



Green AI

sustainable AI solutions

April 2025

table of contents

Acknowledgements	3
Foreword	4
Introduction	5
Asking the Right Questions (ARQ) on sustainability	6
Asking the Right Questions (ARQ) on green AI	18
Green cloud computing and its role in sustainable AI	28
Efficient data management techniques for green AI	38
Green AI in Edge computing: Leveraging localised intelligence for sustainability	47

acknowledgements

Author

Ying Shaowei

Chief Scientist, NCS



Ying Shaowei coordinates the efforts of NCS' centres of innovation and centres of excellence, including leading one or more of these centres. He builds a global network of technology experts and research institutions which NCS leverages to accelerate client development and delivery work.

Shaowei champions selected innovation programmes or practices by providing executive sponsorship and directing investment funding. He also provides advice to the NCS corporate strategy on new potential market spaces and technology trends to invest and innovate in. Shaowei plays a key role in shaping the Voice of NCS on emerging technologies, amplifying long-term perspectives on important tech trends that impact businesses.

Contributors

TechCoE

Vijay Nayudu

AI Business Lead,
Tech CoE

Nikita Atkins

AI Business Lead,
Tech CoE

Dr. Sina Moradi

Senior Data Scientist,
Tech CoE

Samin Batra

Research Scientist,
Tech CoE

NEXT

Dr. Sunil Sivadas

Practice Lead,
NEXT Gen Tech

Cyber

Lim Hoon Wei

Principal Director,
Cyber Special Ops and R&D

NCS Sustainability

Jayne Leow

Sustainability Director,
NCS

Teck Chun Pang

Green Software Lead,
NCS

Editorial support

Georgia Swanson

Director of Strategy
& Growth, NCS NEXT
Global Innovation

Mark Addy

Insights Lead,
NCS NEXT,
Global Innovation

Deirdre Chong

Visual Designer,
NEXT Open Innovation

foreword

The benefits of information and communications technology (ICT) that have accrued to businesses, governments, and individuals over the past couple of decades are undeniable. The productivity, cost, and time-saving benefits of cloud computing over the past decade and artificial intelligence (AI) over the past two years have been particularly transformative. Yet, the power consumption required and the emissions that will be generated to further advance and scale these technologies raise cost versus benefit questions. An industry and society-wide consensus is forming that the current methods used for technology development and operation are unsustainable.

Recognising the need to reduce the carbon footprint of the technology that it uses and produces, NCS has taken steps to 'green' its processes and policies across multiple domains. We've put together several papers that review the challenges that must be overcome to achieve future sustainability, the steps we are taking at NCS, and the recommendations we have for organisations that share our desire to limit the environmental impact of the critical technologies that we all depend on. In this report, we share our technology sustainability expertise with you:

green AI

Asking the Right Questions (ARQ) on sustainability

Explains the importance of technology in addressing sustainability issues faced by the Asia-Pacific region and the roles of green AI, green software, and comprehensive sustainability frameworks for addressing the digital-to-carbon footprint. Covers tech-driven strategies that stakeholders can use to further reduce environmental impact and further net-zero objectives.

Asking the Right Questions (ARQ) on green AI

Defines green AI (the development and deployment of AI in ways that prioritise environmental sustainability and energy efficiency, and reduce carbon emissions) and explores strategies for its achievement including 'greening' systems, data, and intelligence. Highlights potential barriers to the widespread adoption of green AI, its immense potential for enabling responsible innovation, and the crucial role it can play in advancing sustainability.

Green cloud computing and its role in sustainable AI ('Greening' systems)

Explains the central importance of green cloud computing and green data centres to the achievement of sustainability goals for the information technology industry. Explores eco-friendly resource usage and the need for redesign of modern networking and discusses the current state of, and future direction required for, sustainable cloud computing and AI technologies.

Efficient data management techniques for green AI ('Greening' data)

Examines innovations in data processing and management in the context of green AI, including data compression, deduplication, and storage optimisation. Outlines the benefits of adopting sustainable data management practices, task scheduling, and resource allocation to achieve energy efficiency, as well as the potential consequences of the over-regulation of AI and data management.

Green AI in Edge computing: Leveraging localised intelligence for sustainability ('Greening' intelligence)

Presents the current state of development for Edge AI and Edge computing technologies and the benefits of local processing in achieving green AI. Presents the challenges inherent in Edge AI, including power consumption, battery life constraints, and the performance versus accuracy trade off, and suggests resolutions that will allow Edge computing to play a key role in green AI adoption.

With the growing awareness of global warming and climate change, the need for more sustainable technology is urgent and apparent. We hope that you find the information contained in this report useful and would welcome the opportunity to explain how NCS can partner with your organisation to help you achieve your own sustainability goals through more sustainable and greener technologies and practices.

introduction

"Digital for Sustainability" encapsulates the powerful convergence of digital technologies and sustainable practices that serves as an imperative catalyst for driving progress across the environmental, social, and governance (ESG) domains. This concept embraces a dualistic perspective: one that focuses on the sustainable evolution of our digital technologies, and another that acknowledges the transformative role of these very technologies in driving sustainability.

On one front, we concentrate our efforts on fostering the sustainability of our digital technologies. This entails evolving our data centres, algorithms, and AI models towards a greener paradigm to reduce emissions, enhance energy efficiency, and curb e-waste. Our vision is a truly green digital world where progress and innovation do not come at the expense of our planet's health.

In parallel, we underscore the indispensable role of these advanced digital technologies in empowering organisations to successfully meet their ESG targets. The capabilities that these digital tools bring, for facilitating precise Scope 3 emissions reporting, bolstering ESG regulations, enabling efficient waste management, and orchestrating smart grids are nothing short of revolutionary.

In an era where sustainability has ascended from a peripheral concern to a core business priority, "Digital for Sustainability" is not just a choice – it's an imperative. Digital technology is pivotal in this context as it offers the potential to transform traditional sectors and deliver substantial environmental benefits. Moreover, aligning digital innovation with sustainable practices allows us to reduce our digital carbon footprint and ensure that our progress doesn't compromise the health of our planet. The digital tools we craft today are the levers that can, and will, drive sustainable transformation on an unprecedented scale.

The NCS TechCoE is actively pioneering the confluence of digital technology and sustainability. We strive to cultivate green data centres, design sustainable algorithms, measure the sustainability of AI models, and eliminate e-waste. However, our vision extends beyond just improving our internal operations.

We are working diligently to leverage digital technologies as powerful tools for achieving broader sustainability goals, including improving ESG reporting, waste management, and smart grids. Our efforts are also focused on solving pressing sustainability issues and advancing towards the United Nations Sustainable Development Goals for water security, food management, and community engagement.

But we can't do it alone. This is a call to action for all NCS employees to get involved. We need your skills, your ideas, and your passion. Together, we can deliver innovative, sustainable solutions and accelerators that transform our business and significantly impact our planet and society. We invite you to join us on this exciting journey. Your collaboration will be vital in pushing the boundaries of what's possible in the Digital for Sustainability space and ultimately help in shaping a more sustainable future for all.



A photograph of three business professionals (two men and one woman) standing in a modern building with a large glass wall and a lush green living wall. The man on the left is wearing a light blue shirt and glasses, the woman in the middle is wearing a light blue shirt, and the man on the right is wearing a dark suit and glasses. They are all looking towards the right. The building has a dark frame with warm yellow lights. The living wall is covered in various green plants and moss.

Asking the Right Questions (ARQ) on sustainability

Nikita Atkins, Sina Moradi

Asking the Right Questions (ARQ) on sustainability

Executive summary

The article highlights the Asia-Pacific (APAC) region's progress in harnessing technology for sustainability while identifying major challenges faced by the region such as scope 3 emissions and managing the digital-to-carbon footprint. It explores the roles of green AI, green software, and comprehensive frameworks in guiding sustainability efforts by businesses, governments, and individuals. Furthermore, the article delves into tech-driven strategies such as enhancing supply chain transparency using blockchain and IoT, leveraging digital platforms for stakeholder engagement, and investing in energy-efficient data centres. By implementing these strategies and fostering collaborative action, stakeholders in the APAC region can work collectively towards reducing their environmental impact, enhancing efficiency, building resilience, and achieving their net-zero goals.

How has the APAC region made progress in harnessing technology for sustainability?

The Asia-Pacific (APAC) region has made significant strides towards sustainability in recent years, as the following shows:

- PwC's Net Zero Economy Index 2022 revealed that APAC is leading the world in the effort to decouple greenhouse gas emissions from economic growth, while also facing more strenuous headwinds. Encouragingly, the Asia-Pacific region achieved a decarbonisation rate of 1.2% on average in 2021. This indicates that the region's energy-related CO2 emissions per unit of gross domestic product generated are falling. In the same period, the world's overall decarbonisation rate was just 0.5%¹.
- Environmental, Social, and Governance (ESG) practices are becoming standard for public and private organisations, with 50% of APAC companies expected to have a sustainability strategy in place by 2024². This is particularly significant for APAC companies as the APAC region's share of global emissions increased from 35% in 2010 to 39% in 2019, which can be attributed to a few factors, including the region's role as a global manufacturing hub, rising income levels, and the rapid growth of Asia Pacific cities³.
- Consumers increasingly prioritise social responsibility and trustworthiness, which can drive loyalty to companies that support environmental and social issues⁴⁻⁵. There is an urgent need to transition towards a more sustainable future, and adopting digital technologies is a critical part of that shift.



Australia and Singapore are two advanced economies in the Asia-Pacific region that have demonstrated strong commitments to advancing environmental sustainability. Both countries have made significant progress in implementing multiple environmental initiatives by leveraging sectoral-led digital transformations. These transformations are leading to greater energy utilisation efficiencies, lower consumption, and more recycling.

The following table provides a comparison of the key environmental sustainability efforts in Australia and Singapore across different areas:

Area	Australia	Singapore
National sustainability targets	Net-zero emissions by 2050, AUD 1.6 billion Technology Investment Roadmap Initiative, National Circular Economy Hub	Singapore Green Plan 2030 (Energy, Waste, Water, and other sectors)
Sustainable energy generation	Renewable energy generation, Virtual Power Plants (VPPs) and Distributed Energy Resources (DERs) (South Australia's VPP ⁶)	Solar energy deployment (SolarNova Programme), green data centres (National Supercomputing Centre)
Carbon emissions reduction	Upgrading the electricity grid to support more renewable power, Technology Investment Roadmap Initiative (Hydrogen Valley Project & CCUS Projects)	Carbon Pricing Bill (Emission Reduction Targets), hydrogen capabilities (Sembcorp's Hydrogen Research), carbon capture & storage solutions (NTU's Delta-NTU Corporate Lab)
Smart grids and energy efficiency	Advanced Metering Infrastructure (AMI rollout by energy distributors), Smart Grid Technologies (Digital transformation of grid operations)	Advanced Metering Infrastructure (Nationwide deployment by 2023), Smart Grid Technologies (SP Group's Intelligent Energy System)
Healthcare	Royal Flying Doctor Service (Telemedicine for remote communities)	Telehealth (Telemedicine solutions for remote patient monitoring), robotic-assisted surgery (Da Vinci Surgical System at KK Women's and Children's Hospital)
Transport	Intelligent transport systems (Next generation traffic management systems), sustainable urban mobility plans (City of Sydney's Movement for People Strategy)	ERP 2.0 (Satellite-based road pricing system), greener transport options (Electric vehicles and charging Infrastructure, cycling and walking networks ⁷)
Green buildings	National Construction Code (Minimum energy efficiency standards), Green Star Certification (Barangaroo South in Sydney)	Building and Construction Authority (BCA) Green Mark Scheme (Marina Bay Sands), Zero Energy Building (ZEB) at BCA Academy
Recycling initiatives	Container Deposit Schemes (Return and Earn in New South Wales), National Recycling and Waste Reduction Act	National Recycling Programme, Cash-for-Trash (Old newspapers and magazines recycling)
e-waste management	National Television and Computer Recycling Scheme (TechCollect)	National Voluntary Partnership for e-waste recycling, Producer Responsibility Scheme (PRS) on e-waste

Table 1. Comparative overview of environmental sustainability initiatives and digital transformations in Australia and Singapore

Australia and Singapore have been successful in incorporating digital transformations across multiple sectors to improve energy efficiency, reduce waste, and promote sustainable practices. From sustainable energy generation to biodiversity conservation, both countries have shown progress in implementing innovative solutions to address environmental challenges. By focusing on sectoral-led digital transformations, they are paving the way towards a more sustainable and eco-friendly future.

What are the major sustainability challenges faced by the APAC region?

While APAC has made material progress in addressing climate-related risks, there are still challenges to overcome. Much more needs to be done on the sustainability front, as APAC's rapid urbanisation and population growth continues relentlessly. The recent disturbing increase in the frequency of natural disaster emergencies, exemplified by the bush fires across Queensland in 2019-2020 ("Black Summer"), and the floodings in east Australia in 2022, underlines the urgent and growing need for sustainable practices and infrastructure to be put in place. Scope 3 emissions and the digital-to-carbon footprint are the two critical areas where additional focus and effort are required by both governments and industry to improve overall sustainability in the APAC region:



Figure 1. Major sustainability challenges in the APAC region

Measuring and curbing Scope 3 emissions

APAC faces a significant challenge regarding the lack of transparency around Scope 3 emissions. These emissions, also known as indirect emissions, occur throughout a company's value chain and include those from purchased goods and services, transportation, and waste disposal. It is crucial to address these emissions as they can make up the majority of a company's carbon footprint, which ultimately contributes to climate change. Unfortunately, many companies in APAC do not report or address their Scope 3 emissions, making it challenging to accurately measure and ultimately reduce the region's carbon footprint.

According to a report by CDP, a non-profit organisation that works with companies to disclose and manage their environmental impact, only 36% of companies in APAC currently disclose their Scope 3 emissions. In contrast, 56% of companies globally disclose their Scope 3 emissions, highlighting the region's need for increased transparency and action⁸. While it is positive to see that a large number of the almost 4,000 APAC companies that responded to CDP in 2021 were reporting greenhouse gas (GHG) emissions, the proportion of companies reporting at least one of their Scope 3 emission categories is unfortunately low, at just 4 out of 10. This indicates that most companies are not yet on the best practice climate journey of accounting for their full value chain footprint⁸. The lack of transparency around Scope 3 emissions in APAC can be attributed to several factors, including limited regulatory requirements, and a lack of awareness and understanding of the issue. Many companies may also face difficulties in tracking and measuring their indirect emissions due to the complexity of their value chains and the involvement of multiple suppliers and partners.

However, the importance of addressing Scope 3 emissions cannot be overstated. As the global focus on sustainability intensifies, companies that do not address their indirect emissions may face reputational risks and struggle to compete in a market where sustainability is becoming increasingly important to consumers and investors. It is essential for businesses in APAC to prioritise transparency and accountability around their Scope 3 emissions. This will require setting clear targets and implementing effective measurement and reporting systems. Governments and industry associations also have a role to play in incentivising organisations, regulating sustainable practices, promoting education, and increasing awareness around the importance of addressing Scope 3 emissions.

Managing the digital-to-carbon footprint

While digital transformation presents opportunities for addressing sustainability challenges, it could also lead to higher energy consumption and increased e-waste. In addition, the advent of generative AI threatens to drive energy consumption much higher.

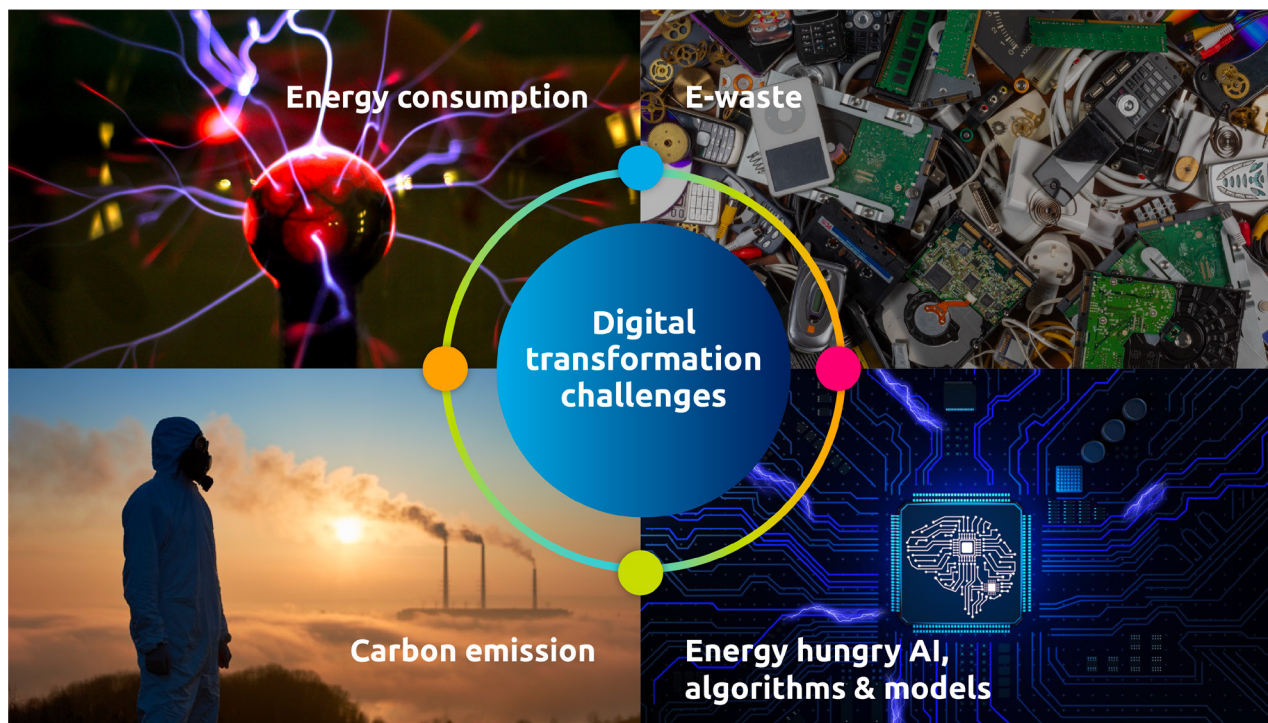


Figure 2. Digital transformation challenges

- **Energy consumption**

Digital transformation brings challenges, including increased energy consumption and carbon emissions from cloud computing and data centres. Despite the numerous benefits, storing data in the cloud has a high energy cost (7 kWh per gigabyte compared with 0.000005 kWh for personal hard disk storage)⁹. According to recent reports, the operational capacity of data centres in APAC has significantly expanded, with a total regional capacity that reached 10.6 GW at the end of 2023 and expected to grow to 26.7 GW by 2028¹⁰. This growth is largely driven by the increasing demand for digital services and the adoption of cloud computing, AI, and other technologies. In Singapore, the overall data centre supply capacity is expected to grow at a CAGR (compound annual growth rate) of 7.2% from 2021 to 2026 (Figure. 3), leading to a moratorium on new data centres as green technologies are being developed¹¹.

Data centre capacity (MW)

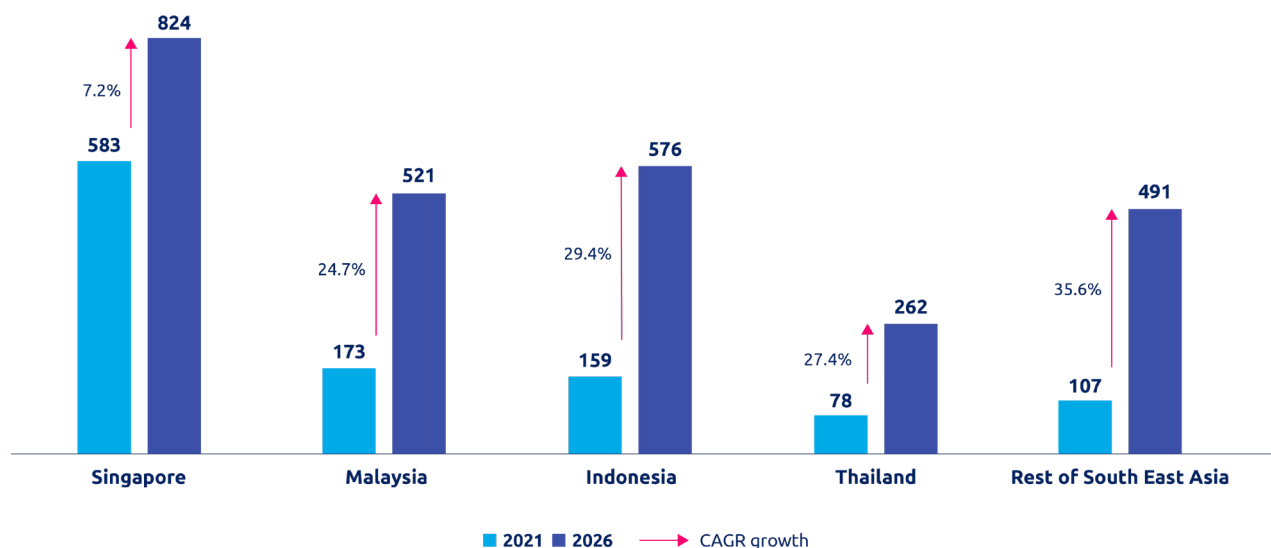


Figure. 3. Overall DC supply capacity projected to grow at CAGR ~19% in Southeast Asia¹¹

The growing volume of data and the need for faster processing speeds will contribute to a rise in energy consumption and carbon emissions.

- **e-waste**

Digital transformation has led to a surge in e-waste, a major environmental concern. The rapid production and replacement of electronic devices results in increased waste disposal and data security challenges. Proper e-waste management is vital, especially in the APAC region, where countries like Greater China contribute significantly to global e-waste. As digital device usage grows, and device lifespans shorten, addressing e-waste becomes increasingly crucial.

- **Energy hungry AI, algorithms and models**

The increasing reliance on AI models raises concerns about their environmental impact, especially given the energy consumption and carbon emissions associated with training Large Language Models (LLMs). The Stanford AI Index reported that the training of the GPT-3 model required approximately 1,300 MWh of power, resulting in 502 tons of CO₂ emissions. That equates to the lifetime emissions of 8 cars, or the yearly emissions of 109 cars, and the energy required to power a US home for more than 120 years¹². The solution is not to discontinue ChatGPT, which serves millions of users, but for organisations to adopt a holistic approach to understand the costs, resources, and benefits of their AI models.

Tech-driven strategies that can be used to tackle sustainability challenges and progress towards net-zero goals

Governments, businesses, and individuals must work together to bring about a combination of policy changes, innovative technological implementations, and individual actions. While technology alone is not going to be sufficient in reducing climate-related risks, it can play an outsized contributory role. The fact that many parts of our economies are undergoing digital transformation means that information and communications technology (ICT) does play a crucial role in reducing energy consumption and addressing environmental issues. The use of ICT and ICT-enabled products and services minimises the negative impact and maximises the positive impact of human behaviour on the environment.

At the same time, ICT itself consumes a huge amount of energy, creating a value paradox. Today, the ICT sector contributes to about 2% of national emissions in Singapore, and this is expected to grow to 8% by 2030. Increasing demand for energy-intensive technologies, such as AI and blockchain, could increase ICT industry emissions to 33% of Singapore's total GHG emissions by 2050¹³. Hence, it is critical that 'Green by Digital' efforts be closely aligned with or offset by 'Green for Digital' initiatives.

In the following sections, we discuss how businesses can harness emerging technologies to advance their Net Zero aspirations, asking what (more) they could do with ICT to:

- Increase the transparency and traceability of supply chains; and
- Enhance energy management and sustainability; and
- Improve communications with stakeholders to encourage more environmentally conscious choices; and
- Leverage green software/AI and green data centres

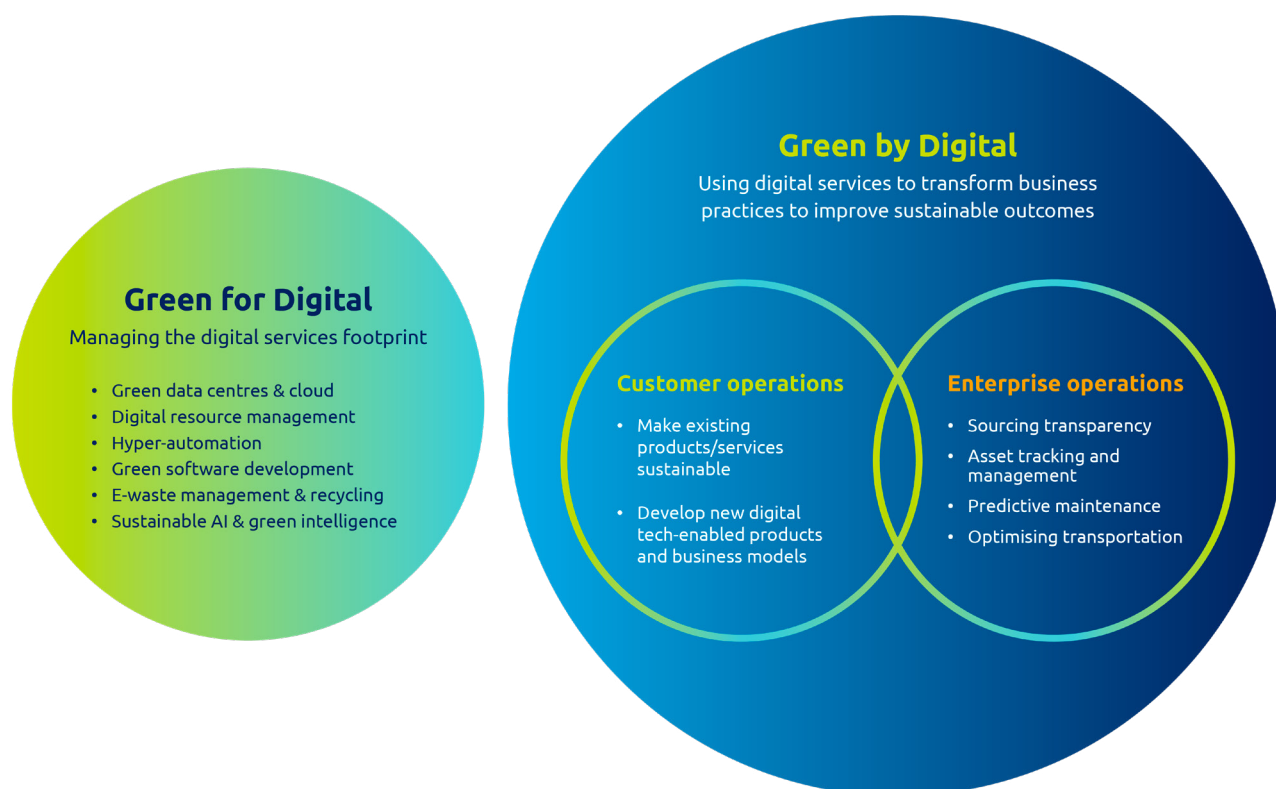


Figure 4. Digital transformation through Green by Digital and Green for Digital approaches

Increasing transparency and traceability of supply chains

Increasing the transparency and traceability of supply chain management is vital for organisations and governments that aim to achieve ESG goals and is particularly important for Scope 3 emission tracking and reporting. Integrating digital technologies such as AI-optimised supply chain management, ML algorithms, and predictive analytics will contribute to waste reduction, optimised inventory management, and sustainable transportation. Emerging technologies, such as blockchain and IoT, can be used to further advance Net Zero goals by creating transparency and traceability, and the ability to audit sustainable practices throughout a supplier's sourcing to manufacturing process. Blockchain enhances supply chain authenticity, ethical standards compliance, and streamlines cross-border transactions tracing. Meanwhile, IoT enables real-time tracking and the ability to optimise transportation routes to reduce emissions. Together, these technologies can enhance environmental sustainability and make smart warehousing and efficient asset management possible.

Enhancing energy management and sustainability

Addressing the issue of increasing energy consumption is crucial for developing innovative solutions to manage energy resources effectively. Implementing smart technologies like IoT and AI can provide transparency and traceability in energy usage. Existing solutions, such as the consumer smart app from Singapore's SP Group (SP App), already incorporate some of these features, and have community engagement as a focus. To further enhance the impact of such solutions, similar apps might include gamifying the experience to foster a stronger sense of community responsibility. Additionally, incorporating an educational component that highlights efficient energy usage best practices can empower users to make more informed decisions about their energy consumption. By building on the foundation of existing smart apps, and integrating advanced IoT and AI technologies, businesses can substantially improve transparency and traceability for better energy management and sustainability. These emerging technologies offer powerful tools for organisations to advance their Net Zero aspirations and contribute to a more environmentally responsible global economy.

Improving communications with stakeholders

Partnership and collaboration among businesses, governments, academia, and civil society are vital for addressing complex sustainability challenges and achieving ESG goals. Digital technologies can facilitate communication, data sharing, and coordination among stakeholders, enhancing the efficiency and effectiveness of sustainability initiatives. Industry-wide sustainability standards can level the playing field and drive progress towards ESG goals.

One of the ways businesses can harness ICT to improve communications with stakeholders and encourage more environmentally conscious choices is by utilising apps and digital platforms. The Breeze app, developed by NCS, promotes green travel by helping users make informed decisions about their transportation choices. The Breeze app not only suggests greener options for travelling from point A to point B, but also calculates the carbon emissions associated with each journey. By providing information on sustainable transport alternatives, as well as comparing and ranking travel options based on their carbon footprints, the app effectively engages stakeholders and influences their decisions towards more sustainable actions. This empowers individuals to actively contribute to reducing their environmental impact and promoting greener travel choices.

Businesses can further improve communications with stakeholders by:

1. Engaging through social media and online forums
2. Implementing AI-driven analytics to tailor sustainability communications
3. Using virtual reality (VR) and augmented reality (AR) to create immersive experiences that foster connections to sustainability goals
4. Developing webinars, podcasts, and e-learning resources for education and thought leadership on ESG topics
5. Establishing open data platforms for real-time access to ESG performance data

By harnessing emerging technologies and ICT, businesses can advance their Net Zero aspirations and foster greater sustainability within industries.

The adoption of green software is equally important for advancing Net Zero aspirations. Green software development methodologies ensure that the software products created and used by businesses have been designed to use resources optimally, consume less energy, and reduce overall environmental impact. By embracing these development and operational practices, businesses can create innovative, high-performing, and environmentally friendly software solutions that advance Net Zero objectives.

In addition, green data centres can significantly reduce e-waste through their use of equipment designed to be more durable and recyclable. This is a crucial aspect of environmental sustainability, as electronic waste is a growing problem worldwide. By prioritising sustainable materials and practices, these centres are leading the way in creating a more sustainable digital economy. Therefore, the importance of green data centres extends beyond energy efficiency. They are at the core of cloud computing and a vital component in the broader push towards environmental sustainability in the digital age.

NCS has been recognised as a leader in cloud professional services (CPS) in the Asia Pacific region according to the IDC MarketScope assessment for 2023-2024 (Figure. 5)¹⁴. This survey revealed that 67% of enterprises in the region now have significant workload deployments on one or more public clouds, underscoring the growing importance of the cloud in enterprise IT.



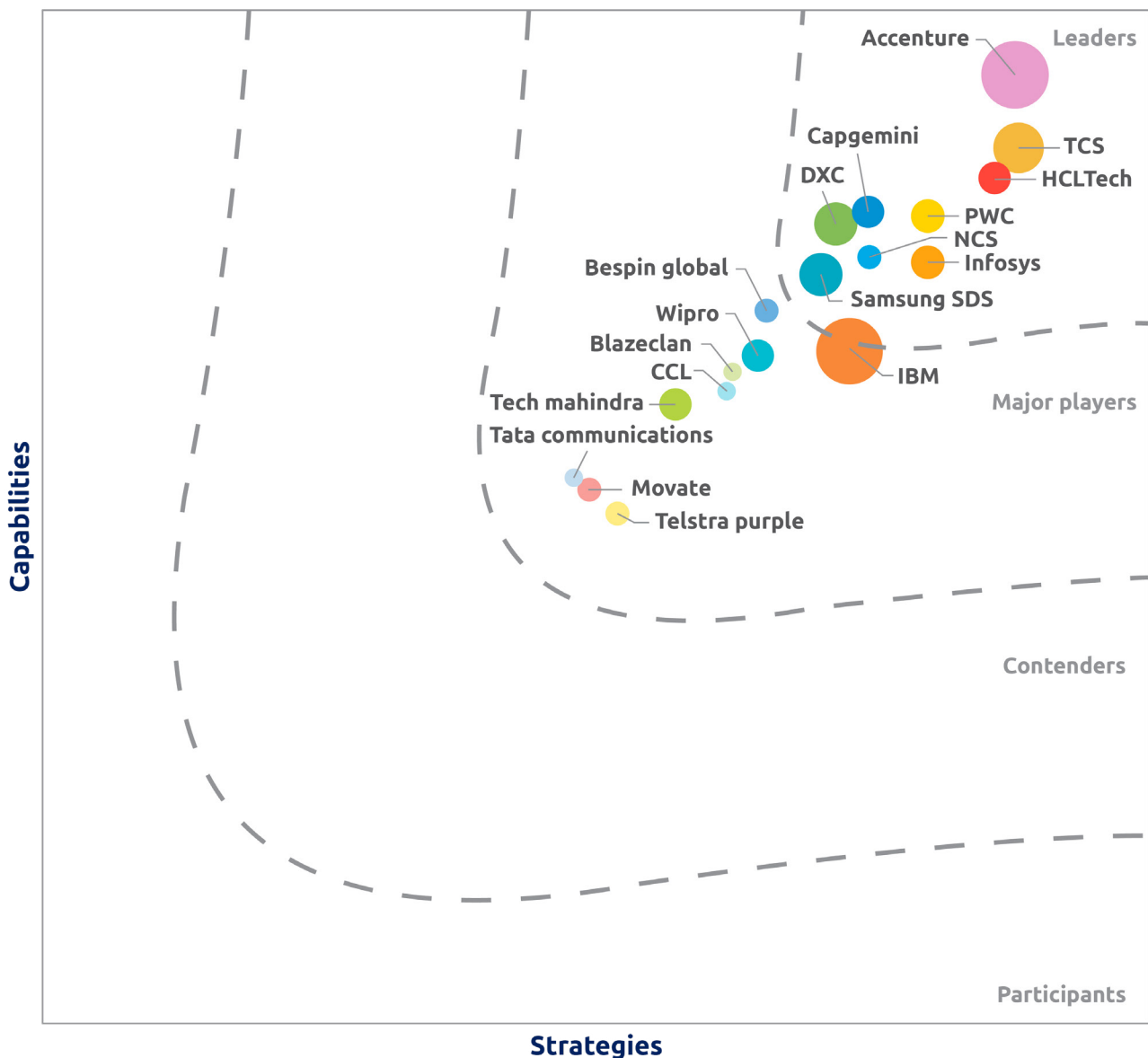


Figure. 5. IDC MarketScape Asia/Pacific cloud professional services 2023-2024¹⁴

One prime example of harnessing emerging technologies to achieve sustainability goals is NCS's Bedok Data Centre. The facility leverages green AI algorithms to monitor and analyse real-time information from various sources on energy usage, cooling systems, and carbon emissions. These insights allow for the development of strategies for reducing energy consumption, improving power usage effectiveness (PUE), and minimising the data centre's environmental footprint. Green software solutions in use at the Bedok Data Centre ensure that processes and operations are optimised, reducing resource wastage, and contributing significantly to the facility's sustainability. The integration of green AI and green software into advanced AI and generative AI technologies is vital for businesses looking to advance their Net Zero aspirations. By adopting these environmentally conscious approaches and embracing a culture of continuous improvement and innovation, NCS's Bedok Data Centre team has made significant strides towards reducing the centre's environmental impact and achieving a more sustainable future. The Bedok Data Centre has become more environmentally conscious, without sacrificing performance or reliability, and has become a leader in energy-efficient and sustainable operations. The facility boasts state-of-the-art energy management systems, advanced cooling technologies, and renewable energy sources, ensuring its minimal environmental impact. Through continuous monitoring, optimisation, and simulation, these technologies enable the Bedok Data Centre operators to minimise resource consumption and waste, contributing significantly to the achievement of sustainability objectives. The Bedok Data Centre serves as a powerful example of how organisations can leverage technology to drive positive change and contribute to the global fight against climate change.

Summary and conclusion

As organisations confront climate change challenges, such as the excessive energy demands of AI models and algorithms and the increase in e-waste, they are increasingly turning to digital services for effective and practical solutions. Strategies encompassing "Green for Digital", and "Green by Digital" have emerged as transformative approaches, enabling organisations to reduce their environmental impact and enhance operational efficiency. These strategies involve sustainable practices like green data centres, e-waste management, hyper-automation, digital resource management, and the use of green AI, green software, green algorithms, and sustainable AI models.

Digital technologies can foster environmental sustainability across industries through digital innovation and strategic partnerships. Organisations can optimise resources, reduce waste, and improve efficiencies by leveraging green AI algorithms, green software, and hyper-automation. Integrating circular economy strategies and effective supply chain management enables a transition towards sustainable and resilient systems.

Beyond environmental impacts, the benefits of sustainable digital practices include financial advantages, such as better operational performance, lower risks, and increased attractiveness to investors. These additional benefits highlight the economic viability of sustainable practices as they can lead to cost savings, increased revenue, and higher return on investment. Moreover, sustainability is essential for both the planet's health and the long-term success of organisations. Public servants, government officials, and private organisation executives should collaborate to share best practices and leverage the latest green technologies to promote sustainable digital practices and ensure a more sustainable future for all. Sustainable business practices can create hundreds of millions of new jobs, increase revenue, save energy, and reduce costs. As sustainability resonates with consumers, adopting sustainable digital practices, including green AI and green software, presents an opportunity to strengthen brand reputation and cater to evolving consumer preferences.

In summary, organisations must embrace sustainable digital practices to achieve environmental, financial, and reputational success. As the world focuses on sustainability, organisations need to integrate "Green for Digital" and "Green by Digital" approaches and solutions into organisational strategy and practices, which are not just beneficial, but essential for long-term prosperity and resilience.



References

1. PwC. (2022). Net Zero Economy Index: Asia-Pacific's transition. Retrieved from <https://www.pwc.com/gx/en/issues/esg/esg-asia-pacific/net-zero-economy-index-asia-pacifics-transition.html>
2. Marshall, C., & Espejo, M. IDC FutureScape: Worldwide Sustainability/ESG 2023 Predictions — Asia/Pacific (Excluding Japan) Implications. International Data Corporation. Retrieved from <https://www.idc.com/getdoc.jsp?containerId=AP50052623>
3. Lit Ping Low, Now or never to rein in emissions: Highlights from IPCC third report and implications in Asia Pacific. Retrieved from <https://www.pwc.com/gx/en/about/pwc-asia-pacific/ipcc-third-report-perspectives-on-asia-pacific.html>
4. Cone Communications. (2017). 2017 Cone Communications CSR Study. Retrieved from <https://www.conecomm.com/research-blog/2017-csr-study>
5. Deloitte. (2020). Deloitte Global Millennial Survey 2020. Retrieved from <https://www2.deloitte.com/gr/en/pages/aboutdeloitte/articles/MillennialSurvey2020.html>
6. Government of South Australia. (n.d.). South Australia's Virtual Power Plant. Department for Energy and Mining. Retrieved from <https://www.energymining.sa.gov.au/consumers/solar-and-batteries/south-australias-virtual-power-plant>
7. National Climate Change Secretariat. (2021, November 9). Joint Media Release by NCCS, MSE, MTI, MOT, MND and MOE: Good progress made on the Singapore Green Plan 2030 as Government accelerates decarbonisation and sustainability efforts. Retrieved from <https://www.nccs.gov.sg/media/press-release/joint-media-release-by-nccs-mse-mti-mot-mnd-and-moe-good-progress-made-on-the-singapore-green-plan-2030-as-government-accelerates-decarbonisation-and-sustainability-efforts/>
8. CDP. (2022). How Companies in Asia Pacific are Preparing for the Net-Zero Economy. Retrieved from https://cdn.cdp.net/cdp-production/cms/reports/documents/000/006/179/original/How_companies_in_Asia_Pacific_are_preparing_for_the_net-zero_economy_EN.pdf?1648712180
9. Adamson, J. (2017). Carbon and the cloud. Retrieved from <https://medium.com/stanford-magazine/carbon-and-the-cloud-d6f481b79dfe>.
10. APAC markets experienced largest data centre growth between 2018-2023 - DC Byte - DCD. Retrieved from <https://www.datacenterdynamics.com/en/news/apac-markets-experienced-largest-data-center-growth-between-2018-2023-dc-byte/>
11. Bill Chang, CEO of Digital Infrastructure Co, Singtel Investor Day (24 August 2023). Retrieved from <https://www.singtel.com/content/dam/singtel/investorRelations/factsheetsAndPresentation/2023/Singtel%20Investor%20Day%202023%20-%20Digital%20InfraCo.pdf>
12. Ghosh, S. (2022, January 4). Report on ChatGPT Model's Emissions Offers Rare Glimpse of AI's Climate Impacts. Truthout. Retrieved from <https://truthout.org/articles/report-on-chatgpt-models-emissions-offers-rare-glimpse-of-ais-climate-impacts/>
13. Energising the ICT sector - meet Dr. Ong Chen Hui of Singapore's IMDA, September 2023. Retrieved from <https://greensoftware.foundation/articles/energising-the-ict-sector-meet-dr-ong-chen-hui-of-singapore-s-imda>
14. Cloud Professional Services "Leaders" in 2023 for Asia/Pacific Revealed by IDC, 30 November 2023. Retrieved from <https://www.idc.com/getdoc.jsp?containerId=prAP51432123>



Asking the Right Questions (ARQ) on green AI

Sina Moradi

Asking the Right Questions (ARQ) on green AI

Executive summary

This article explores the potential for making the growing adoption of AI sustainable through the implementation of green AI's three key strategies. We will discuss the concept of green AI, with a focus on optimising energy consumption, emissions, and resource usage. The article delves into strategies such as greening systems, greening data, and greening intelligence while highlighting barriers that could slow the widespread adoption of green AI. Additionally, we will examine green AI's immense potential for enabling responsible innovation across industries and the crucial role it can play in advancing sustainability amid the far-reaching implications of AI technology.

Introduction

The second digital and fourth industrial revolutions have driven an innovation culture that has fuelled numerous technological developments and breakthroughs¹⁻². Among these advancements, the field of artificial intelligence (AI) has experienced remarkable exponential growth over the last 8 to 10 years³⁻⁴. Today, AI is an in-trend, disruptive technology with countless applications and even more prospects for all industry sectors and areas of life, ranging from health to agriculture, engineering to finance, gaming to transportation, and beyond⁵.

Due to rapid digitalisation and the rising use of technologies like AI, cloud computing, and big data, the need for Information and Communications Technology (ICT) infrastructure and computational resources is rapidly expanding. Current estimates show that the ICT sector contributes to approximately 4% of global CO₂ emissions⁶. By 2040, and projections indicate that the ICT carbon footprint could reach 14%, with data centres contributing to almost half of this growth⁶.

The increasing reliance on AI technologies has resulted in traditional AI models consuming large amounts of energy and, consequently, contributing to increased carbon emissions that exacerbate climate change. To address this issue, the Infocomm Media Development Authority (IMDA) and the Government Technology Agency (GovTech) of Singapore recently unveiled a S\$30m Green Computing Funding Initiative (GCFI) that aims to promote sustainable practices and the efficient use of computing resources in the tech sector⁷. Considering the pressing environmental challenges, the concept of green AI has emerged as a compelling response to balance the need for AI-driven progress with ecological sustainability.

Green AI looks to reshape the development and deployment of AI technologies by focusing on energy efficiency, reduced carbon emissions, and the responsible use of computational resources. By adhering to these principles, organisations that adopt green AI aim to minimise the environmental impact of AI technologies while maintaining their ability to drive progress across various domains.

In this article, we will explore the relevance of green AI in today's world and examine its potential applications, benefits, and challenges. We will discuss how green AI offers effective resolutions to the critical issues faced by traditional AI models and propose ways in which stakeholders can collaborate to promote a greener and more sustainable AI-driven future. This exploration will highlight the importance of embracing green AI to ensure that the remarkable advancements in AI technology are not only transforming industries and sectors, but also contributing to a more environmentally friendly future.

How has the growth of AI models affected the environment and electricity demand?

Advances in computation, including those in hardware, software, and algorithms, have enabled scientific research to progress at unprecedented rates. Machine learning permeates many aspects of society, including economic and social interactions⁸⁻¹¹. In 2020 alone, almost 9 billion compute hours have been used for scientific computing¹², a pace of 24 million hours per day. Yet, the costs associated with large-scale computation are not being fully captured.

Machine learning (ML) models have grown exponentially in size over the past few years¹³, with some algorithms training for thousands of core-hours, and the energy consumption and costs required for ML training have become growing concerns¹⁴. Traditional AI models, such as deep learning algorithms, rely heavily on complex neural networks that require vast amounts of data for training and learning. These models demand substantial computational power, which in turn consumes large amounts of energy¹⁵. In natural language processing (NLP), Strubell et al.¹⁶ found that designing and training translation engines can emit between 0.6 and 280 tonnes of CO₂. While not all NLP algorithms require frequent retraining, algorithms in some fields are retrained daily or weekly, multiplying their energy consumption. As these models grow in both size and complexity, the demand for servers to process the models will grow exponentially (Figure 1)¹⁷.

Days spent calculating at one petaflop per second*, log scale

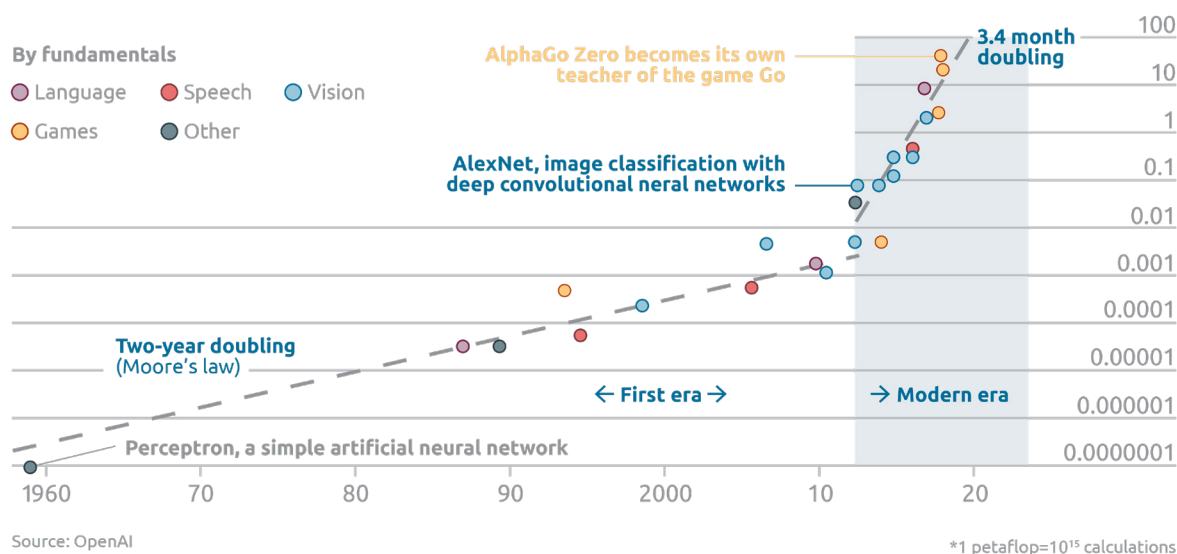


Figure 1. Computing power used in training AI systems¹⁷.

The compute power used for training AI systems has exponentially increased in the era of deep learning¹⁷. The exact total energy required for a single AI model is difficult to estimate as it must include the energy used to manufacture the computing equipment, create the model, and use the model in production. Researchers estimated that developing a generative AI model named BERT, which consisted of 110 million parameters, “consumed energy equivalent to that of a round-trip transcontinental flight for a single individual”¹⁸. The training of GPT-3, which has 175 billion parameters, required 1,287 megawatts hours of electricity, which equates to 552 tons of carbon dioxide, equivalent of 123 gasoline-powered passenger vehicles driven for one year¹⁹. That energy was just for getting the model ready to launch before any consumers started using it. The future development of larger models will substantially increase energy consumption and has raised concerns about the environmental impact of AI advancements.

To estimate the carbon footprint of nearly any computational task in a standardised and reliable way, Green Algorithms (www.green-algorithms.org; Figure 2), a generalisable framework and a freely available online tool, has been developed²⁰. This tool allows users to evaluate their computations or estimate the carbon savings or costs of redeploying them on other architectures.



Figure 2. The Green Algorithms calculator (www.green-algorithms.org).

The growing demand for AI-based applications, and the exponential growth in the use of computational resources due to AI advancements, has resulted in data centres accounting for an increasingly significant portion of global electricity consumption. As these data centres consume power, they generate heat which leads to additional energy consumption for cooling systems to prevent overheating. This vicious cycle contributes to an escalating energy demand that poses considerable challenges to both the environment and the power grid infrastructure. Widely cited forecasts suggest that the total electricity demand by the ICT industry will accelerate throughout the 2020s and that the percentage consumed by data centres will increase (Figure 3) ²¹.

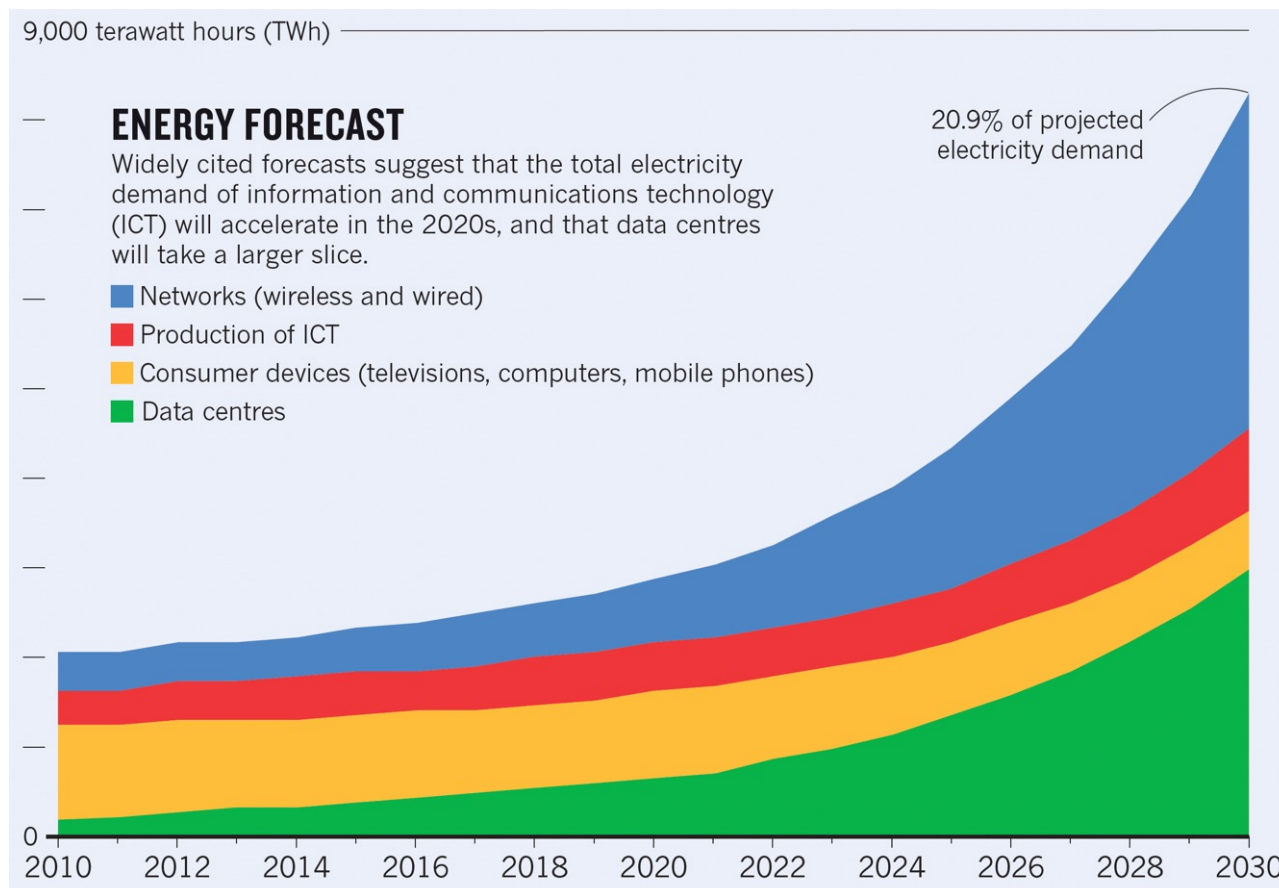


Figure 3. The forecast of electricity demand of ICT.

Figure 3 is an “expected case” projection by Anders Andrae, a specialist in sustainability ICT. Under his ‘best case’ scenario, ICT energy demand grows to just 8% of total electricity demand by 2039, rather than to 21% ²¹.

What environment challenges arise from AI and how can green AI help to mitigate those challenges?

The excessive energy consumption of traditional AI models leads to a series of environmental challenges, with carbon emissions being the most prominent. As most AI-powered data centres rely on energy generated by burning fossil fuels, they contribute to a higher carbon footprint and exacerbate global warming and climate change.

Moreover, the environmental impact of AI is not limited to energy consumption alone. The production and disposal of the hardware needed for AI models, including GPUs, CPUs, and other electronic components, generates significant electronic waste (e-waste). The improper handling of e-waste can result in the release of hazardous materials into the environment, leading to soil, water, and air pollution, which in turn can have severe consequences on human health and ecosystems.

These environmental challenges posed by AI necessitate the development of more sustainable AI technologies. Green AI aims to address these concerns by minimising the environmental impact of AI models while maintaining their potential to drive progress and innovation in various fields. By exploring and adopting green AI principles, stakeholders can promote a more sustainable and environmentally conscious future, ensuring that AI-driven advancements do not come at the expense of the planet’s well-being.

What is green AI?

Green AI refers to the development and deployment of artificial intelligence technologies that prioritise environmental sustainability, energy efficiency, and reduced carbon emissions. The primary principles of green AI involve designing algorithms, models, and hardware that minimise energy consumption, harness renewable energy sources, and lower the environmental impact of AI applications throughout their lifecycles.

Green AI involves integrating energy efficiency and carbon emission reduction into every facet of AI development and deployment. It consists of making environmentally conscious decisions at every stage, from data acquisition to processing, model training, deployment, and monitoring. The goal is for green AI to become synonymous with AI, where energy efficiency, carbon emission reduction, and responsible resource utilisation are intrinsic to all AI projects.

What are the key strategies of green AI?

To successfully address the environmental challenges associated with AI technology, green AI can be divided into three key strategies that target different aspects of AI's impact. These strategies ensure a more comprehensive and effective approach to achieving a sustainable and responsible AI-driven future (Figure 4).



Figure 4. Key strategies of green AI

1. Greening systems

Greening systems involves implementing environmentally friendly practices in the design, manufacture, and operation of ICT, including hardware (Cloud/Data centres, end user devices) and software. The goal is to minimise energy consumption and waste by optimising both hardware and software components for environmental sustainability. This includes:

- Designing energy-efficient hardware that consumes less power and has a longer lifespan
- Incorporating environmentally friendly materials in manufacturing processes
- Optimising compute and storage utilisation to reduce idle resources and maximise efficiency. This involves using AI-driven operations (AIOps) to dynamically allocate and deallocate resources based on real-time demand, ensuring that hardware is actively used when needed and powered down or put into low-power states when not in use. By doing so, we can minimise the energy footprint and extend the life of the hardware.

2. Greening data

Greening data means employing efficient data storage, management, and processing techniques to reduce energy consumption and environmental impact. This strategy aims to minimise the overall energy requirements of AI systems by effectively managing data resources. Key approaches include:

- Reducing data redundancy and implementing data compression techniques to minimise storage needs
- Optimising data processing and management algorithms for energy efficiency
- Implementing storage tiers, with cold storage for data that is not needed after AI model training, to optimise energy usage and reduce environmental impact.

3. Greening intelligence

Greening intelligence revolves around developing and using smart algorithms that optimise energy consumption while balancing accuracy during training and deployment. This strategy emphasises the importance of creating energy-efficient algorithms and models without compromising sustainability goals. Aspects of greening intelligence include:

- Exploring techniques such as pruning, quantisation, and knowledge distillation to simplify models and reduce computational requirements
- Leveraging energy-efficient coding practices to reduce the energy needed to run AI applications
- Designing compact models that require less computation and energy consumption while maintaining accuracy and precision
- Utilising transfer learning techniques to leverage pre-trained models and fine-tuning models for specific tasks to reduce training time, computational resource requirements, and energy consumption compared with training models from scratch.

How can green AI contribute to a more environmentally friendly future?

Green AI adoption has the potential to contribute to a more environmentally friendly future by enhancing the efficiency of hardware, optimising algorithms, and promoting sustainable practices across various sectors.

One critical aspect of green AI research is enhancing the efficiency of the hardware needed for AI modelling. As AI demand grows, the development of more energy-efficient hardware becomes a market advantage. Companies such as NVIDIA, AMD, Intel, Google, and Amazon are creating new generations of chips and hardware to optimise GPU usage and other hardware components for lower energy consumption and greater cost-effectiveness for AI and machine learning.

In addition to more efficient hardware, green AI focuses on developing energy-efficient algorithms and models that require less computation to reduce the energy needed for AI applications. Techniques such as pruning, quantisation, and knowledge distillation can simplify models and decrease computation requirements, resulting in energy savings.

Green AI also emphasises the importance of adjusting the computation location to minimise energy consumption. Edge computing enables local devices to carry out some computations instead of relying on a data centre, reducing energy costs, particularly for data transmission. This approach brings AI benefits to local contexts such as homes, factories, and smart appliances within the Internet of Things (IoT), while also decreasing latency, which is essential for AI applications requiring real-time responses.

Data centre energy consumption can be further reduced through the adoption of efficient power supplies, cooling systems, and improved server utilisation. AI itself can be used to help better manage these factors and to identify opportunities for reducing energy inputs in data centres. For instance, Google used its DeepMind AI to decrease energy consumption for cooling its data centres by up to 40 percent. Incorporating renewable energy into data centre operations also holds promise. Engie recommends that we “exit the classic scenario,” of data centres connected to national grids and water networks and look instead to independent microgrids where green data centre power is drawn from local renewable energy sources²¹. In Montreal, Canada, there are already data centres that run purely on hydroelectricity, and the Citadel Campus in Reno-Tahoe—the largest data centre in the world at 7.2 million square feet—is powered by renewable energy.

In addition, green AI adoption can contribute to a more environmentally friendly future by optimising energy usage in other areas, such as buildings, industry, electricity grid management, and transportation systems. Encouraging sustainable practices, promoting energy-efficient technologies, and ensuring that policymakers and regulators stay aware of the evolving market and its inherent incentives are vital to fostering a greener, more sustainable AI-driven future.

What factors should guide green AI policy development?

Considering the rapidly evolving nature of AI technology, and concerns about its high energy demand, it is crucial for policymakers, technologists, and other stakeholders to evaluate the impact of AI within a proper framework. This requires comparing the energy use and outcomes achievable through AI's computing power to the alternatives. In other words, we must consider the amount of energy that would be required to generate the same outcomes, productivity increases, and economic growth without AI and machine learning, if those gains are even possible. Such an approach highlights both the costs and benefits of the increased energy consumed by AI-based technologies.

Nonetheless, if a market failure (where the market does not account for the environmental costs of AI energy consumption) is identified, any regulatory framework considered should be flexible enough to allow stakeholders to innovate and find market-driven solutions for reducing energy demands. Overly prescriptive rules may quickly become outdated, locking in suboptimal technologies, and unnecessarily impeding newer, more efficient approaches to manage energy consumption. Excessive regulation could also hinder the adoption of novel technologies or the deployment of renewable energy resources that could offer significant economic and social gains, including optimising power grids for more efficient electricity generation and delivery.

Regulating AI technology to curb energy usage is particularly challenging as AI spans various sectors such as healthcare, finance, transportation, and energy, each with existing regulators authorised to intervene in the marketplace. Before creating new regulatory bodies, it would be prudent to identify specific weaknesses, market failures, or challenges related to AI energy consumption that cannot be resolved within existing regulatory frameworks. To do so, policymakers require a deep understanding of the technology, its applications, and the implications of its use. Close collaboration between policymakers, AI researchers, industry stakeholders, and the public will result in a more robust policy framework for managing AI energy usage and the promotion of sustainable practices.

What are the key considerations for a sustainable green AI future?

To ensure a sustainable green AI future, it is essential to address several key considerations:

1. **Understanding AI's energy demand and growth:** A clear comprehension of AI energy demand and growth trajectory is crucial for aligning energy strategies with climate change objectives. This understanding will enable more informed decisions regarding investments in clean energy sources to ensure adequate supplies to meet future energy needs.
2. **Investments in energy-efficient AI algorithms and innovative hardware:** Proactively addressing the energy demands of AI and cloud computing is essential for aligning with long-term sustainability goals. Investments in energy-efficient AI algorithms, innovative hardware, and renewable energy sources will play a pivotal role in mitigating the environmental impact of these transformative technologies.
3. **Collaboration between stakeholders:** Open and informed discussions between AI developers, policymakers, and industry leaders are vital for ensuring a sustainable and resilient future. Developing regulations and incentives that encourage responsible and sustainable use of AI and cloud computing can steer the AI industry towards a greener and more sustainable path.
4. **Balancing AI energy needs with innovation and environmental stewardship:** Amid unprecedented technological advancements, it is crucial to prioritise and balance AI energy needs with innovation and environmental stewardship. By envisioning a future where we embrace clean energy, invest in renewable resources, and incorporate AI energy needs into demand scenarios, we can safeguard the future of the planet for generations to come.



References

1. Morrar, R.; Arman, H.; Mousa, S. The fourth industrial revolution (Industry 4.0): A social innovation perspective. *Technol. Innov. Manag. Rev.* 2017, 7, 12–20.
2. Lee, M.; Yun, J.; Pyka, A.; Won, D.; Kodama, F.; Schiuma, G.; Park, H.; Jeon, J.; Park, K.; Jung, K.; et al. How to respond to the fourth industrial revolution, or the second information technology revolution? Dynamic new combinations between technology, market, and society through open innovation. *J. Open Innov. Technol. Mark. Complex.* 2018, 4, 21.
3. Kirwan, C.; Fu, Z. *Smart Cities and Artificial Intelligence: Convergent Systems for Planning, Design, and Operations*; Elsevier: Oxford, UK, 2020.
4. Shukla, A.; Janmajaya, M.; Abraham, A.; Muhuri, P. Engineering applications of artificial intelligence: A bibliometric analysis of 30 years (1988–2018). *Eng. Appl. Artif. Intell.* 2019, 85, 517–532.
5. Cugurullo, F. Urban artificial intelligence: From automation to autonomy in the smart city. *Front. Sustain. Cities* 2020, 2, 1–14.
6. Simoes, G. Climate crisis and the technology sector, <https://ciandt.com/ca/en-ca/article/climate-crisis-and-technology-sector#:~:text=Current%20estimates%20show%20that%20the,emissions%20of%20the%20aviation%20sector>.
7. IMDA and GovTech unveil new initiatives to drive digital sustainability - Infocomm Media Development Authority <https://www.imda.gov.sg/resources/press-releases-factsheets-and-speeches/press-releases/2024/new-initiatives-to-drive-digital-sustainability>
8. D. Ben-Israel, W. B. Jacobs, S. Casha, S. Lang, W. H. A. Ryu, M. de Lotbiniere-Bassett, D. W. Cadotte, *Artif. Intell. Med.* 2020, 103, 101785.
9. D. M. Dimiduk, E. A. Holm, S. R. Niezgoda, *Integr. Mater. Manuf. Innov.* 2018, 7, 157.
10. G. Choy, O. Khalilzadeh, M. Michalski, S. Do, A. E. Samir, O. S. Pi-anykh, J. R. Geis, P. V. Pandharipande, J. A. Brink, K. J. Dreyer, *Radiology* 2018, 288, 318.
11. S. Athey, in *The Economics of Artificial Intelligence: An Agenda*, National Bureau of Economic Research, Inc., Cambridge 2018, p. 507.
12. XSEDE Impact—Usage Statistics, <https://portal.xsede.org/#/gallery> (accessed: January 2024).
13. I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, MIT Press, Cambridge, MA 2016.
14. R. Schwartz, J. Dodge, N. A. Smith, O. Etzioni, *Green AI. Commun. ACM* 2020, 63, 54.
15. Lannelongue Loïc, Grealey Jason, Inouye Michael, Green algorithms: quantifying the carbon footprint of computation. *Adv. Sci.*, 8 (12) (2021), Article 2100707
16. E. Strubell, A. Ganesh, A. McCallum, Energy and Policy Considerations for Deep Learning in NLP, in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, Jul. 2019, pp. 3645–3650, <https://doi.org/10.18653/v1/P19-1355>.
17. The Economist, the cost of training machines is becoming a problem, <https://www.economist.com/technology-quarterly/2020/06/11/the-cost-of-training-machines-is-becoming-a-problem>, 2020.
18. Is generative ai bad for the environment? <https://changemakr.asia/is-generative-ai-bad-for-the-environment/>.
19. The Rising Energy Demand of AI and Clouds: Unravelling the Environmental Conundrum, <https://agsiw.org/the-rising-energy-demand-of-ai-and-clouds-unraveling-the-environmental-conundrum/>
20. L. Lannelongue, J. Grealey, M. Inouye, Green Algorithms: Quantifying the Carbon Footprint of Computation. *Adv. Sci.* 2021, 8, 2100707.
21. Nicola Jones, how to stop data centres from gobbling up the world's electricity, *Nature* 561, 163-166 (2018) doi: <https://doi.org/10.1038/d41586-018-06610-y>
22. Replace your electricity consumption with 24/7 carbon-free energy generation, generating revenue and achieving high resiliency for your data center <https://www.engie.com/en/campaign/green-data-centers>

An aerial, high-angle photograph of a modern building with a green roof. The roof is covered in a dense layer of green plants and trees, interspersed with rows of solar panels. The building's structure is visible, showing a grid-like pattern of solar panels. The surrounding area is also green, with more trees and vegetation. The overall scene conveys a sense of sustainability and environmental friendliness.

green cloud computing and its role in sustainable AI

Sina Moradi

Green cloud computing and its role in sustainable AI

Executive summary

This article delves into the pivotal role of green cloud computing in attaining sustainability goals by emphasising the importance of green data centres, eco-friendly computer resource usage, decreasing carbon emissions, and the need for redesigning modern networking. Additionally, it provides an overview of the current state and benefits of green cloud computing, addressing environmental concerns, challenges, and future research directions. The article aims to inspire and guide stakeholders, policymakers, and researchers in their pursuit of more sustainable AI and cloud computing technologies.

Introduction

Before delving into the world of green cloud computing, it's essential to understand the environmental challenges posed by traditional data centres. These facilities are power-hungry giants that consume vast amounts of electricity and are responsible for substantial carbon emissions. The environmental impact is significant, given the growing demand for data storage and processing.

The cloud has emerged as a game-changing solution to these challenges. Cloud computing involves the use of remote servers, which are typically more energy-efficient and eco-friendly than traditional data centres. Over the past decade, cloud computing has become a major technological breakthrough, transforming the way organisations operate¹. Cloud computing allows companies to obtain computing resources, software, and applications over the Internet on an as-needed basis, eliminating the need for expensive hardware and software investments². By migrating to the cloud, businesses can reduce their carbon emissions, energy consumption, and the physical space required for on-premises servers. Cloud computing has been widely adopted by a wide range of organisations, from large corporations to small and medium-sized enterprises (SMEs)³.

NCS is a recognised leader in cloud professional services (CPS) in the Asia Pacific region according to the IDC MarketScape assessment for 2023-2024 (Figure 1)⁴. This survey revealed that 67% of enterprises in the region now have significant workload deployments on one or more public clouds, underscoring the growing importance of cloud to Enterprise IT. However, the same survey also revealed that enterprises in the region face critical challenges in their cloud journeys, including a failure to meet their desired performance and operational objectives, hybrid estate complexity, an inability to control costs, and a lack of necessary cloud-related skills, among others. This suggests that there is a lot of value that capable and experienced cloud professional services providers can bring to the cloud-centric migration and modernisation initiatives of enterprises⁴.

The increase in the popularity of cloud technology was due to the benefits it brings to individual consumers and companies. These benefits include flexibility, better disaster recovery, reduced need for investment in ICT resources, optimised collaboration between members of an organisation, and automatic software updates. Cloud computing is attractive to business owners based on its ability to dynamically increase the IT resources available to meet increases in the company's activities and needs.

To meet the increasing demand for computing resources, cloud computing is continuously evolving to provide client companies with well-utilised options such as Software as a Service (SaaS), Platform as a Service (PaaS), Infrastructure as a Service (IaaS), and Container as a Service (CaaS)⁵. These services enable customers to buy specific resources, pay only for their usage, and scale their services on-demand with pay-per-use models.

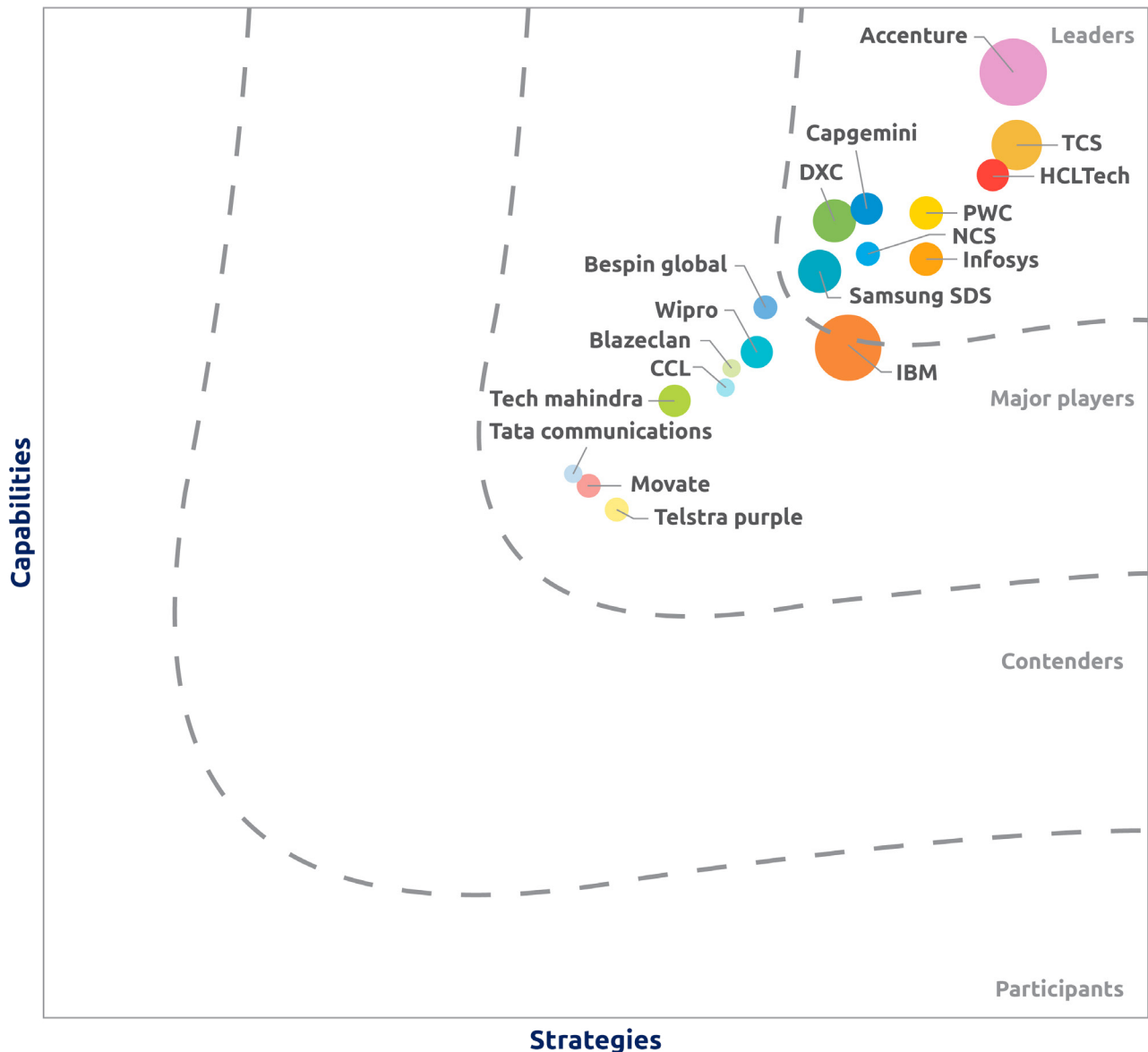


Figure 1. IDC MarketScape Asia/Pacific cloud professional services 2023-2024⁴

Organisations are expected to employ efficient and innovative technologies to maintain optimal levels of productivity within their ecosystems⁶. Cloud computing offers a cost-effective and flexible solution to achieve these objectives by providing access to IT resources⁷⁻⁸. The cloud computing market keeps expanding to achieve automation, streamline operations, reduce costs, and foster business growth.

A survey by Gartner found that 81% of organisations have adopted a multi-cloud IT strategy to use advanced capabilities across clouds, a number that is expected to increase and that confirms that shift towards digitalisation is a competitive imperative⁹. As illustrated in Figure 2, use of cloud computing technologies is expected to increase at an impressive compound annual growth rate (CAGR) through the year 2028¹⁰. This anticipated growth reflects the ongoing emphasis on integrating advanced technologies to enhance operational efficiency and support business expansion. Additionally, there has been a growing emphasis on sustainable growth in the business world¹¹. Organisations are increasingly striving to integrate sustainability into their operations and focusing, not just on economic performance, but also on social and environmental factors¹²⁻¹³. The emergence of cutting-edge technologies, such as cloud computing services, has accelerated this shift. To achieve a balance between economic and operational performance, organisations must improve their capabilities and promote sustainable performance¹⁴. This is where the concept of green cloud computing comes into the picture. Green cloud computing is a concept that merges the power of cloud technology with eco-conscious initiatives. This innovative approach not only streamlines operations but also contributes to a greener, more sustainable future.

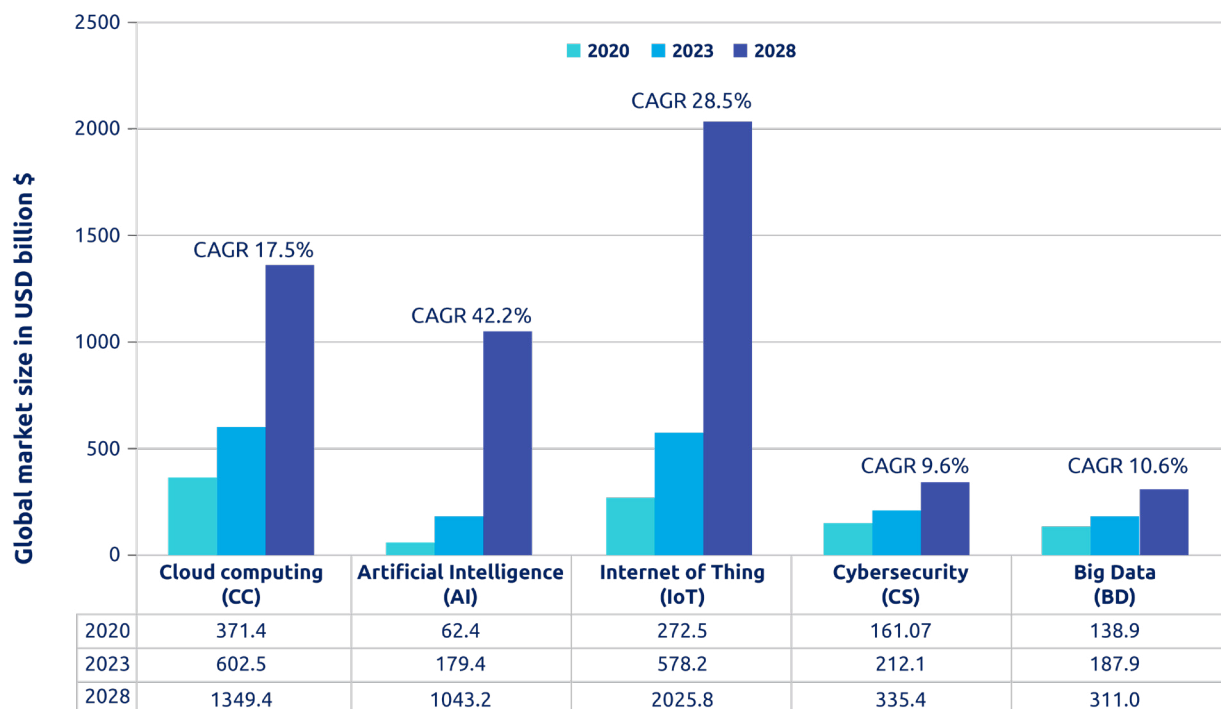


Figure 2. Forecasted CAGR growth of IT sector over the next five years¹⁰.

Interest in studying the influence of cloud computing on the environment is rising as the computing community becomes more aware of green computing. That focus is being driven by estimates that show that the ICT sector is currently responsible for approximately 4% of global CO₂ emissions¹⁵, and that percentage could reach 14% by 2040, with data centres contributing to almost half of this growth¹⁵.

The development of green cloud computing is closely related to the evolution of green data centres because data centres are at the core of cloud computing. The interest in green cloud computing and green data centres is evident in the future strategies of leading providers such as Singapore's Singtel. The design, build, and operation of green and sustainable infrastructure such as green data centres is a key pillar of Singtel's vision (Figure 3)¹⁶.

Vision

To develop and grow the industry’s greenest and most sustainable next generation digital assets



Figure 3. Singtel’s vision on the development of sustainable next generation digital assets¹⁶

Singtel is scaling its regional data centre business, having gained a foothold in new markets beyond Singapore, including Indonesia and Thailand, where it is building a new generation of green, sustainable, and hyper-connected AI-ready data centres. Recently, Singtel also launched its new green and sustainable data centre brand, Nxera, to accelerate the company’s journey to net-zero emissions¹⁷. Figure 4 shows Singtel’s different strategies to develop energy efficient green data centres across the region¹⁶.

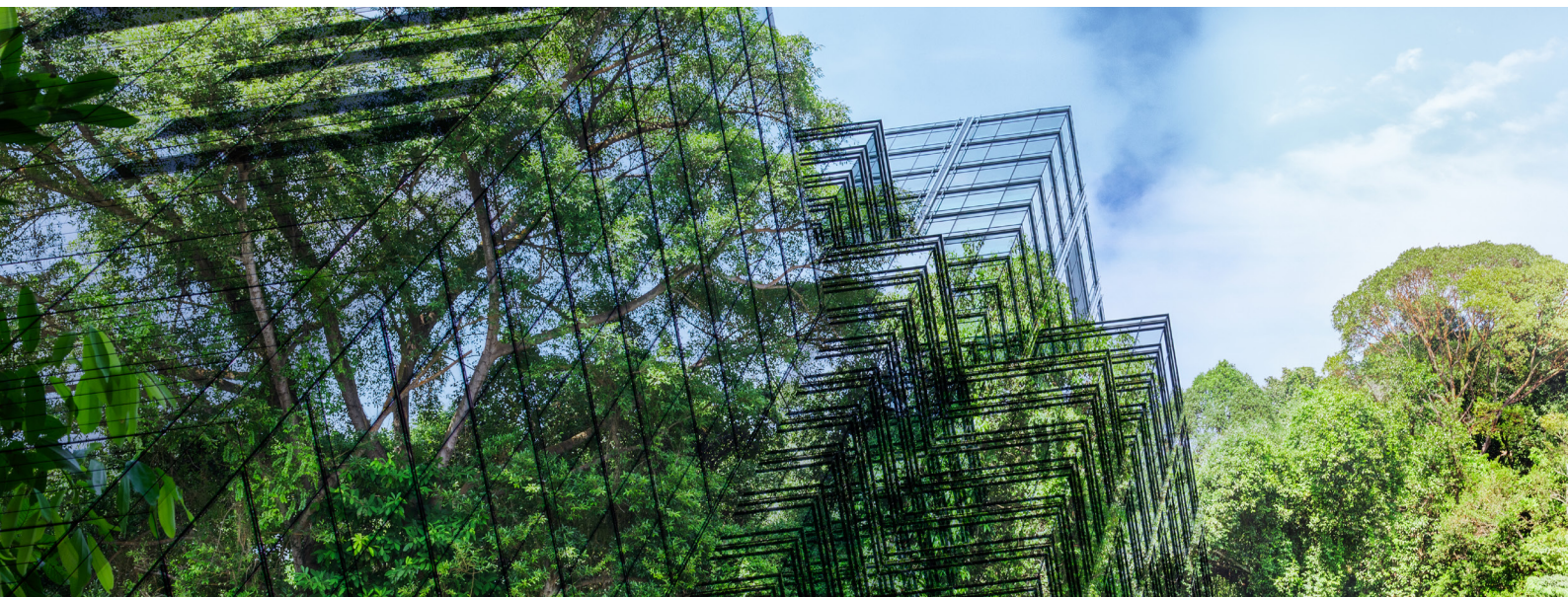




Figure 4. Singtel's different strategies to develop energy efficient green data centres across the region¹⁶.

Green data centres-where energy efficiency is maximised and CO2 emissions and e-waste are minimised, not only for ICT equipment, but across all environmental factors (building, lightning, cooling, etc.) are the basis for current and future green cloud computing. Green computing is not limited to minimising the energy consumption of computer devices. It also concerns the energy consumption of networks and cooling equipment and other environmental issues, such as CO2 emissions, waste management (e-waste), and the overall consumption of natural resources.

Today, we need green technology to solve energy and performance trade-offs. Proper distribution of workloads, and the maintenance of performance by providing access to services anywhere, anytime to users, is one of the cloud's prime objectives. A key challenge for cloud service providers in balancing energy conservation with the performance guarantees that usually form a part of their service level agreements (SLA) with client companies. ¹⁸ Minimising energy usage may cause breaches of Service Level Agreements (SLA) and result in penalties to cloud service providers, and thus we should avoid any SLA violations when considering energy-efficiency systems¹⁸.

The pillars of green cloud computing

Green cloud computing focuses on integrating sustainability throughout the design, implementation, and operation of cloud infrastructure. The following key pillars underpin this approach (Figure 5):

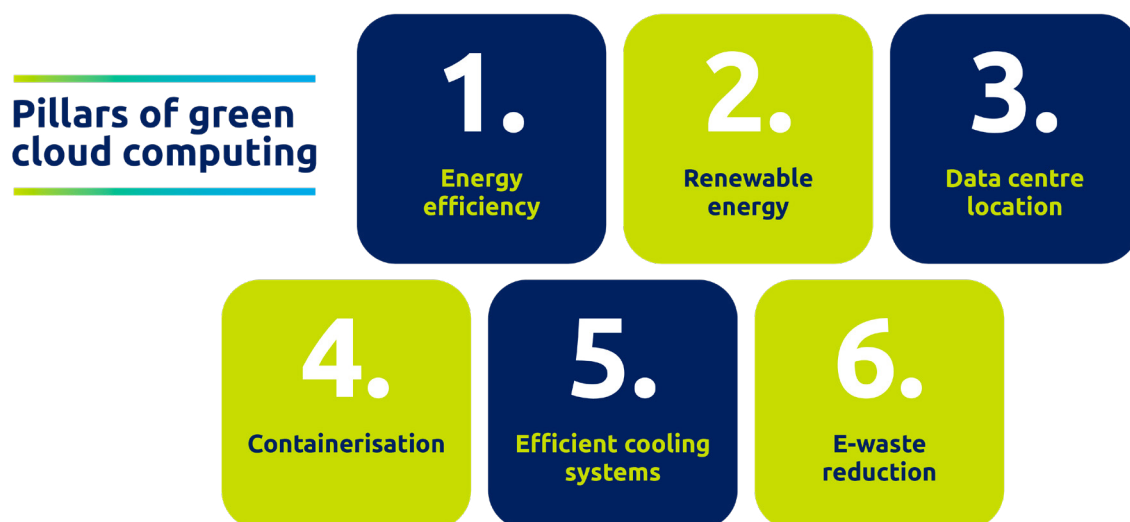


Figure 5. The pillars of green cloud computing.

- 1. Energy efficiency:** A core principle of green cloud computing is prioritising energy-efficient data centres and servers. Facilities are optimised to minimise power usage and waste, resulting in significant energy savings and improved resource utilisation.
- 2. Renewable energy:** Central to green cloud computing is the use of renewable energy sources such as solar and wind power. By leveraging these clean energy options, cloud providers can decrease their dependence on fossil fuels and reduce their overall carbon footprints.
- 3. Data centre location:** The geographic location of data centres plays a crucial role in their sustainability. Situating data centres in regions with access to renewable energy sources allows cloud providers to further enhance their eco-friendly operations.
- 4. Containerisation:** Traditional data centres can be highly power-intensive. However, the adoption of virtualisation and containerisation technologies can help mitigate this issue by decreasing energy consumption. Virtual machines and containers enable hardware to be sub-divided into multiple segments, allowing various operating systems or applications to run on a single server. This configuration minimises the number of servers needed, while maximising resource utilisation, leading to reduced energy usage. Utilising container services such as Kubernetes to run compute workloads on-demand allows for flexible scaling up and down without the need to maintain maximum hardware capacity for occasional peak usage or traffic. This approach lowers overall power consumption and contributes to a reduced carbon footprint for workloads on the cloud.
- 5. Efficient cooling systems:** Data centres generally require extensive cooling systems to maintain appropriate operating temperatures. Green cloud computing emphasises the implementation of more efficient cooling mechanisms to lower overall energy demand.
- 6. E-Waste reduction:** Ensuring responsible disposal and recycling of electronic waste (e-waste) is essential for sustainability. Green cloud computing requires the recycling or disposal of outdated equipment in an environmentally conscious manner to minimise adverse ecological impacts.

Importance and benefits of green cloud computing

Sustainability in cloud computing refers to the adoption of environmentally sustainable practices in the design, deployment, and management of cloud computing systems¹⁹. Although cloud computing provides numerous advantages, including enhanced efficiency, scalability, and reduced costs, it also has substantial environmental impacts, such as energy usage, carbon emissions, and e-waste generation.²⁰⁻²² It is crucial to prioritise sustainability in the domain of cloud computing for multiple reasons:

- **Environmental impact:** Cloud computing has a significant environmental impact due to the energy consumption, carbon emissions, and e-waste generated by data centres. By promoting sustainability in cloud computing, we can reduce these negative impacts and contribute to a more sustainable future²¹
- **Resource efficiency:** Sustainability in cloud computing promotes efficiency by optimising the use of shared computing resources. This can result in lower energy consumption, reduced carbon emissions, and less e-waste²⁰⁻²¹
- **Cost savings:** Sustainable practices in cloud computing can result in cost savings for cloud providers and end-users. For example, energy-efficient hardware and cooling systems can reduce energy consumption and lower operating costs, while recycling and reusing hardware can limit e-waste and lower replacement costs²⁰
- **Social responsibility:** Promoting sustainability in cloud computing is a social responsibility for cloud providers and demonstrates their commitment to reducing their environmental impact and contributing to a more sustainable future. These efforts can also improve their reputations and brand images among environmentally conscious customers²²
- **Regulatory compliance:** Many countries and regions have regulations and standards related to environmental sustainability. Green cloud computing helps companies stay in compliance with these evolving standards and avoid potential legal or financial penalties²⁰
- **Scalability:** Cloud solutions, especially those with green features, offer scalable resources that can be adapted to meet a company's changing needs. This flexibility supports growth without massive increases in environmental impact.

Sustainability is important in cloud computing because it can help to reduce negative environmental impacts, promote resource efficiency, generate cost savings, demonstrate social responsibility, and aid in compliance with regulatory requirements.

Challenges and future outlook

Green cloud computing and other energy-efficient technologies can play a crucial role in reducing the global carbon footprint and creating a more sustainable world. Although many argue that technological advancements contribute to environmental degradation, cloud computing can serve as a solution to address ecological challenges. Green cloud computing enables organisations to achieve a positive bottom line, boost employee efficiency, reinvent business processes, and promote a sustainable environment.

Nevertheless, cloud computing comes with its share of challenges, including internet dependency, data security and privacy concerns, and sometimes, a costly entry barrier. Despite these issues, green cloud computing's advantages far outweigh those of traditional technology, leading to net benefits and better environmental impacts for consumers and businesses.

As an increasing number of customers and companies prioritise environmentally conscious decisions, it becomes essential to recognise the importance of green cloud computing. Data centre providers have accelerated efforts to mitigate their climate impact with energy-efficient cooling systems and AI optimisation, which will continue to improve with technological innovation. Governments can develop policies that incentivise the production and the availability of renewable energy, including through the facilitation of cross border energy trade to power data centres. Policymakers can further raise standards to encourage greener data centres by setting interoperable standards for Power Usage Effectiveness (PUE) and by pushing for green building certification.

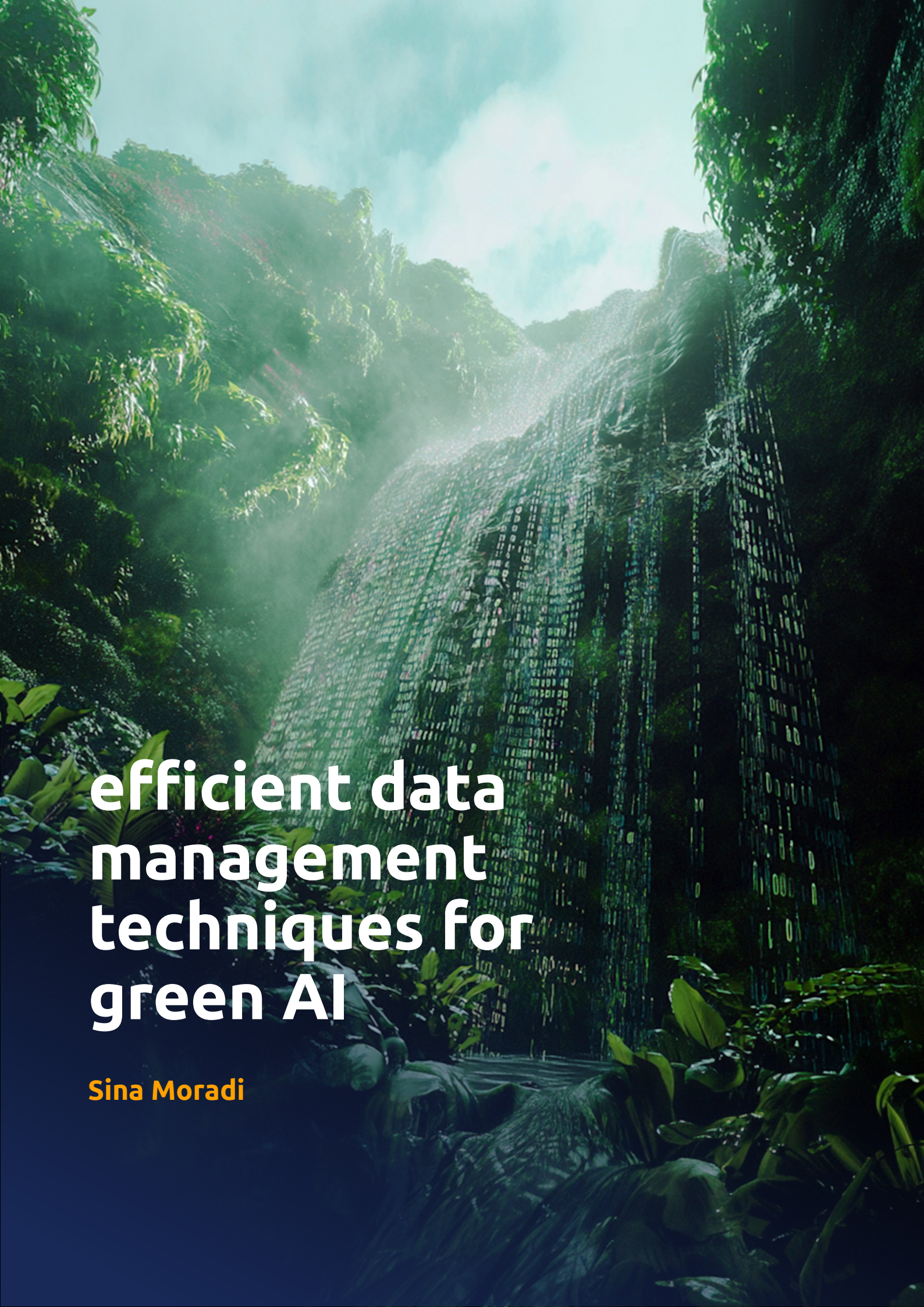
The future of green cloud computing appears promising. With heightened awareness of environmental concerns, businesses and cloud providers will likely invest more in sustainable practices. Furthermore, innovations such as liquid cooling systems and edge computing will contribute to enhanced energy efficiency and the development of cutting-edge technology.

Green cloud computing is not merely a passing trend; it's an essential component for achieving a sustainable future. By adopting environmentally responsible cloud solutions, businesses can diminish their ecological footprint while enjoying economic and operational advantages. The road to a greener, more sustainable, future is in sight and green cloud computing is paving the way.

References

1. Y.A. Qasem, R. Abdullah, Y. Yah, R. Atan, M.A. Al-Sharafi, M. Al-Emran, Towards the development of a comprehensive theoretical model for examining the cloud computing adoption at the organisational level, in: *Recent Adv. Intell. Syst. Smart Appl.*, Springer, 2021, pp. 63–74.
2. L. Tobarra, J.M. Haut, R. Hernandez, R. Pastor-Vargas, A. Robles-Gomez, Analysing the users' acceptance of an IoT cloud platform using the UTAUT/TAM model, *IEEE Access* 9 (2021) 150004–150020, <https://doi.org/10.1109/ACCESS.2021.3125497>.
3. J.P. Shetty, R. Panda, An overview of cloud computing in SMEs, *J. Glob. Entrep. Res.* 11 (2021) 175–188, <https://doi.org/10.1007/S40497-021-00273-2>.
4. CLOUD PROFESSIONAL SERVICES "LEADERS" IN 2023 FOR ASIA/PACIFIC REVEALED BY IDC, 30 NOVEMBER 2023, <HTTPS://WWW.IDC.COM/GETDOC.JSP?CONTAINERID=PRAP51432123>
5. ABDULSALAM, S.; LAKOMSKI, D.; GU, Q.; JIN, T.; ZONG, Z. PROGRAM ENERGY EFFICIENCY: THE IMPACT OF LANGUAGE, COMPILER AND IMPLEMENTATION HOICES. IN PROCEEDINGS OF THE INTERNATIONAL GREEN COMPUTING CONFERENCE, DALLAS, TX, USA, 3–5 NOVEMBER 2014; PP. 1–6.
6. R. EL-HADDADEH, M. OSMANI, N. HINDI, A. FADLALLA, VALUE CREATION FOR REALISING THE SUSTAINABLE DEVELOPMENT GOALS: FOSTERING ORGANISATIONAL ADOPTION OF BIG DATA ANALYTICS, *J. BUS. RES.* 131 (2021) 402–410, <HTTPS://DOI.ORG/10.1016/J.JBUSRES.2020.10.066>.
7. Y.A.M. Qasem, S. Asadi, R. Abdullah, Y. Yah, R. Atan, M.A. Al-Sharafi, A.A. Yassin, A multi-analytical approach to predict the determinants of cloud computing adoption in higher education institutions, *Appl. Sci.* 10 (2020) 4905, <https://doi.org/10.3390/app10144905>.
8. M.A. Al-Sharafi, R.A. Arshah, E.A. Abu-Shanab, Questionnaire development process to measure the SMEs' continuous use behaviour towards cloud computing services, *ACM Int. Conf. Proceeding Ser.* (2019) 50–55, <https://doi.org/10.1145/3316615.3316723>.
9. Gartner. Accelerating Shift to the Cloud Means the Market Opportunity for Providers is Narrowing Available online: <https://www.gartner.com/en/newsroom/press-releases/2022-02-09-gartner-says-more-than-half-of-enterprise-it-spending> (accessed on 21 January 2024).
10. Yenugula, M.; Goswami, S.S.; Kaliappan, S.; Saravanakumar, R.; Alasiry, A.; Marzougui, M.; AlMohimeed, A.; Elaraby, A. Analysing the Critical Parameters for Implementing Sustainable AI Cloud System in an IT Industry Using AHP-ISM-MICMAC Integrated Hybrid MCDM Model. *Mathematics* 2023, 11, 3367. <https://doi.org/10.3390/math11153367>
11. K. Domdouzis, Sustainable cloud computing, in: C. Pattinson, B. Akhgar (Eds.), *Green Inf. Technol.*, Morgan Kaufmann, Boston, 2015, pp. 95–110, <https://doi.org/10.1016/b978-0-12-801379-3.00006-1>.
12. O. Yavuz, M.M. Uner, F. Okumus, O.M. Karatepe, Industry 4.0 technologies, sustainable operations practices and their impacts on sustainable performance, *J. Clean. Prod.* 387 (2023), 135951, <https://doi.org/10.1016/J.JCLEPRO.2023.135951>.
13. S. Gupta, R. Meissonier, V.A. Drave, D. Roubaud, Examining the impact of Cloud ERP on sustainable performance: a dynamic capability view, *Int. J. Inf. Manag.* 51 (2020), 102028, <https://doi.org/10.1016/J.IJINFOMGT.2019.10.013>.
14. S.U. Rehman, H. Elrehail, D. Alshwayat, B. Ibrahim, R. Alami, Linking hotel environmental management initiatives and sustainable hotel performance through employees' eco-friendly behaviour and environmental strategies: a moderated-mediated model, *Eur. Bus. Rev.* 35 (2023) 184–201, <https://doi.org/10.1108/EBR-05-2022-0094/FULL/PDF>.
15. Simoes, G. Climate crisis and the technology sector, <https://ciandt.com/ca/en-ca/article/climate-crisis-and-technology-sector#:~:text=Current%20estimates%20show%20that%20the,emissions%20of%20the%20aviation%20sector>.

16. Bill Chang, CEO of Digital Infrastructure Co, Singtel Investor Day (24 August 2023), <https://www.singtel.com/content/dam/singtel/investorRelations/factsheetsAndPresentation/2023/Singtel%20Investor%20Day%202023%20-%20Digital%20InfraCo.pdf>
17. Singtel advances AI development in Singapore and the region with strategic partners <https://www.singtel.com/about-us/media-centre/news-releases/singtel-advances-ai-development-in-singapore-and-the-region-with>
18. Kitchenham, B.; Brereton, O.P.; Budgen, D.; Turner, M.; Bailey, J.; Linkman, S. Systematic literature reviews in software engineering—A systematic literature review. *Inf. Softw. Technol.* 2009, 51, 7–15.
19. Kumar, R.; Khan, F.; Kadry, S.; Rho, S. A Survey on blockchain for industrial Internet of Things. *Alex. Eng. J.* 2021, 61, 6001–6022.
20. Sharma, R.; Villányi, B. A sustainable Ethereum merge-based Big-Data gathering and dissemination in IIoT System. *Alex. Eng. J.* 2023, 69, 109–119.
21. Goswami, S.S.; Mohanty, S.K.; Behera, D.K. Selection of a green renewable energy source in India with the help of MEREC integrated PIV MCDM tool. *Mater. Today: Proc.* 2022, 52, 1153–1160.
22. Wang, L.; Wang, Y. Supply chain financial service management system based on block chain IoT data sharing and edge computing. *Alex. Eng. J.* 2021, 61, 147–158.



efficient data management techniques for green AI

Sina Moradi

Efficient data management techniques for green AI

Executive summary

This article delves into various techniques for efficient data management in the context of green AI, such as data compression, deduplication, and storage optimisation, while also examining innovations in data processing and management algorithms that aim to improve energy efficiency. The article outlines the benefits of adopting sustainable data management practices, the potential consequences of over-regulation in AI and data management, and the importance of task scheduling and resource allocation strategies to achieve energy efficiency optimisation in AI systems.

Introduction

As the digital and technological revolution continues to surge and expand, enormous amounts of data are being generated daily. It is estimated that by 2025, approximately 463 exabytes of data will be produced every day¹. To put this into perspective, it would take around 237,832 years for a video conference to generate just one exabyte of data. This immense volume of data has led businesses to adopt eco-friendly approaches² and driven the creation of green data centres and sustainable data management practices. Green data centres are designed to employ energy-efficient technologies and practices to reduce energy consumption and carbon emissions, while utilising renewable energy sources. Sustainable data management refers to the processes and techniques used to collect, store, and process data in ways that minimise negative environmental impacts. By combining green data centres with sustainable content and data management, users can significantly reduce their environmental footprints for data storage and processing.

In 2024, we've witnessed an unprecedented whirlwind of interest in landmark AI technology. While the pursuit of 'big' AI persists, it is essential not to overlook the practical implementation of AI for everyday business applications. Accessing AI-driven insights requires time, resource investment, and substantial effort. It is vital to recognise that data, the driving force behind AI, is instrumental in propelling AI adoption. The cleaner the data, the more efficient AI solutions become. Data has the power to optimise these tools, ensuring that they reach their maximum potential.

However, the energy requirements of AI systems are an escalating concern, particularly with the growing utilisation of Deep Learning and other computation-intensive algorithms. This concern has become increasingly apparent with the widespread adoption of Large Language Models (LLMs) that rely on terabytes of data for training, that require considerable computing power and financial investment, and lead to significant carbon dioxide emissions. Consequently, a new sub-field has emerged to make the development and application of AI technologies environmentally sustainable: green AI³.

With the expanding use of AI across diverse sectors, such as manufacturing, finance, and intelligent transportation systems, the leaders in the Asia-Pacific region (APAC) must urgently devise innovative data management solutions that can capitalise on AI's transformative potential while aligning with environmental objectives such as those outlined in the Singapore Green Plan 2030. Efficient data management is crucial for creating energy-conscious AI systems, as data-related operations like storage, retrieval, and pre-processing can consume substantial resources. The results of a recent study showed that a technique such as feature selection can reduce the energy consumption of AI model training by up to 76%, while preserving the performance of the model⁴. This article explores how data management can contribute to energy reduction through techniques like data compression, streamlined data pipelines, and other inventive approaches.

The environmental impact of conventional data management methods

Organisations process vast amounts of data to achieve their business outcomes, making proper data management and governance crucial to their success. As sustainability becomes a focal point in both business and technology, it is essential to design data management strategies with the environment in mind. Conventional data management methods often overlook the environmental impact of massive data centres, leading to an increased carbon footprint. These methods tend to be resource-intensive, consume large amounts of energy (often from non-renewable sources), and lack optimisation levers for usage, resulting in wastage of both compute and storage resources. Consequently, businesses face higher operational costs and struggle to reduce their overall environmental impact.

In data management, the impact is not only driven by compute provisioning factors, but also by practices adopted throughout the data lifecycle. Figure 1 shows three common challenges in conventional data management practices that lead to inefficiencies and wastage:

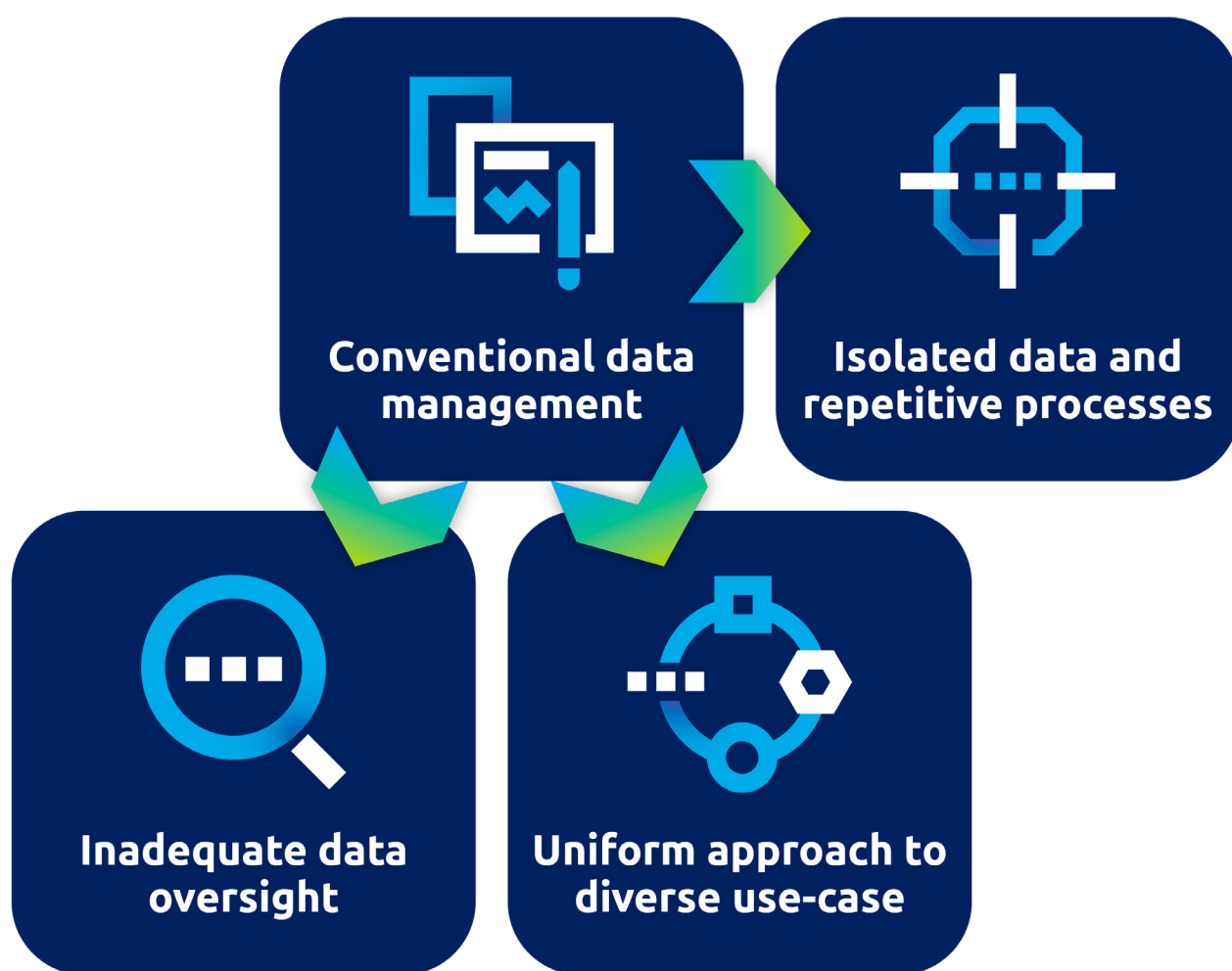


Figure 1. Challenges in conventional data management practices.

1. **Isolated data and repetitive processes:** Data isolation occurs when data is trapped in separate departments or teams within an organisation, causing employees to re-do the same tasks, unaware of each other's progress. This results in data duplication and wastes storage, time, effort, and resources.
2. **Uniform approach to diverse use cases:** Using the same tool for all data management use cases leads to inefficiencies. Different use cases demand specialised solutions for optimal outcomes. For example, when you use a general-purpose database like MySQL for handling both transactional and analytical workloads, you may experience performance bottlenecks due to the diverse nature of these tasks. Instead, using specialised databases such as PostgreSQL for transactional workloads and Apache Cassandra for analytical workloads can lead to better performance, workload isolation, and an improved user experience.
3. **Inadequate data oversight:** The absence of data oversight directly contributes to resource wastage. Without proper supervision and standardisation, data becomes unorganised and duplicated, often leading to data quality issues and inaccurate metrics. Teams spend extra time locating the right datasets, moving data over the network, and correcting data inaccuracies, which requires more storage and processing power and could be avoided with robust data oversight practices.

Conventional data management practices can have a significant adverse environmental impact due to their resource-intensive nature and lack of optimisation. By adopting eco-friendly data management practices, organisations can minimise their environmental footprints, reduce operational costs, and contribute to a greener future.

Green AI and importance of sustainable data management

Green AI is a burgeoning subdomain that concentrates on creating and applying AI technologies in an eco-friendly manner. The primary goal of green AI is to address the energy demands and environmental consequences of AI systems, particularly those employing computation-heavy algorithms and vast quantities of data. A crucial aspect of green AI is sustainable data management, which is vital for minimising the environmental impact of AI systems and ensuring their responsible utilisation.

Sustainable data management involves the diligent control and treatment of data throughout its entire lifecycle, covering data collection, processing, storage, and disposal. The main aim of sustainable data management is to lessen the environmental effects of data management practices, reduce energy consumption, and maximise resource efficiency. Furthermore, sustainable data management focuses on guaranteeing that data is employed in a socially responsible and ethical manner.

The significance of sustainable data management is manifold. First, it assists in reducing the environmental consequences of data management practices. Data centres and other IT infrastructure are responsible for significant energy consumption and carbon emissions. Sustainable data management practices seek to lower energy consumption and carbon emissions by refining data centre design, boosting energy efficiency, and utilising renewable energy sources. Second, sustainable data management aids in decreasing waste and optimising resource usage. Data storage devices, such as hard drives and solid-state drives, consist of rare earth metals and other limited materials. Sustainable data management practices endeavour to minimise waste by ensuring effective data storage and processing, as well as responsible disposal.

Lastly, sustainable data management is essential for guaranteeing the socially responsible and ethical utilisation of data. Data privacy and security are critical concerns for organisations managing sensitive information. Sustainable data management practices aim to protect data from unauthorised access and usage while respecting individuals' privacy rights.

Sustainable data management is a fundamental component of green AI for addressing the environmental, resource, and ethical concerns associated with data management practices. By embracing sustainable data management techniques, organisations can contribute to the development of eco-friendly AI systems and encourage responsible data usage.

Techniques for efficient data management

With the increasing volume of data being generated every day, the need for efficient data management has become more critical than ever before. Techniques for efficient data management involve data compression, data deduplication, storage optimisation, and innovations in data processing and management algorithms (Figure 2).

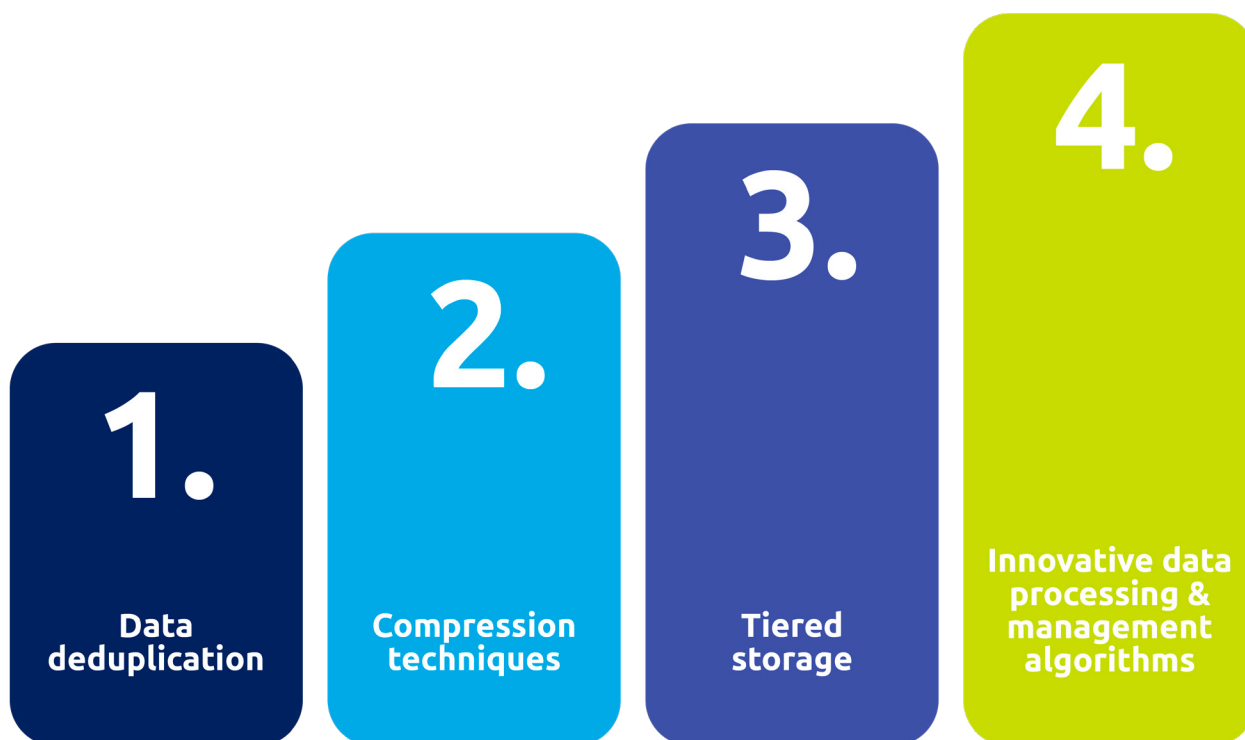


Figure 2. Techniques for efficient data management.

Data deduplication is a technique used to eliminate redundant data by identifying and removing duplicate copies of data so that a single unique instance can be stored. Deduplication can be performed at the file level or the block level. File-level deduplication identifies and removes entire duplicate files, while block-level deduplication identifies and removes duplicate blocks of data within files. Data deduplication can significantly reduce storage requirements and improve storage efficiency.

Data compression is a technique used to reduce the size of data to make it more efficient to store and transmit. In recent years, many data compression techniques have been proposed for big data in cloud computing. These techniques include lossless compression, lossy compression, and hybrid compression. Lossless compression techniques, such as Huffman coding and Lempel-Ziv-Welch (LZW) coding, can achieve high compression ratios but have slower compression and decompression speeds. Lossy compression techniques, such as JPEG and MPEG, can achieve higher compression speeds but sacrifice some data fidelity. Hybrid compression techniques, such as Burrows-Wheeler transform and Run-Length Encoding (RLE), combine the advantages of both lossless and lossy compression techniques⁵.

Storage optimisation techniques are used to improve the efficiency of data storage by organising and managing data in a way that minimises resource usage. Innovative data processing and management algorithms aim to improve energy efficiency by optimising data processing tasks, reducing resource usage, and minimising energy consumption. Examples of such innovations include energy-aware algorithms, adaptive data processing techniques, and machine learning-based optimisation methods.

Through advanced compression algorithms, AI systems can maintain data integrity while minimising storage and energy costs. Additionally, storage optimisation techniques, such as data deduplication and tiered storage, can further reduce energy consumption by minimising redundant data and prioritising energy-efficient storage solutions.

Best practices for sustainable data management

Sustainable data management is essential for creating green AI solutions that balance performance and environmental friendliness. By adhering to best practices in data management, organisations can ensure that that they minimise energy consumption and reduce the environmental impact of AI systems. The following best practices for sustainable data management in green AI should be considered (Figure 3):

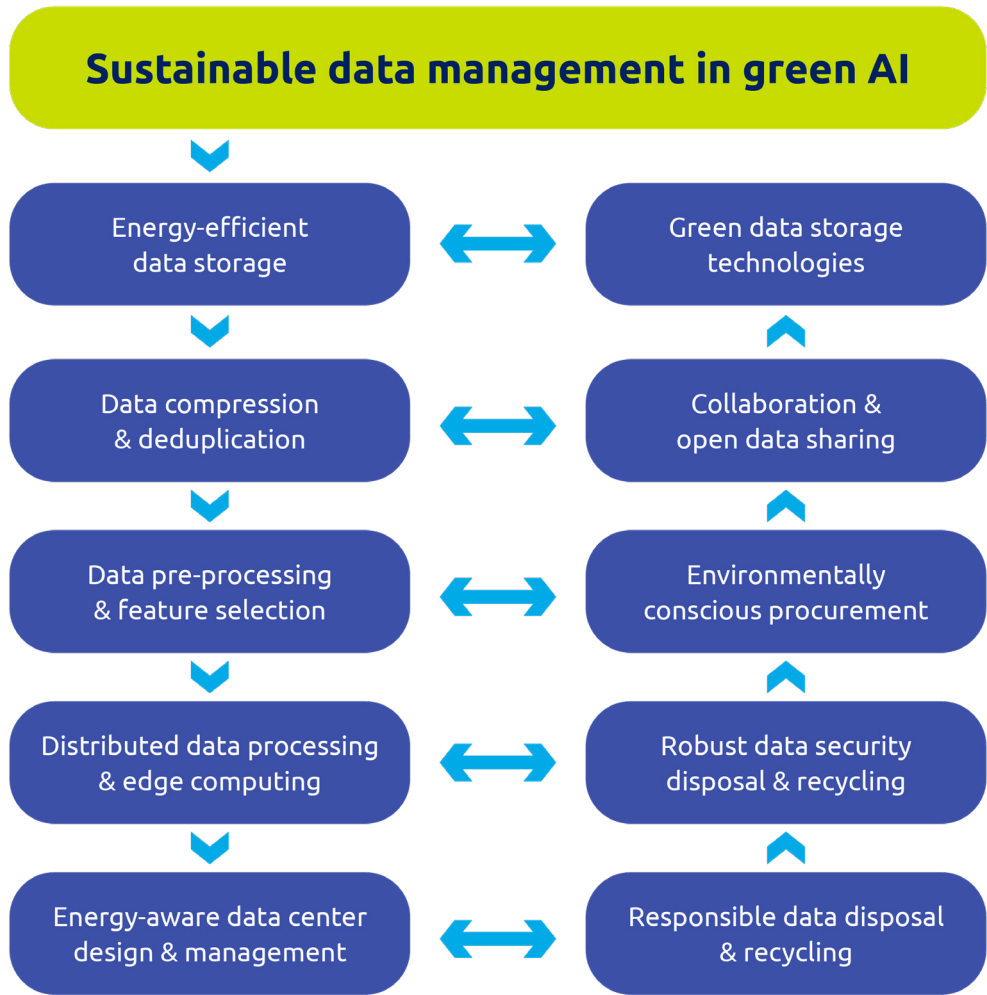


Figure 3. Best practices for sustainable data management in green AI.

1. **Energy-efficient data storage and management**
 - Enhance data centre design and energy efficiency while utilising renewable energy sources to decrease energy consumption and carbon emissions.
 - Apply virtualisation, consolidation, and compression methods to diminish the physical footprint of data centres.
 - Use data lifecycle management strategies to ensure data is stored efficiently to minimise the need for physical storage.
2. **Data compression and storage optimisation**
 - Use advanced data compression algorithms to reduce the size of datasets without compromising data integrity.
 - Implement data deduplication techniques to eliminate redundant data and minimise storage requirements.
3. **Data pre-processing and feature selection**
 - Clean and pre-process data to remove inconsistencies, missing values, and outliers to ensure that AI systems process only relevant information.
 - Apply feature selection techniques to pinpoint the most crucial features in datasets to lower computational complexity and energy consumption during training and inference.
4. **Distributed data processing and Edge computing**
 - Utilise distributed data processing techniques to process data across multiple devices to reduce the energy consumption associated with data transmission and central processing.
 - Implement edge computing to process data closer to the source to reduce data transmission and latency and enhance energy efficiency.
5. **Energy-aware data centre design and management**
 - Design and manage data centres with energy efficiency in mind using techniques such as free cooling, hot aisle/cold aisle containment, and efficient power distribution.
 - Monitor and optimise data centre energy consumption through real-time energy management systems and predictive analytics.
6. **Responsible data disposal and recycling**
 - Implement proper data disposal and storage device recycling to recover valuable materials such as rare earth metals and other limited resources.
 - Use secure data deletion practices to protect data privacy and security during the disposal process.
7. **Robust data security and privacy**
 - Establish strong security measures, including data encryption and access controls, to safeguard data from unauthorised access and use throughout its lifecycle.
 - Emphasise data privacy by incorporating secure data management practices at all stages of data handling.

8. **Environmentally conscious procurement practices**

- Evaluate the environmental and social impact of procured products and services, considering the energy efficiency of IT equipment, the use of renewable energy sources, and the environmental impact of materials in data storage devices.
- Adopt eco-friendly procurement practices to minimise the environmental footprint of data management operations.

9. **Data lifecycle management**

- Implement data lifecycle management policies to ensure that data is stored, processed, and disposed of in an energy-efficient and environmentally friendly manner.
- Regularly review and update data retention policies to minimise the unnecessary storage and processing of outdated or irrelevant data.

10. **Collaboration and open data sharing**

- Encourage collaboration and open data sharing among researchers and organisations to reduce redundant data collection, and processing efforts to minimise energy consumption and environmental impact.
- Utilise data sharing platforms and data repositories that prioritise energy efficiency and sustainability.

11. **Green data storage technologies**

- Invest in energy-efficient storage technologies, such as solid-state drives (SSDs) and low-power hard disk drives (HDDs), to reduce energy consumption in data storage.
- Explore emerging green storage technologies, such as DNA storage and phase-change memory, which have the potential to significantly reduce energy consumption and environmental impact.

By implementing these best practices for sustainable data management, businesses can ensure that AI systems are optimised for energy efficiency while maintaining high performance and ultimately contribute to the development of green AI solutions.

References

1. How much data is generated each day? <https://www.weforum.org/agenda/2019/04/how-much-data-is-generated-each-day-cf4bddf29f/> via @wef
2. Future of Green Business Through Sustainable Data Management, <https://www.allerin.com/blog/future-of-green-business-through-sustainable-data-management>
3. R. Schwartz, J. Dodge, N. A. Smith, and O. Etzioni, "Green AI," Communications of the ACM, vol.63, pp.54–63,112020.
4. 2. R. Verdecchia, et al., "Data-Centric Green AI An Exploratory Empirical Study," in 2022 International Conference on ICT for Sustainability (ICT4S), Plovdiv, Bulgaria, 2022 pp. 35-45.
5. T. Rajkumar Talekar, Efficient data compression techniques for big data in cloud computing: a comparative study, International Research Journal of Modernisation in Engineering Technology and Science, Volume:05/Issue:06/June-2023.

The background of the slide features a digital rain effect, with streams of blue and green binary code (0s and 1s) falling from the top. In the foreground, three people in business attire are shown from behind, standing in a dark, silhouetted landscape and looking up at the digital rain. The overall color palette is dark blue and green.

green AI in Edge computing: Leveraging localised intelligence for sustainability

Sina Moradi

Green AI in Edge computing: Leveraging localised intelligence for sustainability

Executive summary

This article delves into the potential of Edge computing in achieving green AI. By processing AI tasks on local devices and minimising the reliance on data centre resources, Edge computing can provide a more energy-efficient approach to AI applications, particularly those necessitating real-time responses. This article will discuss the current state of developments in Edge AI and Edge computing and highlight specific challenges such as power consumption, battery life constraints, and maintaining performance and accuracy while prioritising energy efficiency. Furthermore, the article will emphasise the critical role of Edge computing in green AI success and explore possible resolutions to overcome the identified challenges.

Introduction

Green AI has emerged as a critical area of research and development in recent years, driven by the rapid growth of artificial intelligence (AI) and its increasing environmental impact. As AI systems become more complex and resource-intensive, their energy consumption and carbon footprints have become a growing concern. Green AI aims to address these concerns through the development of sustainable and energy-efficient AI systems that minimise environmental impact without compromising performance and accuracy. Green AI is not just about reducing the environmental impact of AI systems; it's also about making AI systems more efficient, effective, and cost-effective. The importance of green AI cannot be overstated, as it has the potential to revolutionise the way AI is designed, implemented, and deployed, and to ensure that the benefits of AI can be harnessed in an environmentally responsible manner.

The increasing demand for AI applications across multiple domains, such as IoT devices and smart systems, has led to the need for localised intelligence that can process data and make decisions in real time. This is where edge computing comes into play. Edge computing is a distributed computing paradigm that brings computation and data storage capabilities closer to the sources of data and reduces the need for data to be transmitted to centralised data centres for processing.



In contrast to the cloud computing paradigm, where collected data is fully transmitted to a central server before being analysed and used, edge computing allows for these tasks to be completed on or close to the devices. Decentralised computing itself is not a new paradigm and was prevalent before the advent of the centralised cloud computing paradigm. But it has become more important as data volumes, traffic, analysis, and storage requirements have increased drastically, while bandwidth as well as security remain limiting factors. By processing AI tasks on local devices, and minimising the reliance on data centre resources, Edge computing enables a more energy-efficient approach to AI applications, particularly those necessitating real-time responses. This makes Edge computing a promising solution for achieving green AI, as it can help reduce energy consumption and enable real-time data processing while maintaining the performance and accuracy of AI systems.

In this article, we will explore the potential of Edge computing in achieving green AI, discuss the challenges and complications faced in harnessing Edge computing for green AI, and present possible solutions to overcome these complications. By emphasising the critical role of Edge computing in green AI success, we aim to underline the significance of sustainable and energy-efficient AI systems that leverage localised intelligence for a more sustainable future.

The role of Edge computing in green AI

Edge computing plays a crucial role in green AI by facilitating the creation of eco-friendly and energy-efficient AI systems. By bringing computation and data storage closer to the data sources, edge computing minimises the demand for resource-intensive computing infrastructure, resulting in a substantial decrease in energy usage and carbon emissions. As AI becomes increasingly integrated into our daily lives, it is vital to develop and deploy AI systems in a manner that minimises their environmental impact.

Furthermore, edge computing allows AI systems to function in real-time, which is essential for applications such as self-driving cars, healthcare systems, and industrial automation. Real-time data processing enables AI systems to make quicker decisions, leading to enhanced efficiency and effectiveness. The integration of AI into edge computing boosts productivity and innovation across a wide range of sectors.

To illustrate the practical application of edge computing in achieving green AI, NCS has successfully deployed several projects that leverage localised intelligence. For instance, NCS implemented edge cameras in a patient safety project, which significantly enhanced real-time monitoring and emergency responsiveness while reducing data transmission requirements. Similarly, in another initiative, edge cameras mounted on vehicles provided detailed analytics on traffic patterns directly at the source, which allowed for the optimisation of traffic management and a reduction in carbon emissions. Furthermore, NCS has also developed navigation systems for robots using edge computing, allowing for efficient, real-time decision-making in dynamic environments. These projects not only underscore the potential of edge computing in enhancing performance and sustainability, but also demonstrate the tangible benefits of localised processing across diverse applications.

Edge computing also promotes decentralisation in AI systems to reduce the dependence on large data centres. Decentralised AI systems can lower the risk of downtime and enhance the reliability and availability of AI systems. This decentralised approach contributes to the overall sustainability of AI systems by decreasing the energy consumption and environmental impact associated with large-scale data centres.

Green AI has the potential to benefit numerous industries, including healthcare, manufacturing, transportation, and agriculture. It can help lower energy costs, reduce carbon emissions, and encourage sustainable practices. Some examples of how edge computing can contribute to sustainability are presented in Table 1.

No.	Application area	Edge computing contribution to sustainability
1	Energy consumption	Minimises the need to transmit data over long distances to centralised data centres or cloud servers, reducing energy consumption and associated greenhouse gas emissions.
2	Resource usage	Enables real-time monitoring and analysis of data, allowing for the optimisation of resource usage (such as energy, water, and raw materials) decreasing waste, and lessening the environmental impact of resource-intensive industries.
3	Renewable energy	Facilitates the integration of renewable energy sources into the power grid by enabling real-time monitoring and control of energy consumption to enhance the efficiency and reliability of renewable energy systems and reduce dependence on fossil fuels.
4	Sustainable agriculture	Assists in optimising agricultural production by providing real-time data on soil conditions, crop growth, and weather patterns to help farmers optimise crop yields, minimise waste, and improve the sustainability of agricultural practices.
5	Transportation	Helps optimise traffic flow and reduce congestion by minimising fuel consumption and associated greenhouse gas emissions, and enables real-time monitoring of vehicle emissions, allowing for better management of pollution levels.
6	Building sustainability	Improves energy management in buildings by providing real-time data on energy usage and enabling the optimisation of energy consumption, decreasing energy waste and enhancing the sustainability of buildings.
7	Smart grid management	Contributes to the development of smart grids by enabling real-time monitoring, analysis, for control of energy distribution, optimising energy consumption, reducing peak demand, and facilitating the integration of renewable energy sources.
8	Waste management	Supports waste management systems by providing real-time data on waste generation, collection, and disposal for optimising waste collection routes, reducing fuel consumption, and promoting recycling and waste reduction initiatives.
9	Water management	Helps improve water management by enabling real-time monitoring and control of water usage, distribution, and quality for optimising water consumption, reducing water waste, and ensuring the efficient use of water resources.
10	Wildlife monitoring	Supports wildlife monitoring and conservation efforts by providing real-time data on animal populations, habitats, and movements for optimising conservation strategies, reducing human-wildlife conflicts, and promoting the sustainable use of natural resources.

Table 1. Edge Computing Applications and Their Impact on Sustainability

Table 1 provides a clear and concise overview of the various application areas and how edge computing contributes to sustainability in each of them. The use of edge computing in green AI can contribute to a more sustainable future by reducing energy consumption, optimising resource usage, and promoting eco-friendly practices across different industries.

An essential aspect of green AI in edge computing is its connection to green IoT (G-IoT)¹⁻². The concept of green IoT is defined as energy-efficient methods (for both hardware and software) employed by IoT technologies to either facilitate the reduction of the greenhouse effect in existing applications and services, or to minimise the greenhouse effect's impact on the IoT ecosystem itself³. green IoT and edge AI adoption supports digital circular economy concepts by organisations and society in two main ways. First, by incorporating an open green IoT architecture⁴ where green IoT devices have circularity enabling features (e.g., end-to-end cybersecurity, privacy, interoperability, and energy harvesting capabilities). These devices can work in tandem with edge AI systems to create a more sustainable and efficient ecosystem. Second, by having a network of Edge-AI green IoT connected devices that provides fast smart services and real-time valuable information to the different stakeholders (e.g., designers, end users, suppliers, manufacturers, and investors). Thus, supply chain visibility and transparency of the product, the production system, and the whole business is ensured. Moreover, stakeholders can rely on accurate, real-time information to make the right decisions at the right time to use resources effectively, to improve the efficiency of processes, and to reduce waste. Furthermore, asset monitoring and predictive maintenance can increase product lifetimes. Figure 1 provides an overall view of the main areas impacted by the combined use of green IoT and Edge-AI in the context of green AI⁵.



Figure 1. Edge-AI green IoT main areas and their digital circular life cycle⁵

Challenges and complications in implementing green AI in Edge computing

While Edge computing holds great potential for achieving green AI, there are several challenges and complications that need to be addressed to harness its full potential. In this section, we will discuss these challenges and explore the complexities involved in implementing green AI in Edge computing.

1. **Energy challenges:** One of the primary challenges in Edge AI is managing power consumption and battery life constraints. Edge devices often have limited power resources, and optimising energy consumption is crucial for ensuring the sustainability of AI systems. Balancing energy efficiency with performance and accuracy is a critical aspect of green AI in edge computing. This involves designing energy-aware algorithms and models that can adapt to available power resources and dynamically adjust performance to maintain energy efficiency.
2. **Limited resources:** Edge devices typically have limited computational resources, such as processing power, memory, and storage. Efficient management and allocation of resources, without compromising local intelligence, is essential for implementing green AI on edge devices. This requires the development of lightweight AI models and algorithms that can efficiently utilise available resources while maintaining high performance and accuracy. Techniques such as model compression, pruning, and quantisation can help reduce the resource requirements of AI models to make them more suitable for deployment on edge devices.
3. **Performance and accuracy:** Ensuring that edge computing solutions maintain performance and accuracy while prioritising energy efficiency can be challenging. Developing energy-efficient algorithms and models that provide optimal performance without sacrificing sustainability goals is crucial for green AI success. This may involve exploring novel techniques, such as federated learning and distributed AI, which can help optimise AI performance on edge devices while minimising energy consumption. These approaches can enable edge devices to collaboratively learn and share knowledge, reducing the need for resource-intensive centralised data centres and promoting energy efficiency.
4. **Collaboration and integration:** The need for collaboration between hardware and software developers to optimise energy consumption in edge devices is another challenge. Achieving green AI in edge computing requires a coordinated effort between AI researchers, hardware manufacturers, and software developers to create sustainable AI solutions tailored for edge computing and localised intelligence. This collaboration can lead to the development of energy-efficient hardware and software solutions that optimise energy consumption in edge devices without compromising performance and accuracy. For example, hardware manufacturers can develop low-power processors and memory specifically designed for green AI applications, while software developers can create energy-aware algorithms and models that can efficiently utilise these hardware resources.
5. **Scalability and adaptability:** Edge computing solutions for green AI must be scalable and adaptable to accommodate the growing number of AI applications and the increasing complexity of AI systems. Developing scalable and adaptable solutions that can handle the increasing demands of AI applications while maintaining energy efficiency is a significant challenge. This may involve designing modular and flexible AI architectures that can easily scale and adapt to different edge devices and application requirements. Additionally, edge computing solutions should be capable of handling the dynamic nature of AI workloads, which may require real-time adaptation and reconfiguration to maintain energy efficiency and performance.

Solutions and strategies for green AI in Edge computing

To overcome the challenges and complications in implementing green AI in Edge computing, a number of solutions and strategies can be employed. In this section, we will discuss these potential solutions and explore how they can contribute to the successful implementation of green AI in Edge computing.

1. **Energy-efficient algorithms and models:** Developing energy-efficient algorithms and models for Edge AI is crucial for leveraging localised intelligence for sustainability. Researchers can explore novel techniques, such as model compression, pruning, and quantisation, to optimise AI performance on edge devices while minimising energy consumption. Additionally, adaptive algorithms that can dynamically adjust their performance based on available power resources can help balance energy efficiency with performance and accuracy.
2. **Hardware optimisations:** Implementing hardware optimisations for energy conservation in edge devices, such as low power processors, memory, and storage, specifically designed for green AI applications, can significantly reduce power consumption. Additionally, energy harvesting technologies can be integrated into edge devices to extend battery life and improve overall energy efficiency. These technologies can convert ambient energy sources, such as solar, thermal, or kinetic energy into electrical power to provide a sustainable and continuous power supply for edge devices.
3. **Collaborative efforts:** Encouraging collaboration between AI researchers, hardware manufacturers, and software developers is essential for creating sustainable AI solutions tailored for edge computing and localised intelligence. Joint efforts can lead to the development of energy-efficient hardware and software solutions that optimise energy consumption in edge devices without compromising performance and accuracy. Establishing interdisciplinary teams and fostering open communication channels can help facilitate the exchange of ideas and the development of innovative solutions for green AI in edge computing.
4. **Raising awareness:** Raising awareness about the importance of green AI and the potential of edge computing for achieving sustainability goals through localised intelligence is crucial. Promoting the benefits of green AI and edge computing to stakeholders may encourage them to invest in research and development efforts to create more sustainable AI systems. This can be achieved through educational initiatives, conferences, workshops, and publications that highlight the significance of green AI and the role of edge computing in achieving sustainability goals.
5. **Novel techniques and approaches:** Exploring novel techniques, such as federated learning and distributed AI, can help optimise AI performance on edge devices while minimising energy consumption. These approaches can enable edge devices to collaboratively learn and share knowledge and reduce the need for resource-intensive centralised data centres. By investigating and adopting innovative techniques, researchers and developers can contribute to the advancement of green AI in edge computing and help overcome the challenges associated with limited resources, performance, and accuracy.

Conclusion

Green AI in Edge computing has the potential to revolutionise the way AI systems are designed, implemented, and deployed and ensure that the benefits of AI can be harnessed in an environmentally responsible manner. Leveraging localised intelligence and edge computing with green AI will contribute to a more sustainable future by reducing energy consumption, optimising resource usage, and promoting eco-friendly practices across different industries.

The integration of green AI and Edge computing can lead to significant advancements in various application areas, such as energy management, transportation, agriculture, and smart cities. By adopting energy-efficient algorithms, hardware optimisations, and novel techniques, researchers and developers can create AI systems that are not only powerful and accurate but also environmentally sustainable. This shift towards sustainable AI systems will have far-reaching implications for the global economy, the environment, and society.

Furthermore, fostering collaboration between AI researchers, hardware manufacturers, and software developers is essential for driving innovation in green AI and Edge computing. By working together, these stakeholders can develop comprehensive solutions that address the challenges and complexities associated with implementing green AI in Edge computing. This collaborative approach can accelerate the development and deployment of sustainable AI systems and ultimately contribute to a greener and more energy-efficient future.

Raising awareness about the importance of green AI, and the potential of Edge computing in achieving sustainability goals, is also crucial. By promoting the benefits of green AI and Edge computing, stakeholders may be encouraged to invest in research and development efforts to create more sustainable AI systems. This can be achieved through educational initiatives, conferences, workshops, and publications that highlight the significance of green AI and the role of Edge computing in achieving sustainability goals.

In conclusion, the successful implementation of green AI in Edge computing can pave the way for a more sustainable and energy-efficient AI landscape. By emphasising the critical role of Edge computing in green AI success, we underline the significance of sustainable and energy-efficient AI systems that leverage localised intelligence for a more sustainable future. As AI continues to play an increasingly prominent role in our daily lives, it is imperative that we prioritise the development and deployment of environmentally responsible AI systems to ensure a greener and more sustainable future for all.

References

1. Arshad, R.; Zahoor, S.; Shah, M.A.; Wahid, A.; Yu, H. Green IoT: An Investigation on Energy Saving Practices for 2020 and Beyond. *IEEE Access* 2017, 5, 15667–15681.
2. Albreem, M.A.; Sheikh, A.M.; Alsharif, M.H.; Jusoh, M.; Mohd Yasin, M.N. Green Internet of Things (GIoT): Applications, Practices, Awareness, and Challenges. *IEEE Access* 2021, 9, 38833–38858.
3. Zhu, C.; Leung, V.C.M.; Shu, L.; Ngai, E.C.-H. Green Internet of Things for Smart World. *IEEE Access* 2015, 3, 2151–2162.
4. Askoxylakis, I. A Framework for Pairing Circular Economy and the Internet of Things. In *Proceedings of the 2018 IEEE International Conference on Communications (ICC)*, Kansas City, MO, USA, 20–24 May 2018; pp. 1–6.
5. Fraga-Lamas, P.; Lopes, S.I.; Fernández-Caramés, T.M. Green IoT and Edge AI as Key Technological Enablers for a Sustainable Digital Transition towards a Smart Circular Economy: An Industry 5.0 Use Case. *Sensors* 2021, 21, 5745. <https://doi.org/10.3390/s21175745>

