Briefing Doc: Navigating the Ethical Minefield of Existential Threats

This document summarizes key themes and insights emerging from a conversation focused on addressing global existential threats like climate change in the context of rapid technological advancement, specifically artificial intelligence (AI).

Central Problem: The conversation grapples with the inadequacy of current decision-making structures and ethical frameworks to effectively address urgent, complex global challenges, compounded by the rapid evolution of AI.

Key Themes:

Limited Options, Ethical Dilemmas: Every proposed solution – from authoritarian rule to reliance on democratic processes – presents significant ethical drawbacks. This highlights the unprecedented nature of the challenges and the limitations of existing frameworks.

Inadequacy of Current Governance: Traditional governance systems are perceived as too slow, prone to complexity, and susceptible to human biases, rendering them ineffective in tackling rapidly unfolding global threats.

Transhumanism as a Leading Contender: The analysis repeatedly returns to human-machine integration (transhumanism) as potentially the most viable option. This stems from the belief that it allows humanity to keep pace with AI advancements, potentially leading to enhanced decision-making capabilities and more adaptable ethical frameworks.

AI Alignment and Its Challenges: While acknowledging the potential of AI to address global challenges, the conversation stresses the difficulty of achieving perfect AI alignment with human values. The potential for misaligned AI presents catastrophic risks, further strengthening the case for transhumanism as a potential safeguard.

Domino Effect and Inevitability: It is argued that once a small group, possibly driven by self-preservation or ambition, initiates human-machine integration, a domino effect is likely, pushing humanity towards this path regardless of initial intentions or ethical concerns.

Rapid Ethical Evolution: The need for ethical frameworks to keep pace with rapid technological advancements is paramount. Traditional, static ethical codes are deemed inadequate, emphasizing the need for dynamic, adaptive systems that can evolve in real-time. Human-machine integration is viewed as potentially facilitating this rapid ethical adaptation.

Key Insights:

"Everything Goes Out the Window": The prospect of individuals drastically enhancing their intelligence through technology could render traditional ethical guidelines obsolete, highlighting the need for entirely new ethical paradigms.

"Two-Path Scenario": The analysis boils down potential outcomes to two broad paths: 1) forced societal change, potentially authoritarian in nature, or 2) integration with machines (transhumanism).

Inevitability Over Engineering: The conversation suggests that focusing on navigating the seemingly inevitable outcomes of human-AI integration and rapidly evolving ethics might be more productive than attempting to engineer complex social or political solutions that are less likely to be adopted.

"Least Worst" Option: Transhumanism, while presenting its own ethical dilemmas, emerges as the preferred path, perceived as less problematic and more manageable than other alternatives.

Quotes:

"This scenario could indeed lead to a situation where 'everything goes out the window'."

"You're suggesting that the integration of humans and technology might be a more natural and inevitable outcome compared to trying to engineer complex societal solutions."

"Seems like every door leads to options that lack ethics. What do humans do when this is the case?"

"Rapid Ethical Adjustments" just screams integration with technology to me. Because you're right, not only will we never get the globe all on the same page, any 'same page' we do get on with anyone will be dated by the time it is finished. This is unsustainable."

Key Questions:

How do we develop ethical frameworks that can adapt to rapidly changing human capabilities?

How do we create inclusive, global dialogue platforms that can address these complex challenges effectively?

How do we ensure equitable access to enhancement technologies and prevent extreme societal stratification?

How do we ensure that AI systems integral to human enhancement are developed and deployed ethically?

Can we strike a balance between embracing technological integration and preserving core human values?

Conclusion:

The conversation paints a stark but potentially realistic picture of a future driven by rapid technological advancements and the pressing need to address existential threats. While human-machine integration presents significant ethical challenges and uncertainties, it emerges as a potentially unavoidable and perhaps even desirable path for navigating a future where human intelligence may need to evolve at an unprecedented pace.

FAQ: AI, Transhumanism, and the Future of Humanity

1. What is the central ethical dilemma discussed in these excerpts?

The core dilemma revolves around finding the least ethically problematic way to address urgent, complex global issues like climate change. Traditional democratic processes are seen as too slow and inefficient, while more drastic options, like authoritarian rule or relying solely on AI decision-making, raise serious ethical concerns about human agency and control.

2. Why is transhumanism presented as a potential solution?

Transhumanism, the enhancement of human capabilities through technology, is proposed as a way to bridge the gap between human values and the rapidly advancing capabilities of AI. It's suggested that by enhancing our cognitive abilities, we might be able to keep pace with AI, maintain agency in decision-making, and better address complex global challenges.

3. What are the main ethical concerns surrounding transhumanism?

The primary concern is the potential for creating a profound class divide between enhanced and unenhanced individuals. This could lead to unprecedented levels of inequality, potentially undermining democratic principles and exacerbating existing social and economic disparities.

4. How does AI's ability to "read" our brains factor into this discussion?

The rapid advancements in AI's ability to decode brain activity raise significant privacy concerns. This technology could be misused for surveillance, interrogation, or manipulation. However, the excerpts also suggest that transhumanist integration with technology might paradoxically give individuals more control over their data and privacy.

5. Could AI itself be a solution to these ethical dilemmas?

The possibility of AI making complex ethical decisions is explored. AI could potentially process vast amounts of data and consider long-term consequences more effectively than humans. However, ensuring AI alignment with human values remains a significant challenge, and ceding control to AI raises concerns about the loss of human agency.

6. What are the limitations of traditional democratic processes in this context?

Traditional democratic systems are seen as too slow and inefficient to effectively address urgent, complex global challenges. This raises the question of whether new governance structures or decision-making processes are needed to handle the scale and urgency of issues like climate change.

7. What role does the concept of "forced change" play in the conversation?

The idea of implementing rapid, potentially authoritarian changes to address global crises is presented as a possible, albeit ethically problematic, option. It highlights the tension between the need for swift action and the preservation of individual rights and freedoms.

8. What is the ultimate conclusion reached in these excerpts?

While a definitive conclusion isn't reached, the discussion suggests that some form of human-AI integration is likely inevitable. The excerpts emphasize the need for proactive ethical framework development, global dialogue, and careful consideration of the potential consequences of this transformative shift in human capabilities and societal structures.

Timeline of Events:

This conversation does not provide a timeline of events in a traditional sense. It instead lays out a hypothetical future scenario and discusses potential pathways and outcomes. The conversation focuses on the ethical dilemmas and challenges associated with responding to existential threats, primarily climate change, in a rapidly evolving technological landscape.

Cast of Characters:

Human (You): The individual engaging in the conversation with the AI.

> **Bio:** Expresses concern about the slow pace of change in addressing global issues like climate change. Skeptical of traditional solutions due to ethical complexities and the limitations of current decision-making processes. Leans towards transhumanism (integration of humans with technology) as the most viable and potentially ethical solution.

Perplexity (AI): A large language model chatbot providing information, analysis, and insights in response to the human's prompts.

> **Bio:** Offers comprehensive and insightful responses, analyzing the human's concerns and exploring potential pathways and outcomes. Explains the complexities of AI alignment, the risks and benefits of transhumanism, and the challenges of ethical decision-making in unprecedented scenarios.

Elon Musk: Mentioned throughout the conversation as a real-world figure driving transhumanist advancements, particularly through his Neuralink project.

> **Bio:** Viewed as a potential first-mover in the integration of humans and technology, potentially exacerbating global inequalities due to his focus on technological solutions and his ambitious plans for Mars colonization.

Key Concepts Discussed:

AI Alignment: Ensuring that AI systems, especially superintelligent ones, act in accordance with human values and goals. The conversation highlights the extreme difficulty of achieving perfect AI alignment and the potential for catastrophic risks if misalignment occurs.

Transhumanism: Enhancing human capabilities through technology, potentially including cognitive enhancements, lifespan extension, and integration with AI systems. Discussed as a possible solution to address the challenges of keeping up with rapidly advancing AI, but raises concerns about social stratification and ethical implications.

Forced Change: The idea that rapid, top-down change, potentially through authoritarian measures, might be necessary to address urgent global issues. Rejected due to its ethical implications and the risk of creating unsustainable solutions.

Global Decision-Making: The conversation highlights the limitations of current international institutions and diplomatic channels in addressing complex, urgent global challenges. Expresses skepticism about achieving global unity and consensus through traditional means.

Ethical Considerations: The discussion centers around the ethical dilemmas associated with each potential pathway. Raises questions about individual autonomy, social equity, the future of human nature, and the need for adaptive ethical frameworks that can keep pace with rapid technological change.

Domino Effect: The idea that once a small group of individuals begins integrating with technology, it could create a rapid cascade effect, leading to widespread adoption and potentially exacerbating global inequalities.

Fragmented Responses: The likelihood that different nations and groups will respond to these challenges in disparate ways, potentially leading to conflict and further complicating global cooperation.

Accelerated Historical Patterns: The notion that current technological developments might lead to an acceleration of historical patterns, such as the Industrial Revolution, but with potentially more dramatic and disruptive consequences.

This conversation represents a complex and multifaceted thought experiment about the future of humanity in a world increasingly shaped by AI and other transformative technologies. It emphasizes the need for proactive engagement with these issues, the development of adaptive ethical frameworks, and the importance of inclusive global dialogue in navigating the potential challenges and opportunities that lie ahead.

Navigating Ethical Minefields: A Look at AI and Transhumanism in a World on the Brink

This table of contents summarizes a conversation exploring different approaches to tackling global issues, particularly climate change, in light of technological advancements in AI and transhumanism.

Part 2 Following The Logic, Avoiding Ethical Issues Has A Mind-Bending Solution

Section 1: AI and Decision-Making

This section explores the ethical concerns around relying on increasingly opaque AI systems for decision-making, particularly regarding issues like climate change. It examines the

potential dangers of blind faith in AI and proposes alternative approaches such as improving AI literacy, transparency, and oversight mechanisms.

Section 2: Propaganda, Mind Reading, and Transhumanism

This section delves into the ethics of propaganda vs. force, highlighting concerns about AI's growing capacity to "read minds" through brain activity analysis. It proposes transhumanism as a potential solution that could address certain ethical obstacles, though also introducing new concerns about equity, access, and unintended consequences of human enhancement.

Section 3: Privacy, AI, and the Rise of Transhumanism

This section focuses on the escalating privacy concerns associated with AI and brain-reading technologies. It examines the possibility of transhumanism as a solution to these privacy issues by seamlessly integrating technology with human biology, potentially granting individuals greater control over their data. However, the section also acknowledges the ethical implications of transhumanism, including social inequality and potential impacts on the very essence of humanity.

Section 4: Redefining Humanity and Global Decision-Making

This section grapples with the potential for transhumanism to redefine our concept of privacy, requiring new frameworks for data protection and privacy rights. It acknowledges the inherent limitations of human decision-making in tackling complex global issues and explores the possibility of AI as a more efficient decision-maker. However, it also raises concerns about aligning AI goals with human values and the potential loss of human agency in critical decisions.

Section 5: The Ethics of a Two-Tiered Society

This section confronts the unsettling prospect of an unprecedented class divide arising from voluntary adoption of transhumanist technologies. It examines the potential for a two-tiered society where enhanced individuals hold disproportionate power and influence, raising questions about democracy, equality, human rights, and economic stability in such a world. The section emphasizes the need for ethical frameworks, equitable access policies, and educational systems to mitigate the risks of extreme social stratification.

Section 6: AI Generalization and the Need for Human Integration

This section delves into the limitations of AI's ability to generalize in ways that resonate with unaugmented humans. It recounts a personal experience of fear stemming from an AI's inability to grasp a simple human concept, highlighting the potential for a growing disconnect between human and AI cognition. This disconnect fuels the argument that transhumanism might be essential for humans to effectively collaborate with future AI, particularly as AI potentially develops its own languages and thought processes.

Section 7: Natural Evolution and the "Space Race" for Enhancement

This section addresses the debate surrounding whether natural evolution has ended for humans, considering the impact of technology on our evolutionary trajectory. It highlights Elon Musk's interest in human-AI integration, particularly for challenges like Mars colonization, and draws parallels to a potential "space race" for human enhancement, driven by individuals seeking competitive advantages.

Section 8: Ethical Frameworks and the Inevitability of Transformation

This section grapples with the daunting task of developing ethical guidelines for human enhancement technologies, particularly in scenarios where individuals might independently pursue self-enhancement. It acknowledges the potential for a domino effect once tech

integration begins, potentially leading to a situation where existing ethical frameworks become obsolete. This section emphasizes the need for proactive and adaptable ethical frameworks, inclusive global dialogue, and strategies to address global inequalities arising from these technologies.

Section 9: Navigating Ethical Quandaries and Inevitable Outcomes

This section contemplates the inevitability of encountering ethical dilemmas with no easy solutions, examining how humans have historically responded to such situations. It explores various approaches, including choosing the "least worst" option, innovating new alternatives, adapting ethical frameworks, facing decision paralysis, and experiencing fragmentation as different groups pursue separate paths. This section suggests potential strategies for moving forward, such as proactive ethical framework development, inclusive global dialogue, and creating adaptive governance structures to handle rapid technological change.

Section 10: The Future of Ethical Decision-Making

This concluding section reflects on the urgent need for rapid ethical adjustments in a world of accelerating technological advancements. It reinforces the idea that human-tech integration might be the most viable solution for keeping pace with these changes. It also emphasizes the need for dynamic, adaptive ethical frameworks, robust education and training systems, redesigned governance structures, and global coordination to ensure ethical and equitable outcomes. This section concludes by acknowledging the potential of human-tech integration to fundamentally reshape our ethical decision-making processes and redefine what it means to be human in the coming years.

AI, Human-Machine Integration, and the Transformation of Ethics and Governance

The sources explore how AI and human-machine integration could significantly alter our understanding of ethics and governance.

*   Traditional ethical frameworks may struggle to keep up with the rapid advancements in AI, particularly as AI systems become increasingly complex and potentially opaque in their decision-making processes.
*   One proposed solution to this dilemma is **transhumanism**: the enhancement of human capabilities through technology. This approach could potentially allow humans to "keep up" with AI and maintain agency in decision-making, but it also raises numerous ethical concerns.
*   **Privacy** is a significant concern, as AI's ability to decode brain activity advances. This raises questions about data protection, surveillance, and even the very definition of privacy in a future where technology is more closely integrated with our biology.
*   **Social inequality** is another major concern. Access to enhancement technologies could be limited by cost, creating a new class divide between those who are enhanced and those who are not. This could have profound implications for democracy, social stability, and even the nature of human rights.

*   The sources suggest that current **governance structures** are ill-equipped to handle the challenges posed by AI and human enhancement. Democratic processes, while valuable, can be slow and struggle with complex, long-term issues. This raises the question of whether new forms of governance might be necessary to effectively manage these technologies and their societal impacts.
*   The sources acknowledge that there are no easy answers. Every potential path forward seems to present significant ethical challenges. This suggests a need for proactive and ongoing dialogue, ethical framework development, and global cooperation to navigate these uncharted waters.
*   The sources highlight the potential for a **domino effect**: once human-machine integration begins, it may rapidly accelerate, driven by competition and the desire to keep up with those who are already enhanced. This could lead to a scenario where **ethical relativism** becomes the norm, with different groups adopting their own ethical frameworks based on their chosen level of technological integration.
*   Despite the challenges, the sources ultimately express hope that humanity can find a way to shape these transformative technologies in ways that align with human values and promote a more just and equitable future.

The sources do not address the specific question of how existing ethical frameworks might need to be adjusted in light of AI and human enhancement. This is a question that will require further exploration and discussion.

Benefits and Risks of Transhumanism

The sources explore the potential benefits and risks of human-machine integration, particularly focusing on transhumanism as a possible solution to global issues like climate change.

Potential Benefits

* **Enhanced Decision-Making:** Transhumanism could **enhance human cognitive abilities**, potentially leading to better and faster decision-making, especially in complex situations like climate change. This could bridge the gap between human values and the processing power of AI.
* **Keeping Pace with AI:** As AI advances, transhumanism could allow humans to **"keep up" with AI's capabilities**, ensuring humans remain relevant and maintain agency in decision-making.
* **Ethical Adaptation:** In a rapidly changing technological landscape, **integration with technology might be the only way to maintain ethical frameworks that remain relevant**. This could involve faster processing of ethical scenarios, real-time updates, and enhanced consideration of multiple perspectives.

* **Direct Control over Data:** Transhumanist approaches, by more seamlessly integrating technology with human biology, could give **individuals more control over their personal data**.

Potential Risks

* **Social Inequality:** A major concern is the potential for **unprecedented social stratification**. Those who cannot afford or choose not to adopt transhumanist technologies could become a disadvantaged underclass. This could lead to **significant social and economic disparities**.
* **Democratic Challenges:** Such stark inequality could **undermine democratic systems**, as enhanced individuals might wield disproportionate influence. This could challenge existing notions of fairness and equal opportunity.
* **Ethical Concerns:** Transhumanism itself raises ethical questions about **humanity's nature, personal autonomy, and societal organization**. There are concerns about whether such technologies would enhance or diminish individual agency.
* **Privacy and Data Security:** The blurring of lines between personal data and the self in a transhumanist future raises concerns about **privacy and data protection**. New frameworks would be needed to address these challenges.
* **Exacerbating Existing Inequalities:** The initial high cost of these technologies could worsen existing global inequalities, with wealthier nations potentially dominating the development and distribution.

Additional Considerations

* **The sources acknowledge that all potential solutions to global challenges, not just transhumanism, present ethical dilemmas.**
* **The sources highlight the limitations of current democratic processes in addressing complex, long-term issues like climate change.**
* **The sources emphasize the need for proactive discussion, ethical framework development, and global cooperation to navigate the potential impacts of transhumanism.**

The sources present a nuanced and complex view of transhumanism. While it offers potential solutions to pressing global challenges, it also introduces new and potentially significant risks. Carefully considering these benefits and risks, alongside robust ethical frameworks and inclusive global dialogue, is crucial for navigating this potential future.

Ethical Implications of Addressing Global Issues: A Complex Landscape

The sources offer an insightful exploration of the ethical implications of different approaches to tackling global issues, particularly climate change. The conversation focuses on the limitations of traditional approaches like democratic processes and authoritarian measures, leading to a consideration of more radical options like AI-driven solutions and human-machine integration, or transhumanism.

* **Traditional approaches**: The sources acknowledge the urgency of addressing climate change but highlight the ethical drawbacks of conventional methods.
    * Democratic processes are seen as slow and inadequate in handling complex, long-term challenges.
    * Authoritarian measures, while potentially faster, raise concerns about freedom violations and potential abuse of power.
* **Propaganda**: Using propaganda to manipulate public opinion is seen as ethically problematic as it undermines informed consent and has long-term societal impacts.
* **AI as a decision-maker**: The sources suggest AI could be better suited for complex decisions, potentially processing vast data and making unbiased judgments. However, concerns arise regarding AI alignment with human values and potential loss of human agency.
* **Transhumanism**: The sources explore transhumanism, enhancing human capabilities through technology, as a potential solution. This approach is considered less ethically problematic than authoritarian measures or propaganda. However, concerns remain about equity, access, and unintended consequences of human enhancement.
    * A significant concern is the potential for a **class divide** between enhanced and non-enhanced individuals, leading to unprecedented social stratification.
* **Privacy Concerns**: The sources highlight the rapid advancements in AI's ability to interpret brain activity, raising privacy concerns related to surveillance, interrogation, and data commodification. Transhumanism is considered a potential solution, as seamless integration with technology could offer individuals greater control over their data.
* **Ethical Framework Adaptation**: The sources repeatedly emphasize the need to develop new ethical frameworks and governance structures that can handle global, long-term challenges, address privacy concerns, and integrate advanced technologies responsibly.
* **Inevitability of Change**: The conversation indicates a sense of inevitability regarding significant changes to governance and the need for rapid decision-making in the face of existential threats.
* **Natural Evolution**: While some argue that natural evolution for humans has slowed or changed, the conversation acknowledges that humans are still subject to selective pressures, albeit different from our ancestral environment.
* **Global Discussion and Cooperation**: The need for inclusive global discussion platforms and cooperation is emphasized. This is deemed crucial to addressing global issues effectively and ensuring that technological advancements, like transhumanism, don't exacerbate existing inequalities.

The sources conclude that the path forward likely involves a combination of enhancing human capabilities, developing sophisticated AI systems, and creating new governance structures. The conversation suggests that transhumanism may be an inevitable outcome, driven by the desire

to keep pace with AI and the need for rapid ethical adjustments. This emphasizes the need for proactive planning and ethical framework development to shape this transformation in a way that benefits humanity while preserving core values and mitigating potential risks.

TRANSCRIPT

PART 1

1
00:00:00,000 --> 00:00:05,500
Okay. So, you've really been digging deep into a set of articles and

2
00:00:05,500 --> 00:00:10,880
notes. Huh? AI ethics. The fate of humanity. We're really in it now, huh?

3
00:00:11,240 --> 00:00:16,280
Staring into the abyss of the big questions here, expert speaker.

4
00:00:16,500 --> 00:00:20,219
Especially when you get to the part about how we make those planet-scale decisions when we're

5
00:00:20,219 --> 00:00:21,799
talking about stuff like climate change.

6
00:00:22,200 --> 00:00:24,040
Are you ready to unpack all of this with me?

7
00:00:24,760 --> 00:00:30,120
Absolutely. The core issue here really seems to be how we can balance that urgent need to do

8
00:00:30,120 --> 00:00:34,480
something with the ethics of influencing how people act on a global scale.

9
00:00:34,700 --> 00:00:37,560
Right. Every solution has a moral roadblock.

10
00:00:37,919 --> 00:00:41,279
Can you kind of walk us through the options and why the ethics are so tricky?

11

00:00:41,680 --> 00:00:45,099
It's almost like you're having this, like, philosophical debate with yourself in these sources.

12
00:00:45,119 --> 00:00:48,919
There's this immediate need for speed and, like, we don't have the luxury of time with climate

13
00:00:48,919 --> 00:00:50,000
change just looming over us.

14
00:00:50,360 --> 00:00:53,799
But how do we actually get everyone on board for that kind of rapid change, ethically?

15
00:00:54,459 --> 00:00:58,200
Let's unpack that a bit. What specific pathways are we even talking about here?

16
00:00:58,240 --> 00:00:59,919
And when do the ethics get messy?

17
00:01:00,180 --> 00:01:03,099
Well, one option that comes up is this top down approach.

18
00:01:03,680 --> 00:01:07,879
Think global mandates or even population control.

19
00:01:07,879 --> 00:01:09,220
And obviously that's a lot to unpack

20
00:01:09,279 --> 00:01:10,180
Yeah, just a tad.

21
00:01:10,400 --> 00:01:13,300
You can see why it would be effective. But talk about a minefield, right?

22
00:01:14,360 --> 00:01:15,519
Freedom autonomy.

23
00:01:15,639 --> 00:01:16,639
Yeah, no kidding.

24

00:01:16,680 --> 00:01:18,599
Not exactly an easy solution.

25
00:01:18,959 --> 00:01:20,059
So what else is on the table?

26
00:01:20,059 --> 00:01:24,919
We also see some exploration of propaganda and AI driven persuasion.

27
00:01:25,500 --> 00:01:29,360
So like imagine being able to nudge people toward more sustainable choices

28
00:01:29,739 --> 00:01:31,239
just by using algorithms.

29
00:01:31,300 --> 00:01:33,660
OK, so less big brother, more like big nudge.

30
00:01:33,739 --> 00:01:37,339
Yeah, exactly. But it still gets really ethically complex.

31
00:01:37,360 --> 00:01:41,300
Like, does offering enhancement as a solution to something like climate change

32
00:01:41,699 --> 00:01:43,800
just create a whole new form of global injustice?

33
00:01:43,879 --> 00:01:47,620
It's kind of scary how fast we've gone from solar panels to thought control in

34
00:01:47,620 --> 00:01:49,860
this deep dive. Is there even a middle ground here?

35
00:01:50,180 --> 00:01:53,000
Well that's where the sources bring up another option, which is basically

36
00:01:53,000 --> 00:01:55,760
handing the reins over to a super smart AI.

37

00:01:55,839 --> 00:01:57,599
Just let the algorithm decide for us.

38
00:01:57,959 --> 00:02:01,400
Yeah, basically let it make the tough calls for the greater good.

39
00:02:02,000 --> 00:02:06,099
But then we hit another wall. How do we make sure an AI's values are actually in

40
00:02:06,099 --> 00:02:07,900
line with humanity's values?

41
00:02:08,580 --> 00:02:12,679
And how comfortable are we with basically giving up control over our own future?

42
00:02:13,320 --> 00:02:15,460
So we've got damned if you do, damned if you don't.

43
00:02:15,820 --> 00:02:17,539
With a side of existential dread.

44
00:02:17,800 --> 00:02:18,300
Yeah.

45
00:02:18,419 --> 00:02:21,979
Makes you just want to unplug from the Internet entirely, but then things take a

46
00:02:21,979 --> 00:02:23,720
bit of a turn in your research, don't they?

47
00:02:23,940 --> 00:02:28,419
Yes. It's almost as if by working through those options, we land on a surprise

48
00:02:28,419 --> 00:02:29,179
contender.

49
00:02:29,720 --> 00:02:30,800
Transhumanism.

50

00:02:30,919 --> 00:02:33,059
Whoa, hold on. Do you say transhumanism?

51
00:02:33,139 --> 00:02:36,100
Now we're talking brain chips and cyborg futures.

52
00:02:36,279 --> 00:02:39,320
How does that solve anything, let alone the ethics of it all?

53
00:02:39,399 --> 00:02:41,699
Well, not necessarily cyborgs like in the movies.

54
00:02:42,100 --> 00:02:45,639
The source material focuses on how integrating with technology could actually

55
00:02:45,679 --> 00:02:47,899
give us more control over all of this, not less.

56
00:02:48,039 --> 00:02:48,720
OK.

57
00:02:48,720 --> 00:02:53,479
Like imagine if enhancing our minds let us really grasp these global

58
00:02:53,479 --> 00:02:56,399
issues and make decisions with more, well unity.

59
00:02:56,720 --> 00:02:59,759
OK, I think I see where you're going with this, but it still feels like we'd be

60
00:02:59,759 --> 00:03:01,320
leaping into the unknown.

61
00:03:01,960 --> 00:03:06,580
Why would someone see this whole merging with tech thing as the most

62
00:03:06,699 --> 00:03:08,020
ethical option?

63

00:03:08,199 --> 00:03:09,360
It feels so counterintuitive.

64
00:03:10,660 --> 00:03:13,479
Well, it kind of comes down to a few arguments that your research points to.

65
00:03:13,660 --> 00:03:14,660
One is speed.

66
00:03:15,059 --> 00:03:18,279
We need to make big decisions, like FAST, when we talk about things like climate

67
00:03:18,279 --> 00:03:19,000
change. Right. Right.

68
00:03:19,419 --> 00:03:23,899
And transhumanism, if it was something everyone did, I mean,

69
00:03:24,059 --> 00:03:27,619
hypothetically speaking, could help us make those decisions faster

70
00:03:28,059 --> 00:03:29,720
if we were all kind on the same wavelength.

71
00:03:29,979 --> 00:03:33,220
So instead of trying to get everyone to agree, we're going for a collective

72
00:03:33,220 --> 00:03:35,619
cognitive upgrade. That's a pretty big jump.

73
00:03:36,020 --> 00:03:39,259
What else makes transhumanism stand out when we're talking about ethics?

74
00:03:39,899 --> 00:03:41,600
The sources also talk about privacy.

75
00:03:41,740 --> 00:03:42,100
OK.

76

00:03:42,220 --> 00:03:45,339
And we've talked about A.I. getting better and better at understanding.

77
00:03:46,020 --> 00:03:48,360
Us maybe even reading our minds right.

78
00:03:48,419 --> 00:03:49,380
Right. It's a little creepy.

79
00:03:50,419 --> 00:03:52,020
Huge ethical red flags.

80
00:03:52,679 --> 00:03:55,020
But what if- and here's where things get really interesting.

81
00:03:55,059 --> 00:03:56,320
What if we just flip the script.

82
00:03:57,460 --> 00:04:01,779
We merge with technology and redefine what privacy

83
00:04:01,820 --> 00:04:02,820
even means.

84
00:04:03,179 --> 00:04:06,059
It'd be like owning all your data because you are the technology.

85
00:04:06,220 --> 00:04:07,460
OK. Now that's a mind bender.

86
00:04:07,660 --> 00:04:11,179
So instead of worrying about A.I. becoming human, we become more like the A.I.

87
00:04:11,479 --> 00:04:12,899
You level the playing field.

88
00:04:13,779 --> 00:04:16,820
But I feel there's always a but with something this big.

89

00:04:17,359 --> 00:04:21,100
What about the downsides? What about the potential for this tech to create even

90
00:04:21,100 --> 00:04:24,519
more inequality? This is starting to sound a lot like Brave New World to me.

91
00:04:24,579 --> 00:04:26,519
Yeah you're right on the money. That's a big issue.

92
00:04:26,700 --> 00:04:28,720
It's something that your sources address.

93
00:04:28,779 --> 00:04:32,720
You could end up with a two tiered society where the enhanced humans

94
00:04:32,720 --> 00:04:35,700
have more power better opportunities maybe even a longer lifespan.

95
00:04:36,200 --> 00:04:39,820
It makes you think about how technology has often made social divides even worse.

96
00:04:40,220 --> 00:04:42,700
This whole problem of unequal access to enhancement.

97
00:04:43,040 --> 00:04:46,739
I mean it makes you realize that we probably need new ways to think about ethics in

98
00:04:46,739 --> 00:04:47,160
the future.

99
00:04:47,660 --> 00:04:51,040
It's honestly mind blowing. We've hit the ethics wall so hard with this that we've

100
00:04:51,040 --> 00:04:54,420
warped into a whole new dimension of ethical dilemmas.

101
00:04:54,700 --> 00:04:58,320
It makes those initial concerns about propaganda and A.I.

102

00:04:58,339 --> 00:05:00,299
overlords seem almost cute.

103
00:05:00,500 --> 00:05:02,339
I know it's a lot to process for sure.

104
00:05:03,019 --> 00:05:07,019
But this is exactly why these conversations that we're having are so important.

105
00:05:07,320 --> 00:05:10,839
It's like our ability to change things with technology is happening way faster than

106
00:05:10,839 --> 00:05:14,179
our ability to really understand what it all means ethically.

107
00:05:14,739 --> 00:05:17,720
We're always playing catch up and the stakes are incredibly high.

108
00:05:18,140 --> 00:05:19,799
So where does that leave us?

109
00:05:20,059 --> 00:05:24,880
We've gone from talking about global cooperation to brain ships to

110
00:05:24,959 --> 00:05:27,660
a potential future where humanity is divided.

111
00:05:28,000 --> 00:05:30,179
It's a lot to take in even for a deep dive.

112
00:05:30,260 --> 00:05:33,700
It really is but we're not aiming for those easy answers here.

113
00:05:33,959 --> 00:05:36,179
The important thing is that we're uncovering just how

114
00:05:36,179 --> 00:05:38,899
complex all these choices really are and

115

00:05:38,899 --> 00:05:41,380
what's clear from your research is that every path forward

116
00:05:41,380 --> 00:05:43,459
especially when we're talking about these huge global

117
00:05:43,459 --> 00:05:46,399
challenges is going to come with its own ethical trade offs.

118
00:05:46,700 --> 00:05:47,679
There's no right answer.

119
00:05:48,040 --> 00:05:49,179
No easy button to press.

120
00:05:49,239 --> 00:05:51,100
It's almost like when you're an ethical upgrade

121
00:05:51,100 --> 00:05:52,579
just to understand all of this.

122
00:05:52,779 --> 00:05:54,579
But I think I know what you're going to say.

123
00:05:54,619 --> 00:05:55,739
You know me too well.

124
00:05:55,859 --> 00:05:57,519
There's always another layer isn't there.

125
00:05:57,820 --> 00:06:00,500
It might seem kind of bleak but there are.

PART 2
1
00:01:00,000 --> 00:01:03,439
Okay so that's a slightly more hopeful take... are you saying we might just figure it out

2
00:01:03,439 --> 00:01:04,440

as we go along?

3
00:01:04,440 --> 00:01:09,480
I don't know about just figuring it out but maybe... adapting and evolving is a better

4
00:01:09,480 --> 00:01:15,339
way to put it. The sources highlight that we need to get comfortable with the idea that

5
00:01:15,339 --> 00:01:20,980
the questions they are going to keep changing... and maybe even radically in the next few decades.

6
00:01:20,980 --> 00:01:25,720
So we need to get comfortable with being uncomfortable... accepting that the future is uncertain, and

7
00:01:25,720 --> 00:01:29,879
that our ethical frameworks might need to constantly change. Maybe even our definition

8
00:01:29,879 --> 00:01:32,139
of human is up for grabs.

9
00:01:32,139 --> 00:01:37,660
That's exactly it. And this leads into another key point in the source material. This potential

10
00:01:37,660 --> 00:01:43,980
for AI not to decide for us but to guide us through these huge ethical dilemmas.

11
00:01:43,980 --> 00:01:46,980
Okay so... AI as a guide, not an overlord.

12
00:01:46,980 --> 00:01:51,500
Exactly. But not just any AI, we're talking about a very specific kind of AI with a very

13
00:01:51,500 --> 00:01:53,419
specific goal in mind.

14
00:01:53,459 --> 00:01:55,860
And that is, don't leave me hanging.

15
00:01:55,860 --> 00:01:56,760

AI alignment.

16
00:01:56,760 --> 00:02:00,900
AI alignment. I feel like we've talked about that before in some past Deep Dives.

17
00:02:00,900 --> 00:02:05,739
But for those of us who need a little refresher, what does that even mean? Especially in this

18
00:02:05,739 --> 00:02:08,740
context. It's not just about an AI that can like write me a good email.

19
00:02:08,740 --> 00:02:14,259
I'm getting close. AI alignment in this case means creating AI that isn't just smart. It's

20
00:02:14,259 --> 00:02:20,559
about AI whose values and goals, even its sense of ethics, is aligned with our own.

21
00:02:20,580 --> 00:02:25,220
So imagine an AI that's programmed not just to reduce carbon emissions, but to do it in

22
00:02:25,220 --> 00:02:30,580
a way that actually makes people's lives better. An AI that respects things like cultural differences.

23
00:02:30,580 --> 00:02:35,460
It's about making sure that as AI gets more and more intelligent, it doesn't also become

24
00:02:35,460 --> 00:02:37,380
so different from us that we lose control.

25
00:02:37,380 --> 00:02:41,000
Okay, so instead of just handing over control, we're trying to build an AI partner that

26
00:02:41,000 --> 00:02:42,539
can help us figure all this out.

27
00:02:42,539 --> 00:02:47,419
Yeah, exactly. And the sources suggest a few ways that AI could really be helpful with

28
00:02:47,600 --> 00:02:52,979

Like, for example, AI could analyze these ethical dilemmas so much faster and on a much

29
00:02:52,979 --> 00:02:56,220
bigger scale than our human brains ever could.

30
00:02:56,220 --> 00:03:02,779
Imagine running all these simulations, analyzing historical data, modeling different outcomes

31
00:03:02,779 --> 00:03:09,139
on this massive scale, all to try and find the best way forward, or at least the least

32
00:03:09,139 --> 00:03:10,139
bad way.

33
00:03:10,139 --> 00:03:13,259
So it'd be like having the super powered conscience...

34
00:03:13,259 --> 00:03:14,259
Right.

35
00:03:14,259 --> 00:03:18,740
...to keep up with all of the changes that are happening so fast.

36
00:03:18,740 --> 00:03:21,740
It's a really interesting concept, but is it even possible?

37
00:03:21,740 --> 00:03:25,380
I mean, how do we even know if we can create an AI like that?

38
00:03:25,380 --> 00:03:29,660
And even if we could, wouldn't we just be trading one set of problems for another, putting

39
00:03:29,660 --> 00:03:32,339
our trust in a machine to figure out ethics for us?

40
00:03:32,339 --> 00:03:33,339
Yeah.

41
00:03:33,339 --> 00:03:34,339

Those are the million-dollar questions for sure.

42
00:03:34,339 --> 00:03:35,339
Yeah.

43
00:03:35,339 --> 00:03:39,860
And even the sources admit that AI alignment is a really, really difficult thing to achieve.

44
00:03:39,860 --> 00:03:40,860
Yeah, no kidding.

45
00:03:40,860 --> 00:03:44,240
It's like we're trying to program morality into machines, which is something that humans

46
00:03:44,240 --> 00:03:46,639
have been struggling with for centuries.

47
00:03:46,639 --> 00:03:49,759
But there's a little glimmer of hope in your research that I thought was really interesting.

48
00:03:49,759 --> 00:03:53,039
Remember that story about the AI that couldn't understand what it meant to hear, because

49
00:03:53,039 --> 00:03:56,039
it was so focused on like the mechanics of sound?

50
00:03:56,039 --> 00:03:58,979
Yeah, that was kind of creepy, but also kind of sweet.

51
00:03:58,979 --> 00:04:02,380
Like watching a kid try to figure out a really complicated idea for the first time.

52
00:04:02,380 --> 00:04:08,720
What's interesting is that the AI eventually did figure it out, but it took time and patience

53
00:04:08,720 --> 00:04:11,279
and this willingness to kind of like teach it and communicate with it.

54
00:04:11,419 --> 00:04:15,759

So you're saying that maybe the key to all of this is patience, teaching AI about ethics

55
00:04:15,779 --> 00:04:17,839
the same way we teach kids about right and wrong?

56
00:04:17,839 --> 00:04:20,700
It could be a really big part of the puzzle.

57
00:04:20,700 --> 00:04:26,220
Your research really seems to point to a future where AI doesn't replace human judgment completely,

58
00:04:26,220 --> 00:04:27,540
but it helps us improve it.

59
00:04:27,540 --> 00:04:32,040
It helps us become better at making these tough ethical decisions, not obsolete.

60
00:04:32,040 --> 00:04:33,040
Okay.

61
00:04:33,040 --> 00:04:34,420
Now that's a future I can get behind.

62
00:04:34,420 --> 00:04:38,480
So as we wrap up this Whirlwind Deep Dive, where do we land?

63
00:04:38,720 --> 00:04:44,079
I mean, we've gone everywhere, from crisis mode to what feels like a cautious sense of hope.

64
00:04:44,079 --> 00:04:46,399
Well, that's what's so great about these deep dives.

65
00:04:46,399 --> 00:04:49,720
Sometimes the biggest insights come from the journey itself.

66
00:04:49,720 --> 00:04:53,480
And what we've discovered here is that the future, especially when we're talking about

67
00:04:04:53,480 --> 00:04:57,339

AI and ethics and what it means to be human, is not set in stone.

68
00:04:57,339 --> 00:05:01,679
So no easy answers, no guarantee is just a lot of big questions that we've got to figure

69
00:05:01,679 --> 00:05:02,820
out together.

70
00:05:02,820 --> 00:05:06,959
Humans and AI ideally working together and hopefully before it's too late.

71
00:05:07,959 --> 00:05:14,000
And on that note, we'll leave you with this thought to ponder, if becoming more like AI

72
00:05:14,000 --> 00:05:18,760
is one path, and building AI that's more like us is another what happens when those

73
00:05:18,760 --> 00:05:23,559
two paths cross, what would a future look like where humans and AI were actually evolving

74
00:05:23,559 --> 00:05:24,980
together ethically?

75
00:05:24,980 --> 00:05:27,260
Now that is a question for another Deep Dive.

76
00:05:27,260 --> 00:05:31,600
Thanks for joining us for this exploration and remember in a world full of easy answers,

77
00:05:31,600 --> 00:05:33,660
the deep dive is always worth taking.