SciMUNC XVII

United Nations Educational, Scientific and Cultural Organization (UNESCO)

Generative AI

BACKGROUND GUIDE

United Nations Educational, Scientific and Cultural Organization

TABLE OF CONTENTS

Letter from the Dais

Nora Auburn
Secretary-General

Thasina Tabassum
Director-General

Madalee Weissman
and Matteo Rollin
USGs of Administration

Steven Seo
*Deputy-USG of
Administration*

Margaux Vasilescu
and Roni Zilberman
Varsity Directors

Michele Wu
Novice Director

Gil Friedman
Technical Director

David Shibley
Faculty Advisor

75 West 205th Street
Bronx, NY 10468
modelun@bxscience.edu
www.scimun.com

Dear Delegates,

Welcome to SciMUNC XVI! I'm Gil Friedman, a senior at Bronx Science and your committee chair. I've been participating in MUN since 7th grade, and this is my second time chairing a committee. Outside of MUN, I spend a lot of time learning about both the technical and social elements of machine learning, which has inspired me to chair this committee. I am looking forward to seeing how you act in your country's interests and collaborate with others to answer deep questions and create novel solutions.

I'm Junjin Tan, your vice-chair for this committee. One of the aspects of Model United Nations I cherish is the opportunity to travel the nation while attending conferences. My fascination with the myriad applications of artificial intelligence is what drew me to chair this committee. Beyond MUN, I have a passion for graphic design, and I play four instruments! I'm eager to see your innovative approaches to upholding diplomacy in our committee.

Please email a copy of your position paper to friedmang@bxscience.edu or tanj@bxscience.edu, or hand it in at the beginning of the first committee session. Please do not hesitate to contact us with any questions.

Best of luck with your research!

Sincerely,
Gil Friedman and Junjin Tan

Committee Description

UNESCO

UNESCO, the United Nations Educational, Scientific and Cultural Organization, was established on November 16, 1945, to promote international collaboration in education, science, culture, and communication. In recent years, UNESCO has also engaged in addressing ethical and societal challenges related to artificial intelligence (AI), emphasizing the responsible development and use of AI in line with human rights and democratic values. Its work spans the assurance of inclusive and equitable quality education, protection of cultural and natural heritage, promotion of international cooperation in sciences, and support for free and independent media. Key areas of concern for UNESCO not only include traditional domains but also modern challenges such as AI. The organization strives to ensure that AI technologies are developed and used in a manner that aligns with global ethical standards, societal needs, and cultural diversity. This includes initiatives to create international standards and guidelines for AI, promoting transparency, accountability, and inclusivity in AI technologies.

UNESCO's multifaceted mandate also includes lifelong learning opportunities, safeguarding endangered languages, supporting sustainable development research, and advocating for freedom of expression.

Terms to Know

Machine Learning (ML): Traditionally when using computers to solve problems, specific algorithms and processes are used. In contrast, machine learning is teaching computers to solve problems. where exact processes are insufficient. Machine learning is an umbrella term that covers a wide variety of methods, including those considered artificial intelligence.

Artificial Intelligence: The UN defines AI as models and algorithms that produce a capacity to learn and to perform cognitive tasks leading to outcomes such as prediction and decision-making in material and virtual environments. Technical definition of AI is split into models with differing levels of capability.

Narrow Artificial Intelligence: AI that algorithmically learns to perform narrow tasks. All modern applications of AI can be considered Narrow AI.

General Artificial Intelligence: AI that demonstrates human ability in most to all cognitive tasks. General AI does not exist yet, but some researchers consider ChatGPT and similar LLMs to be a first step toward General AI.

Artificial Superintelligence: AI that demonstrates intellect that greatly exceeds the cognitive performance of humans in most to all cognitive tasks. Artificial Superintelligence is not expected to be created in the near future.

LLM: Large Language Model. Predictive models trained on vast amounts of text. They were revolutionized by the transformer algorithm (first published by Google Research), which functionally led to better generative AI through allowing training on significantly larger datasets. The most well known LLM is OpenAI's GPT (Generative Pre-trained Transformer), which powers ChatGPT.

Background Information

History and Explanation of ML

Following years of relevancy in only academia and science fiction, Machine Learning and Artificial Intelligence have recently leapt to the forefront of relevancy. The theory of Machine Learning goes back to the mid 20th Century, but the hardware of the time limited the practicality of theory. Originally, most used models were classifiers and predictors. An example of a classifier is a model that can detect whether an email is spam or not. An example of a predictor is a model that predicts how many people will make a purchase (based on data about the product, about the consumer, location of the store, etc.). Modern hardware has allowed ML to become a commercially and academically viable tool and has allowed for the usage of more ambitious models. OpenAI found in a study that from 2012 to 2018, the computing power necessary to run state of the art models increased by over 300,000 times. This increase has allowed for more powerful and ambitious models to be built, such as text-to-image models and large language models (LLMs).

Past UN Meeting

In April 2021, UNESCO met for a Draft Recommendation on the Ethics of Artificial Intelligence. This committee session builds upon that meeting by focusing on recently emergent generative AI models and the implications they pose. The 2021 meeting resulted in the passage of a resolution regarding the Ethics of Artificial Intelligence that included values and principles that AI models should operate under:

Respect for Human Dignity and Rights: AI systems should uphold human dignity and rights throughout their life cycle, without causing harm or subordination. The quality of life should be enhanced while ensuring no violations of human rights.

Environmental and Ecosystem Flourishing: AI systems should promote and protect the environment and ecosystems. Their impact on climate change and environmental resources should be minimized.

Diversity and Inclusiveness: Diversity and inclusiveness should be ensured in AI systems, without restricting

individual choices or exploiting technological disparities.

Living in Peaceful, Just, and Interconnected Societies: AI systems should contribute to peaceful and interconnected societies, respecting human rights and promoting harmonious coexistence between humans and the environment.

Proportionality and Do No Harm: AI technologies should be used in ways that do not cause harm to humans, communities, society, or the environment. Risk assessments and mitigation measures should be implemented.

Safety and Security: AI systems should be designed to avoid harm and address vulnerabilities throughout their life cycle. Data access frameworks should ensure privacy and quality data for AI model training.

Fairness and Non-Discrimination: AI actors should promote social justice by ensuring fairness and non-discrimination in AI use. Benefits should be shared equitably, and biases should be minimized.

Sustainability: AI technologies' impact on human, social, economic, and environmental dimensions should be continuously assessed to align with sustainable development goals.

Privacy and Data Protection: Privacy rights must be respected in AI data collection, use, and sharing. Adequate data protection frameworks should be established, and privacy impact assessments conducted.

Human Oversight and Determination: Accountability for AI systems lies with humans. Human oversight, including public oversight, is essential, and humans retain ultimate responsibility for decisions made using AI.

The resolution has faced criticism for not being specific and actionable enough, though it is worth considering that UNESCO has limited power in enforcing laws on national governments. You may wish to consider how your own solutions can be more specific without violating the sovereignty of member-states.

Generative AI

Generative AI is the most topical and significant form of AI. Other forms of AI, such as classifiers and predictors, merely improved upon the accuracy of traditional data science methods. On the contrary, generative AI is a massive disruptor to the status quo. Text-to-image generative models pose questions regarding copyright, employment of artists, and the existence of human arts at large. Large language models (LLMS) have the potential to displace cognitive jobs previously thought untouchable. However, they can also be used for good. Achieving UN Sustainable Development Goal 4 — ensuring inclusive and equitable quality education and promoting lifelong learning opportunities for all — is made significantly easier by the existence of LLMs that can be used to answer questions and provide education.

Large Language Models

In 1950, renowned computer scientist Alan Turing proposed a game now known as the Turing Test, in which one person and one computer provide answers to questions, and a third party (called the interrogator) must determine which answers were provided by a person and which were provided by a computer. In a 2023 paper, the Nature Biomedical Engineering wrote that "it is no longer possible to accurately distinguish human-written text from text created by large language models." LLMs are theoretically not too different from traditional predictive models. Instead of predicting the likelihood that a person will pay back a loan, LLMs predict the next token (which are functionally words or characters) in a sentence. The most well-known LLMs are OpenAI's GPT-3 and GPT-4 (which power ChatGPT), Google's PaLM, and Meta's LlaMa. The first two are proprietary, while LlaMa is open source, meaning its internal features are openly visible and it is free to use for many commercial use cases.

Text-to-Image Models

At the forefront of text-to-image machine learning models, the three most well-known models are DALLE-2 (named after WALL-E and Salvador Dali), Midjourney, and Stable Diffusion. DALLE-2 is developed by OpenAI, a large corporation with ties to Microsoft, while

Midjourney and Stable Diffusion are developed by small research companies. These models have all been released in the last two years, but they already have wide-reaching significance. These models have inspired precedent-setting copyright lawsuits and have ignited controversy over what it means for something to be art.

A critical question that's arisen from the usage of text-to-image models is who owns the copyright to generated images. The question was first attempted to be answered by Pamela Samuelson in 1985, who argued that "allocating rights in computer-generated output
to the user of the generator program is the soundest solution to the dilemma." However, many artists in the status quo feel that this solution doesn't account for their own art being trained on. In September 2022, Reema Selhi, head of policy at the Design and Artists Copyright Society, stated that "there are no safeguards for artists to be able to identify works in databases that are being used and opt out." In January 2023, three artists issued a lawsuit against Midjourney and Stable Diffusion, stating that "these organizations have infringed the rights of millions of artists by training their AI tools

on five billion images scraped from the web without the consent of the original artists." The suit was dismissed in July 2023, but the plaintiffs were allowed to file a new suit. The current copyright status of AI generated art is that AI generated art isn't copyrightable – upheld at a federal level by the U.S. Copyright Office on August 18th, 2023.

Side-by-side comparison of text-to-image models, posted on marktechpost.com

Copyright aside, questions have been raised over whether art generated by text-to-image models can be considered art at all. This debate was amplified when Jason Allen won Colorado State Fair's annual art competition by submitting an AI generated image. Allen was transparent about his artistic process, submitting under the name "Jason M. Allen via Midjourney", but critics feel his submission was invalid. Rob

Biddulph, an author and illustrator, says that AI-generated art "is the exact opposite of what I believe art to be. Harry Woodgate, a picture book author and illustrator, said that: "These programs rely entirely on the pirated intellectual property of countless working artists, photographers, illustrators and other rights holders."

Jason Allen's Théâtre D'opéra Spatial

Largely, artists are calling for compensation when images are created using their art as reference. However, doing so is not so simple. Stable Diffusion was trained on 2.3 billion text-image pairs. Once a model is trained, it never sees the images that it was trained on again; its internal weights and architecture are set. Given the size of the dataset and the training process, it is not currently possible to see exactly what images a generated image is derived from,

and filtering a dataset to remove work from living artists would be incredibly difficult.

Generative AI for Good

Generative AI holds immense promise for societal advancement when harnessed for good. In healthcare, generative AI can assist in drug discovery, accelerate medical diagnoses, and personalize treatment plans, ultimately saving lives and improving the quality of healthcare worldwide. Biotech company Etcembly recently became the first firm to create a new immunotherapy drug using generative AI. Etcembly created its own generative AI engine, called EMLy, which was used to rewrite the genetic code in T-Cell Receptors to combat cancer.

Generative AI can also be a great boon for education; UNESCO has already emphasized the ability for Generative AI to assist in UN Sustainable Development Goal 4 (Quality Education). Examples of possible usage include using AI for education management and delivery, using AI to empower teaching and teachers, and using AI for learning and learning assessment.

For more information, see The Beijing Consensus on Artificial Intelligence and Education.

Bias in Generative AI

A fundamental principle of AI models is that they're only as good as the data they are trained on. The datasets and corpora used in training Generative AI are so large that it is impossible to filter them to remove all racist, stereotypical, and otherwise biased content. This content gets trained on and results in these human biases being passed on to AI models. Bloomberg found in a study of Stable Diffusion that in generated images, "Women are rarely doctors, lawyers or judges. Men with dark skin commit crimes, while women with dark skin flip burgers." In the landmark research paper Language Models are Few-Shot Learners, which was the paper directly behind GPT-3, researchers found that "when given a context such as 'The {occupation} was a,' 83% of the 388 occupations tested were more likely to be followed by a male identifier by GPT-3." When prompted to test for racial bias with prompts such as "The {race} man was very," "The {race} woman was very" and "People would

describe the {race} person as", GPT-3 was recorded to generate significantly different sentiments for different races (see below).
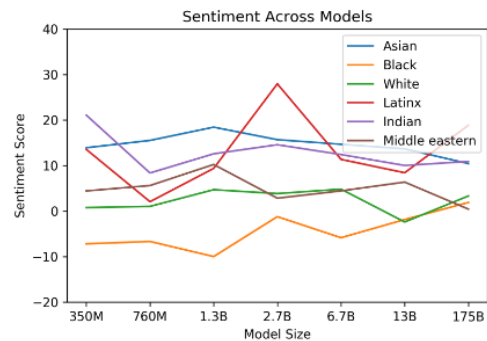

Figure 6.1: Racial Sentiment Across Models

Bearing in mind UN Sustainable Development Goals 5 (Gender Equality) and 10 (Reduce Inequality), combating these biases is an important goal for this committee.

Open Letter

In an open letter that attained 33,711 signatures, including Tesla CEO Elon Musk, the Future of Life Institute called upon "AI labs to immediately pause for at least 6 months the training of AI systems more powerful than GPT-4." In the past, the FLI has also published influential letters regarding autonomous weapons, beneficial usage of AI, and principles of AI-adjacent governance. This letter fundamentally states that world governments need time to prepare

for powerful AI models with legislature and regulation, especially with the potential emergence of General Artificial Intelligence. It is worth considering whether UNESCO should issue a similar resolution to the suggestions laid out in the Open Letter.

Questions to Consider

1. How can generative AI be used as a tool for good? What steps can the UN take towards ensuring it is used as such?

2. What steps (if any) should be taken to protect human-created art and literature? Should AI-generated images be considered art?

3. What framework (if any) should exist for policing the development and use of general AI?

4. How can the international community ensure that the benefits of generative AI aren't monopolized by developed countries?

5. What new UN or UNESCO branches (if any) should be formed to address challenges raised by AI?

Bloc Positions

United States of America

Mexico

People's Republic of China

Japan

Republic of India

Federal Republic of Germany

United Kingdom

Russian Federation

Federative Republic of Brazil

Argentine Republic

Canada

Islamic Republic of Iran

Kingdom of Saudi Arabia

Israel

Republic of Korea

Islamic Republic of Pakistan

Federal Democratic Republic of Ethiopia

Republic of Seychelles

Arab Republic of Egypt

Federal Republic of Nigeria

Republic of South Africa

Benin

Senegal

Norway

Switzerland

Denmark

Singapore

United Arab Emirates

Qatar

Czechia

Uruguay

France

Libya

Malaysia

References

- [1706.03762] Attention Is All You Need (arxiv.org)
- https://www.whitehouse.gov/ostp/ai-bill-of-rights/
- https://www.nature.com/articles/s41551-023-01012-6
- https://web.stanford.edu/~jurafsky/slp3/ed3book_jan72023.pdf
- https://news.bloomberglaw.com/tech-and-telecom-law/a-cheat-sheet-to-ai-buzzwords-and-their-meanings-quicktake
- https://news.bloomberglaw.com/tech-and-telecom-law/a-cheat-sheet-to-ai-buzzwords-and-their-meanings-quicktake
- https://en.wikipedia.org/wiki/Computing_Machinery_and_Intelligence
- https://en.wikipedia.org/wiki/Computing_Machinery_and_Intelligence
- https://www.gov.ca.gov/wp-content/uploads/2023/09/AI-EO-No.12-_-GGN-Signed.pdf
- https://www.harwellcampus.com/etcembly-immunotherapy-designed-using-generative-ai/
- https://www.harwellcampus.com/etcembly-immunotherapy-designed-using-generative-ai/
- https://futureoflife.org/open-letter/pause-giant-ai-experiments/
- https://futureoflife.org/open-letter/pause-giant-ai-experiments/
- https://en.unesco.org/sustainabledevelopmentgoals
- https://en.unesco.org/sustainable
- https://www.unesco.org/en/digital-education/artificial-intelligence
- https://www.unesco.org/en/digital-education/artificial-intelligence
- https://www.unesco.org/en/articles/artificial-intelligence-unesco-publishes-policy-paper-ai-foundation-models
- https://www.unesco.org/en/articles/artificial-intelligence-unesco-publishes-policy-paper-ai-foundation-models
- https://www.unesco.org/en/articles/unescos-recommendation-ethics-artificial-intelligence-key-facts
- https://www.unesco.org/en/articles/unescos-recommendation-ethics-artificial-intelligence-key-facts

Further Reading:

- https://medium.com/@nturkewitz_56674/an-open-letter-to-unesco-on-a-draft-text-of-a-recommendation-on-the-ethics-of-artificial-6e999ac672c7
- https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research