Organisation for Economic Co-operation and Development

DSTI/CDEP/AIGO(2022)3/FINAL

Unclassified English - Or. English

17 November 2022

Directorate for Science, Technology and Innovation
Committee on Digital Economy Policy

Working Party on Artificial Intelligence Governance

MEASURING THE ENVIRONMENTAL IMPACTS OF AI COMPUTE AND APPLICATIONS: THE AI FOOTPRINT

JT03507954

Foreword

This paper was prepared for publication by the OECD Secretariat in consultation with the OECD.AI Expert Group on AI Compute and Climate. The paper was approved and declassified by the Committee on Digital Economy Policy (CDEP) on 28/09/2022.

This report was developed in consultation with experts involved in the Global Partnership on Artificial Intelligence's Working Group on Responsible AI. Its contents reflect the opinions of the GPAI Experts involved and do not necessarily represent the official views of GPAI Members.

Note to Delegations:

This document is also available on iLibrary as:

OECD (2022), "Measuring the environmental impacts of artificial intelligence compute and applications: The AI footprint", OECD Digital Economy Papers, No. 341, OECD Publishing, Paris, https://doi.org/10.1787/7babf571-en.

Abstract

Artificial intelligence (AI) systems can use massive computational resources, raising sustainability concerns. This report aims to improve understanding of the environmental impacts of AI, and help measure and decrease AI's negative effects while enabling it to accelerate action for the good of the planet. It distinguishes between the direct environmental impacts of developing, using and disposing of AI systems and related equipment, and the indirect costs and benefits of using AI applications. It recommends the establishment of measurement standards, expanding data collection, identifying AI-specific impacts, looking beyond operational energy use and emissions, and improving transparency and equity to help policy makers make AI part of the solution to sustainability challenges.

Abrégé

Les systèmes d'intelligence artificielle (IA) peuvent utiliser des ressources informatiques considérables, ce qui pose des problèmes de développement durable. Ce rapport vise à améliorer la compréhension des impacts environnementaux de l'IA, et à aider à quantifier et à minimiser les effets négatifs de l'IA, tout en lui permettant de contribuer à accélérer les initiatives pour le bien de la planète. Le rapport distingue les impacts environnementaux directs liés au développement, à l'utilisation et à la destruction des systèmes d'IA et des équipements connexes, ainsi que les coûts et avantages indirects de l'utilisation d'IA. Ce rapport préconise l'établissement de normes de mesure, l'élargissement de la collecte de données, l'identification des impacts spécifiques à l'IA, la prise en compte de la consommation d'énergie et des émissions autres que celles liées à l'exploitation, ainsi que l'amélioration de la transparence et de l'équité pour aider les responsables de l’action publique à faire de l'IA une solution aux problèmes de durabilité.

Executive summary

The green and digital "twin transitions" offer the promise of leveraging digital technologies for a sustainable future. As a general-purpose technology, artificial intelligence (AI) has the potential not just to promote economic growth and social well-being, but also to help achieve global sustainability goals. AI-enabled products and services are creating significant efficiency gains, helping to manage energy systems and achieve the deep cuts in greenhouse gas (GHG) emissions needed to meet net-zero targets. However, training and deploying AI systems can require massive amounts of computational resources with their own environmental impacts.

The computational needs of AI systems are growing, raising sustainability concerns. While AI can be perceived as an abstract, non-tangible technical system, it is enabled by physical infrastructure and hardware, together with software, collectively known as "AI compute". In the last decade, the computing needs of AI systems have grown dramatically, entering what some call the "Large-Scale Era" of compute. At the same time, according to the International Energy Agency (IEA), data centre energy use has remained flat at around 1% of global electricity demand, despite large growth in workloads and data traffic, of which AI is estimated to represent a small fraction. While this may point to hardware efficiency gains, some researchers note that AI compute demands have grown faster than hardware performance, bringing into question whether such efficiency gains can continue.

The environmental impacts of AI compute and applications should be further measured and understood. Policy makers need accurate and reliable measures of AI's environmental impacts to inform sustainable policy decisions. The 2010 OECD Recommendation on ICTs and the Environment encourages the development of comparable measures of environmental Information ICT impacts. Further, the 2019 OECD Recommendation on Artificial Intelligence underlines that AI should support beneficial outcomes for people and the planet. The 2021 OECD Recommendation on Broadband Connectivity also stresses the need to minimise the negative environmental impacts of communication networks. Yet further efforts are needed to develop measurement approaches specifically focused on AI and its environmental impacts.

The report defines AI compute as *including one or more "stacks" (i.e. layers) of hardware and software used to support specialised AI workloads and applications in an efficient manner*. This definition was developed by the OECD.AI Expert Group on AI Compute and Climate (the "Expert Group") to meet the needs of both technical and policy communities. Informed by the Expert Group and experts involved in the Global Partnership on AI (GPAI), this report synthesises findings from a literature review, a public survey and expert interviews to assess how the environmental impacts of AI are currently measured.

A number of indicators and measurement tools can help quantify the *direct* environmental impacts from AI compute, as well as the *indirect* environmental impacts from AI applications. The report distinguishes between *direct* and *indirect* positive and negative environmental impacts. Direct impacts stem from the AI compute resources lifecycle (i.e. the production, transport, operations and end-of-life stages). Analysis indicates that direct impacts are most often negative and stem from *resource consumption*, such as the use of water, energy and its associated GHG emissions, and other raw materials. Indirect impacts result from *AI applications* and can be either positive, such as smart grid technology or digital twin simulations, or negative, such as unsustainable changes in consumption patterns.

Sustainability and measurement good practices enable efficiency gains in AI compute. Several good practices for sustainable AI exist, such as using pre-trained models, where relevant, and powering data centres with renewable resources. Researchers at the Massachusetts Institute of Technology (MIT) and start-up MosaicML are training neural networks up to seven times faster by configuring AI algorithms to learn more efficiently. Some compute providers are starting to report AI-specific estimates. For example, Google says that its machine learning workloads represented about 15% of its total energy use over the last three years. A large cloud compute provider[3] estimates that between 7-10% of enterprise customers' total spend on compute infrastructure supports AI applications, with 3-4.5% used for training machine learning models and 4-4.5% spent using these models (known as "inference"). Such estimates help quantify AI-specific energy use and associated GHG emissions, while shedding light on how impacts differ according to whether compute is used to train AI models or to use them (inference).

Policy makers must ensure that AI is part of the solution to meet global sustainability targets. A starting point is to address five measurement gaps with policy implications:

1. **Measurement standards for sustainable AI are needed.** Measuring the environmental impacts of AI compute and applications to inform policy decisions would be facilitated by consensus on terminology, standards, consistent indicators and reporting requirements. A comprehensive framework developed by international or inter-governmental standard-setting institutions and international initiatives, as part of a multi-stakeholder process, could enable benchmarking, comparability and compatibility of national AI compute initiatives and their environmental impacts. Organisations such as the OECD could contribute to developing such a framework.

2. **Data collection on the environmental impacts of AI compute and applications should be expanded.** Efforts to collect national, firm and AI model level environmental data should be expanded. National agencies and institutions, and private-sector actors should collect more data using sustainability metrics such as GHG emissions, energy, water and natural resources used for AI compute, and AI applications where possible.

3. **AI-specific measurements are difficult to separate from general-purpose compute**. It is challenging to distinguish compute used for AI from that for other scientific, mathematical and general-purpose ICT needs. Further efforts should be made by governments, national statistical offices, intergovernmental organisations, the private sector, academia and others to disaggregate ICT infrastructure datasets, estimate the share used by AI and explore relevant proxy measures.

4. **Environmental impacts beyond operational energy use and GHG emissions should be considered.** The environmental impacts of AI compute beyond the energy use and carbon footprint of the operations stage (i.e. the production, transport, and end-of-life stages) warrant further research. This includes biodiversity assessments and the impacts of AI compute on other planetary boundaries (e.g. land system change and freshwater use), direct natural resource impacts from manufacturing, transport and end-of-life impacts, and indirect impacts from AI applications.

5. **Efforts are needed to improve environmental transparency and equity everywhere.** Most frameworks and analysis of AI compute are undertaken by experts from advanced economies. With negative environmental impacts anticipated to disproportionally affect emerging economies, further research should focus on ensuring that AI compute and applications support sustainability objectives across a broader range of national contexts, and sharing information and best practices.

Résumé

La « double transition » verte et numérique recèle en elle la promesse d'une mobilisation des technologies numériques au service d'un avenir durable. En tant que technique universelle, l'intelligence artificielle (IA) n'a pas seulement le potentiel de promouvoir la croissance économique et le bien-être social : elle peut aussi contribuer à la réalisation des objectifs mondiaux en matière de durabilité, au sens où les produits et services rendus possibles par l'IA sont des sources de gains d'efficience importants, qui peuvent aider à gérer les systèmes énergétiques et à réaliser les importantes baisses d'émissions de gaz à effet de serre (GES) nécessaires pour atteindre les objectifs de neutralité carbone. Cela étant, l'entraînement et le déploiement des systèmes d'IA peuvent aussi nécessiter des quantités massives de ressources computationnelles ayant elles-mêmes des impacts environnementaux.

Les besoins en puissance de calcul des systèmes d'IA ne font que croître, suscitant des inquiétudes en termes de durabilité. Si l'IA peut être perçue comme un système technique abstrait, non tangible, elle repose tout de même sur des infrastructures physiques, du matériel et des logiciels, désignés collectivement sous le nom de « capacité de calcul nécessaire à l'IA ». La dernière décennie a vu croître de manière spectaculaire les besoins informatiques des systèmes d'IA, nous faisant entrer dans ce que certains appellent « l'ère de la grande échelle ». Dans le même temps, selon l'Agence internationale de l'énergie (AIE), l'utilisation d'énergie par les centres de données est restée stable, à environ 1 % de la demande d'électricité mondiale, et ce, malgré l'augmentation considérable des charges de travail et de la circulation de données. Cette stabilité est peut-être le signe de gains d'efficience au niveau des matériels informatiques, l'IA ne représentant, selon les estimations, qu'une petite fraction de l'utilisation des technologies de l'information et de la communication (TIC) au sens large. Cependant, certains chercheurs relèvent que les demandes de capacité de calcul nécessaire à l'IA ont augmenté plus rapidement que les demandes de performance du matériel, si bien que l'on peut se demander si ces gains d'efficience vont pouvoir se poursuivre.

Les impacts environnementaux de la capacité de calcul nécessaire à l'IA et de ses applications devraient donc être tout à la fois plus finement mesurés et mieux compris. De fait, les responsables de l'action publique ont besoin de mesures précises et fiables des impacts environnementaux de l'IA pour pouvoir prendre, de manière éclairée, des décisions durables. La Recommandation du Conseil de l'OCDE de 2010 sur les technologies de l'information et des communications et l'environnement encourageait déjà l'élaboration de mesures comparables des impacts environnementaux des produits et services des TIC. Par ailleurs, la Recommandation du Conseil de l'OCDE de 2019 sur l'intelligence artificielle souligne que l'IA devrait tendre vers des résultats bénéfiques pour les individus et la planète, notamment favoriser le développement durable. Enfin, la Recommandation du Conseil de l'OCDE de 2021 sur la connectivité à haut débit insiste également sur la nécessité de minimiser les impacts environnementaux négatifs des réseaux de communication. Cependant, des efforts encore plus poussés s'imposent pour élaborer une métrique concernant spécifiquement la capacité de calcul nécessaire à l'IA et ses impacts environnementaux.

Dans le présent rapport, la capacité de calcul nécessaire à l'IA est définie comme *comprenant une ou plusieurs « strates » (ou couches) de matériels et de logiciels utilisés pour étayer, de manière efficience, les charges de travail et applications spécialisées propres à l'IA*. Cette définition a été préparée par le Groupe d'experts OECD.AI sur la capacité de calcul pour l'IA et le climat (ci-après le « Groupe d'experts ») pour répondre aux besoins de la communauté des techniciens comme de celle des décideurs publics. S'appuyant sur les éclairages fournis par le Groupe d'experts ainsi que par des spécialistes du Partenariat mondial sur l'intelligence artificielle (PMIA) chargé du Groupe de travail sur l'IA, le présent rapport dresse une synthèse des conclusions recueillies à l'issue d'un examen de la littérature, d'une étude auprès du public et d'entretiens menés avec des experts afin d'évaluer comment sont actuellement mesurés les impacts environnementaux de la capacité de calcul nécessaire à l'IA et des applications de cette dernière.

Plusieurs indicateurs et outils de mesures peuvent aider à quantifier les impacts environnementaux *directs* de la capacité de calcul nécessaire à l'IA, ainsi que les impacts environnementaux *indirects* des applications de l'IA. Dans le rapport, la distinction est faite entre les impacts environnementaux positifs et négatifs *directs* et *indirects*. Les impacts directs sont imputables au cycle de vie des ressources mobilisées pour la capacité de calcul nécessaire à l'IA (à différents stades : production, transport, activité et fin de vie). Les analyses montrent que les impacts directs sont le plus souvent négatifs et qu'ils sont le résultat de la *consommation de ressources*, par exemple la consommation d'eau, d'énergie (avec les émissions de gaz à effet de serre (GES) qu'elle entraîne) et de matières premières. Les impacts indirects sont dus aux *applications de l'IA* et peuvent être soit positifs, comme la technologie des réseaux intelligents ou les simulations à partir de jumeaux numériques, soit négatifs, comme des évolutions non durables des modes de consommation.

De bonnes pratiques en matière de durabilité et de métrique sont des sources de gains d'efficience au niveau de la capacité de calcul nécessaire à l'IA. De bonnes pratiques pour une IA durable existent déjà, comme le recours à des modèles pré-entraînés, lorsque cela est possible, ou l'utilisation de ressources renouvelables pour l'alimentation électrique des centres de données. Des chercheurs du Massachusetts Institute of Technology (MIT), en coopération avec la start-up MosaicML, arrivent en ce moment à entraîner des réseaux neuronaux sept fois plus vite en configurant des algorithmes d'IA pour qu'ils apprennent de manière plus efficiente. Par ailleurs, certains fournisseurs de capacité de calcul commencent à communiquer des estimations spécifiques à l'IA. Ainsi, Google a fait savoir que les charges de travail d'apprentissage automatique avaient représenté 15 % de sa consommation totale d'énergie au cours des trois dernières années. Un grand fournisseur[3] estime que, sur le total des dépenses de ses entreprises clientes en infrastructures de capacité de calcul, une proportion comprise entre 7 et 10 % est consacrée à des applications de l'IA, entre 3 et 4.5 % à l'entraînement de modèles fondés sur l'apprentissage automatique et entre 4 et 4.5 % à l'exécution de ces modèles (phase dite d'« inférence »). Ces estimations aident à quantifier l'utilisation d'énergie propre à l'IA et les émissions de GES qu'elle génère, et montrent comment les impacts diffèrent selon que la capacité de calcul est utilisée pour entraîner des modèles d'IA ou pour les utiliser (inférence).

Il appartient aux responsables de l'action publique de s'assurer que l'IA fait partie de la solution adoptée pour atteindre les objectifs de durabilité au niveau mondial. On pourrait pour commencer essayer de combler cinq lacunes qui existent actuellement en matière de mesure et qui ont des conséquences en termes d'action publique :

1. **Il faut élaborer des normes de mesure à l'appui d'une IA durable.** La mesure des impacts environnementaux de la capacité de calcul nécessaire à l'IA et de ses applications, de manière à éclairer l'action publique, serait facilité par un consensus sur la terminologie et les normes, par l'élaboration d'indicateurs et d'une métrique cohérents, et par des obligations de diffusion d'informations. L'élaboration d'un cadre global par des institutions internationales ou intergouvernementales ou des initiatives internationales chargées de définir des normes, via un processus impliquant de multiples parties prenantes, pourrait permettre l'évaluation comparative,

la comparabilité et la compatibilité des initiatives nationales en matière de capacité de calcul ainsi que de leurs impacts environnementaux, y compris en ce qui concerne l'IA. Des organisations comme l'OCDE pourraient contribuer à la mise en place d'un tel cadre.

2. **Il faudrait développer le recueil de données sur les impacts environnementaux de la capacité de calcul nécessaire à l'IA et des applications de cette dernière.** Davantage d'efforts doivent être faits pour recueillir des données environnementales au niveau des pays, des entreprises et des modèle d'IA. Les organismes et établissements nationaux et les acteurs du secteur privé devraient collecter davantage de données reposant sur la métrique de la durabilité comme les émissions de GES ou l'énergie, l'eau et les ressources naturelles consommées par la capacité de calcul nécessaire à l'IA et, chaque fois que possible, par les applications de l'IA.

3. **Il est difficile de séparer les mesures concernant spécifiquement l'IA de celles qui concernent la capacité de calcul à visées générales.** Distinguer entre la capacité de calcul utilisée pour l'IA et celle qui sert d'autres besoins informatiques scientifiques, mathématiques et généraux n'est pas une tâche aisée. Les pouvoirs publics, les instituts nationaux de statistique, les organisations intergouvernementales, le secteur privé, les milieux universitaires et les autres acteurs concernés devraient intensifier leurs efforts pour disjoindre les ensembles de données relatives aux infrastructures des TIC, estimer la part de celles qui sont utilisées pour l'IA et étudier la possibilité de recourir le cas échéant à une métrique de substitution.

4. **Les impacts environnementaux autres que la consommation d'énergie opérationnelle et les émissions de GES devraient aussi être pris en compte.** Les impacts environnementaux de la capacité de calcul nécessaire à l'IA allant au-delà de la consommation d'énergie et de l'empreinte carbone de la phase opérationnelle (c'est-à-dire production, transport et fin de vie) méritent d'être étudiés plus avant. Il s'agit notamment d'évaluer les effets sur la biodiversité et les impacts de la capacité de calcul nécessaire à l'IA sur d'autres limites planétaires (par exemple l'évolution des systèmes terrestres ou la consommation d'eau douce), mais aussi les conséquences directes des phases de fabrication, de transport et de fin de vie sur les ressources naturelles, ainsi que les impacts indirects des applications de l'IA.

5. **Des efforts s'imposent pour améliorer la transparence et l'équité environnementales, partout dans le monde.** La plupart des cadres et analyses applicables à la capacité de calcul nécessaire à l'IA sont élaborés par des experts issus de pays développés. Dans la mesure où l'on s'attend à ce que les incidences négatives sur l'environnement affectent de manière disproportionnée les pays en développement, les études qui vont être menées devraient viser en priorité à s'assurer que la capacité de calcul nécessaire à l'IA et les applications de cette dernière soient à même de promouvoir les objectifs de durabilité dans un large éventail de contexte nationaux, et à favoriser l'échange d'informations et de bonnes pratiques.

Acknowledgements

The paper "Measuring the environmental impacts of AI compute and applications: the AI footprint" is based on the work of the OECD.AI Expert Group on AI Compute and Climate (hereafter the "Expert Group") of the OECD Network of Experts, the OECD Secretariat, and was developed in consultation with experts involved in the Global Partnership on Artificial Intelligence (GPAI) Working Group on Responsible Artificial Intelligence (RAI) in 2022. It was prepared under the aegis of the OECD Working Party on Artificial Intelligence Governance (AIGO) and the OECD Committee for Digital Economy Policy (CDEP).

At the time of drafting, the Expert Group was co-chaired by Keith Strier (NVIDIA), Jack Clark (Anthropic) and Jennifer Tyldesley (United Kingdom Department of Digital, Culture, Media and Sport). Sana Khareghani (former United Kingdom Office for AI) and Satoshi Matsuoka (RIKEN Centre for Computational Science, Japan) were formerly co-chairs. Jonathan Frankle (MosaicML), Arti Garg (Hewlett Packard Enterprise) and David Kanter (ML Commons) provided support and guidance to the Expert Group. The GPAI RAI Working Group Project RAISE (Responsible AI Strategy for the Environment) was co-chaired by Raja Chatila (Professor emeritus, Sorbonne Université), Nicolas Miailhe (The Future Society), and Lee Tiedrich (Duke University). Marta Kwiatkowska (Oxford) was a former Project RAISE co-chair. Celine Caira, OECD Digital Economy Policy Division, and Johannes Leon Kirnberger, Consultant on AI and Climate, led the report development and drafting with input of Karine Perset, OECD Digital Economy Policy Division. Gallia Daor, Dirk Pilat, Audrey Plonk, and Andrew Wyckoff, OECD, provided advice and oversight.

The paper benefited significantly from the contributions of those associated with the Expert Group, including input and review by Juan Manuel Ahuactzin (ProMagnus Company), Luis Aranda (OECD), Leonidas Aristodemou (OECD), Aviv Balasiano (Technology Infrastructure in the Israeli Innovation Authority), Gregg Barrett (Cirrus AI), Arnaud Bertrand (ATOS), Pascal Bouvry (LuxProvide), Eliana Cardoso Emediato de Azabuja (Ministry of Science, Technology and Innovation), Landon Davidson (NVIDIA), David Elison (Lenovo), Maria Jose Escobar Silva (Universidad Técnica Federico Santa María), Rebeca Escobar (Federal Telecommunications Institute), Liliana Fernández Gómez (Digital Development Directorate - National Planning Department), Nicole Formica-Schiller (German AI Association, KI-Bundesverband), Garth Gibson (Vector Institute for AI), Alexia González Fanfalone (OECD), Lennart Heim (Centre for the Governance of AI), Taras Holoyad (Federal Network Agency for Electricity, Gas, Telecommunications, Post and Railway), Chen Hui (Infocomm and Media Development Authority), Vijay Janapa Redi (Harvard University John A. Paulson School of Engineering and Applied Sciences), Jan Jona Javoršek (Jožef Stefan Institute), Suzette Kent (Kent Advisory Services), Roland Krüppel (Federal Ministry for Education and Research), Jiwon Lee (Ministry of Technological Innovation and Digital Transition), Drew Lohn (Georgetown University Center for Security and Emerging Technology), Angus Macoustra (Commonwealth Scientific and Industrial Research Organisation), Uptal Manga (IBM), Ulrike Moetzel (Federal Ministry for Digital and Transport), Lorenzo Moretti (Ministry of Technological Innovation and Digital Transition), María Paula Mujica (High Presidential Advisory Office), Alistair Nolan (OECD), Marc-Etienne Ouimette (Amazon Web Services), Manish Parashar (National Science Foundation), Lynne Parker (United States Administration), Anand Rao (PwC), Ghilaine Roquet (New Digital Research Organization), José Gustavo Sampaio Gontijo (Department of Digital Science, Technology and Innovation), Jayne Stancavage (Intel), Dimitris Stogiannis (National Documentation Centre), Lila Tretikov (Microsoft), Bonis

Vasilis (National Documentation Centre), Ott Velsberg (Ministry of Economic Affairs), Verena Weber (OECD), Zee Kin Yeong (Infocomm Media Development Authority of Singapore), and Martin Zagler (Ministry of Industry, Business and Financial Affairs); and Project RAISE members, Eric Badiqué (European Commission), Peter Clutton-Brock (Radiance and Centre for AI and Climate), Virginia Dignum (Umea University), Cyrus Hodes (FoundersX) and David Rolnick (McGill University and Climate Change AI) (see Annex B), and all members of the GPAI RAI Working Group and the project team at the International Centre of Expertise in Montreal of Artificial Intelligence (CEIMIA).

The authors gratefully acknowledge the contributions made by individuals and institutions that took the time to participate in expert interviews (see Annex B), including Priya Donti (Climate Change AI), Lynn Kaack (Hertie School and Climate Change AI), George Kamiya (International Environmental Agency) and David Mytton (Console) for their input and review, as well as Mats Andersson (Lefdal Mine Datacenter), Yuval Bachar (EdgeCloudLink), Thyme Burdon (OECD), Greg Cannon (Amazon Web Services), Antoine Dechezlepretre (OECD), Lea Gimpel (Deutsche Gesellschaft für Internationale Zusammenarbeit), Conor Griffin (DeepMind), John Havens (Institute of Electrical and Electronics Engineers), Colin Herron (Global Water Partnership), Sasha Luccioni (Mila), Nestor Maslej (Stanford University Human-Centered Artificial Intelligence), Dirk Messner (German Environmental Agency), David Mytton (Console), David Patterson (Google), Andrew Prag (OECD), David Rejeski (Environmental Law Institute), Klaus Schimpf (Texas Instruments), Piyawut Srichaikul (Thai National Science and Technology Development Agency Supercomputer Center), Emma Strubell (Carnegie Mellon University), Sims Witherspoon (DeepMind), Chris Watkins (DeepMind), and Jeff Wittich (Ampere Computing). The authors also gratefully acknowledge the contributions made by those who took the time to complete the online survey.

Finally, the authors thank Misha Pinkashov and John Tarver for editing this report and Denisa Bencze, Angela Gosmann, Sebastian Ordelheide and Shellie Phillips for editorial support, the overall quality of the report benefited significantly from their engagement, and Elizabeth Thomas-Raynaud, Andreia Furtado and Kulani Abendroth-Dias, GPAI Secretariat, for administrative and communications support.

Table of contents

Foreword 2

Abstract 3

Abrégé 4

Executive summary 5

Résumé 7

Acknowledgements 10

Acronyms and abbreviations 14

1 Introduction 15

2 Definitions, methodology and limitations 18
- 2.1. What is AI compute? 18
- 2.2. Methodology 20
- 2.3. Limitations 20

3 Review of existing and emerging data and measurement frameworks 22
- 3.1. Direct environmental impacts of AI compute 22
- 3.2. Indirect environmental impacts of AI applications 31
- 3.3. Dual impacts 34

4 Measurement gaps with policy implications 35
- 4.1. Measurement standards for sustainable AI are needed 35
- 4.2. Data collection on the environmental impacts of AI compute and applications should be expanded 35
- 4.3. AI-specific measurements are difficult to separate from general-purpose compute 36
- 4.4. Environmental impacts beyond operational energy use and greenhouse gas emissions should be considered 36
- 4.5. Efforts are needed to improve environmental transparency and equity everywhere 37

5 Conclusion 38

References 39

Annex A. Direct and indirect environmental impact indicators for AI compute and applications 49

Annex B. Experts consulted 50

Annex C. Preliminary AI compute survey results 53

Notes 56

FIGURES

No table of figures entries found.

No table of figures entries found.

TABLES

No table of figures entries found.

BOXES

No table of figures entries found.

Acronyms and abbreviations

AI	Artificial Intelligence
AWS	Amazon Web Services
CO2e	Carbon-dioxide-equivalent
CPU	Central processing unit
CUE	Carbon usage effectiveness
ESG	Environmental, social and governance
GHG	Greenhouse gas
GPAI	Global Partnership on Artificial Intelligence
GPU	Graphics processing unit
ICT	Information and communication technology
IEA	International Energy Agency
IGO	Intergovernmental organisation
IPBES	Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services
IPCC	Intergovernmental Panel on Climate Change
IT	Information Technology
ITU	International Telecommunication Union
ODC	Organic data centre
OECD	Organisation for Economic Co-operation and Development
PUE	Power usage effectiveness
RAISE	Responsible AI Strategy for the Environment
SDG	Sustainable Development Goal
TWH	Terawatt hour
UNEP	United Nationals Environment Program
UNESCO	United Nations Educational, Scientific and Cultural Organization
WUE	Water usage effectiveness

1 Introduction

Artificial intelligence (AI) underpins some of the most promising technological solutions to today's global challenges, including climate change and environmental sustainability. While AI-enabled technologies can create economic efficiency gains and improve well-being, their creation and use should be responsible, trustworthy and support sustainable development (OECD, 2019[1]).

The world's leading environmental scientists agree that humanity is rapidly approaching and exceeding planetary boundaries. Increasingly frequent warnings of planetary emergencies occur as natural systems experience "emergent failures, tipping points and non-linearities" (OECD, 2020[2]) such as the rapid disintegration of the Antarctic ice sheet and resulting acceleration of climate change. The United Nations Environment Programme (UNEP) highlights the "triple planetary crisis" of climate change, biodiversity loss and pollution (UNEP, 2020[3]). Further, the Intergovernmental Panel on Climate Change (IPCC, 2021[4]), the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES, 2019[5]), and established scientific organisations agree that human influence has already caused unprecedented changes in the atmosphere, ocean and biosphere. These emergencies call for rapid and decisive action to stabilise atmospheric greenhouse gas (GHG) concentrations and decrease human pressure on the environment and the planet's ecosystems.

Economies and societies around the world face two emerging trends: the green and digital "twin transitions". Considering these transitions together can offer governments and societies an opportunity to leverage digital transformation for a green future. As a general-purpose technology with applications across sectors, AI can accelerate progress in many domains by creating efficiencies that decrease environmental impacts and lower emissions. However, the training and use of large-scale AI systems can also require massive amounts of processing power, memory, networking, storage and other resources – collectively known as "AI compute" – which can have significant environmental footprints from energy and water use, GHG emissions and end-of-life considerations. AI compute can also have indirect negative environmental impacts through its applications. To harness AI technologies to meet national and global sustainability goals, government, policy makers, academia and private sector actors need accurate and reliable measures of the environmental impacts of AI compute and applications. These include environmental impacts from production, transport, operations, and end-of-life considerations for AI compute, as well as environmental impacts from AI's application.

This realisation came into focus in the AI community in recent years, together with a sense of urgency to address this gap. Several initiatives have emerged to promote the sustainable use of AI and to propose AI regulation to incentivise sustainable AI applications and improve market uptake (European Parliament, 2021[6]). As the first intergovernmental standard on AI, the 2019 OECD Council Recommendation on Artificial Intelligence recommends that stakeholders "proactively engage in responsible stewardship of trustworthy AI in pursuit of beneficial outcomes for people and the planet, such as (…) protecting natural environments" (OECD AI Policy Observatory, 2019[7]). The OECD Framework for the Classification of AI Systems, launched in 2022, places people and the planet at the framework's centre (OECD, 2022[8]). Likewise, the Global Partnership on Artificial Intelligence (GPAI) published its responsible AI strategy for climate action with recommendations for governments to use AI in practice to address the climate emergency (GPAI, 2021[9]).

The UNEP-led Coalition for Digital Environmental Sustainability recognises the need for sustainable digitalisation, such as using planetary digital twins and promoting the circular economy (CODES, 2021[10]). UNESCO's 2021 Recommendation on the Ethics of AI recommends that governments "assess the direct and indirect environmental impact throughout the AI system life cycle" and not use AI systems when there are "disproportionate negative impacts on the environment".

To better understand and measure national AI compute capacity and its environmental impact, the OECD created a dedicated expert group on AI compute and climate (hereafter, "the Expert Group") to work towards developing a measurement framework and collecting data to help policy makers make evidence-based decisions (Box 1). The objective of GPAI's Responsible AI (RAI) Working Group is to contribute to the responsible development, use and governance of human-centred AI systems, including harnessing AI for climate action and biodiversity preservation led by GPAI's Project RAISE (Responsible AI Strategy for the Environment). Experts from the OECD Expert Group and GPAI helped develop this report as an overview of existing and emerging proxies, indicators, frameworks and tools to measure the environmental impacts of AI compute and applications.

Some researchers estimate that the computational capabilities required to train modern machine learning systems, measured in floating-point operations per second (FLOPS), has grown by hundreds of thousands of times since 2012 (OpenAI, 2018[11]). This is likely motivated by the increasing capabilities of large and more compute-intensive AI systems (Kaplan et al., 2020[12]; Hoffmann et al., 2022[13]). Research also notes that compute demands for AI systems, such as processing power, has grown faster than hardware performance, particularly for deep learning applications such as machine translation, object detection, and image classification (Thompson et al., 2020[14]). These trends are explored in further detail in additional OECD work informed by the Expert Group (OECD, forthcoming[15]).

Reducing the environmental impacts of AI is to some extent linked to reducing the environmental impacts of information and communication technology (ICT) systems more generally. For example, it is possible to decrease the environmental impact of data centres – which play an important role in AI's development and use, by integrating more energy-efficient server designs, connectivity architectures, cooling methods, and using 24/7 renewable energy sources. According to the International Energy Agency (IEA), several large network operators have significantly reduced their energy use by improving their networks' energy efficiency. For example, between 2014 and 2019 the telecommunications company Sprint reduced the energy intensity of its network by more than 80%, despite increasing demand, keeping its total network energy consumption flat (IEA, 2021[16]).

For AI specifically, advances in data science that lead to fewer training runs involving smaller data sets and less complex models can bring efficiencies more quickly than updating and modernising physical compute resources and infrastructure such as data centres. For example, researchers at the Massachusetts Institute of Technology (MIT) and at start-up MosaicML are training neural networks up to seven times faster by configuring AI algorithms to learn more efficiently (MosaicML, 2022[17]). Efficiency gains for both compute hardware and software, including algorithms, should be explored to maximise positive sustainability impacts in training and using AI systems.

This report hopes to inspire dialogue between policy makers, researchers, companies and others to agree on indicators and good measurement practices. This evidence base aims to support an improved collective understanding of the impacts of AI compute and applications to help measure and decrease AI's negative environmental effects while enabling its potential to accelerate action for the good of the planet.

Box 1. The OECD Network of Experts on AI and OECD.AI Expert Group on AI Compute and Climate

The **OECD Network of Experts on AI** (ONE AI) provides policy, technical and business input to inform OECD analysis and recommendations. As a multi-disciplinary and multi-stakeholder group, ONE AI also provides the OECD with an outward perspective on AI, serving as a platform for the OECD to share information with other international initiatives and organisations. ONE AI raises awareness about trustworthy AI and sustainability issues, amongst other policy initiatives, particularly where international co-operation is useful.

The **OECD.AI Expert Group on AI Compute and Climate** (hereafter “Expert Group”) advances understanding of AI compute and helps countries build awareness and work towards closing “compute divides” within and between countries. The Expert Group aims to provide actionable and user-friendly evidence on AI compute, including its environmental impacts. In doing so, it seeks to enable policy makers to evaluate current and future national AI compute needs and corresponding capacity.

An AI compute divide can manifest *within* countries between the private sector and academia, as private sector actors often have greater resources and access to AI compute to advance their objectives. An AI compute divide can also manifest and worsen *between* countries, namely between advanced and emerging economies, if governments cannot make informed decisions about investments to fulfil their national AI plans. This opens a gap in countries’ ability to compute the complex AI models that lead to productivity gains in a global digital economy.

The Expert Group aims to support policy makers and practitioners in developing tools and indicators that are measurable at the national level and enable sufficient geographic coverage for benchmarking to take place. Future recommendations resulting from its work will endeavour to be comprehensive, accessible to both technical and non-technical audiences, and dynamic and time-proof, allowing for evolution as compute hardware and software advance (e.g. faster processors, larger memory, next generation networks, quantum computing, etc.).

The Expert Group is co-chaired by Keith Strier (Vice President of Worldwide AI Initiatives at NVIDIA), Jack Clark (Co-Founder of Anthropic) and Jennifer Tyldesley (Deputy Director of Economic Security at the Department of Digital, Culture, Media and Sport, United Kingdom). Sana Khareghani (former Head of the Office for AI, United Kingdom) and Satoshi Matsuoka (Director, RIKEN Centre for Computational Science, Japan) were co-chairs from October 2021 to early 2022. The Expert Group has met virtually every 3-4 weeks since April 2021.

Source: OECD.AI Expert Group on AI Compute and Climate

2 Definitions, methodology and limitations

2.1. What is AI compute?

Alongside data and algorithms, access to computing resources fit for AI is a key enabler for its advancement and diffusion (Figure 1). While data and algorithms receive significant attention in policy circles at the OECD and beyond, the computing resources that make these advancements possible have been largely overlooked. Understanding a country’s needs and capacity for AI compute, and its relationship to AI diffusion, can support policy makers and practitioners in formulating national AI policies and ensuring they have the AI compute necessary to implement national AI plans.

Figure 1. Examples of AI enablers

Source: OECD.AI Expert Group on AI Compute and Climate

While awareness of AI compute as a priority in national AI strategies is growing, its technical nature means that it is poorly understood outside specialised policy communities. Countries are increasingly questioning what environmental impacts result from operating large-scale AI models given their energy and water requirements, but they lack clear and standardised indicators and benchmarks to guide sustainability decisions for public- and private-sector investments in AI. This is why this report focuses on the sustainability impacts of AI compute, rather than the sustainability dimensions of data and algorithms. To take a holistic approach, the report also examines the environmental impacts of AI applications, although this area can be more difficult to measure given the variety of AI applications across products and services.

According to the 2019 OECD Recommendation of the Council on Artificial Intelligence, an AI system is “a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments.” While AI can be perceived as an abstract, non-tangible technical system, it is grounded in and enabled by concrete physical infrastructure and hardware, which requires significant amounts of raw materials and natural resources.

The Expert Group found that, while there is no widely used standard definition of AI compute, its core elements are understood by technical AI experts, developers and practitioners. The Expert Group thus proposes a definition of AI compute that would be accessible to both technical and policy communities: “AI computing resources ('AI compute') include one or more "stacks" (i.e. layers) of hardware and software used to support specialised AI workloads and applications in an efficient manner" (Box 2).

Beyond this general definition, this report refers to specific aspects of AI compute. *AI compute infrastructure* refers to physical systems (i.e. hardware, data centres, etc.) necessary for the development and application of AI. *AI compute capacity* refers to the totality of computing resources that can be used for AI, which can include key considerations that make this capacity effective, such as talent and skills. *AI applications* examine the indirect environmental impacts of deploying AI systems, which implicitly includes AI compute, for example from AI-generated predictions, recommendations or decisions applied in a given context. The *AI compute resources lifecycle* tracks the direct environmental impacts of production, transport, operations and end-of-life considerations for AI compute (Figure 3).

Box 2. Defining and scoping AI compute

Between April 2021 and April 2022, the expert group conducted eight meetings and launched a public survey to inform analytical work on defining and scoping the definition of AI compute. The proposed definition is based on these meetings, interviews with more than 25 experts and survey responses:

> *"AI computing resources ('AI compute') include one or more stacks of hardware and software used to support specialised AI workloads and applications in an efficient manner."*

This definition highlights several key properties central to a common understanding of AI compute:

- **AI compute includes stacks of hardware and software.** AI workloads are not performed by one hardware or software component, but instead one or more "stacks" (i.e. layers) of components. The stacks include storage, memory, networking infrastructure and more, designed to efficiently support AI-specific workloads and applications that run mathematical calculations and process data at scale. The efficient interaction between the hardware and software stacks is crucial for AI compute.
- **AI compute stacks are specialised for AI workloads.** AI training and use are enabled by specialised hardware. For example, graphics processing units (GPUs) are purpose-built for highly parallelised computing, which means that many calculations are carried out simultaneously, making them highly efficient for certain AI model types such as deep learning. AI compute stacks are becoming increasingly specialised, as AI applications, the number of parameters and dataset sizes continue to grow.
- **AI compute requirements can vary significantly.** Depending on the application, AI system lifecycle stage and size of the system, the AI compute needed can vary, from large High-Performance Computing (HPC) clusters or compute hyperscale cloud providers, to smaller data-science laptops and workstations. Consequently, compute requirements will vary significantly based on a country's national AI plans and across the AI system lifecycle.
- **AI compute supports AI workloads and applications in an efficient manner.** AI compute differs from general-purpose compute in that it is capable of supporting AI workloads and applications in an efficient manner (e.g. through optimised execution time and energy usage). This efficiency is critical for conducting AI R&D, using large models and datasets.

Source: OECD.AI Expert Group on AI Compute and Climate

2.2. Methodology

Building on the Expert Group's work to define and scope AI compute, this report synthesises findings from a literature review, public survey and interviews with more than 25 experts to analyse the environmental impacts along the AI compute resources lifecycle and for AI applications. In doing so, this report differentiates between direct and indirect environmental impacts, based on a 2001 report to the OECD on the impacts of ICT on environmental sustainability (Berkhout and Hertin, 2001[18]), reflected in the International Telecommunication Union (ITU) Standard ITU-T L.1410, and Kaack et al.

Direct environmental impacts relate to first-order[1] environmental effects from the AI compute resources lifecycle, which examines impacts from the production, transport, operations and end-of-life impacts of AI compute. *Indirect* environmental impacts relate to second- and third-order environmental effects resulting from the application of AI, including positive impacts on climate action and negative impacts through induced consumption or rebound effects.

Figure 2. Direct and indirect environmental impacts of AI compute and applications

Direct environmental impacts AI compute resources lifecycle

Production	Transport	Operations	End-of-life
• Raw material extraction • Assembly • Manufacturing	• Distribution • Freight transportation • Handling & storage	• Energy consumption • Water consumption • Carbon footprint	• Collection & shipping • Dismantling & recycling • Waste disposal

Indirect environmental impacts AI applications

Positive impacts	Negative impacts
• Beneficial sectoral applications • Climate mitigation and adaptation • Environmental modelling and forecasting	• Harmful sectoral applications • Carbon leakage (net increase in emissions) • Consumption patterns and rebound effects

Source: OECD.AI Expert Group on AI Compute and Climate, literature review, expert interviews

2.3. Limitations

The research for this report revealed several limitations to evidence-based analysis. First, data on the environmental impacts of AI compute is not widely available in a standardised and validated form. As such, this analysis is largely based on existing and publicly available data, and peer-reviewed academic papers. This data limitation is particularly acute for measurements of AI compute water consumption and full lifecycle impacts, as these are currently underexplored and underreported. Second, because the market for AI compute resources is concentrated in a handful of hardware, software and cloud computing companies (Ahmed and Wahed, 2020[19]), disaggregated data on the environmental impacts of AI compute can be difficult to access and viewed as proprietary information in some cases.

Third, data availability limitations also constrained the report's analysis from differentiating between the compute needs of different AI systems, such as symbolic AI or machine learning. Fourth, the report also does not consider in depth the compute needs for processing and cleaning data for AI model training, which occurs at earlier stages of AI training and use. The Expert Group could further examine the compute needs of data for AI in its future work. Lastly, the report does not consider in depth the environmental impacts of edge and Internet-of-Things (IoT) connected devices in the calculation of national capacity as further research is needed to establish what share of AI compute comes from these areas.

A public survey was launched as part of the report's data collection efforts (Annex C). This yielded a sample size of 118 responses as it was targeted to an audience with technical expertise or knowledge of AI compute. The sample size could be expanded as awareness about AI compute and its challenges grows. Future analysis could also benefit from the further participation of government representatives, private-sector entities and academia in systematic data collection efforts. This could be considered in future work following the development of a measurement framework, including indicators and proxies for AI compute and applications.

3 Review of existing and emerging data and measurement frameworks

This section takes stock of existing data and measurement frameworks related to the environmental impacts of AI compute and applications. It distinguishes between direct and indirect environmental impacts, as described in Figure 2. Analysis of direct environmental impacts is guided by the four stages of the AI compute resources lifecycle (Figure 3). The third stage of the lifecycle (AI compute operations) explores the impact of AI compute across the sustainability metrics of energy and water consumption, and GHG emissions. Finally, existing and emerging data and measurement frameworks related to positive and negative indirect environmental impacts resulting from the application of AI, are also explored outside this framework, along with a discussion on dual impacts.

3.1. Direct environmental impacts of AI compute

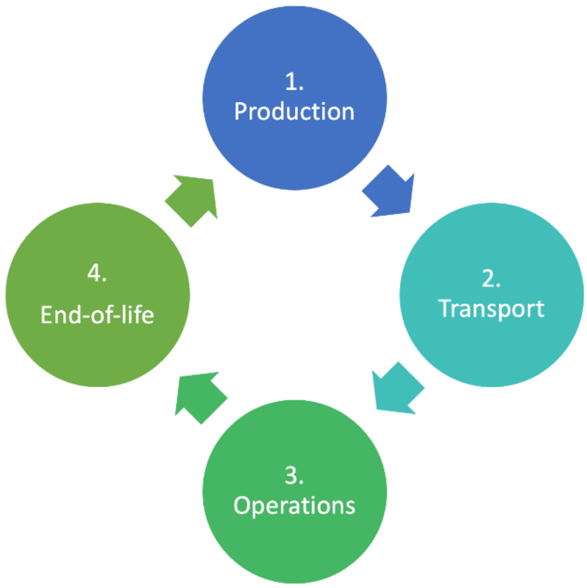
Most indicators and frameworks focus on the direct environmental impacts of AI compute, as opposed to indirect impacts resulting from AI applications. These direct environmental impacts occur along the AI compute resources lifecycle: (1) production, (2) transport, (3) operations and (4) end-of-life (Figure 3). Within this lifecycle, researchers and practitioners largely have focused on impacts from the operations stage where AI compute is used to train or deploy AI systems. While AI applications have both positive and negative impacts, the direct environmental effects of AI compute are largely negative in terms of GHG emissions and resource consumption, with compute infrastructure sometimes requiring large amounts of energy and other material inputs (Barteková and Börkey, 2022[20]).

Guided by this framework, analysis of existing literature revealed that in the first, second and fourth stages of the AI compute resources lifecycle (production, transportation and end-of-life), most indicators and measurements relate to overall ICT equipment and hardware, which are not necessarily AI-specific. AI-specific indicators and measurements were only found for the third stage (operations), in particular related to energy consumption and GHG emissions (Rohde et al., 2021[21]; Sustainability Index for Artificial Intelligence, 2022[22]).

It is important to note that the AI resources lifecycle is a distinct framework from the AI system lifecycle as defined in the 2022 OECD Framework for the Classification of AI Systems, which defines the AI system lifecycle through the following phases: (1) planning and design; (2) collecting and process data; (3) building and using models; (4) verifying and validating the model; (5) deployment, and (6) operating and monitoring the system (OECD, 2022[23]). These frameworks are distinct as they examine different aspects of AI, with the former focusing on environmental impacts AI compute from a natural resources perspective, and the latter focusing on the stages of AI development, application and monitoring.

In addition to the separate stages of the AI compute resources lifecycle, it is important to assess direct environmental impacts from their interconnection, to provide a comprehensive analysis (Wu et al., 2022[24]). An example of such assessment is the Green Cloud Computing methodology from the German Environmental Agency, which looks at the following four impact categories: (1) abiotic resource depletion (i.e. use of minerals and fossil fuels); (2) cumulative energy demand (i.e. use of renewable and non-renewable energy); (3) global warming potential (i.e. impact on climate change); and (4) water consumption (UBA, 2021[25]). Full general lifecycle assessments (as defined in ISO standards 14040:2006 and 14044:2006) could be developed for individual AI compute components, such as the different products that make up AI compute stacks, to assess their environmental impact across lifecycle stages. However, applying existing tools like product lifecycle assessments can be challenging due to the distinct compute needs of AI systems requiring stacks of hardware and software that can vary significantly in their composition and include many different individual and highly specialised products.

Figure 3. Four stages of the AI compute resources lifecycle



Note: This framework is distinct from the AI system lifecycle defined in the 2022 OECD Framework for the Classification of AI Systems as: (1) planning and design; (2) collecting and processing data; (3) building and using models; (4) verifying and validating the model; (5) deployment and (6) operating and monitoring the system (OECD, 2022[23]).
Source: OECD.AI Expert Group on AI Compute and Climate

3.1.1. Production

AI compute production relies on the physical extraction and consumption of natural resources to build computing hardware, including computer chips, semiconductors, graphics processing units (GPUs) and central processing units (CPUs). There are many steps in the production of AI compute hardware and infrastructure, from mining, smelting and refining, to component manufacturing, such as semiconductor fabrication, and assembly. Environmental impacts along this value chain include soil contamination, deforestation, erosion, biodiversity degradation, toxic waste disposal, groundwater pollution, water use, radioactive waste and air pollution (Crawford, 2018[26]). However, few frameworks and indicators differentiate between the computing resources used uniquely for AI and those used for other scientific, academic or industrial applications. Often, data is only available for total ICT infrastructure production and is not disaggregated for AI uses and applications.

Scientific literature reveals different estimates of how much the production stage adds to the environmental impacts of AI compute. Often, production impacts are overlooked due to difficulties in attribution (Henderson et al., 2020[27]), with some scholars arguing that lifecycle assessments would be “impractical at scale and would greatly reduce who can use the method” (Lannelongue et al., 2021[28]). Another reason given for excluding the production stage of AI compute from analysis is its presumed insignificance compared to operational impacts (i.e. stage three of the AI compute resources lifecycle) (Siddik, Shehabi and Marston, 2021[29]). Others underscore that adopting renewable energy procurement and improving energy efficiency in data centres could result in production representing a higher portion of AI compute's environmental footprint over time (Gupta et al., 2020[30]). Non-operational impacts, such as those resulting in the production stage, will thus become relatively more important (IEA, 2021[16]). At present however, many AI compute providers only report operational environmental impacts, excluding environmental impacts from production, which are difficult to estimate.

One estimate of the carbon footprint from the production of data centres globally is 20 megatons of carbon-dioxide-equivalent (CO2e), representing about 15% of total data centre GHG emissions in 2015 (Malmodin and Lundén, 2018[31]). The company Meta estimates that GHG emissions from the production of their data centres are around 30% of total company emissions for 2022 (Wu et al., 2022[24]). But as data centres shift to carbon-free energy sources, their long-term operation will account for fewer emissions and some estimates project the share resulting from data centre production could eventually increase to over 80% (Gupta et al., 2020[30]).

Value-chain GHG emissions arising up- and down-stream of an entity's immediate activities (known as "Scope 3 emissions"[2]) can account for a large proportion of the carbon footprint of cloud computing vendors. According to Meta, their share of value-chain emissions compared to total GHG emissions rose from 44% in 2017 to 99% in 2020 (Meta, 2021[32]), resulting in a "significant embodied carbon cost paid upfront" for new AI-driven system components for data centres (Wu et al., 2022[24]). Similarly, the share of value-chain GHG emissions according to Google increased from 45% in 2016 to over 90% in 2020 (Google, 2022[33]). These increases in value-chain emissions are likely due to higher clean-energy procurement, which lowers the proportion of emissions from energy use. It should be noted that a carbon footprint analysis alone does not reflect the full environmental impact of AI compute production, including the mining of raw materials that can be environmentally damaging and carry significant supply and human rights risks (Crawford, 2021[34]). The environmental and social impacts of extracting raw materials like cobalt, palladium, copper, lithium and aluminium are rising on the agenda of policy makers (European Parliament, 2021[6]) and industry networks (Sustainable Infrastructure Alliance, 2022[35]).

Indicators

Metrics to consider at the production stage of the AI compute resources lifecycle:

- GHG emissions from production, in metric tons of CO2e
- the carbon intensity of production methods, in metric tons of CO2e per unit (e.g. per person or dollar of revenue)
- the share of renewable energy used in production
- the share of recycled or renewable materials used in production

3.1.2. Transportation

As with production, the environmental impacts of transporting AI compute hardware are difficult to disaggregate from those of transporting other ICT hardware, and even from the transportation of non-ICT goods in general. Transport activities can generate various negative environmental impacts, such as air pollution, oil spills, toxic-waste discharges and acoustic pollution (Crawford, 2018[26]; OECD, n.d.[36]). Global distribution, freight transportation, handling and storage depend on fossil fuels, with road-freight vehicles responsible for 2.4 gigatons of CO2e per year (around 6% of global energy-related emissions), and with shipping and aviation each responsible for about 1 gigaton of CO2e per year (around 2.5% of global energy-related emissions) (IEA, 2021[37]).

But only a small fraction of global transportation figures can be attributed to ICT hardware, and an even smaller portion can be attributed to the transportation of AI compute hardware, such as specialised GPU racks. The few estimates available put transport of AI compute hardware at less than 5% of total GHG emissions over an AI system's lifetime, with some estimates as low as less than 1% (Gupta et al., 2020[30]). The differences in estimates could be attributed to the location of hardware production and the corresponding transport distance.

The decarbonisation and sustainability of transportation is a global undertaking. While transportation of AI compute hardware has a small impact compared to other lifecycle stages, opting for sustainable and environmentally responsible transportation methods should be further explored and considered.

Indicators

Metrics to consider at the transportation stage of the AI compute resources lifecycle:

- GHG emissions from transportation of AI compute hardware in metric tons of CO2e
- the carbon intensity of the transport methods in metric tons of CO2e per unit (e.g. per person or dollar of revenue)
- the share of low-carbon energy and/or renewable energy used in these transport methods

3.1.3. Operations

The environmental impacts of operating AI compute primarily relate to energy consumption, GHG emissions and water consumption, which can occur when an AI system is developed, such as through training, and deployed, for example by applying AI models to make predictions, recommendations or decisions (also known as inference). These three environmental impacts are further explored below.

Energy consumption

Compared to the production and transport stages of the AI compute resources lifecycle, energy consumption is well documented for global data centres and ICT at large. Although a range of estimates exist (Banet et al., 2021[38]), researchers and institutions arrive at fairly consistent results. The IEA estimates global data centre electricity demand at 194 Terawatt Hours (TWh) in 2014, or 1% of global electricity demand (IEA, 2017[39]). By 2020, due to large efficiency gains, that estimate had only risen to 200-250 TWh and remained at 1% of global electricity demand (IEA, 2021[16]). Other experts estimate worldwide energy demand by data centres in a similar range: at 190 TWh in 2020, with most coming from cloud and hyperscale operators (Sönnichsen, 2021[40]). Masanet et al. estimate demand at 205 TWh in 2018, or 1% of global electricity consumption, and just a 6% increase from 2010 despite a 550% increase in data centre compute demand over the same time period (Masanet et al., 2020[41]).

Notably, efficiency improvements and the shift to large hyperscale data centres have offset much of the exponential growth in data centre services in the past decade, keeping the estimated energy use of data centres almost flat (IEA, 2021[16]). Further efficiency could be achieved through specialised AI hardware and network architecture, such as breakthroughs in energy efficiency using ultra-low-energy artificial neurons (IEEE, 2022[42]). However, with technological changes in ICT and the acceleration of digital transformation during the COVID-19 pandemic, this trend might vary in the future. Some estimate that data centres could account for 783 TWh by 2030, or around 2.6% of global electricity use which is estimated to reach 30 000 TWh in 2030 (Andrae, 2020[43]).

At the national level, few governments track energy consumption and projected demand by ICT infrastructure and related AI compute resources within their borders. In Denmark, annual demand by data centres is expected to grow from near-0 in 2020 to 5 TWh by 2025 and 7.5 TWh by 2030 (Danish Energy

Agency, 2020[44]). In Ireland, a European hub for data centre operators, electricity consumption by data centres increased 144% from 2015 to 2020, accounting for 11% of metred electricity consumed in the country in 2021 (Central Statistics Office, 2022[45]). Median-demand scenarios estimate that this figure could rise as much as 23% by 2030 (EirGrid, 2021[46]). In the Netherlands, data centre energy use grew from 1.6 TWh in 2017 to 2.7 TWh in 2019, representing 2.7% of national electricity supply (CBS, 2021[47]). In the United States, data centres accounted for 70 TWh or 1.8% of total electricity consumption in 2014 (Shehabi et al., 2016[48]). Estimates for total electricity consumption by data centres in the European Union (EU27) in 2018 range between 42 TWh (European Commission, 2022[49]) and 54 TWh for EU27 countries in addition to the United Kingdom (Montevecchi et al., 2020[50]), or 2-3% of overall electricity consumption.

A Greenpeace report puts Chinese data centre electricity consumption at 161 TWh in 2018, of which 75% came from coal-fired power stations (Greenpeace China, 2021[51]). The Chinese State Grid Energy Research Institute reported that the People's Republic of China's (hereafter "China") data centre energy consumption at 200 TWh in 2020. However, the details of whether such estimates include 5G networks and bitcoin mining are unclear (Chinese State Council, 2021[52]). If accurate, this could indicate that current estimates of global data centre energy consumption are lower-bound estimates, and real energy consumption could be higher (Hintemann, 2018[53]).

Around the world, recent reports have pointed to the demand for data centres contributing to increased pressure on local energy grids. There are reports of new housing developments being halted due to the high electricity demands of data centres (Financial Times, 2022[54]), while some jurisdictions consider a moratorium on new data centre construction due to strained national power supplies and energy grid constraints (The Times, 2022[55]). Climate change and more frequent heatwaves are also reported to add stress on power grids and data centres, which can lead to outages (Google Cloud, 2022[56]).

Many data centre operators report energy consumption, including total electricity consumption, energy efficiency and power usage effectiveness (PUE), renewable energy procurement, facility energy use, and renewable energy consumption (Apple, 2021[57]; Meta, 2021[32]; Google, 2022[33]; Microsoft, 2021[58]; Digital Realty, 2019[59]; Equinix, 2021[60]). ICT companies and large operators are major purchasers of renewable energy and account for almost half of global corporate renewable energy procurement (IEA, 2021[16]). Worldwide associations like the Energy Efficient High Performance Computing Working Group and the Center of Expertise for Energy Efficiency in Data Centers, aspire to drive energy efficiency and share best practices (EEHPCWG, 2022[61]; Center of Expertise for Energy Efficiency in Data Centers, 2022[62]).

Several approaches and frameworks estimate energy consumption by machine learning models. These allow for a more granular assessment of dedicated AI models than do global estimates. The use of standard metrics for reporting efficiency, such as training time and hyperparameter sensitivity, could improve comparability between models (Strubell, Ganesh and McCallum, 2020[63]). The Green500 list ranks the top 500 most powerful supercomputers from around the world according to their energy efficiency (Green500, 2021[64]). Others propose Bayesian deep learning (Welling, 2018[65]) and sustainable federated learning (Guler and Yener, 2021[66]) to make AI more power-efficient. Future machine learning models will need to balance performance and efficiency as demand for AI compute grows with the creation of more complex and large AI models (Desislavov, Martínez-Plumed and Hernández-Orallo, 2021[67]).

Box 3. Optimising energy use to reduce direct environmental impacts at DeepMind AI

Over 40% of DeepMind AI's data centre energy use comes from cooling systems. AI can help operators make data centres more efficient, decreasing energy consumption and resulting GHG emissions. DeepMind's reinforcement learning algorithm was applied at one of Google's data centres and is reported to reduce energy use by 30-40% compared to cooling systems without optimisation.

Source: (DeepMind, 2016[68]; DeepMind, 2018[69]; Zhang et al., 2017[70])

Indicators

Energy consumption metrics to consider at the operations stage of the AI compute resources lifecycle:

- electricity consumption in TWh
- renewable electricity consumption in TWh
- Power Usage Effectiveness (PUE) for total facility power compared to ICT equipment power

Greenhouse gas emissions

The operational carbon footprint of AI compute relates directly to its energy consumption, often from non-renewable energy sources. The global ICT industry (including hardware such as televisions) is estimated to be responsible for 1.8-2.8% of global GHG emissions, while other calculations estimate it as high as 2.1-3.9% (Freitag et al., 2021[71]). A more suitable estimate (excluding televisions) puts the ICT-industry figure at 700 metric tons of CO_2 in 2020, or 1.4% of global emissions (ITU, 2020[72]).

Examining trends for ICT emissions, some researchers claim GHG emissions have plateaued as "footprints are significantly smaller than previously forecasted" (Malmodin and Lundén, 2018[31]). Differences in estimates can be explained by underlying assumptions about the future global energy supply mix, as projections depend largely on the carbon intensity of electricity supply and the availability of renewable energy sources. The ITU and the World Benchmarking Alliance (WBA) estimate that the 150 largest digital companies account for 1.6% of global electricity use, of which 32% comes from renewable sources (International Telecommunication Union and World Benchmarking Alliance, 2022[73]). In the United States, around 0.5% of national emissions are estimated to be from data centres (Siddik, Shehabi and Marston, 2021[29]). More broadly, cloud and hyperscale data centres are estimated to account for 0.1-0.2% of global emissions (Kaack et al., 2022[74]).

AI compute is estimated to account for only a fraction of these numbers. However, few estimates of AI's energy demand exist, with these estimates rarely differentiating between AI training and use (or "inference") workloads. According to one study by Google, its overall energy use for machine learning workloads consistently represented less than 15% of total energy use over the three-year period from 2019 through 2021 (Patterson, 2022[75]). Other estimates use customer spending to approximate the percentage of compute used between AI training and inference workloads. For example, a large cloud compute provider[3] estimates that its enterprise customers spend between 7-10% of their total compute infrastructure expenditure on supporting AI and machine learning applications, broken down between training and use ("inference") at approximately 3-4.5% for machine learning training and 4-4.5% for machine learning inference. This includes about 60% spent on compute platforms featuring hardware accelerators like GPUs, and about 40% spent on CPU-based compute platforms. Such numbers can inform estimates of

AI-specific energy use and associated GHG emissions, while shedding light on how impacts differ according to whether compute is used for AI training or inference.

Many of the world's largest data centre operators announced plans to power them with 100% renewable energy (Dhar, 2020[76]). Amazon announced its intention to power all company operations with renewable energy by 2025 and become carbon-neutral by 2040 (Amazon, 2022[77]). In 2021, it reported to source 85% of its energy from renewable sources across its business, including AWS (Amazon, 2022[77]). Google reports on its emission inventory, compensation for emissions, and carbon intensity, matching 100% of annual electricity use with renewable energy since 2017. Annual electricity matching means purchasing renewable energy to equal annual energy consumption through power-purchase agreements. As annual electricity matching can still contain a significant share of non-renewable resources, Google committed to run on 24/7 carbon-free energy by 2030, meaning that every kilowatt-hour of electricity use is met with carbon-free electricity sources (Google, 2022[33]).

Google also introduced a policy roadmap to achieve completely carbon-free energy (Google, 2022[78]) and a system for "carbon-intelligent" compute management, which generates "carbon-aware" capacity estimates for its data centres (Radovanovic et al., 2021[79]). Similarly, Meta reports on GHG emissions and carbon intensity and announced it has been running carbon-neutral operations as of 2020. Meta plans to achieve net-zero emissions throughout its supply chain by 2030 (Meta, 2021[32]). It also implemented designs for "carbon-aware" data centres under its Carbon Explorer Framework (Acun et al., 2022[80]). Microsoft reports on its emission inventory and carbon-neutrality offsets and aims to become carbon negative by 2030 (Microsoft, 2021[58]).

Major cloud computing vendors offer tools for their customers to estimate carbon-footprints based on their cloud compute usage. AWS introduced a carbon-footprint tool to help customers understand and forecast emissions of cloud services (Amazon Web Services, 2022[81]). This includes a shared-responsibility model in which they provide sustainable infrastructure, efficient cloud usage and workloads while customers are encouraged to run their models efficiently (Amazon Web Services, 2021[82]). Google provides carbon-footprint reports based on the Greenhouse Gas Protocol carbon accounting standards (Google Cloud, 2022[83]). Microsoft offers an emissions impact dashboard showing direct and indirect GHG emissions of its cloud services (Microsoft, 2022[84]). Third-party providers also offer free and open-source tools to estimate cloud GHG emissions (Cloud Carbon Footprint, 2022[85]). It is important to note that these customer-facing tools only estimate operational emissions from AI training and use, and not embodied emissions from the entire AI compute resources lifecycle, such as emissions created during hardware and infrastructure production, transportation and end-of-life.

A number of researchers put forward tools and frameworks to estimate the carbon impact of AI models, such as the Machine Learning Emissions Calculator (Lacoste et al., 2019[86]), a free online tool called Green Algorithms (Lannelongue et al., 2021[28]) and a software package that can be integrated into a Python codebase (CodeCarbon, 2022[87]). One framework, called the experiment-impact-tracker, provides an interface for real-time tracking of GHG emissions and explores how regional energy-grid differences can reduce emissions by up to 30% using low-carbon energy sources (Henderson et al., 2020[27]). Researchers also put forward best practices, including the use of pre-trained models and transfer-learning, where appropriate; efficient machine learning model architectures; processors optimised for machine learning, cloud computing and usage optimisation; and shifting operations to data centres with low-carbon energy availability (Patterson et al., 2021[88]).

Indicators

GHG emissions metrics to consider at the operations stage of the AI compute resources lifecycle:

- GHG emissions in metric tons of CO_2e
- carbon intensity in metric tons of CO_2e per unit (e.g. per person or dollar of revenue)
- carbon usage effectiveness (CUE): the ratio of total CO_2e emissions caused by total data centre energy consumption to the energy consumption of ICT equipment

Water consumption

While many discussions about sustainable compute revolve around energy efficiency and zero-carbon operations, freshwater consumption is a resource with major, often-overlooked environmental impact (Heslin, 2016[89]). AI compute hardware and infrastructure consume water in two major ways: (1) directly, for cooling and (2) indirectly through water use for electricity generation. The production stage of AI compute, such as semiconductor fabrication, can also use large amounts of water. Compared to operational energy use and GHG emissions, water consumption is poorly understood. Only about 33-50% of data centre operators compile and report water-use metrics (Mytton, 2021[90]; Uptime Institute, 2021[91]; Google Cloud, 2022[83]) such as water withdrawal minus water consumption, or water returned to the local water system following use (Microsoft, 2021[58]; Meta, 2021[32]; Siddik, Shehabi and Marston, 2021[29]).

In the United States, data centres are estimated to account for less than 1% of total water consumption (Mytton, 2021[90]). However, they compete with users such as hospitals or agricultural production. The United States data centre industry draws directly and indirectly from 90% of national watersheds and is one of the top ten water-consuming industries in the country. Data centres often cluster in similar geographic areas and many rely on scarce water supplies, particularly in the western United States (Siddik, Shehabi and Marston, 2021[29]).

Water consumption increasingly features in debates around the sustainability of AI compute. European industry associations for data centres and cloud infrastructure list water conservation as a priority (Climate Neutral Data Centre Pact, 2022[92]). Data centres based on liquid cooling could also recover the excess heat for on-site GHGs, nearby buildings or local heating, in what some call an Organic Data Centre (ODC) approach (Karnama, Haghighi and Vinuesa, 2019[93]).

Box 4. Efficient water and energy use to minimise environmental impacts: Lefdal Mine Datacenter

Located inside a former mine in the Sogn og Fjordane region of Norway, the Lefdal Mine Datacenter offers a case study in sustainable data centre architecture that minimises the direct environmental impacts of compute. Using 100% renewable hydropower electricity, the data centre guarantees a PUE of 1.15, close to the generally accepted ideal PUE of 1.0 (ratio of total energy used to the energy delivered to computing equipment). Notably, Lefdal Mine uses very little water in its operations. It also reuses the heat generated by cooling solutions based on the supply of cool sea water. As a signatory of the Climate Neutral Data Centre Pact and the iMasons Climate Accord, its data centre is on track to become fully carbon-neutral.

While its unique location and geographic advantages are hard to replicate, it is an example of the Norwegian Data Centres Strategy goal to become “the world’s most sustainable data centre nation” by highlighting the potential of renewable energy, low water usage and ways to re-use excess heat.

Source: (Lefdal Mine Datacentre, n.d.[94])

Indicators

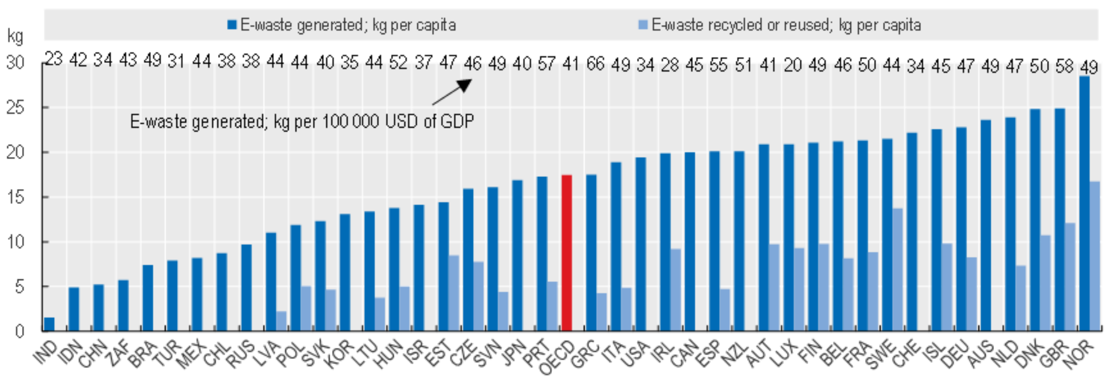
Water-consumption metrics to consider at the operations stage of the AI compute resources lifecycle:

- water withdrawal in cubic meters
- water consumption in cubic meters
- water discharge (quality) in cubic meters
- water usage effectiveness (WUE) in litres per kilowatt-hour
- water withdrawal intensity in cubic meters per unit (e.g. per person or dollar of revenue)

3.1.4. End-of-life

The resources lifecycle of AI compute ends with recycling or disposing of electronic waste (e-waste). The collection, shipping, recovering and disposal of AI compute hardware has environmental and social impacts such as air pollution, acidic and radioactive waste, groundwater pollution, and more (Crawford, 2018[26]). Much of global e-waste disposal is conducted in developing countries, adding to their environmental and social challenges (Forum, 2019[95]). The OECD tracks e-waste, defined as all items of electric or electronic equipment discarded by their owner, and national recycling levels (OECD, 2019[96]). There is a significant gap between e-waste generated and recycled or reused (Figure 4).

Figure 4. E-waste generation and recycling or reuse, 2016



Note: This figure does not distinguish between AI and non-AI e-waste. The figure covers six waste categories: (1) temperature equipment; (2) screens, monitors; (3) lamps; (4) large equipment; (5) small equipment; and (5) small IT and telecommunication equipment. Disaggregating these categories could better estimate AI compute-related impacts.
Source: (OECD, 2019[96]), see Statlink: https://doi.org/10.1787/888933931086

As in the production and transport stages of the AI compute resources lifecycle, most indicators and measurements of end-of-life impacts relate to overall ICT equipment and hardware. ICT infrastructure accounts for about 12 million tons, or 25% of total global electronic waste (Sustainable Infrastructure Alliance, 2022[35]). Increasingly, used AI compute equipment is also sold, repurposed and fed into closed-loop supply chains that re-use or recycle materials (IT Renew, 2022[97]). The European Commission launched the Sustainable Products Initiative, designed to make ICT equipment more durable, reusable, repairable and recyclable, and proposed a digital product passport to track hardware along its lifecycle. Further efforts could be made to incorporate principles of circular design, extended-life hardware, and recycling for AI related equipment (Sustainable Infrastructure Alliance, 2022[35]).

Indicators

Metrics to consider at the end-of-life stage of the AI compute resources lifecycle:

- electronic waste in metric tons
- recycling rate
- electronics disposal efficiency (EDE) percentage
- percentage of electronic waste sent to landfills

3.1.5. Positive direct environmental impacts of AI compute

Some scholars propose the term “green AI” to describe AI systems that do not increase (and ideally, decrease) the environmental costs associated with compute. This contrasts with “red AI”, where accuracy is obtained through the use of massive AI compute resources with little regard for the environmental impact (Schwartz et al., 2019[98]). Since even green AI consumes resources, it still has negative impacts on the environment. In only a handful of cases do positive direct environmental impacts result from AI compute, which are described below. These could be further analysed, expanded and emulated by AI compute providers.

Data centres produce large amounts of excess heat, typically considered “low-grade energy”. This energy usually cannot be repurposed for other activities as the temperatures are too low, typically below 35 degrees Celsius. Instead, excess heat is often directed into cooling towers. Several methods have been proposed to recover this heat, for example by combining the operation of a data centre and an onsite greenhouse or transferring it to local energy networks (Karnama, Haghighi and Vinuesa, 2019[93]; ReUseHeat, 2017[99]).

In terms of water consumption, several cases show the quality of wastewater released from data centres to actually be higher than when it was drawn from the source (Siddik, Shehabi and Marston, 2021[29]). However, this wastewater often must be treated after use, which consumes electricity and produces emissions, rendering the net sustainability impacts unclear.

3.2. Indirect environmental impacts of AI applications

The indirect environmental impacts of AI compute, which result through the application of AI, are a growing area of research and analysis. The application of AI and its structural and behavioural effects can cause both positive and negative indirect environmental impacts. However, given the multiplicity of AI applications and their indirect impacts’ diffuse nature, these are much harder to quantify than direct impacts. To date, assessments of indirect environmental impacts are largely qualitative, but their potential large-scale effect on climate action and planetary health provide a compelling case for the further development of analytical and measurement frameworks.

3.2.1. Negative indirect environmental impacts of AI applications

AI applications can cause negative environmental impacts. For example, AI applications can exacerbate the negative environmental impacts of the mining, extractive and manufacturing sectors, where advanced AI applications can be used upstream for finding and extracting minerals or fossil fuels, midstream for transport and material storage, and downstream for product refining. While AI can increase efficiencies that support sustainability efforts, these applications can also work to increase net GHG emissions instead of lowering them. For example, AI recommender systems used in e-commerce could increase consumption in unsustainable ways. Companies have also been reported to use the cloud to shift IT-related emissions from mandatory reporting requirements (e.g. on-premise) to voluntary reporting categories (e.g. outsourced cloud services) to "hide greenhouse gas emissions in the cloud" (Mytton, 2020[100]).

AI applications can also have complex, systemic effects on the environment and human behaviour. For instance, researchers observe that efficiency gains can be offset by "rebound effects" that cancel out positive sustainability impacts (Paul et al., 2019[101]). Rebound effects such as the "Jevons Paradox" occur when efficiency gains through technological progress are offset by acceleration in resource consumption (Giampietro and Mayumi, 2018[102]). Some researchers also observe a low willingness among consumers to pay for energy-efficient AI applications (König, Wurster and Siewert, 2022[103]). This led some researchers to point out the limitations of relying on efficiency increases alone, asserting that the compute demands of AI in its current form make future progress "economically, technically, and environmentally unsustainable" (Thompson et al., 2020[14]). Applying behavioural insights (e.g. recognition of the effect of different consumer biases on behaviour) to this area may assist in designing more effective policies (e.g. labelling schemes or mandatory information disclosures) that nudge consumers towards more energy efficient AI applications (OECD, 2017[104]).

3.2.2. Positive indirect environmental impacts of AI applications

Rolnick et al. highlight the areas in which AI applications can have a positive impact on climate action. These include electricity systems, transportation, buildings and cities, industry, farms and forests, GHG removal, climate prediction, societal impacts, solar geoengineering, individual and collective action, education and finance (Rolnick et al., 2022[105]). Energy system operators increasingly leverage AI-supported digital twin simulations to optimise energy systems and other environmental parameters. A digital twin can be defined as "a digital representation of a real-world entity or system…that mirrors a unique physical object, process, organization, person or other abstraction" (Gartner, n.d.[106]). On a larger scale, AI is leveraged in initiatives such as the European Space Agency's Destination Earth project to digitally model the Earth's environmental systems (Box 5).

Box 5. Digital twin for forecasting and resilience against climate change at Destination Earth

Understanding the interactions between human activities and natural phenomena is key to combating climate change and the biodiversity crisis. As part of the European Commission's Green Deal, Destination Earth (DestinE) is developing a digital twin of the Earth, supported by AI capacity and the EU's High-Performance Computing Joint Undertaking. Implemented by the European Space Agency, the European Centre for Medium-Range Weather Forecasts (ECMWF) and the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT), Destination Earth will enable dynamic Earth simulations, improve prediction capabilities and inform European environmental policy-making.

Source: (European Commission, 2022[107]; European Commission, 2022[49]; ECMWF, 2022[108])

Several consulting firms put forward estimates of the environmental impacts of AI applications, although their methodologies vary and often are not publicly disclosed (Kaack et al., 2022[74]). Boston Consulting Group (BCG) estimate that AI applications could help eliminate 2.6 to 5.3 gigatons of GHG emissions (5-10% of total emissions) and generate USD 1.3 to 2.6 trillion in value in revenues and cost savings through corporate sustainability by 2030 (BCG, 2021[109]). Capgemini estimated in 2020 that AI could reduce GHG emissions by 16% and improve power efficiency by 15% by 2025 (Capgemini, 2020[110]). According to PricewaterhouseCoopers (PwC), AI could increase global GDP 4.4% and reduce GHG emissions by 4%, or 2.4 gigatons, by 2030 (PWC, 2018[111]).

AI can enable and contribute to the 17 UN Sustainable Development Goals (SDGs), including climate action (SDG 13), life below water (SDG 14) and life on land (SDG 15) (Vinuesa et al., 2020[112]). Examples include facilitating climate analysis and forecasting, promoting energy conservation, improving GHG absorption and carbon storage, as well as decarbonising carbon-intensive sectors such as transport and agriculture (AI4SDGs, 2022[113]). Beyond climate action, AI applications can also be used for biodiversity conservation, healthy oceans, water security, clean air and weather and disaster resilience (World Economic Forum, 2018[114]). For example, AI applications have been leveraged to prevent wildlife destruction by supporting local conservation efforts (Gomes, 2019[115]) and could help cities detect costly water leaks in municipal water systems (Cody, Harmouche and Narasimhan, 2018[116]).

Use cases and tools have emerged in recent years to promote the use of AI for environmental sustainability. For instance, AI can be applied towards improved precipitation “nowcasting” — the high-resolution forecasting of precipitation for weather-related decision-making (Ravuri et al., 2021[117]) — and long-term forecasts and sustainability insights for agricultural production, which can increase climate resilience and agricultural productivity (ClimateAI, 2022[118]). AI is starting to be more widely used by organisations, municipalities and regions to measure, simulate and reduce the environmental footprint of supply chains (Barteková and Börkey, 2022[20]). AI can also be leveraged to scale the transition to a resource-efficient and circular economy, "decoupling economic activity from natural resource use and its environmental impacts" (Barteková and Börkey, 2022[20]). Finally, AI applications can accelerate scientific research and breakthroughs in the development and deployment of sustainable technologies. For example, researchers predict that a type of AI called deep reinforcement learning could accelerate the development of nuclear fusion, a potentially transformative carbon-free technology for the world’s energy demands (Degrave et al., 2022[119]).

It is important to note that most benchmarks, frameworks and impact assessments for the indirect environmental impacts of AI compute remain high-level and qualitative. Some researchers put forward approaches for measuring the environmental impacts of ICT overall, such as a methodology for assessing the effects induced by ICT services (Bergmark et al., 2020[120]), the assessment of indirect environmental effects of digitalisation based on a time-use perspective (Bieser and Hilty, 2018[121]), and an ICT-for-sustainability framework (Hilty and Aebischer, 2015[122]) that was extended specifically to machine-learning-related GHG emissions (Kaack et al., 2022[74]). Nevertheless, AI-specific methodologies for indirect environmental impacts remain rare – a key gap and limitation identified by this report.

Box 6. Tracking global greenhouse gas emissions through Climate TRACE

National GHG emission tracking historically relies on self-reporting and bottom-up estimates. The Climate TRACE coalition of AI analytics organisations leverages over 300 satellites and 11 000 sensors to quantify emission sources through AI algorithms. Emission sources include oil and gas production, shipping, aviation, forestry and more. Climate TRACE is relevant in more than 100 countries that lack accurate emission inventories of the past five years.

Source: (GPAI, 2021[9]; Climate Trace, 2022[123]; Carbon Tracker, 2020[124])

Indicators

Key consideration for the measurement of negative and positive indirect environmental impacts associated with AI compute applications include:

- Impact assessment for net environmental cost or benefit of using an AI application (i.e. the energy/water/carbon/resources saved compared to energy/water/carbon/resources used)

3.3. Dual impacts

Like many technologies, the use of AI compute resources and applying AI itself can have both positive and negative effects, providing an opportunity to harness the benefits while minimising the costs to achieve a net positive environmental impact (Cowls et al., 2021[125]). Machine learning and cloud computing can "help reduce stress on the environment in specific domains", but "raise concerns about environmental and climate impact", including use of energy and natural resources, and waste disposal (EPFL International Risk Governance Center, 2022[126]). Addressing these dual impacts is an area for further study, requiring policy frameworks, tools and measurements to assess the net impacts of AI on the environment, developed through inter-disciplinarily efforts by scientists, policy makers, technologists and others (Stein, 2020[127]).

4 Measurement gaps with policy implications

The literature review, public survey and expert interviews point to five gaps in knowledge and data needed to understand and assess the environmental impacts of AI compute and applications, along with policy implications: (1) measurement standards for sustainable AI are needed; (2) data collection on the environmental impacts of AI compute and applications should be expanded; (3) AI-specific measurements are difficult to separate from general-purpose compute; (4) environmental impacts beyond operational energy use and GHG emissions should be considered; and (5) efforts are needed to improve environmental transparency and equity everywhere.

4.1. Measurement standards for sustainable AI are needed

Measurement of the environmental impacts of AI compute and applications is limited by a lack of common terminology, recognised standards, consistent indicators and metrics, and varying or optional reporting requirements. Terminology used to describe the environmental impacts of AI is largely heterogeneous: terms like “green AI”, “computational sustainability” or “sustainable AI” are often used interchangeably, and various actors define environmental impacts differently in different contexts (Schwartz et al., 2019[98]; Hilty and Aebischer, 2015[122]).

However, developing common terminology would not necessarily lead to comparable data and results unless accompanied by widely used or mandatory reporting requirements. Specific regulations, standards and policies (including tax policies) are underdeveloped in this area compared to other environmental, social and governance (ESG) reporting requirements (OECD, 2020[128]). Common measurement standards will need to reflect a holistic understanding of the environmental impacts of AI compute throughout its lifecycle and AI applications. Focusing only on select indicators could have unintended consequences. For instance, PUE is critical but other parameters are important, like investments in data centre environmental resilience or in energy-saving ICT equipment (Lawrence, 2020[129]). Considerations could also be given to the compliance costs of such measurement standards.

A comprehensive framework developed by international or inter-governmental standard-setting institutions and international initiatives, as part of a multi-stakeholder process, could enable benchmarking, comparability and compatibility of national compute initiatives and their environmental impacts, including for AI. Organisations such as the OECD could contribute to developing such a framework.

4.2. Data collection on the environmental impacts of AI compute and applications should be expanded

Measuring the environmental impacts of AI compute and applications requires national-, company- and model-level data. National government agencies and institutions should expand data collection on GHG emissions, and energy, water and other resource consumption by AI compute. National statistics offices and environmental agencies in countries such as Denmark, Ireland and the Netherlands have started to disaggregate and publish data on the electricity usage of ICT infrastructure and data centres.

Many private-sector companies report environmental metrics, with ESG data collection and reporting having improved and accelerated in recent years. However, much disclosure remains voluntary in the absence of clear reporting requirements, and corporate environmental transparency varies widely.

At the level of AI models, researchers have started to measure the energy and carbon impact of their models and include this information in research papers. This trend can be encouraged by sharing best practices, reporting data on more granular levels (such as differentiating between AI training and inference) and encouraging reporting requirements at AI research institutions, organisations and private-sector entities developing large-scale AI models.

4.3. AI-specific measurements are difficult to separate from general-purpose compute

Another limitation to assessing the environmental impacts of AI compute and applications is the complexity of the measurements themselves. This relates to the challenge of distinguishing general-purpose compute and ICT activities from those related solely to AI, and difficulty in identifying the share of AI compute-related production and transport activities globally. AI-specific indirect positive and negative environmental effects are even harder to account for as they are often embedded in a variety of applications and complex systems, making disaggregation difficult. Further, increasing use of AI compute edge devices, such as mobile smartphones and IoT connected devices, complicate the measurement of AI-specific compute, as many of these devices have applications and uses in addition to their AI compute capabilities.

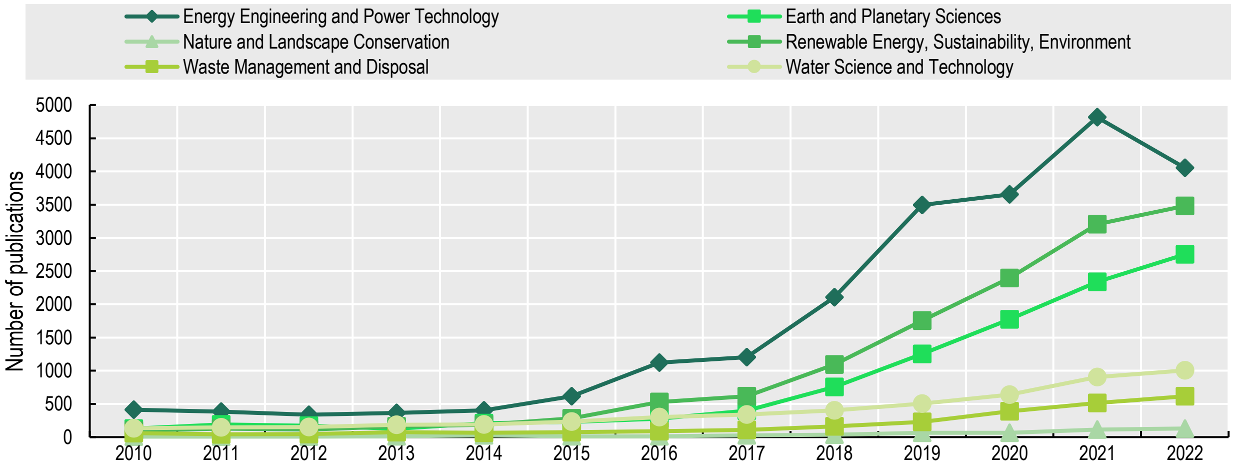
Many articles and publications attempting to measure the environmental impacts of AI compute equate measurements with data centres and compute hardware, or even the ICT industry in general. For policy makers and data centre operators, it can make sense to assess the overall impact of ICT on the environment, with AI embedded as an integral part of these technologies. Further efforts by governments, national statistical offices, intergovernmental organisations, academia and others should identify gaps where AI-specific metrics would be useful. Proxy measures, qualitative indicators and the feasibility of disaggregating national ICT infrastructure datasets should be explored to estimate the share of compute infrastructure used for AI. As AI becomes more widely used across sectors, with models increasing in size and data requirements, computational needs are expected to grow significantly, making measurement ever more important.

4.4. Environmental impacts beyond operational energy use and greenhouse gas emissions should be considered

Environmental impacts of AI compute should be explored beyond the current focus on operational energy consumption and the resulting carbon footprint. These include assessing the impact on planetary boundaries such as biodiversity and the lifecycle impacts of production, including resource impacts like water consumption and rare earth mining, transport and end-of-life impacts.

Much discussion of the environmental impacts of AI compute focuses on energy efficiency and renewable-energy. Environmental AI scientific publications related to these areas significantly outnumber others, such as water science and technology and waste management and disposal (Figure 5). Energy and climate warrant immediate attention and action, but broader environmental sustainability has connected dimensions (such as nature and landscape conservation or ocean sciences) and should also be considered. The OECD’s 2009 Digital Economy Paper “Towards Green ICT Strategies” notes that “environmental impact categories such as biodiversity, water or land use are rarely targeted” by policy makers (OECD, 2009[130]), which appears to remain a policy gap today. Making the operational stage of AI compute carbon-neutral is an important and urgent objective, but operators and policy makers should work in parallel on reducing impacts on ecosystems and ensure that AI compute and applications contribute to holistic environmental action.

Figure 5. Sample of environmental AI scientific publications by subject area



Note: Original visualisation on the OECD AI Policy Observatory powered by JSI using data from Elsevier (Scopus), accessed on 18/10/2022. For more information visit: www.oecd.ai
Source: (OECD AI Policy Observatory, 2022[131])

4.5. Efforts are needed to improve environmental transparency and equity everywhere

There are concerns about an AI compute divide manifesting between private and public actors, and between advanced and emerging countries. This could run counter to efforts to promote environmental equity – generally understood as equal protection from environmental hazards and equal access to environmental benefits. Analysis of the top ten countries in AI research on the environment shows that the United States, EU27 and China lead in the number of publications (OECD AI Policy Observatory, 2022[131]). One aim of the Expert Group is to understand the AI compute divide between and within countries. Important elements of environmental equity and transparency are sharing best practices, measuring and ensuring access to the AI compute ecosystem, and implementing “sustainability by design”. While advanced economies and the private-sector often drive initiatives to reduce the energy and carbon impact of AI compute, diverse perspectives from actors in emerging economies could further enrich and benefit the discussion (Birhane et al., 2021[132]).

Initiatives such as “FAIR Forward – Artificial Intelligence for All” enable knowledge-sharing and access to environmental data in Ghana, India, Rwanda, South Africa and Uganda (Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ), 2022[133]). Sharing best practices on the sustainable design and operations of AI compute and applications contribute to global knowledge dissemination (Patterson, 2022[75]; Microsoft, 2018[134]) and the development of skills for efficient and sustainable management of AI compute and its responsible application.

5 Conclusion

To achieve global sustainability targets, AI must be part of the solution. From energy efficiency gains to the discovery and scale-up of clean technologies, AI-enabled innovation can contribute to finding solutions that countries need to meet global sustainability targets. However, as AI applications become more diverse and as the computing needs of AI systems grow, they leave environmental footprints that must also be measured and taken into account. Through a literature review and expert consultation, this report attempts to evaluate the state of measurement data and tools currently available to quantify the direct and indirect environmental impacts of AI. In doing so, it shows that there are gaps between the data and tools that are available to measure AI's environmental footprint and reliable measures policy makers need to inform sustainable policy decisions.

Several national statistics offices, environmental agencies, and private sector companies have started to disaggregate and publish relevant environmental data. More holistic analysis requires a wide array of stakeholders doing the same, using a common framework to examine the environmental impacts of AI systems. By creating and tracking AI-specific measures of compute, sharing best practices, and supporting new and innovative AI applications for fighting climate change, countries can ensure that AI is trained and deployed in the most sustainable way possible, while minimising negative impacts, for the good of the planet.

References

Acun, B. et al. (2022), "A Holistic Approach for Designing Carbon Aware Datacenters". [80]

Ahmed, N. and M. Wahed (2020), "The De-democratization of AI: Deep Learning and the Compute Divide in Artificial Intelligence Research". [19]

AI4SDGs (2022), *Principles on AI for Climate Action*, https://ai-for-sdgs.academy/principles-on-ai-for-climate-action. [113]

Allwood, J. et al. (2014), *Glossary. In: Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge Univserity Press, https://www.ipcc.ch/site/assets/uploads/2018/02/ipcc_wg3_ar5_annex-i.pdf. [152]

Amazon (2022), *Amazon sustainability*, https://sustainability.aboutamazon.com/ (accessed on 25 April 2022). [77]

Amazon (2022), "Delivering Progress Every Day, Amazon's 2021 Sustainability Report", https://sustainability.aboutamazon.com/2021-sustainability-report.pdf. [150]

Amazon Web Services (2022), *Customer Carbon Footprint Tool*, https://aws.amazon.com/?nc2=h_lg (accessed on 25 April 2022). [81]

Amazon Web Services (2021), *Sustainability Pillar. AWS Well-Architected Framework*, https://docs.aws.amazon.com/wellarchitected/latest/sustainability-pillar/wellarchitected-sustainability-pillar.pdf#sustainability-pillar (accessed on 25 April 2022). [82]

Andrae, A. (2020), "Hypotheses for primary energy use, electricity use and CO2 emissions of global computing and its shares of the total between 2020 and 2030", *WSEAS Transactions on Power Systems*, Vol. 15, https://doi.org/10.37394/232016.2020.15.6. [43]

Andrae, A. and T. Edler (2015), "On Global Electricity Usage of Communication Technology: Trends to 2030", *Challenges*, Vol. 6/1, https://doi.org/10.3390/challe6010117. [136]

Apple (2021), *Environmental Progress Report*, https://www.apple.com/environment/pdf/Apple_Environmental_Progress_Report_2021.pdf. [57]

Avgerinou, M., P. Bertoldi and L. Castellazzi (2017), "Trends in Data Centre Energy Consumption under the European Code of Conduct for Data Centre Energy Efficiency", *Energies*, Vol. 10/10, https://doi.org/10.3390/en10101470. [137]

Banet, C. et al. (2021), *Data centers & the grid: greening ICT in Europe*, https://cerre.eu/wp-content/uploads/2021/10/211013_CERRE_Report_Data-Centres-Greening-ICT_FINAL.pdf. [38]

Barteková, E. and P. Börkey (2022), “Digitalisation for the transition to a resource efficient and circular economy”, *OECD Environment Working Papers* 192, https://doi.org/10.1787/19970900 (accessed on 29 April 2022). [20]

BCG (2021), *Reduce Carbon and Costs with the Power of AI*, https://www.bcg.com/publications/2021/ai-to-reduce-carbon-emissions (accessed on 25 April 2022). [109]

Belkhir, L. and A. Elmeligi (2018), “Assessing ICT global emissions footprint: Trends to 2040 & recommendations”, *Journal of Cleaner Production*, Vol. 177, https://doi.org/10.1016/j.jclepro.2017.12.239. [144]

Bergmark, P. et al. (2020), *A Methodology for Assessing the Environmental Effects Induced by ICT Services: Part II: Multiple Services and Companies*, https://doi.org/10.1145/3401335.3401711. [120]

Berkhout, F. and J. Hertin (2001), “Impacts of information and communication technologies on environmental sustainability: speculations and evidence”, *Report to the OECD*, https://www.oecd.org/science/inno/1897156.pdf. [18]

Bieser, J. and L. Hilty (2018), “An Approach to Assess Indirect Environmental Effects of Digitalization Based on a Time-Use Perspective”, https://doi.org/10.1007/978-3-319-99654-7_5. [121]

Birhane, A. et al. (2021), “The Values Encoded in Machine Learning Research”. [132]

Capgemini (2020), *Climate AI. How artificial intelligence can power your climate action strategy*, https://www.capgemini.com/fr-fr/wp-content/uploads/sites/2/2020/11/Climate-AI_Final.pdf (accessed on 25 April 2022). [110]

Carbon Tracker (2020), *New Global Coalition of Tech, Climate Groups Will Combine AI and Satellites to Monitor GHG Emissions Worldwide in Real Time*, https://carbontracker.org/climatetrace/. [124]

CBS (2021), *Electricity supplied to data centres, 2017-2019*, https://www.cbs.nl/en-gb/custom/2020/51/electricity-supplied-to-data-centres-2017-2019 (accessed on 25 April 2022). [47]

Center of Expertise for Energy Efficiency in Data Centers (2022), *Who We Are*, https://datacenters.lbl.gov/who-we-are. [62]

Central Statistics Office (2022), *Data Centres Metered Electricity Consumption 2020*, https://www.cso.ie/en/releasesandpublications/ep/p-dcmec/datacentresmeteredelectricityconsumption2020/keyfindings/ (accessed on 25 April 2022). [45]

Chinese State Council (2021), *Green data centers in focus*, http://english.www.gov.cn/statecouncil/ministries/202112/09/content_WS61b13edac6d09c94e48a1f81.html. [52]

Climate Neutral Data Centre Pact (2022), *The Green Deal needs Green Infrastructure*, https://www.climateneutraldatacentre.net/ (accessed on 25 April 2022). [92]

Climate Trace (2022), *Radical Transparency for Global Emissions*, https://climatetrace.org/. [123]

ClimateAI (2022), *Solutions Across The Agriculture Value Chain*, https://climate.ai/solutions/ (accessed on 25 April 2022). [118]

Cloud Carbon Footprint (2022), *Cloud carbon emissions measurement and analytics tool*, https://www.cloudcarbonfootprint.org/ (accessed on 25 April 2022). [85]

CodeCarbon (2022), *Track and reduce CO2 emissions from your computing*, https://codecarbon.io/index.html#about (accessed on 25 April 2022). [87]

CODES (2021), *Synthesis report: a Digital Planet for Sustainability*, https://www.sparkblue.org/system/files/2021-05/Zero%20Order%20Draft_A%20DIGITAL%20PLANET%20FOR%20SUSTAINABILITY_%206_May%202021.pdf (accessed on 25 April 2022). [10]

Cody, R., J. Harmouche and S. Narasimhan (2018), "Leak detection in water distribution pipes using singular spectrum analysis", *Urban Water Journal*, Vol. 15/7, https://doi.org/10.1080/1573062X.2018.1532016. [116]

Coroama, V. et al. (2020), *A Methodology for Assessing the Environmental Effects Induced by ICT Services: Part I: Single Services*, https://doi.org/10.1145/3401335.3401716. [148]

Cowls, J. et al. (2021), "The AI gambit: leveraging artificial intelligence to combat climate change—opportunities, challenges, and recommendations", *AI and Society*, https://doi.org/10.1007/s00146-021-01294-x. [125]

Crawford, K. (2021), *Atlas of AI*, https://yalebooks.yale.edu/book/9780300209570/atlas-ai. [34]

Crawford, K. (2018), *Anatomy of an AI*, https://anatomyof.ai. [26]

Danish Energy Agency (2020), *Denmark's Climate and Energy Outlook 2020*. [44]

DeepMind (2018), *Safety-first AI for autonomous data centre cooling and industrial control*, https://www.deepmind.com/blog/safety-first-ai-for-autonomous-data-centre-cooling-and-industrial-control. [69]

DeepMind (2016), *DeepMind AI Reduces Google Data Centre Cooling Bill by 40%*, https://www.deepmind.com/blog/deepmind-ai-reduces-google-data-centre-cooling-bill-by-40. [68]

Degrave, J. et al. (2022), "Magnetic control of tokamak plasmas through deep reinforcement learning", *Nature* 602, pp. 414–419, https://www.nature.com/articles/s41586-021-04301-9. [119]

Desislavov, R., F. Martínez-Plumed and J. Hernández-Orallo (2021), "Compute and Energy Consumption Trends in Deep Learning Inference". [67]

Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) (2022), *FAIR Forward – Artificial Intelligence for All*, https://www.giz.de/expertise/html/61982.html (accessed on 25 April 2022). [133]

Dhar, P. (2020), "The carbon impact of artificial intelligence", *Nature Machine Intelligence*, Vol. 2/8, https://doi.org/10.1038/s42256-020-0219-9. [76]

Digital Realty (2019), *Environmental performance*, https://www.digitalrealty.com/environmental-social-and-governance-report-2019-highlights/environmental-performance. [59]

Eaton (2021), *Data Centers and Decarbonization*, https://www.eaton.com/content/dam/eaton/company/news-insights/energy-transition/documents/bnef-eaton-statkraft-data-center-study-en-us.pdf. [138]

ECMWF (2022), *Destination Earth*, https://www.ecmwf.int/en/about/what-we-do/environmental-services-and-future-vision/destination-earth. [108]

EEHPCWG (2022), *Energy Efficient High Performance Computing Working Group: Sustainably supporting science through committed community action*. [61]

EirGrid (2021), *All-Island Generation Capacity Statement*, https://www.eirgridgroup.com/site-files/library/EirGrid/208281-All-Island-Generation-Capacity-Statement-LR13A.pdf. [46]

Ensmenger (2013), "Computation, Materiality, and the Global Environment.", *IEEE Annals of the History of Computing*, Vol. 35/3, https://ieeexplore.ieee.org/document/6617646. [135]

EPFL International Risk Governance Center (2022), *Ensuring the environmental sustainability of emerging technologies*, https://www.epfl.ch/research/domains/irgc/ensuring-the-sustainability-of-emerging-technology/ (accessed on 25 April 2022). [126]

Equinix (2021), *Sustainability Report Highlights*, https://sustainability.equinix.com/wp-content/uploads/2021/05/Sustainability-Report-Highlights-FY2020.pdf (accessed on 25 April 2022). [60]

European Commission (2022), *Destination Earth – new digital twin of the Earth will help tackle climate change and protect nature*, https://ec.europa.eu/commission/presscorner/detail/en/IP_22_1977. [49]

European Commission (2022), *Sustainable Products Initiative*, https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12567-Sustainable-products-initiative_en. [107]

European Parliament (2021), *The role of artificial intelligence in the European green deal*, https://www.europarl.europa.eu/RegData/etudes/STUD/2021/662906/IPOL_STU(2021)662906_EN.pdf. [6]

Federal Register (2022), "Request for Information on the Energy and Climate Implications of Digital Assets", *Federal Register*, Vol. 87/58, https://www.federalregister.gov/documents/2022/03/25/2022-06284/request-for-information-on-the-energy-and-climate-implications-of-digital-assets (accessed on 29 April 2022). [146]

Financial Times (2022), *West London faces new homes ban as electricity grid hits capacity*. [54]

Forum, W. (2019), "A New Circular Vision for Electronics: Time for a Global Reboot", https://www3.weforum.org/docs/WEF_A_New_Circular_Vision_for_Electronics.pdf. [95]

Freitag, C. et al. (2021), *The real climate and transformative impact of ICT: A critique of estimates, trends, and regulations*, https://doi.org/10.1016/j.patter.2021.100340. [71]

Gartner (n.d.), *Gartner Glossary: Digital Twin*. [106]

Giampietro, M. and K. Mayumi (2018), "Unraveling the complexity of the Jevons Paradox: The link between innovation, efficiency, and sustainability", *Frontiers in Energy Research*, Vol. 6/APR, https://doi.org/10.3389/fenrg.2018.00026. [102]

Gomes, C. (2019), "Computational sustainability: Computing for a better world and a sustainable future", *Communications of the ACM*, Vol. 62/9, https://doi.org/10.1145/3339399. [115]

Google (2022), *A Policy Roadmap for 24/7 Carbon-Free Energy*, https://www.gstatic.com/gumdrop/sustainability/policy-roadmap-carbon-free-energy.pdf. [78]

Google (2022), *Environmental Report 2021*. [33]

Google Cloud (2022), *Carbon Footprint reporting methodology*, https://cloud.google.com/carbon-footprint/docs/methodology (accessed on 25 April 2022). [83]

Google Cloud (2022), *Incident Report*, https://status.cloud.google.com/incidents/fmEL9i2fArADKawkZAa2. [56]

GPAI (2021), *Climate change & AI: Recommendations for government action*, https://gpai.ai/projects/responsible-ai/environment/climate-change-and-ai.pdf. [9]

Green500 (2021), *Green500 November 2021 Release*, https://www.top500.org/lists/green500/2021/11/. [64]

Greenpeace China (2021), *Clean Cloud 2021: Tracking Renewable Energy Use in China's Tech Industry*, https://www.greenpeace.org/static/planet4-eastasia-stateless/2021/04/03a3ce1a-clean-cloud-english-briefing.pdf (accessed on 25 April 2022). [51]

Greenpeace USA (2017), *Clicking Clean: Who is Winning the Race to Build A Green Internet?*. [147]

Guler, B. and A. Yener (2021), "Sustainable Federated Learning". [66]

Gupta et al. (2020), "Chasing Carbon: The Elusive Environmental Footprint of Computing", https://arxiv.org/pdf/2011.02839.pdf. [30]

Henderson, P. et al. (2020), "Towards the systematic reporting of the energy and carbon footprints of machine learning", *Journal of Machine Learning Research*, Vol. 21. [27]

Heslin, K. (2016), *Ignore Data Center Water Consumption at Your Own Peril*, Uptime Institute, https://journal.uptimeinstitute.com/dont-ignore-water-consumption/ (accessed on 21 April 2022). [89]

Hilty, L. and B. Aebischer (2015), "Ict for sustainability: An emerging research field", *Advances in Intelligent Systems and Computing*, Vol. 310, https://doi.org/10.1007/978-3-319-09228-7_1. [122]

Hintemann, R. (2018), *Efficiency gains are not enough: Data center energy consumption continues to rise significantly*, Borderstep Institute for Innovation and Sustainability, https://www.borderstep.de/wp-content/uploads/2020/04/Borderstep-Datacenter-2018_en.pdf (accessed on 2 May 2022). [53]

Hoffmann, J. et al. (2022), "Training Compute-Optimal Large Language Models". [13]

Hogan, M. (2015), "Data flows and water woes: The Utah Data Center", *Big Data & Society*, Vol. 2/2, p. 205395171559242, https://doi.org/10.1177/2053951715592429. [142]

IEA (2021), *Data Centres and Data Transmission Networks*, International Energy Agency (IEA_. [16]

IEA (2021), *Transport. Improving the sustainability of passenger and freight transport*, https://www.iea.org/topics/transport. [37]

IEA (2017), *Digitalization & Energy*, https://iea.blob.core.windows.net/assets/b1e6600c-4e40-4d9c-809d-1d1724c763d5/DigitalizationandEnergy3.pdf. [39]

IEEE (2022), "IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—I: REGULAR PAPERS", https://ieeexplore.ieee.org/document/9782468/metrics#metrics. [42]

International Telecommunication Union and World Benchmarking Alliance (2022), "Greening digital companies: monitoring emissions and climate commitments", https://www.itu.int/en/ITU-D/Environment/Documents/Publications/2022/Greening-Digital-Companies-22June2022.pdf. [73]

IPBES (2019), *Global Assessment Summary for Policymakers*, https://ipbes.net/news/global-assessment-summary-policymakers-final-version-now-available (accessed on 12 April 2022). [5]

IPCC (2021), *AR6 Climate Change 2021: The Physical Science Basis*, https://www.ipcc.ch/report/ar6/wg1/#SPM (accessed on 12 April 2022). [4]

IT Renew (2022), *Energy isn't the only renewable resource*, https://www.itrenew.com. [97]

ITU (2020), *L.1470 : Greenhouse gas emissions trajectories for the information and communication technology sector compatible with the UNFCCC Paris Agreement*, https://www.itu.int/rec/T-REC-L.1470-202001-I/en (accessed on 2 May 2022). [72]

Kaack, L. et al. (2022), "Aligning artificial intelligence with climate change mitigation", *Nature Climate Change*, Vol. 12, pp. 518–527, https://doi.org/10.1038/s41558-022-01377-7 (accessed on 25 April 2022). [74]

Kaplan, J. et al. (2020), "Scaling Laws for Neural Language Models", *arXiv:2001.08361 [cs, stat]*. [12]

Karnama, A., E. Haghighi and R. Vinuesa (2019), "Organic data centers: A sustainable solution for computing facilities", *Results in Engineering*, Vol. 4, https://doi.org/10.1016/j.rineng.2019.100063. [93]

König, P., S. Wurster and M. Siewert (2022), "Consumers are willing to pay a price for explainable, but not for green AI. Evidence from a choice-based conjoint analysis", *Sage Journals*, https://journals.sagepub.com/doi/full/10.1177/20539517211069632. [103]

Lacoste, A. et al. (2019), "Quantifying the Carbon Emissions of Machine Learning". [86]

Lannelongue et al. (2021), "Green Algorithms: Quantifying the Carbon Footprint of Computation", https://onlinelibrary.wiley.com/doi/full/10.1002/advs.202100707. [28]

Lawrence, A. (2020), *PUE: The golden metric is looking rusty*, Uptime Institute, https://journal.uptimeinstitute.com/pue-the-golden-metric-is-looking-rusty/ (accessed on 25 April 2022). [129]

Lefdal Mine Datacentre (n.d.), *One of the greenest data centres in the world*. [94]

Malmodin, J. and D. Lundén (2018), "The Energy and Carbon Footprint of the Global ICT and E&M Sectors 2010-2015", *Information and Communications Technologies (ICT) for Sustainability*, Vol. 9/10, https://doi.org/10.3390/su10093027. [31]

Masanet, E. et al. (2020), "Recalibrating global data center energy-use estimates", *Science*, Vol. 367/6481, https://doi.org/10.1126/science.aba3758. [41]

Meta (2021), *2020 Data Disclosure*, https://sustainability.fb.com/wp-content/uploads/2022/06/Meta-2021-Sustainability-Report.pdf. [32]

Microsoft (2022), *Emissions Impact Dashboard*, https://www.microsoft.com/en-us/sustainability/emissions-impact-dashboard?activetab=pivot_2:primaryr12 (accessed on 21 April 2022). [84]

Microsoft (2021), *2020 Environmental Sustainability Report*, https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RWyG1q (accessed on 25 April 2022). [58]

Microsoft (2020), *The carbon benefits of cloud computing*, https://www.microsoft.com/en-us/download/details.aspx?id=56950. [139]

Microsoft (2018), *The Carbon Benefits of Cloud Computing: a Study of the Microsoft Cloud*. [134]

Montevecchi, F. et al. (2020), *Energy-efficient cloud computing technologies and policies for an eco-friendly cloud market : final study report*, European Commission, Directorate-General for Communications Networks, Content and Technology, https://data.europa.eu/doi/10.2759/3320. [50]

MosaicML (2022), *MosaicML ResNet Part 2: Research Deep Dive*, https://www.mosaicml.com/blog/mosaic-resnet-deep-dive. [17]

Mytton, D. (2021), "Data centre water consumption", *npj Clean Water*, Vol. 4/1, p. 11, https://doi.org/10.1038/s41545-021-00101-w. [90]

Mytton, D. (2020), *Hiding greenhouse gas emissions in the cloud*, https://doi.org/10.1038/s41558-020-0837-6. [100]

NVIDIA (2021), *Climate change 2020*, https://www.nvidia.com/content/dam/en-zz/Solutions/csr/NVIDIA_Corporation_CDP_Climate_Change_Questionnaire_2020.pdf. [140]

OECD (2022), "OECD Framework for the Classification of AI systems", *OECD Digital Economy Papers*, No. 323, OECD Publishing, Paris, https://doi.org/10.1787/cb6d9eca-en. [8]

OECD (2022), *OECD Framework for the Classification of AI systems*, OECD Publishing, Paris. [23]

OECD (2020), *Confronting Planetary Emergencies - Solving Human Problems*, https://www.oecd.org/about/secretary-general/confronting-planetary-emergencies-solving-human-problems-october-2020.htm. [2]

OECD (2020), *ESG Investing: Practices, Progress, and Challenges*. [128]

OECD (2019), *Measuring the Digital Transformation*, Organisation for Economic Co-Operation and Development (OECD), https://doi.org/10.1787/888933931086. [96]

OECD (2019), "Recommendation of the Council on Artificial Intelligence", *OECD Legal Instruments*, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 (accessed on 12 April 2022). [1]

OECD (2017), "Use of Behavioural Insights in Consumer Policy", *OECD Science, Technology and Industry Policy Papers* No.36, https://doi.org/10.1787/c2203c35-en. [104]

OECD (2009), *Towards Green ICT Strategies*, http://www.observatorioabaco.es/biblioteca/docs/65_OECD_WP_155_2009.pdf. [130]

OECD (forthcoming), *Building National Compute Capacity for Artificial Intelligence (AI): Existing measurement tools and preliminary findings*. [15]

OECD (n.d.), *Transport and the Environment*, https://www.oecd.org/environment/transport-and-environment.htm. [36]

OECD AI Policy Observatory (2022), *Data and trends*, https://oecd.ai/en/. [131]

OECD AI Policy Observatory (2019), *Inclusive growth, sustainable development and well-being (Principle 1.1)*, https://oecd.ai/en/dashboards/ai-principles/P5. [7]

OECD AI Policy Observatory (2019), *OECD AI Principles overview*, https://oecd.ai/en/ai-principles. [145]

OECD Going Digital Toolkit (2021), *Share of businesses purchasing cloud services*, https://goingdigital.oecd.org/indicator/25. [149]

OpenAI (2018), *AI and Compute*, https://openai.com/blog/ai-and-compute/. [11]

Patterson, D. (2022), “The Carbon Footprint of Machine Learning Training Will Plateau, Then Shrink”, *TechRxiv*, https://www.techrxiv.org/articles/preprint/The_Carbon_Footprint_of_Machine_Learning_Training_Will_Plateau_Then_Shrink/19139645/1 (accessed on 21 April 2022). [75]

Patterson, D. et al. (2021), “Carbon Emissions and Large Neural Network Training”. [88]

Paul, C. et al. (2019), *Rebound effects in agricultural land and soil management: Review and analytical framework*, https://doi.org/10.1016/j.jclepro.2019.04.115. [101]

PWC (2018), *How AI can enable a Sustainable Future*, https://www.pwc.co.uk/sustainability-climate-change/assets/pdf/how-ai-can-enable-a-sustainable-future.pdf (accessed on 21 April 2022). [111]

Radovanovic, A. et al. (2021), “Carbon-Aware Computing for Datacenters”. [79]

Ravuri, S. et al. (2021), “Skilful precipitation nowcasting using deep generative models of radar”, *Nature* 597, pp. 672–677, https://doi.org/10.1038/s41586-021-03854-z. [117]

ReUseHeat (2017), *Demo Sites: Braunschweig*, https://www.reuseheat.eu/brunswick/. [99]

Rohde, F. et al. (2021), “Nachhaltigkeitskriterien für künstliche Intelligenz”, https://www.ioew.de/fileadmin/user_upload/BILDER_und_Downloaddateien/Publikationen/2021/IOEW_SR_220_Nachhaltigkeitskriterien_fuer_Kuenstliche_Intelligenz.pdf. [21]

Rolnick, D. et al. (2022), “Tackling Climate Change with Machine Learning”, *ACM Computing Surveys*, Vol. 55/2, pp. 1-96, https://doi.org/10.1145/3485128. [105]

Schiff, D. et al. (2020), “IEEE 7010: A New Standard for Assessing the Well-being Implications of Artificial Intelligence”, https://doi.org/10.1109/SMC42975.2020.9283454. [143]

Schwartz, R. et al. (2019), “Green AI”. [98]

Sevilla, J. et al. (2022), “Compute Trends Across Three Eras of Machine Learning”. [151]

Shehabi, A. et al. (2016), “United States Data Center Energy Usage Report | Energy Technologies Area”, *Berkeley Lab* June. [48]

Siddik, M., A. Shehabi and L. Marston (2021), “The environmental footprint of data centers in the United States”, *Environmental Research Letters*, Vol. 16/6, https://doi.org/10.1088/1748-9326/abfba1. [29]

Sönnichsen (2021), *Energy demand in data centers worldwide from*, https://www.statista.com/statistics/186992/global-derived-electricity-consumption-in-data-centers-and-telecoms/. [40]

Stein, A. (2020), “Artificial Intelligence and Climate Change”, *Yale J. on Reg.*, Vol. 37/890, https://scholarship.law.ufl.edu/facultypub/996/ (accessed on 21 April 2022). [127]

Strubell, E., A. Ganesh and A. McCallum (2020), “Energy and Policy Considerations for Modern Deep Learning Research”, *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34/09, pp. 13693-13696, https://doi.org/10.1609/aaai.v34i09.7123. [63]

Sustainability Index for Artificial Intelligence (2022), “SustAIn”, *Sustainable AI in Practice*, https://algorithmwatch.org/en/wp-content/uploads/2022/06/SustAIn_Magazine_2022_EN.pdf. [22]

Sustainable Infrastructure Alliance (2022), *The Roadmap to Sustainable Digital Infrastructure by 2030*, https://sdialliance.org/roadmap. [35]

The Times (2022), *Doubt over tech giants’ EUR 2bn data investment*. [55]

Thompson, N. et al. (2020), “The Computational Limits of Deep Learning”, *arXiv:2007.05558 [cs, stat]*. [14]

Thompson, N. et al. (2020), “The Computational Limits of Deep Learning”. [141]

TOP500 (2022), *June 2022 List*, https://www.top500.org/lists/top500/2022/06/. [153]

UBA (2021), *Green Cloud Computing*, Fraunhofer-Institut für Zuverlässigkeit und Mikrointegration (IZM), Berlin, https://www.umweltbundesamt.de/sites/default/files/medien/5750/publikationen/2021-06-17_texte_94-2021_green-cloud-computing.pdf (accessed on 21 April 2022). [25]

UNEP (2020), *The triple planetary crisis: Forging a new relationship between people and the earth*, https://www.unep.org/news-and-stories/speech/triple-planetary-crisis-forging-new-relationship-between-people-and-earth. [3]

Uptime Institute (2021), *Global Data Center Survey 2021*, https://uptimeinstitute.com/uptime_assets/4d10650a2a92c06a10e2c70e320498710fed2ef3b402aa82fe7946fae3887055-2021-data-center-industry-survey.pdf (accessed on 21 April 2022). [91]

Vinuesa, R. et al. (2020), *The role of artificial intelligence in achieving the Sustainable Development Goals*, https://doi.org/10.1038/s41467-019-14108-y. [112]

Welling (2018), *How algorithmic advances make power-efficient AI possible*, https://www.qualcomm.com/news/onq/2018/08/13/how-algorithmic-advances-make-power-efficient-ai-possible. [65]

World Economic Forum (2018), *Harnessing Artificial Intelligence for the Earth*, https://www3.weforum.org/docs/Harnessing_Artificial_Intelligence_for_the_Earth_report_2018.pdf (accessed on 21 April 2022). [114]

Wu, C. et al. (2022), "Sustainable AI: Environmental Implications, Challenges and Opportunities", https://arxiv.org/pdf/2111.00364.pdf. [24]

Zhang, X. et al. (2017), *Cooling Energy Consumption Investigation of Data Center IT Room with Vertical Placed Server*, https://doi.org/10.1016/j.egypro.2017.03.581. [70]

Annex A. Direct and indirect environmental impact indicators for AI compute and applications

Direct environmental impacts AI compute lifecycle

Production	Transport	Operations	End-of-life
• Carbon dioxide emissions from production *metric tons CO2e* • Carbon intensity of production methods *metric tons of CO2e / unit** • Renewable energy used for production *%* • Recycled or renewable materials used for production *%*	• Carbon dioxide emissions from transport *metric tons CO2e* • Carbon intensity of transport methods *metrics tons of CO2e / unit** • Low-carbon energy and/or renewable energy used for transport *%*	• **Energy** o Electricity consumption *TWh* o Renewable electricity consumption *TWh* o Power Usage Effectiveness (PUE) *total facility power / IT equipment power* • **Carbon** o Carbon dioxide emissions *metric tons CO2e* o Carbon intensity *metric tons CO2e / unit** o Carbon Usage Effectiveness (CUE) *carbon emissions caused by total facility power / IT equipment power* • **Water** o Water withdrawal *cubic meters* o Water consumption *cubic meters* o Water discharge (quality) *cubic meters* o Water Usage Effectiveness (WUE) *litres / kwh* o Water withdrawal intensity *cubic meters / unit**	• Electronic waste *metric tons* • Recycling rate *%* • Electronics disposal efficiency (EDE) *%* • Electronic waste sent to landfill *%*

*Unit e.g. per person or dollars of revenue

Indirect environmental impacts AI applications

- Net environmental cost or benefit of AI applications
  energy, water, carbon dioxide, etc. resources saved / used

Annex B. Experts consulted

Table B.1. OECD.AI Expert Group on AI Compute and Climate Co-chairs, members and observers, September 2022

Name	Title	Organisation	Group / Delegation
Ahuactzin, Juan Manuel	Research & Development Director	ProMagnus Company	Business
Aranda, Luis	Policy Analyst	OECD	Secretariat
Aristodemou, Leonidas	Analyst	OECD	Secretariat
Balasiano, Aviv	VP and Head of the Division	Technology Infrastructure in the Israeli Innovation Authority	Israel
Barrett, Gregg	CEO	Cirrus AI	Business
Bertrand, Arnaud	Chief Technical Officer and Senior Fellow	ATOS	Business
Bouvry, Pascal	Co-CEO	LuxProvide	Business
Caira, Celine	Economist/Policy Analyst	OECD	Secretariat
Cardoso Emediato de Azabuja, Eliana	General-Coordinator of Digital Transformation	Ministry of Science, Technology and Innovation	Brazil
Clark, Jack	[ONE AI Chair] Co-founder	Anthropic	Business
Elison, David	Senior AI Data Scientist	Lenovo	Business
Escobar Silva, Maria Jose	Associate Professor	Universidad Técnica Federico Santa María	Civil Society and Academia
Escobar, Rebeca	Head of Studies Center	Federal Telecommunications Institute	Mexico
Fernández Gómez, Liliana	Advisor	Digital Development Directorate - National Planning Department	Colombia
Formica-Schiller, Nicole	Board Member	German AI Association (KI-Bundesverband)	Germany
Frankle, Jonathan	Chief Scientist / Assistant Professor of Computer Science	Mosaic ML and Harvard University	Civil Society and Academia
Garg, Arti	Chief Strategist, AI Solutions	Hewlett Packard Enterprise	Business
Gibson, Garth	President	Vector Institute for AI	Civil Society and Academia
González Fanfalone, Alexia	Economist	OECD	Secretariat
Heim, Lennart	Research Affiliate	Centre for the Governance of AI	Civil Society and Academia
Holoyad, Taras	Standards expert	Federal Network Agency for Electricity, Gas, Telecommunications, Post and Railway	Germany
Hui, Chen	Assistant Chief Executive	Infocomm and Media Development Authority (IMDA)	Singapore
Janapa Redi, Vijay	Associate Professor	Harvard University John A. Paulson School of Engineering and Applied Sciences	Civil Society and Academia
Javoršek, Jan Jona	Head of Networking Infrastructure Centre	Jožef Stefan Institute	Civil Society and Academia
Kanter, David	Executive Director	ML Commons	Business

Kent, Suzette	Business Executive, Former Federal Chief Information Officer of the United States	Kent Advisory Services	United States
Khareghani, Sana	Head (former)	UK Office for AI	United Kingdom
Kirnberger, Johannes Leon	AI and climate expert	Consultant on AI and Climate	Consultant
Krüppel, Roland	Electronics and Autonomous Driving; Supercomputing	Federal Ministry for Education and Research	Germany
Lee, Jiwon	Policy Officer	Ministry of Technological Innovation and Digital Transition	Italy
Lohn, Drew	Senior Fellow	Georgetown University Center for Security and Emerging Technology	Civil Society and Academia
Macoustra, Angus	CTO, Head of Scientific Computing	Commonwealth Scientific and Industrial Research Organisation (CSIRO)	Australia
Manga, Uptal	VP and Senior Partner	IBM Global Business Services	Business
Matsuoka, Satoshi	Director	RIKEN Center for Computational Science	Japan
Moetzel, Ulrike	Economist/political scientist	Federal Ministry for Digital and Transport	Germany
Moretti, Lorenzo	Innovation Policy Coordinator to the Minister	Ministry of Technological Innovation and Digital Transition	Italy
Mujica, María Paula	Advisor on Digital Transformation, Management and Compliance	High Presidential Advisory Office	Colombia
Nolan, Alistair	Senior Policy Analyst	OECD	Secretariat
Ouimette, Marc-Etienne	Global Lead. AI Policy	Amazon Web Services	Business
Parashar, Manish	Director, Office of Advanced Cyberinfrastructure (OAC)	National Science Foundation	United States
Parker, Lynne	Deputy CTO of the United States of America (former)	United States Administration	United States
Perset, Karine	Head of Unit, OECD.AI	OECD	Secretariat
Rao, Anand	Global AI Leader	PwC	Business
Roquet, Ghilaine	Vice President of Strategy and Planning	New Digital Research Organization (NDRIO)	Civil Society and Academia
Sampaio Gontijo, José Gustavo	Director	Department of Digital Science, Technology and Innovation	Brazil
Stancavage, Jayne	Global Executive Director, Digital Infrastructure Policy	Intel	Business
Stogiannis, Dimitris	Head of the Research, Development and Innovation (RDI) Statistics Unit	National Documentation Centre	Greece
Strier, Keith	[ONE AI Chair] Vice President	NVIDIA Worldwide AI Initiative	Business
Tretikov, Lila	CVP, Deputy Chief Technology Officer	Microsoft	Business
Tyldesley, Jennifer	[ONE AI Chair] Deputy Director, Economic Security	Department for Digital, Culture, Media and Sport (DCMS)	United Kingdom
Vasilis, Bonis	Technical Manager, Team Leader, Senior Software Architect and Technical Coordinator of European Projects	National Documentation Centre	Greece
Velsberg, Ott	Chief Data Officer	Ministry of Economic Affairs	Estonia
Weber, Verena	Head of Communication Infrastructures and Services Policy Unit	OECD	Secretariat
Yeong, Zee Kin	Assistant Chief Executive	Infocomm Media Development Authority of Singapore	Singapore
Zagler, Martin	Danish Business Authority	Ministry of Industry, Business and Financial Affairs	Denmark

Note: Member biographies are available on OECD.AI

Table B.2. GPAI Project RAISE co-chairs and members consulted

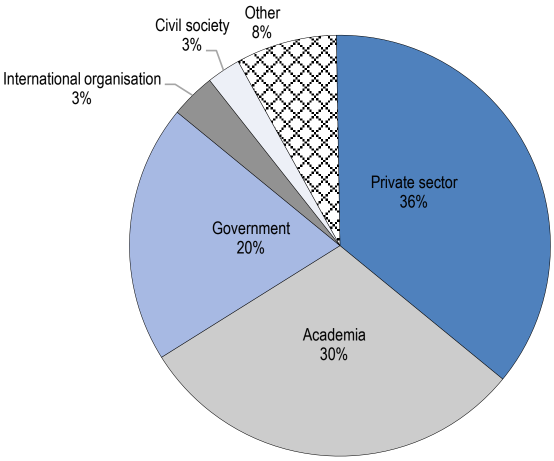
Name	Title	Group
Badiqué, Eric	Member	European Commission, Advisor
Chatila, Raja	Co-Chair	Global Partnership on AI (GPAI) Responsible AI Strategy for the Environment (RAISE)
Clutton-Brock, Peter	Member	Radiance and Centre for AI and Climate
Dignum, Virginia	Member	Umea University
Miailhe, Nicolas	Co-Chair	Global Partnership on AI (GPAI) Responsible AI Strategy for the Environment (RAISE)
Hodes, Cyrus	Member	FoundersX
Rolnick, David	Member	McGill University and Climate Change AI
Tiedrich, Lee	Co-Chair	Global Partnership on AI (GPAI) Responsible AI Strategy for the Environment (RAISE)

Table B.3. Other experts consulted

Name	Organisation
Andersson, Mats	Lefdal Mine Datacenter
Bachar, Yuval	EdgeCloudLink
Burdon, Thyme	OECD, Science, Technology and Innovation Directorate
Cannon, Greg	Amazon Web Services
Dechezlepretre, Antoine	OECD, Science, Technology and Innovation Directorate
Donti, Priya	Carnegie Mellon University and Climate Change AI
Gimpel, Lea	Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ)
Griffin, Conor	DeepMind
Havens, John	Institute of Electrical and Electronics Engineers
Herron, Colin	Global Water Partnership
Kaack, Lynn	Hertie School and Climate Change AI
Kamiya, George	International Environmental Agency
Luccioni, Sasha	Mila
Maslej, Nestor	Stanford University Human-Centered Artificial Intelligence (Stanford HAI)
Messner, Dirk	German Environmental Agency
Mytton, David	Console
Patterson, David	Google
Prag, Andrew	OECD, Environment Directorate
Rejeski, David	Environmental Law Institute
Schimpf, Klaus	Texas Instruments
Srichaikul, Piyawut	National Science and Technology Development Agency Supercomputer Center (ThaiSC)
Strubell, Emma	Carnegie Mellon University
Watkins, Chris	DeepMind
Witherspoon, Sims	DeepMind
Wittich, Jeff	Ampere Computing

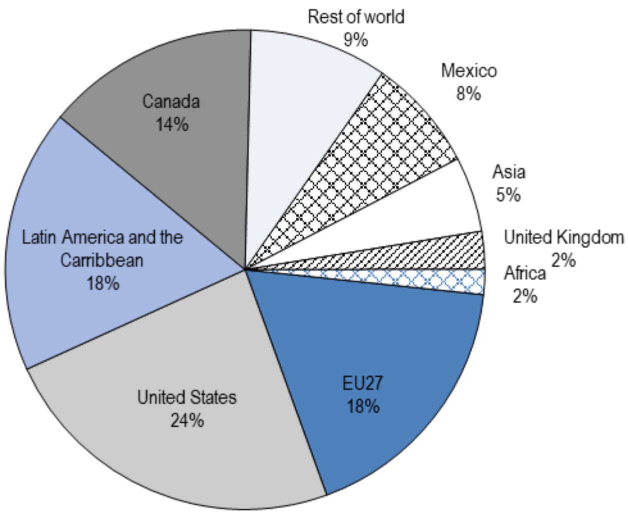
Annex C. Preliminary AI compute survey results

Figure C.1. What kind of organisation do you represent?



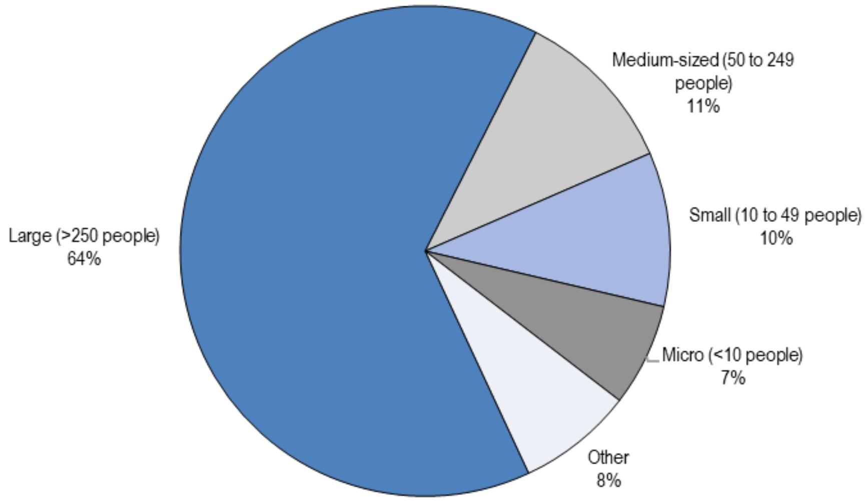
Note: Of 118 respondents who partially or fully completed the survey, 116 respondents answered this question.
Source: OECD.AI Expert Group on AI Compute and Climate, survey on AI compute (March-April 2022)

Figure C.2. Geographic distribution of OECD AI compute survey respondents



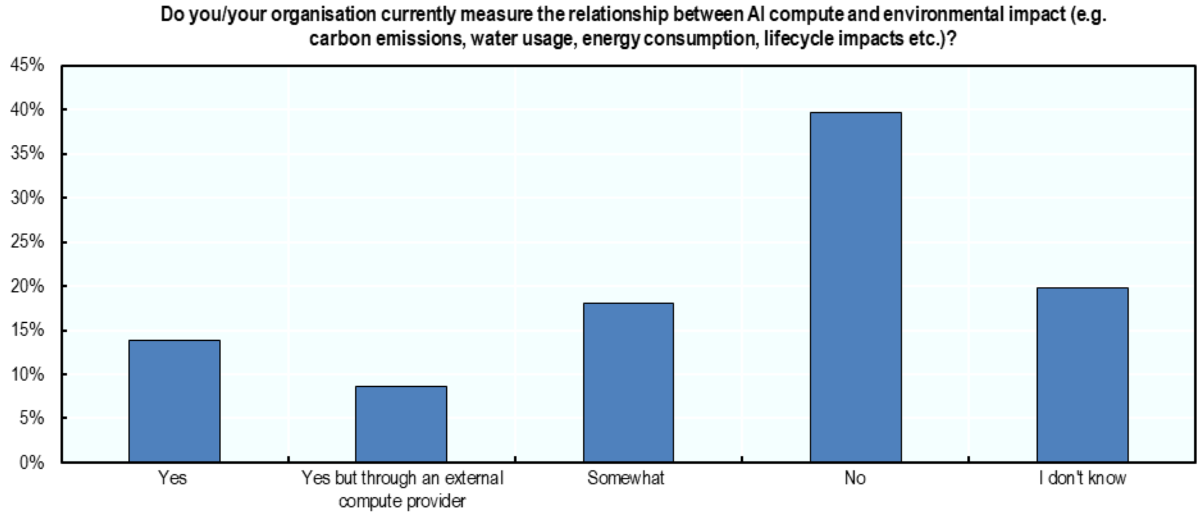
Note: Of 118 respondents who partially or fully completed the survey, 118 respondents answered this question.
Source: OECD.AI Expert Group on AI Compute and Climate, survey on AI compute (March-April 2022)

Figure C.3. Organisation or enterprise size of OECD AI compute survey respondents



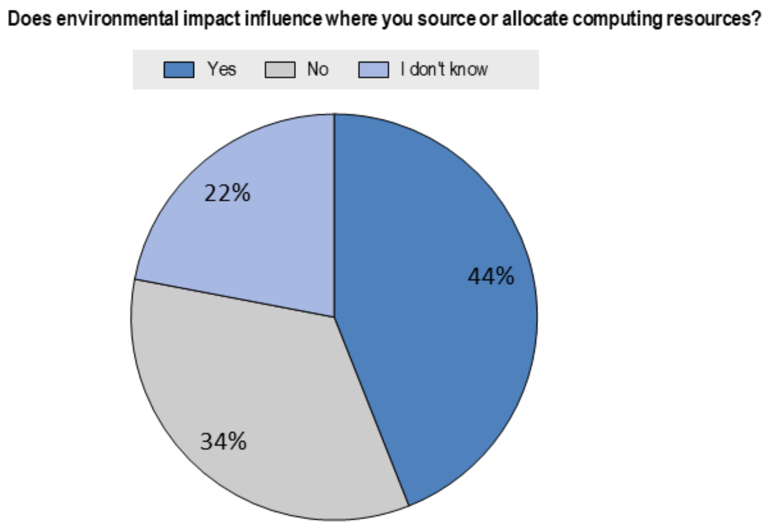
Note: Of 118 respondents who partially or fully completed the survey, 118 respondents answered this question. According to OECD, small and medium-sized enterprises (SMEs) employ fewer than 250 people. SMEs are further subdivided into micro enterprises (fewer than 10 employees), small enterprises (10 to 49 employees), medium-sized enterprises (50 to 249 employees). Large enterprises employ 250 or more people.
Source: OECD.AI Expert Group on AI Compute and Climate survey on measuring AI compute (March-April 2022)

Figure C.4. Environmental impact measurement activities by OECD AI compute survey respondents



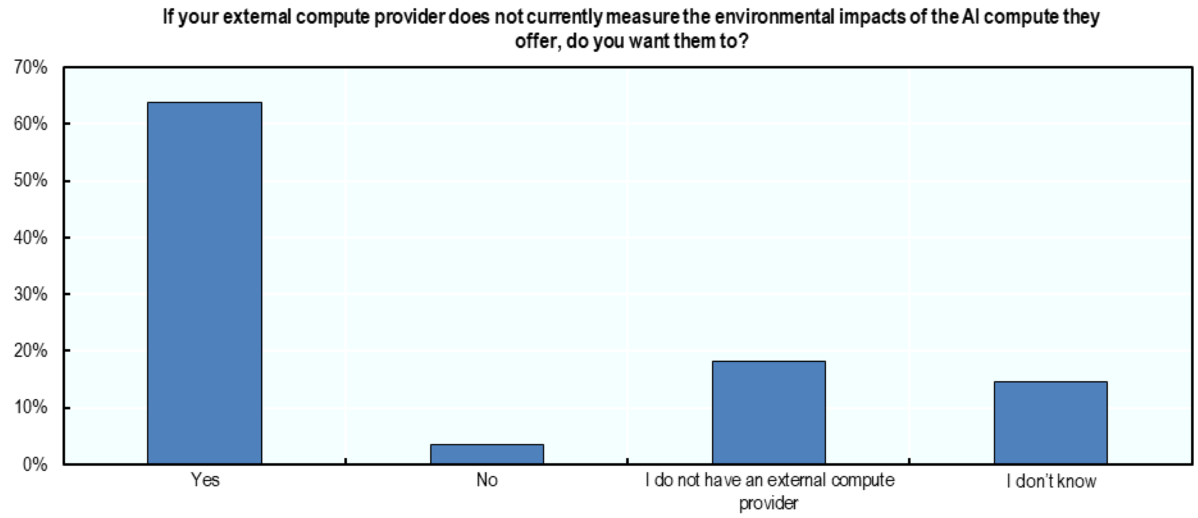
Note: Of 118 respondents who partially or fully completed the survey, 116 respondents answered this question.
Source: OECD.AI Expert Group on AI Compute and Climate, survey on AI compute (March-April 2022)

Figure C.5. Influence of environmental impact on OECD AI compute survey respondents



Note: Of 118 respondents who partially or fully completed the survey, 116 respondents answered this question.
Source: OECD.AI Expert Group on AI Compute and Climate, survey on AI compute (March-April 2022)

Figure C.6. Desire for environmental impact measurement by external providers of OECD AI compute survey respondents



Note: Of 118 respondents who partially or fully completed the survey, 116 respondents answered this question.
Source: OECD.AI Expert Group on AI Compute and Climate, survey on AI compute (March-April 2022)

Notes

[1] According to Berkhout and Hertin (2001), "first order impacts: direct environmental effects of the production and use of ICTs (resource use and pollution related to the production of ICT infrastructure and devices, electricity consumption of ICT hardware, electronic waste disposal); second order impacts: indirect environmental impacts related to the effect of ICTs on the structure of the economy, production processes, products and distribution systems; the main types of positive environmental effects are dematerialisation (getting more output for less resource input), virtualisation (the substitution of information goods for tangible goods) and 'demobilisation' (the substitution of communication at a distance to travel); third order impacts: indirect effects on the environment, mainly through the stimulation of more consumption and higher economic growth by ICTs ('rebound effect'), and through impacts on life styles and value systems."

[2] According to the IPCC, "Scope 1, Scope 2, and Scope 3 emissions: Emissions responsibility as defined by the GHG Protocol, a private sector initiative. 'Scope 1' indicates direct greenhouse gas (GHG) emissions that are from sources owned or controlled by the reporting entity. 'Scope 2' indicates indirect GHG emissions associated with the production of electricity, heat, or steam purchased by the reporting entity. 'Scope 3' indicates all other indirect emissions, i.e., emissions associated with the extraction and production of purchased materials, fuels, and services, including transport in vehicles not owned or controlled by the reporting entity, outsourced activities, waste disposal, etc." (Allwood et al., 2014[152]).

[3] A large cloud compute provider does not wish to be attributed by name due to commercial confidentiality concerns.