Original Article

Multi-Objective Optimization of Thermal Management Systems for Artificial Intelligence Server Racks Using Exergy Analysis and Life-Cycle Assessment

Chinta Mani Tiwari [1] , Prashant K. Srivastava [2] , Shrishti Rastogi [1,]

1. *Department of Mathematics, Maharishi University of Information Technology (MUIT), Lucknow, IND*
2. *Department of Mechanical Engineering, Ambalika Institute of Management and Technology, Lucknow, IND*

Received: February 22, 2026 | Review began: April 05, 2026 | Review ended: June 25, 2026 | Published: July 20, 2026

Abstract

Artificial intelligence (AI) workloads have drastically increased heat generation in current data centers, thus efficient and environmentally friendly cooling systems have become a necessity. Traditional thermal management (TMS) system design methods only focus on optimizing one performance metric like energy efficiency and overlook the environmental impacts throughout the system life cycle. This paper develops a multi-objective optimization framework for an AI server rack TMS by simultaneously maximizing exergy efficiency and minimizing the life-cycle environmental impact. Through second-law thermodynamics and a life-cycle assessment, a physics-based mathematical model is created. The Non-Dominated Sorting Genetic Algorithm II is used to solve the optimization problem and identify the feasible optimum under the imposed thermodynamic, economic, and environmental constraints.

The optimization results yielded a feasible optimum with a scaled exergy-performance index of 329.34 and a life-cycle impact of 159881.57 normalized impact units. The exergy-performance value is not interpreted as a conventional percentage efficiency; rather, it represents the value of the adopted exergy objective function under the present mathematical formulation. In the present formulation, the knee-point solution coincided with this optimum, indicating that the best thermodynamic and environmental performance was achieved at the same operating condition. The proposed framework offers a structured preliminary model-based framework for sustainable thermal management design of AI server racks.

Categories: Sustainable Technologies, Thermal and Fluid Systems, Environmental and Sustainable Engineering

Keywords: thermal management system, exergy analysis, life-cycle assessment, multi-objective optimization, nsga-ii, ai server cooling

Introduction

Background and motivation

The explosive expansion of artificial intelligence (AI) and machine learning has brought about a complete change in the computing infrastructure of the modern world. High-performance computing (HPC) systems, as well as AI server racks, are now generating power densities at a level never seen before, often more than 30-50 kW per rack, which makes their thermal management and energy efficiency very challenging [1,2].

Therefore, the development of an efficient thermal management system (TMS) is of paramount importance to maintain the system's reliability, keep the performance at the highest level, and ensure the operation is environment-friendly.

Traditional cooling methods in data centers have mostly been based on air cooling, mainly because of its straightforwardness and their low cost initially. Nevertheless, air cooling is no longer able to meet the needs of AI-driven workloads due to the limited amount of heat it can transfer and the high level of exergy destruction resulting from large temperature gradients [3]. Hence, advanced cooling methods such as liquid cooling, direct-to-chip cooling, and hybrid TMSs are increasingly being highlighted.

Meanwhile, worldwide issues such as climate change and carbon emissions have brought even more attention to energy-efficient and eco-friendly engineering systems. It is estimated that data centers consume about 1-2% of the world's total electricity, and the cooling system is responsible for a large part of this energy consumption [4].

Therefore, it is critically important to develop TMSs capable of providing both improved thermodynamic performance and minimal environmental impact over the whole system life cycle.

Limitations of conventional energy-based design

Traditionally, most thermal system designs have been assessed using first-law (energy-based) efficiency indicators, such as the coefficient of performance or energy utilization effectiveness. These indicators measure how well energy is conserved but they do not reflect the quality of the energy or the irreversibility of the processes in real life [5]. Consider, for instance, two cooling systems that consume the same amount of power, but one generates significantly more entropy and irreversible losses than the other. This kind of difference cannot be identified with a purely energy-based analysis. This shortcoming of energy-based methods becomes quite significant in systems with very high temperature gradients like AI server cooling, where irreversibilities are the major concern. Exergy analysis, which is grounded on the second law of thermodynamics, has been suggested as a more thorough performance assessment method to address this drawback. Exergy represents how much maximum useful work can be extracted from a system relative to its surroundings, thus it inherently takes into account the irreversibility and degradation of energy [6]. Exergy-based approaches have been employed with success for refrigeration systems, heat exchangers, as well as data center cooling infrastructure [7].

Environmental sustainability and life-cycle perspective

Exergy analysis is a great tool for determining thermodynamic efficiency, but it does not really show the whole picture when it comes to environmental impacts of material extraction, manufacturing, transportation, operation, and disposal of TMSs. Those impacts can account for a substantial portion of the total impact, especially in the case of liquid cooling systems that use copper, aluminum, and high-power pumping devices during their operation.

Life-cycle assessment (LCA) is a standardized approach through which environmental impacts of the entire life cycle of a product or system can be measured [8]. Different research studies have shown that systems that have higher operational efficiency can have higher embodied environmental impacts due to material-intensive designs [6,9].

Consequently, a life-cycle approach is indispensable to prevent burden shifting, where environmental advantages in one phase are neutralized by increased impacts in another. Combining LCA with thermodynamic analysis allows for a holistic sustainability assessment.

Need for multi-objective optimization

It is a well-known fact that exergy efficiency and life-cycle environmental impact represent contrasting goals. For instance, increasing the size of heat exchangers or the flow rate of coolant might improve exergy efficiency but at the same time lead to an increase in material usage, pumping power, and embodied environmental impact. On the other hand, reducing material usage may degrade thermal performance and increase irreversibility. These competing effects cannot be addressed adequately through single-objective optimization methods. Instead, multi-objective optimization (MOO) is used to identify the best compromise solution between exergy efficiency and life-cycle environmental impact under the imposed design constraints [10]. Evolutionary algorithms have been shown to be particularly effective for solving complex, nonlinear, and multi-modal optimization problems in exergy-based systems [7]. Among these methods, the Non-dominated Sorting Genetic Algorithm II (NSGA-II) is one of the most widely used because of its computational efficiency, elitism, and ability to maintain solution diversity [11].

Research gap

An extensive literature review has been done and the following points can be considered as a research gap. The majority of literature related to data center and server cooling has concentrated on energy efficiency or power use effectiveness, neglecting exergy-based performance evaluation, The environmental impact assessment is usually limited to operational energy consumption only and the embodied environmental impacts are overlooked, Few research works integrate exergy analysis and LCA into optimization framework, There are almost no articles on the application of multi-objective evolutionary optimization to TMSs specifically designed for AI server racks, Exploration of how decision-making methodologies can be used to select an optimal design from pareto-optimal solutions is still to a large extent an open question. These gaps represent there is a demand for a holistic framework that can integrate thermodynamic efficiency, environmental sustainability, and design trade-offs altogether.

Objectives of the present study

The main goal of this study was to develop a MOO framework for the design of TMSs for AI server racks. Specifically, the study aimed to develop a thermodynamic model of the TMS using exergy analysis, model life-cycle environmental impacts by accounting for both embodied and operational energy effects, combine the thermodynamic and environmental models into a unified MOO problem, implement the NSGA-II evolutionary algorithm to solve the problem, analyze the trade-offs between exergy efficiency and life-cycle impact (LCI), and identify an optimal knee-point design for practical implementation.

Scope of the study

The research covered the following: Parameters of the TMS design directly connected to AI server racks, trade-offs between thermodynamics and the environment, the use of evolutionary MOO techniques, programming in Python, and using the pymoo library. While experimental validation is outside the limits of this specific numerical study, the developed framework is generic and it can be adapted to experimental or industrial-scale systems.

Materials And Methods

This part discusses a comprehensive, step-by-step mathematical modeling of TMS that was assumed for the research. The modeling framework aims at measuring the system performance referring to the exergy efficiency and the life-cycle environmental impact. It also considers the physical, operational, and economic limitations.

The developed model combines: thermodynamic heat transfer relations, exergy analysis based on the second law of thermodynamics, LCA principles, engineering constraints relevant to real-world thermal systems. The mathematical formulation serves as a base for the MOO process outlined in the previous section.

System description

The TMS that is being discussed in this case is a liquid-assisted, heat exchanger-based cooling system for extremely high heat flux electronic components such as AI servers or HPC units. System Components: The components of the system are as follows: heat exchanger (cold plate or liquid-cooled heat sink), coolant flow loop with pump, inlet and outlet temperature control, structural and material components, which result in embodied environmental impact. Figure *1* illustrates the system.

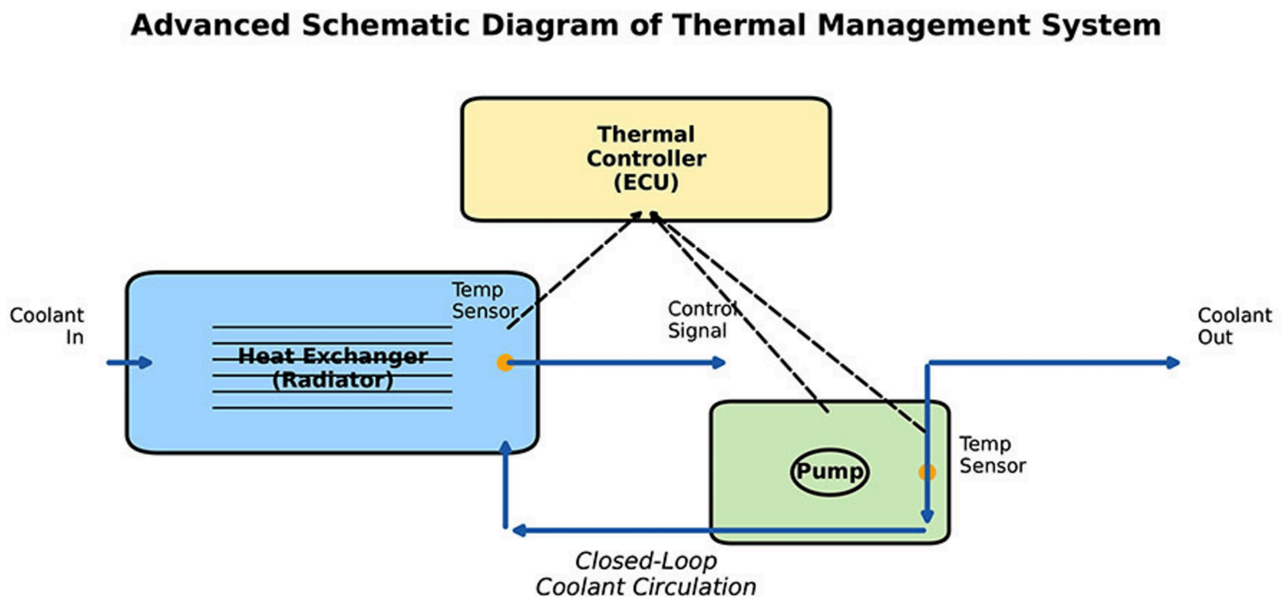

FIGURE 1: Schematic diagram of the thermal management system (TMS) showing heat exchanger, coolant flow, and pump

Modeling assumptions

To avoid complicating the mathematical formulation too much while at the same time staying physically realistic, the following assumptions are made: The system is considered to be under steady-state operation, The thermophysical properties of the coolant are considered to be constant, heat loss to the surrounding is not taken into account, only one-dimensional heat transfer is considered, the pump efficiency is an element of an empirical power correlation, the environmental reference temperature is assumed to be constant, and manufacturing impacts are linearly related to component size and material factor. Such assumptions are commonly adopted in thermal system modeling [5,12].

Design variables

The system is described by five continuous decision variables as shown in Table *1*. Together, these variables determine the system's thermal performance, exergy destruction, and environmental impact. The physical meaning and allowable range of the design variables listed in Table *1* define the optimization search space.

Variable	Description	Unit
A	Heat exchanger area	m^2
D	Pipe diameter	m
$\dot{m}$	Coolant mass flow rate	kg/s
T_{in}	Inlet coolant temperature	K
M_f	Material factor	–

TABLE 1: Design variables of the thermal management system

Heat transfer modeling

Outlet Temperature Estimation

The coolant outlet temperature is estimated experimentally by an empirical correlation for heat transfer:

$$T_{out} = T_{in} + \frac{Q_{load}}{\dot{m}c_p + \varepsilon}$$

In the present study, the server-rack thermal load is taken as $Q_{load} = 15\,\mathrm{kW}$, and ε is a small numerical constant introduced for stability.

Heat Transfer Rate

The heat transfer rate of the coolant is defined as:

$$\dot{Q} = \dot{m} \cdot c_p \cdot (T_{out} - T_{in})$$

where c_p is the coolant's specific heat capacity.

Pump power model

An empirical nonlinear model of pump power is:

$$W_{pump} = k \cdot \dot{m}^3$$

where k denotes a constant specific to the system that accounts for frictional losses and pump characteristics [13].

Exergy analysis

Reference Environment

The temperature of the reference environment is set at:

$$T_0 = 298\,\mathrm{K}$$

Exergy Associated with Heat Transfer

$$\dot{E}_Q = \dot{Q}\left(1 - \frac{T_0}{T_{out}}\right)$$

Exergy Efficiency Objective

Using the exergy associated with heat transfer, the exergy efficiency is written as

$$\eta_{ex} = \frac{\dot{Q}\left(1 - \frac{T_0}{T_{out}}\right)}{W_{pump} + \varepsilon}$$

It should be noted that the exergy-related quantity used in this study is treated as a performance index for optimization. Therefore, the numerical value reported from the optimization should not be interpreted as a conventional percentage efficiency. Instead, it represents the relative exergy-performance value obtained from the adopted objective-function formulation.

$$f_1(x) = -\eta_{ex}$$

Life-cycle impact modeling

The model of LCI, which is used in this work as a simplified parametric environmental indicator, is not a full-fledged ISO-compliant LCA. The system boundary covers embodied impact due to the size of the heat exchanger, the components related to the coolant flow, material factor, and the energy used during operation of the pumping system over the assumed service life. Detailed material inventory databases, emissions from manufacturing processes, transport, maintenance and end-of-life disposal are not provided. Thus, the reported LCI values are only normalized units of the impacts for comparative optimization in the current model.

Embodied Environmental Impact

$$E_{embodied} = \alpha A^{1.3} + \beta \dot{m}^{1.2} + \gamma$$

Operational Environmental Impact

$$E_{operational} = W_{pump} \times 24 \times 365 \times N$$

where N is the operational lifetime in years.

Total Life-Cycle Impact Objective

$$f_2(x) = (E_{embodied} \cdot M_f) + E_{operational}$$

Constraint formulation

Thermal Constraint

$$g_1(x): \quad T_{out} \leq T_{max}$$

Volume Constraint

$$g_2(x): \quad A \cdot D \leq V_{available}$$

Cost Constraint

$$g_3(x): \quad 2000A + 500W_{pump} \leq C_{budget}$$

Complete optimization problem statement

The MOO problem statement is given as:

$$\min\left[f_1(x), f_2(x)\right]$$

with the condition:

$$g_i(x) \leq 0, \quad i = 1, 2, 3$$

$$x_L \leq x \leq x_U$$

where x_L and x_U are the lower and upper bounds of design variables, respectively.

Optimization methodology

By using the mathematical model presented in the previous section, a nonlinear, constrained, MOO problem involving conflicting objectives, namely the maximization of exergy efficiency and the minimization of life-cycle environmental impact, was formulated. Because of the nonconvexity and nonlinear constraints of the problem, classical gradient-based optimization methods are not well suited for obtaining reliable compromise solutions. Therefore, this study employed an evolutionary multi-objective optimization (EMOO) approach based on the NSGA-II, followed by a knee point-based decision-making procedure to identify the most suitable design within the feasible objective space.

Nature of the Optimization Problem

The problem formulation depicts that the optimization problem has the following features: It is a multi-objective problem (two conflicting objectives), contains nonlinear objective functions, nonlinear inequality constraints, decision variables are continuous, and convexity of the problem is not guaranteed explicitly.

Population-based evolutionary algorithms are well suited to such problems because they can explore the feasible design space globally and identify suitable compromise solutions in a single run [10].

Multi-Objective Optimization Framework

Pareto Optimality: We say, in MOO, that a solution x_a dominates another solution x_b if:

$$\forall i: f_i(x_a) \leq f_i(x_b) \quad \text{and} \quad \exists j: f_j(x_a) < f_j(x_b)$$

A solution is pareto-optimal when no other feasible solution dominates it. The collection of all pareto-optimal solutions is known as the pareto front, which illustrates the trade-offs between objectives [14].

Selection of NSGA-II Algorithm

Rationale for Choosing NSGA-II: NSGA-II was picked for the following reasons: Efficient non-dominated sorting, explicit diversity preservation using crowding distance, elitism to hold on to the best solutions, proven convergence characteristics, and extensive use and benchmarking in thermal and energy system optimization. NSGA-II has been utilized extensively in exergy-based as well as sustainability-driven optimization research [15,16].

Overview of NSGA-II: NSGA-II works with a group of candidate solutions (population) which are evolved using genetic operators. The main sequence of steps is: generation of initial population, measurement of fitness, ranking based on dominance, calculation of crowding distance, selection, crossover and mutation, and elitist replacement. A diagram representation can be found in Figure 2.

How to cite this article:
Tiwari C, Srivastava P K, Rastogi S (July 20, 2026) Multi-Objective Optimization of Thermal Management Systems for Artificial Intelligence Server Racks Using Exergy Analysis and Life-Cycle Assessment. Cureus J Eng 3 : es44388-026-00174-8.
DOI https://doi.org/10.7759/s44388-026-00174-8

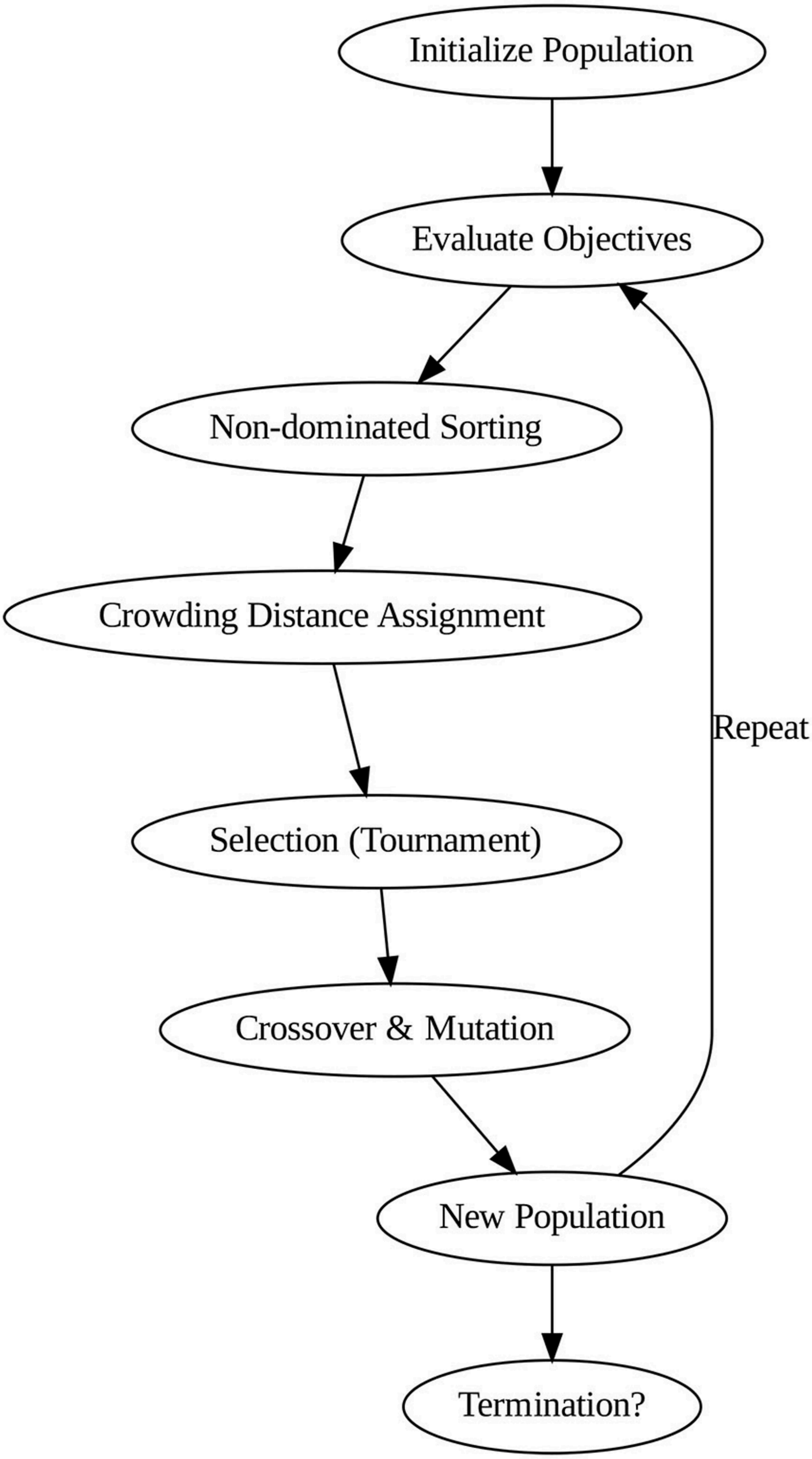
Initialize Population
Evaluate Objectives
Non-dominated Sorting
Crowding Distance Assignment
Selection (Tournament)
Crossover & Mutation
New Population
Termination?
Repeat

FIGURE 2: Flowchart of the NSGA-II optimization process

NSGA-II, Non-Dominated Sorting Genetic Algorithm II

Constraint Handling Strategy

Constraints are handled using a constraint-domination principle, which means: If a solution is feasible, then it dominates an infeasible one, among infeasible solutions, the smaller violation of constraints is preferred.

The total amount of constraint violation is obtained for a solution as follows:

$$CV(x) = \sum_{i=1}^{N_c} \max(0, g_i x)$$

This method gives the assurance that feasible areas are increasingly pointed out during the evolution process [11].

Genetic Operators

Crossover: Simulated Binary Crossover is used for recombination, which behaves like single-point crossover in continuous spaces. The offspring diversity is controlled by a distribution index [17].

Mutation: Polynomial mutation is applied to bring in diversity and prevent premature convergence. Mutation probability is set inversely proportional to the number of decision variables.

Algorithm Parameters

Table *2* presents a summary of the parameters of NSGA-II used in this research.

Parameter	Value
Population size	100
Number of generations	200
Crossover probability	0.9
Mutation probability	1/6
Random seed	1

TABLE 2: NSGA-II parameters

The parameters have been set based on the recommendations found in the literature and preliminary convergence tests [10,18].

Implementation Using pymoo

The optimization algorithm was implemented in the pymoo Python library, which is a flexible framework for EMOO [19]. The problem class extends pymoo. core. problem. Problem, and objective and constraint evaluations are defined in vectorized form.

Generation of Pareto Front

After running the NSGA-II algorithm, the feasible solution space is explored in terms of exergy efficiency and life-cycle environmental impact: exergy efficiency (to be maximized) and life-cycle environmental impact (to be minimized). The corresponding results in the objective space are visualized in Figure 3, from which the feasible optimum is identified.

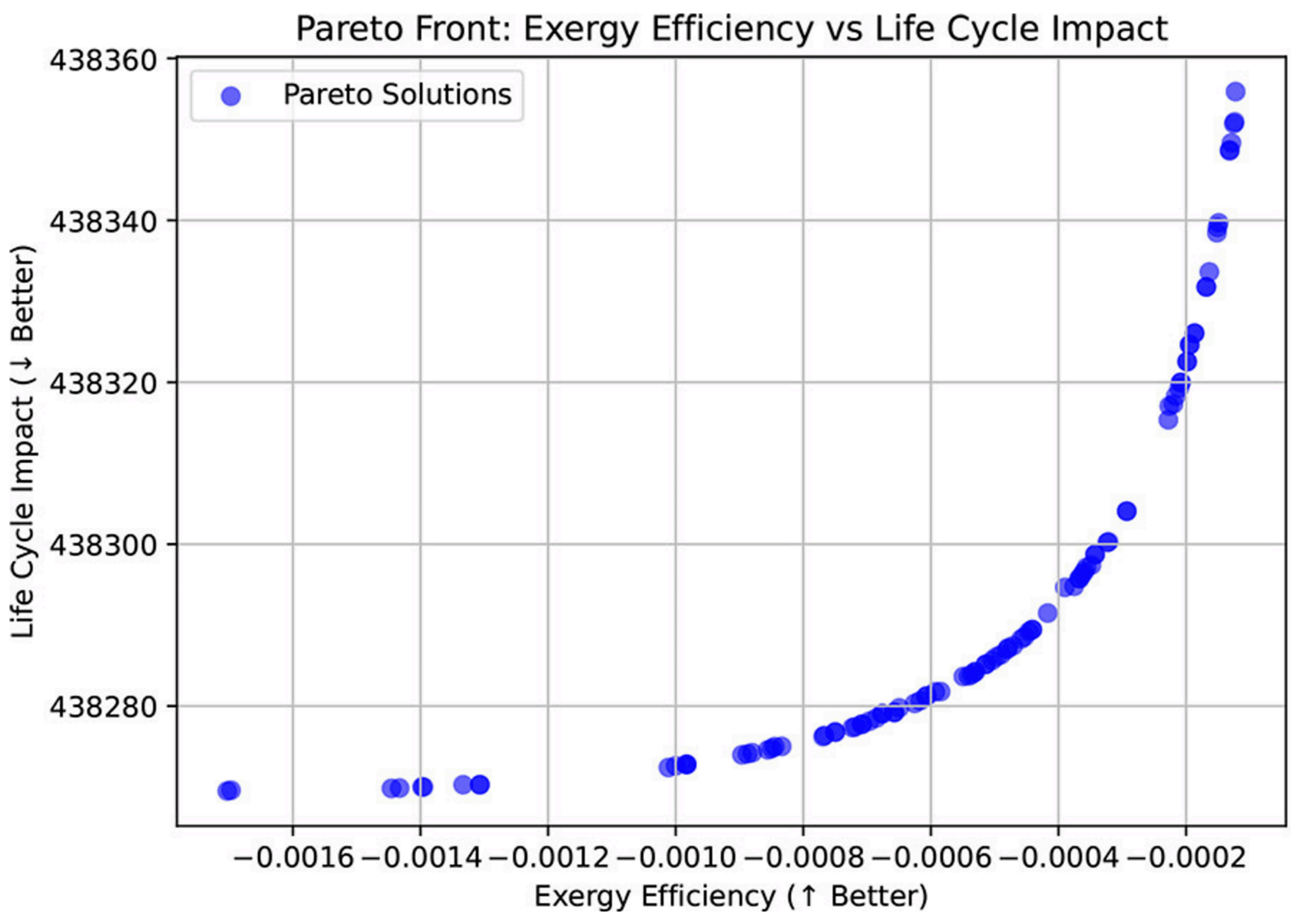

FIGURE 3: Pareto front showing trade-off between exergy efficiency and life-cycle impact

Need for Decision-Making

Although MOO provides insight into the interaction between conflicting objectives, engineering design ultimately requires the selection of a single operating point. Therefore, a multi-criteria decision-making procedure is needed to identify the most appropriate solution. Among the available approaches, the knee point method was adopted in this study because it is objective and physically interpretable.

Knee Point Identification Method

The knee point is a solution where one objective can only be slightly improved at the cost of the other heavily deteriorating. This point represents the best trade-off. Mathematically, the knee point is the one that is closest to the ideal point in normalized objective space [20].

Normalization of Objectives: The feasible pareto solutions are normalized as follows:

$$F_i^{norm} = \frac{F_i - F_i^{min}}{F_i^{max} - F_i^{min}}$$

Distance-Based Knee Selection: The Euclidean distance from the ideal point is determined as follows:

$$d_j = \sum_{i=1}^{N_o} \left(F_{ij}^{norm}\right)^2$$

The knee point corresponds to:

$$j^* = \arg\min d_j$$

This method is quite common in energy and sustainability optimization studies [11,21].

Selected Optimal Design

The knee point solution provides: high exergy efficiency, acceptable environmental impact, and zero constraint violation. The solution is regarded as the optimal design for real-world implementation and is examined thoroughly in Results and Discussion section.

Results And Discussion

This part of the paper illustrates and benchmarks the innovations in MOO for the TMS. The method developed in the previous section was enacted through NSGA-II to identify the pareto front of exergy efficiency and life-cycle environmental impact under physical, economic, and thermal constraints.

The discussion focuses on: numerical verification of the mathematical model, convergence behavior of the optimization algorithm, pareto front characteristics, feasible versus infeasible solution regions, knee point-based optimal design selection, and engineering and sustainability implications of the obtained results.

Numerical verification of the model

To ensure the reliability of the Python implementation, the numerical results were verified against the governing analytical heat-balance equation. In the present study, numerical verification was carried out by comparing the Python-computed heat-transfer rate with the corresponding analytical thermal load for a set of representative operating cases. For this purpose, multiple verification cases were generated by varying the imposed thermal load while keeping the remaining input parameters within the physically relevant operating range of the thermal management system. For each verification case, the heat-transfer rate predicted by the Python model was obtained from

$$\dot{Q} = \dot{m}c_p(T_{out} - T_{in})$$

The comparison is presented in Figure *4*, where the Python-computed values are plotted against the analytical values together with the line of perfect agreement. It can be observed that all verification points lie essentially on the agreement line, indicating that the numerical implementation reproduces the governing energy-balance equation with negligible error.

This result confirms that the developed Python code is numerically consistent with the first-law heat-balance formulation adopted in the mathematical model. Because the present work is a numerical optimization study, the purpose of this verification was to check consistency with the governing analytical model rather than to provide experimental validation. Therefore, Figure *4* serves as a direct verification that the computational implementation accurately preserves the imposed thermal-load balance. It is important to clarify that this verification does not represent experimental validation of the physical cooling system. Rather, it confirms that the Python implementation is consistent with the analytical equations used in the present numerical model. Experimental validation, Computational Fluid Dynamics (CFD) comparison, or benchmarking against industrial cooling data is recommended as future work.

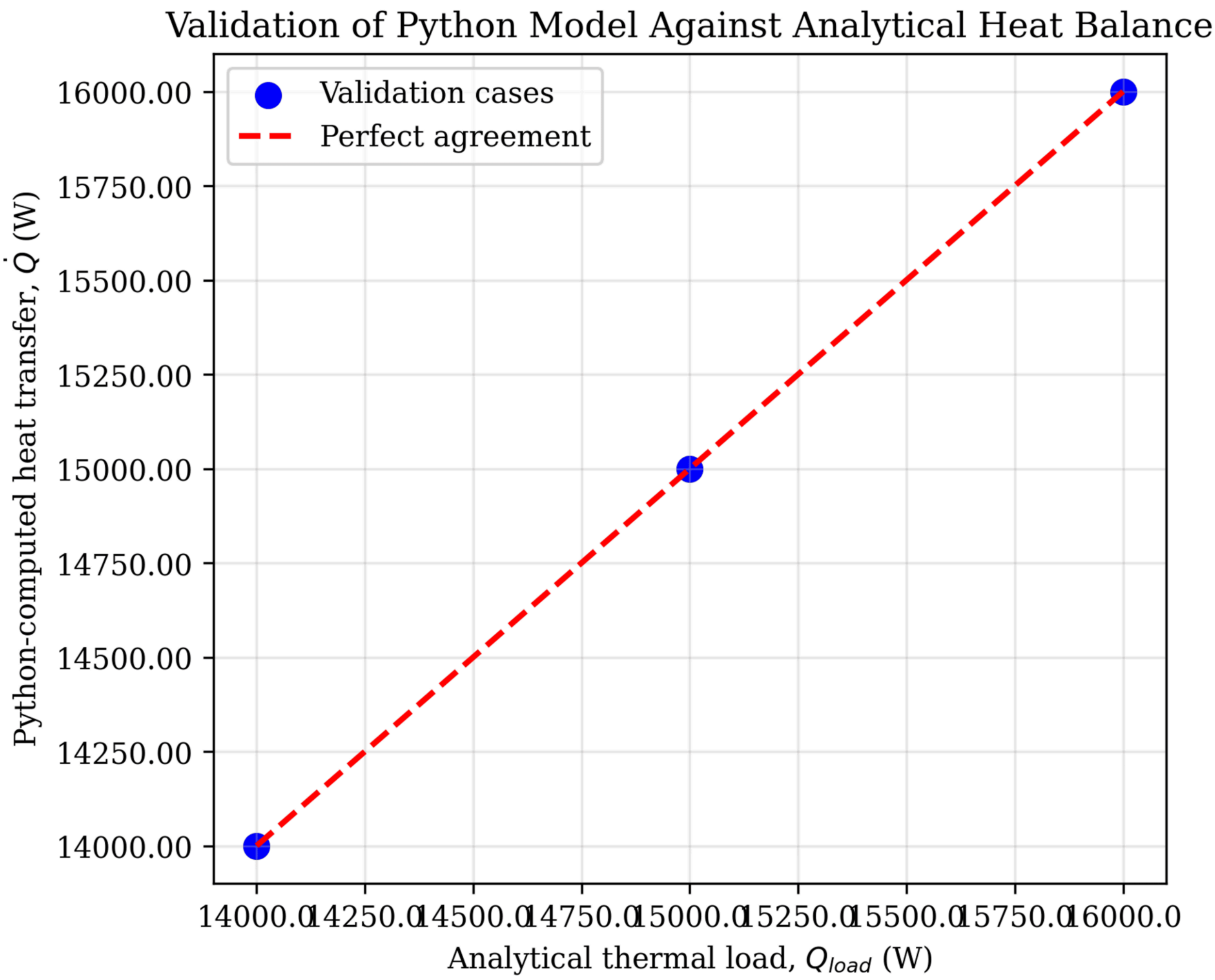

FIGURE 4: Numerical verification of the Python model against the analytical heat-balance relation. The Python-computed heat transfer rate is compared with the analytical thermal load for representative verification cases, while the dashed diagonal line indicates perfect agreement between the two results

In addition to the heat-balance verification, the exergy-performance values obtained from the Python program were also checked against the analytical exergy expression,

$$\eta_{ex} = \frac{\dot{Q}\left(1 - \frac{T_0}{T_{out}}\right)}{W_{pump} + \varepsilon}$$

The results exhibited physically meaningful and thermodynamically consistent trends. In particular, as the coolant mass flow rate increased from 0.19 to 0.23 kg/s, the pump power increased from 7.41 to 13.14 W, whereas the exergy efficiency decreased from 71.37 to 28.40. This behavior indicates that, beyond a moderate operating range, the increase in pumping-power requirement becomes more dominant than the gain in useful thermal exergy. Hence, the Python implementation is not only mathematically consistent with the analytical model but also physically meaningful for the present optimization study.

How to cite this article:
Tiwari C, Srivastava P K, Rastogi S (July 20, 2026) Multi-Objective Optimization of Thermal Management Systems for Artificial Intelligence Server Racks Using Exergy Analysis and Life-Cycle Assessment. Cureus J Eng 3 : es44388-026-00174-8. DOI https://doi.org/10.7759/s44388-026-00174-8

Parametric sensitivity study

To further understand system behavior near the feasible operating region, a parametric study was carried out by varying the coolant mass flow rate ($\dot{m}$) and inlet temperature (T_{in}), while keeping the remaining design variables fixed at near-optimal values. The corresponding results are reported in Table 3. The LCI values reported in the parametric study are separate sensitivity-case results and should not be confused with the optimized LCI value of 159881.57 normalized impact units. The optimum value corresponds to the final feasible NSGA-II solution, whereas the parametric values are used only to examine the response of the model under selected operating conditions.

T_{in}(K)	$\dot{m}$ (kg/s)	W_{pump}	Exergy efficiency (η_{ex})	LCI (units)
290	0.19	7.41	71.37	324744.21
290	0.21	10.0	44.38	438370.15
290	0.23	13.14	28.40	575837.35

TABLE 3: Parametric study results at near-optimal conditions

LCI, life-cycle impact

The results show that the coolant mass flow rate is one of the most influential operating parameters. When $\dot{m}$ was increased from 0.19 to 0.23 kg/s at $T_{in} = 290\ \mathrm{K}$, the pump power increased from 7.41 to 13.14 W, corresponding to an increase of about 77.4%. Over the same range, the LCI increased from 324744.21 to 575837.35 units, whereas the exergy efficiency decreased from 71.37 to 28.40. These results indicate that excessive coolant flow leads to a strong pumping-energy penalty, which outweighs the thermodynamic benefit.

A temperature-based sensitivity analysis was also performed. At a fixed mass flow rate, increasing the inlet temperature increased the outlet temperature and affected the available thermal exergy. This confirms that both hydraulic and thermal operating conditions influence system performance, although the effect of the coolant flow rate is more dominant in the present model.

Overall, Table 3 confirms that the best thermal-management performance is achieved not at the extreme values of the operating variables, but within a balanced intermediate region that limits both exergy destruction and life-cycle burden. The sensitivity trends are also visualized in Figure 5, which clearly shows the opposing variation of exergy efficiency and LCI with increasing coolant mass flow rate.

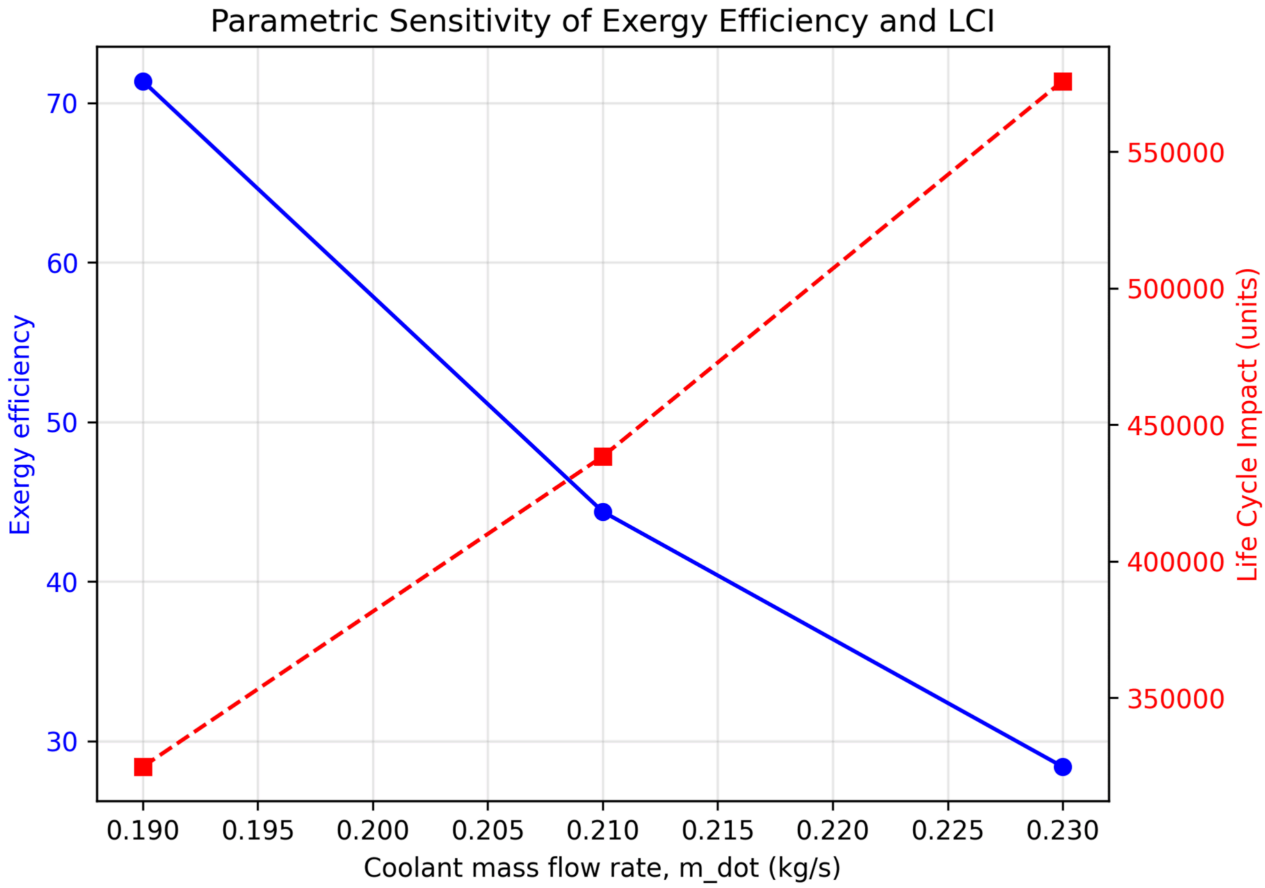

FIGURE 5: Parametric sensitivity of exergy efficiency and life-cycle impact with respect to coolant mass flow rate. Increasing flow rate increases pumping-related environmental burden while reducing exergy efficiency beyond the moderate operating region

LCI, life-cycle impact

Convergence characteristics of NSGA-II

The history of NSGA-II convergence is demonstrated in the figure as the decrease of average and minimum constraint violation per generation. At first, the majority of the population violated the constraints, especially the cost and thermal ones. Over time, the population was more and more populated with solutions that were feasible. The convergence history indicates that feasible solutions were identified at an early stage of the optimization process, after which the algorithm focused on improving objective-space performance rather than feasibility alone. This behavior confirms that the feasible design region was reached efficiently during the optimization process. This kind of convergence pattern agrees with that reported in previous research works where NSGA-II was used for constrained energy systems [11,18].

Pareto front analysis

Pareto Front Structure

The final optimization results after 200 generations are depicted in Figure *3*, where exergy efficiency is plotted against the life-cycle environmental impact. In the present formulation, the feasible solutions converged to a single dominant optimum rather than a broad Pareto trade-off curve. The optimization yielded a scaled exergy-performance index of

329.34 and a corresponding LCI of 159881.57 normalized impact units. This indicates that, under the adopted constraints and model assumptions, the thermodynamic and environmental objectives were simultaneously satisfied at the same operating condition.

The coincidence of the feasible optimum and the knee-point solution suggests that the current design space is strongly restricted by the thermal, cost, and geometric constraints. As a result, the optimization does not produce a wide spread of competing solutions, but instead identifies a narrow feasible region with a single preferred design. This result still remains useful from an engineering viewpoint, since it identifies the most suitable operating condition for the TMS within the imposed design limits.

Feasible vs Infeasible Solutions

Moreover, Figure *3* further separates the feasible from the infeasible solutions. The infeasible solutions mainly gather at the extremes of exergy efficiency and environmental impact where the temperature, volume, or cost constraints are violated. This serves as a good reminder of the role that constraints play in any engineering optimization. Even though from a thermodynamic standpoint certain designs may seem optimal, when their practicality in terms of spatial, economic, or safety considerations is taken into consideration, they are revealed as impractical [15].

Interpretation of objective trends

Exergy Efficiency Behavior

In the present model, exergy efficiency decreases beyond a moderate coolant flow rate because the pumping-power requirement increases faster than the useful thermal exergy gain. This behavior is consistent with the numerical verification and parametric study results, where increasing the coolant mass flow rate caused a rise in pump power but a reduction in the exergy-performance index. Thus, after a certain operating level, the thermodynamic benefit of additional flow becomes marginal, while the hydraulic penalty becomes dominant. The classical exergy theory explains this phenomenon, stating that minimum entropy generation does not necessarily imply maximum useful thermal performance [5].

Life-Cycle Impact Behavior

The environmental impact over the life cycle is mainly due to: embodied energy of materials (heat exchanger size and material factor) and long-term operational energy consumption of the pump. High mass flow rate and large surface area designs lead to a substantial increase in both embodied and operational impacts. Similar patterns have also been observed in the life-cycle studies of cooling and HVAC systems [16,22].

Knee point identification

Rationale for Knee Point Selection

In MOO, a knee point generally represents a solution at which further improvement in one objective leads to a relatively large deterioration in the other. For the present study, the knee point was identified using the distance-to-ideal-point method in the normalized objective space, as described in the previous section. However, in the current optimization run, the feasible design space converged to a single dominant solution. Therefore, the knee point identification procedure served mainly as a confirmation that the obtained optimum also represents the best compromise between thermodynamic performance and life-cycle environmental impact.

Knee Point Location

The knee point was identified at: scaled exergy-performance index: 329.34, LCI: □ 159881.57, constraint violation: zero. Unlike a widely distributed pareto front, the present optimization converged to a single dominant feasible solution. Therefore, the knee point coincided with the optimum point, showing that the most favorable thermodynamic and environmental performance was achieved simultaneously within the feasible design region.

Optimal design variables at knee point

The optimal design variables corresponding to the knee point solution are listed in Table 4, which summarizes the final recommended operating condition.

Variable	Value	Unit
Heat exchanger area, A	0.50	m^2
Pipe diameter, D	0.0832	m
Mass flow rate, $\dot{m}$	0.15	kg/s
Inlet temperature, T_{in}	300.0	K
Pump power, W_{pump}	3.65	W
Material factor, M_f	1.0	--

TABLE 4: Optimal design variables at knee point

These values indicate a compact heat-exchanger configuration, a moderate pipe diameter, a relatively low coolant mass flow rate, and low pumping power. Together, they represent the most favorable thermodynamic and environmental operating condition obtained within the feasible design space.

Constraint satisfaction analysis

At the knee point, all constraints were satisfied:

Thermal constraint: The outlet temperature remained below the maximum allowable limit.

Volume constraint: The system geometry remained within the available design space.

Cost constraint: The estimated total cost remained within the prescribed budget.

This confirms that the selected design is feasible and suitable for practical implementation within the assumptions of the present model.

Engineering implications

Comparison with Conventional Designs

Conventional TMSs are often designed primarily to maximize heat-removal capability, with limited consideration of life-cycle environmental impact. The present results show that a compact and low-flow design can simultaneously satisfy thermal requirements and reduce the overall environmental burden. This observation is consistent with the growing emphasis on sustainable thermal design for data centers and AI hardware systems [23].

Sustainability Perspective

The integration of exergy analysis with LCA provides a broader basis for design evaluation than energy-only methods. A design that appears thermally efficient may still impose a higher environmental burden due to pumping energy, material usage, or long-term operational demand. Therefore, the present framework demonstrates the importance of combining thermodynamic and environmental criteria in the optimization of next-generation AI server cooling systems [12].

Sensitivity and robustness discussion

The optimization results indicate that the feasible design space is strongly restricted by the imposed thermal, cost, and geometric constraints. As a result, the final solution converged to a single dominant feasible optimum. This suggests that, within the present formulation, the selected operating condition is robust in the sense that no competing feasible design offered a superior compromise between exergy efficiency and LCI.

Limitations of the study

This research - despite being useful in its framework - has the following limitations:

The steady-state heat transfer model was used.
Heat losses to the surroundings were ignored and the properties of the coolant were assumed to be constant.
There was no experimental, CFD, or industrial validation of the model.
Simplified parametric coefficients were used to represent the LCI model instead of the detailed material inventory database.
End-of-life treatment and transportation impacts and manufacturing process specific impacts were not considered.
Variables that were selected for the design space included heat-exchanger area, pipe diameter, mass flow rate, inlet temperature, and material factor.
Future works should account for other practical variables such as coolant type, geometry of the heat-exchanger, flow configuration, pump type, and thermal interface materials.

These limitations suggest several directions for future work, including higher-fidelity thermal modeling, experimental validation, CFD-based benchmarking, and the inclusion of additional design objectives. The principal findings of the present study are summarized in Table *5*.

Finding	Description
Feasible optimum	Scaled exergy-performance index of 329.34 with LCI of 159881.57.
Constraint satisfaction	Thermal, volume, and cost constraints all satisfied.
Knee point	Coincided with the dominant feasible optimum.
Design implication	Compact low-flow design favored under current assumptions.

TABLE 5: Summary of major findings

Conclusions

This study shows that sustainable thermal management of AI server racks should not be treated only as an efficiency-maximization problem. Instead, a combined thermodynamic, environmental, and optimization-based framework is needed. The present optimization yielded a feasible optimum with a scaled exergy-performance index of 329.34 and a

LCI of 159881.57 normalized impact units. These findings demonstrate that the proposed framework can identify a thermodynamically consistent and environmentally informed design solution for AI server rack cooling under practical engineering constraints.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Shrishti Rastogi, Chinta Mani Tiwari, Prashant K. Srivastava

Acquisition, analysis, or interpretation of data: Shrishti Rastogi, Chinta Mani Tiwari, Prashant K. Srivastava

Drafting of the manuscript: Shrishti Rastogi, Chinta Mani Tiwari, Prashant K. Srivastava

Critical review of the manuscript for important intellectual content: Shrishti Rastogi, Chinta Mani Tiwari, Prashant K. Srivastava

Supervision: Chinta Mani Tiwari, Prashant K. Srivastava

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Data Availability Statements

The datasets (and/or code) supporting this study are available from the corresponding author upon reasonable request.
AI Declaration: The authors utilized Large Language Models (LLMs) solely for grammatical correction and language polishing to enhance clarity. No AI was used to generate scientific data, results, or conclusions.

Acknowledgements

We, the authors, would like to extend our sincere thanks to the organizers of the 1st International Conference on Robotics, Automation & Mechanical Engineering (ICRAME-2026), which was organized on January 30–31, 2026 by the Department of Mechanical Engineering, Khwaja Moinuddin Chishti Language University (KMCLU), Lucknow. This conference has given us a great opportunity to present our work, share our ideas, and have fruitful discussions with renowned academicians and experts from industry. We are thankful to the conveners, coordinators, and the entire organizing team for their hard work and efforts in making this conference a success. We would also like to thank the participants and researchers who have contributed to the success of this conference and have motivated us to work further toward advancements in the field of robotics, automation, and mechanical engineering.

References

1. Huigui H, Haomin Z, Sheng X, Canbing L, Chunhua H: Optimizing energy consumption for data centers. Renewable and Sustainable Energy Reviews. 2016, 58:674-691. 10.1016/j.rser.2015.12.283
2. Shehabi A, Smith S, Sartor D, et al.: United States Data Center Energy Usage Report. Lawrence Berkeley National Laboratory. 2016, 10.2172/1372902

3. Iyengar M, Schmidt R: Analytical modeling for thermodynamic characterization of data center cooling systems. Journal of Electronic Packaging. 2009, 131:021009. 10.1115/1.3103952

4. IEA: Data Centres and Data Transmission Networks. International Energy Agency. 2023,

5. Bejan A, Tsatsaronis G, Moran MJ: Thermal Design and Optimization. John Wiley & Sons, New York; 1995.

6. Rosen MA, Dincer I: Exergy as the confluence of energy, environment and sustainable development. Exergy, An International Journal. 2001, 1:3-13. 10.1016/s1164-0235(01)00004-8

7. Rao RV, Patel V: Multi-objective optimization of heat exchangers using a modified teaching-learning-based optimization algorithm. Applied Mathematical Modelling. 2013, 37:1147-1162. 10.1016/j.apm.2012.03.043

8. International Organization for Standardization: Environmental Management - Life Cycle Assessment - Principles and Framework (ISO 14040). International Standard, Geneva, Switzerland; 2006. https://www.iso.org/obp/ui/#iso:std:iso:14040:ed-2:v1:en.

9. Deng L, Babbitt CW, Williams ED: Economic-balance hybrid LCA extended with uncertainty analysis: case study of a laptop computer. Journal of Cleaner Production. 2011, 19:1198-1206. 10.1016/j.jclepro.2011.03.004

10. Deb K: Multi-Objective Optimization using Evolutionary Algorithms. John Wiley & Sons Inc, Chichester, UK; 2001.

11. Deb K, Pratap A, Agarwal S, Meyarivan T: A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation. 2002, 6:182-197. 10.1109/4235.996017

12. Dincer I, Rosen MA: Exergy: Energy, Environment and Sustainable Development. Newnes, Oxford, UK; 2012.

13. White FM: Fluid Mechanics. McGraw Hill, New York, USA; 2011.

14. Miettinen K: Nonlinear Multiobjective Optimization. Springer, New York, NY; 1998. 10.1007/978-1-4615-5563-6

15. Rosen MA, Dincer I: Exergy analysis of waste emissions. International Journal of Energy Research. 1999, 23:1153-1163. 10.1002/(SICI)1099-114X(19991025)23:13<1153::AID-ER545>3.0.CO;2-Y

16. Sun Z, Wang R, Lu J, Guo G, Fan Y, Li K: Multi-objective Optimization of Integrated Energy System Based on Weighted Fuzzy Algorithm. Chinese Control Conference. 2021, 5782-5787. 10.23919/CCC52363.2021.9549859

17. Deb K, Agrawal RB: Simulated Binary Crossover for Continuous Search Space. Complex Systems. 1995, 9:

18. Coello Coello CA, Lamont GB, Van Veldhuizen DA: Evolutionary Algorithms for Solving Multi-Objective Problems. Springer, New York, NY; 2007. 10.1007/978-0-387-36797-2

19. Blank J, Deb K: Pymoo: Multi-Objective Optimization in Python. IEEE Access. 2020, 8:89497-89509. 10.1109/access.2020.2990567

20. Oyama A, Fujii K, Shimoyama K, Liou M: Pareto-Optimality-Based Constraint-Handling Technique and Its Application to Compressor Design. 17th AIAA Computational Fluid Dynamics Conference. 2005, 10.2514/6.2005-4983

21. Branke J, Deb K, Dierolf H, Osswald M: Finding knees in multi-objective optimization. Parallel Problem Solving from Nature. Yao X et al. (ed): Springer, Berlin, Heidelberg; 2004. 722-731. 10.1007/978-3-540-30217-9_73

22. Raugei M, Sgouridis S, Murphy D, et al.: Energy Return on Energy Invested (ERoEI) for photovoltaic solar systems in regions of moderate insolation: A comprehensive response. Energy Policy. 2017, 102:377-384. 10.1016/j.enpol.2016.12.042

23. ASHRAE (Firm): Thermal Guidelines for Data Processing Environments. ASHRAE, 2021.

How to cite this article:
Tiwari C, Srivastava P K, Rastogi S (July 20, 2026) Multi-Objective Optimization of Thermal Management Systems for Artificial Intelligence Server Racks Using Exergy Analysis and Life-Cycle Assessment. Cureus J Eng 3 : es44388-026-00174-8. DOI https://doi.org/10.7759/s44388-026-00174-8