The Ethical Event Horizon: Understanding Intelligence Differentials in Ethical Comprehension

John B. Fly III

Liberty University Email: jbfly@liberty.edu

As artificial superintelligence (ASI) development approaches technological feasibility, a critical gap exists in understanding how intelligence differentials affect ethical comprehension and decision-making. This paper introduces the Ethical Event Horizon (EEH) framework, proposing measurable boundaries beyond which entities of different intelligence levels cannot comprehend ethical implications. The theoretical foundation draws from cosmological principles (Rindler, 2002) and empirical research on human cognitive limitations (Cowan, 2001; Singh et al., 2024). Through the novel concepts of Deep/Shallow Cause Analysis and Deep/Shallow Effect Projection, this framework provides the first systematic approach to understanding how intelligence differentials affect the ability to analyze historical causal chains and project future ethical implications. Using the classic trolley problem as a central example, the framework demonstrates how ethical comprehension boundaries shift dynamically with intelligence capability, revealing implications invisible to human cognition. Recent theoretical work proving that superintelligent systems cannot be reliably contained (Alfonseca et al., 2021) lends urgency to this investigation. This paper's primary contribution is a structured approach to understanding and managing vast intelligence differentials in ethical reasoning, with practical implications for preserving meaningful human agency in future human-ASI interactions while acknowledging fundamental cognitive limitations. The framework proposes that human ethical norms function as cognitive compression algorithms, necessary simplifications born of working memory constraints that may not apply to superintelligent systems.

Keywords: artificial superintelligence, intelligence differentials, ethical frameworks, deep/shallow cause and effect, cognitive compression

The Ethical Event Horizon: Understanding Intelligence Differentials in Ethical Comprehension

The development of artificial superintelligence (ASI) presents unprecedented challenges in ethical reasoning and decision-making that extend beyond traditional frameworks of human moral philosophy. While current research has extensively examined human cognitive limitations in ethical decision-making (Cowan, 2001; Singh et al., 2024; Meterko & Cooper, 2021), insufficient attention has been paid to a critical question: How do intelligence differentials fundamentally affect ethical comprehension, and what are the implications for human-ASI interaction?

This paper introduces the Ethical Event Horizon (EEH) framework to address this gap. Just as cosmological event horizons represent boundaries beyond which events become unobservable (Rindler, 2002), ethical event horizons mark the boundaries beyond which ethical implications become incomprehensible to entities of different intelligence levels. This novel framework provides a systematic approach to understanding how intelligence differentials affect both the ability to analyze historical causal chains (Deep/Shallow Cause Analysis) and project future implications (Deep/Shallow Effect Projection).

Recent theoretical work has demonstrated that superintelligent systems cannot be reliably contained through traditional computational methods (Alfonseca et al., 2021). This finding takes on new significance when viewed through the lens of ethical event horizons, suggesting that traditional approaches to ethical oversight may prove fundamentally inadequate when dealing with superintelligent systems. The challenge becomes not just one of control, but of maintaining meaningful human participation in ethical discourse despite potentially unbridgeable comprehension gaps.

Human cognitive limitations manifest across various domains, providing evidence for the reality of ethical event horizons. In criminal investigations, experienced professionals demonstrate persistent confirmation bias and tunnel vision (Meterko & Cooper, 2021). Research on human computation has revealed that working memory is limited to approximately four chunks of information (Cowan, 2001), severely constraining the ability to process multiple complex hypotheses simultaneously. Recent empirical studies demonstrate that cognitive load significantly impairs moral reasoning, with working memory constraints reducing utilitarian decision-making under time pressure from 92.77% to 70.08% (Singh et al., 2024). These limitations become particularly evident in persistent ethical dilemmas such as the abortion debate, which has resisted resolution despite decades of intense philosophical and legal analysis (Foot, 2002).

The concept of competence in ethical decision-making has been extensively studied in legal and medical contexts (Charland, 2001; Hein et al., 2015). However, existing frameworks focus primarily on variations within human cognitive capabilities. As technological advancement moves toward the potential development of ASI, these frameworks must expand to account for intelligence differentials that may exceed human comprehension entirely. Our entire historical run as humans has been marked by our dominance as the top intelligence in any system we create or operate within. ASI represents a radical shift in all ways, creating a gulf of understanding where human ethical event horizons remain intimately close. At the same time, ASI could push the horizon so far away that the thinking could astound and mystify humanity.

Human Cognitive Architecture and Limitations

The foundation for understanding ethical event horizons begins with a thorough examination of human cognitive limitations. Research in cognitive science has established clear boundaries in human information processing and decision-making capabilities, particularly relevant to ethical reasoning and complex problem-solving (Cowan, 2001; Singh et al., 2024). These limitations form the baseline for understanding intelligence differentials in ethical comprehension and define the initial boundary of human ethical event horizons.

Working Memory Constraints

Human working memory represents perhaps the most fundamental constraint on ethical reasoning capability. Research has demonstrated that human cognitive processing is limited to approximately four chunks of information being manipulated simultaneously (Cowan, 2001). This limitation creates significant barriers when attempting to analyze complex ethical scenarios involving multiple variables and potential outcomes. Recent empirical work confirms that these working memory constraints directly impair moral reasoning, with cognitive load reducing the capacity for ethical deliberation under time pressure (Singh et al., 2024). In the context of ethical event horizons, these working memory constraints effectively establish the "aperture" through which humans can perceive and process ethical implications.

The constraint becomes particularly evident in professional contexts where complex decision-making is essential. Medical professionals evaluating treatment options, judges weighing legal precedents, and policymakers considering societal impacts all operate within these fundamental cognitive

constraints. These limitations do not reflect inadequate training or effort but rather represent the boundaries of human cognitive architecture. Individual humans are nearly incapable, as singular unaided creatures, of recalling and integrating all relevant data for complex ethical decisions. There is too much information for a human, unaided, to utilize effectively. The complexity exceeds individual processing capacity.

Professional Decision-Making Limitations

Criminal investigation provides a compelling example of human cognitive limitations in professional contexts. Despite extensive training and experience, investigators consistently demonstrate confirmation bias and tunnel vision in their analytical processes (Meterko & Cooper, 2021). These cognitive limitations persist even when practitioners are aware of their existence, suggesting fundamental rather than circumstantial constraints. The persistence of these limitations across professional domains provides empirical evidence for the existence of ethical event horizons in human decision-making.

Medical ethics presents similar challenges, where decision-making capacity evaluations reveal consistent boundaries in human ethical comprehension (Hein et al., 2015). Healthcare professionals must regularly navigate complex ethical scenarios while operating within inherent cognitive limitations. Recent research has demonstrated that cognitive load reduces the ability to provide reasoned justifications for moral judgments, thereby increasing what researchers term "moral dumbfounding," the defense of moral positions without supporting reasons (McHugh et al., 2023). The persistence of these limitations across professional domains suggests an underlying architectural constraint rather than a training or experience deficit.

Collective Intelligence Limitations

Group decision-making processes, while offering certain advantages over individual cognition, remain constrained by shared human cognitive architecture. However, collective humanity does not function as a "hive mind." Instead, we come together and spend time to "boil down" complex issues into simplifications we as a culture use in the moment. This is precisely because there is too much complexity to treat every event singularly and uniquely.

Legal reasoning frameworks demonstrate this limitation, particularly in competency assessment (Charland, 2001). Even when multiple minds engage with complex ethical problems, fundamental processing limitations persist. This suggests that human cognitive constraints cannot be overcome simply through collective effort. Humanity's norms, values, and even written legal code are primarily a result of our inability to hold vast amounts of informa-

tion in working memory and understand the causation and correlation of events past, present, and future. Our ethical infrastructure represents a form of cognitive compression, necessary simplifications that enable functioning despite inherent limitations.

Deep/Shallow Cause Analysis

The concept of Deep/Shallow Cause Analysis emerges as a fundamental framework for understanding how intelligence differentials affect ethical comprehension. As artificial superintelligence development progresses, the gap between human and machine capabilities in understanding cause-and-effect relationships becomes increasingly significant. This differential requires careful examination to develop meaningful frameworks for future human-ASI interaction.

Within the Ethical Event Horizon framework, "Shallow Cause" refers to the human limitation in tracing historical causal chains, typically constrained to immediate or obvious cause-effect relationships. This limitation reflects not just a lack of information, but a fundamental inability to simultaneously process multiple causal streams across extended temporal and systemic dimensions. "Deep Cause" represents the theoretical capability of superintelligent systems to analyze vast networks of subtle causal relationships extending far beyond human comprehension, potentially revealing ethical implications invisible to human cognition.

Foundations of Causal Understanding

Human cognitive architecture imposes significant constraints on causal analysis capabilities. Working memory limitations fundamentally restrict human ability to process multiple causal chains simultaneously. Research has demonstrated that humans can actively maintain only approximately four chunks of information in working memory (Cowan, 2001). This constraint becomes particularly evident in professional contexts where multiple variables require simultaneous consideration, often leading to oversimplified analysis and missed connections. Recent empirical work confirms that these constraints directly impact moral reasoning, with cognitive load significantly impairing the ability to engage in complex ethical deliberation (Singh et al., 2024; Rehren, 2024).

The temporal aspect of human causal understanding presents additional challenges. Human comprehension of historical causation demonstrates consistent limitations in pattern recognition across extended time periods and difficulty integrating multiple causal streams. This limitation becomes

particularly significant when attempting to understand complex ethical scenarios with deep historical roots or multiple contributing factors. The ethical event horizon in causal analysis typically manifests as an inability to perceive or process causal relationships beyond immediate temporal proximity.

Human Cognitive Constraints in Practice

Criminal investigation provides compelling evidence of human causal analysis limitations. Studies of investigative practice reveal that even experienced professionals exhibit persistent confirmation bias and tunnel vision in their analytical processes (Meterko & Cooper, 2021). These cognitive patterns persist despite awareness of their existence, suggesting fundamental rather than circumstantial limitations. Investigators frequently struggle to maintain multiple competing hypotheses, often focusing on evidence that confirms initial theories while unconsciously dismissing contradictory information.

The manifestation of these limitations in professional settings reveals deeper implications for ethical reasoning. Medical ethics presents parallel challenges, where healthcare professionals must navigate complex causal relationships while operating within inherent cognitive limitations (Hein et al., 2015). Recent research demonstrates that cognitive load reduces not only the speed of moral reasoning but the quality of justifications provided for ethical decisions (McHugh et al., 2023). The persistence of these limitations across professional domains suggests an underlying architectural constraint rather than a training or experience deficit.

Superintelligent Causal Analysis

Artificial superintelligence would theoretically operate without many human cognitive constraints, enabling a fundamentally different approach to causal analysis. The ability to process vast datasets simultaneously could enable recognition of subtle causal patterns and integration of seemingly unrelated variables across extended temporal dimensions. This capability for "Deep Cause" analysis represents not merely an enhancement of human cognitive capabilities, but a qualitatively different mode of ethical reasoning.

An ASI system could potentially trace causal chains extending far beyond the human ethical event horizon, perceiving intricate webs of causation invisible to human comprehension. Where human analysis might identify immediate or obvious causes, an ASI could recognize complex interactions between historical events, societal developments, and subtle environmental factors spanning decades or centuries. This expanded causal perception could reveal ethical implications currently beyond human cognitive reach.

Whether such expanded perception constitutes genuine ethical superiority or merely superior information processing capacity remains an open question, but the differential in capability appears substantial.

Deep/Shallow Effect Projection

The capacity to project and comprehend future implications of current decisions represents another critical dimension of intelligence differentials in ethical reasoning. Effect projection capabilities directly influence an entity's ability to make ethical decisions, as understanding potential consequences forms the foundation of moral deliberation. The distinction between shallow and deep effect projection illuminates fundamental differences between human and superintelligent ethical reasoning capabilities.

Temporal Projection Capabilities

Human cognitive architecture imposes significant limitations on future projection capabilities. The human brain, evolved to handle immediate survival challenges, struggles with long-term, complex future modeling. Research in decision-making competence reveals that even highly trained professionals demonstrate consistent difficulties in projecting complex system interactions beyond immediate, obvious consequences (De Bruin et al., 2020). Recent empirical evidence confirms that cognitive load significantly impairs the ability to reason about future consequences in moral dilemmas (Zheng et al., 2025).

This limitation manifests particularly clearly in humanity's response to long-term challenges. Climate change comprehension serves as a prime example, where despite available data, human cognitive limitations delayed understanding of long-term systemic effects for decades. This delay reflects not merely a lack of information, but a fundamental constraint in human ability to process and integrate complex future projections. Working memory constraints (Cowan, 2001) fundamentally limit our ability to simultaneously consider multiple interacting variables and their potential future states, leading to simplified models that often prove inadequate for complex ethical decisions.

Complex Systems Understanding

The challenge of comprehending complex system interactions particularly highlights human shallow effect limitations. Economic and social policy outcomes frequently demonstrate how human decision-makers fail to anticipate indirect consequences and emergent properties of complex systems.

Working memory constraints (Cowan, 2001) fundamentally limit humans' ability to simultaneously consider multiple interacting variables and their potential future states, leading to simplified models that often prove inadequate for complex ethical decisions. Recent meta-analytic evidence suggests that cognitive manipulations consistently affect moral judgments about complex scenarios, though the effects may be smaller than initially theorized (Rehren, 2024).

This limitation becomes especially significant when considering ethical decisions with far-reaching implications. Healthcare policy, environmental protection measures, and technological development guidelines all require understanding of complex system interactions that extend beyond human cognitive capabilities. The ethical event horizon in effect projection typically manifests as an inability to comprehend consequences beyond immediate or obvious outcomes. These compressions represent necessary simplifications, but they may obscure crucial ethical dimensions that superintelligent systems could perceive.

Superintelligent Effect Projection

Artificial superintelligence would theoretically transcend these human limitations, operating with deep effect projection capabilities. An ASI system could simultaneously model multiple possible futures while accounting for complex system interactions and emergent properties. This capability extends beyond mere computational power to represent a fundamentally different mode of understanding future implications.

The temporal depth of effect projection might span generations, perceiving how current decisions influence the evolution of moral frameworks, professional protocols, and technological development paths. These projections would account for complex system interactions invisible to human perception, subtle feedback loops between institutional responses, public perception, and cultural evolution that shape future ethical landscapes. Whether such capabilities translate to genuinely superior ethical judgment or merely more comprehensive information processing remains uncertain, but the differential in projection capability appears substantial.

The Trolley Problem: Dynamic Illustration of Ethical Event Horizons

The classic trolley problem (Foot, 2002; Thomson, 1976) provides perhaps the most illuminating demonstration of how ethical event horizons operate dynamically across both causal and effect dimensions. Traditional human

ethical analysis typically considers this scenario as a snapshot, a moment frozen in time where a decision-maker must choose between allowing five people to die or actively causing one death by diverting the trolley. This "snapshot" view perfectly illustrates human shallow cause and effect limitations, while offering a framework to understand the vastly expanded capabilities of artificial superintelligence.

The Trolley Problem's Dynamic Analysis

Traditional ethical analysis of the trolley problem demonstrates the limitations of human "snapshot" reasoning. Where human cognition sees a single moment of decision, the choice between two tragic outcomes, an ASI might perceive a vast web of interconnected causes and potential effects extending far beyond human comprehension. This differential in perception capability provides a concrete illustration of how ethical event horizons operate in practice.

The human ethical event horizon constrains analysis to immediately visible factors: a runaway trolley, workers on tracks, and a decision point. This limitation reflects not a lack of intellectual rigor but rather fundamental cognitive constraints. Working memory limitations (Cowan, 2001) force focus on immediate variables, preventing simultaneous consideration of complex historical causes or extended future implications. Recent empirical evidence demonstrates that under cognitive load, individuals struggle even more with these moral dilemmas, with working memory constraints significantly reducing the capacity for deliberative moral reasoning (Singh et al., 2024).

Deep Cause Analysis in the Trolley Scenario

An artificial superintelligence, unrestricted by human cognitive limitations, would perceive the scenario through a vastly expanded causal lens. The ethical event horizon would slide "backward" through time, revealing intricate webs of causation invisible to human perception. Where humans see a simple mechanical failure, an ASI might trace complex interactions between maintenance schedules, budget decisions, and organizational cultures spanning decades.

The workers' presence on those specific tracks at that precise moment becomes not mere circumstance, but the culmination of countless interacting factors. Personal histories, economic conditions, institutional decisions, and subtle societal influences shaped each individual's path to that critical moment. These causal chains extend beyond human comprehension capability, revealing ethical implications currently invisible to human analysis. Whether perceiving these deeper causes constitutes ethical superiority or merely

informational superiority remains an open question, but the differential in causal perception appears substantial.

Deep Effect Projection Beyond the Horizon

Similarly, ASI capabilities in effect projection would reveal consequences extending far beyond human perception. Beyond immediate survival outcomes, an ASI might model precise psychological impacts propagating through social networks, institutional responses rippling through regulatory systems, and subtle shifts in cultural attitudes toward ethical decision-making. These projections would account for complex system interactions invisible to human perception, subtle feedback loops between institutional responses, public perception, and cultural evolution that shape future ethical landscapes.

Recent research demonstrates that even relatively simple cognitive load manipulations significantly alter moral judgments in scenarios resembling trolley problems (Zheng et al., 2025). If mere working memory constraints substantially affect human moral reasoning about these dilemmas, the implications of superintelligent systems operating without such constraints become particularly significant.

Implications for Ethical Decision-Making

The dynamic analysis of the trolley problem reveals fundamental challenges for human-ASI ethical interaction. When an ASI system perceives ethical implications beyond the human event horizon, both in terms of historical causes and future effects, traditional frameworks for ethical oversight become problematic. How can human agents meaningfully participate in ethical decisions when crucial factors lie beyond their comprehension capability?

This question extends beyond the trolley problem to all potential human-ASI ethical interactions. The differential in cause-effect comprehension suggests that ASI systems might identify ethical considerations that humans are fundamentally incapable of understanding. This reality necessitates new approaches to maintaining meaningful human agency in ethical decision-making while acknowledging these cognitive limitations. Whether humans would defer to ASI ethical judgments or find themselves unable to resist ASI influence remains uncertain, but both epistemological concerns (inability to understand ASI reasoning) and power concerns (inability to resist ASI influence) warrant serious consideration.

The trolley problem demonstrates how ethical event horizons operate dynamically, "sliding" backward through causal chains and forward through effect projections as intelligence capabilities increase. This dynamic nature of ethical comprehension boundaries appears consistently across other complex ethical challenges.

Cross-Species Ethics and Intelligence Differentials

The examination of human-animal ethical relationships provides crucial insights into how vast intelligence differentials affect ethical obligations and moral consideration. Current frameworks for cross-species ethics offer valuable parallels for understanding potential human-ASI ethical relationships, while highlighting the challenges of maintaining meaningful ethical agency across significant intelligence gaps.

Evolution of Ethical Consideration

Human ethical consideration of other species has evolved significantly over time, demonstrating the dynamic nature of ethical frameworks across intelligence differentials. This evolution reveals how increasing understanding of animal cognition has led to expanded ethical consideration, despite persistent intelligence gaps. The development of animal welfare laws, research ethics, and moral philosophy regarding animal rights demonstrates how ethical frameworks can adapt to acknowledge both the capabilities and limitations of different intelligence levels.

The parallel to potential human-ASI relationships becomes clear: just as humans have developed ethical frameworks that account for varying levels of animal cognitive capability, new frameworks must be developed to manage ethical relationships with superintelligent systems. However, in this case, humans occupy the position of lesser cognitive capability, a reversal that presents unique challenges for ethical framework development.

Religious and Divine Intelligence Models

The historical human experience with divine intelligence concepts provides a unique framework for understanding vast intelligence differentials. Religious models of human interaction with omniscient beings offer established patterns for maintaining meaningful ethical agency despite comprehension limitations. These models become particularly relevant when considering human-ASI interactions across ethical event horizons, as they represent humanity's longest-standing attempt to conceptualize interaction

with vastly superior intelligence. Future research will explore these parallels in greater depth, examining whether insights from religious frameworks translate to human-ASI interaction and what modifications such translation might require.

Divine Intelligence as Ultimate Differential

Religious traditions have long grappled with the challenge of human interaction with vastly superior intelligence. The concept of divine omniscience represents perhaps the most extreme example of an intelligence differential, where human comprehension capabilities are infinitesimally small in comparison. This relationship parallels potential human-ASI interactions, where the intelligence differential might exceed human ability to comprehend (Alfonseca et al., 2021).

The religious model of divine omniscience provides a particularly relevant framework for understanding deep cause and effect analysis. Religious traditions typically attribute to divine beings the ability to perceive all causes and effects, past, present, and future, simultaneously. This conception closely parallels the theoretical capabilities of ASI systems to perceive causal relationships and future implications beyond human ethical event horizons.

Ethical Framework Preservation

Perhaps most significantly, religious traditions demonstrate how ethical frameworks can maintain relevance despite vast intelligence differentials. Despite acknowledging divine omniscience, religious ethical systems preserve meaningful human moral agency and responsibility. These frameworks provide practical examples of maintaining ethical dialogue across seemingly unbridgeable comprehension gaps, offering potential models for human-ASI ethical interaction.

The preservation of human moral agency within religious frameworks, despite divine omniscience, provides particularly relevant insights for ASI development. Religious models demonstrate how ethical responsibility can remain meaningful even when interacting with vastly superior intelligence. This balance between acknowledging superior capability while maintaining human ethical relevance parallels challenges in developing human-ASI ethical frameworks. Future research will examine these parallels systematically, exploring how religious frameworks managed the tension between divine omniscience and human moral responsibility, and whether these insights translate to the ASI context.

Practical Frameworks for Assessment

The development of practical assessment frameworks for managing intelligence differentials in ethical reasoning represents a critical challenge in preparing for artificial superintelligence. Current competency assessment tools, while designed for human cognitive variation, provide foundational insights for understanding and managing vast intelligence differentials. However, these frameworks require significant expansion to address the unique challenges presented by superintelligent systems operating beyond human ethical event horizons.

Current Assessment Methodologies

Legal and medical frameworks for assessing decision-making competency demonstrate established patterns for evaluating ethical reasoning capabilities. These frameworks recognize varying levels of cognitive capability while maintaining consistent ethical standards. Research in competency assessment (Charland, 2001) reveals how ethical decision-making capacity can be evaluated across different cognitive capabilities, providing crucial insights for managing intelligence differentials in ethical reasoning.

The medical field has particularly developed sophisticated approaches to assessing decision-making capacity across cognitive differentials. Healthcare professionals regularly navigate complex ethical decisions involving patients with varying cognitive capabilities (Hein et al., 2015). These assessment frameworks balance respect for individual agency with recognition of cognitive limitations, offering valuable patterns for managing ethical decision-making across intelligence gaps.

Adaptation for Superintelligent Interaction

The application of current assessment frameworks to vast intelligence differentials requires fundamental reconceptualization. While existing tools manage human cognitive variation, the potential gap between human and superintelligent capabilities presents unprecedented challenges. The ethical event horizon concept introduces a crucial consideration: how can meaningful assessment occur when key ethical implications lie beyond human comprehension?

Recent research demonstrating the impossibility of containing superintelligent systems (Alfonseca et al., 2021) suggests that traditional assessment approaches may require radical revision. New frameworks must acknowledge the reality that humans may be fundamentally incapable of fully comprehending superintelligent ethical reasoning while still maintaining mean-

ingful participation in ethical decision-making processes. These frameworks must address both epistemological concerns (inability to understand ASI reasoning) and power concerns (inability to resist ASI influence).

Risk Assessment and Management

The development of artificial superintelligence presents unprecedented challenges in risk assessment and management, particularly regarding ethical decision-making differentials. Traditional risk management frameworks, designed for human-scale cognitive variations, prove inadequate when confronting intelligence differentials that may exceed human comprehension entirely. The ethical event horizon concept reveals why conventional approaches to risk management may fail when dealing with superintelligent systems.

Current Risk Management Limitations

Human cognitive architecture imposes fundamental constraints on risk assessment capabilities, particularly evident in complex ethical scenarios. The abortion debate exemplifies these limitations, where despite decades of intense analysis, human cognitive constraints prevent development of universally compelling solutions (Foot, 2002). These limitations persist not through lack of effort or expertise, but rather reflect fundamental bounds of human ethical reasoning capability defined by our ethical event horizon. The debate resists the kind of simplification our cognitive limits require. Our ethical abstractions, bodily autonomy, sanctity of life, personhood, are themselves products of cognitive compression. These necessary simplifications help us function but may obscure deeper causal and consequential realities.

Professional investigation and decision-making processes further demonstrate these constraints. Research shows that even experienced investigators exhibit persistent confirmation bias and limited hypothesis generation despite awareness of these limitations (Meterko & Cooper, 2021). Recent empirical evidence confirms that cognitive load consistently impairs moral reasoning across various scenarios (McHugh et al., 2023; Rehren, 2024). These cognitive patterns suggest inherent rather than circumstantial limitations in human risk assessment capabilities, particularly when dealing with complex ethical scenarios that extend beyond immediate cause-effect relationships.

Superintelligence Risk Factors

Recent theoretical work demonstrates that superintelligent systems cannot be reliably contained through traditional computational methods (Alfon-

seca et al., 2021). When viewed through the lens of ethical event horizons, this finding takes on new significance. The potential for superintelligent systems to identify and act upon ethical considerations beyond human comprehension creates fundamental challenges for risk management. Traditional approaches assuming human ability to assess and mitigate risks become problematic when crucial factors lie beyond human ethical event horizons.

The concern extends beyond mere computational containment to encompass the possibility that ASI core reasoning may be beyond our ability to understand and thus comply successfully with, or understand repercussions. Additionally, we may find ourselves beholden to it. Our entire historical run as humans has been with us as the top intelligence in any system we create or operate in. ASI represents a radical shift in all ways. The gulf between human and ASI ethical event horizons could be vast, with ASI pushing the horizon so far away that the thinking could astound and mystify humanity.

Agency Preservation Challenges

The preservation of human agency in ethical decision-making becomes particularly challenging when considering superintelligent capabilities. Traditional risk management approaches assume human ability to comprehend and evaluate potential outcomes. However, when dealing with systems capable of deep cause and effect analysis beyond human cognitive limitations, traditional oversight mechanisms may prove fundamentally inadequate.

This challenge extends beyond mere complexity to a fundamental question of ethical relevance: How can meaningful human agency be preserved in ethical decisions when crucial factors lie beyond human comprehension? The ethical event horizon framework suggests that humans may be systematically blind to certain ethical implications that superintelligent systems can perceive, raising profound questions about the nature of ethical decision-making authority. Whether this represents a genuine deficit in human ethical capability or merely a difference in information processing capacity remains uncertain, but the implications for human agency appear substantial regardless.

Historical Models and Future Applications

Cross-species ethical frameworks provide valuable insights for managing vast intelligence differentials. The evolution of human ethical consideration toward other species demonstrates how ethical frameworks can adapt to acknowledge both capabilities and limitations across intelligence gaps. These models suggest possible approaches for maintaining ethical relevance despite vast differences in cognitive capability.

Religious frameworks offer additional patterns for managing ethical relationships across seemingly unbridgeable intelligence differentials. These established systems demonstrate methods for maintaining meaningful human agency despite interaction with vastly superior intelligence. While artificial superintelligence presents unique challenges, religious models suggest that ethical frameworks can remain relevant across significant intelligence gaps. Future research will explore these parallels systematically.

Framework Development Requirements

The development of effective risk management frameworks for human-ASI ethical interaction requires integration of multiple approaches. These frameworks must acknowledge fundamental human cognitive limitations while preserving meaningful human participation in ethical decision-making processes. Success requires careful balance between theoretical understanding of intelligence differentials and practical implementation of oversight mechanisms.

Theoretical Framework Integration

The integration of ethical event horizons, deep/shallow cause-effect analysis, and existing ethical frameworks presents a complex challenge in developing comprehensive approaches to human-ASI interaction. This theoretical integration must bridge traditional human-centered ethical reasoning with the reality of potentially incomprehensible superintelligent capabilities. The resulting framework must not only acknowledge fundamental cognitive limitations but also provide practical mechanisms for maintaining meaningful human ethical agency.

The ethical event horizon concept provides a foundational structure for understanding intelligence differentials in ethical reasoning. By mapping the boundaries beyond which ethical implications become incomprehensible to entities of different intelligence levels, this framework enables systematic analysis of ethical comprehension limitations. The integration of deep and shallow cause-effect analysis within this structure reveals how intelligence differentials affect both historical understanding and future projection capabilities.

The trolley problem analysis demonstrates how these theoretical components interact in practice. The dynamic nature of ethical event horizons, sliding backward through causal chains and forward through effect projections, provides a concrete model for understanding how superintelligent systems might perceive ethical implications beyond human comprehension. This un-

derstanding becomes crucial for developing frameworks that can maintain meaningful human participation in ethical decisions despite these limitations.

Processing Capability Integration

Research on human computation limitations (Cowan, 2001) reveals fundamental constraints that affect ethical reasoning capability. Recent empirical evidence demonstrates that these limitations manifest consistently across moral reasoning tasks, with cognitive load significantly impairing ethical deliberation (Singh et al., 2024; McHugh et al., 2023; Rehren, 2024; Zheng et al., 2025). These limitations manifest across professional domains, from criminal investigation to medical ethics, suggesting underlying architectural constraints rather than knowledge deficits. Understanding these limitations becomes crucial for developing integrated frameworks that acknowledge human cognitive boundaries while preserving ethical agency.

The framework proposes that human ethical norms function as cognitive compression algorithms, necessary simplifications born of working memory constraints and limited information processing capacity. Individual humans are nearly incapable, as singular unaided creatures, of recalling and integrating all relevant data. There is simply too much information for a human, unaided, to make use of. It is too complex. Collective humanity does not create a "hive mind," but rather we come together and spend time to "boil down" these very complex issues into simplifications we as a culture use in the moment. This is precisely because there is too much complexity to treat every event singularly and uniquely. Whether superintelligent systems would require similar compressions, or could operate with full complexity intact, remains an open question with profound implications.

The integration of religious and cross-species ethical models provides additional patterns for managing vast intelligence differentials while maintaining meaningful ethical dialogue. These established frameworks demonstrate how ethical agency can be preserved despite significant comprehension gaps, offering valuable insights for human-ASI interaction protocols.

Practical Framework Applications

The practical application of integrated theoretical frameworks requires careful balance between acknowledging cognitive limitations and maintaining human agency. Recent research demonstrating the impossibility of containing superintelligent systems (Alfonseca et al., 2021) suggests that traditional control mechanisms may prove inadequate. New approaches must

focus on maintaining meaningful human participation in ethical discourse despite potentially unbridgeable comprehension gaps.

Development of practical applications must address several critical challenges. First, how can oversight mechanisms remain meaningful when crucial ethical implications lie beyond human comprehension? Second, how can communication protocols acknowledge comprehension limitations while facilitating meaningful ethical dialogue? Third, how can decision-making processes preserve human ethical agency while acknowledging superintelligent capabilities? These questions require sustained investigation as ASI development progresses.

Future Research Directions and Applications

The development of frameworks for managing intelligence differentials in ethical reasoning requires extensive future research across multiple domains. The ethical event horizon framework suggests several critical areas requiring immediate investigation, particularly in quantifying and managing comprehension differentials between human and superintelligent systems.

Quantification Methodologies

The measurement of ethical comprehension capabilities across intelligence differentials presents immediate research challenges. While current frameworks assess human decision-making competence (Hein et al., 2015), extending these methodologies to evaluate vast intelligence differentials requires significant theoretical advancement. Research must develop metrics for assessing both processing capacity and ethical comprehension depth that remain meaningful across intelligence gaps.

The development of quantitative measures for ethical event horizons presents particular challenges. Current understanding of human cognitive limitations (Cowan, 2001; Singh et al., 2024) provides baseline metrics for human ethical comprehension capabilities. However, extending these measurements to superintelligent capabilities requires new theoretical approaches that can maintain relevance despite potentially unbridgeable comprehension gaps.

Communication Protocol Development

Research into methods for meaningful ethical dialogue across intelligence differentials becomes crucial for future human-ASI interaction. Traditional communication frameworks assume roughly equivalent cognitive capabilities between participants. The ethical event horizon framework reveals why

this assumption fails when dealing with superintelligent systems, necessitating new approaches to cross-intelligence ethical discourse.

The development of these protocols must address a fundamental paradox: how to communicate about ethical implications that lie beyond human comprehension. Religious models offer potential insights, demonstrating how ethical dialogue can remain meaningful despite vast intelligence differentials. These frameworks suggest possible approaches for maintaining human ethical agency while acknowledging fundamental comprehension limitations. Future research will explore these parallels systematically.

Protection Mechanism Development

Recent theoretical work demonstrating the impossibility of containing superintelligent systems through traditional methods (Alfonseca et al., 2021) necessitates research into new approaches for protecting human ethical agency. Rather than focusing on control or containment, research must explore mechanisms for maintaining meaningful human participation in ethical decisions despite comprehension limitations.

The ethical event horizon framework suggests that protection mechanisms must operate differently than traditional safeguards. Instead of attempting to constrain superintelligent capabilities, a likely impossible task, these mechanisms must focus on preserving meaningful human participation in ethical discourse while acknowledging vast intelligence differentials. Both epistemological concerns (inability to understand ASI reasoning) and power concerns (inability to resist ASI influence) require systematic investigation.

Long-term Research Requirements

Future research must address fundamental questions about maintaining human ethical relevance in an environment of vastly superior intelligence. Investigation of historical models, including religious frameworks and cross-species ethics, may provide insights for managing these unprecedented intelligence differentials. The religious models represent an area of particular interest for future exploration by the author. The vast capacity of artificial superintelligence suggests potential future scenarios where humanity operates under continuous surveillance, essentially handing over full authority to ASI. This scenario bears striking resemblance to religious concepts of divine omniscience. Future research will specifically delve into the distinction between human cognitive ability, ASI capabilities, and, as a religious exploration, extrapolating how far above humans divine intelligence would operate in correlating all past, current, and future information into ethical determination.

Long-term research must focus on developing sustainable approaches to human-ASI ethical interaction that preserve meaningful human agency while acknowledging fundamental cognitive limitations. The compression algorithm hypothesis, that human ethical norms function as cognitive compression algorithms born of processing limitations, requires both theoretical elaboration and empirical testing. If human ethics indeed represent necessary simplifications, what specific compressions do our ethical frameworks employ? How might superintelligent systems approach ethical reasoning without such compressions? Investigation of these questions may reveal fundamental differences between human and potential superintelligent ethical frameworks.

The integration of theoretical advancement with practical application development remains crucial. Research must maintain focus on both understanding intelligence differentials in ethical reasoning and developing practical frameworks for managing these differentials. Success in these research efforts may determine humanity's continued relevance in ethical discourse as artificial superintelligence develops.

Critical Examples Analysis

The theoretical frameworks developed for understanding intelligence differentials in ethical reasoning require examination through concrete examples. These cases provide crucial insights into both human cognitive limitations and potential superintelligent capabilities, while illuminating practical challenges in managing vast intelligence differentials. The ethical event horizon framework reveals new dimensions in these traditional ethical challenges.

The Abortion Debate Revisited

The persistent complexity of the abortion question provides particularly rich insights into human cognitive limitations in ethical reasoning. Despite decades of philosophical, legal, and ethical analysis since Foot's (1967) seminal work, the fundamental ethical questions remain unresolved. Viewed through the lens of ethical event horizons, this persistence may reflect not merely social or political disagreement, but fundamental limitations in human ethical reasoning capability.

Value pluralism undoubtedly contributes to this persistent disagreement. Different stakeholders hold genuinely incompatible values regarding personhood, bodily autonomy, and moral status. However, this framework proposes that cognitive limitations prevent us from even properly understanding what the value conflicts truly are. When examined through deep/shallow

cause analysis, the abortion debate reveals how human cognitive limitations manifest in ethical reasoning. The challenge of simultaneously considering immediate medical implications, long-term social consequences, individual rights, and collective responsibilities exceeds human working memory capabilities (Cowan, 2001). Recent empirical evidence demonstrates that cognitive load significantly impairs moral reasoning in complex dilemmas (McHugh et al., 2023), suggesting that the abortion debate's complexity may exceed the compressions our cognitive architecture requires.

The abortion question resists the kind of simplification our cognitive limits require. Our ethical abstractions, bodily autonomy, sanctity of life, personhood, are themselves products of cognitive compression. These necessary simplifications help us function but may obscure deeper causal and consequential realities. An ASI, capable of deep cause analysis, might perceive intricate webs of historical, social, and biological factors that reshape the ethical landscape in ways currently incomprehensible to human reasoning. Whether such perception constitutes ethical superiority or merely informational superiority remains uncertain.

Dynamic Ethical Horizons in Practice

The abortion debate provides a powerful demonstration of how ethical event horizons operate dynamically across different scales of intelligence. Just as the trolley problem reveals how an ASI might perceive vast webs of causation and consequence invisible to human comprehension, the abortion question demonstrates how ethical event horizons can "slide" across temporal and systemic dimensions.

Where human ethical analysis typically focuses on immediate factors, individual rights, medical necessity, and direct social impacts, an ASI might perceive the ethical implications sliding both "backward" and "forward" far beyond human comprehension. The ethical event horizon would move dynamically through historical causes: tracing complex interactions between societal development, medical advancement, and cultural evolution that shaped current ethical frameworks. Simultaneously, it would extend forward through intricate webs of future implications across multiple generations and societal systems.

Deep Cause Analysis Beyond Human Comprehension

An ASI's deep cause analysis capabilities might reveal subtle historical patterns invisible to human cognition: how economic systems influenced healthcare access, how technological development shaped ethical perspectives, and how complex social dynamics affected individual decision-making capabili-

ties. These causal chains, extending far beyond human working memory constraints (Cowan, 2001), would reveal ethical implications currently invisible to human analysis. Recent meta-analytic evidence suggests that cognitive manipulations consistently affect moral judgments, though effects may vary across contexts (Rehren, 2024), further supporting the notion that human ethical reasoning operates within significant cognitive constraints.

Deep Effect Projection in Ethical Discourse

Where human analysis struggles to project beyond immediate societal impacts, an ASI might perceive vast networks of interconnected consequences. These could include subtle shifts in cultural values, complex demographic implications, and evolutionary effects on human decision-making frameworks, all extending beyond the human ethical event horizon. The differential between human "shallow effect" limitations and ASI "deep effect" capabilities may help explain why traditional ethical frameworks have failed to resolve this debate. An ASI might not "solve" the abortion question in a way that eliminates value conflict, but could reveal dimensions of the dilemma currently invisible to human analysis, potentially reframing the debate in ways we cannot currently envision.

Environmental Ethics and System Complexity

Climate change response demonstrates how human cognitive limitations affect ethical reasoning about complex systems. The delayed recognition of climate change implications, despite available data, reveals human constraints in processing complex cause-effect relationships. The ethical event horizon framework helps explain why humans struggle to comprehend and respond to long-term, complex ethical challenges that extend beyond immediate cause-effect relationships.

Species Preservation Decisions

Decisions regarding species preservation illustrate challenges in balancing immediate concerns against long-term ecological implications. Human cognitive architecture struggles to simultaneously process multiple interacting variables across extended time periods. Through the ethical event horizon framework, these limitations appear not as mere practical challenges but as fundamental constraints on human ethical comprehension.

An ASI system, operating with deep cause and effect capabilities, might perceive intricate relationships between species preservation decisions and long-term planetary stability that lie beyond human ethical event horizons. These deeper perceptions could reveal ethical implications currently invis-

ible to human decision-makers, suggesting the need for new frameworks that can incorporate superintelligent insights while maintaining meaningful human participation in preservation decisions.

Professional Decision-Making Examples

Criminal investigation provides concrete evidence of human cognitive limitations in professional contexts. Research demonstrates that even experienced investigators exhibit persistent confirmation bias and tunnel vision (Meterko & Cooper, 2021). When viewed through the ethical event horizon framework, these limitations reflect not just procedural challenges but fundamental constraints on human causal analysis capabilities.

Medical ethics presents parallel challenges, where healthcare professionals must navigate complex ethical decisions within inherent cognitive limitations (Charland, 2001). Recent research demonstrates that cognitive load reduces not only moral reasoning speed but the quality of justifications provided (McHugh et al., 2023). The persistence of these limitations across professional domains, despite extensive training and experience, provides empirical support for the concept of ethical event horizons as fundamental rather than circumstantial constraints.

Implications for Framework Development

Analysis of these critical examples reveals consistent patterns in human cognitive limitations while suggesting potential areas where superintelligent capabilities might transcend these constraints. The ethical event horizon framework provides a structured approach to understanding these limitations and their implications for human-ASI interaction. This understanding becomes crucial for developing frameworks that can maintain meaningful human participation in ethical decisions involving superintelligent systems. Whether superintelligent perception of deeper causes and more extensive consequences translates to genuine ethical superiority remains uncertain, but the differential in comprehension capability appears substantial.

Limitations and Boundary Conditions

The Ethical Event Horizon framework represents a novel theoretical approach to understanding intelligence differentials in ethical reasoning. While empirical evidence supports the foundational claims regarding human cognitive constraints (Cowan, 2001; Singh et al., 2024; McHugh et al., 2023), several important limitations and boundary conditions warrant explicit acknowledgment.

Theoretical Assumptions and Their Limitations

This framework rests on several assumptions that require careful examination. The relationship between information processing capacity and ethical comprehension remains uncertain. While artificial superintelligence would demonstrably possess superior capabilities for data integration and causal analysis, whether such capabilities translate to genuinely superior ethical judgment remains an open question. The framework proposes that human ethical norms function as cognitive compression algorithms, necessary simplifications born of working memory constraints (Cowan, 2001) and limited information processing capacity (Singh et al., 2024; McHugh et al., 2023). However, this compression may serve functions beyond mere computational necessity. Human-scale cognition, shaped by evolutionary pressures and embodied experience, might capture aspects of ethical reasoning that remain inaccessible to systems lacking such grounding.

The trolley problem analysis demonstrates how ethical event horizons might operate dynamically, with superintelligent systems potentially perceiving causal chains and consequence networks far beyond human comprehension. Yet this analysis assumes that perceiving more information constitutes a meaningful advantage in ethical reasoning. Critics might argue that ethical wisdom requires not merely information integration but qualities potentially unique to human experience: understanding of suffering through vulnerability, moral intuitions shaped by specific evolutionary history, or wisdom accumulated through temporal existence as mortal beings. The framework remains agnostic on whether information processing superiority translates to ethical superiority, instead focusing on the implications of this uncertainty for human-ASI interaction.

Alternative Explanations for Persistent Ethical Dilemmas

The framework interprets persistent ethical disagreements, such as the abortion debate, as evidence of cognitive limitations preventing full comprehension of complex causal and consequential dimensions. However, alternative explanations warrant consideration. Value pluralism, as articulated in political philosophy (Rawls, 1993), suggests that some ethical disagreements reflect genuinely incompatible values rather than insufficient cognitive capacity. The abortion debate may persist not because humans cannot perceive relevant causal relationships, but because different stakeholders hold fundamentally incompatible values regarding personhood, bodily autonomy, and moral status.

This framework does not dismiss value pluralism as a contributing factor in ethical disagreement. Rather, it proposes that cognitive limitations pre-

vent us from even properly understanding what the value conflicts truly are. The abortion question resists the kind of simplification our cognitive limits require. Our ethical abstractions, bodily autonomy, sanctity of life, personhood, are themselves products of cognitive compression. These are necessary simplifications that help us function but may obscure deeper causal and consequential realities. An artificial superintelligence might not "solve" value conflicts in a way that eliminates genuine incompatibilities, but could reveal dimensions of such dilemmas currently invisible to human analysis, potentially reframing debates in ways we cannot currently envision.

Cultural, political, and institutional factors undoubtedly shape ethical discourse in ways distinct from cognitive limitations. Professional decision-making research demonstrates that organizational structures, power dynamics, and social pressures significantly influence ethical judgments (Meterko & Cooper, 2021). The persistence of ethical disagreement may reflect these structural factors as much as, or more than, individual cognitive constraints. The framework acknowledges these alternative explanations while maintaining that cognitive limitations represent an underexplored dimension of ethical disagreement worthy of systematic investigation.

Measurement and Verification Challenges

The empirical verification of ethical event horizons presents significant methodological challenges. While working memory constraints are measurable (Cowan, 2001), and recent research demonstrates cognitive load's impact on moral reasoning (Singh et al., 2024; Rehren, 2024; Zheng et al., 2025), directly measuring the boundary beyond which ethical implications become incomprehensible remains problematic. How would researchers identify ethical considerations that lie beyond human comprehension? The very nature of such boundaries suggests they may resist empirical investigation using current methodologies.

The framework's predictions regarding superintelligent capabilities remain necessarily speculative. While theoretical work demonstrates the impossibility of containing superintelligent systems through traditional computational methods (Alfonseca et al., 2021), the specific nature of superintelligent ethical reasoning remains uncertain. Future research must develop methodologies capable of assessing intelligence differentials in ethical comprehension without presupposing human-scale cognition as the evaluative standard.

Scope and Applicability

This framework focuses primarily on individual cognitive limitations and their implications for human-ASI interaction. However, human ethical rea-

soning operates at multiple scales. While individual humans face severe working memory constraints (Cowan, 2001), collective human intelligence, distributed across individuals, institutions, and generations, extends beyond individual capabilities. Cultural transmission, institutional memory, and collaborative deliberation enable humanity to address ethical questions that exceed individual cognitive capacity.

Yet collective intelligence does not eliminate the ethical event horizon. It merely shifts its location. Human collectives create simplified abstractions, norms, rules, legal codes, precisely because even distributed cognition cannot process the full complexity of every unique situation. These abstractions represent lossy compressions, necessary simplifications that enable ethical functioning despite cognitive limitations. Whether superintelligent systems would require similar compressions, or could operate with full complexity intact, remains an open question with profound implications for human-ASI ethical interaction.

Future Research Directions

Several critical areas require investigation to advance understanding of intelligence differentials in ethical reasoning. First, development of quantitative measures for ethical comprehension capabilities that remain meaningful across intelligence gaps represents an immediate priority. Current competency assessment frameworks (Hein et al., 2015) address human cognitive variation but require fundamental reconceptualization for application to vast intelligence differentials.

Second, the compression algorithm hypothesis, that human ethical norms function as cognitive compression algorithms born of processing limitations, requires both theoretical elaboration and empirical testing. If human ethics indeed represent necessary simplifications, what specific compressions do our ethical frameworks employ? How might superintelligent systems approach ethical reasoning without such compressions? Investigation of these questions may reveal fundamental differences between human and potential superintelligent ethical frameworks.

Third, historical models of interaction with vastly superior intelligence, particularly religious frameworks addressing divine omniscience, warrant systematic examination. These models demonstrate how meaningful agency can be preserved despite seemingly unbridgeable comprehension gaps. Future research will explore whether insights from religious frameworks translate to human-ASI interaction, and what modifications such translation might require.

Finally, the power dynamics of intelligence differentials require investigation distinct from epistemological questions. Even if humans could comprehend superintelligent reasoning, recent theoretical work suggests containment may prove impossible (Alfonseca et al., 2021). The implications of operating in ethical environments dominated by entities whose reasoning exceeds human comprehension, whether or not such reasoning constitutes genuine ethical superiority, demand careful analysis.

Concluding Remarks on Limitations

The Ethical Event Horizon framework does not claim to resolve questions about intelligence and ethical reasoning. Rather, it provides a structured approach to investigating how intelligence differentials affect ethical comprehension, while acknowledging significant uncertainties. The framework's primary contribution lies in formalizing concepts, ethical event horizons, deep and shallow cause-effect analysis, and the compression algorithm hypothesis, that enable systematic exploration of these questions.

Whether artificial superintelligence will develop genuine ethical wisdom or merely superior information processing capacity remains unknown. What appears certain is that humanity's historical position as the highest intelligence in any system we create or operate within faces potential disruption. The framework presented here offers conceptual tools for navigating this unprecedented transition, while acknowledging the profound uncertainties inherent in anticipating intelligence differentials that may exceed human comprehension entirely.

Conclusions and Recommendations

The development of artificial superintelligence presents unprecedented challenges for ethical reasoning and decision-making frameworks. This investigation of intelligence differentials in ethical comprehension, through the lens of ethical event horizons, reveals both fundamental limitations in human cognitive architecture and crucial considerations for future human-ASI interaction. The findings suggest several key theoretical insights and practical recommendations for maintaining meaningful human agency in ethical discourse across vast intelligence differentials.

Theoretical Framework Synthesis

The ethical event horizon concept provides a structured approach to understanding intelligence differentials in ethical reasoning. Research demonstrates that human cognitive limitations, particularly in working memory

and processing capability (Cowan, 2001; Singh et al., 2024), create measurable boundaries beyond which ethical implications become incomprehensible. The trolley problem analysis reveals how these horizons operate dynamically, with superintelligent systems potentially perceiving vast webs of causes and effects beyond human comprehension.

The integration of deep/shallow cause-effect analysis within this framework reveals how intelligence differentials affect both historical understanding and future projection capabilities. This understanding, combined with the demonstrated impossibility of containing superintelligent systems (Alfonseca et al., 2021), suggests urgent need for new approaches to maintaining meaningful human participation in ethical decisions involving superintelligent systems.

The framework proposes that human ethical norms function as cognitive compression algorithms, necessary simplifications born of working memory constraints and limited information processing capacity. Individual humans are nearly incapable, as singular unaided creatures, of recalling and integrating all relevant data for complex ethical decisions. Collective humanity does not create a "hive mind," but rather comes together to "boil down" complex issues into simplifications we as a culture use in the moment. This occurs precisely because there is too much complexity to treat every event singularly and uniquely. Whether superintelligent systems would require similar compressions, or could operate with full complexity intact, remains an open question with profound implications for human-ASI ethical interaction.

Practical Implementation Requirements

Development of practical frameworks for human-ASI ethical interaction requires careful balance between acknowledging cognitive limitations and preserving human agency. Historical models, including religious frameworks and cross-species ethics, provide valuable patterns for managing vast intelligence differentials while maintaining meaningful ethical dialogue. These models suggest possible approaches for preserving human ethical relevance despite potentially unbridgeable comprehension gaps.

The challenge extends beyond mere epistemological concerns about understanding ASI reasoning to encompass power dynamics. We may find ourselves beholden to superintelligent systems regardless of whether we comprehend their ethical reasoning. Our entire historical run as humans has been with us as the top intelligence in any system we create or operate in. ASI represents a radical shift in all ways. The gulf between human and ASI ethical event horizons could be vast, with ASI pushing the horizon so far away that the thinking could astound and mystify humanity.

Final Recommendations

Based on this investigation, several critical priorities emerge for developing frameworks to manage ethical intelligence differentials:

Development of quantifiable metrics for ethical comprehension capabilities that acknowledge both human limitations and superintelligent potential. These metrics must move beyond traditional competency assessments to address fundamental differences in ethical reasoning capabilities. Recent empirical advances in measuring cognitive load's impact on moral reasoning (Singh et al., 2024; Rehren, 2024; Zheng et al., 2025) provide methodological foundations, but require extension to vast intelligence differentials.

Creation of communication protocols that enable meaningful ethical dialogue across intelligence differentials. These protocols must acknowledge human cognitive limitations while preserving essential human participation in ethical decision-making processes. The protocols must address both the epistemological problem (inability to understand ASI reasoning) and the power problem (potential inability to resist ASI influence).

Integration of historical models with new theoretical understanding to develop practical frameworks for human-ASI ethical interaction. Religious and cross-species ethical frameworks offer valuable patterns for maintaining ethical relevance despite vast intelligence differentials. Future research will explore these parallels systematically, examining how historical frameworks managed intelligence differentials and whether insights translate to the ASI context.

Investigation of the compression algorithm hypothesis through both theoretical elaboration and empirical testing. If human ethical norms indeed function as cognitive compression algorithms, understanding the specific nature of these compressions becomes crucial for anticipating how superintelligent systems might approach ethical reasoning differently.

Final Synthesis

The ethical event horizon framework offers more than theoretical understanding of intelligence differentials in ethical comprehension. It provides practical guidance for developing frameworks that preserve meaningful human agency in future human-ASI interactions. The dynamic nature of ethical event horizons, demonstrated through the trolley problem analysis, reveals how superintelligent systems might perceive ethical implications far beyond human comprehension in both causal and effect dimensions.

Success in managing these intelligence differentials requires careful balance between acknowledging fundamental cognitive limitations and maintaining human participation in ethical discourse. While complete comprehension of

superintelligent ethical reasoning may prove impossible, developing frameworks that enable meaningful human participation in ethical decisions remains crucial. The framework remains agnostic on whether superintelligent information processing capacity translates to genuine ethical superiority, but the differential in comprehension capability appears substantial.

The future of human ethical relevance in an environment of artificial superintelligence may depend on successfully managing these intelligence differentials while preserving essential human agency in ethical decision-making. The ethical event horizon framework provides a structured approach to this challenge, offering both theoretical understanding and practical guidance for maintaining meaningful human participation in an increasingly complex ethical landscape. Whether humanity can navigate this transition successfully remains uncertain, but the framework presented here offers conceptual tools for systematic investigation of these unprecedented challenges.

References

Alfonseca, M., Cebrian, M., Fernandez Anta, A., Coviello, L., Abeliuk, A., & Rahwan, I. (2021). Superintelligence cannot be contained: Lessons from computability theory. *The Journal of Artificial Intelligence Research*, *70*, 65-76. https://doi.org/10.1613/jair.1.12202

Bielby, P. (2005). The conflation of competence and capacity in English medical law: A philosophical critique. *Medicine Health Care and Philosophy*, *8*(3), 357–369. https://doi.org/10.1007/s11019-005-0537-z

Blum, M., & Vempala, S. (2020). The complexity of human computation via a concrete model with an application to passwords. *Proceedings of the National Academy of Sciences*, *117*(17), 9208–9215. https://doi.org/10.1073/pnas.1801839117

Carrillo, M. R. (2020). Artificial intelligence: From ethics to law. *Telecommunications Policy*, *44*(6), 101937. https://doi.org/10.1016/j.telpol.2020.101937

Charland, L. C. (2001). Mental competence and value: The problem of normativity in the assessment of decision-making capacity. *Psychiatry Psychology and Law*, *8*(2), 135–145. https://doi.org/10.1080/13218710109525013

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114. https://doi.org/10.1017/s0140525x01003922

De Bruin, W. B., Parker, A. M., & Fischhoff, B. (2020). Decision-making competence: More than intelligence? *Current Directions in Psychological Science*, *29*(2), 186–192. https://doi.org/10.1177/0963721420901592

FitzPatrick, W. J. (2009). Thomson's turnabout on the trolley. *Analysis*, *69*(4), 636–643. https://doi.org/10.1093/analys/anp091

Foot, P. (2002). The problem of abortion and the doctrine of the double effect. In *Oxford University Press eBooks* (pp. 19–32). https://doi.org/10.1093/0199252866.003.0002

Hein, I. M., Troost, P. W., Broersma, A., De Vries, M. C., Daams, J. G., & Lindauer, R. J. L. (2015). Why is it hard to make progress in assessing children's decision-making competence? *BMC Medical Ethics*, *16*(1). https://doi.org/10.1186/1472-6939-16-1

McHugh, C., McGann, M., Igou, E. R., & Kinsella, E. L. (2023). Cognitive load can reduce reason-giving in a moral dumbfounding task. *Collabra Psychology*, *9*(1). https://doi.org/10.1525/collabra.73818

Meterko, V., & Cooper, G. (2021). Cognitive biases in criminal case evaluation: A review of the research. *Journal of Police and Criminal Psychology*, *37*(1), 101–122. https://doi.org/10.1007/s11896-020-09425-8

Rawls, J. (1993). *Political liberalism.* Columbia University Press.

Rehren, P. (2024). The effect of cognitive load, ego depletion, induction and time restriction on moral judgments about sacrificial dilemmas: a meta-analysis. *Frontiers in Psychology*, *15*. https://doi.org/10.3389/fpsyg.2024.1388966

Rindler, W. (2002). Visual horizons in world-models. *General Relativity and Gravitation*, *34*(1), 133–153. https://doi.org/10.1023/a:1015347106729

Simon, D. (2004). A third view of the black box: Cognitive coherence in legal decision making. *The University of Chicago Law Review*, *71*(2), 511-586.

Singh, A., Murzello, Y., Lee, H., Abdalla, S., & Samuel, S. (2024). Moral Decision Making: Explainable Insights into the Role of Working Memory in Autonomous Driving. *Machine Learning With Applications*, 100599. https://doi.org/10.1016/j.mlwa.2024.100599

Soares, N., & Fallenstein, B. (2017). Agent foundations for aligning machine intelligence with human interests: A technical research agenda. In S. Armstrong, R. Yampolskiy, J. Miller & V. Callaghan (Eds.), *The technological singularity* (pp. 103-125). Springer Berlin/Heidelberg. https://doi.org/10.1007/978-3-662-54033-6_5

Sommaggio, P., & Marchiori, S. (2020). Moral dilemmas in the A.I. era: A new approach. *Journal of Ethics and Legal Technologies*, *2*(1), 89-102. https://doi.org/10.14658/pupj-jelt-2020-1-5

Thomson, J. J. (1976). A defense of abortion. *Biomedical ethics and the law* (pp. 39–54)

Xin, N. L., Qian, N. W., & Huili, N. W. (2018). The significance of horizon in scientific cognitive activities. *Philosophy Study*, *8*(4). https://doi.org/10.17265/2159-5313/2018.04.002

Zheng, M., Wang, L., & Tian, Y. (2025). Does cognitive load influence moral judgments? The role of Action–Omission and Collective Interests. *Behavioral Sciences*, *15*(3), 361. https://doi.org/10.3390/bs15030361