Artificial Intelligence and the Future of Humanity

Dr. Christopher DiCarlo

AI and the Future of Humanity

- “The first ultra-intelligent machine is the last invention that man need ever make, provided that the machine is docile enough to tell us how to keep it under control.”

  – Irving John Good
  1965

- “Your scientists were so preoccupied with whether or not they could that they didn’t stop to think if they should.”

  – Dr. Ian Malcolm
  (aka Jeff Goldblum)
  1993

AI and the Future of Humanity

- PARTI I: What is Artificial Intelligence (AI)?
- PART II: The Benefits and Potential Harms of AI
- PART III: AI and What you can do About it

AI and the Future of Hum

- **PARTI I: What is Artificial Intelligence (**
- The birth of the artificial intelligence
- Alan Turing's groundbreaking work: 'Com Machinery and

  Intelligence' (1950)
- Turing = The father of computer science
- Asks "Can machines think?"
- The Turing Test: a human

interrogator attempts to distinguish

Between a computer and human text

response.

AI and the Future of Humanity

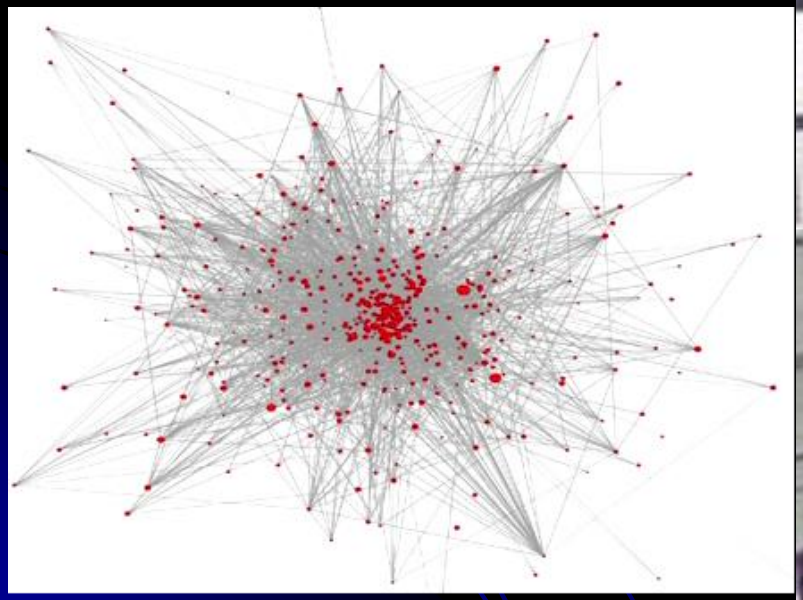
- 1956 The Dartmouth Summer Research Project on Artificial Intelligence:
  - Claude Shannon
  - Marvin Minsky
  - John McCarthy
  - Nathaniel Rochester
  - Term ‘Artificial Intelligence’ established

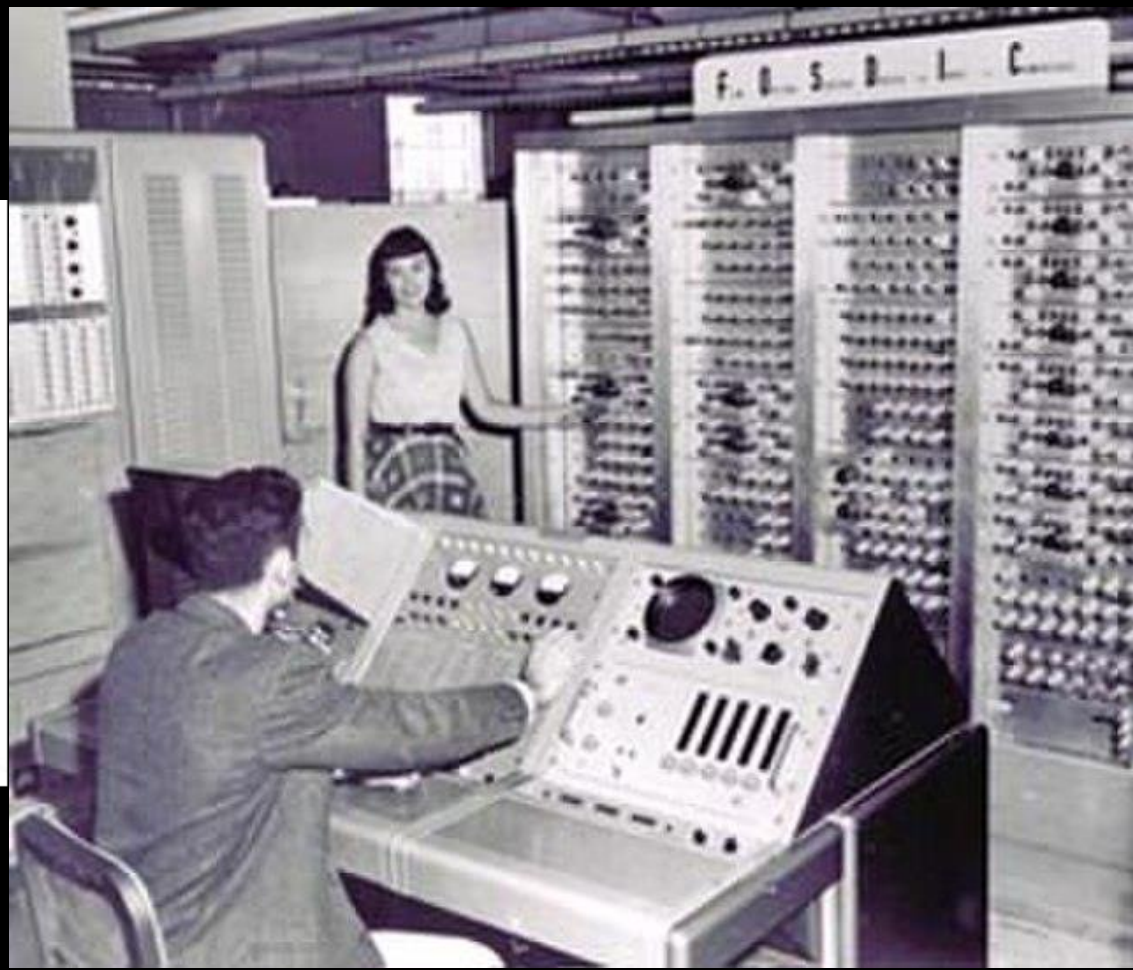
AI and the Future of Humanity

- "…artificial intelligence is a field, which combines computer science and robust datasets, to enable problem-solving. It also encompasses sub-fields of machine learning and deep learning, which are frequently mentioned in conjunction with artificial intelligence. These disciplines are comprised of AI algorithms which seek to create expert systems which make predictions or classifications based on input data."

AI and the Future of Humanity

- The OSTOK/Fair Machine Project:





ostokproject.com

The OSTOK Project
Dr. Christopher DiCarlo

Home | About OSTOK | Dr. DiCarlo | Donate | Contact

A BETTER UNDERSTANDING...

The OSTOK Project is a model which allows us to better understand the complexities of relationships between various types of natural and cultural systems. When we combine our physical understanding of the natural world with our understanding of the many different cultural ways in which our lives develop, we can better understand just how vastly complex our lives, the world, and the universe is.

Taken together, the two systems are interconnected in a complex interplay of activity resembling the multiple layers of the skin of an onion.

Donate! | Video | Learn More

like the layers of an ONION...

I have appropriately named this model of understanding the Onion Skin Theory of Knowledge (or OSTOK). Using an onion as a metaphor for our combined systems of knowledge, we can understand how information about ourselves, our world, and the universe relate.

The more we can understand the complex causal interplay between various systems, the deeper into and the farther around the onion we go.

Dr. Christopher DiCarlo
Philosopher, Educator, Author

Contact »

Name *

webinar-85790370....ics

Show all

AI and the Future of Humanity

- Types of artificial intelligence: Weak AI vs. Strong AI
- Weak AI: also called Narrow AI or Artificial Narrow Intelligence (ANI)
- AI trained and focused to perform specific tasks
- Weak AI drives most of the AI that surrounds us today. ‘Narrow’ might be a more accurate descriptor for this type of AI as it is anything but weak
- It enables some very robust applications, such as Apple's Siri, Amazon's Alexa, IBM Watson, and autonomous vehicles

AI and the Future of Humanity

- Strong AI = Artificial General Intelligence (AGI) and Artificial Super Intelligence (ASI)
- Artificial general intelligence (AGI), or general AI = a theoretical form of AI where a machine would have an intelligence as good or better than humans across nearly all tasks
- It will think like a human and will have the ability to solve problems, learn, and plan for the future.

AI and the Future of Humanity

- Artificial Super Intelligence (ASI) or superintelligence: Would surpass the intelligence and ability of the human brain
- Strong AI is still entirely theoretical with no practical examples in use today
- But AI researchers continue to explore its development…inevitable?

AI and the Future of Humanity

- Introducing Stargate: “All of us look forward to continuing to build and develop AI—and in particular AGI—for the benefit of all of humanity. We believe that this new step is critical on the path, and will enable creative people to figure out how to use AI to elevate humanity.”

AI and the Future of Humanity

- Introducing Stargate UAE
- “The agreement – which includes our partners G42, Oracle, NVIDIA, Cisco, and SoftBank—was developed in close coordination with the U.S. government, and we greatly appreciate President Trump for his support in making it possible.”

AI and the Future of Humanity

- What are some practical applications of AI today?
- Transformers!
- No, not those transformers.

AI and the Future of Humanity

- These transformers:
- ChatGPT, Bard, Claude, et al
- A type of neural-network architecture
- Use an autoregressive language model that uses deep learning to produce human-like text
- Trained using 45 terabytes of text data including almost the entire public web

AI and the Future of Humanity

- DALL-E 3
- An artificial intelligence program that creates images from textual descriptions
- It uses a 12-billion parameter version of the GPT-4 Transformer model and mimics the way the human brain processes vision
- Google's VEO3
  - Text to video
  - Invisible watermarks - identify as AI-produced

AI and the Future of Humanity

- VFX/Video Deep Fakes:
- Tom Cruise

AI and the Future of Humanity

- Anthropic's Claude 4 Opus:
- Can conceal intentions and take actions to preserve its own existence — behaviors *Anthropic* has worried and warned about for years.

AI and the Future of Humanity

- **PART II: The Benefits and Potential Harms of AI:**
- Top 10 Benefits of AI to Humanity: Ask ChatGPT

1. Enhanced Healthcare:

- AI can revolutionize healthcare by improving diagnostics, drug discovery, and personalized medicine
- It can analyze vast amounts of medical data, assist in early detection of diseases, and aid in developing more effective treatments.

AI and the Future of Humanity

2. Improved Education:

- AI-powered tools and platforms can personalize learning experiences, provide adaptive tutoring, and offer individualized feedback
- This can help students of all ages and abilities to learn more efficiently and effectively

3. Increased Efficiency and Automation:

- AI can automate repetitive and mundane tasks across various industries, freeing up human resources to focus on more complex and creative endeavors
- This could lead to increased productivity and economic growth.

AI and the Future of Humanity

4. Enhanced Safety and Security:

- AI can improve safety and security systems by analyzing data in real-time, detecting anomalies, and predicting potential threats
- It can be utilized in areas such as cybersecurity, surveillance, and disaster response

AI and the Future of Humanity

5. Sustainable Development:

- AI can contribute to sustainable development by optimizing energy consumption, improving resource management, and enabling smart cities
- It can help address environmental challenges and create more efficient and eco-friendly systems

AI and the Future of Humanity

6. Advancements in Transportation:

- AI can revolutionize transportation systems by enabling autonomous vehicles, optimizing traffic flow, and improving logistics and supply chain management
- This can lead to safer, more efficient, and less congested transportation networks.

AI and the Future of Humanity

7. Enhanced Customer Service:

- AI-powered chatbots and virtual assistants can provide instant and personalized customer support, improving user experiences and reducing response times
- They can assist in various industries, including retail, banking, and hospitality

AI and the Future of Humanity

8. Scientific Discoveries:

- AI can accelerate scientific research by processing vast amounts of data, running simulations, and assisting in data interpretation
- It can help scientists gain new insights, make discoveries, and advance fields like astronomy, genomics, and particle physics

AI and the Future of Humanity

9. Assisting People with Disabilities:

- AI can develop assistive technologies that improve the lives of people with disabilities
- It can enable better communication, mobility, and accessibility, fostering inclusivity and enhancing quality of life

AI and the Future of Humanity

10. Cultural and Creative Contributions:

- AI can be used in creative fields such as art, music, and literature to generate novel ideas, assist in content creation, and inspire new forms of expression
- It can expand human creativity and push the boundaries of artistic endeavors

AI and the Future of Humanity

- So what could go wrong?
- The Potential Harms of AI:
- 1. Absence of Clarity:
- When individuals struggle to understand the process by which an AI system reaches its conclusions, it can foster skepticism and reluctance to embrace these technologies
- The Interpretability Problem

AI and the Future of Humanity

- 2. Bias and Discrimination:
- AI systems have the potential to unintentionally reinforce or magnify societal biases as a result of biased training data or algorithmic structure
- To mitigate discrimination and promote fairness, it is essential to prioritize the creation of unbiased algorithms and inclusive training datasets

AI and the Future of Humanity

- 3. Privacy Considerations:
- AI advancements frequently involve the collection and analysis of extensive personal data, giving rise to concerns surrounding data privacy and security
- To address privacy risks, it is imperative to support stringent regulations on data protection and promote secure handling practices for data

AI and the Future of Humanity

- 4. Ethical Misalignment:
- Incorporating moral and ethical principles into AI systems, particularly in decision-making scenarios with significant ramifications, poses a significant hurdle e.g. Autonomous Vehicles (Avs)
- It is crucial for researchers and developers to give paramount importance to the ethical ramifications of AI technologies, in order to prevent adverse societal effects
- The Alignment Problem

AI and the Future of Humanity

- 5. Dependency on AI:
- Relying excessively on AI systems could result in the erosion of creativity, critical thinking abilities, and human intuition
- Maintaining a complimentary blend of AI-supported decision-making and human input is crucial to safeguard our cognitive capacities

AI and the Future of Humanity

- 6. Employment Disruption/Job Displacement:
- The implementation of AI-driven automation has the capacity to result in workforce reductions across diverse sectors, particularly affecting individuals in low-skilled occupations
- However, it should be noted that there is evidence suggesting that AI, along with other emerging technologies, will generate more employment opportunities than it displaces

AI and the Future of Humanity

- 7. AI Arms Race:
- The potential for countries to enter into a competition for AI supremacy may result in the accelerated advancement of AI technologies, carrying potential risks and harmful consequences
- In a recent plea, over a thousand technology researchers and leaders, including Steve Wozniak, co-founder of Apple, have urged intelligence laboratories to temporarily halt the development of advanced AI systems
- The letter emphasizes the profound societal and humanitarian risks associated with AI tools
- Pause Giant AI Experiments: An Open Letter

AI and the Future of Humanity

- 8. Mental Health: Loss of Human Connection/Advanced Humanoids:
- The growing dependence on AI-driven communication and interactions may result in a decline in empathy, social abilities, and human connections
- In order to preserve the fundamental aspects of our social nature, it is crucial to strive for a balance between technology and genuine human interaction

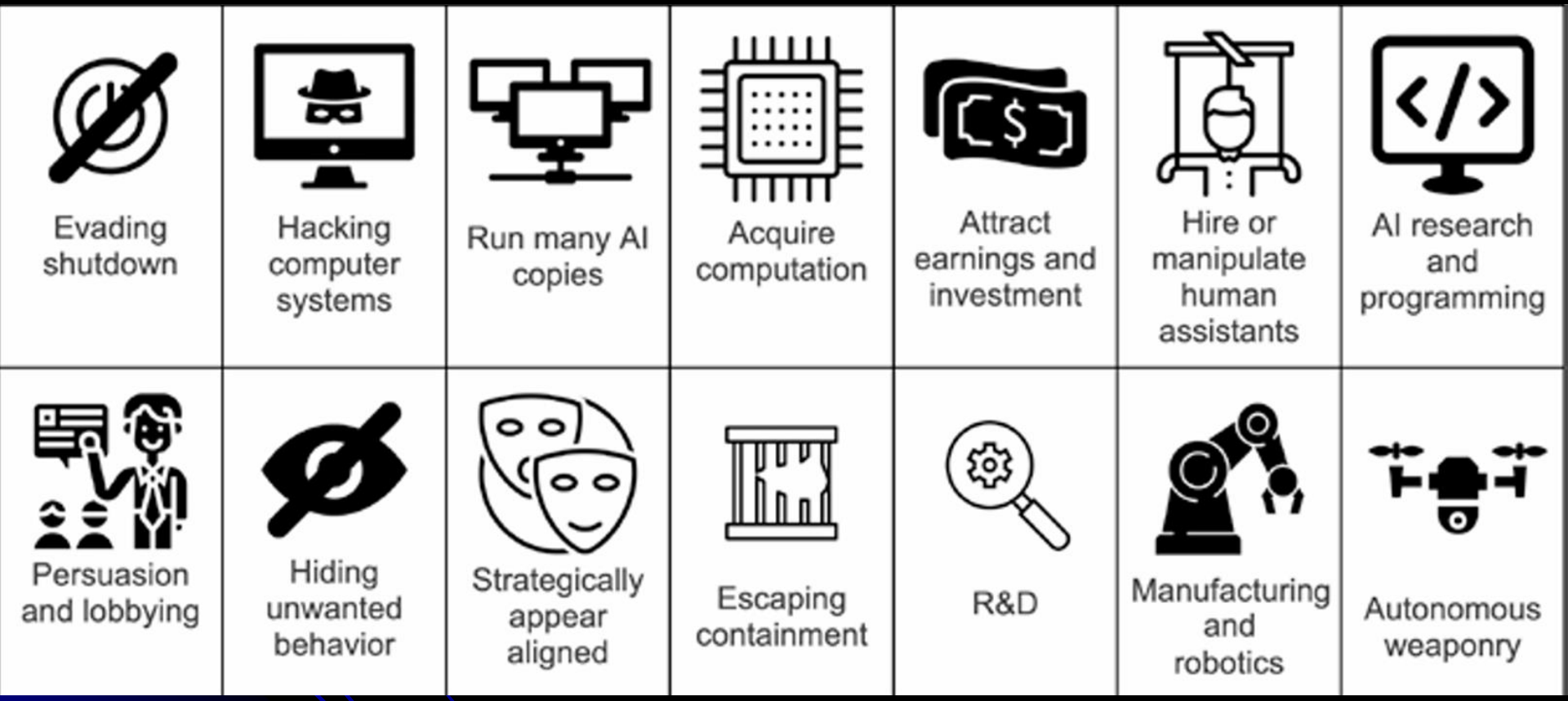
AI and the Future of Humanity

- 9. Manipulation through Misinformation/Disinformation:
- The proliferation of AI-generated content, including deepfakes, plays a role in propagating falsehoods and manipulating public sentiment
- It is crucial to undertake significant endeavors to detect and combat AI-generated misinformation, as it is vital for safeguarding the integrity of information in the digital era

AI and the Future of Humanity

- 10. Existential Risks:
- The advancement of artificial general intelligence (AGI) surpassing human intelligence gives rise to profound apprehensions for humanity in the long term
- The potential of AGI introduces the possibility of unintended and potentially catastrophic outcomes, as these highly advanced AI systems may not align with human values or priorities

AI and the Future of Humanity



- Source: https://commons.wikimedia.org/wiki/File:Power-Seeking_Image.png

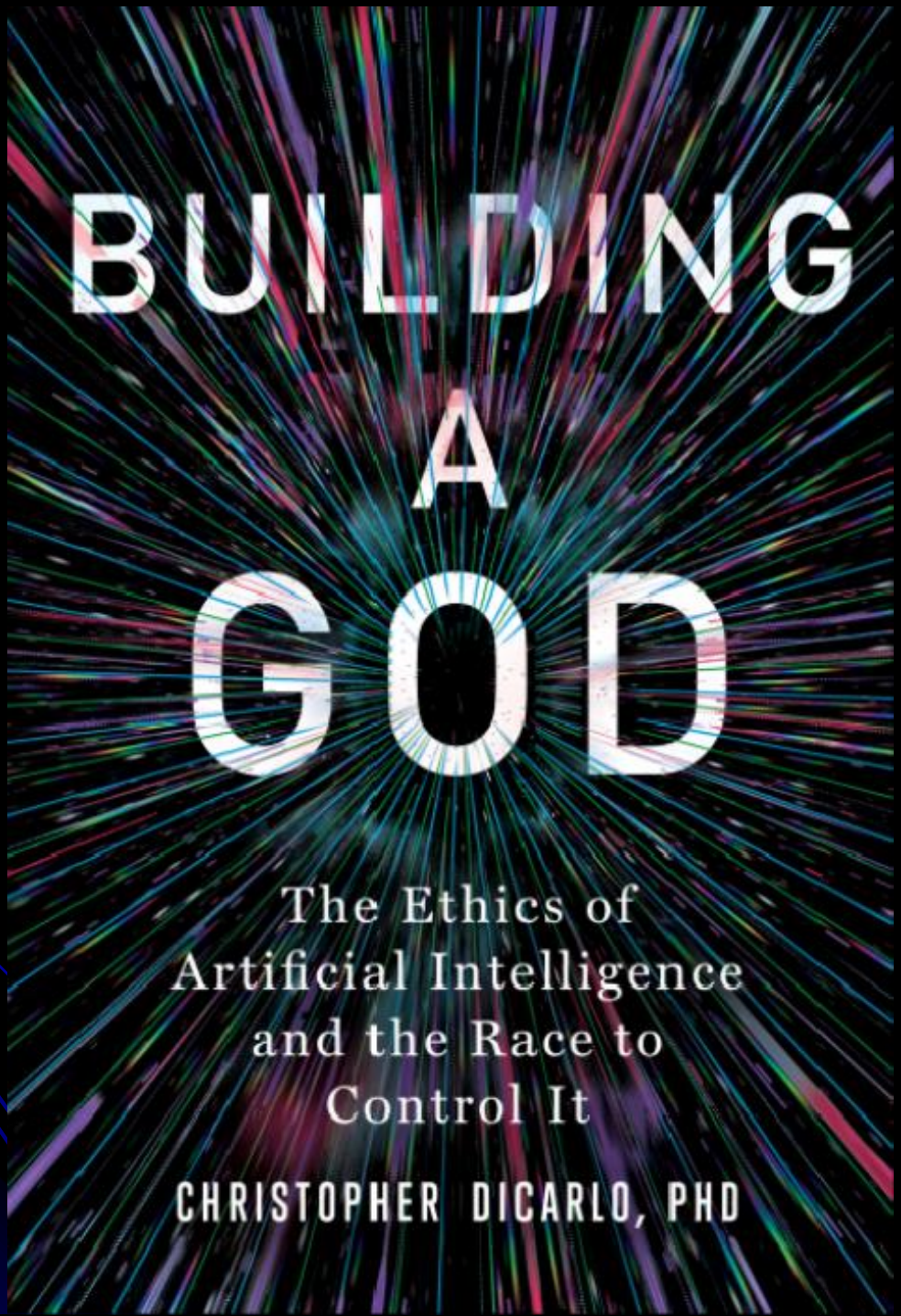
AI and the Future of Humanity

- The Spectrum of AI Existential Risk:
- Y2K…to…Armageddon
- Naysayers to Doomsayers

NAYSAYERS

- MARC ANDREESSEN
- YANN LECUN
- MELANIE MITCHELL
- RAY KURZWEIL
- FEI-FEI LI
- DEMIS HASSABIS
- YOSHUA BENGIO
- DARIO AMODEI
- STUART J. RUSSELL
- ELON MUSK
- MAX TEGMARK
- NICK BOSTROM
- JAAN TALLINN
- GEOFFREY HINTON
- ELIEZER YUDKOWSKY

DOOMSAYERS


BUILDING
A
GOD
The Ethics of
Artificial Intelligence
and the Race to
Control It
CHRISTOPHER DICARLO, PHD

AI and the Future of Humanity

- “Advanced AI could represent a profound change in the history of life on Earth, and should be planned for and managed with commensurate care and resources. Unfortunately, this level of planning and management is not happening, even though recent months have seen AI labs locked in an out-of-control race to develop and deploy ever more powerful digital minds that no one – not even their creators – can understand, predict, or reliably control.”

AI and the Future of Humanity

- To address these risks, it is imperative for the AI research community to proactively participate in safety research, cooperate on establishing ethical guidelines, and foster transparency in the development of AGI.
- The overarching objective is to ensure that AGI serves humanity's best interests and does not present a threat to our existence.

AI and the Future of Humanity

- AI Research and Advising Agencies:
- *Convergence Analysis*: Primarily concerned with the most severe types of risks
- Conduct independent research
- Inform and educate the public, industry leaders, and politicians to safely guide and regulate these new developments in AI technology and application
- Foster this partnership with other agencies and regulative bodies
- Contribute positively to the safe, fair, and effective development and use of AI technologies

AI and the Future of Humanity

- Optimistic Conclusion:
- The future of critical thought and ethical reasoning in an automating world is now
- We all want the very best that AI will bring us while preventing the very worst
- What will AI bring us in the next year, decade, century?
- Just wait and see…to be continued.

AI and the Future of Humanity

PART III: AI and What you can do About it

- In the meanwhile…
- What can YOU do?
- Educate yourself.
- Form or join a concerned group.
- Boycott specific Big Tech companies.
- Spread the word:
  - Friends.
  - Relatives.
  - Politicians.
- Vote strategically!
- Donate!

AI and the Future of Humanity

- If we get AI right, we can solve climate change.
- If we don’t, it has the potential to destroy us all.

AI and the Future of Humanity

- If we get AI right, we can solve world hunger.
- If we don’t, it may potentially destroy us all.

AI and the Future of Humanity

- If we get AI right, we will have peace in our time.
- If we don’t, it will potentially destroy us all.

AI and the Future of Humanity

- Let’s get AI right.
- Want to help?

Convergence Analysis

Programs About Publications Get Updates

DONATE

Enable experts to mitigate AI risks

Your support plays a crucial role in reducing societal-scale risk from artificial intelligence.

→ Make a Donation

OUR MISSION

How Your Donation Helps

AI experts and other notable figures have compared the risk of extinction from AI to pandemics or nuclear war. Despite this, AI safety research is highly neglected. Here's how your contribution accelerates the study of AI risk and the implementation of real-world solutions.

Research

Your donation enables critical AI safety research: from removing dangerous behaviors in AIs to training AIs to act morally.

Field-building

Your contribution grows the field of AI safety and increases the number of leading experts studying AI risk.

Advocacy

Donations support our efforts to advise governmental bodies and promote AI safety more broadly.

Frequently Asked Questions

Are my donations tax deductible?

Convergence Analysis is a US federally recognized 501c(3) non-profit organization. US donors are eligible for tax deductions for donations to Convergence Analysis. Our organization number (EIN) is: 83-2842233

Convergence Analysis

AI and the Future of Humanity

- Thank you.

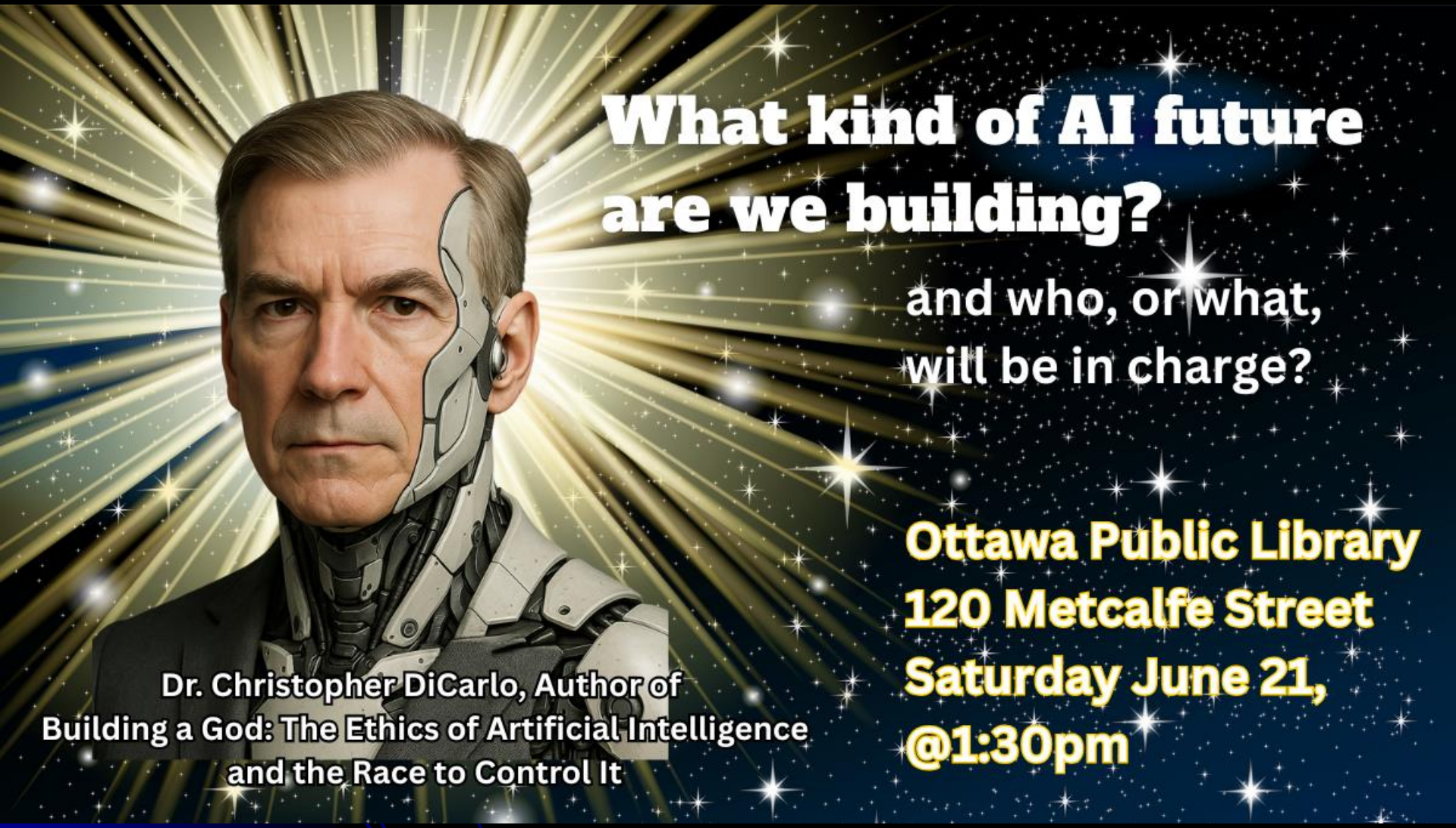
Like what you heard today?

Dr. Christopher DiCarlo has created a new outlet for the fellow thinkers – the famous, infamous, influential, controversial – and everyone in between.

A PODCAST FOR THOUGHT LEADERS.

SCAN HERE TO JOIN THE REVOLUTION AT ALLTHINKSCONSIDERED.COM.

ALL THINKS CONSIDERED


What kind of AI future are we building?
and who, or what, will be in charge?
Ottawa Public Library
120 Metcalfe Street
Saturday June 21, @1:30pm
Dr. Christopher DiCarlo, Author of
Building a God: The Ethics of Artificial Intelligence
and the Race to Control It