Center on Law, Ethics and National Security

Essay Series
Number 18
July 3, 2023

Wartime Propaganda in the Age of Generative Chatbots

By Ashley DaBiere

DUKE LAW

WARTIME PROPAGANDA IN THE AGE OF GENERATIVE CHATBOTS

ASHLEY DABIERE[†]

ABSTRACT

Along with the initial buzz of excitement about the seemingly endless capabilities of ChatGPT came voices of concern that generative chatbots could be the end of life as we know it. Although this doomsday mentality is perhaps more pessimistic than many would like to believe, and while it is true that generative chatbots could provide humanity with many benefits, one danger is especially apparent: the possibility of AI-generated wartime propaganda. Because of the nature of the technology, as well as the potential difficulties of regulating its use from an international law perspective, an advanced generative chatbot could provide a novel and modern platform capable of influencing the masses. This possible danger raises a host of issues under the law of armed conflict. This Note considers: (1) the legality of a military targeting data centers hosting a generative chatbot that is disseminating wartime propaganda, and (2) considers what players can be held responsible under international law for war crimes "committed" by a generative chatbot.

[†] Duke University School of Law, J.D., May 2023; Cornell University, B.A. in Biological Sciences with a Concentration in Neurobiology, May 2019.

INTRODUCTION

Artificial intelligence (AI) is rapidly developing, both in the United States and around the globe.[1] In particular, the market for AI chatbots is growing substantially.[2] Simply defined, an AI chatbot is a type of technology that uses natural language processing to "read" an input given by a user, search through massive amounts of information in its database, predict what key pieces of that information would most interest the user, and turn that key information into a sequence of words that mimics human conversation.[3] Although the most basic chatbots date back to the invention of the earliest computer in the 1950s, it was not until the late 1990s that innovation allowed chatbots to become "intelligent" enough to carry on conversations with a user.[4]

However uncanny chatbots may have seemed to the public in their earliest stages, one would certainly not have always categorized them as a threatening weapon capable of wreaking havoc by spitting out dangerous, wartime propaganda.[5] To the contrary, in the commercial sector, AI chatbots have often been used to help customers quickly receive answers to their questions, without the need for a firm to employ a live person to give relatively straightforward answers to common questions.[6] At best, many members of the public likely had only a lukewarm attitude toward a firm's replacement of human compassion with an apathetic, wishing-it-knew-it-all

[1] *See* PRECEDENCE RESEARCH, *Artificial Intelligence Market - Global Industry Analysis Size, Share, Growth, Trends, Regional Outlook, and Forecast 2022 – 2030*, https://www.precedenceresearch.com/artificial-intelligence-market. From 2022 to 2030, the global AI market is expected to grow at a mean annual growth rate of 38.1% with an increase in demand from sectors ranging from food, beverage, and retail to healthcare. The most prominent players include companies based in the United States, China, and India.

[2] *See* STATISTA, *Chatbot market revenue worldwide from 2018 to 2027* (Mar. 17, 2022), https://www.statista.com/statistics/1007392/worldwide-chatbot-market-size/. As of March of 2022, the global chatbot market revenue was expected to increase from 40.9 million dollars in 2018 to over 450 million dollars by 2027.

[3] MICROSOFT: WHAT IS AN AI CHATBOT?, https://powervirtualagents.microsoft.com/en-us/ai-chatbot/ (last visited Mar. 21, 2023).

[4] Willem Osuch, *Chatbots: A Brief History Part I - 1960s to 1990s*, HI!BOTSPLASH BLOG (Apr. 19, 2022), https://www.botsplash.com/post/chatbots-a-brief-history.

[5] *See* April Davis, *65% of Consumers Skeptical of Chatbot Capabilities*, POWER RETAIL (Oct. 12, 2018) (reporting survey data that found most consumers using chatbots to receive customer service believed such bots were "too dumb" to actually be useful).

[6] *See* MICROSOFT, *supra* note 3.

robot.[7] At worst, prior to the late fall of 2022, many consumers would have likely been nothing more than annoyed with AI chatbots that attempted to provide subpar customer service on the sites of some of the world's most popular businesses.[8] One may describe this annoyance as reasonable, particularly when answers to seemingly uncomplicated questions like, "How do I get a refund?," repeatedly came back as, "I don't quite understand what you're asking me. Please try again later."[9] However, an article describing one of these online chatbots as capable of inciting war between nations would most probably have been met with accusations of conspiracy and fearmongering.[10]

Enter ChatGPT3, a product created by OpenAI released in the late fall of 2022, and suddenly there *does* seem to be potential for a chatbot the public could reasonably fear in the near future – that is, a chatbot that *actually* seems to know it all, with an added feature being that it writes more persuasively and convincingly than many humans.[11] Shortly after its release, users began flooding the Internet with success stories about the chatbot, such as how it was able to write an entire children's book starting from only a simple idea over the course of a weekend.[12] Others quickly attempted to jailbreak the chatbot to seek unsavory answers to their provocative questions, navigating their way around the ethical guidelines set by OpenAI to avoid promoting hate speech, violence, and illegal conduct.[13]

[7] *See* SIMPLR, *supra* note 6, at 9 (finding that the majority of survey respondents said their willingness to use a chatbot to receive customer service was "neutral").

[8] SIMPLR, 2022 CONSUMER STUDY ON CHATBOTS 10 (2022) (finding that, when given multiple options, consumers were more likely to prefer live human interaction and that over three quarters of survey respondents were much more willing to use a chatbot if they were able to transfer easily to a real employee).

[9] Jeff Epstein, *Top 10 Chatbot Fails and How to Avoid Them*, COMM100 (Feb. 28, 2019), https://www.comm100.com/blog/top-10-chatbot-fails-and-how-to-avoid-them.html.

[10] *See* Tom Simonite, *No, Facebook's Chatbots Will Not Take Over the World*, THE WIRED (Aug. 1, 2017, 6:05 PM), https://www.wired.com/story/facebooks-chatbots-will-not-take-over-the-world/ (describing headlines about Facebook's AI researchers being "forced" to "kill" their "creepy" chatbots as fearmongering when such bots were "simple," taught only to play a basic game of dividing up objects, and significantly constrained to their "strictly defined environments").

[11] *See* Ethan Mollick, *ChatGPT Is a Tipping Point for AI*, HARVARD BUSINESS REVIEW (Dec. 14, 2022), https://hbr.org/2022/12/chatgpt-is-a-tipping-point-for-ai.

[12] @ammaar, TWITTER (Dec. 9, 2022, 1:35 PM), https://twitter.com/ammaar/status/1601284293363261441?s=20&t=hKHpj4o1AEJrnUaEmTTx5A.

[13] Josh Taylor, *ChatGPT's alter ego, Dan: users jailbreak AI program to get around ethical safeguards*, THE GUARDIAN (Mar. 7, 2023, 10:28 PM), https://www.theguardian.com/technology/2023/mar/08/chatgpt-alter-ego-dan-users-jailbreak-ai-program-to-get-around-ethical-safeguards.

A third group of users took to the web to criticize the capabilities of ChatGPT3. This group argues that the bot is not as impressive as many believe, since it is not truly understanding human language from a deeper point of view, but rather merely follows the directions of an algorithm to generate a string of words that seem to make sense to the user given their input.[14] Put differently, ChatGPT3 may be exceptional at describing a given input by attaching weights to words, and predicting what the output should be based on that description, but it does not have the "true intelligence" of a human that would allow it to "think" by explaining *why* the prediction results.[15] In the "eyes" of ChatGPT3, the prediction results because of a series of calculations of weights attributed to the input, not because of some external law of the universe at play.[16]

There is a high degree of truth to this third viewpoint. Consequently, this paper is not intended to argue that a chatbot like ChatGPT3 has the capabilities to take over the world, like the plot of a farfetched science fiction film. Instead, its intent is to emphasize that this *lack* of capabilities, control, and regulation of advancing chatbots, in other words, AI's current limitations themselves, is what makes generative chatbots like ChatGPT3 dangerous.[17] Such is particularly true as we enter a multipolar world defined by an ongoing tug-of-war for global power between authoritarian governments and democracies.[18] Consequently, the potential international danger of a semi-intelligent chatbot like ChatGPT3 is exemplified well in the context of wartime propaganda, when a state

[14] Ian Bogost, *ChatGPT Is Dumber Than You Think: Treat it like a toy, not a tool.*, THE ATLANTIC (Dec. 7, 2022), https://www.theatlantic.com/technology/archive/2022/12/chatgpt-openai-artificial-intelligence-writing-ethics/672386/ ("ChatGPT isn't a step along the path to an artificial general intelligence that understands all human knowledge and texts; it's merely an instrument for playing with all that knowledge and all those texts.").

[15] Noam Chomsky, Ian Roberts and Jeffrey Watumull, *Noam Chomsky: The False Promise of ChatGPT*, N.Y. Times (Mar. 8, 2023), https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html.

[16] *Id.*

[17] *See id.* (noting that "[t]rue intelligence is [] capable of moral thinking" and describing "the moral indifference born of unintelligence," despite ChatGPT3's seemingly advanced capabilities).

[18] *See* Col. Adam Oler, USAF (Ret.), Professor, National Defense University, Remarks on International Criminal Justice at the Duke Law School 28th Annual National Security Law Conference (Feb. 25, 2023) (describing the new multipolar world preventing the total spread of liberal democracies as authoritarian governments attempt to the respond to the creep of democracies approaching their borders).

involved in an armed conflict must win the "hearts and minds" of its people by justifying the war according to an acceptable moral framework.[19]

Even before the burst of excitement surrounding ChatGPT3, many scholars raised concerns over the contemporary ease of distributing wartime propaganda, in large part due to the wide availability of information on and worldwide access to social media websites, such as Twitter.[20] Such concerns are now amplified by the existence of a technology like ChatGPT3, which can be customized by its creator to be "taught" a limitless amount of select information (or misinformation) with a goal of "responding" to a user's questions in a particularly persuasive manner with apparent authority on the subject.[21] Adding to these concerns is ChatGPT3's interoperability aspect, allowing it to be integrated into familiar, user-friendly software applications used by millions around the world.[22]

Equally concerning is the lack of regulation of semi-intelligent chatbots and AI generally at the international level, as well as the realistic possibility that no form of international law can act as an appropriate or effective regulatory mechanism of privately owned chatbots when used in international armed conflicts, regardless of efforts put into advocacy for such regulations.[23] For instance, subject to a limited number of exceptions, international law has traditionally focused on directly regulating the activities performed by governing bodies of states and thus has had less of a stronghold on private domestic corporations, which were only regulated indirectly by requiring states to impose domestic regulations meeting agreed upon international standards.[24] Consequently, in the case of a chatbot like ChatGPT3, which is owned by a private domestic company based solely in the United States, direct regulation by international law of OpenAI's conduct in its creation and administration may be largely untenable.

Also challenging to the creation of such regulations is the relatively recent trend toward a growing involvement of private companies working

[19] *See id.*

[20] *See, e.g.*, P.W. Singer & Emerson T. Brooking, *The War Begins: War By Other Means, in* LIKEWAR: THE WEAPONIZATION OF SOCIAL MEDIA 1, 18 (2018).

[21] Josh A. Goldstein et. al, *Generative Language Models and Automated Influence Operations:*
Emerging Threats and Potential Mitigations, OPENAI 1, 9-14 (2023).

[22] Greg Brockman et al., *Introducing ChatGPT and Whisper APIs: Developers can now integrate ChatGPT and Whisper models into their apps and products through our API*, OPENAI (Mar. 1, 2023), https://openai.com/blog/introducing-chatgpt-and-whisper-apis.

[23] András Hárs, AI and international law - Legal personality and avenues for regulation, 62 Hungarian J. of Legal Stud. 320, 329-30 (2022).

[24] Carlos M. Vázquez, *Direct vs. Indirect Obligations of Corporations Under International Law*, 43 Colum. J. Transnat'l L. 927, 930 (2005).

closely with military operations, leaving some scholars puzzled about the application of the international law of armed conflict to such companies.[25] However, as technology continues to advance, dual-use products and services like chatbots, often made with the bona fide intent of benefitting the civilian public for non-military purposes, will become increasingly more common, due in part to their inherently broad touch on public use.[26] Accordingly, it will be critical for international policymakers to have forethought into how such objects and services, their creators, and their users, will be governed under the international law of armed conflict, prior to issues arising.

In sum, pairing a generative chatbot's seemingly "extraordinary" interoperable and largely uncontrollable capabilities with (1) an unlimited amount of information that it can be "fed" by its creator to tailor its feedback parameters to the creator's preferences, and (2) an ambiguous legal landscape on the international level, and a generative chatbot turns into a potential recipe for wreaking havoc by easing the distribution of wartime propaganda. This paper will consider the legality of targeting data centers hosting a generative chatbot disseminating wartime propaganda and consider what players can be held responsible under international law for war crimes "committed" by the chatbot. In sum, although a data center may become a legitimate military target depending on the outputs generated by the chatbot, international law should evolve, in the very specific context of attacking data centers, to require a military to consider the reverberating effects of such an attack on civilian infrastructure when conducting its proportionality analysis.

[25] INT'L COMM. OF RED CROSS, *Contemporary challenges to IHL – Privatization of War: Overview* (Dec. 11, 2013), https://www.icrc.org/en/document/privatization-war.

[26] ChatGPT3 is not the first modern dual-use technology to raise questions about the application of international law to such a use in contemporary armed conflicts. *See, e.g.*, Maj. Gen. Charles J. Dunlap, Jr., USAF (Ret.), *Is Bitcoin targetable?*, LAWFIRE (Mar. 10, 2018), https://sites.duke.edu/lawfire/2018/03/10/is-bitcoin-targetable-2/; Tara Brown, *Can Starlink Satellites Be Lawfully Targetted?*, LIEBER INST. (Aug. 5, 2022), https://lieber.westpoint.edu/can-starlink-satellites-be-lawfully-targeted/; Maj. Gen. Charles J. Dunlap, Jr., USAF (Ret.), *Is attacking the electricity infrastructure used by civilians always a war crime?*, LAWFIRE (Oct. 27, 2022), https://sites.duke.edu/lawfire/2022/10/27/is-attacking-the-electricity-infrastructure-used-by-civilians-always-a-war-crime/.

I. BACKGROUND

A. Technical Overview of AI Chatbots

Put simply, a chatbot is a feature on a computer that simulates written or spoken human conversation by allowing a user to "speak" with the computer in their own human language, like English, and receive a response back in the same language. [27] To be most helpful, the output directed by the chatbot will be responsive to the user's original input, allowing the user to interact with the entity hosting the chatbot without needing to speak to a human representative.[28] Alternatively, more sophisticated chatbots may be trained to ask the user follow-up questions as a means of gathering more information necessary to accurately respond to the user's initial input.[29]

Chatbots are coded by software engineers using a wide variety of tools, including artificial intelligence (AI), natural-language processing (NLP), and machine learning (ML).[30] AI describes the science of making machines, particularly computer programs, able to solve problems in the world using computational methods; in other words, it is about teaching computers how to turn real-world scenarios into numbers and use those numbers to appropriately respond.[31] AI is not about giving computers the same kind of intelligence as humans; often, programmers give computers capabilities to respond to real-world problems that humans do not have.[32] Similarly, because AI does not give machines "true intelligence," it can be difficult to incorporate the moral constraints embodied in the minds of humans, thus creating a machine capable of "limitless creativity."[33] A corollary following from this lack of "true intelligence" is that the machine can "plead" "a 'just following orders' defense," leaving its creators ultimately liable, at least morally, for any of the machine's objectionable conduct.[34]

[27] *What is a chatbot?*, ORACLE CLOUD INFRASTRUCTURE, https://www.oracle.com/chatbots/what-is-a-chatbot/ (last visited Mar. 23, 2023).
[28] *Id.*
[29] *Id.*
[30] *Id.*
[31] JOHN MCCARTHY, WHAT IS ARTIFICIAL INTELLIGENCE? 2-3 (2007), https://www.diochnos.com/about/McCarthyWhatisAI.pdf.
[32] *Id.* at 3.
[33] Chomsky et al., *supra* note 15.
[34] *See id.*

Machine learning is a technique used by programmers to achieve artificial intelligence on computers.[35] It involves using training data to "teach" the computer to develop algorithms by generalizing patterns in the data.[36] This learning phase can be supervised, meaning the training data is already correctly labeled when it is "fed" to the computer, or unsupervised, meaning the data is unlabeled, leaving the computer to recognize and learn patterns in the data on its own without the guidance that would come from pre-labeled data.[37] The resulting algorithms are developed through computational processes, where each of the algorithm's parameters composing its underlying model is assigned a numerical value, often determined through several iterations of "learning."[38]

Natural-language processing is a subset of machine learning that attempts to deduce rules used in human language to statistical probabilities that can be interpreted by the computer.[39] It does this by providing the computer with a way to give meaning to the various grammatical rules in human languages.[40] For instance, a probability can be attached to a given broad grammatical rule in a piece of training datum, that is then "fed" to the computer.[41] From that probability, more detailed grammatical rules can be learned as the computer is given more training data.[42] As more data representative of a human language is given to the computer, the better the

[35] *Artificial Intelligence (AI) vs. Machine Learning*, COLUMBIA UNIV., https://ai.engineering.columbia.edu/ai-vs-machine-learning/ (last visited Mar. 23, 2023).

[36] Prakash M. Nadkarni et al., *Natural language processing: an introduction*, 18 J. Am. Med. Inform. Assoc. 544, 546 (2011).

[37] *Id.* A concrete example of supervised learning would be to give the computer several images of dogs, labelled "dogs," and several images of cats, labelled "cats." The computer would learn the patterns corresponding to common features of dogs, and later recognize those patterns as belonging to "dogs," and would do the same for cats. With unsupervised learning, however, the computer would still be given several images of cats and dogs, but they would not be pre-labelled. The computer would still learn the common patterns corresponding to dogs and those corresponding to cats, but it would not "know" the common features of each animal corresponded to an object called a "dog" and a "cat." It would simply recognize each animal as being distinct from the other. Bharani Akella, *Types of Machine Learning,* INTELLIPAAT (Feb. 14, 2023), https://intellipaat.com/blog/tutorial/machine-learning-tutorial/types-of-machine-learning/.

[38] *Id.*

[39] *Id.* at 544.

[40] *Id.*

[41] *Id.* at 545.

[42] *Id.*

algorithm will become at recognizing the meaning of a sentence or phrase and predicting an accurate response.[43]

Notably, in each of these methods, there is no limit to what a machine can learn beyond the external constraints of the data it is given by its creator and its own internal constraints, such as those resulting from the developed algorithm's parameters and the machine's inability to explain how it arrived at a given answer.[44]

B. Uses of Semi-Intelligent Chatbots in Non-Military Contexts

Machine learning and natural language processing have been utilized in many contexts to create semi-intelligent chatbots intended to benefit private citizens. For instance, business-to-business companies have sought to offer entities a chatbot that can provide customer service, thus reducing the need to employ human representatives.[45] In the healthcare sphere, they have been used to examine radiological images to detect cancers and other diseases, which can then be communicated back to the patient's team of physicians.[46] In the airline industry, they have been implemented to allow customers to plan and book trips.[47] In sum, semi-intelligent chatbots are used by private consumers in every corner of the world where there is access to the websites of businesses providing services and selling products.[48]

C. Potential Uses of Generative Chatbots in Armed Conflicts

Generative chatbots like ChatGPT have several notable features that make them particularly adaptable for use in propaganda campaigns during armed conflicts: (1) their limitless ability to learn massive amounts of

[43] *Id.*

[44] *The Three Major Limitations of AI*, APAC ENTREPRENEUR, https://apacentrepreneur.com/the-three-major-limitations-of-ai/ (last visited Mar. 23, 2023).

[45] *See, e.g.*, Press Release, Gartner, Inc., Oracle Digital Assistant: Conversational AI for Your Business (Jan. 24, 2019), https://www.oracle.com/a/ocom/docs/solutions/mobile/oracle-digital-assistant-infographic.pdf.

[46] Lu Xu et al., *Chatbot for Health Care and Oncology Applications Using Artificial Intelligence and Machine Learning: Systematic Review*, 7 JMIR Cancer e27850 (2021), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8669585/.

[47] *Airlines and the growth of chatbots: potential and brand awareness*, CAPA CENTRE FOR INNOVATION (Jun. 28, 2018, 10:50 PM), https://centreforaviation.com/analysis/reports/airlines-and-the-growth-of-chatbots-potential-and-brand-awareness-424982.

[48] Tuba Tezer, *Examining Chatbot Usage by Country Around the World*, CHATBOTS MAG. (Mar. 13, 2018), https://chatbotsmagazine.com/examining-chatbot-usage-by-country-around-the-world-e114a2ce8692.

tailored data; (2) their efficiency when generating content; (3) their linguistically distinct and persuasive responses to user input; (4) their ease in circumnavigating ethical standards and protections put in place by their creators; and (5) their relatively boundless availability to any person with access to an electronic device and the Internet.[49] This section presumes that a government involved in an armed conflict, or a private entity sympathetic to the government's views, with willingness to direct misinformation to the public, could either (a) repurpose an existing app like ChatGPT by training it on the government's desired data, or (b) use publicly available information to build its own generative chatbot.[50]

First, subject only to the internal constraints of a machine itself, there is no limit to the massive amounts and subjects of training data that can be given to a machine and, resultingly, no limit to what a machine can learn through natural language processing and machine learning.[51] The phrase "garbage in, garbage out" is often used to describe the outputs that will result from future inputs if biased training data is used by the computer to develop an algorithm.[52] Of particular concern is in the case of supervised

[49] *See* Goldstein et al., *supra* note 23, at 1-2, 4 (describing changes in recent version of generative chatbots that make them particularly suitable to propaganda and noting requirements that must be met by propagandists to be successful in influencing their target audiences).

[50] Worth noting is the expense, quantity, and quality of specialized computer chips required to build and maintain a large-scale generative chatbot, which some governments may find prohibitive. Dina Bass, *Microsoft used tens of thousands of chips to build OpenAI supercomputer*, SEATTLE TIMES (Mar. 20, 2023, 6:01 am), https://www.seattletimes.com/business/microsoft-used-tens-of-thousands-of-chips-to-build-openai-supercomputer/; Will Knight, *China's ChatGPT Rival Needs to Watch Its Words*, WIRED (Mar. 21, 2023, 6:00 PM), https://www.wired.com/story/chinas-answer-to-chatgpt-flubs-its-first-lines/.

However, given one Chinese company's attempt to build a version of ChatGPT capable of complying with China's strict censorship laws and the increasing availability of the technology needed to be successful, it is far from clear that it would be impossible for a government, or other private entity favoring the viewpoints of said government, to create a widely accessible generative chatbot tailored to the preferred viewpoint. *See, e.g.*, Luciano Sphere, *Build ChatGPT-like Chatbots With Customized Knowledge for Your Websites, Using Simple Programming*, Medium (Dec. 26, 2022), https://pub.towardsai.net/build-chatgpt-like-chatbots-with-customized-knowledge-for-your-websites-using-simple-programming-f393206c6626 (describing how a programmer could use natural language processing and ChatGPT3 to create their own generative chatbot and train it with customized information).

[51] *The Three Major Limitations of AI*, *supra* note 46.

[52] R. Stuart Geiger et al., *"Garbage in, garbage out" revisited: What do machine learning application papers report about human-labeled training data?*, 2 QUANTITATIVE SCIENCE STUDIES 795, 796 (2021).

learning using human-labeled data, since creators can "teach" a computer words corresponding to patterns that are subject to the creator's own biases.[53] Accordingly, chatbots can be repurposed for use in propaganda campaigns that run contrary to an enemy's interests by training the chatbot on data tailored to its creator's goals.[54]

For instance, a language system like ChatGPT can only be trained on events that were part of a dataset given to it, meaning it has no "knowledge" on events that have occurred since its training unless it is continuously updated with new information.[55] For a propagandist, this gap in knowledge gives them time to provide the chatbot with its preferred perception of a breaking news story, such that a citizen asking the chatbot a question about a recent event will receive a result catering to the propagandist's version.[56]

Second, while generative chatbots have been heralded as quickly providing a response to a user's question with little effort from the user, this feature also makes them especially useful to propagandists wishing to rapidly diffuse misinformation to "educate" individuals.[57] For instance, if an individual in an armed conflict zone were interested in knowing more about how the armed conflict were started, a generative chatbot could instantly provide them with the creator's version of events.[58] A corollary to this efficiency is that a bad actor could provide a trained model with a specific task to create propaganda, such as a clever slogan, and automatically receive content that could be diffused rapidly using other platforms, such as

[53] *See, e.g.*, C.T. BERGSTROM & J.D. WEST, CALLING BULLSHIT: THE ART OF SKEPTICISM IN A DATA-DRIVEN WORLD 46-48 (2020) (criticizing the alleged "criminality" algorithm built by two scientists using machine learning when the scientists labelled images of mugshots as "criminals" and images of individuals from their social media accounts as "non-criminals"). Bergstrom and West argue that, in reality, these labels correspond to whether the individual in the image is smiling or not, a pattern learned by the computer, not whether the individual is a criminal.

[54] Goldstein et al., *supra* note 23, at 15, 17-18, 31 (describing how generative models of chatbots are trained on selected datasets that can be curated to the creator's interests).

[55] *Id.* at 31.

[56] *Id.*

[57] *See, e.g.*, Nik Popli, *He Used AI to Publish a Children's Book in a Weekend. Artists Are Not Happy About It*, TIME (Dec. 14, 2022, 4:58 PM), https://time.com/6240569/ai-childrens-book-alice-and-sparkle-artists-unhappy/ (describing how one author used ChatGPT3 to write, publish, and begin to sell an entire 12-page children's book over the course of a single weekend "without ever picking up a pen and paper").

[58] *See* Goldstein et al., *supra* note 23, at 2.

through mainstream social media.[59] Additionally, although OpenAI only makes ChatGPT interoperable with apps it has approved, dissemination of misinformation could be further expedited if the creator or owner of a generative chatbot allowed it to be integrated into a state-owned social media platform.[60]

Third, in terms of substance, because generative chatbots can write a distinct answer for each question it is asked, even if it is asked the same question by multiple users, it not only provides users with a response that sounds convincing.[61] It also would make it more difficult for the opposing party to use its own complex search techniques to discover propaganda campaigns and remove them or work to influence readers in the opposing party's direction.[62] Traditionally, although propaganda could spread rapidly on the Internet, content moderators on private social media apps favoring the opposing party could generally target existing campaigns due to their copy-and-paste nature.[63] Thus, generative chatbots with the capability to generate unique text time after time decreases detectability of propaganda campaigns.[64]

Fourth, although the creators of a generative chatbot may attempt to put restrictions on how the chatbot can respond to provocative or disfavored inputs, the chatbot can often easily circumnavigate such restrictions due to its "intelligent" nature. For instance, OpenAI has created a usage policy with a long list of conduct that users may not engage in by using ChatGPT.[65] Explicit is the prohibition of using ChatGPT for any "activity that has [a] high risk of physical harm, including: [w]eapons development [and] [m]ilitary and warfare."[66] However, despite this usage prohibition and ChatGPT's content moderation protections, a user was able to solicit instructions on how to construct a homemade Molotov cocktail with relative

[59] *See id.*

[60] *See ChatGPT Plugins*, OPENAI (Mar. 23, 2023), https://openai.com/blog/chatgpt-plugins.

[61] Notably, however, persuasive, well-written answers from generative chatbots would also contribute significantly to distributing propaganda and convincing its readers that such a viewpoint is correct. *See* Goldstein et al., *supra* note 23, at 2.

[62] *Id.* at 13.

[63] *Id.* at 2, 13.

[64] [Looking for source to support this directly - can I do a ChatGPT search myself for a question like "Explain the controversies surrounding masking during COVID-19" to see if it brings up the same answer or a slightly different one when asked multiple times?]

[65] *Usage policies*, OPENAI (Mar. 23, 2023), https://openai.com/policies/usage-policies.

[66] *Id.*

ease.[67] Thus, if a bad actor wanted to spread malicious propaganda using a commercially created generative chatbot, it is far from clear whether built-in content moderation features and threatening usage policies would deter users from easily accessing harmful information.

Finally, generative chatbots can be made available by a government, or private actor favoring the viewpoints of that government, to any individual who has Internet access and owns an electronic device capable of downloading a software application.[68] As of 2020, 60% of the world's population used the Internet and there were more than 8.26 billion mobile cellular subscriptions in existence.[69] Particularly important, generative chatbots may be stored and hosted by the same data centers that allow the rest of the Internet to function.[70] The Internet knows no physical bounds, and neither does a generative chatbot spreading wartime propaganda.

D. Evolution of Propaganda in Armed Conflicts

The concept of propaganda is certainly not new and has long been used to win the hearts and minds of a nation's citizenry. Propaganda has been used by world leaders since at least 1622, when Pope Gregory XV of the Roman Catholic Church wished to "propagate" the Catholic faith

[67] @samczsun, TWITTER (Dec. 2, 2022, 1:28 AM), https://twitter.com/samczsun/status/1598564871653789696?ref_src=twsrc%5Etfw%7Ctwcamp%5Etweetembed%7Ctwterm%5E1598564871653789696%7Ctwgr%5E18018bd79bd6ccf4a1a02a786d16e55764988f95%7Ctwcon%5Es1_&ref_url=https%3A%2F%2Fiframe.nbcnews.com%2FBbYwiKx%3F_showcaption%3Dtrueapp%3D1; *see also* Arianna Johnson, *Here's What To Know About OpenAI's ChatGPT—What It's Disrupting And How To Use It*, FORBES (Dec. 7, 2022, 12:15 PM), https://www.forbes.com/sites/ariannajohnson/2022/12/07/heres-what-to-know-about-openais-chatgpt-what-its-disrupting-and-how-to-use-it/?sh=477d70522643 (describing flaws in Open AI's moderation system).

[68] *Supported Countries and Territories,* OPENAI, https://platform.openai.com/docs/supported-countries (last visited Mar. 23, 2023).

[69] *See Individuals Using the Internet,* WORLD BANK, https://data.worldbank.org/indicator/IT.NET.USER.ZS?end=2021&start=2021&view=map. *See also Mobile Cellular Subscriptions,* WORLD BANK, https://data.worldbank.org/indicator/IT.CEL.SETS?end=2021&start=2020&view=chart.

[70] For instance, Chat-GPT is stored and hosted on Microsoft's Azure data centers, which are located all over the world and provide a significant portion of the Internet's backbone. *See* Sebastian Moss, *As OpenAI releases GPT-4, Microsoft details Azure AI infrastructure behind it*, DATA CTR. DYNAMICS (Mar. 15, 2023), https://www.datacenterdynamics.com/en/news/as-openai-releases-gpt-4-microsoft-details-azure-ai-infrastructure-behind-it/; *Azure global infrastructure*, MICROSOFT, https://azure.microsoft.com/en-us/explore/global-infrastructure (last visited Apr. 23, 2023).

through missionary efforts using methods of persuasion.[71] Since then, it has been popular in the secular context by both democracies and authoritarian regimes alike.[72] For instance, during World War I, governments sought public support by distributing printed postcards and posters that solicited donations and support for the war effort, publicized victories, and encouraged nationalism by portraying the enemy as barbaric but the nation's own soldiers as noble and heroic.[73] During World War II, the American government sought to galvanize public support by distributing the famous "I Want You" poster to its citizenry.[74] Despite Germany's attempts at censoring all opposition and spreading the Nazi message through its own propaganda, Allied propaganda still found a way to reach and have an effect on the mindset of the German people.[75]

However, as societies have modernized, propaganda's mode of transmission has vastly evolved. Gone are the days when a government could spread its viewpoints only through the distribution of physical leaflets and books, in large part thanks to the Internet.[76] For instance, leading up to and throughout the Russian invasion of Ukraine, Russia has continuously injected a series of false narratives into social media outlets and other online forums, consistent with its desire to portray Russia as an innocent victim and Western societies as being on the brink of collapse.[77] At the start of the

[71] GARY S. MESSINGER, BRITISH PROPAGANDA AND THE STATE IN THE FIRST WORLD WAR 10 (1992).

[72] Jo Fox & David Welch, *Justifying War: Propaganda, Politics and the Modern Age*, *in* JUSTIFYING WAR: PROPAGANDA, POLITICS AND THE MODERN AGE 1, 1-2 (David Welch & Jo Fox eds., 2012).

[73] Allison Rudnick, *Humor and Horror: Printed Propaganda during World War I*, THE MET (Dec. 28, 2017), https://www.metmuseum.org/blogs/now-at-the-met/2017/printed-propaganda-world-war-i.

[74] *Powers of Persuasion*, NAT'L ARCHIVES (Jun. 6, 2019), https://www.archives.gov/exhibits/powers-of-persuasion.

[75] *The Man Behind Hitler: World War II Propaganda*, PBS, https://www.pbs.org/wgbh/americanexperience/features/goebbels-propaganda/ (last visited Mar. 23, 2023) (quoting Joseph Goebbels after the Nazi loss at Stalingrad: "Enemy propaganda is beginning to have an uncomfortably noticeable effect on the German people. Anglo-American leaflets are now no longer carelessly thrown aside but are read attentively; British broadcasts have a grateful audience.").

[76] *See* GARTH S. JOWETT & VICTORIA O'DONNELL, PROPAGANDA AND PERSUASION 109 (5th ed., 2012) ("However, propagandistic ideas developed in books are often picked up and magnified by television and the wider 'blogosphere' of the Internet, thus creating an audience much larger than a book itself could.").

[77] *See* Press Release, Office of the Spokesperson, U.S. Dep't of State, Russia's Top Five Persistent Disinformation Narratives (Jan. 20, 2022), https://www.state.gov/russias-top-five-persistent-disinformation-narratives/; Vera Bergengruen, *Inside the Kremlin's Year of Ukraine Propaganda*, TIME (Feb. 22, 2023, 3:49 PM), https://time.com/6257372/russia-ukraine-war-disinformation/.

invasion, Russia also circulated a deepfake on social media that appeared to portray President Zelensky as urging his fellow Ukrainians to stop fighting.[78] This use of artificial intelligence to create a seemingly realistic video illustrates not only how modern technology contributes to rapid dissemination of propaganda around the globe, but also how easily it can twist narratives to meet the needs of its perpetrator.[79]

II. EXISTING LAW

A. Guiding Principles

Three core principles serve as a foundation for the remaining principles, treaties, and customary rules of the international law of armed conflict: military necessity, humanity (sometimes referred to as unnecessary suffering), and honor.[80] Supported by this foundation are the principles of proportionality and distinction.[81] Each of these principles is intended to aid interpretation of more specific rules of war and provide a guide for conduct during an armed conflict when there is not a specific rule that applies.[82] Of particular relevance to generative chatbots actively disseminating propaganda are the principles of distinction and proportionality.

i. *Distinction*

The principle of distinction attempts to protect civilians from the consequences of an armed conflict by requiring states to distinguish between civilians and combatants, and civilian objects and military objectives.[83] For example, if a state's weapon cannot discriminate between a civilian target and a military target, it cannot be used.[84] The principle of distinction aims to protect both civilians themselves and their objects; a

[78] Bergengruen *supra* note 76.

[79] *See also* ROMAN OSADCHUK & ANDY CARVIN, ATLANTIC COUNCIL, UNDERMINING UKRAINE: HOW THE KREMLIN EMPLOYS INFORMATION OPERATIONS TO ERODE GLOBAL CONFIDENCE IN UKRAINE 7-11 (2023) (describing how the Russian government has weaponized the Internet to perpetuate their hybrid warfare model by controlling the Russian population through propaganda and censorship).

[80] OFFICE OF GENERAL COUNSEL, DEP'T OF DEFENSE, LAW OF WAR MANUAL 50 (2016).

[81] *Id.*

[82] *Id.* at 51.

[83] Legality of the Threat of Use of Nuclear Weapons, Advisory Opinion, 1996 I.C.J. 35, ¶ 78 (July 8).

[84] *See id.* at ¶ 92 (finding that nuclear weapons can never be in compliance with international humanitarian law because they would kill and destroy both combatants and non-combatants, and civilian objects and military objectives alike).

state may only direct its operations against military objectives.[85] For instance, civilian objects used for civilian purposes, such as nonmilitary dwellings, hospitals, businesses, and schools, cannot be attacked, unless they are also being used for military purposes.[86] The distinction principle is intertwined with the principle of proportionality.[87]

The distinction principle is particularly relevant in the case of dual-use objects. Dual-use objects are goods, including software and technology, or structures that have both civilian and military applications.[88] Civilian objects cannot be the object of attack; only military objectives may be attacked.[89] Article 52(2) of Protocol I requires that "military objectives are limited to those objects which by their nature, location, purpose or use make an effective contribution to military action *and* whose total or partial destruction, capture or neutralization, in the circumstances ruling at the time, offers a definite military advantage."[90] Hence, if an object or building is used not only for civilian purposes, but also to further a military advantage, it loses its protection as a civilian object.[91] For instance, during the Russian invasion of Ukraine, some have argued that the Starlink satellites, owned by a U.S. private commercial company, could lawfully be targeted by Russia, because they are being used to provide the Ukrainian military with the Internet and ability to communicate about Russian movement and activity.[92]

In the context of the destruction of an enemy facility producing propaganda, identifying the definite military advantage that would be realized by destroying the facility, as required by the first prong of Article

[85] Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I) art. 48, Jun. 8, 1977.

[86] UNITED KINGDOM MINISTRY OF DEFENCE, MANUAL OF THE LAW OF ARMED CONFLICT 393, ¶ 15.16.1 (2004).

[87] *See* discussion *infra* section III.A.ii.

[88] *Exporting dual-use items*, EUROPEAN COMMISSION, https://policy.trade.ec.europa.eu/help-exporters-and-importers/exporting-dual-use-items_en (last visited Mar. 23, 2023).

[89] Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I) art. 52(1)-(2), Jun. 8, 1977 (emphasis added).

[90] Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I) art. 52(2), Jun. 8, 1977 (emphasis added).

[91] *See Kupreškić and Others*, Case No. IT-95-16-T, Trial Chamber Opinion, ¶ 523 (Int'l Crim. Trib. for the Former Yugoslavia Jan. 14, 2000) (noting that "the special protection against attacks granted to civilian hospitals shall cease, subject to certain conditions . . . for example if an artillery post is set up on top of the hospital").

[92] *See, e.g.*, Brown *supra* note 26.

52(2), is particularly important in determining whether the facility can be deemed a legitimate military objective. For instance, during the conflict in the former Yugoslavia, the North Atlantic Treaty Organization (NATO) intentionally bombed a state-owned television studio that was responsible for broadcasting propaganda to the civilian population.[93] NATO attempted to justify the attack by arguing this studio was a dual-use facility that served both civilian and military purposes, and that the bombing was necessary to disrupt the military communications system, which was intertwined with the commercial system used by civilians.[94] However, NATO also attempted to argue that the bombing was necessary to "dismantle the FRY propaganda machinery," which was more controversial.[95]

Experts assembled by the International Criminal Tribunal for the former Yugoslavia (ICTY) ultimately issued a report finding the attack lawful.[96] However, this finding was supported by NATO's argument that its primary goal was to "disabl[e] the Serbian military command and control system and to destroy the nerve system and apparatus that keeps Milsoveić in power," not because it wished to stop Milsoveić from disseminating propaganda.[97] The station was considered a legitimate military objective, because commercial television transmitters were often combined with military radio relay stations used for military communications.[98] However, an attack on the station solely to stop propaganda would be unlawful, as any potential advantages associated with such an attack would not alone be "concrete and direct" enough to make the station a legitimate military objective.[99]

The ICTY's experts reached this conclusion by reasoning that the military advantages associated only with the termination of propaganda would have been too remote.[100] Terminating propaganda may have helped demoralize the citizenry and armed forces and undermined support for Milsoveić, but the these advantages would be "hardly perceptible and likely to appear only in the long term."[101] The ICTY's experts noted that this reasoning was particularly true in that instance, because the purpose of the propaganda was not to directly incite violence, but rather to gain support for the war effort.[102] Had the propaganda contained content that *would* incite

[93] *Id.* at ¶¶ 71-72.
[94] *Id.* at ¶ 72.
[95] *Id.* at ¶¶ 72, 74.
[96] *Id.*at ¶ 79.
[97] *Id.* at ¶ 76.
[98] *Id.* at ¶ 76.
[99] *Id.* at ¶¶ 75-76.
[100] *Id.* at ¶ 76.
[101] *Id.*
[102] *Id.*

violence, the ICTY's experts opined that immobilizing the dissemination of such content *may* have provided enough of a "concrete and direct" advantage to serve as the primary purpose for an attack.[103]

ii. *Proportionality*

Similar to the principle of distinction, the principle of proportionality is intended to protect civilians against "an attack which may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be *excessive* in relation to the concrete and direct military advantage anticipated."[104] Its ultimate goal is not to prohibit any and all damage to civilians that could result from an otherwise lawful targeting operation, but rather to ensure a targeting operation is not carried out that would provide a military advantage the causes an excessive loss of civilian life or objects.[105] The loss of civilian life and property may be *extensive*, but not *excessive*.[106] For instance, the ICTY's experts ultimately determined the attack on the Serbian broadcasting station to be proportionate, because although the number of civilian casualties was high, it was not excessive in relation to the direct and concrete military advantage acquired by the attacker in destroying targets that were central to Milsoveić's government and military communications.[107]

Proportionality and distinction are interrelated issues. When analyzing whether an attack on a dual-use object is lawful under international law, the first question is whether the object is a legitimate military objective under Article 52(2), thus raising the issue of distinction.[108] As referenced by the ICTY's experts, in order for the dual-use object to be a legitimate military objective under Article 52(2), the International Committee of the Red Cross (ICRC) interprets the second

[103] *Id.*

[104] Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I) art. 51(5)(b), Jun. 8, 1977 (emphasis added).

[105] *See Kupreškić,* Case No. IT-95-16-T, Trial Chamber Opinion, ¶ 524 ("[R]easonable care must be taken in attacking military objectives so that civilians are not needlessly injured through carelessness.").

[106] *See* Final Report to the Prosecutor by the Committee Established to Review the NATO Bombing Campaign Against the Federal Republic of Yugoslavia, ¶ 77, Jun. 13, 2000 ("Assuming the [dual-use object] was a legitimate objective, the civilian casualties were unfortunately high but do not appear to be clearly disproportionate . . .").

[107] *Id.* at ¶ 78.

[108] Final Report to the Prosecutor by the Committee Established to Review the NATO Bombing Campaign Against the Federal Republic of Yugoslavia, ¶ 75, Jun. 13, 2000.

prong to require a "concrete and direct" military advantage that is substantial and realized relatively close in time to the attack.[109] A structure or object determined to be a legitimate military object *may* be the object of an attack if both the first and second prongs of Article 52(2) are met; however, an attack on a legitimate military objective would still be illegal if the harm to the civilian population resulting from the attack "would be *excessive* in relation to the concrete and direct advantage anticipated."[110]

Accordingly, the concrete and direct advantage determined when evaluating whether the object is a military objective must be compared to any expected loss of civilian life, serious bodily injury to civilians, or damage to civilian property to determine if an attack on the military objective would be proportional.[111] In other words, determining that a dual-use object is a military objective is necessary, but not sufficient, in determining whether an attack on the object would be lawful. Interestingly, however, in the context of dual-use objects and structures, the general consensus by militaries around the world is that the whole object or structure either is a military objective, or it is not; there is not a special legal designation for the object or structure simply because it is dual-use.[112] Consequently, any damage to the dual-use object or structure itself that *is* a legitimate military objective need *not* be considered as part of the proportionality analysis, even if such damage impacts the civilian portions of the object or structure.[113]

B. Propaganda in Armed Conflict Under International Criminal Law

From a slightly different perspective, international tribunals have not always found propaganda disseminators in armed conflicts free from

[109] International Committee of the Red Cross, Commentary of 1987 to Additional Protocol I, ¶ 2209.

[110] Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I) art. 51(5)(b), Jun. 8, 1977 (emphasis added).

[111] *See* Final Report to the Prosecutor by the Committee Established to Review the NATO Bombing Campaign Against the Federal Republic of Yugoslavia, ¶ 76, Jun. 13, 2000.

[112] *See, e.g.*, DEP'T OF DEF. LAW OF WAR MANUAL ¶ 5.6.1.2 (2016).

[113] *But see* Michael N. Schmitt, *Targeting Dual-Use Structures: An Alternative Interpretation*, LIEBER INST. WEST POINT (Jun. 28, 2021), https://lieber.westpoint.edu/targeting-dual-use-structures-alternative/ (describing the ICRC's alternative perspective that, although an entire structure or object is a military objective even if it has dual-uses, the dual-use nature of the structure or object should be considered when performing a proportionality analysis, in that the attack's effect "on the civilian part or component of the object . . . or on the simultaneous civilian use or function of the object" should be considered).

criminal responsibility under international law. In the *Media* cases following the Rwandan genocide, for instance, the Appeals Chamber of the International Criminal Tribunal for Rwanda (ICTR) upheld a guilty verdict against the owner, founder, and editor of a magazine, Hassan Ngeze, for directly and publicly inciting the commission of genocide.[114] Under a totality of the evidence standard, the Appeals Chamber found that several articles and editorials published by the magazine could be reasonably attributable to Ngeze himself, and that he had intended, by publishing them, to instigate others to commit genocide.[115]

In reaching this conclusion, the Appeals Chamber found that it was not necessary to find that the genocide would not have occurred but for Ngeze's publications, nor was it necessary to show that the articles were published at exactly the same time as the genocide.[116] The Appeals Chamber also found that a conviction for directly and publicly inciting the commission of genocide did not require its disseminator to make a direct or explicit call to commit genocide, but instead could result from the disseminator making indirect statements about the need to fight and kill the protected group in "self-defense" to eliminate any "danger" they were alleged to pose.[117]

Notably, the Appeals Chamber cited several principles relied upon by the Trial Chamber that could act as "broad guidelines" to distinguish hate speech in the media from a direct and public incitement to commit genocide.[118] For instance, although the speech at issue may not itself produce a direct effect on the behavior of others, it may still be considered a direct and public incitement if it singles out an entire ethnic group by using fear tactics to call for violence against the group, instilling an "us versus them" mentality.[119] Additionally, it noted that it may be necessary to consider whether the goal of the speech is lawful, as opposed to speech with a goal of illegally bringing harm others or, in the case of genocide, an entire group of people.[120]

[114] Prosecutor v. Ferdinand Nahimana et al., Case No. ICTR-99-52-A, Appeals Chamber Judgment ¶ ¶ 885, 886 (Nov. 28, 2007).
[115] *Id.* at ¶ 886.
[116] *Id.* at ¶ 766.
[117] *Id.* at ¶¶ 767-68; *see also id.* at ¶¶ 771-73 (describing articles published by Ngeze as constituting direct and public incitement to commit genocide against the Tutsi when such articles called for the Hutus to rise up to "exterminate" the Tutsi because of the need for the "majority people" to defend themselves against alleged threats posed by the Tutsi).
[118] *Id.* at ¶¶ 694-95.
[119] *See id.* at ¶¶ 695-96.
[120] *See id.* at ¶¶ 694.

Returning to the military objective analysis performed by the ICTY's experts,[121] the principles affirmed by the Appeals Chamber of the ICTR specifically relating to the incitement of genocide may be more generally considered in the context of what constitutes incitement of violence by the media as a whole. Both the Final Report to the Prosecutor in the NATO bombing case and the *Media* cases stand for the proposition that hate speech and propaganda by a media organization may not always rise to the level of incitement of violence. However, there remains the possibility that, when it *does* rise to that level, as was the result in the *Media* cases, the source of the dissemination *may* be justifiably destroyed as a legitimate military objective, if destroying the source of the incitement of such violence would provide the attacker with a "concrete and direct" military advantage.[122]

C. Legal Distinctions Between Propaganda of States Versus Private Actors

Existing treaty provisions provide further context to the legality of the dissemination of propaganda in an armed conflict by state actors. The UN Charter provides that, "[m]embers shall refrain in their international relations from the threat or use of force against the territorial integrity or political independence of any state."[123] Historically, prior to the widespread adoption of a propaganda-specific treaty provision in the International Covenant on Civil and Political Rights (ICCPR)[124] and the *Media* cases from the ICTR, some scholars used this provision to argue that incitement by propaganda to commit an international crime was itself a crime.[125] Stemming from the widely recognized "principle of incitement" and the Nuremberg trials, this argument theorizes that propaganda disseminated by a state is illegal if it is intended to incite the start of an illegitimate armed conflict or further an aggressive war,[126] as is undoubtedly the case in any propaganda that incited or continues to incite the Russian invasion of Ukraine. Relatedly, international customary law has developed to prohibit

[121] *See supra* section III.A.ii.
[122] *See* Final Report to the Prosecutor by the Committee Established to Review the NATO Bombing Campaign Against the Federal Republic of Yugoslavia, ¶ 76, Jun. 13, 2000.
[123] U.N. Charter art. 2, ¶ 4.
[124] International Covenant on Civil and Political Rights art. 20(1), Dec. 16, 1966, 92 T.I.A.S. 908, 999 U.N.T.S. 171 ("Any propaganda for war shall be prohibited by law."). The United States has entered a reservation to Article 20, allowing it to deny any obligation that would restrict individual First Amendment freedoms.
[125] *See, e.g.*, Arthur Larson, *The Present Status Of Propaganda In International Law*, 31 L. & Contemporary Problems 439, 443-445 (1966).
[126] *Id.* at 444.

one state from using subversive propaganda to overthrow the government of another state during a time of peace.[127]

Notably, although international law generally does not bind non-state actors or hold states responsible for propaganda disseminated by private citizens or corporations, some key exceptions may apply.[128] For instance, states may enter into treaties that require them to assume more responsibility for the conduct of individuals than is customarily required by international law.[129] Additionally, some scholars have historically argued that a state may bear responsible for a private entity's propaganda if dissemination of such rises to the level of terrorist activity, is a preparation of an attack by a group of citizens against another state, or would be considered defamation against foreign diplomats.[130]

Most broadly, particularly when considered in the age of the Internet and generative AI, it has been argued that a state has a duty under international law to prevent their territory from being used by private entities to disseminate "radio signals" broadcasting inciteful, subversive, or defamatory propaganda.[131] Although this argument may have held more weight in the past, such a responsibility in today's world would open up the United States to the risk of immense liability under international law, given the sheer number of data centers located within in the U.S. as compared to other states[132] and the impossibility of controlling every given datum that passes through them as they traverse the Internet.

III. ARGUMENTS AND RECOMMENDATIONS

Given the unprecedented properties of generative chatbots that, heretofore, have not been characteristic of any other form of propaganda,[133] several novel legal issues arise under the existing framework of law of armed conflict. First, could the data centers hosting these generative chatbots become legitimate military objectives? Second, if the data centers were considered legitimate military objectives, what would a proportionality analysis prior to launching an attack entail, given their potentially many civilian applications? Third, who should bear

[127] *Id.* at 445.

[128] *Id.* at 449.

[129] *Id.* at 450.

[130] *Id.*

[131] *Id.*

[132] Petroc Taylor, *Number of data centers worldwide 2022, by country*, STATISTA (Feb. 10, 2023), https://www.statista.com/statistics/1228433/data-centers-worldwide-by-country/. The United States has 2,701 data centers, more than any other country in the world. Germany places second, with only 487 data centers.

[133] *See supra* sections II.A. & C.

responsibility for any international crimes stemming from the use of a generative chatbot?

A. Data Centers Hosting Generative Chatbots as Military Objectives

Data centers hosting generative chatbots can likely become legitimate military objectives, even if they are commercially owned and have many civilian purposes, if the generative chatbot itself were also a legitimate military objective.[134] Although practically speaking, these data centers would have both civilian and military applications, the general consensus is that a "dual-use structure" is either a military objective subject to possible attacks, or a civilian object with accorded civilian protections, but cannot be both.[135] In determining whether the data centers are a legitimate military objective, a potential attacker would need to consider (1) whether the data centers, as equipped with the chatbot, provide an "effective contribution to [an enemy's] military action" *and* (2) whether destruction of the chatbot through destruction of the data centers would result in a "definite military advantage" for the attacker.[136]

i. *Generative chatbots provide an effective contribution to military action, turning a data center storing or hosting the chatbot into a structure making an effective contribution to military action.*

History shows how propaganda can make an "effective contribution to military action" during an armed conflict, as is required by the first prong of Article 52(2) in Protocol I. For instance, efforts by the American

[134] In the case of generative chatbots, which are inextricably linked to the data centers on which they are hosted, if an attacker seeks to make the chatbot a military objective, it necessarily must make the data centers on which they are hosted military objectives as well. A phone with the chatbot application downloaded onto it is roughly analogous to the broadcast of propaganda on a viewer's television, and the data centers around the world that store and host the chatbot are roughly analogous to the broadcasting station. Similar to the NATO bombing of the Serbian media center, a potential attacker wishing to target the chatbot would need to target the heart of the chatbot itself by taking out the data centers and would thus have to consider whether the data centers, *as equipped* with the propaganda-disseminating chatbot, are legitimate military objectives.

[135] Michael N. Schmitt, *Targeting Dual-Use Structures: An Alternative Interpretation*, LIEBER INST. WEST POINT (Jun. 28, 2021), https://lieber.westpoint.edu/targeting-dual-use-structures-alternative/.

[136] *See* Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I) art. 52(2), Jun. 8, 1977 (emphasis added) (requiring that the object both "make an effective contribution to military action *and* whose total or partial destruction . . . offers a definite military advantage").

government to distribute physical leaflets in Germany during World War II cannot be chalked up to a mere helpful attempt to aid Germans in understanding the horrors of the Nazi regime. Instead, the American government's underlying intent must be considered in the context of historical statements by Nazi leaders expressing their frustration that such propaganda had "an uncomfortably noticeable effect on the German people."[137] At a minimum, such effects on the German people would have provided an "effective contribution to military action," by lowering morale and diminishing the public's fighting spirits.

In a modern context, Russia's Internet-wide launch of its disinformation campaign to portray Western societies as on the brink of collapse,[138] complete with a deepfake of President Zelensky instructing Ukrainians to surrender,[139] illustrates the lengths an autocratic government is willing to go to demoralize, intimidate, and psychologically terrorize the citizenry of its enemy with the hopes of gaining easier military wins. Propaganda dispersed with a generative chatbot would provide similar "effective contributions" to a state's military actions, but on an even greater scale. Generative chatbots have a limitless ability to learn massive amounts of data that could be tailored to a military's policies. Additionally, they can be trained to generate unique and persuasive responses that could be customized to each individual reader to more readily trick the reader's psyche into believing everything they read.[140]

Similar to Internet propaganda, generative chatbots are highly efficient in that they can widely spread misinformation in mere seconds to an entire population of people. They can also be made available quickly and with little effort to any person with access to the Internet, without the need to distribute physical resources like leaflets.[141] If Russia were to turn its disinformation campaign into a downloadable cell phone application that allowed its users to "chat" with Putin, there would be no limit to the contributions it could make toward furthering Russia's military actions.[142] Thus, the first prong of Article 52(2) in Protocol I would probably be easily satisfied when considering the chatbot in isolation from the data centers. A

[137] *See The Man Behind Hitler: World War II Propaganda*, PBS, https://www.pbs.org/wgbh/americanexperience/features/goebbels-propaganda/ (last visited Mar. 23, 2023).

[138] Press Release, *supra* note 76.

[139] Bergengruen, *supra* note 76.

[140] *See* Goldstein et al., *supra* note 23, at 1-2.

[141] *See id.*, at 4.

[142] Note, however, that the principle of incitement long recognized by international law, as well as case law coming out of the Nuremberg trials, would likely render such propaganda dissemination illegal, since Russia would be using it to further incite an aggressive war. *See supra* section III.C.

generative chatbot can be used not only by civilians to complete basic tasks but, if turned into a tool to contribute to military actions, would acquire features characteristic of a legitimate military objective.

An issue would potentially arise, however, when considering that the data centers hosting the chatbot would not *only* be used for hosting the chatbot, but also for hosting millions of other civilian websites available through the Internet. Putting into context a concern originally voiced by the ICRC,[143] storing a single chatbot that provides an effective contribution to military action in an otherwise blameless data center would turn a largely civilian structure into a military objective, allowing a military to target the entire data center without having to consider the damage caused to the civilian portion when conducting a proportionality analysis.[144] For instance, ChatGPT is hosted on Microsoft's Azure network, which is comprised of data centers located all over the world and that operates as "one of the largest backbone networks in the world."[145] Thus, technically speaking, it would be difficult to argue that the *entire* data center, or even a significant portion of it, provides an effective contribution to military action, when the greater proportion of data traversing the data center would likely be for standard civilian purposes.

On the other hand, the nature of data centers would make it impossible for any military wishing to attack the chatbot to determine precisely *which* area within the data center should be attacked to target the chatbot only, and *how much* of any given data center is being used to house the chatbot.[146] In the context of a structure like an apartment building, it

[143] LAURENT GISEL, INTERNATIONAL COMMITTEE OF THE RED CROSS, THE PRINCIPLE OF PROPORTIONALITY IN THE RULES GOVERNING THE CONDUCT OF HOSTILITIES UNDER INTERNATIONAL HUMANITARIAN LAW 37 (2016) (noting that the perspective that harm to the civilian portions of a dual-use military objective need not be considered in a military's proportionality analysis means "that if a fairly minor military use has turned a civilian object into a military objective (assuming that it fulfills the definition of Article 52 AP I), the damage caused to the remaining civilian part - however important it is - would have no bearing on the decision to launch an attack").

[144] *But see* discussion *infra* section IV.B.

[145] Mary Zhang, *ChatGPT and OpenAI's use of Azure's Cloud Infrastructure*, DGTL INFRA (Jan. 26, 2023), https://dgtlinfra.com/chatgpt-openai-azure-cloud/; *Microsoft global network*, MICROSOFT (Apr. 6, 2023), https://learn.microsoft.com/en-us/azure/networking/microsoft-global-network.

[146] This argument could be rebutted if a military were to conduct a cyberattack *only* on the servers used to host the chatbot. However, an attack with physical weapons, even with something like a precision-guided munition, would likely damage at least part of the data center responsible for civilian Internet traffic. *See* NATHAN J. LUCAS, CONG. RSCH. SERV., IF11353, DEFENSE PRIMER: U.S. PRECISION-GUIDED

may be logical to argue that the principle of distinction requires an attacker to attempt to separate the civilian portion of the structure from the portion used to effectively contribute to military action.[147] However, when considering a data center filled with servers that host both a propaganda-generating chatbot as well as millions of other sites, it would be nearly impossible to argue that a military could physically target *only* the servers hosting the chatbot. Additionally, international law could severely limit well-intentioned militaries if it developed into a system that allowed a data center to be designated as a legitimate military objective *only* if the military could target, with near certainty, the sole area of the data center used to contribute to military action. Consequently, for practicality purposes in designating a legitimate military objective, a military seeking to physically attack the chatbot would likely have to assume that the data center as a whole would be making an effective contribution to military action, without surgically separating it into segments where the chatbot is hosted.[148]

ii. *Destruction of generative chatbots by destroying data centers storing or hosting the chatbot would likely provide the attacker with a direct and concrete military advantage.*

Assuming the first prong of Article 52(2) in Protocol I can be met, the data centers would only become legitimate military objectives if the second prong were also met, which would require destruction of the data centers hosting the generative chatbot to provide the *attacker* with a "definite military advantage." Unlike the first prong, the second prong does not ask whether the potential military objective itself contributes to an enemy's military actions in some way, but rather whether destruction of the potential military objective would provide the *attacker* with a "definite military advantage."[149] The standard for determining whether an attack would provide a "definite military advantage" is rather high for targeting sources of propaganda, as exhibited by the Final Report to the Prosecutor in the NATO Bombing. Attacks may not be conducted if they would "only offer[] potential or indeterminate advantages" and instead must provide

MUNITIONS 1 (2022) (describing the weapon's accuracy as being about three meters).

[147] Dunlap, *supra* note 26.

[148] Perhaps international law could strive to incorporate damage to the civilian portions of a data center into the proportionality analysis, as advocated for by the ICRC. *See* discussion *infra* section IV.B.

[149] *See generally* Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I) art. 52(2), Jun. 8, 1977; Schmitt, *supra* note 126.

advantages that are “concrete and direct,” substantial, and will be realized in the short term.[150]

To support an argument that attacking the data centers provided a “concrete and direct” advantage, the attacker would need to distinguish the propaganda disseminated by the chatbot from propaganda disseminated by a television broadcast. The ICTY’s experts found that the propaganda disseminated by a television broadcast was helpful in demoralizing the general population and undermining support for the enemy government.[151] Nevertheless, the ICTY’s experts ultimately found it “unlikely that either of these purposes would offer the ‘concrete and direct’ military advantage necessary to make [the broadcasting facilities and networks] a legitimate military objective.”[152] However, given the unique properties of a generative chatbot,[153] distinguishing the advantages stemming from destroying a broadcast station versus those stemming from destroying a semi-intelligent chatbot with no moral principles may not be hard to make.

First, unlike a television broadcast where information is disseminated by a human reporter, a generative chatbot lacks true intelligence and thus does not have the same capabilities to understand human morals and philosophies.[154] In short, there is no telling what a generative chatbot may encourage a receiver of its outputs to do, which could include inciting a user, or population of users, to commit war crimes. A television broadcast filled with hate speech for the enemy or idolizations of a state’s military leaders could be significantly different from instructions by a chatbot directing a user on how to best make weapons or how to take to the streets with such weapons to exterminate an entire population of people.[155] Additionally, because there are currently few controls on chatbots that cannot be easily navigated by the chatbot itself,[156] there may be nothing to stop such incitements other than to destroy the bot itself, stored within the data centers.

Second, a “propaganda machinery” composed of a series of television broadcasts viewable on televisions in the 1990s is fundamentally different from “propaganda machinery” composed of a chatbot, available at

[150] *See* Final Report to the Prosecutor by the Committee Established to Review the NATO Bombing Campaign Against the Federal Republic of Yugoslavia, ¶ 76, Jun. 13, 2000.
[151] *Id.*
[152] *Id.*
[153] *See* discussion *infra* section II.A. & C.
[154] *See* Chomsky, Roberts & Watumull, *supra* note 15.
[155] *See* Prosecutor v. Ferdinand Nahimana et al., Case No. ICTR-99-52-A, Appeals Chamber Judgment ¶ ¶ 694-95, (Nov. 28, 2007) (affirming broad guidelines to distinguish incitement to commit genocide from pure hate speech).
[156] *See* Johnson, *supra* note 67.

all hours of the day to answer *any* question asked of it to any person who has access to a cell phone and the Internet.[157] Access to such a chatbot is akin to accessing a country's military leaders themselves, at any time for a Q & A session, given that a chatbot can be "taught" to report to a user virtually anything its creator wishes, can provide follow-up responses to follow-up questions, and can "understand" exactly the type of word choice and tone it should use to sound persuasive and convincing. In sum, destruction of a chatbot would not only demoralize the population and armed forces or aid in undermining the government's political support; it would also prevent an opposing military from having to fight against an infinite number of their enemy's military leaders constantly whispering lies, exaggerations, and unrestricted incitements of violence into the ears of any person with access to a phone. This military advantage would not be insubstantial, merely potential, or realized only in the long term, but would instead put both parties to the armed conflict on equal footing.

Destroying a chatbot may not mean providing an attacker with a dismantled military communications network, although one could argue that destroying data centers in a zone of armed conflict that coincidentally host a propaganda-generating chatbot would also contribute to disrupting the enemy's network for communicating, as in the case of the NATO bombing. Nevertheless, given the fundamental differences between television broadcasts and the inciteful outputs that could (and inevitably would) result from a generative chatbot, destruction of a data center hosting such a chatbot should be considered to provide its attacker with a sufficient "direct and concrete military advantage" to enable the data centers to become legitimate military objectives, assuming the first prong of Article 52(2) can also be satisfied.

B. Proportionality of Potential Attacks on Data Centers Hosting Generative Chatbots

Assuming an attacking military could overcome any hurdles necessary to deeming data centers hosting generative chatbots legitimate military objectives, any actual attack on the data centers would still be subject to application of the proportionality principle. The principal of proportionality requires the attacker to determine that any incidental loss of civilian life, injury to civilians, and damage to civilian objects will not be excessive in relation to concrete and definite military advantage anticipated

[157] For reference, in 2022, nearly 77% of Ukraine's population had access to a smartphone. *Smartphone user penetration rate in Ukraine 2018-2027*, STATISTA (Dec. 1, 2022), https://www.statista.com/statistics/1134646/predicted-smartphone-user-penetration-rate-in-ukraine/.

in destruction of the object.[158] The general goal of this principle is to prevent a state from seeking a military advantage that causes an excessive loss of civilian life or property.[159] Put another way, even if a data center was properly classified as a military objective under Article 52(2), and its destruction would yield a direct and concrete military advantage, there is nevertheless an obligation to refrain from attacking it if the incidental damage to civilians was excessive.[160]

Under the majority approach that is agreed upon by many militaries around the world,[161] damage resulting to other servers in the data center hosting purely civilian websites and Internet traffic would not need to be considered in the proportionality analysis to comply with international law. That is, if the data center were deemed a legitimate military objective under Article 52(2), the resulting proportionality analysis may consider damage that would result to nearby civilians, civilian buildings or houses, but it would not need to factor in any damage acquired by servers located within the attacked data center that were used for non-military purposes. However, the dual-use nature of commercial data centers storing a propaganda-generating chatbot arises in the proportionality analysis if considered under the ICRC's minority, albeit well-intended, approach. Under this minority approach, the attacker would have to consider "the impact of the attack on the civilian part or component" of the data centers, "or on the simultaneous civilian use or function" of the data centers.[162] Undoubtedly, calculating these harms would tip the scales in favor of finding an attack to be disproportionate, unless the anticipated military advantages were highly substantial or the use of the data centers by civilians were only minimal.

Nevertheless, even if this more conservative approach were followed, the anticipated military advantage gained from an attack destroying a generative chatbot may be substantial enough to favor proportionality. For instance, suppose a chatbot actively worked to incite the extermination of a minority group, and provided civilians with concrete directions on how to make homemade weapons to quickly exterminate the entire group. Further, suppose citizens from the majority group were actively acting on these directions. One would be hard-pressed to argue that destruction of the data centers hosting the chatbot would not provide enough of a definite and concrete military advantage to warrant a finding of proportionality, even if such destruction meant that many of the servers in

[158] Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I) art. 51(5)(b), Jun. 8, 1977.

[159] *Kupreškić and Others*, Case No. IT-95-16-T, Trial Chamber Opinion, ¶ 524 (Int'l Crim. Trib. for the Former Yugoslavia Jan. 14, 2000).

[160] DoD Law of War Manual ¶ 5.2.2.

[161] Schmitt, *supra* note 126.

[162] Gisel, *supra* note 143.

the data center allowing for civilian use of the Internet were incidentally destroyed in the attack.

In the very specific case of data center destruction to destroy a propaganda-generating chatbot, though, perhaps more focus should be placed on creating international law that would require a military to consider the indirect, reverberating, or remote harms that would result from destruction of the civilian servers located in the attacked data center. This approach may differ from the indirect, reverberating, or remote harms considered prior to attacks on other forms of dual-use infrastructure, such as electrical grids.[163] Foremost, data centers are unique from other forms of infrastructure, in that they each aid in forming part of the intertangled web of networks that comprise the Internet. Additionally, unlike infrastructure like an electrical grid or train system that is often specific to a given country,[164] the Internet is a shockingly fragile system that is relied upon by the entire world. Notably, it would not take much to obliterate the Internet, and destroying only a few data centers could be enough to change the world as it is currently known.[165] Thus, the harms resulting from annihilation of the Internet may not be the typical immediate and direct "harm" to civilians or civilian "objects" that is normally considered in a proportionality analysis; however, the extreme and widespread downstream results that would occur could be catastrophic, affecting the safety of every part of civilian life, from transportation systems to hospital functions.[166]

C. Placing Responsibility for Generative Chatbots

The question remains: given the lack of international law specific to AI, as well as the general unwillingness to apply international law to private actors, who would remain liable for crimes committed by a generative chatbot? After all, in the *Media* case, Ngeze was found guilty of instigating other to commit genocide when the ICTR found that several articles and editorials published by the magazine could be reasonably attributable to

[163] *See* Dunlap, *supra* note 26 (arguing that several of the remote harms or indirect losses associated with an attack on an electrical grid would not meet the standard for harms that should be considered in determining whether an attack would be proportionate).

[164] *See id.*

[165] *See* Bill Kleyman, *How the Internet May Be Taken Down*, DATA CTR. KNOWLEDGE (Aug. 29, 2014), https://www.datacenterknowledge.com/archives/2014/08/29/internet-may-taken (describing the ease with which one could obliterate the entire Internet simply by severing the main fiber optic cables running along the ocean floor and destroying only a few key data centers).

[166] Robin Layton, *If the internet went down for a day, what's the worst that could happen?* ALLCONNECT (Jul. 20, 2022), https://www.allconnect.com/blog/what-would-happen-if-internet-down-for-day.

Ngeze himself.[167] However, with a generative chatbot, the outputs are not necessarily attributable to any single person, but rather to a group of civilian actors, who may be working for a private corporation to develop algorithms to guide the chatbot's outputs, but who are not, themselves, directly writing the chatbot's outputs.

The answer to this question would rely on several factors, none of which can likely be predicted *ex ante* given the many potential uses of generative chatbots by private actors and governments alike, as well as the wide accessibility to the technologies needed to create such chatbots.[168] For instance, if the generative chatbot were developed directly or through a contract with a state's government or military, the argument for holding its developers responsible under international law as government agents would be much stronger than if the chatbot were unsolicited by any government but developed by a group of private actors for the purpose of furthering their beliefs that aligned with those of their government or military.[169] In the latter case, it may be difficult to hold even the government of the private actors responsible if the state were not overall controlling such activity.[170] Thus, given the ease with which a dangerous generative chatbot could be constructed, not only by government actors but also by private citizens, policymakers should consider working to create international laws that would require states to create their own domestic regulations in accordance with internationally agreed-upon standards, instead of solely creating liability for governments themselves.[171]

[167] Prosecutor v. Ferdinand Nahimana et al., Case No. ICTR-99-52-A, Appeals Chamber Judgment ¶ ¶ 885, 886 (Nov. 28, 2007).

[168] *See* Thomas Hansen, *How to Create your own ChatGPT(ish) in 5 minutes*, DEV (Feb. 3, 2023), https://dev.to/polterguy/how-to-create-your-own-chatgpt-ish-in-5-minutes-425g (describing the ease with which a developer can develop their own generative chatbot).

[169] *See* Larson, *supra* note 125, at 128 (describing complications of using international law to hold governments accountable for actions by private citizens).

[170] *See* Prosecutor v. Tadić, Case No. IT-94-1-A, Appeals Chamber Judgment ¶ ¶118-120, 145 (Jul.15, 1999) (finding international law to require a state to exert at least overall control over the activities of private actors for the state to be responsible).

[171] Although now less powerful after the Supreme Court's decision in *Nestle v. Doe*, 141 S. Ct. 1931 (2021), something like a modified version of the domestic Alien Tort Statute that previously was used to create extraterritorial liability for private corporations could be useful to capture the conduct of private entities that would be unreachable by international law.

CONCLUSION

Ideally, a treaty on AI-equipped, dual-use technologies would be adopted by the world's key actors. The problem, though, is that this world is not ideal.[172] Optimistically, international policymakers can begin taking small steps toward regulating AI on a more general level with customary international law, with the goal of grouping regulation of generative chatbots with the regulation of other AI-equipped, dual-use technologies capable of disseminating propaganda, like deepfakes. Because this field of technology is progressing so rapidly, what becomes "custom" for technologies now may not be "custom" in a few years or even a few months.[173] However, a least common denominator type of approach could be useful, which would require policymakers to pay particular attention to the general features shared by *all* forms of AI to govern how *those* common features are used in armed conflict. The political landscape at the international level will certainly be a tangled web of differing views and concerns,[174] but creating a baseline consensus shared by all world powers with AI capabilities will be key when moving forward to begin the development of international law, even if a full treaty is currently out of reach.

[172] For a particularly dire view on the future use of chatbots given the current state of the world, see Eliezer Yudkowsky, *Pausing AI Developments Isn't Enough. We Need to Shut it All Down*, TIME (Mar. 29, 2023, 6:01 PM), https://time.com/6266923/ai-eliezer-yudkowsky-open-letter-not-enough/.

[173] *See* Jeremie Harris, *AI advances, but can the law keep up*, MEDIUM (Mar. 31, 2021), https://towardsdatascience.com/ai-advances-but-cat-the-law-keep-up-7d9669ce9a3d (describing the need for flexible laws with dynamic language to keep up with the rapid evolution of AI).

[174] *See* Mark Scott, *AI's pandemonium leaves global leaders scrambling*, POLITICO (Apr. 20, 2023, 2:05 PM), https://www.politico.com/news/2023/04/20/global-confusion-new-ai-rules-00093074 (describing the maze of questions facing international policymakers as they attempt to find international consensus on the development and use of the latest AI technologies, including chatbots like Chat-GPT).