



Sustainable AI: AI for sustainability and the sustainability of AI

Aimee van Wynsberghe¹

Received: 29 January 2021 / Accepted: 8 February 2021
© The Author(s) 2021

Abstract

While there is a growing effort towards AI for Sustainability (e.g. towards the sustainable development goals) it is time to move beyond that and to address the sustainability of developing and using AI systems. In this paper I propose a definition of Sustainable AI; Sustainable AI is a movement to foster change in the entire lifecycle of AI products (i.e. idea generation, training, re-tuning, implementation, governance) towards greater ecological integrity and social justice. As such, Sustainable AI is focused on more than AI applications; rather, it addresses the whole sociotechnical system of AI. I have suggested here that Sustainable AI is not about how to sustain the development of AI per se but it is about how to develop AI that is compatible with sustaining environmental resources for current and future generations; economic models for societies; and societal values that are fundamental to a given society. I have articulated that the phrase Sustainable AI be understood as having two branches; AI for sustainability and sustainability of AI (e.g. reduction of carbon emissions and computing power). I propose that Sustainable AI take sustainable development at the core of its definition with three accompanying tensions between AI innovation and equitable resource distribution; inter and intra-generational justice; and, between environment, society, and economy. This paper is not meant to engage with each of the three pillars of sustainability (i.e. social, economic, environment), and as such the pillars of sustainable AI. Rather, this paper is meant to inspire the reader, the policy maker, the AI ethicist, the AI developer to connect with the environment—to remember that there are environmental costs to AI. Further, to direct funding towards sustainable methods of AI.

Keywords Sustainable AI · AI ethics · Sustainability · Responsible AI · Sustainable development

1 Introduction

There is little doubt that Artificial Intelligence (AI) is, and will continue to, transform the world. However, the power for positive change that AI brings holds the possibility for negative impacts on society. The recent explosion of AI, made possible by ever-rising amounts of data and computing power, has given rise to the field of AI ethics—the study of ethical and societal issues facing developers, producers, consumers, citizens, policy makers, and civil society organizations. The first wave of AI ethics focused on what AI might do (e.g. Superintelligence) and amounted to the ethics of fanciful scenarios of robot uprisings [6]. The second wave of AI ethics addressed the practical concerns of

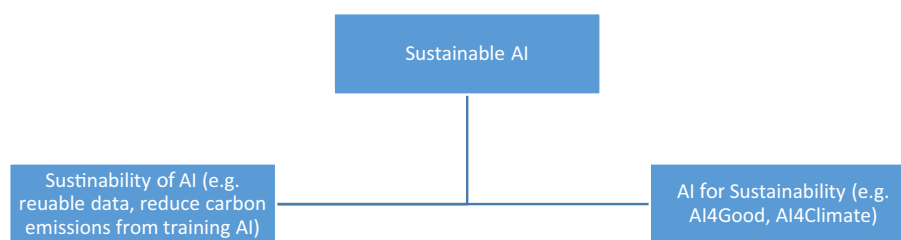
machine learning (ML) techniques: the black-box algorithm and the problem of explainability [9, 16], the lack of equal representation in training data and the resulting biases in AI models [2, 7], and the increase in facial and emotion recognition systems infringing on citizen's rights (e.g. privacy) [4]. It is time to usher in the third wave of AI ethics, one that confronts the environmental disaster of our time head-on and actively seeks to engage academics, policy makers, AI developers and the general public with the environmental impact of AI.

This third wave must place *sustainable* development at its core. While there is a growing movement to direct AI usage towards 'good' uses (i.e. AI4Good), towards the sustainable development goals, for example, it is time to move beyond that and to address the sustainability of developing and using AI systems in and of themselves. A well-known study by Strubell et al. illustrated that the process of training a single, deep learning, natural language processing (NLP) model (GPU) can lead to approx. 600,000 lb of carbon dioxide emissions [17]. Compare this to familiar consumption and

✉ Aimee van Wynsberghe
aimeevanrobot@protonmail.com

¹ Alexander von Humboldt Professor for Applied Ethics of AI, University of Bonn, Bonner Talweg 57, 53113 Bonn, Germany

Fig. 1 Sustainable AI as sustainability of AI vs AI for sustainability



you're looking at roughly the same amount of carbon dioxide emissions produced by five cars over the cars' lifetime. Other studies have shown that 'Google's AlphaGo Zero generated 96 tonnes of CO₂ over 40 days of research training which amounts to 1000 h of air travel or a carbon footprint of 23 American homes' [1, 15]. In a time when the world must commit itself to reducing carbon emissions, one has to ask if the emissions from algorithms that can play games (or do other menial tasks) is really worth the cost.

Added to this, AI is not a technology that is restricted to manufacturing or healthcare alone. It promises to be as pervasive as the internet or smartphones. This is a technology for which we cannot afford to ignore the environmental costs. For these reasons, I suggest we turn our attention to sustainable AI. I suggest we (AI ethicists) begin a movement towards sustainable AI as a way of connecting the dots between AI (its production, development and usage) and the environment, for the general public, AI developers and policymakers alike.

In the following paper, I will outline the concept of Sustainable AI as an umbrella term to cover two branches with different aims and methods; *AI for sustainability* vs *sustainability of AI*. I will show that *AI for sustainability* holds great promise but is lacking in one crucial aspect; it fails to account for the environmental impact from the development of AI. Alternatively, the environmental impact of AI training (and tuning) sits at the core of the *sustainability of AI*. *Sustainability of AI* is focused on sustainable data sources, power supplies, and infrastructures as a way of measuring and reducing the carbon footprint from training and/or tuning an algorithm. Addressing these aspects gets to the heart of ensuring the *sustainability of AI* for the environment. Finally, I argue that the study of Sustainable AI is imminently needed with resources directed to its understanding and development.

2 What is Sustainable AI?

The term, or phrase, 'Sustainable AI' is in its infancy. In fact, to my knowledge this is the first academic article with the explicit aim to propose a definition of Sustainable AI and to argue for its prominence. To begin, I suggest that 'sustainable AI' is a field of research that applies to the technology

of AI (the hardware powering AI, the methods to train AI, and the actual processing of data by AI) and the application of AI while addressing issues of AI sustainability and/or sustainable development. I suggest further that Sustainable AI deals not exclusively with the implementation or use of AI but ought to address the entire life cycle of AI, the sustainability of the: design, training, development, validation, re-tuning, implementation and use of AI.

Under this umbrella term, there is a crucial distinction to be made; *AI for sustainability* versus the *sustainability of AI* (see Fig. 1). The former branch—*AI for sustainability*—is somewhat more developed with the well-known not-for-profit organization "AI4Good".¹ In this branch, the goal is to explore the application of AI to achieve sustainability in some manner of speaking, for example, AI and machine learning (ML) to achieve the United Nations Sustainable Development Goals (SDGs). Here, AI/ML is a tool to make affordable and clean energy, SDG 7 for example, available for a greater segment of the global population. Seeing as approx 600 million people on this planet are currently lacking access to modern electricity² this is of course a 'good' goal to have. However, at what cost? It needs to be known that to train or tune (re-fine) an AI/ML model requires a considerable amount of energy and researchers have already posed the question if the "energy spent training a neural network might better be allocated to heating a family's home" [17]. Accordingly, Sustainable AI cannot restrict itself to the use of *AI for sustainability*.

Hence, I propose that sustainable AI encompass another branch of research, this branch is somewhat underdeveloped, under-researched and under-funded. This area deals with assessing the *sustainability of AI* or, the sustainable development of AI/ML itself. Thus, the *sustainability of AI* is not solely focused on how to apply AI for sustainable banking, energy consumption or healthcare; rather, this branch of sustainable AI is concerned with how to measure the sustainability of developing and using AI models, e.g. measuring of carbon footprints, computational power for training algorithms, etc. To be sure, Sustainable AI must address both of

¹ <https://ai4good.org/> (accessed Jan 28, 2021).

² See <https://ai4good.org/ai-for-sdgs/goal-7-affordable-clean-energy/> (accessed Jan 28, 2021).

these branches. Put another way, it should be clear that AI *for* sustainability cannot be achieved without simultaneously addressing the sustainability *of* AI.

3 What is the sustainable in ‘Sustainable AI’?

In an effort to clarify ambiguities found in the abundance of literature on sustainable development, Mensah [13] conducted a systematic literature review on the topic and argues that “the entire issue of sustainable development centres around inter- and intragenerational equity anchored essentially on three-dimensional distinct but interconnected pillars, namely the environment, economy, and society” [13, 1].

Although first derived from economics [13, 14], sustainable development has more recently been defined by the World Commission on Environment and Development as “development that meets the needs of the present without compromising the ability of future generations to meet their own needs” [13, 8]. As such, sustainable development embodies the tension between innovation with the equitable distribution of resources across society from one generation to the next.

Further, sustainable development is a “development paradigm as well as the concept that calls for improving living standards without jeopardising the earth’s ecosystems or causing environmental challenges such as deforestation and water and air pollution” [13, 6]. From this pervasive definition, it has been put forth that sustainable development is a mechanism through which ‘society can interact with the environment’ [13]. Thus, it is more than a concept but it is also appealed to as a movement of sorts.

Sustainable development has also been described as a ‘visionary and forward-looking paradigm’ [13, 9] with the three pillars upon which the concept of sustainable development rests: economic sustainability, social sustainability, and environmental sustainability. Thus sustainable development embodies not only the tension between innovation and equitable resource distribution but also the tension between serving the needs of the environment, economy and society.

Given what we have just read about sustainable development and the literal translation of sustainability being “a capacity to maintain some entity, outcome or process over time [5] then what could sustainable AI mean? I suggest that ‘Sustainable AI’ carry with it the complexities of sustainable development as it relates to AI. Thus, sustainable AI must also embody the tension between innovation in AI for sustainable development goals as well as explicitly targeting the sustainability of AI training and usage. Furthermore, I suggest that the global discussion on AI explicitly address, not only human rights and/or ethical issues, but also the tension between serving the needs of the environment, economy, and society.

In short, the AI which is being proposed to power our society cannot, through its development and use, make our society unsustainable. This thought should cause us to realize that society’s use of AI is a choice. It is not predetermined that AI must be developed and/or used for any- and everything; rather, it is a choice that society (e.g. developers, industry leaders, consumers, citizens, policy makers) make. This latter point, against technological determinism, also means that societies must choose carefully about the values they wish to safeguard (e.g. privacy, dignity, fairness, justice) and must work to ensure that AI is not developed in a way that renders those values unsustainable.

4 Environmental sustainability of AI

Renowned AI ethicist Mark Coeckelbergh proposes ‘AI for Climate’ suggesting that we use AI “for dealing with environmental and climate problems” [8, 5]. Given my earlier distinction of sustainable AI, I would label this as AI *for* sustainability. And while I couldn’t agree more, I also believe it is necessary to focus our attention on the sustainability *of* AI. This change in framing is paramount as it means that one can no longer talk about AI for Climate or AI4Good without at the same time addressing the impact that developing a particular AI model will have on environmental sustainability.

Thankfully there are a select few studying and developing AI models who are already bringing attention to the issue and outlining areas in need of further study. In a 2019 paper by Strubell et al. [17], it is argued that there are both financial and environmental costs to Deep Learning (DL) models for natural language processing (NLP). The financial costs were attributed to hardware and electricity or cloud compute time (which raised ethical issues as to who has access to such hardware etc.) whereas the environmental costs were attributed to “the carbon footprint required to fuel modern tensor processing hardware” [17, 1]. The authors acknowledge the energy required to power the hardware for training such models is considerable given that training happens over the course of weeks or even months [17].

The authors also point out that while it is possible to obtain some of the required energy from “renewable or carbon credit-offset resources, the high energy demands of these models are still a concern since (1) energy is not currently derived from carbon-neutral sources in many locations, and (2) when renewable energy is available, it is still limited to the equipment we have to produce and store it, and energy spent training a neural network might better be allocated to heating a family’s home” [17, 1]. This last point should strike a chord for all of us; as said before, there are approx 600 million people in the world without access to modern electricity and instead of prioritizing the provision of electricity to these homes, we are prioritizing the training

of AI models that can beat the world champion at the game of Go (AlphaGo). It is time for such calculations to be made explicit and to be evaluated in an open forum.

The Strubbel et al. [17] paper goes on to show that ‘tuning’ (aka re-purposing or refining) an AI model is more expensive than training a model to begin with. This kind of finding is crucial for the policy makers to understand to make decisions concerning the proportionality of certain AI methods compared with its intended application. Meaning, it is time for policy makers to govern AI at a more detailed level and suggest that certain methods, for example tuning a deep learning NLP model, should not be permitted for ethically charged tasks like recruitment of new employees or prediction of employees who may be on the verge of quitting. The reason being that the costs to environmental sustainability are simply too great to justify such a menial (not to mention ethically problematic) application. This could also cause society to pause when a particular AI model will be used in an application context which will require constant tuning. As society evolves in its communication, transportation, and social habits, old AI models will need constant tuning to continue to be effective. These costs must be added to any proportionality calculation.

One of the final recommendations from Strubbel et al.³ is that “authors should report training time and sensitivity to hyperparameters” as “this will enable direct comparison across models” (Strubell, Ganesh, and McCallum 2019, 5). To date, there are two possible tools available for calculating emissions: the ‘Machine Learning Emissions Calculator’ for estimating the carbon footprint of GPU compute through specifying hardware type, hours used, cloud provider, and region [12], Anthony et al. [3]; and, the ‘experiment-impact-tracker’ framework “for tracking real-time energy consumption and carbon emissions, as well as generating standardized online appendices” [10, 1]. Each of these approaches aims at the ‘mitigation of carbon emissions and reduction of energy consumption’ to facilitate the sustainable development of ML [10].

If we recall from the introduction, studies have shown that ‘Google’s AlphaGo Zero generated 96 tonnes of CO₂ over 40 days of research training which amounts to 1000 h of air travel or a carbon footprint of 23 American homes’ [1, 15]. Or, training one large NLP model (aka a transformer), with neural architecture search, resulted in over 600,000 CO₂e(lbs), roughly the equivalent of carbon emissions of five cars (over the lifetime of the car) [17]. These numbers are overwhelming to read. What’s worse is that we have only

a few studies to call on to learn numbers like this. In other words, we need more studies to fully grasp the extent to which these findings can be supported or refuted. With tools like the ‘machine learning emissions calculator’ [12] and the ‘experiment-impact-tracker’ [10] this should no longer be the case. This means the tools to track carbon emissions are there, however, greater incentives are needed to encourage researchers and industry developers to measure and report such findings.

Building on the work of Henderson et al., Anthony et al. propose ‘carbontracker’, “a tool for tracking and predicting the energy consumption and carbon emissions of training DL models” [3, 2]. Not only does the ‘carbontracker’ tool allow for the generation of carbon impact statements but it also allows for the model training to be “stopped, at the user’s discretion, if the predicted environmental cost is exceeded” [3, 2]. Thus, the ‘carbontracker’ provides the possibility that if a training exceeds a responsible use of energy consumption, or generation of carbon emissions, training of the model can be stopped. Again, this is the kind of tool that should be known to policy makers to create governance mechanisms for limiting the amount of carbon emissions, with the tools to end the training when emissions reach an unacceptable threshold.

In short, while the use of AI for achieving sustainability is to be applauded there are many reasons for which the environmental costs of AI, the sustainability of AI, need to be studied and made transparent to the AI community, consumers, and policy makers. More to the point, “the carbon emissions that occur when training DL models are not irreducible and do not have to simply be the cost of progress within DL” [3, 3].

5 Towards Sustainable AI

Distinguished authors such as Klein [11] write about the dangers of climate change and the need for consumer and industry habits to change. Climate activists such as Greta Thunberg work tirelessly to raise awareness of the need for systematic political reform to repair the damage to our planet. Politicians are taking note and acting. The European Commission has enacted ‘A European Green Deal’⁴ and the United States has re-joined the Paris Climate Agreement⁵ with plans for greater commitments to tackle climate change. Environmental resilience is a global issue at the heart of many policy and industry decisions so why not at

³ To be sure, the other two recommendations in the paper are: “academic researchers need equitable access to computation resources”, and “researchers should prioritize computationally efficient hardware and algorithms”. (Strubell, Ganesh, and McCallum 2019, 6).

⁴ See https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal_en (accessed Jan 28, 2021).

⁵ See <https://www.whitehouse.gov/briefing-room/statements-releases/2021/01/20/paris-climate-agreement/> (accessed Jan 28, 2021).

the heart of every AI ethics discussion, or every AI discussion, period? While the world argues over which principles to adopt, out of more than 200 sets of AI ethics principles, large tech companies are increasing their use of computational power and are increasing their demand for data and its storage in data centres across the world that require energy and cooling systems.

At the time of this paper, early 2021, it is unquestionable that any pervasive global production of products (or innovation) demands attention for its impact on the environment. Mass farming has shown to have an impact on biodiversity; mass clothing production has shown to have an impact on the water supplies of the world; and, electronic waste has shown to leak chemicals and poison into the water and soil where it is dumped (often in poorer countries). If we are to believe the promises from industry and policy makers, that AI/ML will be used on a global scale across public and private sectors, then AI/ML promises to be pervasive across sectors. Already we see applications to aid medical practitioners in their diagnoses, to assist legal officials in their judgements, to assume a portion of tasks in human resources for recruiting new employees. The list goes on and on. Hence, this is not a technology that we can afford to ignore the environmental impacts of. When it comes to AI, attention to the environmental impact is mandatory.

To do this, first, AI must be conceptualized as a social experiment conducted on society [18]. This is a technology we still have much to learn about. With the experimental nature of AI made explicit it is then imperative that ethical safeguards are put in place to protect people and planet.

Second, we need sustainable AI taskforces in governments who are actively engaged in seeking out expert opinions of the environmental impact of AI. From this, appropriate policy to reduce emissions and energy usage can be put into effect. Governance schemes to facilitate sustainable development of AI should be put in place: companies and public institutions should be required to provide carbon emissions reports for all training and tuning of AI systems; and, funding directed to SMEs actively pursuing sustainable AI innovation—and not just AI for sustainability but the sustainability of AI approaches.

Third, public and private Institutions, for example the European Commission as part of their regulatory options for AI,⁶ should create a ‘proportionality framework’ to assess whether training or tuning of an AI model for a particular task is proportional to the carbon footprint, and general environmental impact, of that training and/or tuning. This is

especially true with AI projects used by the state. We must require tech companies developing (training, re-training) AI models to use tools such as the ‘Carbon Tracker’ proposed by Anthony et al. [3] not only to track the carbon footprint of training a certain model but to predict the carbon footprint of training a certain DLM so as to stop model training if the predicted environmental cost is exceeded.

6 Conclusion

In this paper I propose a definition of Sustainable AI; Sustainable AI is a movement to foster change in the entire lifecycle of AI products (i.e. idea generation, training, re-tuning, implementation, governance) towards greater ecological integrity and social justice. As such, Sustainable AI is focused on more than AI applications; rather, it addresses the whole sociotechnical system of AI. I have suggested here that Sustainable AI is not about how to sustain the development of AI per se but it is about how to develop AI that is compatible with sustaining environmental resources for current and future generations; economic models for societies; and societal values that are fundamental to a given society.

I have articulated that the phrase Sustainable AI be understood as having two branches; AI *for* sustainability and sustainability *of* AI. The former has received a great deal of attention in recent years, the latter appears to be a hidden part of the development process.

I proposed that Sustainable AI takes sustainable development at the core of its definition with three accompanying tensions between AI innovation and equitable resource distribution; inter and intra-generational justice; and, between environment, society, and economy. This paper is not meant to engage with each of the three pillars of sustainability (i.e. social, economic, environment), and as such the pillars of sustainable AI. Rather, this paper is meant to inspire the reader, the policy maker, the AI ethicist, the AI developer to connect with the environment—to remember that there are environmental costs to AI.

Funding Open Access funding enabled and organized by Projekt DEAL. This research is supported by the Alexander von Humboldt Foundation.

Compliance with ethical standards

Conflict of interest The author confirms there is no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are

⁶ See https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf for more, “On Artificial Intelligence-A European approach to excellence and trust” (Accessed on Jan 28, 2021).

included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Agravente M.: MIT Moves toward Greener, More Sustainable Artificial Intelligence. In: Habitat (blog). <https://inhabitat.com/mit-moves-toward-greener-more-sustainable-artificial-intelligence/> (2020). Accessed 15 May 2020
2. Angwin, J., Jeff L.: Machine Bias. Text/html. ProPublica. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> (2016) Accessed date 23 May 2016.
3. Anthony LFW, Kanding B, Selvan R.: Carbontracker: Tracking and Predicting the Carbon Footprint of Training Deep Learning Models. *ArXiv:2007.03051* (2020).
4. Barrett, L.F., Adolphs, R., Marsella, S., Martinez, A.M., Pollak, S.D.: Emotional expressions reconsidered: challenges to inferring emotion from human facial movements. *Psychol. Sci. Public. Int.* (2019). <https://doi.org/10.1177/1529100619832930>
5. Basiago, A.D.: Economic, social, and environmental sustainability in development theory and urban planning practice: the environmentalist. Kluwer Academic Publishers, Boston (1999)
6. Bostrom N., Yudkowsky E. The Ethics of Artificial Intelligence. *The Cambridge Handbook of Artificial Intelligence*, 316–334 (2014).
7. Buolamwini, Joy, and Timnit Gebru. 2018. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." In *Conference on Fairness, Accountability and Transparency*, 77–91. PMLR. <http://proceedings.mlr.press/v81/buolamwini18a.html>.
8. Coeckelbergh, M.: AI for climate: freedom, justice, and other ethical and political challenges. *AI Ethics* (2020). <https://doi.org/10.1007/s43681-020-00007-2>
9. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., et al.: AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Mind. Mach.* **28**(4), 689–707 (2018). <https://doi.org/10.1007/s11023-018-9482-5>
10. 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. *ArXiv:2002.05651* (2020).
11. Klein, N.: *On Fire: The (Burning) Case for a Green New Deal*. Simon and Schuster, New York (2020)
12. Lacoste, A., Luccioni, A., Schmidt, V., Dandres T.: Quantifying the Carbon Emissions of Machine Learning. (2019)
13. Mensah, J.: Sustainable development: meaning, history, principles, pillars, and implications for human action: literature review. *Cogent. Soc. Sci.* **5**(1), 1653531 (2019). <https://doi.org/10.1080/23311886.2019.1653531>
14. Pigou, A.: *The Economics of Welfare*. Macmillan, London (1920)
15. Preedipadma. 2020. "New MIT Neural Network Architecture May Reduce Carbon Footprint by AI." *Analytics Insight*, April 2020. <https://www.analyticsinsight.net/new-mit-neural-network-architecture-may-reduce-carbon-footprint-ai/>.
16. Robbins, S.: A misdirected principle with a catch: explicability for AI. *Minds. Mach.* (2019). <https://doi.org/10.1007/s11023-019-09509-3>
17. Strubell, E., Ganesh, A., McCallum, A.: Energy and Policy Considerations for Deep Learning in NLP. In *ArXiv:1906.02243* (2019).
18. Wynsberghe, A van.: Artificial Intelligence: From Ethics to Policy | Panel for the Future of Science and Technology (STOA) | European Parliament. Study. Panel for the Future of Science and Technology. Brussels, European Union: Scientific Foresight Unit (STOA). [https://www.europarl.europa.eu/stoa/en/document/EPRS_STU\(2020\)641507\(2020\)](https://www.europarl.europa.eu/stoa/en/document/EPRS_STU(2020)641507(2020)).

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.