FORUM:	Commission on Science and Technology for Development
ISSUE:	Measures to Develop strategies against the Spread of Misinformation by Artificial Intelligence
STUDENT OFFICER:	Seung Woo Ham
POSITION:	President of Commission on Science and Technology for Development

Introduction

In an age defined by an unprecedented surge in technological advancements of all kind, Artificial Intelligence (AI) has emerged as a powerful tool, reshaping various aspects of our lives. The world still remembers the day back in 2016 when Google's AlphaGo developed by Deepmind defeated the world champion Go player Lee Sedol, when Tesla introduced the first ever non-excludable fully autonomous self-driving car, and when OpenAI introduced ChatGPT half a year ago. However, alongside its numerous benefits, AI also brings forth an alarming challenge—the rapid spread of misinformation. Misinformation, fueled by AI-driven algorithms and automated processes, poses a significant threat to society, eroding trust, distorting realities, and sowing discord among individuals. As we navigate the complex landscape of information dissemination, it is imperative to develop effective strategies to combat the proliferation of falsehoods by leveraging the very technologies that enable their propagation.

This essay examines the problem of AI-driven disinformation and its effects on society, emphasizing the demand for all-encompassing solutions to counteract it. It analyzes advances in machine learning and natural language processing to look at how AI and false information interact. The article also looks at how AI-driven disinformation affects society, emphasizing how vital it is to protect the integrity of the information ecosystem. Technology limits, ethical issues, and the constant cat-and-mouse game between disseminating and refuting false information are significant obstacles. The essay makes a number of recommendations, such as technology advancements, media literacy instruction, critical thinking, and ethical AI development. Commission on Science and Technology for Development, a committee under the United Nations to promote the development and use of science, technology, and innovation (STI) to support sustainable development
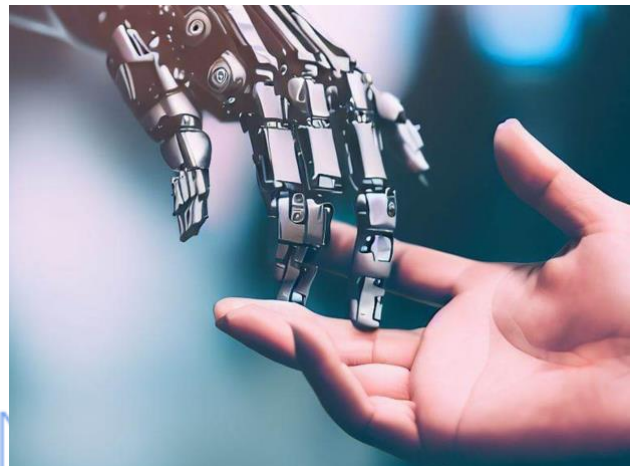

Figure 1. AI and Human

efforts worldwide, is undertaking a task to promptly build frameworks to restrain the spread of AI driven falsehoods.

Background

When the artificial intelligence was first introduced by John McCarthy in 1950s, it wasn't as popular or even productive for commercial use. It was a mere concept that prospected to take a prolonged period to be developed. Through 1970s and 80s, movements to develop the model started, and as inferred, it wasn't just one person, but they were several professionals who thrusted themselves into this AI game, believing a rosy future. Entering the 90s, AI industry was ever flourishing by all means, now recalled as the time of "AI Boom". XCON, the first expert system came into the commercial market, was introduced;

first driverless car that could accelerate up to 50 mph was created; the first strategy managerial advisory system so called the Alacrity was commercially launched; and so forth. Today, the world still remembers the day Deep blue beat the world's chess champion Gary Kasparov, when NASA succeeded to land two man-less rovers onto Mars, and when Steve Jobs introduced the first iPhone. But just like there's no sun that never sets, things started to get irrational.
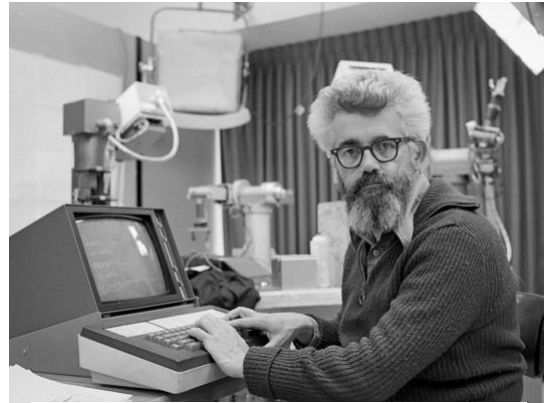

Figure 2. John McCarthy first coiled the term "AI"

Scientists and programmers did not forecast these man-made intelligence models could prepare immense measures of information, including both dependable and untrustworthy sources, and intimidate human lives. As a result, entering 2010s, the rapid dissemination of information, including false information, has increased on social media platforms before the white hackers could respond. It was difficult to control the spread of AI-generated content because it is easily shared and amplified on social media, reaching a large audience. Also, it was impossible for humans to fact-check in real-time at the rate that AI can produce content. The problem with AI was certain; they evolve. Once a program that detects the AI driven false information or virus develops, a new AI model is developed as a response which denies the detection system. Simulated intelligence models like GPT-3, for instance, have been prepared to create text that intently emulates human language, making it hard for pursuers to recognize simulated intelligence-produced content and human-composed content. This trait increases the likelihood that false information will be accepted as true.

Problems Raised

Erosion of Trust and Credibility

The erosion of trust and credibility in information sources is one of the principal problems originating from artificial intelligence produced falsehood. Deluding and bogus substance spread through artificial intelligence driven algorithms can sabotage public trust in media sources, foundations, and, surprisingly, logical exploration. This loss of trust has a number of repercussions, including putting democracy and governance at risk, making decisions more difficult, and slowing scientific progress.

Misinformation can sway public opinion, manipulate elections, and undermine the democratic process. Simulated intelligence-controlled deception crusades, intended to spread bogus accounts and disinformation, can energize social orders, disturb talk, and compromise the uprightness of decisions. The foundation of democracy is undermined when individuals are exposed to information that is biased or fabricated. As a result, an informed citizenry is replaced by one that is misinformed.

Social polarization and harm

Social polarization, hate speech, and harmful ideologies have been undermined by the dissemination of falsehood information by AI. Artificial intelligence algorithms frequently focus on content that elicits strong emotional responses. Echo chambers that reinforce extreme views, ideological biases, and conspiracy theories can be created by this tendency and misinformation. It prompts the radicalization of people and fills social divisions, subverting cultural concordance. Simulated intelligence created misinformation can work with the spread of disdain discourse, unfair philosophies, and unsafe generalizations. Platforms that use AI algorithms might promote and spread content that reinforces biases inadvertently, which could exacerbate social divisions and contribute to the marginalization of particular groups. Hostility, discrimination, and inequality thrive as a result.

Deception, particularly when sincerely charged, can negatively affect people's psychological well-
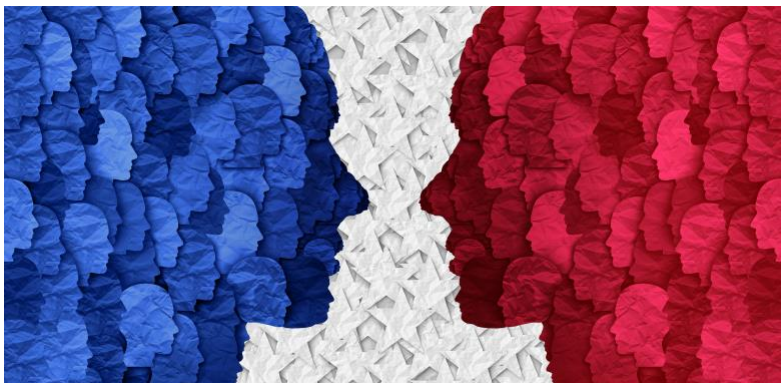

Figure 3. Social polarization Issue

being and prosperity. Anxiety, confusion, and stress can result from constant exposure to false narratives created by AI. Additionally, people may make poor decisions that jeopardize their health as a result of misinformation about public health or medical treatments.

International Actions

Development of AI-based misinformation detection systems

To handle the spread of misinformation, specialists and associations have been chipping away at man-made intelligence models fit for recognizing and hailing deluding or misleading data. These frameworks use Natural Language Processing (NLP) procedures and AI calculations to investigate text, pictures, and recordings. They expect to recognize solid and temperamental sources, distinguish coherent errors or irregularities, and survey the validity of cases made. For example, associations like OpenAI have put resources into innovative work to make computer based intelligence models that can recognize and moderate the spread of falsehood. These models are prepared on huge datasets of solid data and deception to learn examples and marks of false or deceiving content. Platforms can automatically identify and flag potential false information by utilizing such AI models, allowing human fact-checkers to examine the information and take the necessary actions. Additionally, AI models that are capable of detecting deepfake content have been developed. Deepfake content is created by manipulating or fabricating videos or images with a realistic appearance, often celebrities', using artificial intelligence. Deepfake

identification frameworks utilize man-made intelligence algorithms to break down visual and hearable prompts, distinguishing indications of control to help recognize and battle engineered falsehood. In fact, these models have been detected false images and videos that are circulating in the web and specific social media with the collaboration of the headquarters.

Figure 4. AI driven misinformation detection program

Collaborations and partnerships

Tackling the spread of deception requires coordinated effort among different partners, including innovation organizations, specialists, truth checkers, and policymakers. It is possible to share expertise, data, and resources through collaborative efforts, resulting in a more comprehensive and efficient strategy. For instance, associations like the Partnership on AI (PAI) unite industry pioneers, common society gatherings, and scholastic foundations to work on the whole on tending to difficulties presented by man-made intelligence, including deception. Through joint efforts, these substances can pool their assets, coordinate examination endeavors, and foster prescribed procedures for battling falsehood spread by man-made intelligence innovations.

Key Players

The Russian Federation

Russia has been accused for taking part in data fighting and utilizing disinformation missions to further its political plan. The Internet Research Agency (IRA) is a Russian savage ranch that has been engaged with spreading deception and publicity through virtual entertainment stages. It has been connected to endeavors to impact the 2016 United States presidential election by spreading disruptive substance and planting dissension among the American public. The Russian government, through state-controlled news sources and online stages, has been blamed for advancing misleading stories and paranoid ideas. These endeavors expect to subvert popularity-based organizations, undermine different nations, and make disarray among people in general.

People's Republic of China

China is one of the countries that put stringent restriction on web surfing and social media, blocking most of the world's popular social media and web engines from its mainland. The 50 Cent Party takes up a big part that promotes to post pro-government propaganda by manipulating online discussions. Because the country's media outlets are tightly monitored by the government, misleading information and attacks that are prominent on media are effectively blocked and resolved. Moreover, China is one of the member states that put heavily into AI advances for observation and restriction purposes. According the CNN report, China has recently unveiled an updated list of guidelines to manage burgeoning AI industry due to the latest development of Chat GPT and other AI tools. Although there exists some disputes over the action that China takes to trespass privacy of their citizens, it is an undeniable fact that China is the only safety zone from AI driven misinformation as of 2023.

Figure 5. Chinese government regulates AI misinformation

Possible Solutions

Establish AI Ethics Guidelines and Regulatory Frameworks

The UNCSTD could work towards developing comprehensive AI ethics guidelines and regulatory frameworks specifically targeting the spread of misinformation. These guidelines could contain the responsible use of AI technologies and the importance of transparency, accountability, and fairness in AI systems. By the same token, the regulatory frameworks are expected to provide legal restriction and protection in terms of the misuse of AI and could include measures such as mandatory audits of AI
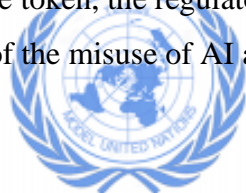
algorithms used by organizations, penalties for non-compliance, and mechanisms for reporting and addressing AI-driven misinformation. These guidelines and frameworks will incorporate governments, AI developers, researchers, and civil society organizations to ensure they are comprehensive, effective, and globally applicable.

Promote AI Literacy and Media Literacy Programs

To address the spread of misinformation by AI, it is essential to equip individuals with the ability to distinguish the reliability and authenticity of the information from the AI sources with their critical sense. Our Commission on Science and Technology for Development should ceaselessly aid member nations, specifically LEDCs, with AI literacy and media literacy programs to educate the public about the complicatedness of AI technologies and the potential risks of dealing with computers.

These programs could be integrated into educational curricula at various levels, focusing on digital literacy, critical thinking, and media literacy skills. Additionally, the UNCSTD could collaborate with technology companies, media organizations, and civil society groups to develop public awareness campaigns that promote responsible AI use, raise awareness about misinformation, and provide tools and resources to verify and fact-check information.

Glossary

Artificial Intelligence

The theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages

Deepfake

a synthetic media that have been digitally manipulated to replace one person's appearance convincingly with that of another

Digital literacy

The ability to access, manage, understand, integrate, communicate, evaluate, and create information safely and appropriately through digital technologies

Natural Language Processing (NLP)

The branch of Artificial Intelligence concerned with giving computers the ability to understand text and spoken words in much the same way human beings can

Sources

"5 Big Problems With OpenAI's ChatGPT." *MUO*, 22 Dec. 2022, www.makeuseof.com/openai-chatgpt-
biggest-probelms/.

"AI and Disinformation in the Russia-Ukraine War." *Enterprise AI*,
www.techtarget.com/searchenterpriseai/feature/AI-and-disinformation-in-the-Russia-Ukraine-
war.

"AI Literacy Program." *AI Center of Excellence*, **15 Mar. 2019, aicoe.ai/ai-literacy-program/.**

"AI-generated Deepfakes Are Moving Fast. Policymakers Can't Keep Up." *NPR*, 27 Apr. 2023,
www.npr.org/2023/04/27/1172387911/how-can-people-spot-fake-images-created-by-artificial-
intelligence.

Akhtar, P., Ghouri, A.M., Khan, H.U.R. *et al*. " Detecting fake news and disinformation using artificial
intelligence and machine learning to avoid supply chain disruptions." *Ann Oper Res* **327**, 633–
657 (2023). https://doi.org/10.1007/s10479-022-05015-5.

"Artificial Intelligence in Automated Detection of Disinformation: A Thematic Analysis." *MDPI*,
3 June 2023, www.mdpi.com/2673-5172/4/2/43.

Cassauwers, Tom. "Can artificial intelligence help end fake news?" *European Commission*,
15 Apr. 2019, ec.europa.eu/research-and-innovation/en/horizon-magazine/can-artificial-
intelligence-help-end-fake-news. Accessed 24 July 2023.

Childs, Martin. "John McCarthy: Computer Scientist Known As the Father of AI." *The Independent*,
1 Nov. 2011, www.independent.co.uk/news/obituaries/john-mccarthy-computer-scientist-known-
as-the-father-of-ai-6255307.html.

"China, Russia Target Audiences Online With Deep Fakes, Replica Front Pages." *VOA*, 23 Mar. 2023,
www.voanews.com/a/china-russia-target-audiences-online-with-deep-fakes-replica-front-pages-
/7018918.html.

"Ethics Guidelines for Trustworthy AI." *Shaping Europe's Digital Future*, **digital-
strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.**

He, Laura. "China Takes Major Step in Regulating Generative AI Services Like ChatGPT | CNN

   Business." *CNN*, 14 July 2023, edition.cnn.com/2023/07/14/tech/china-ai-regulation-intl-

   hnk/index.html.

Oukropec, Jindřich. "A(I)dding Fuel to Fire: How AI accelerates Social Polarization through

   Misinformation." ***House of Ethics*, 5 June 2023, www.houseofethics.lu/2023/06/05/aidding-**

   fuel-to-fire-how-ai-accelerates-social-polarization-through-misinformation/.

"What is the History of Artificial Intelligence (AI)?" *Tableau*, www.tableau.com/data-insights/ai/history.