

Article

The AI Carbon Footprint and Responsibilities of AI Scientists

Guglielmo Tamburrini

DIETI, Department of Electrical Engineering and Information Technology, University of Naples Federico II, Via Claudio 21, 80125 Napoli, Italy; guglielmo.tamburrini@unina.it

Abstract: This article examines ethical implications of the growing AI carbon footprint, focusing on the fair distribution of prospective responsibilities among groups of involved actors. First, major groups of involved actors are identified, including AI scientists, AI industry, and AI infrastructure providers, from datacenters to electrical energy suppliers. Second, responsibilities of AI scientists concerning climate warming mitigation actions are disentangled from responsibilities of other involved actors. Third, to implement these responsibilities nudging interventions are suggested, leveraging on AI competitive games which would prize research combining better system accuracy with greater computational and energy efficiency. Finally, in addition to the AI carbon footprint, it is argued that another ethical issue with a genuinely global dimension is now emerging in the AI ethics agenda. This issue concerns the threats that AI-powered cyberweapons pose to the digital command, control, and communication infrastructure of nuclear weapons systems.

Keywords: AI ethics; AI carbon footprint; environmental responsibilities of AI scientists; global issues in AI ethics



Citation: Tamburrini, G. The AI Carbon Footprint and Responsibilities of AI Scientists. *Philosophies* **2022**, *7*, 4. <https://doi.org/10.3390/philosophies7010004>

Academic Editor: Peter Boltuc

Received: 10 November 2021

Accepted: 31 December 2021

Published: 5 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

There is hardly any aspect of human life that artificial intelligence (AI) has not changed or may not change in the near future. The pervasive impact of AI over the last decade has been powered by machine learning (ML) technologies in general, and deep neural networks (DNN) in particular. Learning AI systems are having an increasing role to play in commerce, industry, finance, the management of public and private services, communication and entertainment, security, and defense.

The increasing pervasiveness of AI technologies and systems drives a growing apprehension about the carbon footprint of AI models obtained by means of ML techniques, and their potential impact on climate warming. This article focuses on the little explored ethical issue of distributing fairly among involved actors prospective responsibilities concerning the growing AI carbon footprint. And it builds on various contributions to understanding sources of the AI carbon footprint and related mitigating actions [1–6], which are here used as epistemic starting points for the ethical analysis of involved responsibilities and the outline of corresponding ethical policies.

In Section 2, the significance of the ethical issue addressed here is emphasized by reference to whistleblowing estimates of the carbon footprint of selected AI systems and corresponding calls for more systematic evaluations. In Section 3, major involved actors are identified—including AI scientists, AI industry, AI infrastructure providers—and the problem of pinpointing their environmental responsibilities is introduced. In Section 4, distinctive responsibilities of AI scientists in the way of climate warming mitigation actions are disentangled from responsibilities of other involved actors. In Section 5, nudging interventions are suggested to implement these prospective responsibilities. These interventions presuppose modifying the very idea of what counts as a “good” result in AI, moving on from the goal of improving accuracy to the goal of pursuing the accuracy of AI systems in combination with greater computational and energy efficiency. The corresponding good practices that are suggested leverage on the time-honored AI tradition of research pursued

in the framework of competitive games. In Section 6, it is pointed out that greenhouse gases emissions have climate effects anywhere and everywhere on the planet, impacting most or even all members of present generations, nations, future generations, and the rest of nature. It is argued on this ground that climate warming ethical issues, in view of their genuinely global reach, take up a rather unique position in the AI ethics agenda, which has been mostly concerned with issues of a more local character. In Section 7, on the basis of the local–global distinction concerning the AI ethics agenda, it is argued that another ethical issue with a genuinely global dimension is emerging there, which brings about moral responsibilities of AI scientists that are akin to those carried by physicists in connection with the development of nuclear weapons. This novel ethical issue concerns maturing cyberweapons powered by AI technologies and the threats that these cyberweapons are likely to pose to the digital command, control, and communication infrastructure of nuclear weapons systems. Section 8 concludes.

2. AI Ethics and Estimates of the AI Carbon Footprint

Widespread attention to the growing environmental impact of AI and its carbon footprint has been significantly stimulated by estimates of computational and electricity resources that are required to train selected AI models by ML methods. Strubell and co-workers selectively focused in their carbon footprint analysis on AI models for natural language processing (NLP). In addition to the training of a variety of off-the-shelf AI models, they considered downstream training processes that one needs to adapt and fine-tuning these AI models to perform new NLP tasks [1]. Electricity consumption and greenhouse gases (GHG) emissions of these systems were extrapolated from an estimate of computational training costs. Notably, the training of an NLP Transformer model, based on a DNN architecture, was estimated to produce GHG emissions equivalent to those of five average automobiles throughout their lifecycle. The GHG emissions of the BERT_{LARGE} model—which was trained using GPU (Graphic Processing Units) as a specialized hardware accelerator—were estimated to be equivalent to those of a commercial flight between San Francisco and New York [1]. The publication of these estimates had a significant whistle-blowing effect. Behind the ethereal and intangible appearances of AI information processing, its material character and consequences were exposed, raising specific concerns about the environmental impact of the distinctive processes involved in the development of AI systems.

Shortly after the publication of these data, the 2020 White Paper on AI released by the European Commission (EC) called for actions moving beyond the collection of impressive but admittedly anecdotal data about the training of selected AI systems. Indeed, the more general problem was raised there of assessing the carbon footprint of each individual AI system and of the AI sector as a whole: “Given the increasing importance of AI, the environmental impact of AI systems needs to be duly considered throughout their lifecycle and across the entire supply chain, e.g., as regards resource usage for the training of algorithms and the storage of data.” [7] (p. 2). Motivated by the increasing pervasiveness of AI technologies and systems, this recommendation calls for a mapping of the environmental impact of AI systems (a) throughout their lifecycle, that is, from development and introduction to withdrawal, and (b) across the entire supply chain, which involves multiple research and industrial actors, including AI researchers, AI industry, providers of information and communication (IC) infrastructures, electricity suppliers.

The overarching inquiry solicited by the EC is crucial to understand how big the AI carbon footprint is and how significant is the individual contribution of various AI segments, from research to industry, to produce this carbon footprint. It is needed, moreover, to identify effective policies and countervailing actions, aimed at curbing the carbon footprint of the AI sector and to mitigate its impact on climate warming. Last but not least, assessing the AI carbon footprint and identifying its major sources is essential from the special perspective of environmental ethics. This knowledge is needed to identify and fairly distribute environmental responsibilities among the heterogeneous groups of

actors that are involved, from AI researchers and AI industry to providers of information and communication (IC) infrastructures and electricity suppliers. These environmental responsibilities are both prospective and retrospective. Prospective responsibilities concern those who are in the position to act and reduce the current carbon footprint of AI research and business. Retrospective responsibilities concern corresponding omissions and other environmentally irresponsible behaviors arising—all things considered—from the use of AI technologies and systems. Altogether, one can hardly question the importance of estimating the AI carbon footprint and identifying its sources. However, providing realistic estimates presupposes an inventory of the wide variety of relevant factors and the development of suitable carbon footprint metrics and models.

3. AI Carbon Emissions Sources and Related Responsibilities

A relevant factor motivating concerns about present and prospective AI carbon footprints is the steadily growing size of AI learning models that are based on DNN. The size of a DNN is usually measured by reference to the total number of weights associated to the connections between its neural nodes. In the NLP domain, the large version of the BERT model mentioned above contains about 350 million of these parameters, the DNN system GPT-2 (where GPT stands for Generative Pre-trained Transformer) contains about 1.5 billion parameters, and 175 billion is the figure reached by the more recently released GPT-3. This accelerating trend in the development of larger AI models is jointly explained by the increased availability of computing resources, the improved accuracy achieved by larger models and the special emphasis placed by research and industry on enhanced system accuracy. However, improved AI model performances attained using larger models come with increased energy consumption. This approach to achieving increased model accuracy does not go well with the goal of curbing the overall AI carbon footprint.

Another source of environmental concern is the growing size of training sets and the growing number of hyperparameter experiments. The latter enable one to explore how the performances of learning systems change by tuning DNN hyperparameters such as the number of learning cycles and the network learning rate. Again, expansions in both training sets and hyperparameter experiments are usually motivated by the achievement of better DNN model performances. However, expansions of both kinds come with increased energy consumption [2] (p. 58). Thus, major trends that are being observed in AI research and development (R&D)—recourse to hyperparameter experiments, increasing DNN model and training set size—are aimed at achieving increased model accuracy, but do not go well with the goal of curbing the overall AI carbon footprint. From the viewpoint of environmental ethics, an evident tension emerges between entrenched behaviors in the AI research community and climate warming mitigation objectives. Thus, one is led to ask whether it is ethically justified to pursue academic and industrial R&D solely focusing on AI system accuracy, if better model accuracy comes with increased energy costs due to the use of ever larger models, training sets and epochs, massive hyperparameters experiments.

The R&D activity of setting up from scratch or adapting AI models reveals only some facets of the AI carbon footprint problem. Another significant aspect concerns the use of AI systems after training is completed. In connection with the above estimates of computational and electricity resources required to *train* some selected AI models for NLP [1], Patterson and co-workers emphasized that major digital companies—unlike AI academic researchers that are mostly engaged in model development, training, and testing—apply and use AI models that are fully operational for prediction, decision-making, and inference. In fact, up to 90% of computational and energy consumption costs faced by these companies and attributed to AI models have been estimated to flow from their post-training use. Accordingly, to evaluate sensibly the AI carbon footprint, one must carefully attend to fully operational uses of AI models for prediction, decision-making, and inference [3]. From an environmental ethics viewpoint, this finding leads one to distinguish between prospective responsibilities of different groups of involved actors. AI researchers must specifically alleviate tensions that may arise between climate warming mitigation goals

on the one hand, and training and other experimental practices aimed at achieving better AI model accuracy on the other hand. In addition to this, AI companies must attend to and curb the carbon footprint of fully operational AI systems. Actors of both kinds must additionally attend to the computational costs of algorithms and programs used for AI model training and inference.

Additional factors to consider for the purpose of providing credible estimates of the AI carbon footprint comprise the energetic efficiency of the infrastructure formed by the variety of ICT systems supporting AI systems training and use. These factors notably include the energetic efficiency of the processors on which AI algorithms and programs run and the energy supply mix of datacenters and other providers of computing resources [3] (p. 2). Clearly, as one takes into account the broader landscape of AI infrastructures, prospective responsibilities extend beyond AI scientists and industry, reaching larger groups of actors within the ICT sector.

One should be careful to note that the list of relevant factors to consider and their attending responsibilities does not come to an end here. Comprehensive assessments of the AI carbon footprint require an examination of wider interaction layers between AI technologies and society. These interaction layers arise from AI-induced societal changes occurring in work, leisure, and people's consumption patterns. Current estimates point to a pervasive and rapid impact of AI across the spectrum of human activities. However, wider interaction layers between technological developments and societal patterns have proven difficult to encompass in connection with other technologies and systems, and their environmental consequences have been correspondingly difficult to identify and measure [8].

In the light of current expectations about wider interaction layers between AI technologies and society, a “concerted effort by industry and academia”—invoked to attain substantive reductions of the AI environmental cost [1] (p. 5)—appears to be a necessary, but still insufficient step to effectively curb the AI carbon footprint. Moreover, a wide variety of actors must play a role in this concerted effort: AI researchers and engineers, universities and research centers, AI firms and providers of ICT infrastructures. Hence, it may become quite difficult—both in practice and in principle—to set apart which responsibilities pertain to which community of involved actors. For example, AI scientists may point to AI industry pressing need for ever more accurate models as an excusing reason for the use of ever larger models and training sets. In its turn, AI industry may shift the prevalent responsibility burden on electrical energy supply chain actors who are still prevalently relying on fossil fuel sources. Thus, an instance of the many-hands responsibility problem in environmental ethics [9] looms large on AI carbon footprint reduction efforts. It is a bitter and well-known fact that major political negotiations about climate warming mitigation actions have often floundered in similar buck-passing games.

4. Disentangling the Environmental Responsibilities of AI Scientists

Acknowledging that significant stumbling blocks hinder a thorough allocation of responsibilities to reduce the AI carbon footprint does not entail that any such allocation effort is invariably bound to fail. Interestingly, this allocation problem is being debated within the AI research community, and distinctive roles and responsibilities for AI researchers to reduce the AI carbon footprint are being proposed, which disentangle these from roles and responsibilities of other involved actors, including commercial firms relying on already trained and fully operational AI models for inference, prediction, and decision-making, private and public datacenters, providers of cloud computing resources, and electricity producers.

To begin with, one must tackle the problem that the total amount of GHG emissions that one may sensibly attribute to the development or adaptation of some learning AI model (more briefly, for now on, an AI research result) depends on both spatial and temporal factors. The computing activities that one carries out to achieve an AI research result is invariant neither with respect to where the underlying computing activities take place nor with respect to when they take place. Indeed, one may obtain the electrical energy needed

to carry out the required computing activities from various providers, which differ from each other in the way of their electricity supply mix, formed by variable relative proportions of fossil fuel, alternative, and renewable sources. Moreover, the electricity supply mix of each provider may change over time (e.g., at night or during daylight). Accordingly, direct estimates of GHG emissions are unsuitable to draw fair comparisons between the carbon footprint reduction efforts of AI researchers working asynchronously or in distinct locations, and to disentangle corresponding responsibilities from the responsibilities of electric energy producers and suppliers [2].

As an alternative to direct estimates of GHG emissions attributed to each AI result, one may try and use electricity consumption estimates which are blind to the energy supply mix. However, these estimates are in their turn sensitive to the kinds of processors and, more generally, hardware resources that scientists use to train AI models. Accordingly, gross electricity consumption is unsuitable to identify and compare fairly the carbon footprints reduction efforts of AI researchers relying on different hardware resources, and to disentangle corresponding responsibilities from the responsibilities of private or public datacenters.

To overcome the drawbacks of direct measures of either GHG emissions or gross electricity consumption—concerning the disentanglement of AI research responsibilities from those of energy producers and datacenter administrators—it has been suggested that a more suitable metrics should identify the computational efficiency of AI research and its results. In this vein, Schwartz and co-workers proposed AI researchers to report “the total number of floating-point operations (FPO) required to generate a result”, on the grounds that FPO estimates of the amount of work performed by a computational process are agnostic with respect to both the energy supply mix and the energetic efficiency of hardware infrastructures [2] (p. 60). More generally, any sensible measure of computational efficiency to correlate, albeit indirectly so, to the AI carbon footprint would enable the AI research community to identify distinctive responsibilities for climate warming mitigation actions based on the development of computationally more efficient methods and systems.

Unlike the pursuit of increased computational efficiency, additional actions that AI researchers may undertake to curb the AI carbon footprint depend on knowledge of what datacenters administrators, cloud computing providers, electricity producers and other involved actors do. These additional actions include the choice of the more energy-efficient computing hardware and datacenters, the choice of shifting the training of and experimenting with their AI models towards low carbon intensity regions, and the choice of suitable times of the day to train their AI models, insofar as the carbon intensity of any given region may change throughout the day [4]. Software tools are being made available to predict and estimate the carbon footprint of AI results, taking into account computational efficiency, and both energy-efficient uses of datacenters and electrical energy supplies [4–6].

5. Promoting Environmentally Responsible AI Research

Measures of computational efficiency enable one to identify specific responsibilities of AI researchers, and knowledge of what other involved actors do enables them to identify a variety of additional good practices in the way of AI carbon footprint mitigation. However, how is the taking on of these responsibilities and implementation of good practices effectively encouraged?

One approach involves modifying the idea of what counts as a “good” result in AI. The development or the tuning of an AI model which enables one to go significantly beyond present accuracy benchmarks is normally considered a good result in AI, independently of how computationally expensive is to train this model and make the required experiments with it. A significant example of this kind of result appraisal is the above-mentioned transition, in the NLP domain, from the BERT model containing about 350 million parameters, to GPT-2 containing about 1.5 billion parameters, and then to the GPT-3 model containing 175 billion parameters. Environmentally thriftier research across the AI research community may be encouraged by prizing results which combine better system accuracy

with greater computational and energy efficiency. It was suggested in this connection to set up leaderboards listing best results and practices [5].

The identification of benchmarks suitably combining accuracy with energy efficiency provides a basis for “environmentally sustainable AI” competitions, which may leverage on a long tradition of AI research pursued in the frame of competitive games. These tournaments may eventually achieve the reputation of other AI undertakings which took the form of competitions and played prominent roles in the development of AI and its research programs. Major competitions of this sort have revolved around computational chess, Go, and poker. Many Grand Challenges promoted by DARPA have taken the form of competitive games among AI or robotic systems. The RoboCup initiative uses the game of soccer for developing skills and intelligence of robotic systems. Results achieved by teams of robots competing in robotic soccer tournaments are used to identify benchmarks inspiring further research and the development of new generations of robotic soccer teams [10]. By the same token, AI competitions prizing the energetic efficiency of AI systems may foster good practices in AI research and even attract the interest of AI industry, in view of economic advantages flowing from reductions of electricity supply costs. The allure of competitive games may bring about an increased capability to attract students and junior researchers, raising their awareness on AI carbon footprint, and to introduce the study of technological approaches to the reduction of this footprint in computer science graduate programs.

How distant from the current reality of AI research is the scenario of an environmentally virtuous AI research? It is a fact that the search for computationally efficient solutions to research problems is not a prevailing goal in today’s ML research. This much can be gleaned from a random sample of 60 papers presented at recent top-level AI conferences: a large majority of these papers were found to target model accuracy only, without taking computational efficiency into account [2] (p. 56). In another random sample of 100 papers from the 2019 NeurIPS proceedings, one paper only “measured energy in some way, 45 measured runtime in some way, 46 provided the hardware used, 17 provided some measure of computational complexity (e.g., compute-time, FPOs, parameters), and 0 provided carbon metrics.” [5] (p. 6). These findings suggest that improvements in task performance accuracy are pursued without taking notice of environmental costs. This neglect for environmental costs in AI research is often transmitted downstream in university study programs, where AI projects and master theses produce large amounts of minor and practically inconsequential accuracy results which fail to make it into AI conferences and publications. Accordingly, a significant departure from prevalent research and educational goals is required to raise environmental awareness and introduce corresponding good practices in AI research and academic communities.

One may introduce competitive games, research awards and recognitions as nudging interventions to enhance environmental awareness and foster related good practices in AI academic and more broadly research communities. However, one may raise the reasonable doubt that these nudging interventions, if any, will be able to gain sufficient traction, eventually leading to the accomplishment of significant objectives in the way of reducing the AI carbon footprint on a temporal scale which is meaningfully related to the overall goal of limiting global warming in the XXI century to 1.5 °C [11]. If each community is required to do its share to implement this overall goal, one may well consider mandatory environmental policies for AI research as an alternative to nudging interventions and voluntary participation into environmentally virtuous AI competitions. Thus, one may introduce AI-research carbon quotas, in the absence of a swift and widespread endorsement of climate warming mitigation actions by the AI research community. However, clearly, this restrictive policy would impact negatively on research freedom, raising the additional ethical challenge of an equitable distribution of bounded computing resources among AI stakeholders [12]. Moreover, this policy may negatively affect AI’s vast potential to support climate warming mitigation actions in other spheres of human activity, insofar as AI research may substantively contribute to identify approximate solutions to a wide

variety of energy use optimization problems, ranging from the energetic efficiency of buildings to the reduction of transportation needs, the planning of travel routes, and the efficiency of supply chains in industry production and food consumption [13].

To sum up. Applications of AI research promise to drive climate warming mitigation actions across a variety of economic and social activities. At the same time, however, AI research is an integral part of the climate crisis problem. This much is conveyed by recent—admittedly perfectible—estimates of the AI carbon footprint. Suitable measures of computational costs arising from AI research are needed to foster a better understanding of AI's environmental impact and to identify distinctive environmental responsibilities of the AI research community. AI competitions prizing computational efficiency and the establishment of leaderboards may encourage environmentally virtuous research attitudes within the AI research community. However, the need for mandatory policies may emerge too, if the prevailing goal of prizing accuracy in current AI research will not be willingly and timely replaced by more comprehensive goals combining accuracy with energy and computational efficiency.

6. The AI Carbon Footprint and Global Ethical Issues

AI's role is steadily growing in both climate warming and related mitigation efforts. The ethical issues arising from this growing role for AI research and industry concern a truly global phenomenon. Regardless of their source, GHG emissions have climate effects anywhere and everywhere on the planet. The corresponding ethical issues of responsibility and fairness impact individuals, nations, future generations, and the rest of nature [14]. For this reason, climate warming ethical issues are a major novelty to appear in the AI ethics agenda in view of their genuinely global reach. This much can be gleaned by contrast, if one looks at the list of items included in comprehensive documents [15,16] and overarching review articles [17,18] spanning over the wide variety of issues that are now being addressed in AI ethics. To exemplify, consider from this perspective the EU proposal for regulating AI [16]. This influential document classifies as “high risk” from ethical and legal standpoints a wide range of AI application domains, such as access to education, vocational training and employment, the management of migration, asylum and border control, access to essential services, and public benefits. Most of these issues are local, in the sense that an AI system operating in one of these domains raises ethical concerns, at each given point in time, about the good and the fundamental rights of a limited fraction of persons within the human population. Thus, for example, information processing biases possibly embedded into an AI system supporting college admissions procedures may lead to discriminations affecting the good and the rights of rejected college applicants [19], and raises ethically justified concerns about the life plans of other prospective college applicants. Similarly, AI decision-making concerning bank loans, job hirings, career advancement, migration and asylum management, access to unemployment benefits, and other public or private services affect or raise concerns about the good and fundamental rights of limited portions of the humankind at each given point in time.

In contrast with this local character of most ethical issues in the AI ethics agenda, there are ethical issues that are genuinely global, so that the good and the fundamental rights of most—and possibly even of all members of the human species—are involved at least at some given point in time. Notably, pandemic infections like SARS-CoV-2 raise—in addition to formidable medical, economic, and political problems—some genuinely global ethical issues, which concern the physical integrity, life, well-being, right to work, and education of most members of the human species. Indeed, as of December 2021, only a handful of countries—Nauru, Turkmenistan, and Tuvalu, in addition to North Korea—had no reported COVID cases.

The history of humankind is scattered with the waxing and waning of other global ethical issues, in the sense here specified of ethical issues affecting at some given point in time the good and the fundamental rights of most, and possibly all members of the human species. The Spanish flu pandemic posed a global ethical issue back in 1918–1919. The

1976 Rowland–Molina hypothesis identified the major cause of atmospheric ozone layer depletion in the use of chlorofluorocarbons (CFCs). This anthropically induced effect might have deprived the humankind and other living entities of protection from exposure to solar UV radiation, thereby triggering another potentially global ethical issue. Effective international efforts to decrease the use of chlorofluorocarbons (CFCs) from the 1980s onward appear to have successfully curbed this specific global threat.

Additional global ethical issues are presently at stake in connection with both the anthropically induced climate crisis [20] and the threat of a nuclear world war represented by the very existence of nuclear arsenals [21]. The climate crisis has now entered the AI ethics agenda in connection with both GHG emissions attributed to AI systems and AI's potential contribution to the identification of sensible solutions to energy use optimization problems. Some ethically relevant implications of the AI current and potential impact on climate warming have been analyzed in this article. However, what about the threat of a nuclear world war? Are this threat and its ethical implications significantly related to current AI developments? The answer to this question, as we shall presently see, is likely to be a resounding “yes”.

7. AI Cyberweapons and Nuclear Weapons: A Global Ethical Issue on the Rise

Ethical debates on the militarization of AI have been so far mostly concerned with the ethical implications of developing and deploying autonomous weapons systems (AWS). These weapons systems are capable to select and engage targets without further intervention by a human operator after their activation [22,23]. AI technologies play a crucial role in the development of ever more sophisticated AWS, by enabling perceptual, deliberative, and action planning capabilities that an AWS needs to perform the tasks of target selection and attack.

Normative debates about AWS have been basically concerned with local ethical issues. These issues notably concern (i) AWS causing breaches of *jus in bello* norms of just war theory and international humanitarian law (IHL), thereby affecting the rights and the welfare of their victims [24–26]; (ii) the difficulty of selectively attributing responsibilities for IHL breaches to the persons taking on responsibilities and decision-making roles in AWS operation [27,28] (iii) affronts that AWS targeting decisions make to the human dignity of its victims [29]. All of these are local ethical issues according to the distinction introduced in the previous section: only the good and the rights of AWS potential victims are selectively at stake (by (i) and (iii)), in addition to the duties of persons and institutions who are responsible for AWS operation (by (ii)). However, the growing number of tasks that AI systems autonomously perform is giving rise to a new global issue, concerning the impact of AI systems on threats of worldwide nuclear conflicts and its ethical implications.

The connection between the AI ethics agenda and threats of worldwide nuclear conflicts is emerging from the growing role of AI in cyberspace [30] in general, and in cyberconflicts [31] in particular. It was pointed out that “parties to armed conflicts frequently deploy cyber weapons and, recognizing the competitive advantages afforded by autonomy, States are developing—or perhaps have already developed—autonomous cyber weapons for use in armed conflict” [32] (p. 646). AI-powered cyber weapons can in principle use their adaptive intelligence and learning capabilities to identify and exploit without human intervention the software vulnerabilities of other digitalized military systems. Are autonomous cyber weapons (ACW from now on) genuine AWS? What happens to the normative debate about AWS if one counts ACW as some special sort of AWS? Even more important for our present concerns: are the involved ethical issues bound to stay local?

The above mentioned 2012 Directive of the US Department of Defense [22]—which first introduced the functional condition on a weapons system to count as autonomous—leaves aside any consideration of machine autonomy in the cyberspace. At the same time, however, no explicit restriction in terms of warfare domains is introduced there. ACW and their targets inhabit the cyberspace, thereby differing in the way of operational domain from other AWS, including autonomous robotic sentries, loitering munitions, or swarms

of autonomous armed drones. However, they are no different from other AWS in their capability to independently select and attack targets. It is therefore reasonable to conclude that an ACW is a special sort of AWS.

Like cyber weapons operated by humans, ACW can potentially target surveillance and reconnaissance military systems, weapons system requiring software resources to be operated, software systems serving intelligence and command-and-control purposes at military headquarters. By replacing teams of skilled engineers in orchestrating cyber attacks, ACW are likely to accelerate the pace of cyberwarfare beyond human response capabilities, enabling the delivery of cyber attacks on larger scales, and making cyber threats more persistently available in the cyberspace [33]. This possibility aggravates concerns expressed about interactions between AWS which would accelerate the pace of conflict beyond human cognitive capabilities [34]. Indeed, ACW may target AWS releasing their force in traditional warfare domains. Accelerating the pace of cyberwarfare may lead to runaway interactions between ACW and AWS. Moreover, the rise of ACW may aggravate existing cyberthreats to nuclear weapons and related nuclear command, control, and communication (NC3) systems. Cyberattacks directed to nuclear defense systems could lead to false warnings of nuclear attacks by the enemy, disrupt access to information and communication, damage nuclear delivery systems, and even enable the hacking of a nuclear weapon [35]. Therefore, cyberattacks on nuclear defense systems raise new daunting threats for peace and a global ethical issue concerning the very persistence of human civilizations.

The maturing of these technological possibilities has far-reaching ethical implications, involving the responsibilities of AI scientists on account of their privileged epistemic position. Right after World War II, many physicists felt it was their moral obligation to make public opinion and political decision-makers aware of the existential threat posed by nuclear weapons and the nuclear arms race starting during the cold war. Later, chemists and biologists played a pivotal role in international debates and diplomatic efforts leading to international treaties banning the development, production, stockpiling and use of chemical and biological weapons of mass destruction. Today, AI scientists must make public opinion and political decision-makers aware of the threats to peace and stability posed by the maturing of ACW, up to and including their impact on NC3 systems, and the existential threats for human civilization that may emerge from ACW targeting nuclear defense systems. They must face hard moral choices concerning their active participation in or support of ACW research.

8. Conclusions and Future Work

It has been widely emphasized that applications of AI research promise to drive climate warming mitigation actions across a variety of economic and social activities, insofar as AI research may substantively contribute to identify solutions to optimization problems, which range from the energetic efficiency of buildings and transportation needs to the efficiency of supply chains in industry production and food consumption. At the same time, however, it is widely recognized that AI research is an integral part of the climate crisis problem. This is witnessed by recent estimates of the AI carbon footprint, which are here used as epistemic starting points for the ethical analysis of involved responsibilities and the outline of corresponding ethical policies. Clearly, more comprehensive and accurate measures of computational costs arising from AI research are needed to develop a better understanding of AI's environmental impact and to pinpoint the ways in which each one of the involved actors contributes to the AI carbon footprint and to reduce its impact. It was argued here, however, that the admittedly incomplete and imprecise state of knowledge about the AI carbon footprint suffices to disentangle distinctive responsibilities of AI scientists concerning their research activities. At the same time, what is presently known suffices to promote nudging interventions by the AI research community, leveraging on the AI long tradition of pursuing research in the framework of competitive games, and prizing computational efficiency along with accuracy of novel AI systems.

The emergence of the AI carbon footprint problem has motivated the opportunity of introducing, in the final part of this article, the distinction between local and global issues in connection with the AI ethics agenda. The AI carbon footprint raises ethical issues of an unprecedented global reach in the AI ethics agenda. However, the list of global issues in the AI ethics agenda is likely to expand soon in view of AI growing pervasiveness across and within each domain of human activity. Indeed, it was pointed out that another ethical issue with a genuinely global dimension is emerging in view of maturing AI-powered cyberweapons. These jeopardize the integrity and availability of the digital command, control, and communication infrastructure of nuclear weapons systems, thereby posing new threats to nuclear stability. Future work, which goes clearly beyond the scope of this article, will be devoted to identifying and analyzing the responsibilities of AI scientists, emerging in connection with this novel global issue in the AI ethics agenda, and ranging from awareness-raising communication—addressed to the public opinion and political decision-makers alike—to moral choices concerning their active participation in AI cyberweapons R&D activities.

Funding: This research was partially funded by Italian National Research Project PRIN2020, grant 2020SSKZ7R.

Institutional Review Board Statement: Not applicable.

Acknowledgments: The author is grateful to Daniele Amoroso, Fabio Fossa, Giuseppe Trautteur and three anonymous reviewers for their helpful comments and suggestions.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Strubell, E.; Ganesh, A.; McCallum, A. Energy and Policy Considerations for Deep Learning in NLP. *arXiv* **2019**, arXiv:1906.02243.
2. Schwartz, R.; Dodge, J.; Smith, N.A.; Etzioni, O. Green AI. *CACM* **2020**, *63*, 54–63. [[CrossRef](#)]
3. Patterson, D.; Gonzales, J.; Le, Q.; Liang, C.; Mungia, L.M.; Rotchild, D.; So, D.; Texier, M.; Dean, J. Carbon Emissions and Large Neural Network Training. *arXiv* **2020**, arXiv:2104.10350.
4. Anthony, L.F.W.; Kanding, B.; Selvan, R. Carbontracker: Tracking and Predicting the Carbon Footprint of Training Deep Learning Models. *arXiv* **2020**, arXiv:2007.03051.
5. Henderson, P.; Hu, J.; Romoff, J.; Brunskill, E.; Jurafsky, D.; Pineau, J. Towards the Systematic Reporting of the Energy and Carbon Footprints of Machine Learning. *J. Mach. Learn. Res.* **2020**, *21*, 1–43.
6. Lacoste, A.; Luccioni, A.; Schmidt, V.; Dandres, T. Quantifying the Carbon Emissions of Machine Learning. 2019. Available online: <https://deepai.org/publication/quantifying-the-carbon-emissions-of-machine-learning> (accessed on 2 January 2022).
7. European Commission. White Paper on AI. A European Approach to Excellence and Trust. 2020. Available online: https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en (accessed on 2 January 2022).
8. Williams, E. Environmental effects of information and communication technologies. *Nature* **2011**, *479*, 354–358. [[CrossRef](#)] [[PubMed](#)]
9. Van de Poel, I.; Fahlquist, J.N.; Doorn, N.; Zwart, S.; Royakkers, L. The problem of many hands: Climate change as an example. *Sci. Eng. Ethics* **2012**, *18*, 49–67. [[CrossRef](#)] [[PubMed](#)]
10. Tamburrini, G.; Altiero, F. Research Programs Based on Machine Intelligence Games. In *Italian Philosophy of Technology*; Chiodo, S., Schiaffonati, V., Eds.; Springer: Cham, Switzerland, 2021; pp. 163–179.
11. IPCC Intergovernmental Panel on Climate Change. Global Warming of 1.5°. 2018. Available online: <https://www.ipcc.ch/sr15/> (accessed on 2 January 2022).
12. Lucivero, F. Big data, big waste? A reflection on the environmental sustainability of big data initiatives. *Sci. Eng. Ethics* **2019**, *26*, 1009–1030. [[CrossRef](#)] [[PubMed](#)]
13. Rolnick, D.; Donti, P.L.; Kaack, L.H.; Kochanski, K.; Lacoste, A.; Sankaran, K.; Ross, A.S.; Milojevic-Dupont, N.; Jaques, N.; Waldman-Brown, A.; et al. Tackling Climate Change with Machine Learning. *arXiv* **2019**, arXiv:1906.05433.
14. Gardiner, S.M. *A Perfect Moral Storm: The Ethical Challenge of Climate Change*; Oxford University Press: Oxford, UK, 2011.
15. IEEE. Ethically Aligned Design. A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems. 2018. Available online: <https://ethicsinaction.ieee.org/#series> (accessed on 2 January 2022).
16. European Commission. Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonized Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts. 2021. Available online: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206> (accessed on 2 January 2022).
17. Gordon, J.-S.; Nyholm, S. Ethics of Artificial Intelligence. 2021. Available online: <https://iep.utm.edu/ethic-ai/> (accessed on 2 January 2022).

18. Müller, V.C. Ethics of Artificial Intelligence and Robotics. 2021. Available online: <https://plato.stanford.edu/archives/sum2021/entries/ethics-ai/> (accessed on 2 January 2022).
19. O’Neil, C. *Weapons of Math Destruction*; Penguin Books: London, UK, 2017.
20. IPPC Intergovernmental Panel on Climate Change. Climate Change 2021. Available online: <https://www.ipcc.ch/report/ar6/wg1/> (accessed on 2 January 2022).
21. Cerutti, F. *Global Challenges for the Leviathan. A Political Philosophy of Nuclear Weapons And Global Warming*; Rowman and Littlefield: Lanham, MD, USA, 2007.
22. U.S. Department of Defense. Autonomy in Weapons Systems. 2012. Available online: <https://www.hrw.org/news/2013/04/15/review-2012-us-policy-autonomy-weapons-systems> (accessed on 2 January 2022).
23. International Committee of the Red Cross. Views of the International Committee of the Red Cross on Autonomous Weapon System. Convention on Certain Conventional Weapons. In Proceedings of the Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS), Geneva, Switzerland, 11–15 April 2016.
24. Amoroso, D. *Autonomous Weapons Systems and International Law*; Edizioni Scientifiche Italiane and Nomos Verlag: Napoli, Italy, 2020.
25. Scharre, P. *Army of None. Autonomous Weapons and the Future of War*; W.W. Norton & Co.: New York, NY, USA, 2018.
26. Umbrello, S.; Wood, N.G. Autonomous weapons systems and the contextual nature of hors de combat status. *Information* **2021**, *12*, 216. [[CrossRef](#)]
27. Amoroso, D.; Tamburrini, G. Autonomous Weapons Systems and Meaningful Human Control: Ethical and Legal Issues. *Curr. Robot. Rep.* **2020**, *1*, 187–194. [[CrossRef](#)]
28. Umbrello, S. Coupling levels of abstraction in understanding meaningful human control of autonomous weapons: A two-tiered approach. *Ethics Inf. Technol.* **2021**, *23*, 455–464. [[CrossRef](#)]
29. Amoroso, D.; Tamburrini, G. Toward a Normative Model of Meaningful Human Control over Weapons Systems. *Ethics Int. Aff.* **2021**, *35*, 245–272. [[CrossRef](#)]
30. *The Ethics of Cybesecurity*; Christen, M.; Gordijn, B.; Loi, M. (Eds.) Springer: Cham, Switzerland, 2020.
31. Taddeo, M.; Floridi, L. Regulate artificial intelligence to avert cyber arms race. *Nature* **2018**, *556*, 296–298. [[CrossRef](#)] [[PubMed](#)]
32. Buchan, R.; Tsagourias, N. Autonomous Cyber Weapons and Command Responsibility. *Int. Law Stud.* **2020**, *96*, 645–673.
33. Heintz, C. Maturing autonomous cyber weapons systems: Implications for international security cyber and autonomous weapons systems regimes. In *Oxford Handbook of Cyber Security*; Cornish, P., Ed.; Oxford University Press: Oxford, UK, 2021; Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3255104 (accessed on 2 January 2022).
34. Altmann, J.; Sauer, F. Autonomous Weapon Systems and Strategic Stability. *Survival* **2017**, *59*, 117–142. [[CrossRef](#)]
35. Lin, H. Cyber risk across the U.S: Nuclear enterprise. *Tex. Natl. Secur. Rev.* **2021**, *4*, 108–120.