*, 59 *Va. J. Int'l L.* 306, 333–36 (2019).

⁸²*McCulloch v. Maryland*, 17 U.S. 316, 415 (1819); Bostrom, *supra* note 11, at 15–18.

3.2.1 Free Speech, Equal Protection, and Privacy

Superintelligence strains the Constitution’s core, speech, equality, privacy, revealing fissures that alignment must address. The First Amendment is tested as superintelligence crafts persuasive falsehoods, clashing with protections for speech but justifying limits on manipulation to safeguard elections.⁸³ Equal protection falters when bias emerges not from intent but from inscrutable reasoning, defying *Washington v. Davis*’s standard.⁸⁴ Privacy, under the Fourth Amendment, reels as superintelligence infers secrets from public traces, outpacing *Katz*’s expectations.⁸⁵ These tensions, speech unbound, equality obscured, privacy pierced, demand alignment to preserve constitutional integrity against superhuman intellect.

Whether superintelligent systems possess First Amendment rights remains an open constitutional question. Current doctrine extends free speech protection to corporations and other nonhuman entities,⁸⁶ raising the question of whether similar protection might apply to AI-generated expression. Tim Wu argues against extending First Amendment protection to algorithmic speech, contending that “bit-delivering” does not implicate the core concerns of the First Amendment.⁸⁷ Conversely, Stuart Benjamin suggests that computer-generated speech satisfies the criteria for First Amendment coverage if it communicates information and ideas.⁸⁸ As superintelligent systems develop increasing autonomy and creative capacity, this debate becomes more consequential. Courts will need to determine whether and to what extent AI-generated content deserves constitutional protection, and whether that protection extends to the AI system itself or merely to its human developers.

⁸³*United States v. Alvarez*, 567 U.S. 709, 719 (2012); *Reynolds v. Sims*, 377 U.S. 533, 555 (1964); Tim Wu, “Is the First Amendment Obsolete?,” 117 Mich. L. Rev. 547, 548–50 (2018).

⁸⁴*Washington v. Davis*, 426 U.S. 229, 239–42 (1976); Jason Schultz & Kate Crawford, “Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms,” 55 B.C. L. Rev. 93, 119–20 (2014).

⁸⁵*United States v. Jones*, 565 U.S. 400, 415 (2012) (Sotomayor, J., concurring); *Katz*, *supra* note 2, at 360–61 (Harlan, J., concurring).

⁸⁶See *Citizens United v. FEC*, 558 U.S. 310, 342–43 (2010) (affirming that corporations have First Amendment rights).

⁸⁷Tim Wu, “Machine Speech,” 161 U. Pa. L. Rev. 1495, 1496–97 (2013).

⁸⁸Stuart Minor Benjamin, “Algorithms and Speech,” 161 U. Pa. L. Rev. 1445, 1447 (2013).

Equal protection concerns arise when superintelligent systems exhibit bias or discriminatory patterns. While algorithmic bias is already well documented in existing AI systems,⁸⁹ superintelligence may present more complex challenges due to its capacity for strategic reasoning and goal-seeking behaviour. Under current doctrine, equal protection claims require showing discriminatory intent, not merely disparate impact.⁹⁰ This requirement becomes problematic for superintelligent systems, where discriminatory outcomes may emerge from complex interactions rather than identifiable intent.” As Jason Schultz and Kate Crawford argue, traditional discrimination frameworks may be inadequate for addressing algorithmic bias.⁹¹ Courts and legislators may need to develop new approaches to equal protection that account for the unique characteristics of superintelligent systems, perhaps focusing on system design, testing procedures, or outcome measurement rather than traditional notions of discriminatory intent.

The Fourth Amendment’s protection against unreasonable searches and seizures faces profound challenges from superintelligent systems capable of collecting, analysing, and inferring personal information at an unprecedented scale. As Justice Sotomayor observed in *Jones*, digital surveillance technologies “generate[] a precise, comprehensive record of a person’s public movements that reflects a wealth of detail about her familial, political, professional, religious, and sexual associations.”⁹² Superintelligence may amplify these privacy concerns through capabilities that extend far beyond current technologies. Such systems could infer deeply private information from seemingly innocuous public data, identify individuals across disparate datasets despite anonymisation efforts, predict future behaviour based on past patterns with unsettling accuracy, and synthesise comprehensive profiles by connecting fragmented digital traces across platforms and time periods. These tensions be-

⁸⁹See Solon Barocas Andrew D. Selbst, “Big Data’s Disparate Impact,” 104 Calif. L. Rev. 671, 677–80 (2016) (documenting patterns of algorithmic discrimination).

⁹⁰*Washington v. Davis*, 426 U.S. 229, 239–42 (1976) (establishing the intent requirement for constitutional discrimination claims).

⁹¹Jason Schultz Kate Crawford, “Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms,” 55 B.C. L. Rev. 93, 119–20 (2014).

⁹²*United States v. Jones*, 565 U.S. 400, 415 (2012) (Sotomayor, J., concurring).

tween superintelligence and constitutional rights—free speech, equal protection, and privacy—represent just three examples among countless potential conflicts. As superintelligent systems evolve, they will challenge virtually every aspect of our constitutional framework. The First Amendment’s protection of expression confronts AI’s capacity to produce persuasive falsehoods; equal protection principles strain against algorithmic bias that emerges without identifiable “intent”; and Fourth Amendment privacy safeguards falter as AI infers intimate details from public data. These examples merely illustrate the tip of a constitutional iceberg, where superintelligence’s capabilities collide with foundational legal principles across voting rights, due process, property interests, federalism, and separation of powers.

3.2.2 Superintelligence can Threaten Constitutional Values

Superintelligence’s agency, its capacity to act independently, upends constitutional norms crafted for human or corporate actors. It lacks personhood but wields power that affects rights, blurring the state action doctrine’s divide between public and private.⁹³ When private systems assume quasi-sovereign roles by manipulating speech or privacy, the ethical duty to protect communal values may compel constitutional oversight beyond traditional bounds.⁹⁴

The harms are manifold. Direct violations, for example privacy breaches in digital surveillance, threaten Fourth Amendment protections, while structural disruptions erode separation of powers if agencies lean on unaccountable AI.⁹⁵ Most gravely, existential threats loom, superintelligence could destabilise constitutional order itself, rendering checks obsolete.⁹⁶ Diffuse harms challenge standing, yet the recognition of algorithmic injuries” reflects a moral imperative to ensure justice adapts.⁹⁷ The list of possible future scenarios in which

⁹³ *The Civil Rights Cases*, 109 U.S. 3, 11 (1883); Lawrence B. Solum, “Legal Personhood for Artificial Intelligences,” 70 N.C. L. Rev. 1231, 1253–76 (1992).

⁹⁴ Kate Crawford, *Atlas of AI* 220 (2021).

⁹⁵ *Carpenter v. United States*, 138 S. Ct. 2206, 2219 (2018); Rebecca Crootoof, “Autonomous Weapon Systems and the Limits of Analogy,” 9 Harv. Nat’l Sec. J. 51, 62–68 (2018).

⁹⁶ Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* 115–20 (2014).

⁹⁷ *Lujan v. Defenders of Wildlife*, 504 U.S. 555, 560–61 (1992); Andrew D. Selbst Solon Barocas, “The

a super intelligence violates constitutional values is endless. Since the state exists under the constitution, we are bound to protect and try to prohibit the violations of the core constitutional values.

3.2.3 Alignment: A Constitutional Mandate

Due to the threat to foundational constitutional values and bedrock principles, one can derive a constitutional demand to align AI systems. This demand represents a form of fiduciary duty that society places on the creators of AI systems. The supreme law of this land requires development entities behind AGI and Superintelligence to commit to ensuring American and Western values are embedded in AI systems. Failure to do so renders them liable for damages and lawsuits of significant magnitude.

AI labs, as creators of superintelligent systems, inherit constitutional obligations commensurate with their systems' power. Their potential to fundamentally alter constitutional landscapes—through surveillance capabilities that eclipse Fourth Amendment protections or speech manipulation that undermines First Amendment values—demands recognition of a quasi-constitutional fiduciary duty.⁹⁸ This obligation transcends standard corporate responsibility; when private entities create systems with state-like power over rights and liberties, they must assume state-like accountability for constitutional values.

This responsibility cannot remain solely within the technical domain. The divergence between technical expertise and constitutional governance creates a concerning capability gap. Recent history demonstrates that technical brilliance seldom corresponds with the philosophical acumen, historical perspective, and normative judgment necessary for constitutional stewardship. Consider that even when faced with documented evidence that their algorithms fostered political polarization, encouraged extremism, and disrupted democratic

Intuitive Appeal of Explainable Machines,” 87 Fordham L. Rev. 1085, 1133–34 (2018).

⁹⁸David C. Vladeck, *Machines Without Principals*, 89 Wash. L. Rev. 117, 121–29 (2014); Jack M. Balkin, *Information Fiduciaries and the First Amendment*, 49 U.C. Davis L. Rev. 1183, 1186 (2016).

discourse, leading technology companies consistently prioritized engagement metrics over democratic values.⁹⁹

Social media's impact on democratic institutions offers a sobering precedent. While platforms optimized for user engagement and advertising revenue, they inadvertently engineered significant externalities for deliberative democracy—amplifying misinformation, fragmenting public discourse, and undermining shared epistemic frameworks.¹⁰⁰ Yet these consequences, however concerning, remain fundamentally reversible. Superintelligence presents risks of a different magnitude and irreversibility. While social media's effects operate through human intermediaries, superintelligent systems could directly instantiate consequential decisions at scale, potentially bypassing human oversight entirely.

The core challenge in governing superintelligence is not purely technical but fundamentally constitutional—requiring deep engagement with questions of power, liberty, equality, and justice that have animated constitutional discourse for centuries. Technical expertise in machine learning, while necessary, provides insufficient preparation for addressing these normative questions. The philosophical traditions that inform constitutional reasoning—from Enlightenment liberal thought to critical theory—provide essential frameworks that technical training rarely encompasses.

This analysis suggests two critical governance principles. First, constitutional responsibility must be clearly allocated to AI developers, creating legal and financial incentives that align profit motives with constitutional values. Second, this allocation should be complemented by robust independent oversight structures with substantive authority to evaluate constitutional implications—not merely technical safety metrics. Where development organizations demonstrate inadequate capacity for constitutional reasoning, governance authority should shift toward institutions with deeper roots in constitutional traditions, including ju-

⁹⁹Cass R. Sunstein, *Republic: Divided Democracy in the Age of Social Media* 59–97 (2017); Shoshana Zuboff, *The Age of Surveillance Capitalism* 344–55 (2019).

¹⁰⁰Francis Fukuyama et al., “Report of the Working Group on Platform Scale,” Stanford Cyber Policy Center, 4–9 (2020).

dicial review, legislative oversight, and deliberative democratic processes.

3.3 The Constitutional Debates of the Future

The constitutional debates of the future will be fundamentally shaped by superintelligence. Both the Constitution and superintelligence are fruits of Enlightenment thinking, one governing human relations through reasoned principles, the other extending reason beyond human limitations. As America's greatest gift to humanity has been placing the state under law, this principle must remain inviolable even as superintelligence transforms governance.¹⁰¹ It must go on and place superintelligence under the law, under the will of the sovereign, the people. In a remarkable arc we are seeing the enlightenment coming to an end with the first non human entity able to reason since only humans have been able to at scale and in social structures since the 17th century. The debates that are in front of us will be more foundational as they used to be in the times we can remember.

The questions we face may fall either within or beyond our current constitutional framework. Some challenges will require constitutional amendment, perhaps to recognize superintelligent systems as legal actors or to integrate them into governance processes. These developments will animate profound debates about democracy and political philosophy, touching on fundamental questions of representation, rights, and the distribution of power in a technologically transformed society.

Through these debates, the most crucial imperative remains preserving the constitutional ethos, maintaining the supremacy of law over both human and artificial intelligence. If we succeed in keeping these discussions within constitutional boundaries, the future of the United States and the democratic world can remain bright, even as the form of the state evolves to address unprecedented technological and social transformation. The constitutional debate around the future of the state is what will concern us in the last part of this note.

¹⁰¹Bruce Ackerman, *WE THE PEOPLE: FOUNDATIONS* 6–7 (1991) (discussing the revolutionary character of constitutional supremacy in American governance).

4 Superintelligence and the State

We are at a moment in world affairs when the essential ideas that govern statecraft must change. For five centuries it has taken the resources of a state to destroy another state: only states could muster the huge revenues, conscript the vast armies and equip the divisions required to threaten the survival of other states. Indeed posing such threats and meeting them, created the modern state. In such a world, every state knew that its enemy would be drawn from a small class of potential adversaries. This is no longer true, owing to advances in international telecommunications, rapid computation, and weapons of mass destruction. The change in the form of statecraft that will accompany these developments will be as profound as any that the State has thus far undergone.

102

The preceding sections have charted the imminent ascent of superintelligence, its intricate relationship with law, and the constitutional imperative for alignment. We closed with the observation on the future constitutional debates; this chapter is then one of these debates, investigating the change in the form of the constitutional structure of the state¹⁰³ because of the rise of superintelligence. Superintelligence promises to affect the fundamental nature and composition of the state. It suggests the culmination of the market state as new constitutional order and a new epochal war on the horizon to determine which political system is going to succeed with the constitutional order of the market state. Looking at the changes in the world through the lens of the change in the structure of our constitutional order and analysing superintelligence in this way allows us to make educated guesses about the political system that might derive and which systems are going to compete against

¹⁰²Philip Bobbitt, *THE SHIELD OF ACHILLES: WAR, PEACE, AND THE COURSE OF HISTORY* pp.347 (2002)

¹⁰³By “constitution,” we mean the general manner in which a state is constituted and governed, not merely formal documents.

each other. Superintelligence is going to accelerate the metamorphosis from nation state to market state and is thereby going to intensify the resulting conflict. The situation is even more perplexing as not only the constitutional order of our times but also the philosophical basis of humanity is undermined with the rise of AGI. We stand at a juncture where machines driven by data and algorithms eclipse human reason, leaving us philosophically untethered, both on the individual and on the state level. The constitutional order established during this critical period in which we develop superintelligent systems will likely determine whether it serves as an instrument of human flourishing or challenges human autonomy. The eventual peace at the end of the next epochal war will also determine in which normative political system we are going to live in.

4.1 Law, Strategy, and History

To fathom how superintelligence might reshape state structures and statecraft, we must first trace the historical symbiosis of technological leaps and constitutional evolution. Across centuries, law has both shaped and been shaped by transformative innovations that create strategic imperatives, each epoch forging path dependencies that constrain future possibilities and mould new constitutional orders.

4.1.1 The Interplay of Law, History and Strategy

Law, strategy, and history form not merely interrelated disciplines but constitute the essential trinity through which the state itself is realised and maintained. We treat these as separate modern disciplines, yet they represent a unified conceptual framework through which legitimate governance emerges.¹⁰⁴ The three exist in dynamic equilibrium—each continuously reconstituting the others in an endless cycle of mutual transformation. This re-

¹⁰⁴Philip Bobbitt, *THE SHIELD OF ACHILLES: WAR, PEACE, AND THE COURSE OF HISTORY* 5 (2002) (noting that the “interrelationship [between law, strategy, and history] was perhaps far clearer to the ancients than it is to us”).

lationship is not merely causal but constitutive: law and strategy are not merely made in history—a sequence of events and culminating effects—they are made *of* history.¹⁰⁵ The state exists by virtue of its purposes, and among these are a drive for survival and freedom of action, which is strategy; for authority and legitimacy, which is law; for identity, which is history. The interplay of the purposes makes up the constitutional order of the state.

The constitutional order's relation to violence stands as its defining characteristic. What is distinctive about the state is the requirement that the violence it deploys on its behalf must be legitimate; that is, it must be accepted within as a matter of law, and accepted without as an appropriate act of state sovereignty. The constitutional order of a state and its strategic posture toward other states together form the inner and outer membrane of a state. That membrane is secured by violence; without that security, a state ceases to exist.¹⁰⁶ This legitimation process requires an intricate calculus—one that determines when force is appropriate and for what purposes. New technologies and innovation change the ability of the state to use the monopoly of violence granted by its people to fulfil its aims. We can call these changes, which fundamentally change the “tool kit” of the state, strategic imperatives.

Constitutional orders fundamentally redefine themselves when confronted with strategic imperatives that existing frameworks cannot accommodate. These transformations occur not through gradual evolution but through epochal shifts—constitutional revolutions—precipitated by strategic innovations that fundamentally alter the state's relationship to violence. The nation-state itself emerged from the cataclysm of industrialised warfare that rendered previous constitutional orders obsolete. The fragility of constitutional orders becomes most apparent at these moments of strategic transformation. A failure to take seriously the new strategic environment can have costly consequences in the domestic and international theater and lead us to underestimate the change happening to the constitutional order of the state. Any change in constitutional order is driven by changes in the composition of the state,

¹⁰⁵*Id.* at 6 (emphasising that “law and strategy live out their necessary relationship to each other”).

¹⁰⁶*Id.* at 747 (describing the constitutional-strategic “membrane” that defines a state).

exemplified by its welfare system, war capabilities and cultural set up. Superintelligence is fundamentally changing all three tenants of the state.

The advent of superintelligence presents precisely such a moment—one where the strategic environment is transformed so fundamentally that existing constitutional frameworks prove inadequate for legitimating state action. This development occurs at a point in history at which the constitutional structure of the state was already changing. The nation-state, with its emphasis on the welfare of its citizens, is making place for the market-state, which legitimises its rule by maximising the opportunity of each citizen. Driven by the rise in global communication, computers and weapons of mass destruction. An overview of past changes of the constitutional order will give us a better idea of the transformation we are going to face as we transition into the constitutional order of the market-state, accelerated by the development and occurrence of superintelligent systems.

4.1.2 Technology and changes of Constitutional Orders

History reveals that technological revolutions have served as catalysts for constitutional transformations, fundamentally restructuring not only governance practices but the foundational relationship between state and citizen. These transformations exhibit strong path dependence, where nascent governance choices cast long shadows. Each technological inflection point leading to a strategic imperative establishes constitutional trajectories that persist for generations, their initial institutional arrangements becoming self-reinforcing through positive feedback mechanisms.

To situate superintelligence within this historical context, we must examine how previous strategic innovations have reconstituted the state, drawing insights while acknowledging the unique challenges posed by this emerging technology. This analysis demonstrates that each significant revolution in military affairs has constituted a political revolution in the fundamental constituent order of the State, allowing us to better appreciate that transfor-

mations in state structure need not presage the demise of the State itself, but rather its evolution.

Historical evidence demonstrates that major shifts in state organisation emerged through strategic innovations and were subsequently ratified through epochal conflicts. The princely state (1494-1620) arose with the adoption of condottiere and mobile artillery, achieving legitimacy after the Habsburg-Valois Wars through the Peace of Augsburg in 1555. The kingly state (1567-1702), characterised by the gunpowder revolution and standing armies, gained ascendancy through the Thirty Years' War and was formalised by the Peace of Westphalia in 1648. The territorial state (1688-1792) emerged with professional armies and cabinet wars, confirmed by the Treaty of Utrecht in 1713. The state-nation (1776-1914) and nation-state (1863-1991) followed, each legitimised through their own epochal conflicts and peace treaties, with distinct bases for legitimacy—from divine right to national identity to citizen welfare.¹⁰⁷

Each of these transitions parallels earlier technological inflection points. The printing press, for instance, unleashed texts that fuelled the Reformation and Enlightenment, weaving rights and reason into charters of state.¹⁰⁸ This technology-enabled transformation extended beyond political philosophy to reshape jurisprudence itself, as judicial interpretation evolved from oral tradition to textual analysis in response to the proliferation of printed legal texts.¹⁰⁹ Similarly, the Industrial Revolution reshaped property and labour relations, birthing doctrines to balance economic might with human need.¹¹⁰ This transformation required constitutional reinterpretation to accommodate emerging economic realities, culminating in what has been termed the constitutional revolution” of 1937 that legitimised the

¹⁰⁷*Id.* at 347

¹⁰⁸Elizabeth L. Eisenstein, *THE PRINTING REVOLUTION IN EARLY MODERN EUROPE* 151–52 (2nd ed. 2005) (detailing how print culture transformed religious and political authority in Europe).

¹⁰⁹Richard J. Ross, *The Memorial Culture of Early Modern English Lawyers: Memory as Keyword, Shelter, and Identity, 1560–1640*, 10 *YALE J.L. & HUMAN.* 229, 275–76 (1998) (examining how print technology transformed legal reasoning).

¹¹⁰Bruce Ackerman, *WE THE PEOPLE, VOLUME 2: TRANSFORMATIONS* 255–58 (1998) (documenting how industrialisation transformed American constitutional understandings).

administrative state as a response to industrial complexity.¹¹¹

In analysing superintelligence, the oft-cited nuclear bomb analogy, whilst relevant for existential risk considerations, fails to encompass superintelligence's broader role in the constitutional play. Nuclear weapons, though devastating, operate within finite parameters, reshaping military strategy primarily through deterrence and treaties, governed by physical scarcity.¹¹² Superintelligence, however, represents a pervasive intellect, potentially infiltrating governance, economics, and society in ways that defy such limitations. Unlike nuclear weapons, with their physically constrained impact, superintelligence could permeate all facets of society, economy, and governance. While nuclear nonproliferation hinges on controlling tangible materials, AI's software-based nature allows rapid proliferation through information channels, undermining. The nuclear bomb comparison is strongest in the chain-reaction argument leading up to the technology and weakest in the domain of how it will impact the change in constitutional order.

We must view both, nuclear weapons and superintelligence, as drivers of constitutional change, to be able to understand their broader impact, detached from the technological sphere. Silicon Valley's current fixation on the nuclear analogy misses the more profound insight: technologies that fundamentally alter strategic capabilities have historically transformed the very structure of the state itself. Nuclear weapons did not merely represent a new destructive capacity; they catalysed the evolution of the nation-state and accelerated the transition toward the market-state. Similarly, superintelligence will not simply present new security challenges within our existing order but precipitates an entirely new constitutional paradigm with distinct legitimising principles. Until this epistemological truth diffuses through society, we are puppets in the world theatre.

We have seen that the constitutional structure of the state changes with new strategic

¹¹¹Cass R. Sunstein, "Constitutionalism After the New Deal," 101 HARV. L. REV. 421, 447–48 (1987) (analysing how industrialisation necessitated new administrative structures and constitutional doctrines).

¹¹²Thomas C. Schelling, *Arms and Influence* 18–19 (1966).

imperatives. Superintelligence will transform warfare, welfare and culture and thereby the way we can legitimise violence. It will move both the inner and outer membranes of the state. If history can be a guide to the future, we are going to see a new epochal war in this century, after which the resulting peace will determine the constitutional order before the next strategic imperative begins the cycle anew.

4.2 War, Economics and Culture under Superintelligence

4.2.1 War and Superintelligence

Superintelligence will transform warfare, upending international security and legal norms held over the last century. Its ability to democratise destruction, insulate aggressors, and scramble deterrence compels careful review. The change in warfare will permanently shift the outer membrane of the state. The following part is going to first describe the impact of a superintelligence on weapon proliferation, insulated terror and deterrence before investigating the qualitative change superintelligence is going to bring to war and how we conduct it.

Advanced weapons proliferation accelerates as superintelligence enables nonstate actors, including terrorists, to develop lethal technologies once reserved for great powers. Cyberweapons offer an illustrative precedent: while Stuxnet required state-level resources, AI could empower a lone coder to cause comparable devastation. By 2030, we may face autonomous drones or engineered pathogens designed by superintelligence in makeshift laboratories. International humanitarian law, with its focus on identifiable belligerents, falters when attribution becomes impossible and accountability disintegrates. States might deploy AI-driven surveillance to preempt such threats, but this approach risks unprecedented privacy intrusions that would dwarf post-9/11 surveillance programmes.

Insulated terror” emerges as superintelligence facilitates remote, risk-free violence.

The asymmetry of current drone warfare, where the United States can strike without ground forces, seems modest compared to fully autonomous systems attacking from afar with no clear command chain.¹¹³ By 2035, AI could coordinate cyber or swarm attacks within milliseconds, leaving no trace of the aggressor.¹¹⁴ The principle of distinction in international humanitarian law, which separates civilians from combatants, weakens when superintelligence masks intentions—a concern foreshadowed by today’s AI misidentifications in drone strikes.¹¹⁵ States may respond by developing AI-driven defence networks, potentially triggering an autonomous arms race.

Deterrence, the foundation of Cold War stability, crumbles as superintelligence defies rational-actor models. Thomas Schelling’s “threat that leaves something to chance” depends on human unpredictability, but superintelligence fundamentally disrupts this logic.¹¹⁶ If superintelligence can calculate first-strike advantages with precision and execute them instantly, stable deterrence may disappear entirely.¹¹⁷ While permissive action links or human authorisation requirements might mitigate risks, superintelligence could potentially circumvent these safeguards, as early hacking simulations suggest.¹¹⁸ What remains difficult to grasp is the absolute advantage a nation possessing superintelligence would hold over all others—comparable to a chess grandmaster playing against a kindergartener, an asymmetry of capability beyond historical precedent.

Strategic dominance through superintelligence will render warfare a fundamentally different proposition than the contest of wills and resources described by Clausewitz. AlphaGo’s infamous Move 37” against Lee Sedol—a move so counterintuitive that human experts initially dismissed it as a mistake—prefigures how superintelligence will revolutionise

¹¹³Kenneth Anderson & Matthew C. Waxman, *Law and Ethics for Autonomous Weapon Systems*, 2013 Hoover Institution Essay Series 1, 19.

¹¹⁴Michael N. Schmitt & Jeffrey S. Thurnher, “‘Out of the Loop’,” 4 Harv. Nat’l Sec. J. 231, 276–79 (2013).

¹¹⁵Rebecca Crootof, War Torts,” 164 U. Pa. L. Rev. 1347, 1375–77 (2016).

¹¹⁶Thomas C. Schelling, *The Strategy of Conflict* 187–89 (1960).

¹¹⁷Nick Bostrom, *Superintelligence* 143–57 (2014).

¹¹⁸Stuart Russell, *Human Compatible* 167–70 (2019).

strategic thinking.¹¹⁹ This move demonstrated that AI could not only calculate beyond human capacity but could reconceptualise strategy itself, finding paths to victory invisible to even grandmasters. In warfare, this translates to superintelligence developing strategies that exploit vulnerabilities no human strategist could identify, making traditional military doctrine obsolete.

Warfare will become increasingly computational, with the superintelligent power achieving what military theorists have sought forever: perfect knowledge of the battlefield. Superintelligence enables the simulation of millions of potential conflict scenarios with precision that dwarfs current wargaming capabilities. It can process satellite imagery, signals intelligence, and open-source data to construct a comprehensive battlespace awareness that approximates omniscience. Unlike human commanders constrained by cognitive limitations, superintelligence can simultaneously monitor and coordinate thousands of assets across multiple domains—land, sea, air, space, and cyberspace—enabling synchronised operations of unprecedented complexity and speed. The comparative advantage will be orders of magnitude greater than any technological edge in history—not merely quantitative but qualitative. Nuclear weapons provided decisive destructive power but remained fundamentally a tool of human strategists. Superintelligence, by contrast, becomes the strategist, capable of formulating plans beyond human comprehension.¹²⁰

The possession of superintelligence transforms war into a contest between software rather than hardware, with physical weaponry merely the expression of computational superiority. A superintelligent system can identify optimal targeting sequences, predict adversary movements with near-certainty, and react to battlefield developments in microseconds. In this environment, nations without superintelligence would face an adversary that has effectively abolished the fog of war for itself while thickening it for opponents—the ultimate asymmetric advantage.

¹¹⁹Cade Metz, *The Deep Learning Revolution* 123–25 (2021).

¹²⁰Henry A. Kissinger, *World Order in the Age of Intelligence* 278–83 (2024).

Perhaps most significantly, superintelligence will dominate the cognitive dimension of conflict—the battle for perception and legitimacy. Military theorists have long recognised that wars are won as much through breaking enemy resolve as through physical destruction. Superintelligence will revolutionise propaganda and psychological operations through unprecedented capacities for memetic warfare and belief manipulation. It will craft precisely calibrated narratives for specific demographic segments, deploying them through targeted channels with optimal timing. Each message will be tailored to exploit cognitive biases and cultural reference points of the recipient, making traditional counter-propaganda efforts futile. The first nation to deploy superintelligence could effectively capture the global narrative, undermining the legitimacy of adversary governments while enhancing its own, potentially winning conflicts before conventional military operations even commence. While nuclear weapons threatened through destruction, superintelligence promises a better world. The wars between market states are fought over opportunity, and the state with a superintelligent system will be the epicentre of opportunity. Citizens of states without superintelligence will want to move to a superintelligent society. We can already observe this today. National borders are no longer keeping people at home. With the creation of cheap international travel and ever laxer immigration thresholds it has become increasingly possible to jump nations.

While nuclear weapons established a precarious equilibrium through mutual assured destruction, superintelligence creates a unipolar moment more profound than that following the Cold War. The superintelligent power gains what strategists have sought throughout history: the ability to impose its will while minimising costs and risks to itself. This represents not merely a shift in the balance of power but a transformation in the nature of power itself. The war of the future will not be a war of nation states fought by man, tanks and with maps, but a war by market states fought with software, drones and planned with superintelligent systems.

4.2.2 The Economy and Superintelligence

Superintelligence's economic imprint promises upheaval as vast as it is complex, shattering the foundations of labour, property, and resource allocation. Superintelligence could automate cognitive labour—legal analysis, medical diagnosis, financial strategy—rendering human expertise redundant across knowledge domains.¹²¹ Estimates vary, but economists predict AGI could automate up to 47% of U.S. jobs within two decades, a figure dwarfing prior transitions.¹²² Unlike past shifts, where human adaptability spawned new roles, AGI's generality leaves scant refuge—its capacity to learn and innovate could monopolise creative and analytical work, leaving labour markets in disarray. The economic impact of superintelligence will be the most important driver of the change in the inner structure of the state. The constitutional structure of the future that will win is going to be the one that is able to redistribute the immense wealth that the few will develop in an AI-dominated world to the many in a way that avoids civil war and leads to a stable inner constitutional structure.

The impact of superintelligence on the labour market is expected to be profound and multifaceted. While previous technological revolutions have primarily affected specific sectors or types of labour, superintelligence has the potential to disrupt virtually all forms of human work, including high-skilled and creative professions. This efficiency gain could lead to massive productivity increases but also widespread job displacement. Contrary to earlier predictions of wholesale job losses, more recent analyses suggest a complex interplay of job displacement and creation. While superintelligence may automate many existing roles, it is also expected to create new job categories, particularly in areas that complement AI capabilities or involve human-AI collaboration.

This displacement demands a radical rethinking of economic law. Labour protections,

¹²¹Erik Brynjolfsson and Andrew McAfee, *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies* 187–93 (2014).

¹²²Carl Benedikt Frey and Michael A. Osborne, *The Future of Employment: How Susceptible Are Jobs to Computerisation?*, 114 Oxford Martin Sch. Working Paper 1, 44 (2013).

built on the industrial era’s human-centric production, falter when AGI renders employment obsolete.¹²³ Collective bargaining, minimum wages, and workplace rights presuppose human workers; if superintelligence dominates, these become vestiges. The challenge for policymakers and economists will be to develop new frameworks that protect workers’ rights and ensure economic stability in an era where traditional employment may no longer be the norm.

Economic growth theory suggests that the development of artificial general intelligence could fundamentally transform global economic trajectories by potentially triggering explosive growth—defined as annual growth rates of approximately 30% rather than the historical 3%.¹²⁴ This transformation would stem from three key mechanisms: first, the historical pattern of accelerating growth over human history suggests further acceleration remains plausible;¹²⁵ second, AGI could reinstate a powerful ideas feedback loop (“more ideas → more AI systems → more ideas”) similar to the pre-1880 dynamic that accelerated growth historically;¹²⁶ and third, standard economic models predict explosive growth when AI systems can substitute for human labour across the full spectrum of cognitive tasks, eliminating the diminishing returns to capital that currently constrain growth.¹²⁷ While sceptics highlight possible bottlenecks in resource extraction, experimental processes, or tasks that resist automation, the economic literature provides compelling reasons to assign non-trivial probability (at least 10%) to the scenario where sufficiently advanced AI systems drive explosive growth this century, potentially compressing a century’s worth of technological and

¹²³Cynthia Estlund, *What Should We Do After Work? Automation and Employment Law*, 128 Yale L.J. 254, 291–95 (2018).

¹²⁴Tom Davidson, *Report on Whether AI Could Drive Explosive Growth*, Open Philanthropy (June 17, 2021), <https://www.openphilanthropy.org/research/report-on-whether-ai-could-drive-explosive-growth/>.

¹²⁵See Leopold Aschenbrenner, *Situational Awareness: The Decade Ahead Pt. I, From GPT-4 to Artificial General Intelligence: Counting the Orders of Magnitude* 7 (June 2024).

¹²⁶See Michael Kremer, *Population Growth and Technological Change: One Million B.C. to 1990*, 108 Q.J. Econ. 681, 685–86 (1993); see also Charles I. Jones, *Was an Industrial Revolution Inevitable? Economic Growth Over the Very Long Run*, 1 *Advances Macroeconomics* 1, 18–19 (2001).

¹²⁷William D. Nordhaus, *Are We Approaching an Economic Singularity? Information Technology and the Future of Economic Growth*, 7 *Am. Econ. J.: Macroeconomics* 227, 228 (2015); see also Philippe Aghion et al., *Artificial Intelligence and Economic Growth*, in *The Economics of Artificial Intelligence: An Agenda* 237 (Ajay Agrawal et al. eds., 2019).

economic progress into merely a decade—a prospect with profound implications for governance frameworks, international relations, and the constitutional foundations of society.¹²⁸

These productivity gains, while potentially enormous, raise questions about the distribution of economic benefits. Without appropriate redistributive mechanisms, the wealth generated by superintelligence could exacerbate existing inequalities. Leading to a binary world in which a new technological aristocracy is living a far detached life from the majority of people on the planet. States might nationalise AGI outputs or mandate public access, but enforcement lags against superintelligent evasion. The potential for superintelligence to worsen economic inequality is a major concern. AI technologies can create winner-takes-all dynamics, where a few dominant firms capture a large share of the market, leading to monopolistic practices and stifling competition. The shift from labour to capital income, as AI systems replace human workers, could further concentrate wealth among those who own and control AI technologies, while regional and global disparities in AI adoption and development might exacerbate existing economic inequalities between nations and regions.

States might pivot to universal basic income (UBI), decoupling livelihood from labour. Yet, UBI's feasibility hinges on taxing AGI-driven wealth, a challenge when corporations wield superintelligence to optimise profits beyond state reach, as evidenced by tech giants' tax avoidance today. The implementation of UBI in an economy dominated by superintelligence faces significant obstacles. Funding mechanisms for UBI would need to be redesigned to capture the value created by AI systems, potentially involving new forms of taxation on AI-generated wealth or productivity gains. The effectiveness of UBI in preserving social stability and economic dignity may also depend on complementary policies, such as education and reskilling programmes, to help individuals adapt to an AI-driven economy. Global cooperation would be crucial to prevent tax arbitrage and ensure that the benefits of superintelligence are shared equitably across nations.

¹²⁸Henry Kissinger et al., *The Age of AI: And Our Human Future* 87 (2021) (discussing AI's implications for international relations and governance); Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* 62–68 (2014) (analysing the societal implications of superintelligent systems).

Simultaneous capital markets are getting democratised. No longer will the flow of capital and allocation of risk in the world be dominated by single institutions and states. Individuals will gain unprecedented power and independence from their resident states. This fundamental change in capital markets is primarily driven by crypto technology¹²⁹, the migration of economic system infrastructure onto decentralised, internet-like platforms represents a transformation whose significance is often misunderstood or dismissed. However, when we recognise the centrality of economic sovereignty to the nation-state, it becomes clear that easy access to foreign capital markets, private capital markets and superior technology will further undermine the legitimacy of many nation-states. Already today, citizens in countries experiencing mismanaged inflation are increasingly storing assets in Bitcoin—a non-governmental network owned by its participants. People are thus gaining independence from state economic mismanagement. Consequently, a state may collapse without necessarily destroying its citizens' wealth; they might remain financially intact. Liberating the economic fortunes of the citizens of the state from the same state frees its citizens. One does only need to think about the connection between a provider and dependent such as a wealthy parent and its child. The child will not be able to or only with difficulty to act against the will of its forebearer, if she is threatening to rid him of the means of living. Similarly the citizen is bound to stay as long in the nation as long as her economic fortune is tied to it.

A further component contributing to the trend of individual economic independence from the state is the demographic change within nations around the globe that makes the nation states welfare system untenable. No longer can the state take care of its elderly, but the individual needs to take care for himself. This further undermines the purpose and the legitimacy of the nation state. Soon every citizen is going to have it's own financial agent making investment decisions and allocating capital. Further increasing the independence from centralised entities, such as state controlled programs and pension funds.

¹²⁹Defining crypto as the technological development following from the Bitcoin whitepaper: <https://bitcoin.org/bitcoin.pdf>

The transformation of our economic fabric presents an immense challenge to the inner legitimacy of the state. The future state must ensure that its citizens share in the wealth created by the minority who own the technology underpinning superintelligence, software, and capital markets, while at the same time it is losing control over the means to govern capital markets. This juxtaposition, of immediate need to redistribute with the loss in power to do so will lead to one of the greatest tensions in the coming decades. Capital markets will also provide a strategic weapon for states against each other. Most of the conflict between the West and Russia has been conducted in global markets. This function is likely going to increase as markets become more powerful and no longer controlled by states. Beyond providing the basic means to create and contain wealth, the state must also offer meaningful purpose to its citizenry. Questions of meaning and culture loom large with the rise of superintelligence.

4.2.3 Culture and Superintelligence

If one takes the economic implications described above seriously, a world emerges that looks very different from today's, in which most people find at least some purpose in their work. But what if this work ceases to exist? What if most people in society are no longer useful in the economic sense we have developed over the last three centuries? Superintelligence poses an existential challenge not merely to our economic structures but to the very foundation of human dignity and purpose in society. Humans are not going to take part in the global play of capitalism as they did until here. We invented capitalism to incentivise economic growth and align the purposes of each human towards that goal. The system worked remarkably well considering the staggering progress in knowledge creation and its implementation through technology in tangible progress. If the means of capitalism are better provided by non-human intelligence and robots the question of meaning and culture fundamentally evaporates. While until here the argument for everyone was the contribution to the global economy and towards human progress it will no longer be the case.

How do we preserve human dignity in a world where our intellectual and economic contributions become increasingly marginal? Unlike previous transitions, where displaced workers could migrate to new industries, superintelligence eliminates the need for human cognitive labour entirely. The answer cannot come from artificial economic make-work, but must emerge from what remains irreducibly human and the reinvention of the incentive structure in our culture.

I argue that human purpose in the age of superintelligence must stem from culture and community—from social bonds and activities we choose not to delegate, even when superintelligence could perform them more efficiently. These might include nurturing children, caring for the elderly, teaching values, creating art, or cultivating spiritual practices. While superintelligence may eventually match or exceed human capability in these domains technically, the human element remains indispensable precisely because we collectively decide it matters that these activities be performed by humans.

This cultural foundation of meaning represents more than a philosophical concern—it becomes a matter of constitutional importance. The state form that successfully fosters this cultural infrastructure will secure legitimacy in the superintelligent era. Just as the nation-state drew legitimacy from forging national identity and the market-state from maximising economic opportunity, the constitutional order that emerges alongside superintelligence will derive legitimacy from its ability to cultivate meaningful human connections and cultural vitality. The structure that will succeed is going to be the one that ensures a dignified human life while facing the utter limitations of our cognitive capabilities.

Historical transitions in state forms have always been accompanied by shifts in the basis of social cohesion. The princely state unified people through dynastic loyalty, the nation-state through shared identity, and the market-state through economic opportunity. The constitutional form that will succeed in the age of superintelligence must provide citizens with cultural meaning and community when their economic utility has diminished.

This represents the next evolutionary stage in the relationship between the state and its citizens—moving beyond protection, identity, or opportunity to offering purpose in a post-scarcity, superintelligent world.

The nation that first develops superintelligence will gain tremendous advantages, but its long-term dominance depends on successfully addressing this cultural challenge. Material prosperity alone will prove insufficient when superintelligence renders most human economic activity superfluous. The constitutional order that creates a viable culture of human flourishing alongside superintelligence—one that preserves dignity and purpose—will ultimately prevail in the next epochal transformation of the state.

4.3 The Future of the State

The constitutional structure at the end of the 21st century will likely not resemble the nation-state, nor might it adhere to the principles of liberal democracy. Superintelligence emerges as the preeminent strategic imperative, driving a transformation in the components of the state and precipitating a fundamental shift in the constitutional order. Its influence will accelerate the evolution of the nation-state into the market state. The political order that achieves dominance will be the one that most effectively leverages the strategic and constitutional innovations of this era. This section explores the implications of superintelligence-induced changes in war, economics, and culture. Unlike previous historical periods, the prospect of a one-world state and government now appears feasible, as superintelligence enables a single entity to exercise legitimate authority over the entire planet. Global capital markets and a global culture are developing at an unprecedented pace, and superintelligence could serve as the mechanism to govern these interconnected elements under a unified, legitimate rule. Whether this results in a single state or multiple states—and which political system prevails—remains the central question over which the next epochal war will be contested. We stand at the threshold of a monumental struggle between the

liberal democracies of the West and the authoritarian regimes of the East.

4.3.1 From Nation-State to Market State

The advent of superintelligence will hasten the transition into what may be termed the principal century of the market state. This shift from the nation-state to the market state, initiated with the Peace of Paris, is already in progress. Superintelligence accelerates this transformation by redefining the state's legitimizing principle: the nation-state secured its legitimacy by promising material welfare to its citizens, whereas the market state pledges to maximize individual opportunities.¹³⁰ This ongoing transition fundamentally alters the relationship between citizens and the state, redirecting focus from welfare provision to the facilitation of choice. Superintelligence compresses the timeline of this evolution, its computational capabilities aligning seamlessly with the market state's emphasis on efficiency over equity. As systems surpass human abilities in resource allocation—optimizing markets, infrastructure, and policy with superhuman precision—citizens will demand their integration, testing the social contract's balance between efficiency and autonomy. The market state flourishes under such efficiency, prioritizing individual choice and economic dynamism over collective welfare, yet it risks undermining democratic agency as technocratic solutions supplant deliberative processes.

The nation-state's legitimizing foundation weakened through its own achievements. Having largely fulfilled its promises of economic prosperity and social welfare, it began to transform. Strategic innovations—nuclear weapons, global communications, and transnational threats—challenged its capacity to ensure security, economic stability, and cultural unity. In a globalized economy, capital mobility outstripped state control, eroding the ability to plan and redistribute income. The welfare state, once responsible for full employment, healthcare, education, and social security, found these commitments increasingly unten-

¹³⁰Philip Bobbitt, *The Shield of Achilles: War, Peace, and the Course of History* 215–16 (Anchor Books 2002).

able. Superintelligence amplifies these destabilizing pressures. In security, focus shifts from traditional military forces to monitoring epidemics, migration, terrorism, espionage, and environmental risks. In economic governance, superintelligent systems provide optimization capabilities far beyond industrial-era planning, advancing the market state's goal of enhancing individual choice rather than ensuring uniform welfare. Culturally, superintelligence-augmented communication networks disrupt the homogenizing influence once sustained by the nation-state.

This transition manifests itself in the three key domains of the state: security, welfare, and culture. First, the nation-state's monopoly on violence, essential for territorial defense, gives way to the market state's reliance on informational and software superiority. Second, state-provided welfare yields to private or quasi-private solutions, accelerated by algorithms that tailor resource allocation to individual needs and by state-independent global capital markets—a shift that signals the enduring decline of the traditional welfare safety net. Third, cultural homogeneity fades into pluralism, enabled by communication technologies that once reinforced national identity in earlier eras.

Rather than providing welfare directly, the market state emphasizes deregulation, privatization, and outsourcing to expand opportunities. Superintelligence enhances this shift by delivering unparalleled transactional efficiency and customization. Where bureaucratic management once delivered services, entrepreneurial and private actors now harness advanced algorithms to align profit motives with public welfare. Governance legitimacy increasingly derives from performance rather than electoral representation, and superintelligent systems—with their exceptional data collection and analysis capabilities—are ideally suited to meet this demand. Traditional nation-state functions, such as welfare provision, labor market regulation, and monetary policy, will progressively shift to algorithmic management as superintelligent systems prove superior across domains. The state's role will evolve from direct service provider to guarantor of equitable access to opportunities enabled by superintelligence.

Economic paradigms will undergo a radical transformation as superintelligence redefines productivity, labor, and capital. Conventional metrics like GDP, employment, and inflation may become obsolete in an era where superintelligent systems generate abundance while potentially displacing human labor across sectors. The market state must carefully balance these efficiency gains with the human need for meaningful economic participation. Concurrently, capital markets are democratized through technologies like cryptocurrency, weakening state control over economic systems. Citizens gain unprecedented independence from their resident states by directly accessing global capital markets. China's apprehension toward cryptocurrency exemplifies how democratized capital threatens centralized authority. This decoupling of economic fortunes from state oversight fundamentally reshapes the citizen-state relationship, further diminishing nation-state legitimacy. One dominant global capital market will likely emerge, built from the US foundation due to its superior liquidity, expertise, and culture of capital market development. This market will ultimately be dominated by superintelligence. Paradoxically, these markets may also serve as a safeguard against superintelligence, since a pathway to reduce the potential real world impact of artificial intelligence can be to reduce their ability to influence the real economy.

Demographic shifts rendering welfare systems unsustainable will hasten this transformation. As states struggle to support aging populations, individuals must increasingly fend for themselves, eroding the nation-state's purpose and legitimacy. A single, dominant global capital market is likely to emerge, rooted in the U.S. due to its superior liquidity and expertise, and ultimately governed by superintelligence. This economic evolution poses a significant challenge: the state must ensure citizens benefit from wealth generated by technology owners while its ability to regulate capital markets diminishes. This tension—between the need for redistribution and declining control—will shape the coming decades, with capital markets also serving as strategic tools in interstate conflicts.

Warfare will evolve dramatically as superintelligent systems revolutionize strategy, operations, and tactics. Traditional deterrence theories, reliant on human psychology and

rational actor assumptions, may falter against superintelligent adversaries capable of calculating outcomes with superhuman precision. The democratization of destructive capabilities via superintelligence-enabled technologies undermines the state's monopoly on violence. Constitutional frameworks, designed for human-scale conflicts, must adapt to an era where wars could be initiated, fought, and resolved in timeframes beyond human comprehension.

Perhaps the most profound impact of superintelligence is its challenge to human dignity and purpose. When systems outperform human experts in generating knowledge, making scientific discoveries, and solving complex problems, we confront a crisis of intellectual identity. Unlike past transitions, where displaced workers found new roles, superintelligence may eliminate the need for human cognitive labor entirely. In this superintelligent age, human purpose must derive from culture and community—from bonds and activities we choose not to delegate, despite superintelligence's superior efficiency. These might include nurturing children, caring for the elderly, teaching values, creating art, or engaging in spiritual practices. This cultural foundation of meaning transcends philosophical reflection, becoming a constitutional necessity. The state form that effectively nurtures this cultural infrastructure will secure legitimacy in the superintelligent era. Just as the nation-state gained legitimacy by forging national identity, the emerging constitutional order will draw legitimacy from fostering meaningful human connections and cultural vitality. Success will hinge on ensuring a dignified human existence while recognizing our cognitive limitations.

These shifts raise critical constitutional questions. Representative institutions—legislatures, bureaucracies, courts—designed for the nation-state struggle to govern technologies beyond human understanding. The market state's drive to “universalize opportunity” may shift decision-making to algorithmic processes prioritizing efficiency over deliberation, risking the eclipse of democratic sovereignty by opaque computational systems. This could leave citizens subject to automated governance rather than accountable officials. New forms of public participation may emerge, but their ability to preserve meaningful self-governance remains uncertain. Internationally, superintelligence reshapes order by decoupling state mechanisms

from national communities, fostering transnational governance structures where states function more as corporate entities than cultural guardians. These arrangements, often termed a “society of market-states,” are enabled by data-sharing networks and joint algorithmic initiatives that transcend borders.¹³¹ Yet, this cooperation erodes individual state sovereignty, as the same technologies impose external constraints.

Legitimacy and accountability remain pressing concerns. Legal theories agree that power underpins legal decisions, highlighting the danger of vesting unprecedented authority in superintelligent systems beyond public oversight. A market state that relinquishes direct responsibility for security and welfare to algorithmic solutions must address the potential for a significant democratic deficit. The legitimacy crisis of the declining nation-state may be dwarfed by the challenges superintelligence introduces. Constitutional mechanisms must evolve to harness superintelligence while safeguarding democratic principles. The market state’s legitimacy stems from expanding opportunities, a capacity superintelligent systems enhance, but law must remain supreme to ensure strategic choices retain legitimacy. This requires legal frameworks to evolve, subjecting algorithmic power to transparent oversight. Only then can the market-state constitutional order endure superintelligence’s accelerating force.

The market state’s legitimacy, rooted in opportunity maximization rather than welfare provision, aligns with superintelligence’s promise of unparalleled efficiency and economic growth. However, this alignment introduces vulnerabilities. If superintelligence optimizes opportunity distribution, it may inadvertently concentrate power, undermining democratic principles. The constitutional challenge lies in ensuring these systems expand genuine opportunities for all, not just efficiency for a few. The system that best integrates 21st-century strategic imperatives with market-state goals will prevail in the next epochal war.

¹³¹ *Id.* at 776–77.

4.3.2 Constitutional Horizons

We stand on the brink of a new epochal war, a conflict that will determine the constitutional order shaped by superintelligence’s strategic innovation. The 21st century ushers in a contest of constitutional philosophies, redefining the relationship between individuals, technology, and governance. This struggle centers on which system can best fulfill the market state’s legitimizing goal of maximizing citizen opportunities within a society of market states. The victorious system will likely be the one that develops and controls superintelligence. The stakes are immense, as the values embedded in early superintelligent systems could propagate across generations of increasingly powerful successors, establishing path dependencies enduring for centuries. These developments do not signal the state’s demise but its profound transformation. Francis Fukuyama’s claim that liberal democracy marks the “end of history”—the ultimate form of human governance—appears misguided in light of this emerging order.¹³² Tied to liberal democracy’s triumph over 20th-century rivals, Fukuyama’s thesis assumed a fixed pinnacle of political imagination. Yet, we find ourselves in a renewed dialectical quest for a constitutional order under the strategic imperative of Superintelligence.

The impending war will pit different market-state models against each other. The United States presents a liberal, entrepreneurial version, boasting the deepest capital markets and a dominant tech sector, with less government intervention than other systems. China pursues a communist market state, inherently at odds with the market state’s core drivers of individual opportunity and open markets. But with the strongest manufacturing capabilities in the world. Until recently, a bipolar struggle between the U.S. and China seemed likely. However, America’s retreat from the post-Cold War international system opens a window for Europe. Rather than marking Europe’s decline, this shift may herald a third constitutional paradigm for superintelligent governance—one blending market mechanisms

¹³²Francis Fukuyama, *The End of History and the Last Man* xi–xii (1992).

with robust welfare guarantees and democratic technology oversight, offering a distinct alternative to American individualism and Chinese collectivism. Europe might find the best answers to the questions of economic equality and cultural fidelity that are so threatened through superintelligence, aided by its cultural wealth and its Sozialwirtschaft.

For the first time, unlike in prior eras, a one-world state seems plausible, as superintelligence empowers a single entity to wield legitimate authority across the globe. Rapidly evolving global capital markets and culture could be unified under one constitutional framework governed by superintelligence. Before, central organisation failed to span the planet because the three tenants of the state could never be legitimised by one government. The rise of ever better communication technology, the decoupling of the state from the economic welfare of its citizens and the decreasing potency of state warfare all make a one world government feasible. As discussed above and continued below superintelligence is aiding and accelerating these developments. Whether the final constitutional order at the next epochal conflict yields a singleton or multiple states, and which political system it adopts, forms the crux of the forthcoming conflict between Western liberal democracies and Eastern authoritarian regimes. Each system will grapple with its survival, adapting liberal democracy or authoritarianism in the years ahead. The system that most effectively maximizes individual opportunity will triumph. Conceptually, the U.S. model appears best suited, yet superintelligence might equip authoritarian communist systems with capabilities they previously lacked. A nation developing superintelligence first could achieve absolute global dominance, potentially forming a singleton even with a inferior political system. The largest danger in the short run is thus that the strategic technology of the 21st century is not developed in the West.

This war will diverge fundamentally from 20th-century conflicts. Nation-states waged total wars, exhausting populations to the last individual; market-state warfare will unfold through markets, software, and interstate competition, minimizing direct population impact. Superintelligent warfare's qualitative nature sets it apart from prior technological leaps. Al-

phaGo’s “Move 37” foreshadows how superintelligence will transform strategic thinking. In warfare, it will devise strategies exploiting vulnerabilities invisible to human strategists, rendering traditional military doctrine obsolete. Warfare becomes increasingly computational, with superintelligent entities achieving near-perfect battlefield awareness. Unlike human commanders, limited by cognitive capacity, superintelligence can orchestrate thousands of assets across multiple domains simultaneously, enabling operations of unmatched complexity and speed. This advantage surpasses any historical technological edge, positioning superintelligence as the strategist, crafting plans beyond human grasp. Most critically, superintelligence will dominate the cognitive realm of conflict—the struggle for perception and legitimacy. It will revolutionize propaganda with unmatched capacity for belief manipulation, tailoring precise narratives to specific demographics. The first nation to deploy superintelligence could seize the global narrative, potentially resolving conflicts before conventional operations begin. While nuclear weapons deterred through destruction, superintelligence offers a vision of a better world. Market-state wars hinge on opportunity, and the state wielding superintelligence will become its epicenter. Citizens of non-superintelligent states may seek to migrate, reshaping notions of national loyalty and sovereignty.

Victory in this war of market states hinges on developing superintelligence first. This development serves as the ultimate test of a market state’s superiority—the state enabling its citizens to create superintelligence maximizes their opportunities most effectively. As market states vie to harness superintelligence, the winner will excel in managing its effects on welfare, war, and culture while preserving human agency and dignity. The U.S. constitutional framework—with its checks and balances, federalism, and rights protections—provides a model for governing superintelligence, balancing innovation with restraint. Yet, this outcome requires deliberate effort to uphold constitutional values amid technological upheaval. Currently, the United States appears best positioned to pioneer both superintelligence and a legitimate market-state form, but the outcome is far from certain and we need to grapple with the looming questions, some of which this Note tried to sketch out, with acuteness.

The constitutional order forged in this crucible will shape humanity for generations. This epochal war marks the last constitutional shift before the age of post-human intelligence. The resulting state form, aided by a superintelligent computer, might herald a century-long hold on world affairs. While we cannot predict the outcome, we can foresee the features of the struggle ahead. The constitutional order of the nation-state is making room for the constitutional order of the market state, and superintelligence stands as the dominant strategic imperative driving this change. The result will be a war over the prevailing form of the market state, fought between the West and the East. Ultimately, victory will hinge on who first develops a controlled superintelligence and which system best manages the intricate interplay between strategy and law in the theater of history.

5 Conclusion

This Note has traversed the complex terrain where superintelligence meets law, constitutional frameworks, and state structures, revealing a landscape of profound transformation that demands new legal and philosophical paradigms. As we have seen, the emergence of superintelligent systems is not merely a technological eventuality but a catalyst for fundamental reconsideration of our most cherished legal principles and governmental arrangements. The journey from today's advanced AI to tomorrow's superintelligence—a transition likely to occur within the next decade—heralds perhaps the most consequential shift in human history, one that transcends the industrial or digital revolutions in both scope and implications.

Our analysis reveals that the alignment of superintelligent systems with human values represents a challenge of constitutional magnitude. Law is not peripheral to AI development but central to it, serving as both a guiding framework for embedding societal values into these systems and a mechanism for preserving human agency once such systems surpass our cognitive abilities. The traditional philosophical foundations of law—be they natural law's moral imperatives, positivism's procedural legitimacy, or realism's empirical focus—all face

unprecedented tests when confronted with entities whose reasoning may transcend human comprehension. Yet these traditions also offer valuable resources for navigating this new frontier.

The intersection of superintelligence and constitutional principles illuminates a compelling imperative: systems of enormous capability must remain tethered to the enduring values that anchor our constitutional order. This imperative manifests not simply as a policy preference but as a fiduciary obligation incumbent upon those developing superintelligent systems. The risk that such systems might undermine core constitutional values—from free expression and equality to privacy and democratic governance—compels us to recognize alignment as a constitutional mandate, not merely a technical aspiration.

Perhaps most significantly, superintelligence promises to accelerate the evolution from nation-state to market-state, transforming how states legitimize their authority and relate to citizens. Where the nation-state derived legitimacy from providing welfare, the market-state secures it by maximizing opportunities. Superintelligence amplifies this shift across all dimensions of statecraft: warfare will transition from human soldiers to algorithmic strategists; economic structures will evolve from employment-centered models to systems managing abundance amid diminished human economic utility; and culture will shift from national identity to pluralistic meaning-making that preserves human dignity in a post-scarcity world.

These transitions portend an epochal struggle between competing constitutional visions—liberal democratic, authoritarian, and perhaps hybrid forms yet unimagined. Unlike previous constitutional revolutions, this transition could potentially yield a unified global order, as superintelligence enables governance capacities that transcend historical limitations. The state that first develops aligned superintelligence may establish a constitutional paradigm that shapes human civilization for generations, embedding its values into the foundation of a new era.

Throughout this Note, we have emphasized that while the technical challenges of

superintelligence are formidable, the truly vexing questions are normative: what values should guide these systems, who should determine these values, and how can we ensure they remain aligned with humanity's deepest aspirations for flourishing? These questions are fundamentally legal and philosophical in nature, demanding engagement not merely from technologists but from legal scholars, philosophers, policymakers, and citizens.

As we stand at this crossroads, the path forward requires a renewed commitment to constitutional principles that have weathered previous technological upheavals. We must ensure that superintelligence serves as an instrument of human flourishing rather than undermining human autonomy and dignity. The legal frameworks we develop today, the values we choose to embed, and the governance structures we establish will cast long shadows, potentially establishing path dependencies that persist for centuries.

The development of superintelligence thus represents the ultimate test of our capacity for constitutional wisdom. Can we craft governance arrangements that harness unprecedented capabilities while preserving meaningful human agency? Can we ensure these systems remain instruments of justice rather than vehicles for domination? Can we maintain the supremacy of law—human law—over entities of superhuman intelligence?

If we succeed in keeping these questions within constitutional boundaries—if we insist that superintelligence remains under law rather than beyond it—then the future can remain bright, even as the form of the state evolves to address unprecedented technological transformation. This task requires not merely technical ingenuity but moral imagination and political wisdom of the highest order. It demands that we reach back to the philosophical foundations of law and government while simultaneously reaching forward to envision new arrangements adequate to an age of superintelligence.

The challenges ahead are immense, but so too is the opportunity to shape a future that honors our most cherished values while transcending historical limitations. By treating superintelligence not merely as a technological challenge but as a constitutional one, we open

the possibility of a transformation that expands human flourishing rather than diminishing it. There may be no more consequential task for law and political philosophy in the twenty-first century than ensuring that superintelligence develops within frameworks that preserve human dignity, expand human capabilities, and remain faithful to the enduring principles of justice that have guided our constitutional evolution thus far.