ID is 2023-07-0046-PHI-TTS

GPT – A Paradigm Shift for the Twenty-First Century

Joseph R. Carvalko, Jr. *IEEE*

***Abstract*— This paper discusses the rapid advancement of artificial intelligence (AI) in the form of generative pre-trained transformer (GPT) technology such as the latest GPT-4 platform. It highlights the potential of GPT technology to drastically change aspects of society, including creativity, problem-solving, employment, education, justice, medicine, and governance. The author emphasizes the need for policymakers and experts to join in regulating against the potential risks and implications of this technology. The European Commission has taken steps to address the risks of AI through the European AI Act (EIA), which categorizes AI uses based on their potential harm. The legislation aims to ensure scrutiny and control in extreme cases like autonomous weapons or medical devices. However, the author criticizes the lack of meaningful AI oversight in the United States and argues that time has come for government to step in if it desires to make any meaningful change given the technology's (1) rate of diffusion discussed above, (2) multiple products GPT is anticipated to augment, and (3) depths to which it may penetrate daily life, including countless fields of employment.**

***Index Terms*— artificial intelligence, computation and language, deep learning neural networks, NLP, LLM, OpenAI, ChatGPT, BARD, generative AI, generative pretrained transformer, transformer-based AI, European AI Act, EIA, technology ethics.**

"Success in creating AI would be the biggest event in human history. Unfortunately, it might also be the last, unless we learn how to avoid the risks." -Stephen Hawking

I. INTRODUCTION

IF quantum mechanics and computers were the paradigm shifts that defined the twentieth century, artificial intelligence will undoubtedly define the twenty-first. For several weeks I've watched handwringing by everyone from my social media contacts, to TV talking heads and U.S. senators concerned by the latest deep learning neural networks developed by Microsoft, Google, Meta, OpenAI, and others, which generate creative content from natural language descriptions.

The most talked about product in today's news is ChatGPT, which responds to language prompts that result in drafting of emails, essays, poetry, fictional stories, or computer code for executing novel apps.[1] Other products such as Dall-E uses natural language descriptions to produce images of scenes, faces, often in motifs resembling a Dali or Rembrandt artwork. Google Bard is designed more specifically to generate poetry. As a group, these products are considered large language models, or LLMs.

Following ChatGPT's launch in November 2022, it took two months to reach 100 million users. By March 2023, it recorded over 1.5 billion visits. Users of these products are rapidly finding ready-made markets.[2] The technology poses the potential for misuse, for example to intentionally propagate misinformation, spam, phishing, abuse legal and governmental process, and abet fraudulent academic essay writing and social pretexting. A recent example occurred in May, 2023, where a deceptive AI-generated dystopian political advertisement was released by the Republican National Committee, offering a glimpse into how the latest AI tech could be used in next year's U.S. election cycle. The ad prompted Congressmen Yvette Clarke (D-NY) to introduce a bill to require disclosures of AI-generated content in political ads.[3] The power of GPT technology will undoubtedly advance in accuracy and usability. The latest upgrades to ChatGPT, referred to as AutoGPT, claim to create entire websites, conduct market research, and automate similar complex objectives without intermediate human intervention.

II. TECHNOLOGY

Today's LLMs use a neural network architecture called a transformer introduced by a team at Google Brain, in 2017, in connection with its effectiveness in machine translation.[4] Transformers are a type of deep learning model that are designed to process sequential data such as language or time-series data. Subsequently, in 2018, a team at OpenAI reported pre-training a large-scale transformer model using a massive corpus of text data.[5] These initiatives combined to represent a quantum leap in natural language processing, especially demonstrating the successful utilization of parallelization and training on extremely large datasets. Parallelization allows for efficient utilization of computational resources, such as multiple-cores with hyper-threading, or specialized graphical or tensor processing units to operating at the rate of trillions of flops, reducing the time required for training.

A transformer in and of itself comprises a specific type of neural network designed for sequentially processing tokens, such as words in a sentence or data points in a time series.[6] Each token is first embedded into a high-dimensional vector space, which allows the model to compare and manipulate the tokens in meaningful ways. Part of this process importantly includes a self-attention mechanism allowing the transformer to capture relationships between different tokens in the input sequence. For each token, the model calculates a weighted sum of the other tokens in the sequence, where the weights are determined by the similarity between the tokens. This allows

Joseph R. Carvalko, Jr,. is Chair of the Technology and Ethics Research Working Group, of the Interdisciplinary Center for Bioethics, Institution for Social and Policy Studies, Yale University, New Haven, CT, USA (e-mail: joseph.carvalko@yale.edu).

the model to focus on the most relevant parts of the input sequence for each token, and enables it to capture complex dependencies between the tokens.

While it is possible to analyze the input-output relationship to gain some appreciation of a LLM's behavior, explaining precisely how its neural network determines a particular result is impossible, a point I shall elaborate on below as it factors into the potential for producing output having unintended consequences.

The GPT-3 model currently consists of 175 billion parameters and utilizes 96 layers of neural network transformers that operate to compare the input to patterns learned from training data. Transformer layers include self-attention and feedforward networks, which process data and pass results to subsequent layers. Part of the computation process includes estimating the conditional probabilities of generating a word given the previous words in a sequence to predict the next likely word or sequence of words based on context. While techniques like attention maps can provide some insights into which parts of the input text the model focused on, they do not provide a complete explanation of the decision-making process, which utilizes statistical processes.

III. PERFORMANCE

After processing the input sequence through multiple layers of self-attention and feed-forward neural networks, the transformer produces an output sequence that can be used for a variety of tasks, such as language modeling, translation, or sentiment analysis. LLMs do not require supercomputers to run. That said, one cannot ignore that enormous computing power projected by the latest supercomputers will open vast opportunities in the medical, surveillance and military/defense industry, sectors where generative AI applications are bound to have considerable impact in advancing artificial general intelligence (AGI).[7] As the state-of-the-art currently exists, a GPT-4 output already strikingly matches human-level performance, exhibiting signs of AGI.[8] Developers of a GPT class product claim their product can take the written portion of the U.S. multistate bar exam (MBE) and the Graduate Record Exam (GRE) General Test, generating performance levels upwards of 75% and 90% respectively, representative of successful test taker's scores.[9] When supplied a biochemical molecule it turns-on its biochemical expertise to produce variations of the molecule. Concern abounds about the potential of these systems to replace skilled workers in the performance of certain tasks, which now require considerable training, e.g., of artists in the production of graphic arts, of copywriters engaged in the production of advertisements or entertainment. It is also capable of creating new tools for programmers and DIY nonprogrammers, tasked with writing software.

For decades AI has been part of our lives operating below the surface, outside spheres of attention. More recently our attention is drawn to its power to drive social media content, determine access to credit, predict election results, stock performance, weather prediction, sporting outcome prediction, and recommend optimum gambling strategies. It's becoming a central feature in hiring decisions, biomedical analysis, medical device operation, robotic surgery, driverless vehicles, and piloting airplanes.[10,11,12] In all respects modern life runs into AI at some level.

Now more than ever, AI threatens to swallow a chunk of what was once an exclusively human-inspired domain: composition, art, the production of media, and political persuasion. Until now, commentators, educators, elected officials, and government bureaucracies have largely ignored AI and its implications to sway public opinion. But, we ignore its power to persuade at our peril as this paradigm shift will precipitate a reorganization of society on multiple levels, not the least which involves employment, education, justice, medicine, invention, science and government.

Within the next few years GPT will permanently change the nature of creativity or problem solving in mathematics, law, medicine, as well as aiding the physically or psychologically challenged. AI's power and success won't stop we humans from composing, authoring, or inventing, as we are wired to express ourselves in ways that ensure our survival, both materially and aesthetically. Yet, generative AI eventually will foster new inventions. Some of these will take form in utilitarian products and processes, such as new article of manufacture, apparatuses, compositions of matter and creative endeavors as manifested in non-utilitarian objects, such as AI generated human-like avatars, posing as actors, hucksters, and politicians, or as humanoid robots for companionship and commercial utility.[13] Coupled with human contribution to products and processes, societal changes will dwarf the kind of transitions the world experienced going from horse-driven carts to high speed autos, bull horns to television, or snail mail to email.

IV. CALL TO ACTION

As GPT technology becomes accessible to billions of people throughout the world it will increase the potential AI has to change the way societies function. As observed with other world-changing technologies, such as fossil fuels and the Internet, unless we move toward regulation now, soon there may come a time when it will become nigh impossible to effectively plan and manage change.[14] In this regard, I share the clarion call by ethicists, technologist and policymakers that regulation is overdue.[15]

As an ethical and legal principle purveyors of LLM products must ensure that products are helpful, honest, and harmless. As Yampolskiy reminds us, "The unprecedented progress in artificial intelligence (AI) over the last decade, came alongside multiple AI failures and cases of dual use causing a realization that it is not sufficient to create highly capable machines, but that it is even more important to make sure that intelligent machines are beneficial for humanity (citations omitted)."[16] Ethicists have been warning for decades that AI raises the capability claim and the value claim. The first claim considers whether AI can become sufficiently capable of inflicting major damage to well-being. GPT

technologies operate without human supervision.[17] Concerns specific to AI and AGI are well-reported, notably as applied to AI driven autonomous weaponry, the production of BOTs, or the deployment of malicious code.[18] In light of these kinds of applications, GPT-type should be limited in their potential to threaten the value and capability claims, especially via the production of computer code. To this end GPT developers have ongoing programs to develop methods that encode desirable AI behavior in simple and transparent forms, as well as informing our understanding and evaluation of AI decision making.[19,20]

GPT technology because it employs neural networks does not permit a complete understanding about how a decision is made, and thus allow an assessment of its unintended ramifications to health, safety and welfare. In the past, many technologies have been suspected of being potentially harmful to humans or the environment. In some cases it took decades to understand how a technology adversely affected health, such as regarding cigarette smoking or pesticide exposure but overtime science was able to establish the causal connections between a population-wide application of what were physical products and their affect on health. In respect to non-physical, that is intangible products, such as GPT technology, a calculation produces an expression that humans interpret as information. This type of computation/interpretive cause and effect has an ontologically different characterization compared to the cause and effect phenomena that manifest between physical objects. The affect of a chemical on one's physical or psychological condition, cannot be framed in the same way as the cause and effect information has on one's physical or psychological condition. In short, they are different things and the consequences are different, the first being instantiated in physics and chemistry and the other instantiated in a social construction, which is replete with cultural, political and economic implications.

GPT technology does not explicitly store information or provide step-by-step reasoning for their outputs. The models' calculations are distributed across complex neural network architectures, making it challenging to pinpoint exactly how a specific output was generated. As such neural networks pose a special problem in risk assessment as the parameters established through training and algorithmic paths, through which data propagates in reaching a final result, cannot be determined. The risk appertaining to the unknowable issue is aided and abetted by the statistical nature of a GPT model, which limits our understanding of how they arrive at specific outputs. While LLMs can produce impressive results, they lack explicit understanding or reasoning about the content they generate.

There are risks associated with any invention when its inner workings are not fully understood by anyone, including the inventor. This characteristic as applied to AI driven products like GPT has become a field unto itself under the rubric "safety and security."[21] The lack of transparency, in the model's decision-making process, necessitates that users exercise caution and critical evaluation when using GPT outputs in sensitive or high-stakes contexts. Companies such as Anthropic and OpenAI as well as other researchers are actively working on developing methods to improve interpretability and explain-ability of AI models, but it remains an ongoing challenge.

V. ISSUES NEEDING PROMPT ATTENTION

Without understanding how a technology works, its impossible to identify its universe of hazardous or socially objectionable permutations. For instance, although we may reduce the occurrence of threats caused by the production of provably untrue information, we cannot entirely eliminate it. This applies to the impossibility to reliably determine in advance if a particular applications will produce an output that does not conform to a normative capability and value claim. A lack in understanding causation often limits the assignment of responsibility in cases of negative outcomes, creating legal, ethical, and regulatory challenges in determining liability and remediation. There is much to consider but here are three areas of particular concern:

1. Privacy and data protection: GPT technology relies on vast amounts of data, which it acquires from various database. It's been established that training data often includes text from sources, which may contain sensitive, private or proprietary information. If not properly anonymized or stripped of personally identifiable information, the use of this data in training GPT models can result in unintentional exposure of private information. Overall, regulations are needed to protect individuals' privacy and ensure that their data are handled responsibly and securely. Likewise, individuals and entities that produce copyrighted materials must be afforded protection against infringement. This includes instituting regulations regarding data collection, storage, sharing, and consent.

2. Posing and exposing: GPT models can generate text resembling human language, which can potentially lead to adverse inferences of matters private or confidential. For example, if a user interacts with a GPT-based chatbot and provides personal details or discusses sensitive topics, a risk exists that the model might generate responses that at a future point in time indirectly reveal or expose that information. Along related lines, GPT may be employed to generate or inadvertently amplify and propagate false or misleading information, potentially impacting privacy by misrepresenting individuals or groups via fraudulent emails, deepfake content, or targeted advertising based on personal data.

3. Bias and fairness: AI systems can inherit biases from the data they are trained on, leading to unfair outcomes or discrimination.

4. Safety and reliability: Certain GPT-type applications may find their way into applications, such as counselling or medical diagnosis and will have direct impact on human lives, which will require safety standards, testing requirements, and liability frameworks to ensure that these systems are reliable, trustworthy, and do not pose unnecessary risks.[22]

5. Economic and employment impact: GPT has the potential to disrupt industries and reshape job markets, which requires regulation to assist managing these transitions, ensuring fair competition, protecting workers' rights, and promoting responsible adoption of AI technologies.

Overall, regulating the use of GPT-type technology should strike a balance between fostering innovation and protecting societal interests, ensuring that the related technologies are developed and deployed to benefit humanity. Unfortunately, I place little faith in government's ability to alone counter or mollify the adverse consequences of GPT technology. Chief concerns, about a timely government response, stem from: (1) the speed with which the GPT technology is diffusing throughout society, and (2) its power to generate content, both in the non-utilitarian space, such as art, prose, or poetry, and in the utilitarian space of computer code, or invention in the traditional sense, and (3) the inability to comprehend the computational parameters and pathways the technology utilizes in achieving an output.[23]

A lack of understanding on the part of legislators as to how the technology works in the general sense, coupled with the above three concerns limits the effectiveness of regulation in the short term. Technologies such as nuclear power, were understood to have obvious devastating consequences as was demonstrated in 1945, and thus it required relatively little incentive to subject their use to strict regulation. But countless technologies exist that don't immediately manifest their potential for planet altering effects. Fossil fuels, cancer causing chemicals, and social media serve as examples which initially were obscured by a lack of knowledge about their potential to do harm. And often when the potential harms a technology is capable of visiting upon a population become apparent, policymakers have a tendency to ignore the problem because of self-interest, political or otherwise, or large-scale skepticism about whether actions, or even warnings are necessary. Examples are: cigarette smoking, which was found to cause cancer; excessive fossil fuel use, which contributes to climate-change; and in the U.S., the reluctance of many to use masks during the recent COVID-19 pandemic, and the use of assault rifles in the commission of senseless mass murders.

Presuming regulators begin to investigate GPT technology, they need advice from experts in various fields, such as computer science, technology, medicine, social science, economics, intellectual property, and ethics. In the first instance experts will define and specify the various failure modes, e.g., what might harm the public in a particular instance. It is important not to overly constrain the development of the technology, or limit or burden its distribution or application. Never the less the public is entitled to a thorough understanding of the ways in which the technology could harm vital humanitarian interests such as health, safety, welfare, self determination, and creativity.

A preliminary investigation might include a failure mode analysis on GPT technology that identifies potential failure modes, their causes and possible outcomes. Here is partial list of questions, most which are familiar to the engineering community, but may help guide a policy analysis of GPT pitfalls:

1. What are the intended functions and performance requirements of the GPT-type product?

2. What are the possible failure modes that could occur during the GPT-type product's lifecycle?

3. What are the potential causes or factors that could lead to each failure mode?

4. What are the consequences or impacts of each failure mode on the GPT-type product, users, or particular environment?

5. How likely is each failure mode to occur, and what is the severity of its consequences?

6. Are there any existing safeguards or preventive measures in place to mitigate or prevent failure modes?

7. Can the a failure mode be detected or monitored through any means (sensors, inspections, etc.)?

8. What are the potential warning signs or indicators that a failure mode is about to occur?

9. What are the possible actions or countermeasures that can be taken to prevent, detect, or mitigate each failure mode?

10. How can the GPT-type product design or processes be improved to eliminate or minimize failure modes?

Answering these questions may help gain deeper knowledge of potential failure modes allowing for the development of policy strategies to prevent, detect, and mitigate harm.

VI. GOVERNMENTAL REGULATION

The developed world is rife with unbridled commercialization, fierce competition, and political instability, each country through its high-tech establishment pushing the boundaries of technological conquest. However,

with advances in know-how comes responsibility. Powerful tools in the hands of irresponsible agents always threaten the fabric of civilization. Query: will governments heed the warnings and work alongside developers in an effort to advance humane goals, or simply allow AI technology to propagate unconstrained by the value, to "do no harm?"

There have been recent efforts in both Europe and the U.S. to reign in AI initiatives. In October 2022, the White House Office of Science and Technology Policy published The Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People. This is an example of a generalized omnibus type governance applicable to all AI designs and deployment. It does not mandate U.S. policy, but states principles by which the government and industry might find footing. It is discretionary as to its adoption. Other examples of preliminary action are the Department of Defense AI Ethical Principles and Responsible AI Implementation Pathway and the Intelligence Community AI Ethics Principles and Framework and The National AI Initiative Act of 2020, which became law on January 1, 2021. These kinds of overarching principles often gain traction in private industry when it is required that they are adopted as a condition of being awarded government contracts.

In September 2022, the U.S. Food and Drug Administration (FDA) passed regulations dealing with the use of AI in medical devices. Its guidance states that some AI tools should be regulated as medical devices as part of the agency's oversight of clinical decision support software. The guidance includes a list of AI tools that should be regulated as medical devices, including devices to predict sepsis, identify patient deterioration, forecast heart failure hospitalizations, and flag patients who may be addicted to opioids. The FDA recognizes that AI and machine learning particularly have been increasingly incorporated into medical devices because these algorithms are capable of "learning" from experience and improving performance over time.

Specific to AI, the European Commission has identified applications of AI based on their potential for widespread harm and has moved to install the European AI Act (EIA), which addresses risks of specific uses of AI. The EIA applies to AI machine learning, expert and logic systems, and Bayesian or statistical approaches whose outputs "influence the environments they interact with," which includes generative AI products like ChatGPT. The legislation distinguishes four categories of AI use: unacceptable AI risk, high-risk, limited risk, and minimal or no risk. On May 15, 2023, a committee in the European Parliament approved the EIA, which is expected to pass into legislation.

Earlier, the European Parliament resolution of 16 February 2017 offered recommendations to the Commission on Civil Law Rules on Robotics, including stating the principle that Asimov's Laws must be regarded as being directed at the designers, producers and operators of robots, including robots assigned with built-in autonomy and self-learning, since it claims that Asimov's laws cannot be converted into machine code, presumably to prevent a robot from acting against the interest of humanity. The AIA may soften the impact of AI by ensuring that in extreme cases, e.g., autonomous weapons or medical devices, developers of the technology will be subjected to a measure of scrutiny and control. It appears that GPT technology has the potential to sense and express abstractions and human motives, as well as create code, self learn, and therefore lead to the instantiation of the technology into robots, broadly speaking which exhibit autonomous behavior.

Over the course of history, the U.S. has established numerous regulatory agencies to deal with technology. For example, the Federal Communications Commission (FCC) and the FDA regulate communications, drugs, and medical devices, respectively. But when it comes to the more amorphous forms of digital technology, such as data gathering, data security or the reach of the Internet, regulation has remained lethargic. The government has yet to regulate any concrete aspect of social media.

To successfully regulate any technology requires experts in the technology and its application. This has been true for communications as well as medical technology. For example as to drugs the FDA enlists chemists, physicians, statisticians, patients, and policy experts to effectively regulate. The Select Committee on AI, created in June 2018, advises the White House on interagency AI R&D priorities. But neither the Executive Branch nor Congress has yet to assemble any meaningful AI oversight commission. Perhaps it's on its way. This past Spring 2023, the U.S. Senate Committee on the Judiciary, Subcommittee, Artificial Intelligence and Human Rights convened hearings to investigate the general concerns AI poses.[24] These hearing are ongoing, but given the current dysfunction in the U.S. Congress, and the breadth of commercial interests at stake, it may be that attempts to regulate GPT-like technology will not succeed in having any measurable impact for the foreseeable future.

VII. CONCLUSION

LLM-based AI technology, particularly the use of transformers and self-attention mechanisms, enables powerful natural language processing capabilities. It's anticipated that the technology will substantively infiltrate all sectors of the society, affecting the economy, as well as the health, safety and welfare of its citizens. The high potential for GPT technology to change the social, commercial and creative status quo calls for an initiative to carefully consider the value claim and the capability claim as to whether it will always act according to human values, such as "do no harm," which are aligned with those of humanity, and if not whether its actions could cause significant harm. Efforts in both the U.S. and Europe to regulate AI applications have been observed, such as the U.S. FDA's regulation of AI in medical devices and the European AI Act categorizing AI uses and requirements for oversight based on potential harm. The power of LLM-based AI technology coupled with its rapid diffusion into society-at-large, requires a comprehensive plan of oversight and collaboration between government, AI developers and those who will commercialize LLM related products and services. In the U.S. its incumbent upon Congress and the Executive

Branch to heed the warnings and proactively initiate regulation through a new commission to work alongside developers and companies that intend to market LLM applications, to ensure that AI is advanced and controlled in a manner that aligns with humane goals and avoids potential harm.

ACKNOWLEDGMENT

This article is adapted from an online piece published by *Church and State* on the 19 May 2023 http://churchandstate.org.uk/2023/05/ai-a-paradigm-shift-for-the-twenty-first-century/

BIBLIOGRAPHICAL NOTE

Joseph Carvalko is an American technologist, academic, patent lawyer, and writer. As an inventor and engineer he has been awarded eighteen U.S. patents in various fields. He has authored academic books, articles, and fiction throughout his career. Currently he is Chairman, Technology and Ethics Working Research Group, Interdisciplinary Center for Bioethics, Yale University; an Adjunct Professor of Law at Quinnipiac University, School of Law, teaching Law, Science and Technology; member, IEEE, Society on Social Implications of Technology and member of the Publications Board, IEEE Transactions on Technology and Society. His latest book "Conserving Humanity at the Dawn of Posthuman Technology," provides the latest account of AI and genetics from a technical, historical and ethical perspective as well as expectations for its future development.

[1] ChatGPT and Google LaMDA are both language models developed by OpenAI and Google respectively, each having different focuses and capabilities. ChatGPT, developed by OpenAI, is a conversational AI model based on the GPT (Generative Pre-trained Transformer) architecture designed to generate human-like responses to user prompts and engage in meaningful conversations.

[2] The Decoder, July 5, 2023, "ChatGPT reached its preliminary peak with traffic down 10% in June." Available:https://the-decoder.com/chatgpt-reached-its-preliminary-peak-with-traffic-down-10-in-june/

[3] The Verge, May, 2, 2023, Democrat sounds alarm over AI-generated political ads with new bill / After the RNC's dystopian AI-generated attack ad, Rep. Yvette Clarke is calling for more transparency. Available: https://www.theverge.com/2023/5/2/23708310/ai-artificial-intelligence-political-ads-election-rnc-biden.

[4] Vaswani, A., et al, Attention is All You Need. In Advances in Neural Information Processing Systems, pages 6000–6010, 2017. Also see, https://doi.org/10.48550/arXiv.1706.03762.

[5] Radford, Alec and Karthik Narasimhan. "Improving Language Understanding by Generative Pre-Training." (2018).

[6] The LaMDA models have up to 137B parameters and are pre-trained on 1.56T words of public dialog data and web text. See, Cohen, Aaron Daniel, Adam Roberts, Alejandra Molina, Alena Butryna, Alicia Jin, Apoorv Kulshreshtha, Ben Hutchinson et al. "LaMDA: Language models for dialog applications." (2022). [Online]. Available: http://research.google/pubs/pub51115/.

[7] WccFTech, Inflection AI Develops Supercomputer Equipped With 22,000 NVIDIA H100 AI GPUs, July 4, 2023, Available: https://wccftech.com/inflection-ai-develops-supercomputer-equipped-with-22000-nvidia-h100-ai-gpus/?utm_source=www.therundown.ai&utm_medium=newsletter&utm_campaign=could-openai-save-the-world-from-ai.

[8] See, Sparks of Artificial General Intelligence, https://doi.org/10.48550/arXiv.2303.12712

[9] Model Card and Evaluations for Claude Models Dec 19, 2022, https://www-files.anthropic.com/production/images/Model-Card-Claude-2.pdf (Accesses July 14, 2023).

[10] Carvalko, J., "Future of Pharmaco-electronic Medicine," in Quinnipiac Health Law Journal, Vol. 25, No. 3, 2022.

[11] Camacho,D., Luzón,V., Cambria,E., New research methods & algorithms in social network analysis, Future Generation Computer Systems, Volume 114, 2021, Pages 290-293, https://doi.org/10.1016/j.future.2020.08.006. (https://www.sciencedirect.com/science/article/pii/S0167739X20324912).

[12] Park, J., Cheon, M., Hou, S., Lee, O. (2023). Forecasting Election Result via Artificial Intelligence Approach: NLP and Machine Learning. In: Kumar, S., Hiranwal, S., Purohit, S.D.,

Prasad, M. (eds) Proceedings of International Conference on Communication and Computational Technologies . Algorithms for Intelligent Systems. Springer, Singapore. https://doi.org/10.1007/978-981-19-3951-8_57.

[13] Google CEO sounds alarm on AI deepfake videos: 'It can cause a lot of harm' July 10, 2023, https://www.foxbusiness.com/technology/google-ceo-sounds-alarm-ai-deepfake-videos-can-cause-lot-harm.

[14] See, Carvalko, J. Chapters: Societal Repercussions, Policy and Ethics, "Conserving Humanity at the Dawn of Posthuman Technology," (2020), Part VI. Part VII, which provide a roadmap on societal repercussions that will ensue from the confluence of various biological and AI driven technologies, and will require an increasingly important role for government policy and technology ethics.

[15] "The leaders of the ChatGPT developer OpenAI have called for the regulation of "superintelligent" AIs, arguing that an equivalent to the International Atomic Energy Agency is needed to protect humanity from the risk of accidentally creating something with the power to destroy it." https://www.theguardian.com/technology/2023/may/24/openai-leaders-call-regulation-prevent-ai-destroying-humanity#:~:text=The%20leaders%20of%20the%20ChatGPT,the%20power%20to%20destroy%20it.

[16] Yampolskiy, R., Journal of Cyber Security and Mobility, Vol. 11 3, 321–404. doi: 10.13052/jcsm2245-1439.1132, (2022).

[17] It's possible to write a software program that replicates itself, and if that program is a neural network, to learn successful traits gathered from previous generations or iterations where it previously established operating parameters based on weights. AI developed in this way could theoretically improve without needing to be trained from scratch. See, AI Researchers Create Self-Replicating Neural Network, (2018), https://thenewstack.io/ai-researchers-create-self-replicating-neural-network/; Also see, Carvalko, J., The Techno-Human Shell: A Jump in the Evolutionary Gap, pp.95-109. (2013).

[18] Wallach, W. (2015), A Dangerous Master.

[19] Discovering Language Model Behaviors with Model-Written Evaluations, arXiv:2212.09251v1 [cs.CL] (2022).

[20] Anthropic, "Claude's Constitution," (2023), https://www.anthropic.com/index/claudes-constitution, Accessed: 2023-07-08.

[21] Yampolskiy, R., Artificial Intelligence Safety and Security. (2018).

[22] S. Hamdoun, R. Monteleone, T. Bookman and K. Michael, "AI-Based and Digital Mental Health Apps: Balancing Need and Risk," in IEEE Technology and Society Magazine, vol. 42, no. 1, pp. 25-36, March 2023, doi: 10.1109/MTS.2023.3241309.

[23] "Technology diffusion can be defined as the process by which innovations are adopted by a population. Whether diffusion occurs and the rate at which it occurs is dependent on several factors including the nature and quality of the innovation, how information about the innovation is communicated, and the characteristics of the population into which it is introduced. …" [Online]. Available: https://sphweb.bumc.bu.edu/otlt/MPH-Modules/HPM/AmericanHealthCare_Technology-Drugs/AmericanHealthCare_Technology-Drugs2.html.

[24] See, https://www.judiciary.senate.gov/committee-activity/hearings/artificial-intelligence-and-human-rights.