



Responsible Generative AI

An Examination of Ongoing Efforts to Tame This Powerful Technology

2024-1-9

Authors:

Cheranellore
Vasudevan
IBM
vasuv@us.ibm.com

Michael Linehan
OMG
linehan@omg.org

Chuck Byers
OMG/IIC
byers@omg.org

Natalie N. Brooks
IBM
brooksn@us.ibm.com

Luis Freeman
IBM
freemanl@us.ibm.com

CONTENTS

- Abstract 4**
- 1 Introduction..... 4**
- 2 Trustworthiness of Generative AI (GenAI)..... 6**
 - 2.1 Accuracy, Bias, and Hallucination 7
 - 2.2 Filling in the Blanks 7
 - 2.3 What is the Impact?..... 7
 - 2.4 Corrective Steps and Guidelines 8
 - 2.5 Research Trends 9
- 3 Impact on Labor 9**
 - 3.1 Types of Professions That May Get Impacted 10**
 - 3.1.1 Software Development..... 10
 - 3.1.2 Law and Governance 10
 - 3.1.3 Accounting 11
 - 3.1.4 Medical/Health Care..... 11
 - 3.1.5 Design 12
 - 3.1.6 Customer Support 12
 - 3.1.7 Finance and Investment 13
 - 3.1.8 Sports/Games 13
 - 3.1.9 Marketing 13
 - 3.2 Transformation of Workforce 14
 - 3.3 How the Disruption Can Be Managed 15
 - 3.4 Responsible Use Case Scenarios 15
- 4 Impact on IP/Copyrights..... 16**
 - 4.1 Examples of Concern 16
 - 4.2 How Can We Protect? 16
 - 4.3 How GenAI Can Be Useful 17
- 5 Impact on Learning, Cognition, and Intellectual Skills..... 17**
 - 5.1 Impact on Learning and Memorization 17
 - 5.2 Impact on Value of Education and Skills 18
 - 5.3 Evolutionary Impact on Human Cognitive Abilities 18
- 6 Impact on IoT Scenarios 19**
 - 6.1 Background 19
 - 6.2 Smart Sensors 19
 - 6.3 Intelligent Digital Twins 20
- 7 Impact on Environment..... 20**
 - 7.1 Energy Requirements of GenAI 20
 - 7.2 GenAI Energy Estimate 21
 - 7.3 Human Skills and Training needed..... 22
 - 7.4 Data Requirements..... 23
- 8 Need for Regulating GenAI 24**

Responsible Generative AI

8.1 Why Regulations Are Necessary 24

8.2 Current Efforts..... 24

8.3 Innovation Vs. Regulation 25

8.4 Global Governing/Regulatory Organization 26

9 Conclusion 26

10 References 27

11 Acknowledgements..... 31

FIGURES

Figure 1-1: Painting created by Midjourney. Courtesy Smithsonian Magazine, September 2022. 5

Figure 7-1: GenAI hardware infrastructure example. 22

ABSTRACT

The world of AI is undergoing a major transformation with the advent of the new Foundation Models. Foundation Models are, no doubt, very powerful with amazing generative capabilities to produce original lookalikes in various disciplines like literature, art, software, science, scholastic work, and other creative areas. However, the power unleashed by Generative AI technology has triggered several concerns about its limitations and impact on human society and surrounding ecosystems.

Governments, social scientists, technologists including many heads of AI organizations are expressing concerns about the potential negative impacts of Generative AI and the need to regulate this emerging technology. This paper presents detailed investigation and assessments of the concerns centered around Generative AI and discusses the guard rails that need to be put in place for effective and safe use of this technology. Specific topics that are covered include concerns of bias, accuracy and non-transparency of Generative AI Models as well as how this technology impacts labor, IP/copyrights, privacy, education, basic human cognitive skills, and the carbon footprint.

1 INTRODUCTION

The world of AI models is undergoing a significant transformation with the advent of the newly developed Foundation Models. Foundation Models are very powerful with impressive generative capabilities to produce original look-alikes in various disciplines like literature, art, science, scholastic work, and other creative areas. Governments, social scientists, technologists, and even the heads of AI organizations are expressing concerns about the potential misuse of the power unleashed by Foundation Models and Generative AI.

The impacts are in different facets– labor market, IP and copyrights, plagiarism, education, basic human cognitive skills, of course, the carbon footprint. Our aim in this paper is to address these impacts and concerns, whilst remaining hopeful and optimistic about the potential benefits of this technology.

The new AI models¹ that provide generative capabilities are trained on massive data – texts, images, audio, and video and structured data in tables and files. These deeply trained models are not only good at recognizing patterns but can also generate/synthesize new outputs based on what they are trained on. This leads to the capability to compose answers, create summaries, and generate new images, audios, and videos.

While the deep learning transformer models had shown similar capabilities earlier, they reached a threshold limit when trained on massive amount of information, leading to humungous models,

¹ <https://arxiv.org/pdf/2306.17170.pdf>

Responsible Generative AI

with hundreds of billions of parameters. Learning from unlabeled data (recognizing the labels by itself) improved this massive training.

The picture below is an award-winning painting created by Midjourney – a generative AI program.

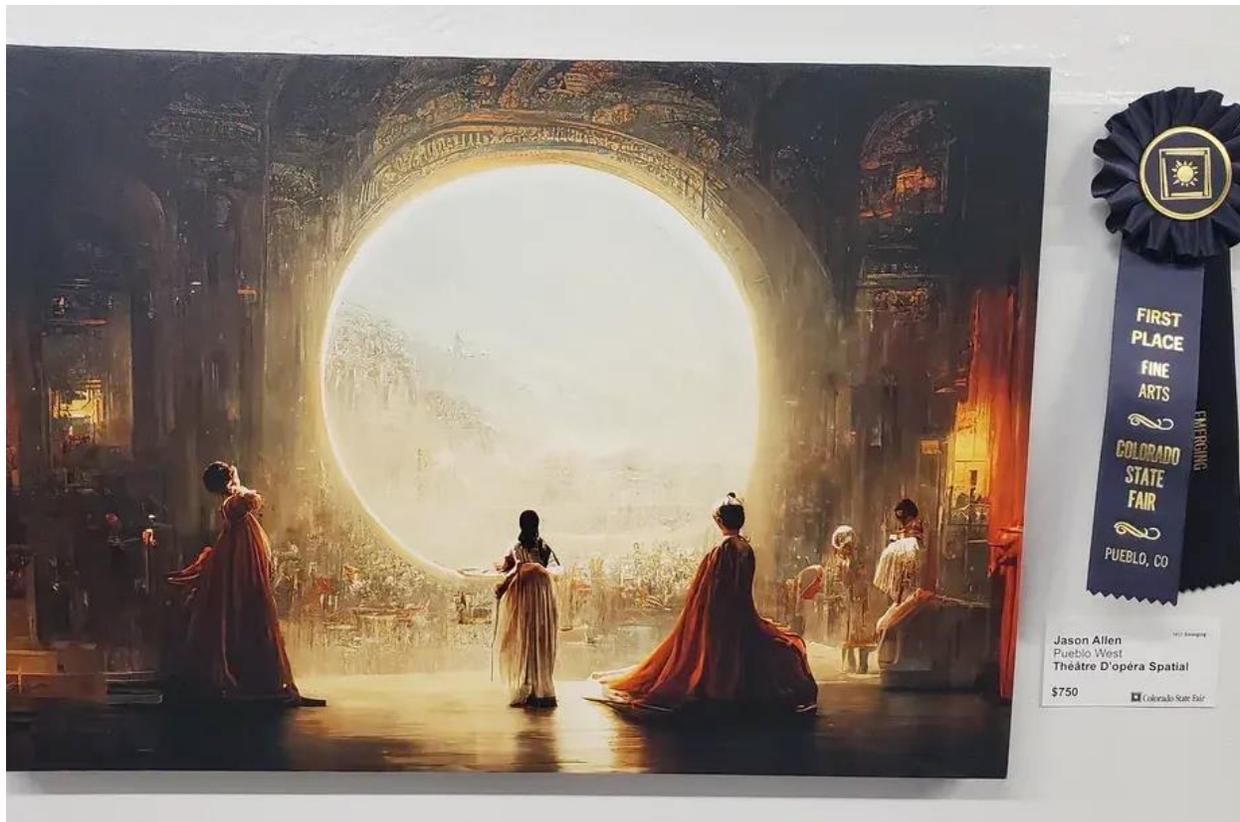


Figure 1-1: Painting created by Midjourney. Courtesy *Smithsonian Magazine*, September 2022.

It was reported that GPT-4 passed the American Uniform Bar Examination, in the 90th percentile. Another report states that some foundation models outperformed many medical graduates in medical school exams. Some of the leading tech companies have demonstrated how bug free, well annotated code can be generated using Generative AI. These are very impressive capabilities of current day AI models. At the same time, it is important to note that “most” Generative AI applications are simply composing from what the model has seen or what it has been trained on.

The behavior of creating something apparently new or answering questions sometimes is good enough to be valuable, but not necessarily a sign of real intelligence which is often demonstrated by traits like ability to learn from context, make decisions, solve problems, and create new artistic, scholastic, and physical artifacts. The fact that these systems are unable to explain their reasoning, and most often they are not based on domain principles, is unsettling.

But this should not be unnerving. After all, a good auto mechanic can repair an engine or even build one without knowing anything about thermodynamics, combustion, or principles of gear mechanisms. Foundation models are no different. They can do an amazing job in solving some

seemingly deep cognitive functions. While the performance is impressive, given that these models show no signs of having domain basics (understanding the science, rules, and principles), there are lots of concerns about their widespread proliferation.

Concerns emanate from the unleashed power to do many things that educated professionals do and the fact that those tasks are done as a backbox. The following sections cover some of areas of concerns such as lack of trustworthiness, unfair impact on labor force, issues with copyrights and IP, impacts on human cognitive skills, and hence the need for appropriate regulation on the development and use of this technology.

As in the case of any fast-paced technologies, the Generative AI domain is quickly evolving and the capabilities of technology and efforts on governing the AI applications may get outdated as this paper gets published. The perspectives presented here are limited to the time at which the data is collected and documented.

2 TRUSTWORTHINESS OF GENERATIVE AI (GENAI)

A major concern with the easy access to GenAI models is their trustworthiness. Can a user trust this AI system? The National Institute of Standards and Technology (NIST) describes the essential components of AI trustworthiness as

- Validity and reliability
- Safety
- Security and resiliency
- Accountability and transparency
- Explain-ability and interpretability
- Privacy
- Fairness with mitigation of harmful bias

These models have the potential to create fake outputs, wrong answers that could be deceptive, addictive, and hence dangerous. Seemingly convincing answers to some set of questions, the power to generate texts and images often paralleling humans have created an illusion of the capabilities of this technology. There are many instances where the outputs of even well-advertised GenAI Models from industry-leading companies are found to be questionable, misleading, and often hallucinating².

² <https://www.ibm.com/topics/ai-hallucinations>

2.1 ACCURACY, BIAS, AND HALLUCINATION

Some recent research papers³ cover detailed experiments on the correctness of different commercially available Large Language Models (LLMs) like ChatGPT, etc.

In some cases, LLMs have shown capabilities of deception including cheating and demonstrating innocence. LLMs can reason their way into using deception as a strategy for accomplishing a task. Sycophantic deception in which chatbots agreeing to a popular conversation has been cited as a tendency⁴.

There are some experimental studies which demonstrate that large language models often repeat common misconceptions since they are trained on language texts created by humans and trained on sometimes false data on the internet. Often the explanations produced by these models, although they seem to be plausible yet misleading, risk trust and safety of their use⁵.

Incorrect answers, unsubstantiated conclusions, toxicity, lack of robustness, unethical solutions, bias-influenced advice, and overconfidence on model performance are some of the concerns expressed by even the model developers and technical leaders⁶.

2.2 FILLING IN THE BLANKS

The LLMs are trained in available data, especially those available on the internet, some of which can be inaccurate, biased, and not proven. When the source itself doesn't reflect the truth, the model trained on the same data obviously cannot be any better. Secondly, the generative part of the LLMs is mostly a composition of the patterns (text, sound, or image) it has seen in the past. This composition happens by a "fill in the blanks" kind of approach from the learned patterns without having any validation of the domain principles, and relationships have no resemblance to the real physics and logic of the domain.

2.3 WHAT IS THE IMPACT?

The lack of trust in GenAI may lead to several risks. An IBM Business Value Institute report lists the following potential risks:

- Increasing legal liability.
- Introducing new vulnerabilities.
- Answers not identifying data sources.
- Generating outcomes inconsistent with society's expectations.

³ <https://arxiv.org/pdf/2308.14752.pdf>

⁴ <https://arxiv.org/pdf/2303.08774.pdf>

⁵ <https://arxiv.org/abs/2305.04388>

⁶ <https://arxiv.org/abs/2306.11698>

Responsible Generative AI

- Spreading misinformation.
- Asserting incorrect information as fact.
- Prejudicial or preferential propositions.
- Tricking and manipulating people.

There are other risks too such as “herd mentality” or monocultural society where everyone thinks and acts the same way because they are advised by the same AI models.

In fact, many AI models have similar risks; however, they are amplified with the generative capabilities of the newer models like ChatGPT, Google Bard, and similar ones. Fortunately, many socio-economic leaders, research scientists, and political leaders are already aware of these although most of them complained that we don’t have adequate regulatory controls and guidelines to manage the safe proliferation of the powerful GenAI models and their potential misuse.

The potential to “hallucinate” (leading users to believe complete non-factual matters) is indeed very alarming. We heard many disturbing stories such as “AI generated photo of Pentagon explosion causes panic on Twitter even affecting stock market⁷.”

Morphing of images and altering videos are not new. The ability of social media to disseminate fake data on an industrial scale makes the ability to easily create fake data extremely dangerous. The new reality is that it has become much easier to create such fake entities – news, audio, or video – very quickly with the generative functions of some AI models. Many political and technology leaders have expressed concern on potential impact on elections with the advent of fake entities which can be created very easily and convincingly using GenAI tools.

2.4 CORRECTIVE STEPS AND GUIDELINES

Installing guard rails with appropriate tooling is the technological answer to the issues of trust, bias, and incorrect results created by GenAI. Harvard Business Review outlines several guidelines to ensure better trust and accuracy. These include organizations and model builders focusing on:

- Training the models on their own data.
- Enabling the users to validate.
- Publishing the sources of data and uncertainties in areas where results are not reliable.
- Explaining why certain results are produced.
- Preventing certain tasks (illegal, unethical, not trained on) from being performed.
- Conducting continuous bias, explain ability, and robustness assessments.
- Respecting data provenance, privacy, and ensuring consent to use.

⁷ <https://www.nytimes.com/2023/05/23/business/ai-picture-stock-market.html>

Responsible Generative AI

- Enabling accessibility to all.
- Defining the application use cases and their scope (just empowering rather than allowing fully autonomic).
- Testing extensively and encouraging feedback.

The parallel approach is having industry guidelines, best practices, and governmental regulations to enable moderated use and appropriate monitoring, to reduce the effect of the misguidance and hallucinatory results of the GenAI. This is discussed in more detail in a later section.

2.5 RESEARCH TRENDS

Researchers and organizations are focusing on new tools, measures, and validations for improving the trustworthiness of GenAI Models. For example, Stanford University Researchers touts a general bias measurement framework – DivDist⁸. The Pentagon’s new Center for Calibrated Trust Measurement and Evaluation initiative⁹ is expected to begin before the end of September focusing on AI trust.

A research paper¹⁰ on evaluating AI Models for “extreme risks,” points to some interesting facts such as having deeper tools/practices to evaluate models for potential threat, bias, etc., could lead to such techniques falling into “rogue” hands to misuse them. This article also points to the fact that the risks may often be ignored because of fierce competition among companies, organizations, and countries to dominate in this power technology area.

Thankfully, there is an urgent and hectic activity by AI researchers and AI companies to develop and deploy technical solutions to the issues of trustworthiness of AI models.

3 IMPACT ON LABOR

Sam Altman, CEO of OpenAI says that this new AI transformation will eclipse the agricultural revolution, industrial revolution, and the internet revolution all put together. Each of the previous “technology revolutions” have disrupted the labor market. The impact of Generative AI technology on the labor market has a major difference.

While automation through technology replaced the physical labor of humans, AI is changing the ways that “white-collar” jobs are performed. Intellectual skills were always the superior asset of an educated and professionally trained labor force which is now under a big threat.

⁸ <https://arxiv.org/abs/2212.11672>

⁹ <https://defensescoop.com/2023/08/29/pentagon-to-launch-pilot-focused-on-calibrated-trust-in-ai/>

¹⁰ <https://arxiv.org/abs/2305.15324>

3.1 TYPES OF PROFESSIONS THAT MAY GET IMPACTED

Professions that will be touched may range from low end reception desk personnel and customer support professionals to high end of cognition such as software programmers, legal attorneys, medical practitioners, and even threatening creative writers, artists, and musicians.

3.1.1 SOFTWARE DEVELOPMENT

In recent decades, software development has been thought of as a “safe profession” despite the onslaught of automation; however, foundational models trained on millions of lines of software code have already demonstrated the capabilities of generating bug-free and well-annotated code samples. That is the “dream comes true” moment for AI researchers, who predicted that AI machines will be able to be self-programmed. Software development is very human-labor centric. Like in any field, run-of-the-mill practitioners in this field too could cause lots of pain by creating erroneous code.

GitHub Copilot is a GenAI tool that helps developers write code. According to this article from Harvard Business Review¹¹, in a six-week pilot at Deloitte with 55 developers using GPT-3's Codex tool, a majority of users rated the resulting code's accuracy at 65% or better. Overall, the Deloitte experiment found a 20% improvement in code development speed.

Brown University researchers recently reported¹² how they leveraged large language models throughout the entire software development process, streamlining and unifying key processes through natural language communication, thereby eliminating the need for specialized models at each phase. CHATDEV, a virtual chat-powered software development application, was successfully mirrored after the established waterfall model with commendable efficiency and cost-effectiveness unveiling fresh possibilities for integrating LLMs into the realm of software development.

3.1.2 LAW AND GOVERNANCE

There are multiple use cases in the legal profession that could dramatically alter the way the lawyers work and impact their professional job market.

Most lawyers - especially junior attorneys - spend their time on three tasks:

- Understanding/creating a brief on a case filed as part of noting the significant points.
- Preparing a formal legal filing which may have a standard template and format.
- Researching similar previous cases, arguments, and verdicts, and associated law points (excerpts from laws prevailing in that locale).

¹¹ <https://hbr.org/2022/11/how-generative-ai-is-changing-creative-work>

¹² <https://arxiv.org/abs/2307.07924>

Responsible Generative AI

In all the above scenarios, generative AI can take over a significant part of the tasks by automatically summarizing case documents, generating new case filing documents, and answering queries from troves of historical legal case incidents and records.

Powerful GenAI applications trained on legal frameworks and historical case diaries can lead to automation of this profession which is currently very human labor oriented. What it would entail is replacing or eliminating the jobs of legal assistants, judicial clerks, and junior attorneys. The profession may also be impacted (if permitted by courts), the plaintiffs and defendants directly preparing the court filings/responses and getting legal advice from professional attorneys.

3.1.3 ACCOUNTING

Accountants spend substantial amounts of time on data entry, transaction categorization and reconciliation. GenAI can automate the process of data entry by extracting information from documents, invoices and receipts eliminating the need for manual entry. It can analyze and categorize transactions based on trained data from previous financial records. GenAI can also assist on the creation of Financial Statements like Income, Balance Sheets, and Cash Flow. From a Tax compliance point of view, GenAI systems can help stay compliant with tax regulations.

While GenAI may not completely replace accounting jobs, it will certainly transform the job roles and distribution of responsibilities. Many Solutions and Tools for Accounting leveraging AI are available today. However, none of them seem to leverage GenAI yet. Tools like Sage and Deloitte TrueVoice use Natural Language Processing to generate financial reports from numerical data and can also provide insights and advice based on the data analysis.

KPMG Clara is an audit platform using machine learning, natural language processing, computer vision, and other technologies to automate audit processes, analyze data, detect anomalies, and provide recommendations.

3.1.4 MEDICAL/HEALTH CARE

According to a report from the National Center for Health Workforce Analysis, the average number of physician shortages across all specialties is projected to increase from 6% in 2025 to 8% by 2035. Some specialties like thoracic surgeons and ophthalmology have projected shortages of 30% and above. The medical field is poised to benefit from GenAI helping reduce those gaps. With the ability to learn specialized data, adapt and automate complex processes GenAI has the potential to develop specialized knowledge for different medical specialties enhancing clinical efficiency and improving patient outcomes.

GenAI can be used to reply to patients' basic questions on health, or medications they are taking, thus freeing physicians to increase the number of patients they can cover¹³. GenAI can provide a

¹³<https://www.forbes.com/sites/forbestechcouncil/2023/06/20/the-impact-of-large-languages-models-on-healthcare/>

second opinion to improve diagnostic accuracy, predict disease progression and support clinical decision making. GenAI can be used to extract missing patient Data from unstructured text such as discharge summaries. In one study¹⁴, a Natural Language Processing Model was able to recover 31% of missing patient data (height, weight, blood pressure), leading to a less biased data set for patient outcome analysis. The medical field, however, comes with its own challenges and limitations like complexity of medical language and medical contexts, patient privacy, regulatory and legal frameworks, and ethical considerations.

3.1.5 DESIGN

Visual design, Media design, and industrial design can be greatly benefited with GenAI capabilities to translate vague high-level instructions on the concept. These include design of visual games, web portals, consumer appliances and gadgets, toys, creative advertisements, posters, logos, emblems, cartoons, building architectural design, and so on, where the human designers can quickly create several experimental patterns/illustrations before finalizing one. Media synthesis involves composing new music, tunes, slogans, and creative articles, poems, and stories.

3D Models for Game Development and Studio Animation are assets that take several weeks of back and forth and hundreds of dollars to produce. With the assistance of GenAI these turnaround times and costs can be reduced substantially.

Examples of 3D tools available today that leverage GenAI for 3D modeling are as follows¹⁵:

Get3D by Nvidia is a representative example which was trained using only simple 2D images. It can generate 3D shapes with high-fidelity textures and robust geometric details. Luma AI's Imagine 3D tool allows you to enter a text input to generate a fully solid 3D Model with a full color texture.

3.1.6 CUSTOMER SUPPORT

One of the jobs that will be impacted the most by GenAI is customer support. The ability of GenAI to comprehend and generate human-like responses will reshape the landscape of customer service.

Expedia has beta launched a new in-app travel planning experience powered by ChatGPT that allows users to ask questions and get recommendations on travel, lodging, and activities.

GenAI lacks the nuanced understanding and emotional intelligence that human agents offer. Instances where customers seek personalized assistance or express complex emotions might

¹⁴ <https://www.nature.com/articles/s41746-022-00590-0>

¹⁵ <https://medium.com/echo3d/7-generative-ai-tools-for-3d-asset-creation-97dd88153b7>

Responsible Generative AI

require a human touch. Striking the right balance between AI-driven efficiency and human empathy remains a key challenge.

A report from Accenture estimates that 80% of Customer Support will be satisfied by GenAI tools. The remaining 20% will require the human factor.

3.1.7 FINANCE AND INVESTMENT

According to this report from Accenture, one multinational bank is using GenAI to transform how it manages volumes of post-trade processing emails—automatically drafting messages with recommended actions and routing them to the recipient. The result is less manual effort and smoother interactions with customers¹⁶.

3.1.8 SPORTS/GAMES

Simulation of game strategies, scenarios, and corrective actions can be hugely benefited by creations of GenAI. A virtual coach based on LLMs will not be a fictitious concept in the near future. This “game simulation” and endless “virtual scenario practices” can be an excellent training tool in the absence or even replacing human coaches.

GenAI will also improve coaching. GenAI will provide the opportunity to analyze an entire game from the sidelines and predict the next best play for the team to win the game. Not only will GenAI provide benefits to coaches but players as well. Giving players access to GenAI tools can allow the players to make real time corrections on the court, the field, and in every arena.

The US Tennis Association in collaboration with IBM’s WatsonX platform has added digital Fan features on USOpen.org and US Open App that provide spoken commentary using GenAI designed to give fans a more insightful and engaging experience when catching up on key moments from the Tournament (GenAI powered draw analysis¹⁷).

Kognia is an AI powered software that can provide answers in a matter of minutes to Questions like “Show me all situations in which a defensive line was unbalanced during all of the 380 matches of last season’s premiere league.” Or “Show me all the off-the-ball movements behind the defensive line player “x” has done in the last year.”

3.1.9 MARKETING

GenAI can help automate content creation (both text and images/video), analyze large amounts of data to identify patterns in consumer behavior leading to more effective marketing strategies, and creation of personalized marketing.

¹⁶<https://www.accenture.com/content/dam/accenture/final/accenture-com/document/Accenture-A-New-Era-of-Generative-AI-for-Everyone.pdf>

¹⁷<https://www.theee.ai/2023/08/22/30536-ibm-usta-enhance-2023-us-open-digital-platforms-with-genai-commentary-ai-draw-analysis/>

Jasper, a marketing-focus version of GPT-3 can produce blogs, social media posts, web copy, sales emails, ads, and other types of customer-facing content. Jasper is used today by individuals, small businesses, and even large companies.

DALL-E 2 and other image generation tools are already being used for advertising. Heinz, for example, used an image of a ketchup bottle with a label similar to Heinz's to argue that "This is what 'ketchup' looks like to AI." Of course, it meant only that the model was trained on a relatively large number of Heinz ketchup bottle photos. Nestle used an AI-enhanced version of a Vermeer painting to help sell one of its yogurt brands.

3.2 TRANSFORMATION OF WORKFORCE

Automation has always disrupted the labor market. What makes this time much different is probably the fact that GenAI affects white-collar jobs more than the blue-collar or physical labor in factory floors and farmlands. All historical transformative periods (agricultural revolution, industrial revolution, information technology (IT) revolution) have disrupted labor. However, the GenAI revolution is indeed touching a wide cross-section of sectors including blue-collar and white-collar professionals and administrators.

However, if one studies history, this may be more of a change in types of jobs rather than there being fewer jobs. Of course, there is going to be a temporary displacement of jobs, however, the enhanced capabilities of GenAI can lead to better modes of work. The tools can be an assistant for us to become more productive, enabling us to produce higher quality work.

The invention of roller painting, the printing press, photocopy machines, autopilot mechanisms, etc., have only improved the economy and the quality of life for most of society. If automation can lead to less working hours, better quality of life, better usage of time, improved productivity and general profitability, it should be a welcome transformation. It is significant to note that such transformation and shifting of labor were observed in history many times. For example, the labor force in agriculture was reduced from 70% to less than 10%.

Technology creates more jobs than it destroys given our past experience. For example, introduction of the personal computer, for instance, has enabled the creation of 15.8 million net new jobs in the United States since 1980, even after accounting for jobs displaced¹⁸. Shifts in employment sectors are painful indeed. However, retraining and reskilling enable better and more compensating employment opportunities. The introduction of Uber as a technology platform for ride-sharing services had some major impacts in traditional taxi services. However, many Uber drivers were employed, some of whom are working on it as a second job.

¹⁸<https://www.mckinsey.com/featured-insights/future-of-work/five-lessons-from-history-on-ai-automation-and-employment>

3.3 HOW THE DISRUPTION CAN BE MANAGED

The pain of disruption can be minimized if the transition to new ways of working is gradual and controlled. There are three ways this can be accomplished:

1. First, identify areas where the need for AI-driven automation is more necessary. Examples include providing one-to-one education, remote diagnoses where experts are not available locally, translation between languages that enable better communication, providing better accessibility for tools and technology.
2. Second, keep the human in the loop in GenAI systems to the extent possible to avoid complete displacement as well as better supervision of the autonomic operation.
3. Third, retrain the workforce and discover new professions and job opportunities. We have seen the transition of the workforce even recently, when telecom companies went completely digital, and eCommerce companies employed better paid employees than local brick and mortar shops. For example, new earning opportunities evolved when Uber and other ride-sharing companies disrupted the transportation industry.

There are several examples where new technology has led to improved productivity, accuracy, and speed. A doctor could serve more patients when some of the mundane HIPAA documentation can be handled autonomously by GenAI applications.

The newly introduced service “ScribeGPT” by Amazon is a good example. Lawyers and doctors who depend on a lot of precedent cases can now get historical information much faster. Instant translation between languages reduces the need to produce documentation such as user manuals in several languages. Teachers can focus on more student interactions and instruction delivery than preparing class notes, finding examples, etc. All these are scenarios where the GenAI systems can enable workers to focus on their much-demanded professional skills rather than administrative and routine tasks.

3.4 RESPONSIBLE USE CASE SCENARIOS

Technological advances cannot be ignored, or capabilities discarded just because of the risks and fear it brings along with them. It is up to the wider community of stakeholders – technologists, businesses, political leaders, social scientists, and other professionals – to collectively think and devise ways in which the power of this technology can be channeled for better use of humanity and nature. We can easily pinpoint several good use cases where the disruption to labor can be minimized. These include using GenAI to:

- Create different lesson plans as appropriate to the comprehension level of students.
- Facilitate individualized tutoring – a dream of educationists.
- Discover and conduct experimental tests (simulation) of new drugs.
- Understand intelligent diagnoses and prognoses of disease conditions.

Responsible Generative AI

- Identify lost persons.
- Quickly detect and warn about changes in the environment through images.
- Answer common questions.
- Enable more natural multi-modal interfaces to machines, reaching to masses through more accurate language translations.
- Transform massive amount of legacy code from one programming language to another (e.g.: COBOL to Java/Python).
- Help students to have a “quick start.”

The above list while not conclusive, points to some beneficial immediate use case scenarios.

4 IMPACT ON IP/COPYRIGHTS

IP/Copyright violation is probably the second most concern triggered by GenAI. Foundation Models are trained using several sources of information assets spanning internet pages, news services, repositories of art, literature, music, audio, and video collections. Many of these are original creations of artists, writers, musicians, journalists, and many such professionals using their talent, imagination, and skills. When Foundation Models trained using these intellectual and artistic works to enable them to output “so called” newly synthesized pieces of art and literature, there arose concerns about originality, copyright violation, and royalty payments.

4.1 EXAMPLES OF CONCERN

The concerns and complaints of copyright violation are emerging and increasingly becoming visible and vocal. Image generators such as Stable Diffusion, Midjourney, or DALL·E can produce remarkable visuals in styles from aged photographs and watercolors to pencil drawings and Pointillism. Getty Images alleged that Stability AI, creators of the popular AI art tool Stable Diffusion, were misusing Getty's photos and images for training the LLMs. The comedian Sarah Silverman has joined a class-action lawsuit against OpenAI and another against Meta accusing the companies of copyright infringement¹⁹, saying they “copied and ingested” her protected work to train their artificial intelligence programs, according to court papers.

4.2 HOW CAN WE PROTECT?

Issues created by technology can be most effectively resolved by technology itself. Protection against copyright violations can be brought about in multiple ways:

- Organizations ensuring that they use data responsibly with proper consent, without violating the privacy, recording the provenance and lineage of data, and acknowledging and publishing the sources.

¹⁹ <https://www.nytimes.com/2023/07/10/arts/sarah-silverman-lawsuit-openai-meta.html>

Responsible Generative AI

- Secondly, identifying what is machine generated versus what is human generated and setting boundaries when it comes to comparisons, references, competitions, and commercial usages.
- Providing traceability and ensuring due credits to the original sources and resources the models rely on.
- Encouraging and enforcing payments of proper compensation and royalty to original creators if their data is commercially exploited.

4.3 HOW GENAI CAN BE USEFUL

Even for artists and creative writers and designers, the availability of GenAI tools can be useful for early experimentation, if properly used. Just like CAD has become an essential tool for professional engineering designers and architects, the new GenAI tools can help amplify the creativity and improve the productivity of professionals in all sectors from, art, literature, to engineering and medicine.

In fact, in a recent development, a set of artists wrote an open letter²⁰ to governments not to regulate AI, as it can outperform them. They said that “generative AI tools are empowering and expressive and it can be used not to duplicate others, but rather to make transformative new works and experiences.”

One other potential benefit would be any new entities created by machines may be out of bounds of copyright protection at least according to US Copyright Laws²¹. The U.S. Copyright Office has already declared, and courts have ruled, that copyrightable works require human authorship. Consequently, content generated solely by a GenAI system is ineligible for copyright protection.

5 IMPACT ON LEARNING, COGNITION, AND INTELLECTUAL SKILLS

From a report by IBM Institute of Business Value (IBV) entitled “Augmented Work for an Automated, AI-Driven World²²,” comes the following quote: “AI won’t replace people – but people who use AI will replace people who don’t.”

5.1 IMPACT ON LEARNING AND MEMORIZATION

In the same way that the advent of calculators did not eliminate the teaching of math at schools (addition, subtraction, multiplication, geometry, etc.), GenAI will not eliminate the need to teach

²⁰<https://creativecommons.org/about/program-areas/policy-advocacy-copyright-reform/open-letter-artists-using-generative-ai-demand-seat-at-table-from-us-congress/>

²¹ <https://www.linkedin.com/pulse/generative-ai-copyright-issues-gilbert-j-villaflor/>

²²<https://www.ibm.com/thought-leadership/institute-business-value/en-us/report/augmented-workforce>

Science, Technology and other subjects. After all, humans will need to be aware or at least have some general knowledge on the subject to properly prompt the GenAI and validate any data provided by it given the Bias and Hallucination issues.

GenAI, with its ability to provide data on demand, will shift the focus of learning from memorization of facts to the general understanding of concepts and ideas.

Using Geography as an example, it will no longer be required to memorize countries, capitals, geographical locations, etc. An overall understanding of the concepts will be enough to prompt the GenAI for answers.

5.2 IMPACT ON VALUE OF EDUCATION AND SKILLS

There is no doubt that GenAI will replace routine or repetitive tasks via automation. However, for those tasks requiring critical thinking or complex problem solving it is evident that the human factor will still be needed and skills like decision making, interpersonal skills, creativity and emotional intelligence will be more important than ever and will be used together with GenAI. This will increase the importance of soft skills like communication, interpersonal skills, emotional intelligence will be more important than ever and at par with STEM skills.

5.3 EVOLUTIONARY IMPACT ON HUMAN COGNITIVE ABILITIES

If we look back at the historical evolution of human society, the level of our cognitive skills and thinking has evolved and transformed as time progresses. Knowledge that is expressed and transferred orally got revolutionized by methods of scribbling down through notations. The printing press and copying machines further changed the way knowledge and information are stored, expressed, and conveyed or transferred.

While memorization was a big part of early knowledge transmissions (from mouth to ear), its significance was reduced with the proliferation of the ability to make multiple copies of written matter. It reduced the reliance on memory power as one could refer to written copies.

It was beneficial too – it helped disseminate knowledge to the masses easily, allowing the scholars could focus on the most essential things. In a similar fashion, the capabilities of GenAI are taking over much knowledge-intensive functions of decision making, question-answering, summarizing/interpreting of large sets of data, the need of some skills gets lower importance while it enables the experts to operate at a higher level – to be more and more accurate and productive. A doctor who probably was spending hours with a patient can now see multiple patients during the same time with an AI assistant.

6 IMPACT ON IOT SCENARIOS

6.1 BACKGROUND

The Internet of Things (IoT) can greatly benefit from the careful application of AI techniques. In IoT systems, various types of sensors measure parameters in the physical world. These readings are digitized and transported by networks to computational capabilities that may exist in implementations like Programmable Logic Controllers (PLCs), edge computers, or cloud resources. Then, the results of these computations are sent over digital networks to actuators that can influence the physical world. AI techniques come in at the compute stage, where the capabilities of AI enable very sophisticated analysis of the sensor data, and equally sophisticated calculation of the best commands to send back to the actuators.

Although Generative AI is usually associated with unstructured content like text or pictures, large-scale IoT systems can often benefit from careful application of generative AI techniques in areas like automated code generation, optimization of complex system parameters through the creation and analysis of multiple “what-if” scenarios, and the generative creation of digital twins and simulation models.

One important aspect of IoT is that the actuators allow software to control the physical world. So, if that software is inaccurate, unsecure, hallucinating or compromised in some way, those actuators can do physical damage in the real world in a way that IT systems without actuators can't – in worst-case scenarios injuring or killing people. This is a great concern in traditional IoT systems that don't use AI, but an even greater concern if we must worry about the potential for bias and hallucination of AI models that drive the actuators.

Great care must be taken at all stages of AI-based IoT deployments to ensure the model data is correct and consistent, that the model building process is accurate, the models are distributed in trustworthy ways, and that the inference engines that perform computations based upon those models are correctly connected and tested. Careful choices must be made to determine which physical systems we can entrust to Generative AI. Taken to extremes, there is concern of what is called the “Skynet Scenario”²³ where a large-scale AI system controls enough weapons to threaten humanity.

6.2 SMART SENSORS

The data from IoT sensors is given twofold use in AI-based systems. First, this data represents a behavioral record of the physical systems the sensors are monitoring in the physical world. This data (perhaps going back years) is ideal for training many classes of AI models.

²³ <https://medium.com/data-driven-fiction/the-skynet-moment-looms-with-chatgpt-5ec9714f2750>

Responsible Generative AI

Sensors are also key to the sense-compute-actuate loop once the model training is complete. Several hazards exist between the physical process and the input to the AI inference engine. Sensors can be incorrectly connected to the physical system (for example, a clogged port on a pressure sensor). Networks can be congested or hacked. AI systems could come in handy to detect and correct many erroneous situations. Generative AI could create synthetic data to supplement or replace missing or untrustworthy sensor data.

6.3 INTELLIGENT DIGITAL TWINS

Digital twins use sophisticated simulations of real-world systems to predict their operations. The simulation testing of digital twins is costly and time-consuming and GenAI holds promise to make them affordable and timely for a variety of potential applications.

A blog by IBM²⁴ says that the use of GenAI increases the power of the digital twin by simulating any number of physically possible and simultaneously reasonable object states and feeding them into the networks of the digital twin. Some interesting use cases cited are:

- LLMs helping the identification of anomalies and damages on utility assets.
- LLMs based on time series data and its co-relationship with work orders, event prediction, health scores.

7 IMPACT ON ENVIRONMENT

7.1 ENERGY REQUIREMENTS OF GENAI

Deep Machine Learning has always been very time consuming and power hungry. While most of the power consumption happens during training (building) the model, the run-time (inferencing) is also expensive especially since there may be several users running the trained model concurrently on different applications.

GPT-3 was reported to have 175 billion parameters and required 355 years of single-processor computing time and consumed 284,000 kWh of energy to train²⁵. Environmental impacts of the carbon footprint caused by LLMs are a major concern²⁶.

GenAI will have a significant energy cost to run the very large computational and networking capabilities it will require for both its model training and inference phases. As GenAI snowballs in popularity, its global energy consumption will probably greatly exceed that of cryptocurrency

²⁴<https://www.ibm.com/blog/will-generative-ai-make-the-digital-twin-promise-real-in-the-energy-and-utilities-industry/>

²⁵ <https://www.nnlabs.org/power-requirements-of-large-language-models/>

²⁶<https://medium.com/@sebastiaan.bollaart/the-environmental-cost-of-llms-a-call-for-efficiency-206cbf352c79>

mining (which is reported by Cambridge University to be 121.36 terawatt-hours (TWh) a year – about the same energy use as the entire country of Argentina²⁷). If care is not taken, this high growth in energy use could lead to significant carbon emissions, and potential energy shortages.

7.2 GENAI ENERGY ESTIMATE

Training the GPT-3 model like the one used by Open-AI's ChatGPT GenAI system is reported to require 3.14×10^{23} Floating Point Operations (FLOPs) – 314 Zetta FLOPs²⁸. This is assumed to be 32-bit floating point operations. Newer AI models, such as GPT4.0 are expected to use at least an order of magnitude more.

The next component of the calculation is to determine the FLOP capacity of a reasonable choice of processor infrastructure. Clusters of high-performance Graphics Processing Units with AI acceleration capabilities are the preferred choice for many AI compute infrastructures. One leading GPU server is the NVIDIA DGX-H100²⁹, which has eight H100 Tensor Core GPU chips, a server rating of approximately 8 Peta FLOPs/second (32-bit float), and a power consumption of 10.2KW.

Finally, the full computational capability of a GPU cluster can't be applied to the model building problem due to inefficiencies in parallel execution, memory conflicts, inter-chip network inefficiencies and many other factors. This reference³⁰ suggests 33% utilization of the available compute cycles is reasonable.

So, dividing the 314 Zetta FLOPs by 3 (because of the 33% efficiency factor) needed to compute the GPT-3 model by the 8 Peta FLOPs per second (32 bit) supplied by a DGX-H100 GPU cluster with 8 GPUs, we see about 118 million system seconds, or approximately 3.7 system years of continuous computation are needed for each learning run of the GPT3 model.

Running a DGX H100 for 3.7 years at 10.2KW requires about 334000 kWh of electric power. Adding an additional 20% for cooling and data center power overhead, the electrical consumption due to a DGX H100 is 400000kWh per model run. Electrical sources in the US average about 0.855 pounds of CO₂ emissions per kWh, so each training run would produce 171 tons of CO₂.

²⁷ <https://www.bbc.com/news/technology-56012952>

²⁸ <https://lambdalabs.com/blog/demystifying-gpt-3>

²⁹ https://resources.nvidia.com/en-us-dgx-systems/ai-enterprise-dgx?ncid=pa-srch-goog-122967#cid=dl32_pa-srch-goog_en-us

³⁰ <https://openai.com/research/ai-and-compute>

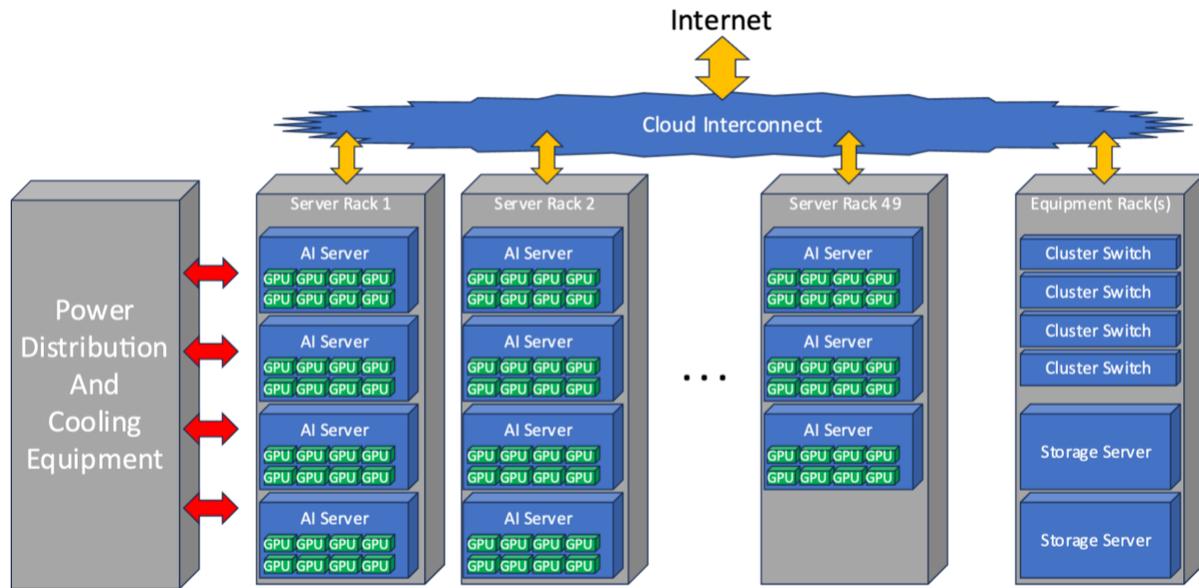


Figure 7-1: GenAI hardware infrastructure example.

It is possible to estimate the large energy investment for a single training run of a popular AI model. See below for one approximate calculation. Of course, there are many variables involved in this, including various performance and efficiency factors that are highly variable and moving targets, so actual systems are likely to be quite different in their resource requirements.

Of course, practical systems would use many (perhaps hundreds) of these DGX H100 systems in parallel. An example of one such architecture is shown in Figure 7-1. To complete a model training, run in a week (which is reasonable to keep the model current, or in multi-client environments), an infrastructure with 195 DGX systems (about 50 racks full) would be required in parallel, along with the storage, switching, power and cooling support equipment, drawing about 2.4MW of power continuously.

Running the Large Language Models (LLMs) in inference mode is also power and other resource (memory) intensive. It is reported that the compute resources for running the inference phase for large AI models account for the same and in some cases bigger than the power requirements of training the model. Imagine the resource impact of multiple instances and use cases running at the same time³¹.

7.3 HUMAN SKILLS AND TRAINING NEEDED

There is a dearth of experienced AI scientists and engineers to work in the area of GenAI and this could hamper the development of guard rails, tools for ethical and responsible frameworks for AI, as most AI engineers could get lured by other work that may result in immediate commercial

³¹ <https://doi.org/10.1787/7babf571-en>

gains. Unmanaged proliferation of AI applications may be adversely impacted by properly trained AI experts. There is a risk of improper understanding and use of the technology if hiring and employment take short-cuts.

As many commercial organizations, governmental agencies, and military are jumping on the AI bandwagon, there is a clear shortage of AI Talent. The demand for AI experts cuts across several industries, not just tech companies. Many companies have yet to adopt AI because there's a shortage of experts with the required skills in the field. Training the labor force on AI has become a focus for many consulting companies. Announcements like “Ernst & Young plans on training its 400,000 employees how to use AI in the workplace³²” and “Indian IT Giant TCS is spending \$1 Billion to train entire staff in AI” are ample evidence of the need for additional trained human power to address the surge in demand in adopting GenAI.

It is obvious that resource requirements to create reasonably large GenAI systems are very high. This means the technology could potentially be controlled by a few powerful organizations with sufficient funds and human resources, and access to the right enabling technology. The syndrome of “Power breeds power” could lead to a few technology giants monopolizing this field and probably lead to unfair influence and hazards of not having enough competition. Google’s dominance in the “search” domain and Amazon’s clout in the eCommerce industry are modern day examples. Monopoly, as in any field, will pose at least three risks – killing innovation, price manipulation, and unfair socio-economic influence if owned/managed by rogue citizens.

7.4 DATA REQUIREMENTS

Large Language Models are hungry not only for power but also for large and diverse amounts of data. Very often, the LLMs are trained on publicly available internet data which includes digitized books, news articles, blogs, social media posts, YouTube videos, and so on. The performance of these LLMs is dependent on how large and diverse the dataset is and how it can use this dataset for training.

However, human-created data has its limitations such as cost, accessibility, privacy and bias. OpenAI and many other leading companies have started making formal contracts with enterprises in news, finance, medical, transportation, and supply chain industry sectors to get access to the enterprise data to augment the public data. Also, they have turned into "synthetic" data³³ - data that is generated by algorithms/ models - to train the LLMs.

In summary, the ever-growing hunger for data poses another concern because of the resources needed to collect, store, and process massive amounts of data.

³² <https://www.businessinsider.com/ey-ernst-young-consulting-invests-ai-strategy-training-model-tools-2023-9>

³³https://www.reddit.com/r/ChatGPT/comments/153wfxk/microsoft_and_openai_test_synthetic_data_to_train/

8 NEED FOR REGULATING GENAI

8.1 WHY REGULATIONS ARE NECESSARY

Malicious use of GenAI in promoting harmful information and beliefs, enforce authoritative control, terrorist activities, military weaponization, unfair control of power leading to exploitation are big risks. Lethal autonomous weapons guided by AI technology can identify and execute targets without human intervention. Such robot driven warfare can trigger irresponsible and ruthless actions leading to quick massive destruction of human lives and property. There are reports that even military tactical planners³⁴ are launching safety practices and guidelines with AI weaponry.

There are several currently visible examples and projected future scenarios of harmful effects of GenAI. Without proper legal and ethical guidelines, the generative power of AI could play havoc in the lives of people in many ways.

8.2 CURRENT EFFORTS

Governmental agencies of many countries and professional organizations have already initiated steps to regulate the responsible usage and managed proliferation of GenAI technology.

Seven US companies met with the White House Team and agreed to self-regulate AI systems³⁵. The points of commitment included.

- Ensuring products are safe before introducing them to the public.
- Building systems that put security first.
- Earning the public's trust.

Also noteworthy is the announcement of the "Blueprint for an AI Bill of Rights"³⁶.

European Commission proposed the first EU regulatory framework for AI³⁷. AI applications that pose an "unacceptable risk" would be banned; high-risk applications in such fields as finance, the justice system, and medicine would be subject to strict oversight.

GenAI, like ChatGPT, would have to comply with transparency requirements:

- Disclosing that the content was generated by AI.

³⁴ <https://defensescoop.com/2023/08/29/pentagon-to-launch-pilot-focused-on-calibrated-trust-in-ai/>

³⁵ <https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/>

³⁶ <https://www.whitehouse.gov/ostp/ai-bill-of-rights/>

³⁷ <https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>

Responsible Generative AI

- Designing the model to prevent it from generating illegal content.
- Publishing summaries of copyrighted data used for training.

China's initial set of rules for GenAI required websites to label AI-generated content, banned the production of fake news, and required companies to register their algorithms and disclose information about training data and performance.

The UK Government proposes a new National Artificial Intelligence Strategy and for an AI Council³⁸ as an independent expert committee that advises the government. Israel's draft policy says the development and use of AI should respect "the rule of law, fundamental rights and public interests and, in particular, [maintain] human dignity and privacy." The Government of India also has stated that it plans to regulate the use of GenAI to ensure that they do not harm digital "Nagariks" (citizens)³⁹. Many other jurisdictions are also preparing their own guidelines or have already released them.

As AI-powered solutions continue to be deployed on a global basis, there will be an impetus and motivation for the guidelines across jurisdictions to have shared concepts in their regulatory compliance requirements.

8.3 INNOVATION VS. REGULATION

A strict "licensing regime" can kill innovation as may be agreed not only by scientists but regulators and users too. At the same time, a free hand to invent, propagate, and promote technology could lead to dangerous disruption to human society spreading chaos, destruction, and anarchy.

Some reservations exist on how and when to regulate GenAI. A group of the prominent Foundation Model creators are resisting any governmental regulations while volunteering to self-regulate⁴⁰.

There seems to be several reasons for the hectic activities in enforcing regulatory controls on GenAI. First, its emergence is relatively sudden as the capabilities reached a threshold with OpenAI's work on GPT models. Secondly, it directly impacted high level cognitive functions which were supposed to be the privilege of the educated, wise, and expert professionals which sounded an alarm bell. Thirdly, the power that GenAI facilitates in supremacy not only in the business world but also in inter-country political competition. The fear of this technology is triggered from the belated acknowledgement of the influence of social media platforms which was ignored in

³⁸<https://www.livemint.com/ai/artificial-intelligence/india-will-regulate-ai-to-ensure-user-protection-11686318485631.html>

³⁹<https://www.livemint.com/ai/artificial-intelligence/india-will-regulate-ai-to-ensure-user-protection-11686318485631.html>

⁴⁰ <https://finance.yahoo.com/news/top-ai-companies-visit-white-090119724.html>

its early development phase. Competition between companies and between countries can be a two-edged sword in protecting technology as well as controlling who, how, and when it can be used.

8.4 GLOBAL GOVERNING/REGULATORY ORGANIZATION

Being convinced about the potential dangers of proliferation of nuclear capabilities from the incidents in World War II, several countries agreed to form an international body - International Atomic Energy Agency (IAEA) - to monitor and regulate the use of nuclear technology. Many leaders – technocrats, administrators, social scientists – are raising the need⁴¹ for a similar global organization to regulate the research and use of GenAI given the potential scale of disruption it represents — as well as issues such as privacy, bias, and even national security.

A group of tech experts met with US Senators in September 2023 to discuss the need for regulating AI⁴² and it wouldn't be surprising if an agency like the Federal Aviation Agency (FAA) is set up in future in the AI domain too.

The “existential risks” of AI are in active discussion. The Global Summit on “Regulating AI” to be hosted by Britain⁴³ has been widely welcomed. There are panic driven actions too such as the “Pause AI” group that demands an immediate pause of development work on AI⁴⁴ and calls for dedicated regulatory authorities, public funding for safety research, and oversight and tracking of powerful systems.

9 CONCLUSION

Historically, technology has provided more benefits to humanity than damage, if properly regulated and managed. There are several examples from the real world that prove the early concerns of technology are often misplaced. The second largest killer of people are automobile accidents which doesn't prevent us from using them. More sophisticated entities like aircraft have less accidents because of stringent regulations and safety measures. Fortunately, there has been just one use of nuclear weapons since their invention.

Could runaway AI systems destroy human civilization? The notion that AI could wipe out human society may be an extreme view. It is definitely a concern that needs to be addressed and the current efforts to define and enforce guardrails for AI are indeed well placed. Regulating AI to encourage responsible uses of AI rather than restricting the research in the field is desirable.

⁴¹ <https://www.safe.ai/statement-on-ai-risk>

⁴² <https://www.npr.org/2023/09/13/1198994746/top-tech-leaders-are-to-meet-with-u-s-senators-on-the-future-of-ai-regulation>

⁴³ <https://www.theguardian.com/technology/2023/jun/07/sunak-hopes-to-bring-biden-on-board-for-ai-safety-summit>

⁴⁴ <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>

Responsible Generative AI

There are several potential uses of AI to improve our productivity, better the decision-making process, improve health care, accelerate literacy, enhance communication and so on.

Besides protecting the power of GenAI from hallucination, bias, improper usages, criminal activities, and other harmful impacts on human society and environment, it is important to promote “responsible” and “helpful” use cases. Some of those are already mentioned in the previous sections. However, it would be appropriate to give some additional focus on “desirable” applications of GenAI.

These can be in many diverse areas with the common threads of:

- Automating mundane tasks which would improve the productivity of people who can devote the time and effort for better tasks.
- Extending the human cognitive capabilities for complex tasks such as new discoveries, closer monitoring, better diagnosis, better communication, enhanced creativity, etc.
- Enabling services that are hitherto not available for people with access deficiencies.
- Enabling services which are not reachable to parts of the world because of affordability, lack of education or awareness.
- Enabling better interfaces to tools, appliances, machines, and applications.

Many use case examples can be found in fields of education, scientific research, medicine, communication, art, manufacturing, transportation, and so on. Leaving the hype and fear mongering aside, as “responsible technologists,” let’s all work together to determine how this powerful new tool can be “tamed” for the benefit of humanity in the coming years.

10 REFERENCES

- [1] Yun-Cheng Wang, “An Overview on Generative AI at scale with Edge Computing,” July 2023. <https://arxiv.org/pdf/2306.17170.pdf>
- [2] "What are AI hallucinations?, 2023. <https://www.ibm.com/topics/ai-hallucinations>
- [3] P. Park, et.al, “AI Deception: A Survey of Examples, Risks, and Solutions,” 2023. <https://arxiv.org/pdf/2308.14752.pdf>
- [4] OpenAI, "GPT-4 Technical Report", OpenAI, 2023. <https://arxiv.org/pdf/2303.08774.pdf>
- [5] M. Turpin, et. al, "Language models don't always say what they think: Unfaithful explanations in chain-of-thought prompting.", arXiv:2305.04388 [cs.CL], 2023. <https://arxiv.org/abs/2305.04388>

- [6] B. Wang, et. al, "Decoding Trust: A Comprehensive Assessment of Trustworthiness in GPT Models", arXiv:2306.11698v1 [cs.CL], 2023.
<https://arxiv.org/abs/2306.11698>
- [7] A. Sorkin, An A.I.-Generated Spoof Rattles the Markets, The New York Times, 2023.
<https://www.nytimes.com/2023/05/23/business/ai-picture-stock-market.html>
- [8] R. Bommasani and P. Liang, "Trustworthy Social Bias Measurement," arXiv:2212.11672 [cs.CL], 2022.
<https://arxiv.org/abs/2212.11672>
- [9] "Pentagon to Launch Pilot focused on Calibrated Trust in AI," 2023.
<https://defensescoop.com/2023/08/29/pentagon-to-launch-pilot-focused-on-calibrated-trust-in-ai/>
- [10] T. Shevlane, "Model evaluation for extreme risks," arXiv:2305.15324 [cs.AI], 2023.
<https://arxiv.org/abs/2305.15324>
- [11] T. Davenport and D. D'Amico, "The AI Advantage," Harvard Business Review, 2022.
<https://hbr.org/2022/11/how-generative-ai-is-changing-creative-work>
- [12] C. Qian, et. al, "Communicative Agents for Software Development," arXiv:2307.07924 [cs.SE], 2023.
<https://arxiv.org/abs/2307.07924>
- [13] E. Ben-Joseph, "The Impact Of Large Languages Models On Healthcare," 2023.
<https://www.forbes.com/sites/forbestechcouncil/2023/06/20/the-impact-of-large-languages-models-on-healthcare/?sh=1d5280326292>
- [14] Shaan Khurshid, et. al., Nature: Digital Medicine, "Cohort design and natural language processing to reduce bias in electronic health records research," April 2022.
<https://www.nature.com/articles/s41746-022-00590-0>
- [15] echo3D, "7 Generative AI Tools for 3D Asset Creation."
<https://medium.com/echo3d/7-generative-ai-tools-for-3d-asset-creation-97dd88153b7>
- [16] Accenture, "A new era of generative AI for everyone," 2023.
<https://www.accenture.com/content/dam/accenture/final/accenture-com/document/Accenture-A-New-Era-of-Generative-AI-for-Everyone.pdf>
- [17] J. Adashek, "IBM, USTA enhance 2023 US Open digital platforms with GenAI commentary, AI draw analysis," The Evolving Enterprise, 2023.
<https://www.theee.ai/2023/08/22/30536-ibm-usta-enhance-2023-us-open-digital-platforms-with-genai-commentary-ai-draw-analysis/>

- [18] Susan Lund and James Manyika, "Five lessons from history on AI, automation, and employment". McKinsey Report, November 2017.
<https://www.mckinsey.com/featured-insights/future-of-work/five-lessons-from-history-on-ai-automation-and-employment>
- [19] "Sarah Silverman Sues OpenAI and Meta Over Copyright Infringement," New York Times, July 2023.
<https://www.nytimes.com/2023/07/10/arts/sarah-silverman-lawsuit-openai-meta.html>
- [20] Creative Commons, "Open Letter: Artists Using Generative AI Demand Seat at Table from US Congress," 2023.
<https://creativecommons.org/about/program-areas/policy-advocacy-copyright-reform/open-letter-artists-using-generative-ai-demand-seat-at-table-from-us-congress/>
- [21] G. J. Villaflor, "Generative AI and Copyright Issues," 2023.
<https://www.linkedin.com/pulse/generative-ai-copyright-issues-gilbert-j-villaflor/>
- [22] IBM Institute for Business Value, "Augmented work for an automated, AI-driven world," 2023.
<https://www.ibm.com/thought-leadership/institute-business-value/en-us/report/augmented-workforce>
- [23] A. Brand, "The Skynet Moment Looms with ChatGPT," 2023.
<https://medium.com/data-driven-fiction/the-skynet-moment-looms-with-chatgpt-5ec9714f2750>
- [24] B. Gandhi, "Will generative AI make the digital twin promise real in the energy and utilities industry?" 2023.
<https://www.ibm.com/blog/will-generative-ai-make-the-digital-twin-promise-real-in-the-energy-and-utilities-industry/>
- [25] nnlabs.org, "Power Requirements to Train Modern LLMs," 2023.
<https://www.nnlabs.org/power-requirements-of-large-language-models/>
- [26] S. Bollaart, "The Environmental Cost of LLMs: A Call for Efficiency," 2023.
<https://medium.com/@sebastiaan.bollaart/the-environmental-cost-of-llms-a-call-for-efficiency-206cbf352c79>
- [27] C. Criddle, "Bitcoin consumes 'more electricity than Argentina,'" 2021.
<https://www.bbc.com/news/technology-56012952>
- [28] J. Li, "OpenAI's GPT-3 Language Model: A Technical Overview," 2020.
<https://lambdalabs.com/blog/demystifying-gpt-3>

- [29] NVIDIA, "NVIDIA DGX H100 - The gold standard for AI infrastructure," 2023.
https://resources.nvidia.com/en-us-dgx-systems/ai-enterprise-dgx?ncid=pa-srch-goog-122967#cid=dl32_pa-srch-goog_en-us
- [30] OpenAI, "AI and Compute," 2018.
<https://openai.com/research/ai-and-compute>
- [31] OECD, "Measuring the environmental impacts of artificial intelligence compute and applications: The AI footprint," OECD Digital Economy Papers, Vols. OECD Publishing, Paris, no. 341, 2022.
<https://doi.org/10.1787/7babf571-en>
- [32] A. Mok, "EY has created its own large-language model," 2023.
<https://www.businessinsider.com/ey-ernst-young-consulting-invests-ai-strategy-training-model-tools-2023-9>
- [33] Reddit, "Microsoft and OpenAI test synthetic data to train LLMs, as web data is "no longer good enough," 2023.
https://www.reddit.com/r/ChatGPT/comments/153wfxk/microsoft_and_openai_test_synthetic_data_to_train/
- [34] J. Harper, "New Center for Calibrated Trust Measurement and Evaluation initiative," DEFENSESCOOP, August 2023.
<https://defensescoop.com/2023/08/29/pentagon-to-launch-pilot-focused-on-calibrated-trust-in-ai/>
- [35] The White House, "FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI," 2023.
<https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/>
- [36] The White House, "Blueprint for an AI Bill of Rights," 2023.
<https://www.whitehouse.gov/ostp/ai-bill-of-rights/>
- [37] EU, "EU AI Act: first regulation on artificial intelligence," European Parliament News, 2023.
<https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>
- [38] N. Potter, "Rule Britannia: Can the UK Set Global AI Standards.," IEEE Spectrum, 2021.
<https://spectrum.ieee.org/rule-britannia-can-the-uk-set-global-ai-standards>

- [39] G. Aulakh, "India will regulate AI to ensure user protection", June 2023.
<https://www.livemint.com/ai/artificial-intelligence/india-will-regulate-ai-to-ensure-user-protection-11686318485631.html>
- [40] D. Coldewary, "Top AI companies visit the White House to make 'voluntary' safety commitments," 2023.
<https://finance.yahoo.com/news/top-ai-companies-visit-white-090119724.html>
- [41] Center for AI Safety, "Statement on AI Risk," 2023.
<https://www.safe.ai/statement-on-ai-risk>
- [42] C. Grisales, "The who's who of the tech world meet with senators to debate plan to regulate AI," 2023.
<https://www.npr.org/2023/09/13/1198994746/top-tech-leaders-are-to-meet-with-u-s-senators-on-the-future-of-ai-regulation>
- [43] P. Walker, "UK not too small to be centre of AI regulation, says Rishi Sunak," 2023.
<https://www.theguardian.com/technology/2023/jun/07/sunak-hopes-to-bring-biden-on-board-for-ai-safety-summit>
- [44] Pause AI, "Pause Giant AI Experiments: An Open Letter," 2023.
<https://futureoflife.org/open-letter/pause-giant-ai-experiments/>

11 ACKNOWLEDGEMENTS

The views expressed in the *IIC Journal of Innovation* are the author's views and do not necessarily represent the views of their respective employers nor those of the Industry IoT Consortium®.

© 2024 The Industry IoT Consortium logo is a registered trademark of Object Management Group®. Other logos, products and company names referenced in this publication are property of their respective companies.

- Return to *IIC Journal of Innovation landing page* for more articles and past editions.