Final Report Submission

Introduction

Artificial Intelligence (AI) systems depend fundamentally on datacenter infrastructure, making their environmental impact inseparable from the energy demands concentrated in these facilities. Across the AI lifecycle — hardware manufacturing, model training, and operational deployment — energy consumption emerges as the central driver of environmental pressure. Hardware production embeds significant upstream energy costs; large-scale training requires prolonged high-intensity computation; and operational deployment generates continuous demand through inference workloads. Although studies quantify certain elements of these phases, the available data remain fragmented, particularly because companies seldom disclose complete energy or carbon metrics associated with their datacenter operations. This lack of transparency limits the ability to assess AI's full environmental footprint from a lifecycle perspective.

Grounded in sustainable technological development, the study interprets AI systems as socio-technical infrastructures whose environmental performance depends on lifecycle decisions. Energy metrics therefore serve not only as indicators of efficiency but also as proxies for broader sustainability outcomes, linking design choices to measurable environmental effects. The project adopts the perspective of Green AI, which argues that AI research must prioritize energy efficiency, transparency, and environmental responsibility rather than focusing solely on maximizing accuracy. As described in the literature, Green AI emerged in response to the escalating environmental footprint of large-scale AI systems, including rising electricity consumption, water usage, and carbon emissions from datacenters. It emphasizes metrics such as PUE (Power Usage Effectiveness), carbon intensity of energy sources, number of parameters, and inference emissions, proposing that sustainability be treated as a core performance dimension (Ferraz, 2024). The adoption of Green AI is therefore central to this study, providing both the conceptual justification and methodological rationale for evaluating AI initiatives through a lifecycle and resource-efficiency lens.

The guiding research question for this module asks which initiatives most effectively reduce the energy consumption of AI systems across the three targeted phases and how these efforts can be characterized as best practices for sustainable technological development. Rather than reproducing implementation details, the analysis focuses on documented, data-center–based interventions that lower operational energy use. These initiatives include industry energy efficiency programs, experimental datacenter infrastructure designs, carbon and energy monitoring tools, and large-scale optimization projects aligned with the challenges identified in recent Green AI research. The objective is to synthesize these findings into a coherent assessment and select the most suitable initiative as the case study for Module 15.

Methodology

This study uses an exploratory and descriptive qualitative design to assess initiatives that reduce AI-related energy consumption in datacenters. The analysis focuses exclusively on three lifecycle phases with the highest energy impact: hardware manufacturing, model training, and operational deployment. The methodological approach follows Yin's guidelines for exploratory case-study research (Chapters 1–4), emphasizing clarity of purpose, defined units of analysis, and consistent procedures.

Research Design

The project is structured as a multi-initiative exploratory case study. While Module 15 will develop a single in-depth case (Open Compute Project), Module 14 compares several initiatives to identify best practices and refine the analytical approach. This follows Yin's recommendation that exploratory studies help shape propositions and confirm the unit of analysis before the final case is selected.

The unit of analysis is any initiative that provides verifiable evidence of reducing datacenter energy use in at least one of the three lifecycle phases. Initiatives are evaluated using four predefined criteria: embodied energy, training energy, inference energy, and optimization potential.

Case-Study Protocol

A formal protocol guides all stages of data collection and analysis to ensure consistency and reliability:

1. Purpose
   Assess how documented initiatives reduce energy demand in datacenter hardware, training workloads, or deployment operations.

2. Case Questions
   How does the initiative affect embodied energy?
   How does it influence training energy consumption?
   How does it reduce inference energy per query?
   What optimization or scaling mechanisms does it provide?

3. Sources of Evidence
   Publicly available technical documentation, sustainability metrics, academic studies, and independent reporting. Evidence must include traceable data and, when possible, corroboration from more than one source.

4. Analytic Strategy
   Apply the four criteria to each initiative and compare results across cases using pattern matching and cross-case synthesis, as outlined by Yin (Chapter 5).

5. Reliability Procedures
   Maintain a structured evidence log and preserve a clear chain of evidence linking sources to extracted data and final findings.

Pilot Test

The pilot test employed Project Natick — a sealed subsea datacenter module deployed off the coast of Scotland in 2018 — to assess whether the case-study protocol could be meaningfully applied to an initiative focused on datacenter energy consumption. The module comprised a 40-foot long pressure vessel containing 12 racks with 864 servers and approximately 27.6 petabytes of storage. The datacenter was powered via a cable connected to on-shore renewable power sources, and submerged at a depth of 117 feet in the Northern Isles, where it remained for two years without physical maintenance or human intervention (Phase 2 retrieval in 2020).

The case-study protocol was applied fully: the unit of analysis was defined as the "underwater datacenter initiative," and data were gathered from publicly available technical documentation, project reports, and external media coverage. Each of the four lifecycle-based criteria — embodied energy, training energy, inference/deployment energy, and optimization potential — was mapped against available evidence. For quantitative claims such as power use or failure rates, independent or third-party reports were sought when possible, in accordance with the protocol's corroboration requirements.

The pilot yielded useful data for some criteria, while revealing limitations for others. In terms of deployment/operational energy, the underwater cooling system resulted in a power usage effectiveness (PUE) of approximately 1.07, indicating that cooling overhead was minimal and most of the supplied power was used by the servers. This outcome demonstrates that operational energy and cooling demand — often major contributors to datacenter energy consumption — can be substantially reduced under certain environmental conditions.

Concerning reliability and optimization potential, the datacenter exhibited a markedly low hardware-failure rate. Over 25 months of unattended operation, just six out of 855 servers failed, whereas a land-based control group of 135 servers experienced a higher failure rate. This suggests that factors such as a sealed nitrogen atmosphere (reducing corrosion risk), stable temperature, and absence of human interference may contribute to longer hardware lifespan and lower maintenance needs, which in turn can reduce lifecycle resource and energy costs.

However, significant evidence gaps emerged regarding embodied energy and training-phase energy. Public documentation did not include detailed data on the energy required for manufacturing, material sourcing, or transport. Moreover, because the underwater datacenter was not explicitly used for large-scale AI model training, no training-specific energy consumption data are available. This limitation illustrates a broader challenge: not all lifecycle phases are equally documented across initiatives, especially when projects do not cover the full AI lifecycle.

Thus, the pilot demonstrated both the strengths and limits of the protocol. For initiatives like Project Natick, it is possible to derive robust insights into deployment-phase energy and infrastructure-related optimization. Yet other phases, such as embodied energy and training workloads, may remain unassessable depending on the available documentation.

Given these results, the protocol was adjusted: for each initiative examined in the full comparative study, the evaluation criteria will be applied selectively, depending on evidence availability, and any data gaps will be explicitly noted. This ensures that findings remain evidence-based, transparent, and adequately qualified.

It must be acknowledged that most of the available evidence regarding Project Natick comes from the organization behind it — hence, the lack of extensive third-party audits or independent verification constrains the ability to confirm some claims, such as reliability or energy efficiency outcomes. Accordingly, results derived from Project Natick should not be generalized as definitive proof of the viability or sustainability of underwater datacenter approaches. Rather, these results primarily validate the protocol's applicability and highlight where evidence is sufficient or lacking.

In sum, the Project Natick pilot validated the structure of the case-study protocol and confirmed that when detailed documentation exists, lifecycle-based energy criteria can be applied in a comparative context. Concurrently, the pilot clarified the necessity of adapting criteria application to each initiative's scope and carefully acknowledging missing data — practices that will guide the full Module 15 analysis.

Triangulation Results

Across all initiatives examined, three major categories emerged:

(1) infrastructure-centered initiatives (Open Compute Project, Google Data Centers, Microsoft Project Natick, Project Suncatcher);

(2) grid and energy-management programs (EPRI DCFlex);

(3) measurement and transparency tools (CodeCarbon, EcoLogits, Eco2AI).

Each category contributes different forms of sustainability impact, allowing for complementary comparison.

Open Compute Project (OCP)

By standardizing server “blocks” and data-center modules, OCP enables reuse, modular upgrades, and cross-vendor interoperability, reducing the embodied energy tied to frequent hardware replacement or vendor-locked systems. Beyond hardware efficiency, OCP’s “Heat Reuse” subproject transforms waste heat — normally a cooling liability — into a usable resource, turning a cost into an asset.

Empirical data underscores the benefits of OCP-aligned modular design: a reference modular system showed a 27% lower carbon footprint compared to a traditional, non-modular setup. OCP-Ready facilities — data centers designed for OCP hardware — optimize power distribution, airflow, and rack density, collectively reducing operational energy waste, cooling overhead, and improving total energy efficiency relative to conventional data centers.

Thanks to its open-source, modular design and broad community adoption, OCP’s sustainability practices are highly replicable and scalable, supporting large scale AI or cloud infrastructure while enhancing lifecycle efficiency through reduced embodied energy, streamlined deployment, and system-wide optimization.

However, OCP primarily impacts infrastructure and hardware; it lowers embodied and operational energy but does not directly affect training or inference energy or algorithmic efficiency. Its value lies in enabling sustainable infrastructure: reduced carbon footprint, lower operational overhead, modular upgrade paths, and systemic reusability — key factors for long-term, large-scale deployment and lifecycle sustainability.

Microsoft Project Natick

Microsoft’s Project Natick represents an unconventional experimental approach: a sealed subsea data center leveraging the ocean’s ambient cold for cooling and renewable grid energy for operation.

Phase II, deployed off the Orkney Islands in Scotland, reportedly achieved a Power Usage Effectiveness (PUE) of approximately 1.07 — significantly better than typical land-based data centers, reflecting highly efficient energy use. Seawater cooling also eliminated the need for traditional evaporative or chiller-based systems, bringing Water Usage Effectiveness (WUE) close to zero and removing freshwater consumption for cooling entirely.

The underwater environment further enhanced reliability: Microsoft reported far fewer server failures compared to equivalent land-based deployments, suggesting stability benefits and reduced risk from environmental variability or human intervention. These results indicate exceptional operational efficiency, resource conservation, and reliability — key aspects of deployment and resource usage efficiency under lifecycle criteria.

However, Natick's experimental nature and discontinuation as of mid-2024 limit its external validity and replicability. Available data focus on general data-center operations rather than AI-specific workloads, so its impact on training or inference energy for large-scale AI remains unclear.

In summary, Natick demonstrates the potential of extreme cooling efficiency and resource savings in a niche context, but it remains primarily a proof of engineering concept rather than a practical model for widespread adoption or AI infrastructure under lifecycle sustainability criteria.

Google Project Suncatcher

Project Suncatcher envisions using space-based solar arrays to power AI workloads, addressing the energy-supply side of AI sustainability rather than hardware or software infrastructure. Within the lifecycle framework, it targets long-term energy supply sustainability and future-oriented optimization potential.

However, Suncatcher remains conceptual, with minimal public data on embodied energy, deployment efficiency, or operational metrics such as inference energy. Without detailed information on system architecture, lifecycle analysis, or performance, it cannot be reliably evaluated against the four lifecycle criteria.

Suncatcher's value, in this context, lies primarily in its vision: expanding the scope of sustainable innovation by incorporating energy-supply considerations into the AI lifecycle. From a case-study perspective, its contribution is speculative, serving to illustrate potential future pathways rather than provide concrete, actionable insights.

EPRI DCFlex

The EPRI DCFlex program (and similar grid-level demand-response initiatives) shifts focus from individual data-center hardware to the temporal coordination of compute demand with renewable energy availability. By aligning workloads — such as AI training jobs — with periods of high renewable generation or lower carbon grid supply, data centers can reduce the carbon intensity of energy consumed.

This approach enhances sustainability not by lowering total energy use, but by shifting when energy is consumed, reducing reliance on fossil-fuel–based peak generation and supporting better integration of renewable resources.

However, DCFlex does not inherently improve per-training or per-inference energy efficiency, nor does it affect embodied energy. Its impact is limited under lifecycle criteria, addressing only carbon intensity and timing, not infrastructure efficiency or absolute energy per computational unit.

In short, DCFlex improves the environmental profile of energy use timing, complementing infrastructure-focused strategies like OCP, but it is insufficient alone to reduce the total energy footprint under a full lifecycle perspective.

CodeCarbon, EcoLogits, and Eco2AI

These tools provide the most transparent sustainability methodology among all initiatives reviewed.

CodeCarbon automatically tracks CO_2 emissions by integrating hardware power measurements with regional carbon-intensity data.

EcoLogits and Eco2AI extend this approach to monitoring at inference and pipeline level, increasing visibility into energy hotspots across the AI lifecycle.

While these tools do not directly *reduce* energy consumption, they significantly strengthen methodological rigor by improving traceability, standardizing emissions reporting, and enabling comparability across models and data centers.

These tools enhance *construct validity* for sustainability assessments and support the development of reproducible reporting standards but have limited impact on material energy savings.

Benchmark and External Sources (Statista, IEA, LCA Literature)

External references provide essential contextualization for evaluating the initiatives:

- Statista's global PUE averages (≈1.2–1.4) contextualize Natick's performance.
- IEA's Electricity 2024 projections highlight increasing pressure on energy grids as AI demand grows, emphasizing the relevance of initiatives like DCFlex.
- Lifecycle studies (LCAs) reinforce the importance of embodied energy, aligning with OCP's hardware-focused sustainability.

These sources function as external reference baselines, enabling consistent comparison across initiatives.

Cross-Initiative Patterns

Applying the cross-case synthesis protocol across all evidence and initiatives, four major patterns emerge:

First, hardware and infrastructure interventions deliver the most material energy and carbon reductions when properly implemented. OCP's modular, open-standard approach and Natick's subsea cooling experiment illustrate that infrastructure-level design — not just software or scheduling tweaks — can substantially influence embodied carbon, operational overhead, and resource efficiency.

Second, transparency and measurement tools (CodeCarbon, EcoLogits, Eco2AI) improve methodological rigor without directly reducing energy demand. Their impact lies in enabling consistent tracking, comparability, accountability, and long-term optimization.

Third, system-level or grid-level initiatives (like DCFlex) address carbon intensity — the temporal and source-based environmental footprint of compute — but do not alter per-workload energy consumption or embodied carbon. They are thus complementary but insufficient alone under a lifecycle-sustainability perspective.

Fourth, replicability and scalability favor open standards and modular design. Initiatives with strong open protocols (OCP) or standardized design conventions offer greater external validity and broad applicability. In contrast, experimental or conceptual initiatives (Natick, Suncatcher) — while valuable as proofs of concept — face limitations in generalizability, widespread deployment, or empirical grounding.

Discussion

Synthesizing across initiatives and criteria, the evidence indicates that the Open Compute Project (OCP) provides the strongest foundation for sustainable, scalable

AI infrastructure under a lifecycle framework. Its modular design reduces embodied carbon, enables reuse and circular hardware practices, and supports energy-efficient, high-density deployments. Public documentation and broad adoption by multiple vendors and operators make OCP replicable at scale, aligning ecological sustainability with practical deployment needs.

Other initiatives play complementary roles. Measurement tools (e.g., CodeCarbon) enhance transparency and accountability, which are essential for verifying sustainability claims. Grid-coordination programs like DCFlex improve carbon intensity by aligning compute demand with renewable supply, but they do not directly reduce energy consumption or improve infrastructure efficiency. Experimental projects like Natick offer valuable lessons in cooling and resource efficiency, yet limited replicability and discontinuation restrict their broader applicability. Visionary concepts like Suncatcher illustrate long-term potential but currently lack empirical validation.

The triangulation results reveal that AI sustainability emerges from interdependent factors across hardware, infrastructure, energy supply, and measurement layers. Hardware and infrastructure interventions generate measurable lifecycle energy reductions: OCP addresses embodied and operational energy, while hyperscale providers like Google optimize airflow, thermal management, and cooling efficiency. Experimental approaches such as Natick reinforce the importance of physical infrastructure design as a key sustainability lever.

Energy-management and grid-alignment programs influence carbon intensity rather than total energy use. DCFlex, for example, shifts workloads to periods of renewable abundance, reducing reliance on fossil-fuel peak generation but not per-unit energy demand. Measurement tools improve comparability and methodological rigor, enabling reproducible reporting, but do not reduce consumption directly.

These distinctions highlight that sustainability outcomes arise through different causal pathways: hardware standards reduce embodied and operational energy, infrastructure optimization enhances deployment efficiency, grid alignment improves carbon intensity, and measurement tools strengthen transparency. Triangulation makes these mechanisms visible, demonstrating that no single initiative addresses the full spectrum of AI's environmental impacts.

Replicability and openness are critical for external validity. Natick achieved exceptional performance but lacked scalability; Google optimizes at hyperscale but publishes limited methodological details; software tools are transparent but hardware-independent. In contrast, OCP combines open documentation, growing adoption, and modular designs, making it a robust, verifiable, and scalable sustainability intervention.

Conclusion

Module 14 established the methodological and analytical foundation for assessing AI's environmental impacts through comparative, multi-initiative analysis. Guided by Yin's principles, the module reinforced construct validity via triangulation, strengthened reliability with transparent protocols, and clarified theoretical propositions guiding the final case study.

The analysis demonstrates that hardware and infrastructure interventions provide the most direct reductions in energy use, grid-alignment programs reduce carbon intensity, and measurement tools improve transparency and accountability. Experimental projects illustrate engineering possibilities, while system-level coordination initiatives like DCFlex show how demand timing can improve environmental outcomes. Among these, only OCP combines transparency, replicability, real-world adoption, and alignment with lifecycle criteria, making it the strongest candidate for Module 15.

Module 14 also served as a critical learning stage, deepening understanding of how sustainability outcomes depend on intertwined technological, infrastructural, and methodological choices. It highlighted data gaps, reinforced the importance of triangulated evidence, and demonstrated the value of a disciplined, replicable research approach.

These insights set the stage for Module 15, enabling a focused, in-depth case study of OCP. With validated criteria, tested analytical frameworks, and cross-case patterns identified, the next module will link lifecycle sustainability outcomes directly to concrete design practices within OCP.