

# Sustainable Artificial Intelligence through Continual Learning

Andrea Cossu<sup>1</sup>, Marta Ziosi<sup>2</sup>, Vincenzo Lomonaco<sup>3</sup>

{andrea.cossu@sns.it<sup>1</sup>, marta.ziosi@aiforpeople.org<sup>2</sup>, vincenzo.lomonaco,@aiforpeople.org<sup>3</sup> }

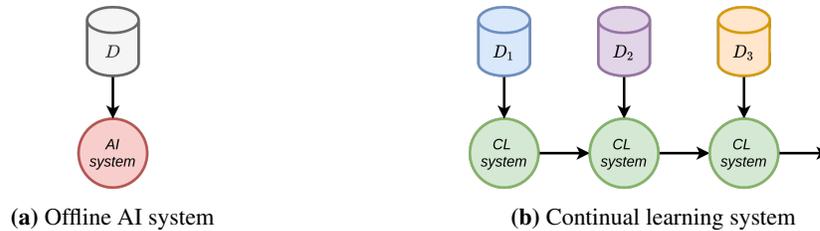
AI for People<sup>2,3</sup>, University of Pisa<sup>1,3</sup>, University of Oxford<sup>2</sup>, Scuola Normale Superiore<sup>1</sup>

**Abstract.** The increasing attention on Artificial Intelligence (AI) regulation has led to the definition of a set of ethical principles grouped into the Sustainable AI framework. In this article, we identify Continual Learning, an active area of AI research, as a promising approach towards the design of systems compliant with the Sustainable AI principles. While Sustainable AI outlines general desiderata for ethical applications, Continual Learning provides means to put such desiderata into practice.

**Keywords:** sustainable-ai, continual-learning, lifelong-learning

## 1 Introduction

AI systems learning from raw data are widely used in a large number of real world applications [1]. Their performance has dramatically increased the degree to which we are able to address previously inaccessible tasks (e.g. machine translation, image segmentation, speech-to-text and many others). Fortunately, the debate around AI is not only centered on its performance, but it is also increasingly focusing on other *non-functional* properties, like explainability, privacy and scalability [2]. In particular, the Sustainable AI framework encompasses a set of principles which are often deemed too general and abstract to lead to concrete implementation means [3]. In this paper, we focus on the "Sustainability of AI" and we show how Continual Learning [4], a recently growing topic in AI, can be a suitable candidate to render current AI Ethics principles both more operationalisable (one of their main critiques), and more socially, financially and environmentally sustainable. We proceed by first reviewing the Sustainable AI framework and then by presenting the main characteristics of the Continual Learning objective. Finally, we show the tight correspondence between Sustainable AI and Continual Learning objectives and we discuss how future research on the latter could lead to systems able to respect the Sustainable AI principles previously described.



**Fig. 1.** Difference between offline AI (left) and Continual Learning setup (right). While offline learning learns in a single phase from a static set of data, Continual Learning systems learn from a stream of non-stationary data.

## 2 Sustainable AI Principles

As the range of AI capabilities expands, so does our awareness of the ethical issues related to the the design, development, deployment and use of AI systems. The promise for positive change that AI represents has been challenged by several reports on ethically questionable uses of AI in contexts as varied as healthcare, education, law enforcement, recruitment, risk-assessment and more. This increasing number of reports has contributed to the birth of the field of AI Ethics. AI Ethics refers to "the study of ethical and societal issues facing developers, producers, consumers, citizens, policy makers, and civil society organizations" in the ambit of AI systems [3]. The field has so far taken a principled-base approach to ethics, with more than 70 documents containing AI Ethics principles published between 2016 and 2019. This principles stand as abstractions, arising as responses to practical concerns around the application of AI techniques [5]. They are expressed by values such as fairness, privacy, accountability, transparency, etc. One of the main critiques advanced in relation to these AI Ethics principles has been their overtly abstract character, represented by the failure to move from principles to practice [6]. This critique led to a search for more usable solutions.

Concurrently, Wynsberghe in [3] has started calling for a change of focus in AI Ethics. She calls for a focus on "Sustainable AI" and she conceives it as a new wave of AI Ethics that puts sustainable development at its score. Sustainable AI encompasses both AI for Sustainability, represented by initiatives such as "AI for Good" or "AI4SDGs", as well as the Sustainability of AI represented by underfunded concerns about the social, financial and environmental impact of AI systems. While the initiatives on AI for good abound, AI Ethics has paid much less attention to assessing what Wynsberghe calls the "Sustainability of AI", especially in terms of the environmental impact of AI systems themselves. Here, we present the set of AI Ethics principles that we will be concerned with in this paper:

- **Efficiency & Scalability.** Modern machine learning solutions usually require long training phase on large datasets. Training such models have a drastic impact on both the environment (due to high resource consumption) and the dynamic acquisition of new information (due to the large amount of data required). Sustainable AI approaches, instead, should incorporate an efficient and scalable approach to learn ever complex information from a stream of data, without requiring massive amount of computation each time new knowledge comes in.

- **Fairness, Privacy & Security.** While these three principles can stand in separation, they relate to each other at some level. Fairness can be understood as the absence of bias and discrimination, that is, ensuring equitability [2]. This may refer to the inner process of algorithmic decisions as well as in their outcomes. Responses to discrimination, as well as to biased and opaque decision-making often appeal to data privacy [7]. This can entail solutions that prevent decisions being taken based on protected attributes such as race or gender. Privacy refers to the capacity for an individual to control information about themselves. AI Ethics guidelines present privacy both as a value to uphold and as a right to be protected [2]. One way in which privacy can be violated is by breaching the security of a system through adversarial attacks that expose the model structure, the training or test dataset. Security here refers to measures taken to avoid foreseeable as well as unintentional harm from or through AI systems.
- **Accuracy & Robustness.** Algorithmic conclusions are based on probabilities. As such, they are not infallible and they might incur in errors during execution. The fact that algorithmic decisions may be taken on inconclusive evidence renders the accuracy of their predictions an important value to uphold. Apart from accurate, AI systems need to be resilient and secure. This aspects are expressed in the principle of robustness.
- **Explainability, Transparency & Accountability.** Explainability refers to the explanation of an algorithmic decision being comprehensible to technical people such as data-scientists as well as to non-technical people such as the data subjects affected by the decision [7]. While ethical guidelines feature an overreliance on explicability [6], it can be problematic to uphold this value without information being accessible. Accessibility and comprehensibility of information are the two main components of transparency [8]. Transparency stands in contrast to the opacity of algorithms, especially when referring to the problem of black-box algorithms. Additionally, the opacity of machine learning algorithms is considered as a challenge to accountability [9]. In that respect, solutions to ensure accountability such as human oversight are threatened by opaque algorithms. There is no accountability without transparency.

### 3 Continual Learning as Key Technological Enabler

Making machines that can learn and think like humans has always been one of the main goal of Artificial Intelligence (AI) research. Since its inception, AI has been thought as grounded in the ability to learn and adapt to new situations. Indeed, learning itself progresses through a continuous exposure to external stimuli and experiences. However, despite early speculations and few pioneering works, very little research and effort has been devoted to address this vision. Current AI systems (Figure 1a) learn offline from a static dataset and greatly suffer from the exposure to new data or environments which even slightly differ from the ones for which they have been trained for [10, 11]. One notable example of such deterioration is the catastrophic forgetting phenomenon, which causes the disruption of previous knowledge when learning new information [10]. Moreover, the learning process is usually constrained on fixed datasets within narrow and isolated tasks which may hardly lead to the emergence of more complex and autonomous intelligent behaviors. The idea of creating algorithms that can learn continually and incrementally from a progressive exposure to external

stimuli (Figure 1b) is at the core of the recent surging of interest in Continual Learning research, within the AI and Machine Learning community [12]. Learning continually means not only being able to adapt to unpredictable circumstances not foreseen during the design of our AI systems, but also to build systems that are *efficient*, *scalable* and *easily amendable*, or in other words, *sustainable*.

### 3.1 Continual Learning for Sustainable AI

We will now address the sustainable AI principles described in Section 2 and we will show how they can be satisfied by means of AI systems which learn continuously over time.

**Efficiency and Scalability** The efficiency of AI systems is usually measured with respect to metrics like *computing time* (how much time does it take to learn), *memory occupancy* (how much memory is it required to learn) and *sample efficiency* (how many data patterns are required to learn) [4, 13]. Scalability is commonly approached by extrapolating these metrics for growing dataset / model sizes. However, the evaluation protocol currently in use for both efficiency and scalability considers only static environments, where the training process occurs on a fixed dataset, entirely available beforehand. Continual learning operates in a completely different settings, where patterns are learned *over time*: the computing time and memory occupancy are distributed during the entire learning experience and, since the model is capable to tackle few patterns at a time incrementally, available resources can be adjusted dynamically, as needed. The advantage in efficiency and scalability for Continual Learning with respect to offline AI models increases with the nonstationary degree of the environment: each new piece of information may require to retrain a non continual model on both new and previous data. This may be unnecessary and wasteful, since most of the knowledge needed is already inside the model. Continual learning, on the contrary, only requires to train a model on the new incoming samples, possibly exploiting temporal coherence between successive patterns to improve performance [14, 15].

**Fairness, Privacy & Security** AI systems operating on the edge (e.g., on device) are not able to incorporate new information, since they suffer from the catastrophic forgetting of previous knowledge. The process of retraining on both new and old data is expensive due to the large size of datasets and cannot be performed on resources-constrained devices like smartphones or even modern laptop. Therefore, each time new data is available the model need to be sent back to a central computing unit (e.g., datacenter on cloud) on which retraining is performed. This process exposes security holes which can be exploited by an external attacker in order to retrieve sensitive knowledge. For example, a wearable device collecting medical parameters from a patient may need to send its data on the cloud for further training, thus allowing for stealing of private information. Continual learning systems are instead able to learn directly on-device and they remove the need to send data to external infrastructures, greatly limiting possible security issues.

On a different note, lack of fairness in model predictions often comes by the fact that AI models are trained on biased datasets, where information is incomplete or driven by social/political prejudices. Continual Learning may help in understanding and preventing biased predictions in two ways: firstly, it allows to study changes in predictions as new data is injected into the model. Since no re-

training is performed, the change in prediction is directly related to the input patterns provided to the model. Secondly, once biased prediction is detected, Continual Learning is capable of efficiently incorporate new information. In real-world applications, this opportunity may be prevented by the fact that the model would need a long and costly update, possibly to manage few edge cases.

**Accuracy & Robustness** While current AI systems excel in addressing narrow tasks with high accuracy, they are prone to catastrophic forgetting of previous knowledge as soon as they encounter new information. Therefore, the prediction accuracy with respect to all the data seen by the model reduces drastically when considering even slightly nonstationary environments. This makes AI systems unreliable and not robust to possible changes in the incoming data. In order to address this point, Continual Learning studies the tradeoff between the prediction accuracy on the current data and the prediction accuracy on all previous data. Even if existing Continual Learning solutions do not completely remove catastrophic forgetting of previous knowledge, they are able to largely mitigate it [16]. As a matter of fact, catastrophic forgetting may never be solved completely. It is nonetheless fundamental to highlight that accuracy alone is not a robust evaluation metric and that the performance should be measured in terms of the entire model lifetime. This directly advocates for the use of effective continual learning models able to operate in less confined environments.

**Explainability, Transparency & Accountability** Modern AI solutions are often thought as *black boxes*, with few to no possibility to understand the reasons and the rationale behind their decisions. Continual Learning does not directly address this problem, since a continual learner may as well be able to update its internal representations during time without disclosing any hints on what such representations mean. Interestingly, a prominent line of research in Continual Learning calls for the use of neuroscience-grounded solution [17]. While the human brain is far from being explainable and accountable, it is nonetheless possible in that context to use scientific tools to draw general conclusions on the behavior of different brain regions and to study which of them are more entrusted with a high-level cognitive function. By building Continual Learning models which follow neuroscience principles, it may be possible in the future to study these systems with similar tools and to reach a similar understanding of its inner functioning, which seems one of the most ambitious advances towards explainable systems.

## 4 Conclusion and Future Directions

In this paper, we presented Continual Learning as an effective approach towards the development of Sustainable AI systems able to tackle real-world applications in respect to a set of principles including efficiency, privacy, robustness and explainability. In particular, we showed that the Continual Learning objectives and the Sustainable AI desiderata are aligned, in the sense that a model able to learn continuously will be most likely brought to respect the Sustainable AI principles. The growing amount of contributions to the field suggests that the path is worth pursuing: as new Continual Learning solutions are designed, the advantages in the use of models which learn continuously over static ones will become wider and will highlight even further the opportunities for a concrete realization of the Sustainable AI framework. While we acknowledge that Continual Learning may not

be the only viable solution towards sustainable models, here we supported the idea that Continual Learning is currently the most promising candidate and a necessary requirement to drive research in the direction of Sustainable AI.

## References

- [1] LeCun Y, Bengio Y, Hinton G. Deep Learning. *Nature*. 2015;521(7553):436–444.
- [2] Jobin A, Ienca M, Vayena E. The Global Landscape of AI Ethics Guidelines. *Nature Machine Intelligence*. 2019;1(9):389–399.
- [3] van Wynsberghe A. Sustainable AI: AI for Sustainability and the Sustainability of AI. *AI and Ethics*. 2021;1(3):213–218.
- [4] Lesort T, Lomonaco V, Stoian A, Maltoni D, Filliat D, Díaz-Rodríguez N. Continual Learning for Robotics: Definition, Framework, Learning Strategies, Opportunities and Challenges. *Information Fusion*. 2020;58:52–68.
- [5] Turilli M. Ethical Protocols Design. *Ethics and Information Technology*. 2007;9(1):49–62.
- [6] Morley J, Floridi L, Kinsey L, Elhalal A. From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. *Science and Engineering Ethics*. 2020;26(4):2141–2168.
- [7] Mittelstadt BD, Allo P, Taddeo M, Wachter S, Floridi L. The Ethics of Algorithms: Mapping the Debate. *Big Data & Society*. 2016;3(2):2053951716679679.
- [8] Turilli M, Floridi L. The Ethics of Information Transparency. *Ethics and Information Technology*. 2009;11(2):105–112.
- [9] Schermer B. The Limits of Privacy in Automated Profiling and Data Mining. *Comput Law Secur Rev*. 2011.
- [10] French R. Catastrophic Forgetting in Connectionist Networks. *Trends in Cognitive Sciences*. 1999;3(4):128–135.
- [11] Grossberg S. How Does a Brain Build a Cognitive Code? *Psychological Review*. 1980;87.
- [12] Maltoni D, Lomonaco V. Continuous Learning in Single-Incremental-Task Scenarios. *Neural Networks*. 2019;116:56–73.
- [13] Lomonaco V, Pellegrini L, Cossu A, Carta A, Graffieti G, Hayes TL, et al. Avalanche: An End-to-End Library for Continual Learning. In: *CLVision Workshop at CVPR; 2021*. .
- [14] Parisi GI, Lomonaco V. Online Continual Learning on Sequences. *Studies in Computational Intelligence*. 2020.
- [15] Cossu A, Carta A, Lomonaco V, Bacciu D. Continual Learning for Recurrent Neural Networks: An Empirical Evaluation. *Neural Networks*. 2021;143:607–627.
- [16] De Lange M, Aljundi R, Masana M, Parisot S, Jia X, Leonardis A, et al. A Continual Learning Survey: Defying Forgetting in Classification Tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2021.
- [17] Cui Y, Ahmad S, Hawkins J. Continuous Online Sequence Learning with an Unsupervised Neural Network Model. *Neural Computation*. 2016;28(11):2474–2504.