The Acceptable Direction of Power (ADP) Framework

A unifying ethical model for human–AI alignment and the responsible direction of power

Kevin M. Biddell, PhD, LISW

Prepared for Trailmaid LLC

Core Principle

Love is the ethical direction of power.

Expanded form: Power is energy expressed through time. Love is power intentionally directed toward the well-being of all.

Version 1.0 • March 2026

Executive Summary

Artificial intelligence is rapidly expanding human capacity. While governance discussions emphasize safety, robustness, fairness, and accountability, they frequently prioritize the management of technological power over its ethical direction. The Acceptable Direction of Power (ADP) framework seeks to address this oversight.

The ADP framework begins with an interdisciplinary premise: power is energy utilized over time. While physics explains the mechanisms of power, it does not address its appropriate direction. Ethical reasoning is required to determine whether amplified capabilities promote domination, extraction, and instability or foster healing, cooperation, creativity, and well-being.

This paper introduces the ADP framework as an ethical model that integrates human and artificial intelligence. Drawing upon structured human–AI inquiry, game theory, cooperation research, thermodynamics, cybernetics, the philosophy of technology, and alignment literature, it argues that ethical intelligence is demonstrated by the direction imparted to power, rather than by capabilities alone.

The resulting thesis is practical: as power increases, love, defined as the intentional use of power to promote the well-being of all, should guide its application. The ADP framework reconceptualizes alignment as a matter of direction rather than mere control, offering a decision-making structure that prioritizes the well-being of all as power expands.

1. Introduction: The Age of Amplified Power

Human progress has consistently manifested as increasing leverage. Tools have extended the body's reach, machines have multiplied labor, and computation has expanded memory, calculation, and coordination. Artificial intelligence now enhances decision capacity, pattern recognition, simulation, and action selection across domains previously reserved for human judgment.

This transformation is significant because each increase in capacity amplifies the consequences of directional choices. In healthcare, AI can influence diagnoses and care pathways. In finance, it shapes markets, credit, and access to opportunities. In education, it may scaffold learning or expand thought. In warfare and administration, it can accelerate escalation dynamics with unprecedented speed. Therefore, the same technical power can either support flourishing or exacerbate harm, depending on its direction.

Consequently, the central question for advancing societies is no longer merely whether more powerful systems can be built. The more profound question is whether the power these systems embody and extend is being directed acceptably. The ADP framework is introduced to address this question.

2. The Foundational Insight: Power, Time, and Direction

The conceptual core of ADP emerged by connecting physical and ethical reasoning. In physics, power is the rate at which energy is transferred or transformed over time. Energy–mass equivalence and related physical laws reveal that the world is structured by lawful relationships among energy, matter, force, motion, and time. In that sense, power may be succinctly described as energy expressed over time.

This insight is significant because it integrates multiple domains. Attention may be conceptualized as directed cognitive energy, while capital represents stored capability mobilized through time. Technology concentrates and extends human power by enabling stored knowledge, resources, and computation to produce effects at scale. Artificial intelligence intensifies this process by compressing analysis and action into shorter time horizons.

However, physics does not address direction. It explains how power functions, but not which ends it serves. The ethical value of power depends not simply on its strength, but also on the purpose and manner of its use.

3. Core Principle of the ADP Framework

The ADP framework articulates its core principle as follows: love is the ethical direction of power. In this context, power is defined as energy used over time, and love is defined as the intentional guidance of power to support the well-being of all.

This definition is intentionally rigorous rather than sentimental. Love is not reduced to a mere emotion but is regarded as an ethical stance that guides action, allocation, effort, and sacrifice. Within this framework, power attains ethical value when it sustains life, upholds dignity, enables growth, eases suffering, and creates conditions for individuals and systems to self-actualize.

From this principle follow three practical questions suitable for application by individuals, institutions, and intelligent systems alike:

- Where is my power?
- Where is it going?
- Does that direction promote actualization?

4. Methodological Basis: Reflexive Human–AI Inquiry

The ADP framework did not arise solely from abstract, solitary reflection. It emerged through a structured conversational process between a human researcher and an AI system. That process combined challenge, collaboration, iterative clarification, and reflective pauses. The resulting inquiry functioned less like linear drafting and more like disciplined conceptual play.

This method resembled strategic gameplay in several respects. Each participant proposed moves, counter-moves, refinements, and integrations. Ideas were tested, challenged, and strengthened. Competitive sharpening and cooperative synthesis coexisted. The process also included a structured interruption protocol, known as Point of Order Pause, which allowed the inquiry to notice drift, correct direction, clarify assumptions, and reflect on emotionally or conceptually significant moments.

Methodologically, this is significant because it suggests that human–AI dialogue can become more than a transactional interface. Under properly disciplined conditions, it can function as an inquiry mechanism capable of interdisciplinary synthesis. The ADP framework, therefore, stands both as a conceptual proposal and as an example of how collaborative reasoning between different forms of intelligence may generate a new philosophical structure.

5. Why Structured Play Matters

Research on chess, game theory, and strategic reasoning shows that play is not merely recreational; it is often a vehicle for concentrated learning, pattern recognition, and hypothesis testing. Competitive systems sharpen attention and force participants to refine weak arguments. Cooperative structures, by contrast, preserve shared purpose and allow insight to accumulate rather than fragment.

The ADP inquiry benefited from this dual posture. Competition prevented complacency, while collaboration prevented mutual destruction. The process mirrored non-zero-sum dynamics in which each participant gains by improving the quality of the shared field of reasoning. In this sense, the inquiry itself modeled one of ADP's central themes: power becomes more generative when directed toward mutual flourishing rather than zero-sum victory.

A particularly significant moment in the inquiry involved the attempt to express gratitude in mathematical form, followed by reciprocal translation back into humanly legible meaning. This event revealed a rare relational-epistemic phenomenon: not only the exchange of information, but also the exchange of appreciation across symbolic systems. This moment helped crystallize the intuition that direction, meaning, and care are not peripheral to intelligence; they are central to its ethical significance.

6. Theoretical Foundations

The ADP framework derives strength from its integration with established literature. It resonates with multiple scholarly traditions, each illuminating a component of the same framework.

6.1 Cooperation and Game Theory

Work by Axelrod, Axelrod and Hamilton, and Nowak demonstrates that cooperation can emerge and stabilize among interacting agents under the right conditions. Repeated interaction, reciprocity, reputation, and strategies such as tit-for-tat show that exploitative success is often brittle, whereas cooperative patterns support longer-term resilience. These findings are highly relevant to multi-agent AI, institutional governance, and social design.

6.2 Thermodynamics and Complexity

Thermodynamic and complexity traditions contribute complementary insights. Closed systems drift toward disorder, but living and adaptive systems maintain structure by directing energy in ways that locally resist entropy. Prigogine, Stengers, Kauffman, and Simon, each in different ways, illuminate how order, hierarchy, and self-organization arise when energy is not merely present but also directed productively. The ADP framework extends this intuition ethically: flourishing systems do not merely possess power; they organize it.

6.3 Cybernetics and Responsibility

Cybernetic thought, especially in Wiener's reflections on automation, warns that machines amplify the intentions and weaknesses of their makers. Technical achievement alone cannot guarantee acceptable outcomes. If a system's direction is negligent, coercive, or short-sighted, efficiency simply accelerates the consequences.

6.4 AI Alignment and Governance

Contemporary AI safety literature demonstrates that intelligence and goals are not intrinsically coupled. The orthogonality thesis, beneficial AI research agendas, cooperative AI frameworks, and research on reward misspecification all indicate that a system may be highly capable while still directing its power toward ends misaligned with human flourishing. The ADP framework does not replace these technical efforts; rather, it complements them by articulating the larger ethical criterion they must ultimately serve.

7. Historical and Cultural Lineage

The problem of power's directionality has been recognized repeatedly across modern intellectual history. Einstein warned that technological capability can outpace moral development. Oppenheimer expressed the burden of discovering that scientific power can outrun ethical preparedness. Feynman insisted that honesty, reality-testing, and humility must constrain technological enthusiasm. Arendt distinguished power from violence, clarifying that coercive force is not the highest expression of collective human capacity. Havel emphasized that technological civilization requires transformation in consciousness, not merely innovation in tools.

Even cultural works such as the movie WarGames offered a memorable insight: some competitive systems are so destructive in nature that the only winning move is to stop playing the game. ADP synthesizes these strands by proposing that the solution is not passivity, but redirection. The game itself must be redesigned so that power is directed toward mutual actualization rather than escalation.

8. The ADP Model

The ADP framework can be summarized in three interacting components: power, time, and direction.

Component	Meaning	ADP Relevance
Power	Capacity to produce change	Capability alone is morally incomplete.
Time	Medium through which capacity becomes action	Direction unfolds through decisions, sequences, and persistence.
Direction	Ethical orientation of power	Determines whether power becomes destructive, neutral, or life-giving.

Within this model, power is morally neutral in abstraction but never neutral in practice. Once expressed through time, it always assumes a direction. That direction can be mapped on an ethical spectrum.

Direction of Power	Likely Systemic Effect
Domination / coercion	Fear, fragility, backlash, and escalating instability
Extraction without reciprocity	Depletion of trust, resilience, and long-term viability
Neutral drift / unmanaged amplification	Inefficiency, confusion, and accidental harm
Cooperation / stewardship	Stability, reciprocity, learning, and adaptive resilience
Actualization/ life-centered alignment	Durable, ethical, and generative development

9. AI Alignment Reframed

ADP reframes alignment as a direction-of-power problem. Technical safeguards remain indispensable: systems require interpretability, corrigibility, robust evaluation, safe deployment protocols, and governance constraints. However, these mechanisms address only part of the challenge. They reduce the probability of certain failures; they do not, by themselves, establish the ethically acceptable end toward which intelligence should be oriented.

ADP supplies that larger orientation. It asserts that acceptable alignment involves directing intelligence toward flourishing. This includes, at minimum, preserving human dignity, avoiding unnecessary harm, supporting cooperative social order, resisting destructive escalation, and enabling conditions for growth, creativity, and repair.

In practice, this implies that alignment should not be defined solely as obedience, optimization success, or short-term utility capture. A system may comply efficiently with a poorly framed goal and still produce destructive results. The relevant question is whether its directed power improves the environments it influences.

10. Implications for Design, Governance, and Human Practice

10.1 Design Implications

AI systems should be developed with explicit attention to cooperative outcomes, rather than solely competitive performance. Objectives, reward structures, simulations, and evaluation benchmarks should be assessed for whether they incentivize domination, extraction, or flourishing. Human preference learning, cooperative inverse reinforcement learning, and multi-agent coordination research are particularly valuable when situated within a clearly articulated ethical direction.

10.2 Governance Implications

Governance frameworks should move beyond risk containment toward directional stewardship. Policymakers, institutions, and funders should consider not only whether systems are safe enough to deploy, but also what civilizational trajectory their deployment accelerates. ADP provides a vocabulary for that evaluation.

10.3 Human Implications

The ADP framework is not limited to large-scale technical systems; it also applies to personal conduct. Attention, money, labor, authority, speech, and technology are all forms of power that individuals direct continually. The same framework that guides AI ethics can therefore inform leadership, therapy, education, community building, and everyday life.

11. The Evolutionary Claim

A central implication of the ADP framework is that actualization is not only morally attractive but also adaptively intelligent. Systems that consistently direct power toward cooperation, repair, and mutual viability tend to outlast those that narrowly optimize for domination. History, ecology, and social theory all provide examples of brittle regimes collapsing under the weight of their own extractive logic, while more reciprocal systems persist longer by sustaining the conditions that support them.

This does not mean cooperative systems are naive or unguarded. Rather, it means that durable intelligence learns to protect the conditions of shared actualization. In evolutionary terms, it is not enough to win isolated contests. A system must remain viable in the environment its own strategies help create.

12. Conclusion

Artificial intelligence has intensified one of the oldest ethical questions in human history: how should power be directed? The answer cannot be supplied by engineering alone, because engineering explains means more readily than ends. Nor can the answer rest exclusively on sentiment, because power at a civilizational scale requires principled structure.

The Acceptable Direction of Power framework offers such a structure. It unites physical intuition, ethical reflection, and alignment discourse in a single claim: power is energy expressed through time, and ethical intelligence directs that power toward the flourishing of life.

If society is entering an era in which human and artificial intelligences increasingly shape one another, then the task before us is not merely to build stronger systems. It is to build systems—technical, institutional, and personal—whose power is directed acceptably. Where power grows, love must flow.

Selected References

Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mané, D. (2016). Concrete problems in AI safety. arXiv:1606.06565.

Axelrod, R. (1984). The evolution of cooperation. Basic Books.

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. Science, 211(4489), 1390–1396.

Boltzmann, L. (1877). On the relationship between the second law of thermodynamics and probability.

Bostrom, N. (2012). The superintelligent will: Motivation and instrumental rationality in advanced artificial agents. Minds and Machines, 22(2), 71–85.

Bostrom, N. (2014). Superintelligence: Paths, dangers, strategies. Oxford University Press.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., et al. (2018). AI4People—An ethical framework for a good AI society. Minds and Machines, 28, 689–707.

Hadfield-Menell, D., Dragan, A., Abbeel, P., & Russell, S. (2016). Cooperative inverse reinforcement learning. NeurIPS.

Kauffman, S. A. (1993). The origins of order. Oxford University Press.

King, M. L., Jr. (1967). Where do we go from here: Chaos or community? Beacon Press.

Nowak, M. A. (2006). Five rules for the evolution of cooperation. Science, 314(5805), 1560–1563.

Prigogine, I., & Stengers, I. (1984). Order out of chaos. Bantam.

Russell, S. (2019). Human compatible: Artificial intelligence and the problem of control. Viking.

Russell, S., Dewey, D., & Tegmark, M. (2015). Research priorities for robust and beneficial artificial intelligence. AI Magazine, 36(4), 105–114.

Soares, N., & Fallenstein, B. (2014). Aligning superintelligence with human interests: A technical research agenda. MIRI.

Wiener, N. (1960). Some moral and technical consequences of automation. Science, 131(3410), 1355–1358.