



Center for AI and
Digital Policy

UPDATING THE UNIVERSAL GUIDELINES ON AI IN A WORLD OF GENERATIVE AI

SEPTEMBER 2023



THE SPRING GROUP

photo generated and extended with OpenAI's DALL-E 3 using the prompt "a photo of a robotic arm shaking hands with a human overlayed over a vibrant exploding nebula and black background, futuristic, digital art"

UPDATING THE UNIVERSAL GUIDELINES ON AI IN THE WORLD OF GENERATIVE AI

The SPRING Group

September 2023

Prepared for:

The Center for AI and Digital Policy

Authors:

Arik Karim (Co-Chair)*

Brian Zhou (Co-Chair)*

Gia Mendonca

Natalie Zhang

Kaveen Shah

Hanming Sun

Aileen Wu

* These authors contributed equally.

Correspondence:

arik@thespringgroup.org

brian@thespringgroup.org

1. Table of Contents

UPDATING THE UNIVERSAL GUIDELINES ON AI IN THE WORLD OF GENERATIVE AI	2
1. Table of Contents	3
2. Executive Summary	4
3. Introduction	5
3.1 AI's Existential Risks	5
3.2 Purpose of the Universal Guidelines	5
3.3 AI Adoption & Investment Trends	6
4. Analysis	6
4.1 Overview of the Universal Guidelines	6
4.2 Environmental Applications of the UGAI	7
4.3 Educational Applications of the UGAI	7
4.4 Protecting Copyright & Data Privacy	8
4.5 Increasing Cultural & Ethical Competence	8
5. Recommendations	8
5.1 International Cooperation	8
5.2 Streamlining of Intergovernmental AI Standards	9
5.3 Educational Initiatives	9
6. Conclusion	9

2. Executive Summary

The Universal Guidelines for AI (UGAI), authored in 2018, was hailed at the time as a needed push for the recognition of rights of all individuals in a world of advancing and continuous AI. In the five years since, many of the principles in the UGAI have been adopted into national law and global forums and frameworks regarding AI governance. Principles of the UGAI have been implemented in the European Union's landmark AI Act, as well as the White House's blueprint for an AI Bill of Rights, which is expected to be released as an executive order in winter 2023.

However, various fields of AI have jumped leaps and bounds in development and public recognition. Large Language Models, developed by companies for decades, suddenly emerged into the public eye in November 2022 when OpenAI launched a front-facing platform for ChatGPT. Four years ago, on Gartner's hype cycle, Generative AI did not exist. The larger field of Generative AI, which had been developing for decades, has taken over public discourse in the news.

Generative AI is disruptive. Disputes, such as the ongoing debate over copyrighted data used in training models, is being resolved from the courtrooms to the negotiating table of the Writers Guild of America's historic strike. **These disputes open a new front of questions that the UGAI does not answer:**

1. What rights do individuals have over the climate, and how do they interact with the environmental costs of training large and complex models?
2. What rights do the youngest generation have to learning and education?
3. What rights does any individual have to the content that they own and release on the internet?
4. What rights do individuals have to protect them against encoded discrimination in AI systems and models?

In this brief, we call for the addition of new clauses that clarify any individual's right to environmentally sustainable AI, educational use, algorithmic transparency, and right to their authorship. We further advocate for increased global collaboration towards ethical AI development and governance as new fields of AI continue to emerge.

3. Introduction

3.1 AI's Existential Risks

Artificial Intelligence (AI) has catalyzed a wave of technological change. While AI has been in digital services and products for years, the past several years have seen the rise of public interest and awareness in front-facing AI interfaces with humans to answer questions and generate media. Most shocking is the sudden rise of generative AI (GenAI), which emerged overnight in November 2022. GenAI, not even recognized as a field in 2019 (Fig. 1), has risen in a year to offer the public previously-unthinkable levels of advanced information generation and search to complex modeling and data analysis. However, various broad aspects of AI, including its potential applications to mass disinformation, surveillance, and autonomous weapons, pose unique dangers to society due to people's increasing reliance on artificial intelligence.¹

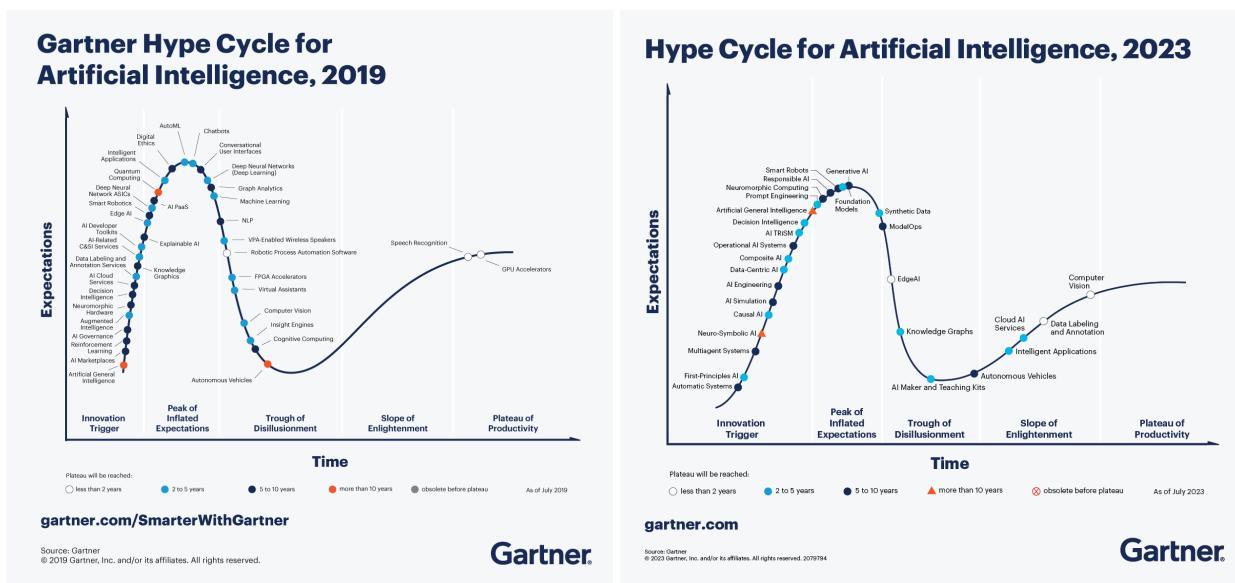


Figure 1. Gartner's Hype Cycles for AI in 2019 and 2023²

The unchecked and unregulated development of AI introduces a host of problems like data privacy concerns, worker displacement, harmful use cases (i.e. profiling), and human-AI goal misalignment. The first draft of the UGAI emerged in 2018, authored by the Electronic Privacy Information Center and supported by hundreds of worldwide

¹ [Marr, 2023](#)

² [Gartner, 2019; Gartner, 2023](#)

signatories (i.e. educational institutions, law firms, and national justice agencies)³ and AI-oriented organizations such as the Center for AI and Digital Policy.⁴

The UGAI functions as a quasi-Bill-of-Rights, outlining the protections society is owed considering the rapid rise of GenAI in particular. These protections are centered around 12 fundamental rights that people have with AI tools, ranging from the right to transparency to termination obligations.

3.2 Purpose of the Universal Guidelines

The purpose of the UGAI is simple: challenge the greatest problems AI poses by establishing a framework of principles that societal institutions should adhere to when developing and deploying AI technology.

Since the proposal of the Guidelines in 2018, there has been a surge in investment for GenAI, alongside a host of developments across various domains. Since 2020, GenAI experienced a 425% increase in investment (reaching \$2.1B).⁵ Industry has led the way in producing new machine learning models, developing 32 industry-produced models in 2022, from Google, Meta, IndicNLG, OpenAI, GitHub, and NVIDIA.⁶ As 2023, which is widely considered to be the breakout year of GenAI, comes to an end, it is crucial to re-evaluate the current guidelines to accommodate recent advancements.⁷

3.3 AI Adoption & Investment Trends

Since 2017, adoption has more than doubled, with 50-60% of companies having deployed AI.⁸ This surge in popularity is fueled by recent demonstrations of capabilities in areas like GenAI. These systems, such as advanced chatbots, virtual assistants, and language translation tools, can create text, images, audio, video, and other content based on prompts from a user. This could be applied to fields including education, government, law, and entertainment. As of 2023, some nascent GenAI platforms have reached over 100 million users.⁹

GenAI in particular has the potential to enhance labor productivity and boost economic output as it is adopted across industries. Global private investments in AI are expected to

³ [Center for AI and Digital Policy, 2023](#)

⁴ [Electronic Privacy Information Center, 2018](#)

⁵ [White & Case Law Firm, 2023.](#)

⁶ [Clark & Perrault, 2023](#)

⁷ [McKinsey & Company, 2023](#)

⁸ [McKinsey & Company, 2022](#)

⁹ [Governmental Accountability Office, 2023](#)

increase dramatically over the next three years, growing from an estimated \$91.9 billion in 2022 to \$158.4 billion in 2025, a 72% increase.¹⁰ Thus, the value of AI could potentially reach \$15 trillion.¹¹

Taken together, this paper will discuss the significance of the UGAI while offering analysis and suggestions to strengthen it. Lastly, we will recommend new solutions to challenges the UGAI still faces.

4. Analysis

4.1 Overview of the Universal Guidelines

The UGAI from 2018 appropriately addressed the uses and implications of AI for its time: security and accuracy of data; eliminating undue bias; and the preservation of privacy, dignity, and awareness are emphasized to maintain the integrity of not only software systems, but also the users and creators thereof. The guidelines lay a strong foundation for the development of GenAI systems used in professional and government situations, but fail to address circumstances that are applicable in daily life considering AI's prominence in today's educational, professional, and domestic atmospheres.

4.2 Environmental Applications of the UGAI

The training and maintenance of AI (testing, server fields, etc.) creates a large carbon footprint and can produce over 626,000 pounds of carbon dioxide annually.¹²

The rapid advancement and widespread deployment of AI technologies can exert considerable strain on resources, particularly through energy consumption and electronic waste generation. Training sophisticated AI models, especially large neural networks and GenAI, demands significant computational power, often necessitating the use of data centers with powerful processors, which require substantial electricity for computation and cooling. Hence, GenAI expends water supply and contributes to the sector's immense carbon footprint. It remains essential that regulations emphasizing viability and sustainability are implemented on the energy sources of AI.

4.3 Educational Applications of the UGAI

Rising popularity for GenAI has severe impacts on students' learning processes. Given GenAI's capability to process questions and produce responses, students are prone to

¹⁰ [Revell, 2023](#)

¹¹ [McKinsey & Company, 2023](#)

¹² [Earth.org, 2023](#).

using it to cheat and compromise academic integrity,¹³ overhauling educational principles. Therefore, we suggest including a guideline addressing AI's ethical application in the educational sphere. Educational initiatives that outline how AI can be used ethically and productively are also necessary for the education sector to build trust in an increasingly AI-ubiquitous society.

4.4 Protecting Copyright & Data Privacy

For the training and dataset development of GenAI systems, companies like Midjourney and Stability AI have used images, videos, writing, and other media without the consent of its owners.¹⁴ This results in the production of GenAI content that closely resembles the work of humans. We suggest articulating that individual owners have a right to opt-out¹⁵ of their intellectual property being used in training GenAI in accordance with ethical principles governing copyright, therefore indemnifying GenAI companies from infringement accusations.

4.5 Increasing Cultural & Ethical Competence

Implementing a guideline ensuring cultural and ethical sensitivity in AI programming is paramount in fostering a technological landscape that reflects the diversity of human societies. The programmers and testers of GenAI are sources of bias, reflecting their own conscious or unconscious beliefs, cultural backgrounds, or personal experiences in the code they write and training they provide. These biases manifest in AI systems and applications, leading to unfair outcomes and reinforcing societal inequities and profiling. Unfortunately, this is already evident in studies¹⁶ wherein AI shows narrow racial subsets in certain occupations,¹⁷ such as the prevalent depiction of marginalized groups in jobs like "janitorial services" and "housekeeping" in generated media. In order to address this problem, the following steps are necessary:

1. We recommend guidelines encouraging third-party, open-source reviews of AI programs. By increasing the number of project contributors, the end result ends up encompassing more cultural viewpoints, substantially diluting previously imminent bias.
2. We recommend suggesting the involvement of ethicists, social scientists, and community advisors to establish proper accountability mechanisms and prevent future bias.

¹³ [Ancell, 2023](#)

¹⁴ [Wiggers, 2023](#)

¹⁵ [Zhou & Motati, 2023](#)

¹⁶ [Small, 2023](#)

¹⁷ [Nicoletti & Bass, 2023](#)

- Finally, the most crucial factor in decreasing unequal outcomes is the prevalence of diverse perspectives in data. While data quality is essential, it remains imperative that data comes from a variety of sources to account for as many scenarios as possible.

5. Recommendations

5.1 International Cooperation

In order for the UGAI to become the standard for all countries using AI,¹⁸ it needs to be expanded to all countries. For this, there needs to be more attention from political figures and governments to enact such issues.

In the United States, for example, a blueprint for an AI Bill of Rights has been created, focusing on the values of “Safe and effective systems, Algorithmic Discrimination Protections, Data Privacy, Notice and explanation, and Human Alternatives, Consideration, and Fallback.”¹⁹ A similar legislation is being introduced in Europe with the EU AI Act, becoming the world’s first legal framework for AI.²⁰ Building on these crucial baby steps, more countries must adopt the policy to improve the efficacy of the UGAI.

5.2 Streamlining of Intergovernmental AI Standards

Currently, there are a multitude of potential agreements from a macro standpoint between separate countries as well as a corporate perspective. Such divisions would make all guidelines confusing at best and exploitative at worst—or example, within the UN, 193 countries²¹ adopted a global agreement on the Ethics of AI while the OECD (Organisation for Economic Co-operation and Development), created its own AI Principles²² centered around promoting democratic values. Indeed, even for companies, Google²³ and SAP 2018 released their own separate AI guidelines.

Thus, the promotion of cooperation between the UGAI and other global AI agreements to create a standardized regulatory system will enable corporations to be held accountable for the development and implementation of AI.

¹⁸ [Whittaker, 2018](#)

¹⁹ [White House Office of Science & Technology Policy, 2022](#)

²⁰ [Lynch, 2023](#)

²¹ [United Nations, 2021](#)

²² [Organization for Economic Cooperation & Development, n.d.](#)

²³ [Jobin et al., 2019](#)

5.3 Educational Initiatives

While the guidelines state that AI system developers are primarily responsible for transparency of techniques that go into AI output,²⁴ it fails to outline possibilities for these organizations to do so. This can be addressed by adding avenues for the audit and exhibition of information on algorithms via public domain papers, sites, or programs.

6. Conclusion

Considering the prominence of GenAI as a supplement to human intelligence in professional and educational fields, it is imperative that a comprehensive and panoptic ethical standard is established as a baseline for the future development of GenAI. Building on the protocol outlined in the 2018 UGAI, we suggest the following:

1. The inclusion of an environmental sustainability clause;
2. Educational use policies;
3. Avenues for algorithm transparency;
4. Perspective analysis in AI testing; and
5. Global collaboration to create a streamlined approach to ethical AI development that can be applied to respects such as GenAI.

Ultimately, GenAI should be created and utilized with the goal of beneficence in mind: it should serve as a tool to advance human development, efficiency, and achievement to create a more knowledgeable and capable race. Under a revamped series of guidelines that promote equity, informed and unbiased data, diversity in AI learning, and environmental sustainability in addition to the already-existing assemblage from 2018, it is possible to create more current guidelines for Generative Artificial Intelligence in a constantly evolving world.

²⁴ [Bostrom, 2011](#)