NOKIA

A transparent and standards-based way to assess the environmental impact of AI systems

White paper

Mireille Farzan and Susanna Kallio

This paper discusses a framework for assessing the environmental impact of an AI system. For the purpose of looking at how different life cycle stages contribute to the overall environmental impact, Nokia mapped the AI system life cycle to the environmental life cycle assessment method. This ENVironmental Impact Assessment of AI systems (ENVIAA) framework shows the importance of evaluating each life cycle stage separately to add transparency into where the environmental impacts are coming from rather than using a bundling approach. We also applied the ENVIAA framework for high-level estimation of the impacts from low and high availability of renewable energy from an ICT equipment perspective. Our evaluation concludes that low availability of renewable energy contributes to proportionally high GHG emissions in the use stage, which are increased by adding AI system computational processes due to increases in energy consumption. For high availability of renewable energy, the raw materials, transportation, and end-of-life treatment of the hardware have a higher proportional impact on GHG emissions, which is not affected by the addition of AI system computation unless major hardware changes are needed for introducing AI functionality.

Contents

Introduction 3

AI system life cycle 4

Modeling AI system life cycle 4

Energy consumption 6

Energy efficiency 6

AI for sustainability 7

Summary of AI system life cycle representation 7

The environmental impact assessment of AI framework 8

Introduction to life cycle assessment 8

Mapping AI system life cycle into LCA 9

Impact from energy consumption 9

Embodied environmental impact of AI systems 10

Impact from energy source 10

Other environmental impacts than GHG emissions 12

Conclusions 13

Abbreviations 13

References 14

Introduction

The environmental impacts of AI systems represent a subcategory of a wider discussion on sustainable AI, which includes also responsible, trustworthy and ethical AI (more on these principles here [1]). This broader discussion aims to identify ethical, social, governance, and environmental aspects of AI systems with the aim to increase the benefits of developing and using AI systems and decrease the potential unwanted outcomes that can arise from this activity. The environmental impact discussion relates to how the development and use of AI systems impacts the environment, the climate and nature.

There is a growing trend to develop and implement more and more AI solutions throughout multiple industries. The growth in this technology trend makes a strong case for evaluating the environmental aspects related to it, since both the development and use of AI systems can be a highly energy consuming process, which, depending on the source of energy used, can result in massive environmental impact [2, p. 217], [3]. The growth rate of energy consumption by AI systems is estimated to be annually around 30%. By 2028, AI could consume more energy than the entire country of Iceland used in 2021 [4]. This speaks to the environmental impacts of AI systems.

Sustainability in software development can refer to two aspects: product longevity or software development that considers its environmental, social and economic implications [5]. By developing software with sustainability in mind, we can expand the sustainability impact related to the use of it. The main focus of this paper is looking at sustainability from the point of view of the environmental impact of the full life cycle, including production and use of AI systems. This viewpoint is often called environmentally sustainable AI for its consideration of the environmental impacts of AI systems.

Alternately, there is an “AI for sustainability” viewpoint, defined as the application of AI systems to enable sustainability goals and the environmentally sustainable development of different areas and industries. [2, p. 214]. This includes using AI systems to reduce the unwanted environmental impacts of a function or process. Here AI can be used to optimize energy intensive functions and reduce the use of energy in the reference system or to track and reduce emissions, pollution, or other aspects that contribute to environmental impact. We have included some points from this viewpoint as well in this paper.

The positive sustainability outcomes that are reached through implementing AI are to be coupled with environmentally sustainable methods of designing and using AI systems. AI environmental considerations are relevant throughout the entire AI system life cycle. These environmental aspects and impacts are linked to the inception, design, development, deployment, operation, and retirement of the AI system.

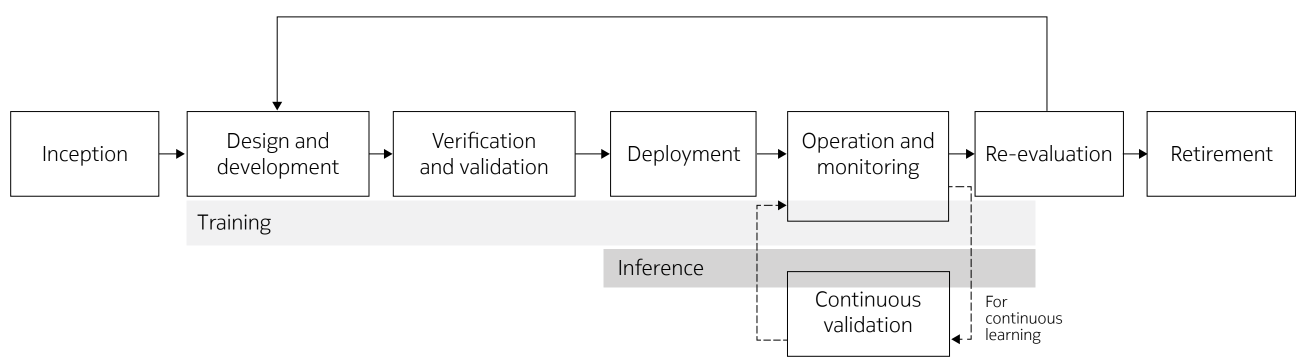
In this paper, we look at ways to assess the environmental impact of AI systems. We created the ENVIAA framework by first looking at the AI system life cycle that depicts important stages in the evolution of an AI system and, second, by mapping these stages onto the standardized environmental life cycle assessment (LCA) method. In the end, we explore the impact of two scenarios in relation to energy source, one with low availability of renewable energy, the other with high availability.

AI system life cycle

Modeling AI system life cycle

An AI system life cycle describes the evolution of the AI system and can include various important stages that occur throughout the life cycle of the AI system from inception to retirement. The stages mentioned in the AI system life cycle model can be technical processes, governance processes (such as organizational project enabling processes, risk management processes [6, p. 9-13], or AI ethics review stages [7, p. 4]), or focus on any other stages of the system's evolution that are selected for closer evaluation. Depending on the intended purpose of the life cycle description, life cycles can depict different stages in different levels of detail. A life cycle description can include all stages from inception to retirement or a subset of these stages. There is a standardized general approach to AI system life cycles (ISO 5338, see [6]), but there has yet to be a unified AI system life cycle model that can be used to point out different elements of different AI systems.

Figure 1. AI system life cycle highlighting stages where training and inference can occur



A life cycle model can include stages such as inception, design and development, verification and validation, deployment, operation and monitoring, continuous validation, re-evaluation and, finally, retirement. These stages can be separated into further substages. For example, the design and development stage can include substages such as acquiring training data, data preparation, algorithm selection, and model training [6, pp. 22-25]. The deployment stage can be separated into metrics evaluation, reviews and operationalization [7, p. 4]. In this paper, our approach to the AI system life cycle includes the stages depicted in figure 1, which also highlights the stages where training and inference can occur depending on the AI system.

Training and inference

Some stages of an AI system life cycle can be linked to higher energy consumption than others. While an AI system life cycle model is not an environmental impact assessment method, observations of energy consumption can be important when looking at the environmental impacts of AI systems. This can be done when the energy consumption of a life cycle stage is compared to that of another. For instance, some interesting stages for machine learning systems involve training and inference.

For machine learning systems, training can refer to the process model training. Depending on the system, this can mean using the training data to establish internal representations or to identify underlying functions that map to wanted outcomes. For example, this can mean using training data that includes pictures of cats to teach the model to recognize cats from new images. Training can be linked to multiple stages in an AI system life cycle. It can be linked to the design and development stage of the system when it refers to model training on training data. Retraining can be a part of a validation stage in connection to model updates and the continuous monitoring of the model and its performance. This can be especially relevant if there are deviations from wanted outcomes or in cases of continuous learning. These processes can also be seen as maintenance activities that ensure the model produces wanted outcomes over time.

Where the initial training phase relates to training data, retraining and continuous training can be linked to production data, which is data acquired during the operation stage of an AI system. The role of this data is especially important for systems with continuous learning. Depending on how an AI system life cycle is depicted, and, depending on the system, these additional training phases can be scattered along the validation, operation and/or monitoring stages of the AI life cycle.

Model training is one of the AI system life cycle substages that has been linked to needing substantial computing power. This is because, in some cases, a training model can require weeks' or months' worth of hardware energy consumption at a time. In addition to training time, other aspects such as the energy consumption of specific hardware solutions as well as model sensitivity to hyperparameters (configuration variables for the model) can affect the power consumption [8].

Inference in AI systems can refer to a process executed by the system or its result. This process or result is derived from a rule, a model, a feature or raw data [9]. Inference can be a prediction, conclusion or an outcome produced by the system. Examples of inference can be samples of text or images produced by an AI system, predictions made by a system, or when an AI system identifies or recognizes elements such as images from new information. From the AI system life cycle perspective, this type of inference can take place in the deployment, operation and monitoring stages.

Inference has also been suggested as one of the AI system life cycle stages that contributes to considerable amounts of energy consumption in comparison to other stages. The inference stage can be an ongoing process where the system continuously applies its training to new information. This can result in substantial energy consumption if the system is continuously in use, for example, in cases where the system is continuously used to produce text, images, predictions, or other outcomes. Here, energy consumption is linked to the power draw of the hardware during inference.

Comparing the energy consumption of training to that of inference can have varying results depending on the specific AI system and how it is used. For some systems, training only occurs in the design and development stage of the AI system life cycle. This can be followed by a significantly longer period of operation and monitoring where inference occurs millions of times a day, every day. In this example, due to a shorter one-time training time in comparison to a longer ongoing inference period, the energy consumption of training can be negligible in size compared to that of inference. However, in cases of machine learning research where more time is spent on training the system or when different periods of training occur throughout the life cycle of the system, the energy consumption of the training phase can dominate over other processes throughout the AI system life cycle. In addition to these two examples, where either training or inference dominates the energy consumption, it can also be a mixed case of high energy consumption both in training and inference due to a wide variety of factors. Therefore, a case-by-case evaluation is needed for different AI systems focusing on the system's design phase and intended use to determine energy consumption in both training and inference.

Energy consumption

The energy consumption of an AI system, measured in kilowatt-hours (kWh), is tied to the energy required to power the hardware that the computation runs on. This consumption strongly depends on the stage of the AI system life cycle, the type of hardware that is used and how it is used [8]. It is also heavily dependent on the type of AI system and its intended use. One way of estimating energy consumption is by looking at the average power draw of key components such as CPUs, GPUs or NPUs. Key factors for the energy consumption are the number of parameters in a model, model efficiency, and the power usage effectiveness of the data center [10, p. 120]. To get a wider picture, energy calculations also need to include the additional energy required to support the compute infrastructure and site related energy consumption, including the energy used for cooling [8, p. 2].

An important factor about AI system hardware is that it can be dedicated solely to the functions of one AI system, used as shared hardware for multiple AI systems, or used for AI and other non-AI related computational processes simultaneously. Here, it is important to consider how much of the energy consumption is attributable to a specific AI system and how much is linked to other AI systems or non-AI related processes. To evaluate the environmental impact of a single AI system in case of shared hardware, an allocation of the hardware to the different systems running on it needs to be conducted, e.g. by time allocation estimating or measuring how many hours each of the systems is using the hardware. From the allocated time slot in the computing (percentage of the total running time), the dedicated share from the total hardware energy consumption can be calculated. However, keeping the computational resources available even when not in direct use has some idle mode energy consumption, although rather minimal compared to the active mode consumption.

Energy efficiency

An energy efficient software is one that is able to produce wanted outcomes with minimal energy consumption. Energy efficiency can be evaluated at different stages, for example, the development and/or the operation stage of the software [5, p. 8]. One important aspect to leveraging energy efficient AI development methods is being able to measure and predict the energy consumption of specific AI development methods. Not only can this weigh in method selection considerations, but it also allows designers to determine potential upcoming environmental costs and offers options such as stopping model training when predefined environmental costs are being exceeded [2, pp. 216 -217]. In addition, accurate accounting in this area enables accurate cost-benefit analysis, raises stakeholder awareness, drives mitigation efforts, and helps stakeholders to achieve their environmental goals. **On the other hand, lack of accounting methods means that designers are unaware of the impacts of their models and might not pursue mitigation measures** [3, p. 7].

A key topic is reducing energy consumption and emissions linked to AI development. One approach is through re-use. The re-use or reproducibility of models and code can provide an approach for spending less resources on the same or similar products. This approach is enabled by researchers and developers releasing and sharing code and models, when it is appropriate. This can happen by releasing them to the community or internally within a company. It reduces the environmental impact, energy use and greenhouse gas (GHG) emissions of replicating resources or results [3, p. 23].

The AI community can also benefit from shared information about the most energy efficient combination of hardware, software and algorithms. This information can include metrics such as carbon emissions, energy use, runtime, or cost, and it may provide insights on whether some methods are more energy intensive than others or whether there are no significant differences in terms of sustainability [3, p. 17]. Other approaches to reducing the environmental impact linked with AI design can be found in optimized code for lower energy consumption, selecting methods by considering training time, and continuous verification to ensure efficiency. Other solutions to reduce energy consumption are optimizing the training data to train with less data and scheduling retraining less frequently when the model accuracy is still good enough.

AI for sustainability

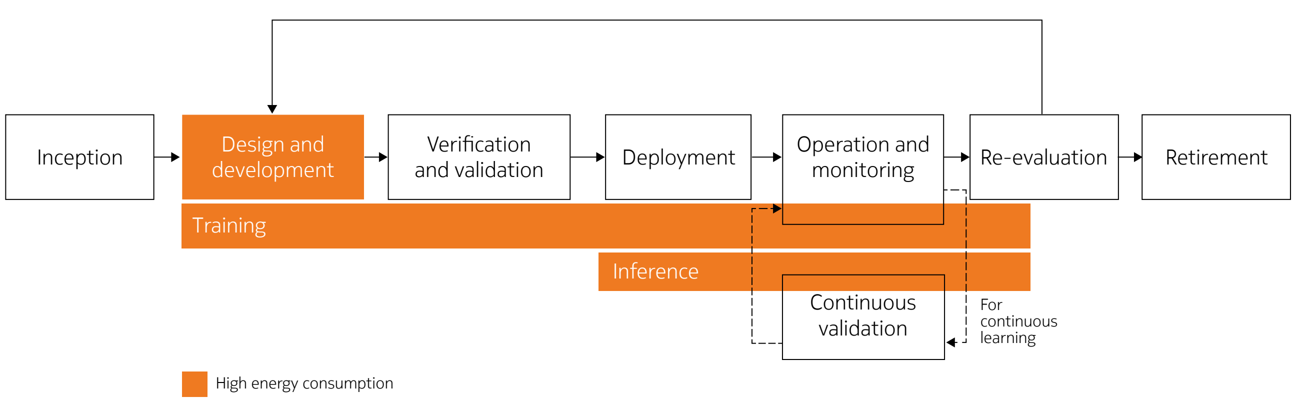
A traditional cost-benefit analysis can be based on weighing the estimated revenue generated by the model or AI system against, for example, the cost of electricity. Another approach can be weighing the emissions saved by the system against the emissions generated. This can be the case when the emissions generated during the lifetime of an AI embedded product are lowered by using the AI system. For example, AI systems can be used to reduce the energy consumption in ICT network infrastructure. By analyzing network traffic and the correlated energy and resource use, AI can be used to learn and predict when and where network service is needed and to identify unused resources that consume energy and can be switched off. This increases energy efficiency by optimizing energy use in active and passive network equipment. The targeted outcome is reduced network energy consumption while maintaining network performance.

Using AI systems to reduce energy consumption is sometimes considered separately from the energy that is consumed when designing and using an AI system. When AI systems are designed for increasing energy efficiency of the reference system, it becomes important to measure the energy consumption during the development and operation stages of the AI system, when possible, as well as the energy consumption of the process before AI is embedded for energy efficiency, and the reduced energy consumption that results from using the AI system. As the development of AI systems contributes to total energy consumption, adding AI systems to pre-existing processes should be justified from an energy consumption perspective as well as to assess the total benefit or environmental impact.

Summary of AI system life cycle representation

This section has looked at the AI system life cycle and its connection to the assessment of energy consumption. The AI system life cycle representation can be useful for describing different stages and processes included in the AI system's evolution and for highlighting the more energy intensive stages of the life cycle. **Depending on the AI system and its use case, the most energy hungry stages and processes are typically training and inference, and, to a lesser extent, design and development.** Other stages, like retirement or inception, are likely to consume less energy. This is depicted in figure 2 by coloring the high energy consumption processes with orange.

Figure 2. AI system life cycle with high energy consumption processes marked with color



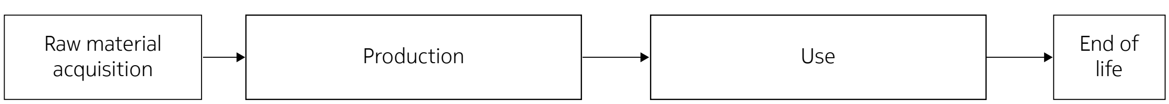
However, the AI system life cycle does not provide enough information for assessing the full environmental impact of an AI system, since it merely depicts the different stages and processes within the AI system life cycle but does not provide methodological guidance on how to assess environmental impact. On the other hand, environmental impact assessment methodology for traditional software is not directly applicable to AI systems due to the added complexity that AI technology introduces. For this, a holistic approach is needed. We suggest the ENVIAA framework that incorporates an environmental life cycle assessment and maps to it the AI system life cycle. This will be the topic of the next section.

The environmental impact assessment of AI framework

Introduction to life cycle assessment

Environmental life cycle assessment (LCA) is an assessment method created and standardized for assessing the environmental impact of a product's full life. The International Organization for Standardization (ISO) published standard ISO 14040 in 2006 [11]. It is a generic environmental assessment method to evaluate and improve environmental performance applicable for all sectors and products globally and used widely across a range of industries from manufacturing and chemicals to agriculture and transport and also includes ICT. In 2014, International Telecommunication Union Telecommunication Standardization Sector (ITU-T) produced and published the ICT-specific LCA standard L.1410 [12], which is based on ISO 14040 and adds further ICT-specific details and guidance for LCA practitioners. The LCA assessment method looks into four main life cycle stages: raw material acquisition, production, use, and end-of-life treatment, as shown in figure 3 [11].

Figure 3. Life cycle assessment stages



The raw material acquisition stage is very specific to the material production for the hardware. This considers the environmental impacts from mining and extraction and further processing of the material into a form that can be utilized in the production of the assessed product. The production stage includes manufacturing a product to a sellable condition or finalizing it to be used for its intended purpose. Typically, manufacturing includes processing input materials and assembling parts and components to form the final product. The use stage covers installation and the use of the product for its intended purpose as well as maintenance during the product's use. The final disposal of the product, including material recycling, is at the end-of-life treatment stage. Additional processes like design, transportation, repair, reuse, and refurbishment are included in that life cycle stage where the process takes place.

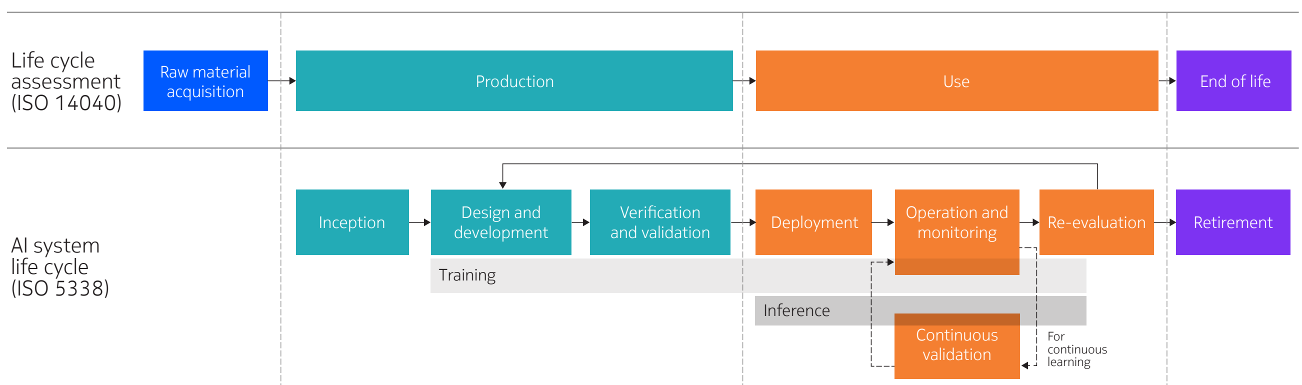
LCA method studies all processes and unit processes within the four main life cycle stages. System boundaries need to be clearly set to get valuable results from using the method. In the method, all the material and energy flows within the product life cycle are identified and associated with selected environmental impact category and its indicator. The most often used environmental impact category is global warming potential, also called climate change or greenhouse gas (GHG) emissions.

When looking at AI systems, the mapping of AI system life cycle stages into the LCA stages is intuitive and looks simple but is not straightforward. Below we present the ENVIAA framework and propose how this mapping can be done. Additionally, we present different scenarios for pointing out the general hot spots in AI systems from an environmental impact perspective.

Mapping AI system life cycle into LCA

Looking at the two models, AI system life cycle and environmental life cycle assessment, there are some commonalities but also use case dependent processes that are not easily mapped. Figure 4 illustrates the ENVIAA framework with our proposed life cycle stage mapping.

Figure 4. The ENVIAA framework



Inference can be easily mapped to the LCA use stage, as well as re-training during use. We mapped continuous validation also to the LCA use stage, since it is expected to happen after the AI system is in operation and use. The challenge is where to map the AI training process in the LCA stages. Depending on the use case scenario, training could be either in production or use stage from an LCA perspective. For AI solutions that are sold as working products with all AI features included (e.g., a smartphone), training can be mapped to part of the LCA production stage. On the other hand, an AI system that is provided to the end customer with embedded AI functionality or capability and the customer is customizing the AI system for their intended use, training can be considered to belong to the LCA use stage. **For the sake of general applicability of this framework, we mapped training across LCA production and use stages.** This use case dependent mapping leads to ambiguity in embodied environmental impact, which in LCA means environmental impact from all other life cycle stages except the use stage. Another alternative for mapping training would have been to map it either to LCA production or use stage, thus fixing the applicability of this framework to only those use cases which comply.

Impact from energy consumption

Now that we have mapped the AI life cycle stages and processes to the environmental LCA methodology stages, we can start to look at the GHG emissions from AI systems. The GHG emissions from the AI system's computational part can be assessed by looking at the emissions linked to energy consumption. These emissions originate from the production of the energy that powers the AI system. Different energy sources contribute to different levels of emissions with fossil fuels contributing to significantly higher emission levels than renewable energy sources [13, p. 19].

GHG emissions for an AI system can come from multiple processes in the AI system life cycle. When looking at the emissions linked to the AI system development stage, an important role is played by the energy consumption, training time and frequency, and the carbon intensity of the energy in use [3, p. 39]. As there are many data points to look at, it is also important to notice that one can severely over or underestimate energy use and related carbon emissions when extrapolating from partial information [3, p. 15].

When all means to reduce the energy consumption of the AI system have been taken (examples discussed above), the GHG emissions related to the development and use of AI systems can be further reduced through utilizing energy with lower GHG emissions. Cloud-providers located in regions with low carbon intensity energy can help lower the emissions associated with the AI system. The carbon intensity of the region or cloud-provider is tied to the energy production method that is used. Here, moving model training to regions with low carbon intensity can reduce the emissions associated with developing the system. This approach can be further supported with statistical information on the emissions associated with different cloud provider regions [3, p. 8]. Regional carbon intensity gives the average value for all energy consumption in that region. More accurate impact can be calculated when the actual energy source in the consumption site is known and used in the environmental impact assessment. In addition, there might be cases where there is a carbon intensity difference between daytime and nighttime energy use. In these cases, scheduling training during low carbon intensity time, e.g., in the daytime with solar energy, would lower contribution to GHG emissions [3, p. 12].

Additional environmental impacts of AI systems come from the life cycle of the hardware and components that enable computation. This additional impact includes the energy use and emissions from hardware raw material acquisition, manufacturing, use, as well as end of life treatment [3, p. 4].

Embodied environmental impact of AI systems

The embodied emissions are full life cycle emissions of a product excluding use stage emissions, so in general everything associated with manufacturing, transportation and end of life treatment. One study suggests that embodied emissions of AI systems are becoming dominant in the full life cycle emissions [14]. That suggestion assumes that model training contributes to embodied emissions. Another study concludes that in the future embodied emissions from hardware manufacturing will become the major source of GHG emissions [15]. That study assumes that data centers use only renewable energy, while hardware manufacturing would use only partly renewable energy. As shown by these examples, there is no common approach on how to communicate conclusions about embodied emissions. **For that reason, it would always be good to report transparently the study's assumptions together with its conclusions.**

To overcome different interpretations of embodied environmental impact, **we propose to keep the impact from AI system development and initial training separate from hardware manufacturing and to report environmental impact from training in the LCA production stage separate from the embodied environmental impact**, especially in cases when training expands beyond development and validation stages. By doing so, we should enable more transparent environmental impact assessment and better see the hot spots and areas for improvement from an environmental impact perspective.

Impact from energy source

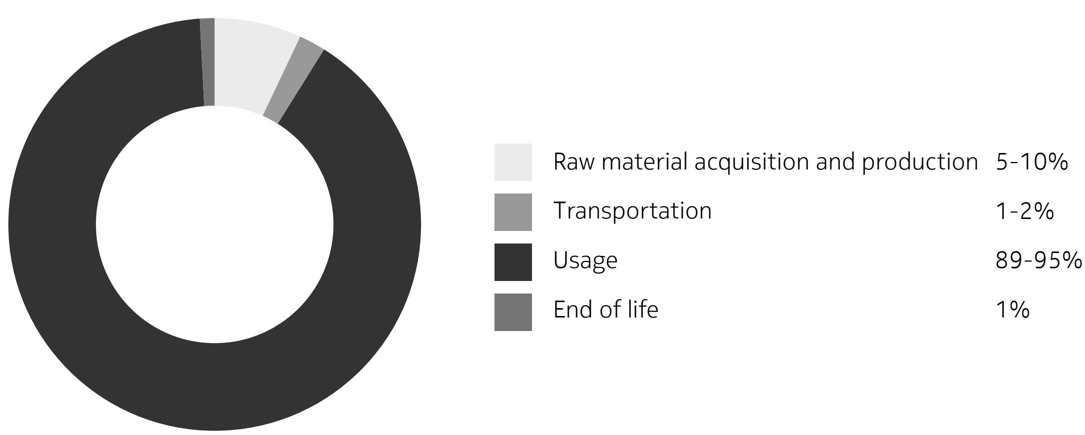
When considering GHG emissions from an AI system, the energy source and emission factors related to it play major roles. When using 100% renewable energy, the impact from energy consumption is minimized. In that case, even the energy hungry processes in AI cause only low GHG emissions. As companies are more and more moving to renewable energy in their operations, many also request their suppliers to move to renewable energy, thus reducing GHG emissions in the whole value chain—expanding also to hardware manufacturing and raw material production. The question that remains is whether the energy production can provide enough renewable energy for all due to the ever-increasing demand and intermittent availability of these energy sources.

Currently, there is not enough renewable energy available to satisfy all global energy demand, and it is not easily available 24/7, e.g., due to the intermittency of sun light or wind. Therefore, we consider next two extreme example scenarios: 1) when there is little or no renewable energy available and 2) when 100% renewable energy is available for all energy needs. The reality today and for the near future lies between these extremes with roughly 20-30% renewable energy available as a global average [16], [17], [18] and unevenly distributed. As a third example scenario, we explore below a case of environmental impact screening.

Scenario 1: Low availability of renewable energy

For typical ICT equipment, use stage energy consumption dominates the total lifetime environmental impact over the embodied emissions, when relatively low levels of renewable energy are available. Figure 5 shows an example of GHG emission shares in different life cycle stages based on the product LCA of typical configurations for mobile, fixed and core optic network products using the global average energy emission factor.

Figure 5. An example showing operator's network products' share of GHG emissions in different life cycle stages [19]



When AI system computation is added, it becomes important to look at the most energy hungry aspects of the AI system, which are typically training and/or inference. The ENVIAA framework, introduced earlier for the life cycle stage mapping exercise, showcases how these AI stages can be mapped to the LCA use stage and partly overlap with the production stage, adding to the total energy consumption. At the same time, they increase the proportional share of the use and production stages even more from the above mentioned 89-95% and 5-10%.

In cases of low availability of renewable energy sources, it becomes important to assess the additional energy consumption added by the implementation of AI systems. When considering positive or negative environmental impact compared to the reference system without AI functionalities, the addition of the AI system should be justified by a total benefit analysis to cover the additional environmental impact, especially for use cases where energy saving is targeted by introducing AI functionality into the product.

Scenario 2: High availability of renewable energy

The GHG emissions from the energy use will be minimal when 100% renewable energy is used both in product manufacturing and use stage. This expands also to the impact from the AI computational processes. This would be an ideal case for AI when we would not need to worry about the energy consumption of the AI system either in training or in inference processes. In this exceptional case, the environmental impact from the raw materials, transportation, and end-of-life treatment of the product's hardware will have the biggest impact in the GHG emissions. Targeting 100% renewable energy in the whole value chain will leave emissions from the materials and transportation as the main source of GHG emissions. These are mostly hardware related and, therefore, the AI impact on top of the hardware does not increase emissions, unless a major hardware update is needed due to the introduction of the AI functionality. Even in this ideal case, we still need to consider the environmental impact from building the supporting energy infrastructure for the unlimited renewable energy generation and distribution.

Scenario 3: Environmental impact screening

A more realistic example of the use of the ENVIAA framework is screening for the main environmental impact and planning actions to reduce it. Let's take as an example the environmental impact from energy consumption. The assessment of energy consumption during different AI processes and life cycle stages, together with the related emission factors, gives us hot spots with major environmental impact, for which relevant mitigation actions can be planned. If a long training process is one of the hot spots, this process could be moved to a location where more renewable energy or energy with lower carbon intensity is available. Other actions could be shortening the training time, if feasible, and/or scheduling training for a time or season when energy with lower carbon intensity is available.

Other environmental impacts than GHG emissions

Current discussions on AI systems' environmental impact mostly focus on GHG emissions as an indicator of the global warming potential, sometimes called "climate change". The main focus in this paper has also been on GHG emissions. However, there are many other environmental impact categories. The environmental LCA method provides indicators also for other environmental impacts from a product, such as land use, water use and resource depletion. **For AI, the main impacts outside of GHG emissions come from raw material extraction and production required for data center infrastructure, computing hardware and energy generation.** More studies and discussions are needed on these other environmental impacts in the future.

The physical material used both in hardware manufacturing and energy generation consumes resources from the Earth. At the end of life, waste from electrical and electronic equipment, called e-waste, creates pollution. To overcome these challenges, substitute materials are being introduced that have fewer environmental impacts and new refining and production processes are being explored. **Circularity is being introduced to reduce e-waste and recover materials that otherwise would need to be extracted from the ground.** Recycled material is becoming more available as industry adopts and governments mandate circular economy practices, although more progress is needed in all these areas.

Conclusions

This paper looks at ways to assess the environmental impact of AI systems. We introduced the ENVIAA framework by first looking at the AI system life cycle that depicts important stages in the evolution of an AI system, and second, by mapping these stages onto the environmental life cycle assessment (LCA) method. In assessing the AI system life cycle, it is suggested that, within the AI system, training and inference are the most important stages where energy consumption is concerned. In the ENVIAA framework, we suggest separating the environmental impact of the AI system development and initial training from that of the hardware manufacturing process in embodied emissions. This adds transparency and helps clarify the most important hot spots and areas for improvement from an environmental impact perspective.

The ENVIAA framework can be used also for high-level estimations. We explored two simplified example scenarios in relation to energy sources. In the first scenario, it is concluded that during low availability of renewable energy, the use stage of the product contributes the most to the environmental impact. This high impact is increased with the addition of AI system computation and possibly expanded to training (depending on the use case) due to the increased energy consumption. A second scenario is assessed with high availability of renewable energy. Here, it is suggested that the environmental impact from the raw materials, transportation, and end-of-life treatment of the product have the biggest impact that is not increased by the addition of AI system computation, unless major component and material enhancements are needed due to AI functionality.

While AI for sustainability can enable reduction of energy consumption of the reference system, for example, using AI systems to increase energy efficiency, we suggest that the addition of AI systems should be done mindfully by checking AI system full life cycle environmental impact.

Abbreviations

AI	Artificial intelligence
CPU	Central processing unit
ENVIAA	Environmental impact assessment of AI systems
ICT	Information and communication technology
GHG	Greenhouse gas
GPU	Graphical processing unit
ISO	International Organization for Standardization
ITU-T	International Telecommunication Union - Telecommunication Standardization Sector
LCA	Life cycle assessment
NPU	Neural processing unit

References

[1] Nokia, "Responsible AI," Nokia Bell Labs, 2024. Online:https://www.bell-labs.com/research-innovation/ai-software-systems/responsible-ai/

[2] Van Wynsberghe, A., "Sustainable AI: AI for sustainability and the sustainability of AI." AI and Ethics 1(3), pp. 213-218, 2021. Online: https://link.springer.com/article/10.1007/s43681-021-00043-6

[3] Henderson, P. et al, "Towards the systematic reporting of the energy and carbon footprints of machine learning," The Journal of Machine Learning Research 21(1), pp. 10039-81, 2020. Online: https://jmlr.org/papers/volume21/20-312/20-312.pdf

[4] Ammanath, B., "How to manage AI's energy demand — today and in the future," World Economic Forum, Apr 2024. Online: https://www.weforum.org/agenda/2024/04/how-to-manage-ais-energy-demand-today-tomorrow-and-in-the-future/

[5] Haider, W. et al, "Factors influencing sustainability aspects in crowdsourced software development: a systematic literature review." Journal of Software: Evolution and Process, 36(6), pp. e2630, 2023. Online: https://doi.org/10.1002/smr.2630

[6] International Organization for Standardization, "Information technology — Artificial intelligence — AI system life cycle processes," ISO Standard No. 5338, 2023. Online: https://www.iso.org/standard/81118.html

[7] De Silva, D. and Alahakoon, D., "An artificial intelligence life cycle: From conception to production," Patterns 3(6), Jun 2022. Online: https://doi.org/10.1016/j.patter.2022.100489

[8] Strubell, E. et al, "Energy and Policy Considerations for Modern Deep Learning Research," Proceedings of the AAAI Conference on Artificial Intelligence, 34(09), pp. 13693-13696. Online: https://doi.org/10.1609/aaai.v34i09.7123

[9] International Organization for Standardization, "Information technology, Artificial intelligence — Artificial intelligence concepts and terminology," ISO Standard No. 22989:2022), Jul 2022. Online: https://www.iso.org/standard/74296.html

[10] Maslej, N. et al, "AI Index 2023 Annual Report," Stanford University, Institute for Human-Centered AI, Apr 2023. Online: https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index-Report_2023.pdf

[11] International Organization for Standardization, "Environmental management — Life cycle assessment — Principles and framework," ISO 14040:2006, Jul 2006. Online: https://www.iso.org/standard/37456.html

[12] ITU-T, "Methodology for environmental life cycle assessments of information and communication technology goods, networks and services," Recommendation ITU-T L.1410, Dec 2014. Online: https://www.itu.int/ITU-T/recommendations/rec.aspx?rec=12207

[13] Weisser, D., "A Guide to Life-Cycle Greenhouse Gas (GHG) Emissions from Electric Supply Technologies," Energy 32(9), pp. 1543-1559, Sep 2007. Online: https://doi.org/10.1016/j.energy.2007.01.008

[14] Wu, C.-J., et al, "Sustainable AI: Environmental Implications, Challenges and Opportunities," Jan 2022. Online: https://doi.org/10.48550/arXiv.2111.00364

[15] Gupta, U., et al, “Chasing carbon: The elusive environmental footprint of computing,” 2021 IEEE International Symposium on High-Performance Computer Architecture (HPCA), Seoul, Korea (South), Feb-Mar 2021, pp. 854-867. Online: https://ieeexplore.ieee.org/document/9407142

[16] IEA, “Pathways for the energy mix,” World Energy Outlook 2023, IEA, 2023. Online: https://www.iea.org/reports/world-energy-outlook-2023

[17] Ritchie, H. and Rosado, P., “Energy Mix,” Our World in Data, 2020. Online: https://ourworldindata.org/energy-mix

[18] Enerdata, “Share of renewables in electricity production,” World Energy & Climate Statistics – Yearbook 2024, 2024. Online: https://yearbook.enerdata.net/renewables/renewable-in-electricity-production-share.html

[19] Nokia, “People and Planet report,” Nokia web site, 2020. Online: https://www.nokia.com/sites/default/files/2021-04/Nokia_People_and_Planet_Report_2020.pdf

About Nokia

At Nokia, we create technology that helps the world act together.

As a B2B technology innovation leader, we are pioneering networks that sense, think and act by leveraging our work across mobile, fixed and cloud networks. In addition, we create value with intellectual property and long-term research, led by the award-winning Nokia Bell Labs.

Service providers, enterprises and partners worldwide trust Nokia to deliver secure, reliable and sustainable networks today – and work with us to create the digital services and applications of the future.

Nokia OYJ
Karakaari 7
02610 Espoo
Finland
Tel. +358 (0) 10 44 88 000

Document code: CID214115 (August)