AI ethics in action:
the case of the EU's ethics review procedure

AIDA AICET2025

Dr Mihalis Kritikos, EGE-European Commission

European approach to artificial intelligence

Communication **'Artificial Intelligence for Europe'** (European Commission, 2018),

Declaration of cooperation on AI (European Commission, 2018),

Coordinated Action Plan on the development of AI in the EU

White Paper on AI – **A European Approach to Excellence and Trust**

Main axes of EU's involvement

LEGAL GOVERNANCE OF AI: AI ACT/EU PACT/

INNOVATION: FUNDING OF AI RESEARCH

AI FACTORIES

COORDINATED PLAN ON AI INTERNATIONAL POLICY WORK

ETHICAL GOVERNANCE OF AI: PRINCIPLES AND GUIDELINES

CODES OF PRACTICE/ STANDARDS

- HGAI / EU GUIDELINES ON AI/ALTAI LIST
- HUMAN RIGHTS IMPACT ASSESSMENT METHODOLOGY
- AI ETHICS APPRAISAL MECHANISM
- AI ETHICS STANDARDS AND TRAINING ACTIVITIES
- EU-FUNDED RESEARCH PROJECTS ON AI VALUES/ETHICS
- OPERATIONAL GUIDELINES /TOOLBOXES

Ethics Guidelines for Trustworthy AI

7 core requirements:

1. **Human agency and oversight**
2. **Technical robustness and safety**
3. **Privacy and Data Governance**
4. **Transparency**
5. **Diversity, non-discrimination and fairness**
6. **Societal and environmental wellbeing**
7. **Accountability**

The EU’s ethical approach for AI in research

- AI as a **self-standing ethics issue** in Horizon Europe

- **‘Ethics by design’ approach** integrated into **Horizon Europe** ethics appraisal process.

- **Dedicated evaluation of ‘robustness’ of AI in scientific evaluations** of **Horizon Europe** proposals

- **Trustworthy and Ethical AI Research** is the general approach for Horizon Europe

European Commission

8 ARTIFICIAL INTELLIGENCE	YES/NO		Information to be provided	Documents to be provided/kept on file
Does this activity involve the development, deployment and/or use of Artificial Intelligence-based systems?	☐	☐	1) Explanation as to how the participants and/or end-users will be informed about: - their interaction with an AI system/technology (if relevant); - the abilities, limitations, risks and benefits of the proposed AI system/technique; - the manner in which decisions are taken and the logic	1) Detailed risk assessment accompanied by a risk mitigation plan (if relevant). These must cover the development, deployment and post-deployment phases. 2) Copies of ethics approvals (if relevant).

Artificial intelligence

E.g.: *Could the AI system/technique stigmatise or discriminate against people (based on sex, race, ethnic/social origin, age, disability, sexual orientation, religion, political affiliation, etc.)?*

→Applicants must explain how potential bias, discrimination and stigmatization **could arise**.

→Applicants must indicate **how it will be addressed**.

- E.g., testing for algorithmic bias during the detailed development phase by using counterfactual evaluation methods
- E.g., testing whether the system becomes unequally functional for different end-users, and adapt design to ensure that interface design is universally accessible

→ **Key ethical requirements**:

- ⇨ People must be made **aware** that they are interaction with an AI system, its abilities and limitations, risks and benefits
- ⇨ Mechanisms for **human oversight**, **transparency** and **auditability** must be ‘built in’ the AI system
- ⇨ AI-system must be designed to **avoid bias** in input data and algorithmic design
- ⇨ Compliance with **data protection and privacy principles** (e.g., data minimisation) must be demonstrated
- ⇨ The **impact** of the developed and/or used AI system/technique on the individual, society and environment must be **carefully evaluated and any possible risk of harm must be avoided**.

Guidance on Ethics By Design for AI

European Commission

Ethics By Design and Ethics of Use Approaches
for Artificial Intelligence

Version 1.0
25 November 2021

Builds on the work of the **Independent High-Level Expert Group on AI** and their **'Ethics Guidelines for Trustworthy AI'**

Addressing ethical issues during research and development

Ethics guidance

- 'Ethics by design' requirements:
  - People must be made aware that they are interacting with an AI system, its abilities and limitations, risks and benefits
  - Mechanisms for human oversight, transparency and auditability must be 'built in' the AI system
  - AI-system must be designed to avoid bias in input data and algorithmic design
  - Compliance with data protection and privacy principles (e.g., data minimisation) must be demonstrated
  - The impact of the developed and/or used AI system/technique on the individual, society and environment must be carefully evaluated and any possible risk of harm must be avoided.

Ethics By Design and Ethics of Use Approaches for Artificial Intelligence

Version 1.0
25 November 2021

European Commission

EU Operational Guidelines in ethics /Publication foreseen for June 2025

Guidance Note on *human-centered AI: algorithmic bias and fairness*

Guidance Note on *AI and informed consent*

Guidance Note on *AI Ethics Audits and Checks*

Guidance Note on *Human Rights Impact Assessment*

International initiatives on the ethics of AI

UNESCO
'Recommendation on the Ethics of Artificial Intelligence' adopted by all 193 Member States.

OECD AI principles

(first set of intergovernmental policy guidelines on AI)

G7

Hiroshima AI Process Comprehensive Policy Framework

The Bletchley Declaration

Global AI Safety Summit

Council of Europe Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law

WHO

Global Initiative on AI for Health

Guidelines on the use of Generative AI for research

- **Reliability -** This includes aspects related to verifying and reproducing the information produced by the AI for research. It also involves being aware of possible equality and non-discrimination issues in relation to bias and inaccuracies.
- **Honesty** - This principle includes disclosing that generative AI has been used.
- **Respect -** Responsible use of generative AI should take into account the limitations of the technology, its environmental impact and its societal effects (bias, diversity, non-discrimination, fairness and prevention of harm). This includes the proper management of information, respect for privacy, confidentiality and intellectual property rights, and proper citation.
- **Accountability** - This includes responsibility for all output a researcher produces, underpinned by the notion of human agency and oversight.

Guidelines on the use of Generative AI for research

- Builds upon the ALLEA Code (2023) and Guidelines for trustworthy AI (2019)
- Concrete recommendations for researchers, research organisations, and research funders.
  - Full accountability for the content
  - Due citation and disclosure
  - Avoiding and caution for plagiarism when using generative AI
  - Refrain from using generative AI for peer review or evaluating research proposals! (To avoiding bias, faults, and protecting confidentiality of unpublished work)

Communication on boosting startups and innovation (January 2024)

- AIM: **to support Artificial Intelligence startups and SMEs.**
- Several measures to support European startups and SMEs **in the development of trustworthy AI that respects EU values and rules.**
- Sets out a **strategic investment framework in trustworthy AI** for the Union to capitalize on its assets, in particular its world-leading supercomputing infrastructure, and to foster an innovative European AI ecosystem.
- **"GenAI4EU:"** to stimulate the uptake of generative AI across the Union's key strategic industrial ecosystems and that will encourage the development of large open innovation ecosystems

AI Continent Action Plan (April 2025)

Aim: to make Europe a global leader in AI.

- The Action Plan focuses on developing trustworthy AI technologies to enhance Europe's competitiveness while safeguarding and advancing our democratic values. It aims to bring the benefits of AI to various sectors such as healthcare, education, industry, and environmental sustainability.
- Actions to build large-scale AI data and computing infrastructures, increase access to high-quality data, foster AI adoption in strategic sectors, strengthen AI skills and talent, and facilitate the implementation of the AI Act.
- Establishment of AI Factories and Gigafactories, the InvestAI Facility and AI Skills Academy.
- .

Framework of AI EU initiatives

AI Continent

Apply AI Strategy

AI in Science Strategy

Data Union Strategy

Cloud and AI Development Act

AI Gigafactories

....

RAISE
(Resource for AI Science in Europe)

AI in science strategy - Political and Policy context

European Commission Political Guidelines 2024-2029

- ‘**A new plan for Europe’s sustainable prosperity and competitiveness**’ and with ‘**research and innovation at the heart of our economy**’

  Draghi report: focus on advanced technologies (such as AI) to regain competitiveness.

  Letta report: position the EU as global AI hub and win the race of AI applications

  Competitiveness Compass: reducing the innovation by gap with the technologies of the future, in particular AI.

Council conclusions PL presidency on AI in science (23rd May)

Global geopolitical context: Digital sovereignty, investments on AI, role of Big Tech

Commissioner Zaharieva’s mission letter:

*You will put forward **a strategy to increase the uptake of AI by European scientists** and help set up a European AI Research Council where we can pool all of our resources.*

The ‘AI in science’ opportunities for the EU

- Science is one of the areas with more potential benefits in the application of AI: accelerate scientific discovery, tangible societal benefits, new business opportunities…
- The EU is well placed in science to capitalise AI (scientists, research infrastructures and scientific data ecosystem, open science...)
- In the use of AI in science, there is room for major technical innovations in AI. Complex and challenging problems requiring new approaches.
- Leadership in AI in science could keep the EU as a scientific leader and a key actor in the global AI race.

Goals of the ‘AI in science’ strategy

Use AI to make EU science more impactful and productive

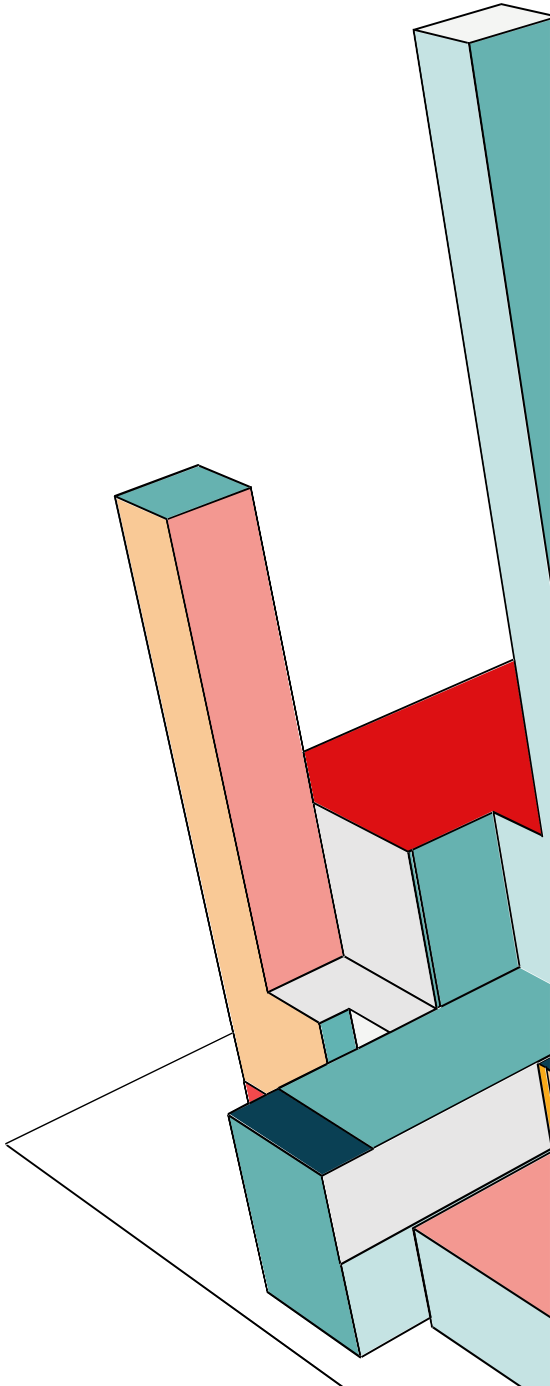
- The EU will produce more and better science
- Accelerate and advance science facilitating innovations to drive economic growth and tackle societal benefits (health, climate change…).
- Keep the EU as a leader in science.

Accelerating the uptake: More researchers using AI as a tool, and more advanced and complex research with AI

Managing the impacts on the scientific process

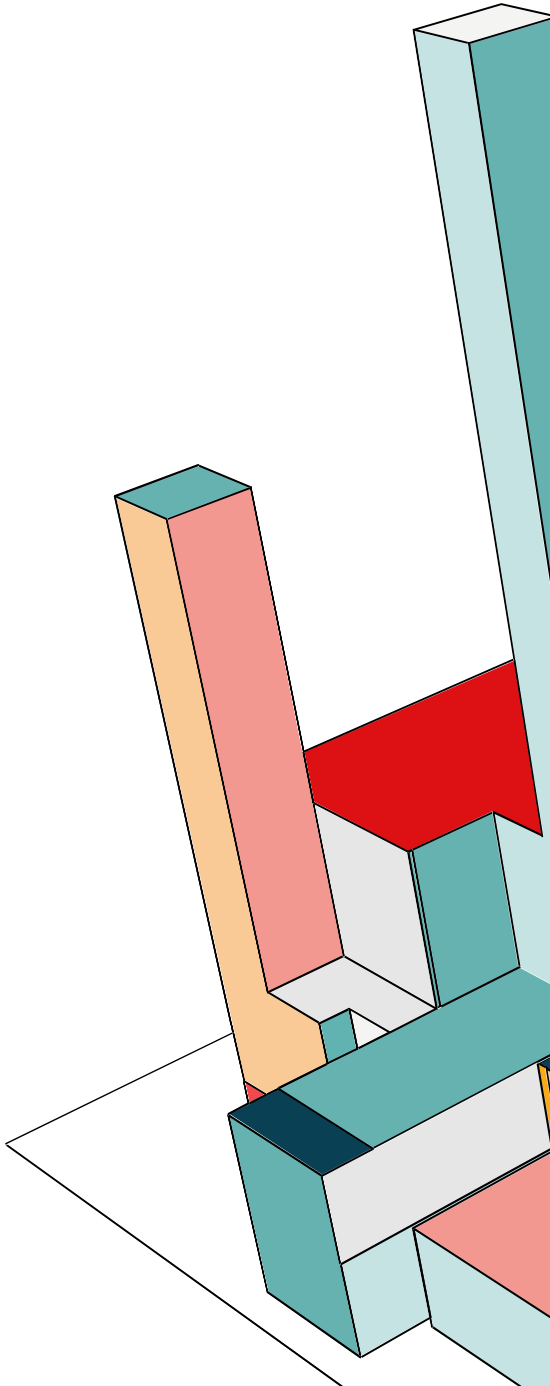
HOW DOES AI AFFECT THE ENVIRONMENT?

- **Large-scale AI deployments pose several environmental concerns.**
- Most AI servers are stored in data centres, which produce electronic waste and can contain toxic chemicals, such as mercury and lead.
- Data centres consume **vast amounts of electricity, creating greenhouse gas emissions.**
- They also require **large amounts of water for construction and to cool the electrical components.**
- Global AI demand is expected to consume **4.2-6.6 billion cubic meters of water by 2027**, surpassing Denmark's total annual water withdrawal of 4-6 billion cubic meters.



IS THE DIGITAL ECONOMY ONLY VIRTUAL?

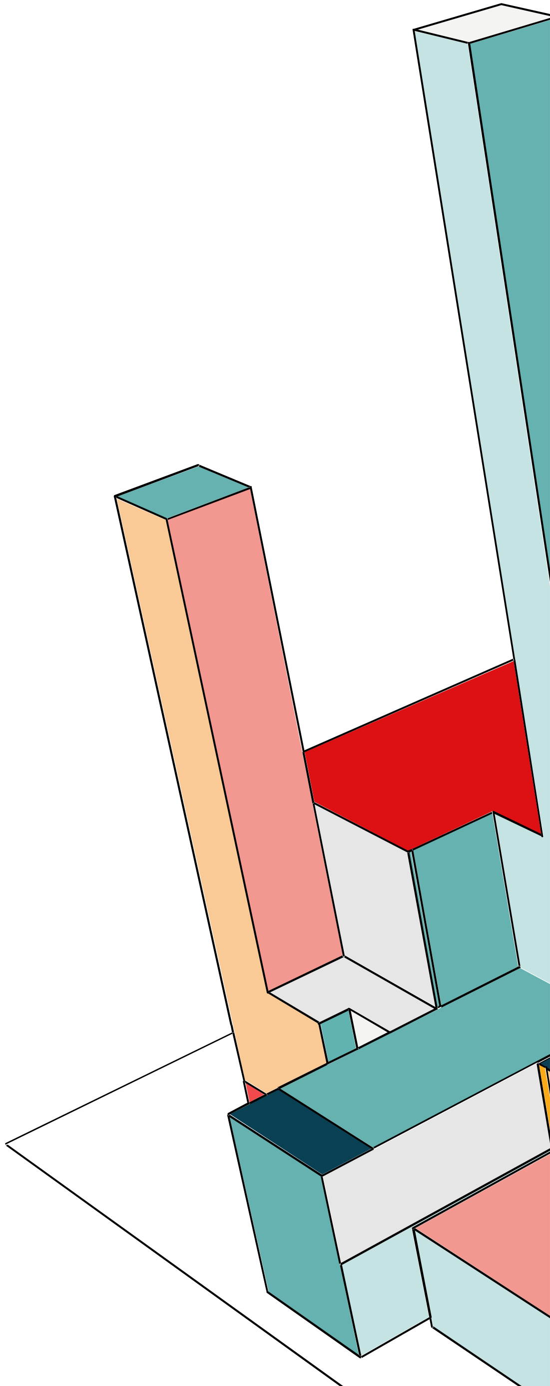
- Although the digital economy is sometimes viewed as virtual or in the "cloud", it is nonetheless **highly reliant on physical resources and raw materials.**
- **Digital devices, hardware and infrastructure are made of plastics, glass, ceramics, and various minerals and metals.**
- **Data centres rely on minerals and rare elements,** which are often mined unsustainably. Making a 2-kilogram computer requires approximately 800 kg of raw materials.
- Evaluating AI's hardware life cycle is complex because each stage has its environmental impact, **from mining and extraction to transportation, energy and water consumption, and e-waste generation.**



SOFTWARE AND HARDWARE

The life cycle of AI can be seen through two lenses: software and hardware.

- The software life cycle involves data collection and preparation, model development, training, validation, deployment, inference, maintenance and retirement.
- The hardware life cycle involves the production of computer chips, including graphical processing units (GPUs) essential to the training and inference, and the construction and operation of data centres.
- It starts with raw material extraction, through manufacturing, shipping and data centre construction, and ends with operation, maintenance and disposal of e-waste.



CHALLENGES AND RISKS OF AI IN CLIMATE ACTION

- **Energy Consumption:**

AI systems require significant energy to operate, and a growing portion of AI's energy use comes from data centers, which themselves can contribute to emissions.

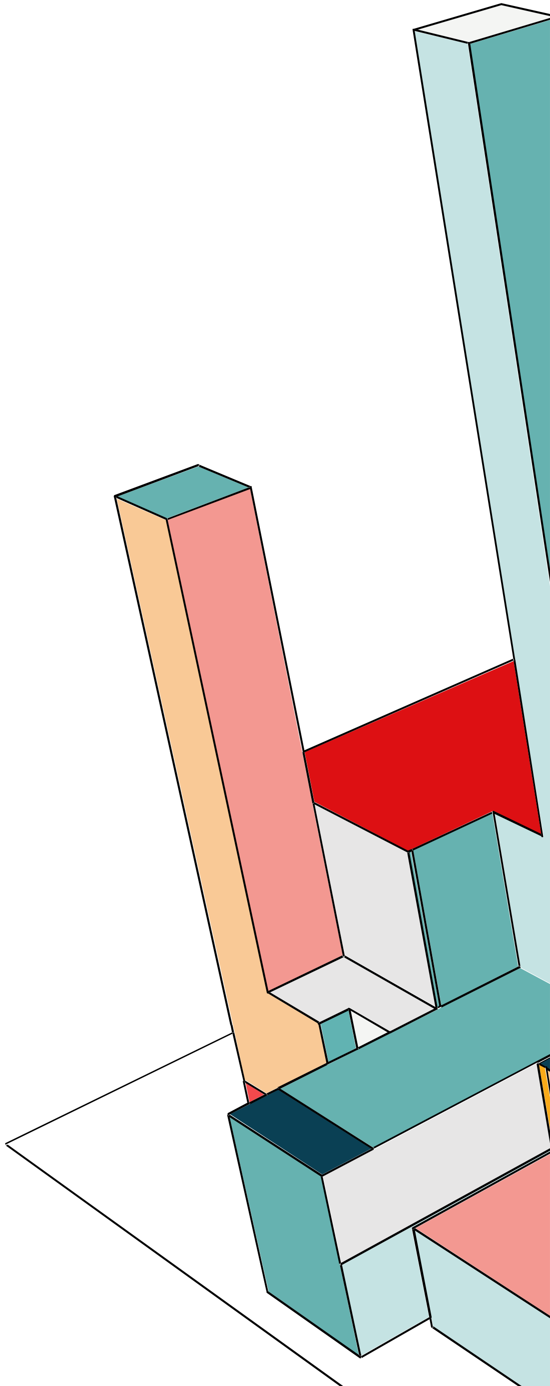
Energy-intensive AI training: Training advanced AI models uses a lot of electricity, increasing energy demand.

E-waste and short device lifespans: The rapid turnover of digital devices continues to contribute to global waste problems.

Growing internet and data traffic: Increased streaming, cloud computing, and blockchain operations demand ever-expanding data infrastructure.

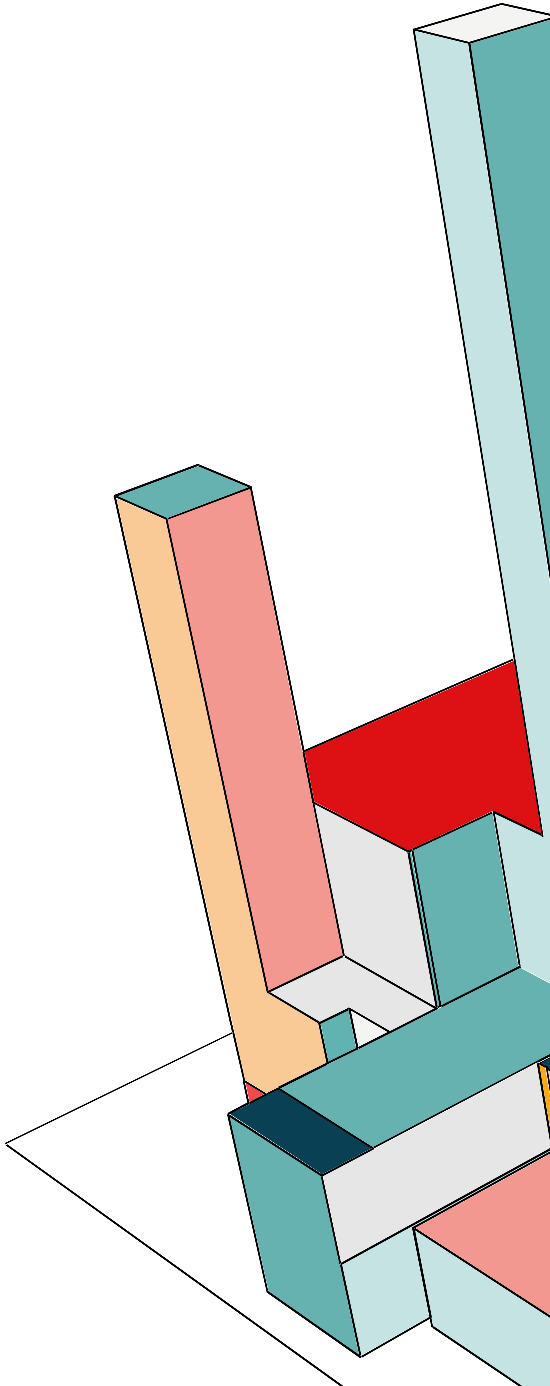
EU POLICY RESPONSES

- Need to advance EU's 'twin transition' strategy, which blends digitalisation with green goals through the Green Deal Industrial Plan, aimed at boosting clean technology manufacturing and cutting down on fossil fuel usage.
- The new Competitiveness Compass sets out Europe's long-term vision, emphasising digital innovation and sustainability as key to future economic strength and prosperity.
- The Clean Industrial Deal aims to support the transition to a more sustainable and competitive industry by leveraging digitalization and clean technologies, including AI.



EU POLICY RESPONSES

- Sustainable development is listed as one of the 7 Requirements in EU's Principles for Trustworthy AI
- Existing instruments include:
- the Ecodesign Regulation on servers and data storage products
- the EU Code of Conduct on Data Centre Energy Efficiency
- the EU Green Public Procurement criteria for data centres, server rooms and cloud services



EU POLICY INITIATIVES

- The Commission Proposal for a Directive on energy efficiency (recast) introducing new elements to improve the energy efficiency and sustainability monitoring of data centers;

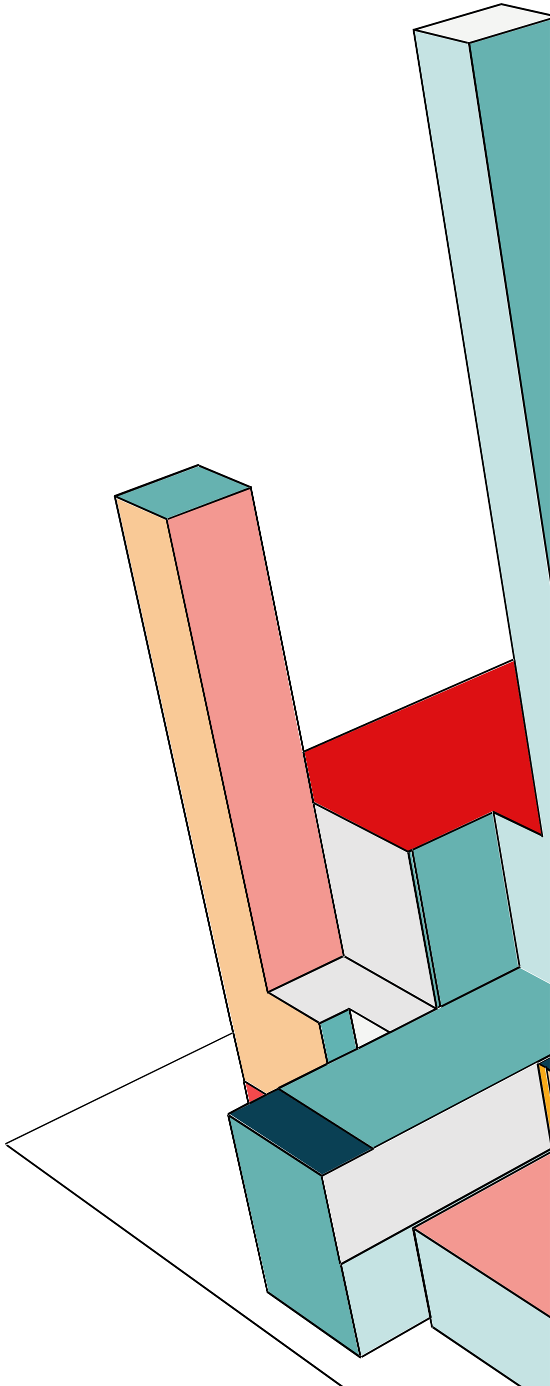
- The Taxonomy Regulation and its Delegated Act adopted in July 2021, which sets the framework for investments to be qualified as sustainable and whose delegated act is currently under finalisation, has a section on data centres;

UN Global Compact

The UN Global Compact advocates for a principles-based approach to responsible business, anchored on respecting and supporting human rights, the environment, labour and anti-corruption.

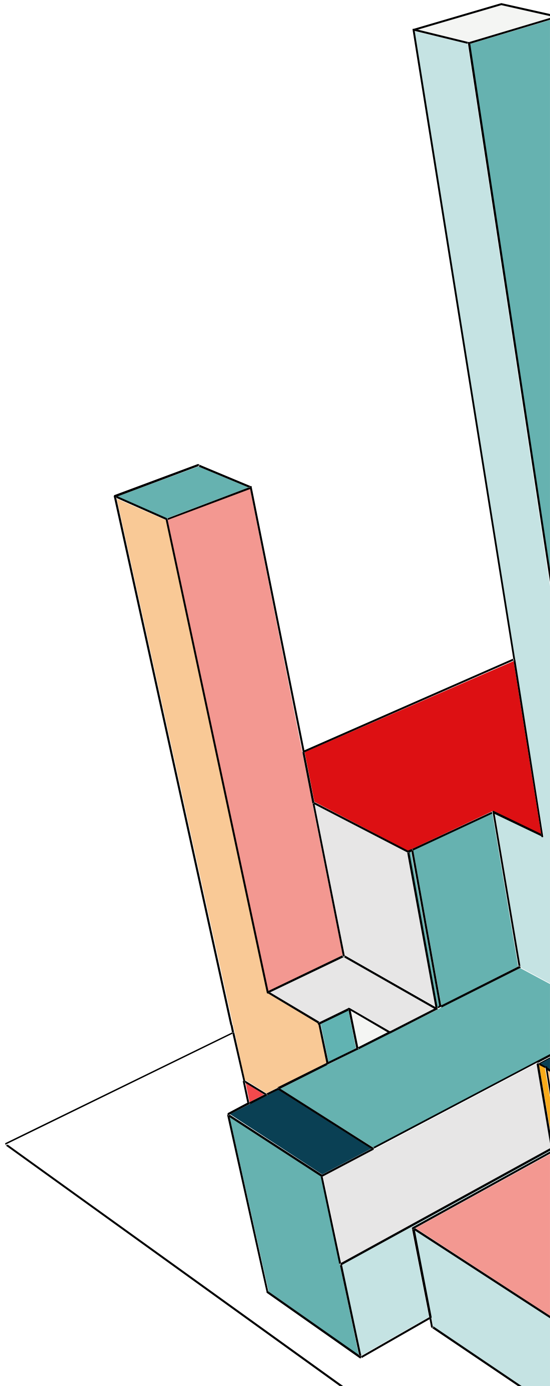
The UN Global Compact further challenges companies to leverage AI to maximize their contributions to the 2030 Agenda.

These frameworks can be used to help assess and mitigate risks around AI development and implementation and to guide companies in leveraging AI applications to advance in key areas where they can have the greatest impact on sustainable development.



UN ENVIRONMENT PROGRAMME

- Countries develop **standardized methods to measure AI's environmental footprint.**
- Governments **plan to develop regulations requiring companies to disclose the environmental impact of AI-based products and services.**
- **Tech companies aim at making AI algorithms more energy-efficient, reducing their energy demand while recycling water and reusing components where feasible.**
- Countries encourage organizations **to use renewable energy and carbon offset to green their data centres.**
- AI-related policies should also be integrated into broader environmental regulations.



ISSUES NOTE
21 September 2024

UN environment programme

Artificial intelligence (AI) end-to-end:

The environmental impact of the full AI life cycle needs to be comprehensively assessed

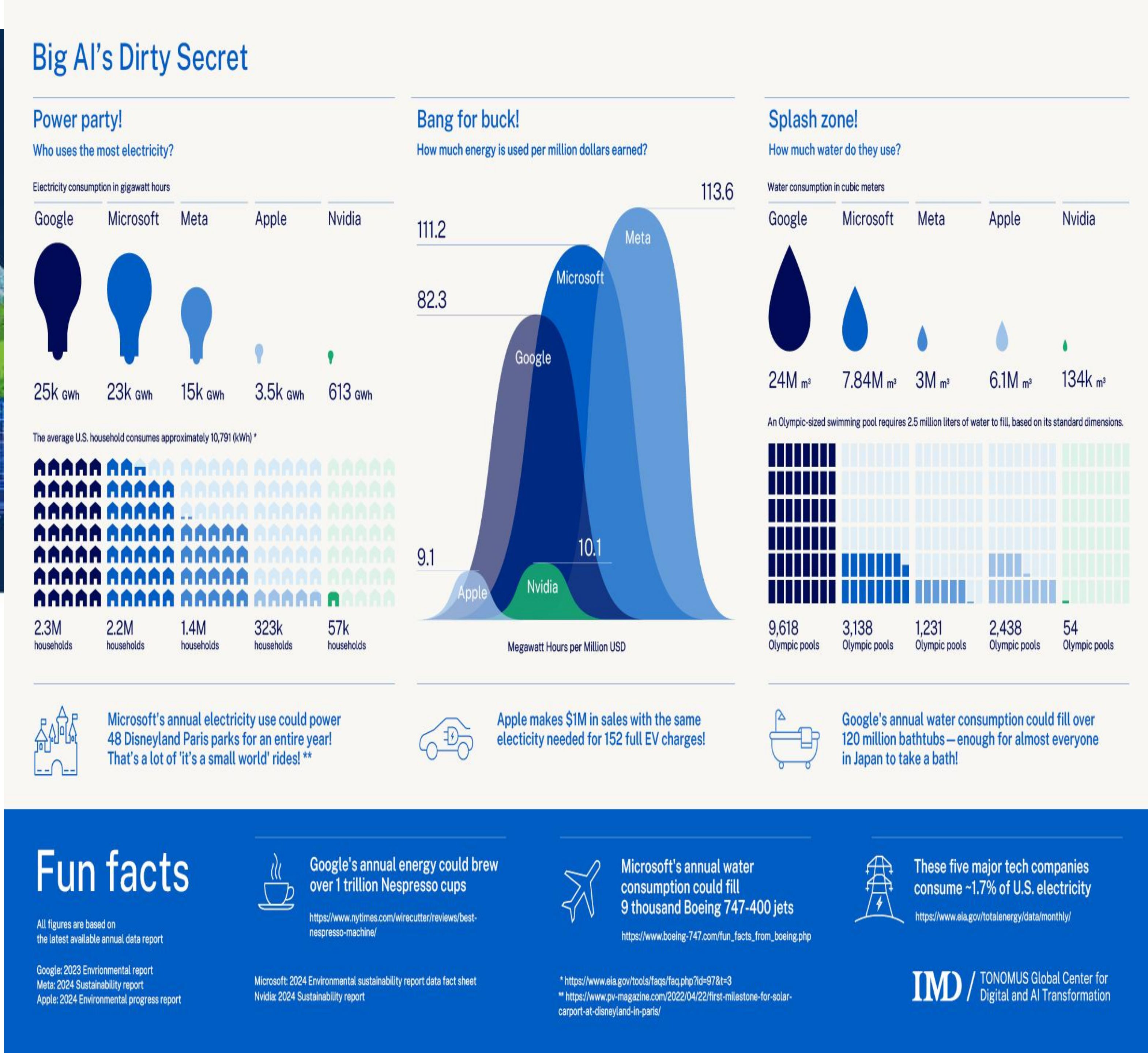
The United Nations Environment Programme (UNEP) is the leading global environmental authority that sets the global environmental agenda, promotes the environmental dimension of sustainable development within the United Nations system, and serves as an authoritative advocate for the global environment with a mandate to keep under review the world environmental situation.

Against this mandate, UNEP has been requested by UN Member States to consider the environmental dimensions of digital technologies, assessing their opportunities to enable environmental sustainability and the impact they can have on the environment.

This note outlines key areas identified by UNEP regarding the environmental impact of Artificial intelligence (AI) across its life cycle. The note aims to inform Member States, civil society, the private sector and the public, while encouraging the research community to develop and use scientific methods to allow the objective measurement of AI's environmental footprint.

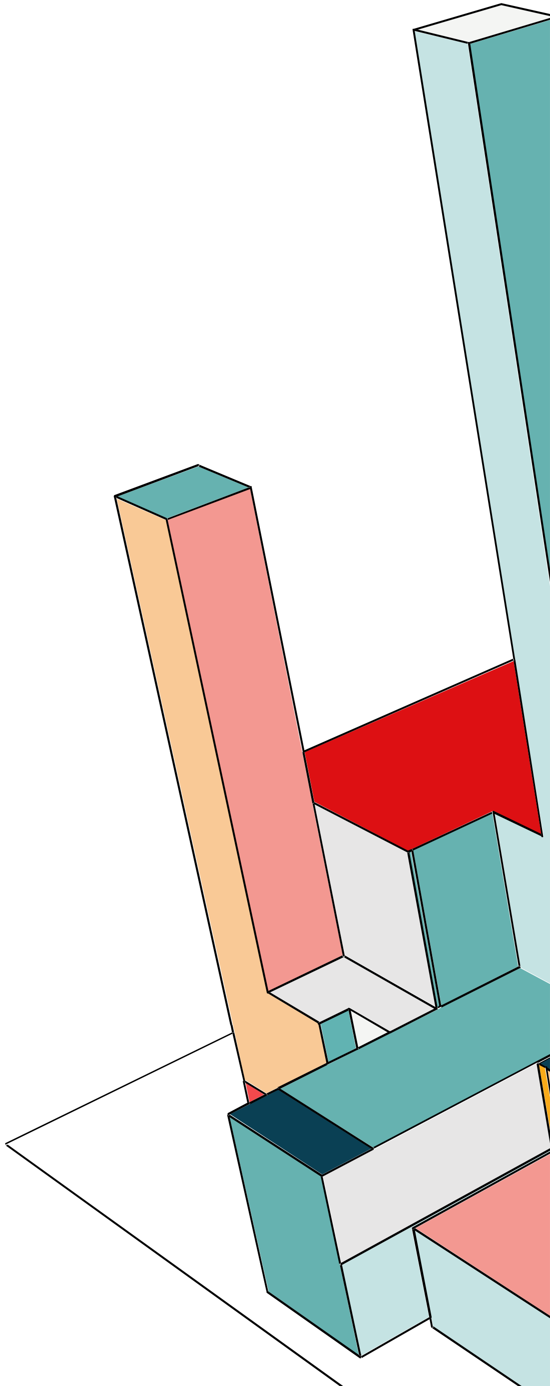
[1] This Issues Note provides a review of the latest literature on specific topics that are of relevance to UNEP's mandate as outlined in "The future we want: outcome of the Conference on Sustainable Development, Rio de Janeiro, Brazil, 20–22 June 2012, paragraph 88". It also presents a set of agreed approaches and recommendations regarding UNEP's communication of the subject matter.

[2] The Ministerial Declaration adopted at the sixth session of the United Nations Environment Assembly (UNEA-6) emphasized the importance of leveraging emerging technologies and closely monitoring their development to ensure they contribute to sustainability. The declaration also stressed the need to ensure that digitalization is inclusive, equitable and sustainable.



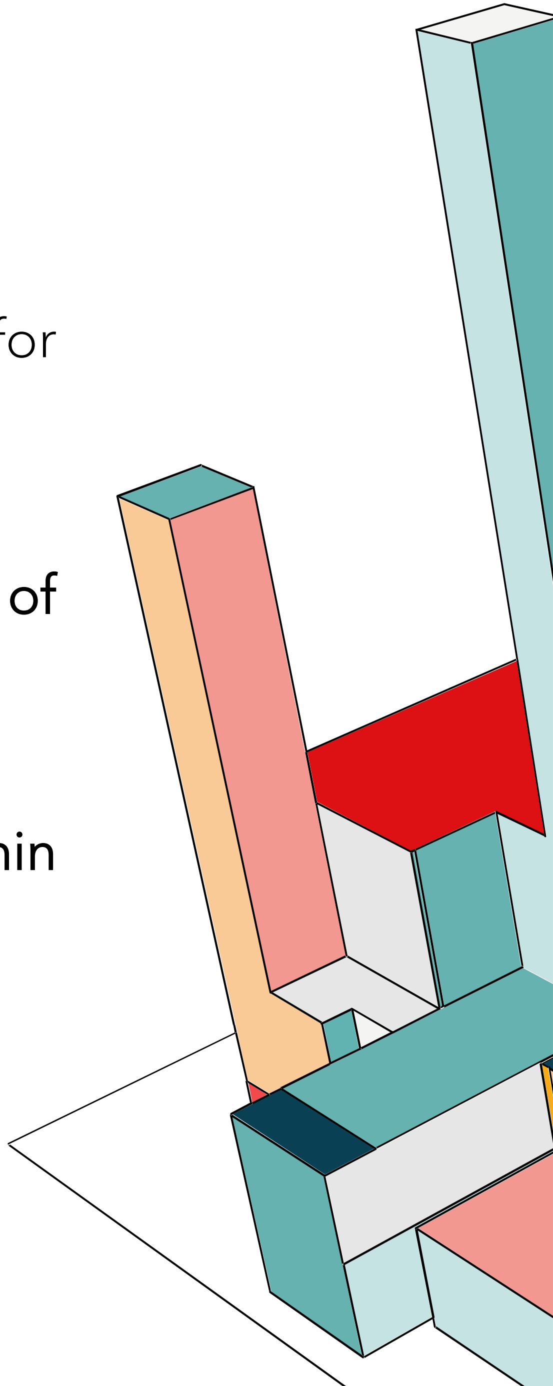
CHALLENGES

- How to measure and confront the environmental consequences of our digital lifestyles.
- The environmental impact of digitalization is a global issue, but its effects are not evenly distributed.
- Developing countries, often rich in the resources needed for digital technologies, bear a disproportionate burden of its costs while reaping limited benefits.
- Discarded smartphones, laptops, screens and other electronic devices grew by 30 per cent between 2010 and 2022, reaching 10.5 million tons globally.



CHALLENGES

- Need to measure the impact of the extraction of minerals essential for digital technologies on environmental and social sustainability.
- Unsafe working conditions, environmental degradation and exploitation of vulnerable communities, including children.
- Need for greater transparency and responsible sourcing practices within the digital supply chain
- Human rights issues / benefit sharing/ corruption phenomena



SUSTAINABLE GEN AI

MEANS

DEVELOP AND DEPLOY GEN AI RESPONSIBLY

Companies must ensure that the **means** used to develop and deploy Gen AI are ethical and transparent. The private sector can do this through adopting both responsible processes and governance.

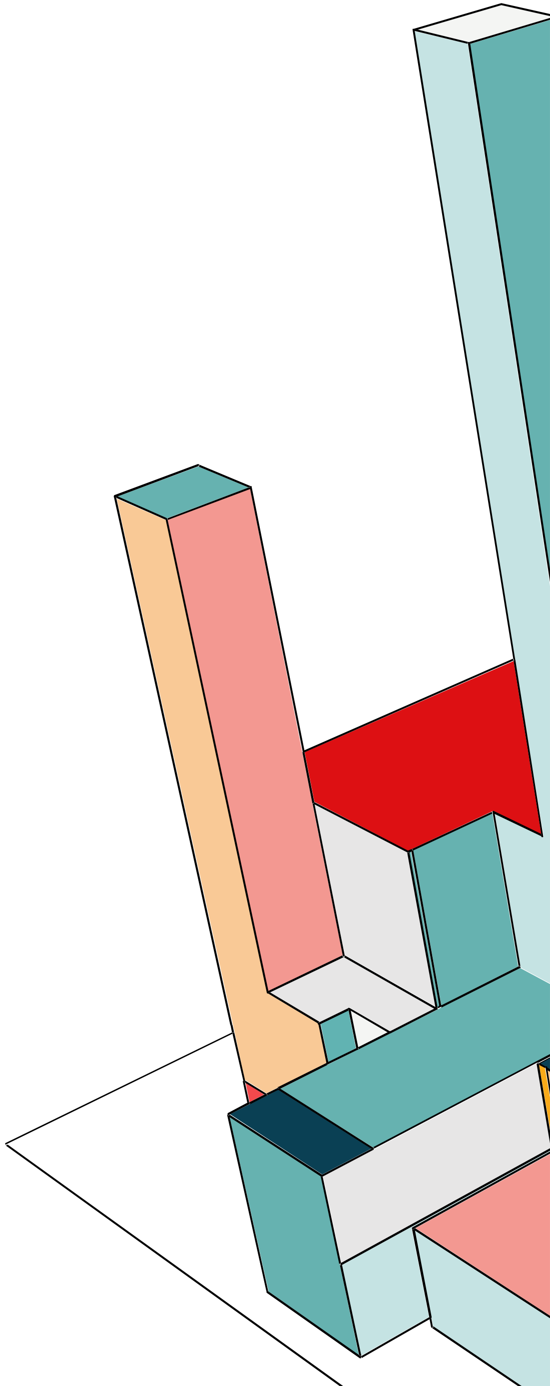
ENDS

USE GEN AI TO ADVANCE SUSTAINABLE DEVELOPMENT

Companies must also consider the **ends** for which Gen AI is deployed. The private sector can accelerate and amplify SDG action through applying Gen AI in sustainable development action areas.

STRATEGIES FOR DEVELOPING SUSTAINABLE AI POLICIES

- What guidance is available to guide the sustainable development and application of AI technologies?
- How can governments and global institutions confirm that AI systems contribute positively to the SDGs?
- EU AI Act **touches on sustainability considerations associated with the design, training, and use of AI.**
- The way in which the environmental sustainability of AI ought to be assessed and minimised **will be clarified by the AI Office** and **individual Member States in forthcoming codes of conduct and harmonized standards.**



Concluding thoughts

01

VALUE ALIGNMENT: IS IT FEASIBLE/DESIRABLE?

02

ETHICAL GOVERNANCE OF AI: FROM DESIGN TO RESEARCH ASSESSMENT AND AT WHICH LEVEL?

03

PUTTING PRINCIPLES/REQUIREMENTS INTO PRACTICE

04

ALIGNING COMPUTATIONAL PRACTICES/INFRASTRUCTURES PRACTICES TO THE REQUIREMENTS OF TRUSTWORTHY AI