The paradox of success in evolutionary and bioinspired optimization: Revisiting critical issues, key studies, and methodological pathways

Daniel Molina [a],*, Javier Del Ser [b,c], Javier Poyatos [a], Francisco Herrera [a]

[a] *Department of Computer Science and Artificial Intelligence, Andalusian Research Institute in Data Science and Computational Intelligence (DaSCI), University of Granada, Granada, 18071, Spain*
[b] *TECNALIA, Basque Research & Technology Alliance (BRTA), Derio, 48160, Spain*
[c] *Department of Mathematics, University of the Basque Country (UPV/EHU), Leioa, 48940, Spain*

ARTICLE INFO

ABSTRACT

Evolutionary and bioinspired computation are crucial for efficiently addressing complex optimization problems across diverse application domains. By mimicking processes observed in nature, like evolution itself, these algorithms offer innovative solutions beyond the reach of traditional optimization methods. They excel at finding near-optimal solutions in large, complex search spaces, making them invaluable in numerous fields. However, both areas are plagued by challenges at their core, including inadequate benchmarking, problem-specific overfitting, insufficient theoretical grounding, and superfluous proposals justified only by their biological metaphor. This overview recapitulates and analyzes in depth the criticisms concerning the lack of innovation and rigor in experimental studies within the field. To this end, we examine the judgmental positions of the existing literature in an informed attempt to guide the research community toward directions of solid contribution and advancement in these areas. We summarize guidelines for the design of evolutionary and bioinspired optimizers, the development of experimental comparisons, and the derivation of novel proposals that take a step further in the field. We provide a brief note on automating the process of creating these algorithms, which may help align metaheuristic optimization research with its primary objective (solving real-world problems), provided that our identified pathways are followed. Our conclusions underscore the need for a sustained push towards innovation and the enforcement of methodological rigor in prospective studies to fully realize the potential of these advanced computational techniques.

1. Introduction

Bioinspired computation, a prominent area within metaheuristic optimization research [1], focuses on the design of computational techniques and optimization algorithms that draw inspiration from biological processes and behaviors observed in nature. This inspiration allows such algorithms to solve complex optimization problems in a more efficient and robust manner than traditional solvers. Evolutionary computation [2] plays a central role in this research area, finding its inspiration in principles from Darwin's Theory of Evolution to tackle intricate optimization problems that often surpass the capabilities of traditional approaches [3–5]. Both fields have seen significant practical advancements [6], demonstrating their potential to improve other models, from classic Machine Learning (ML) pipelines [7] to deep neural networks [8,9].

Recent studies have anticipated the optimization challenges posed by the advent of General-Purpose Artificial Intelligence Systems (GPAIS) [10], in which the flexibility of bioinspired optimization approaches can address the complexity of GPAIS for several learning tasks. Examples include Generative Adversarial Networks in zero-shot learning [11], prompt evolution for Large Language Models [12], and evolutionary algorithms for transfer learning and pruning in deep neural networks [13].

The surge in publications that present new bioinspired optimization algorithms has motivated efforts to systematically classify this vast body of literature in different taxonomies [14,15], as well as comprehensive overviews of the topic [16–18]. Several of these contributions have been continuously updated over time, leading to more sophisticated taxonomies that incorporate descriptions of novel overviews and methodologies. A notable example is [19], whose latest version maintained in [20] has systematically collected, analyzed, and classified more than 500 bioinspired solvers to date.

---

* Corresponding author.
E-mail addresses: dmolina@decsai.ugr.es (D. Molina), javier.delser@tecnalia.com (J. Del Ser), poyatosamador@gmail.com (J. Poyatos), herrera@decsai.ugr.es (F. Herrera).

Although the explosive growth and success of bioinspired computation have established it as a cornerstone of metaheuristics, they have ironically given rise to a *paradox of success*. The influx of academic interest and sustained flow of contributions have not only accelerated advancements, but also increased the concerns of part of the community around critical issues and bad practices noticed in the area.

Prior to this recent surge, the community had concerns about accepting new proposals due to insufficient experimentation, which often failed to demonstrate significant differences. In particular, [21] raised a critical voice about the acceptance of new algorithms based on experimental results that might not be significant for many reasons. This suggests that there might not be a direct correlation between the level of interest in an algorithm within the research community and its ease of being published.

In [22], the authors state that a new proposal should offer at least one advantage – such as speed, performance, robustness, or innovation – to demonstrate its contribution to the field over previous heuristics, considering innovation as valuable algorithm novelty approaches. In the following, we will define the "novelty" of an algorithm as innovation from an algorithmic standpoint, rather than from a bioinspiration perspective. The seminal work by Sörensen [23], along with other contemporary manifests focused on criticizing specific solvers, was the first of its kind to expose that a metaphor is in no way a factor guaranteeing the novelty of a bioinspired solver. Since then, the number of critiques of such practices has increased sharply, highlighting the lack of novelty of certain algorithms or exposing poor methodological practices in benchmarking and evaluation [24,25]. Remarkably, the importance of this growing corpus of critiques has motivated the publication of an open letter formally rejecting noninnovative proposals, signed by front-line experts in this research area [26].

On a positive note, the research community is addressing shortcomings in bioinspired algorithms by proposing methodological guidelines, quantifying differences between algorithms and their operators [27], establishing rigorous comparison guidelines [28], automating algorithm design [29], and developing appropriate benchmarks [30]. These initiatives aim to promote best practices and drive transformation in the field, ensuring that bioinspired and evolutionary computation continue to advance robustly in years to come.

In this context, this manuscript presents an overview of the weaknesses and criticisms that have appeared in the literature over the years, delving into the rationale given by the authors for their claims. In addition, it also analyzes the different methodological pathways that have been put forward to address these criticisms and to improve the field. By examining the two sides of the coin, the problems and the corresponding solutions, our study provides a comprehensive analysis of the aforementioned *paradox of success*: a plethora of publications presenting new solvers and/or applications using bioinspired or evolutionary algorithms, but unfortunately flooded by algorithmic designs do not really offering a significant contribution to the area, mainly due to a lack of novelty beyond the biological metaphor inspiring its design. As will be discussed later, it is time to separate the wheat from the chaff, starting from a positive recapitulation of the current state of the area in this regard, and ending with referential guidelines to drive efforts in the field toward impactful scientific advancements and practical applications. This recapitulation is indeed the overarching goal of this work.

The paper is structured according to this overall goal. First, Section 2 addresses the problems, examining critical issues known in the bioinspired and evolutionary computation fields and the key studies that expose them. Next, Sections 3, 4 and 5 showcase the efforts made by the community to address these critical problems through the development of proper methodological pathways for the benefit and progress in the area. Next, Section 6 focuses on the automated design of metaheuristics algorithms as a research area that can overcome the shortcomings discussed previously. Finally, Section 7 draws the main conclusions from this study.

2. Known critical issues: From lack of innovation to low-quality experimental studies, and benchmarking

As mentioned in the introduction, the number of bioinspired solvers has grown significantly in recent years, owing to methods that draw inspiration from natural behaviors or processes. However, many proposals lack innovation and experimental rigor, hindering the field's progress. It is paradoxical that the field's success in terms of publications and proposals is what threatens its real progress. In agreement with the title of this paper, this section examines the issues leading to this *paradox of success*: the lack of algorithmic innovation (Section 2.1), the neglect of critical aspects of experimentation (Section 2.2), and the failure to use appropriate benchmarks (Section 2.3).

2.1. The illusion of novelty in metaheuristic design: On the lack of algorithmic innovation

The plethora of bioinspired algorithms available poses a significant challenge in choosing the best solver for an optimization problem. In this context, the work of Molina et al. [19] proposes a dual taxonomy according to their inspiration and algorithmic behavior. They highlight that:

"*In summary, although in the last years many nature-inspired algorithms have been proposed by the community and their number grows steadily every year, more than half of the proposals reviewed in our work are incremental, minor versions of only three very classical algorithms (Particle Swarm Optimization, Differential Evolution, and Genetic Algorithms). We therefore conclude that the large number of natural and biological sources of inspiration used so far to justify the design of new optimization solvers has not led to significantly disruptive algorithmic behaviors*".

In other words, many recently developed algorithms, inspired by natural phenomena or processes/behaviors observed in other fields of knowledge, frequently recycle existing ideas without introducing algorithmically new methods or approaches.

This has been a concern for almost a decade. As was discussed in [23], the fallacies of "novel" metaphor-based methods and the vulnerability of bioinspired research emphasized the need to impose good research practices in the area. In the author's own words, "*most "novel" metaheuristics based on a new metaphor take the field of metaheuristics a step backward rather than forward*. That study is a "*call for a more critical evaluation of such methods*" because "*there are more than enough novel methods, and there is no need for the introduction of new metaphors just for the sake of it*".

Before Sorensen's manifesto was published, other researchers [31, 32] had also addressed the lack of innovation of specific bioinspired solvers, claiming that the diversity of their natural inspirations did not extend to mathematical differences with previous evolutionary algorithms. Weyland [31] (together with a series of follow-up rebuttals and responses [33,34]) provided a comprehensive review of *Harmony Search* harmony, and critically analyzed its developments over the past decade. He argued that this bioinspired solver, often seen as a significant innovation, is essentially a special case of Evolution Strategies. However, this assertion was later called into question by Hasançebi et al. [35]. The second paper [32] focused on a bioinspired approach, called *Black Hole Optimization* blackhole, demonstrating that it is, in fact, a "*simplification of the well-known Particle Swarm Optimization with inertia weight*". Furthermore, the authors express their concern about the fact "*that such an approach may jeopardize achievements in the field of Evolutionary Computing*", since Black Hole Optimization is a simplification that produces favorable results in specific benchmark tests.

Later, the work in [36] discussed this phenomenon in relation to the number of papers published, arguing that "*the new population-based nature-inspired algorithms are released every month and, basically, they*

have nothing special and no novel features for science". Their conclusions are aligned with the taxonomy published in [19], because "*our research revealed that the process of the new population-based nature-inspired algorithm possesses the behavior of the swarm intelligence paradigm*", revealing that the category with the highest number of contributions is Swarm Intelligence.

Thereafter, in [24] a meticulous review of the trends of new algorithms was performed, providing additional criticisms that apply to most of these proposals:

1. The existence of a physical analogy, as there are some examples of bioinspired algorithms that present a phenomenon that does not exist in nature.
2. The similar inspiration and even duplicate methods, due to the existence of groups of algorithms very similar to established solvers, with subtle differences lying on slight modifications to their algorithmic components.
3. Multiple techniques drawing inspiration from the same idea and with authors publishing multiple variants over time under this premise. Interestingly, this observed practice escalates to research groups publishing "novel" algorithms that solve the same set of functions and problems.
4. New natural-inspired algorithms provide added value when their contribution is crucial. There are some cases in which the introduction of novel algorithms is necessary, and in those cases, the development of novel algorithms is considered a suitable solution.

In addition, the study in [24] examines the circumstances in which it is necessary to develop a new bioinspired algorithm. Two particular cases merit further attention, as they exemplify the underlying postulates that drive our analysis:

• "*when it manages to solve a problem in a way superior to the way it was solved by previous competitive approaches*"; or when
• "*more intelligent mechanism has been incorporated within this new nature-inspired algorithm that seems to render it more efficient than others*".

They agree on the importance of its application to real-world problems and highlight the fact that "*the main goal from now on should be the attempt to upgrade and fine-tune the existing nature-inspired algorithms, to obtain competitive solutions in real-world applications*".

The considerable number of proposals that fail to introduce genuinely new concepts not only fail to contribute to the field, but also present a significant challenge for researchers trying to stay up-to-date, due to the difficulty of identifying relevant proposals among the plethora of irrelevant proposals. Fortunately, several papers have attempted to identify non-innovative proposals, which we will describe in Section 3.1.

Although it is beneficial to have resources to identify ineffective proposals, this is not a sufficient solution. It is essential to understand the underlying factors that contribute to the exponential growth of the number of proposals. In this context, Campelo and Aranha [25] remark that a possible cause is the pressure to *publish or perish*, which is perceived as a low-effort, low-risk process with high potential rewards. Some authors have built professional careers from developing an upsurge of metaphor-based methods. The authors of this paper maintain a regularly updated collection of bioinspired proposals in the well-known Evolutionary Computation Bestiary.[1] In their words, "*the phenomenon of metaphor-inspired methods has had a negative impact on the field, wasting the work of scientists and reviewers on methods that continuously reinvent the wheel, hiding sloppy or dubious practices*". In this context, the aforementioned *publish or perish* phenomenon has been taken to new extremes, all in an effort to gain more influence in the field. As

---

[1] Evolutionary Computation Bestiary, http://fcampelo.github.io/EC-Bestiary/, accessed on January 10th, 2025.

described by the authors, "*although the critical tone of the Bestiary is clearly stated in the repository, we are often contacted by authors of "novel" metaphor-based metaheuristics requesting that their work be listed. It has never been clear to us whether these authors do not understand the critical tone of the page or if they assume that any exposition, however critical, would be a net positive for their work*".

For this reason, and to avoid falling into this bad practice, in [26] the authors wrote a letter to "calling on all editors in chief in the field to adapt their editorial policies" to reject the publication of *allegedly novel* metaphor-based bioinspired algorithms. This letter underscores that the authors of new proposals have succeeded in publishing them in journals related to the metaphor itself, without a comprehensive and principled editorial assessment of the computational advantages and benefits of the presented algorithm: "*For instance, if the metaphor is the mating behavior of bats, the authors will attempt to publish it in specialized journals on bats or animal mating behavior; metaheuristics inspired by improvising musicians will go for journals on music*". For these reasons, almost a hundred important researchers in the area agreed to sign this letter and accept the publication of novel bioinspired algorithms under the following postulates:

1. "*present their method using the normal, standard optimization terminology*";
2. "*show that the new method brings useful and novel concepts to the field*";
3. "*motivate the use of the metaphor on a sound, scientific basis*"; and
4. "*present a fair comparison with other state-of-the-art methods using state-of-the-art practices for benchmarking algorithms*".

Despite the critical manifests expressed by several authors over the years, concerns regarding these issues continue to persist today. This is evinced by several works published during this year. For example, the study in [37] examines various natural mechanisms and certain "*commonalities of nature-inspired intelligent computing paradigms. These commonalities provide a solid algorithmic foundation to avoid designing unreasonable metaphors*".

Another work in the same line is [38], that analyzes various bioinspired algorithms, focusing on Arithmetic Optimization Algorithm's RRO mathematical formulation, implementation, and lack of unbiased comparisons. The authors conclude that "*the current trajectory of focusing on novelty over theoretical foundations jeopardizes the credibility and long-term sustainability of advancements in the field. Hence, it becomes imperative for researchers to direct their efforts toward developing and substantiating theoretical foundations for metaheuristic algorithms*". Another study that exposes the lack of novelty of a specific modern bioinspired solver is [39], which performs a comparison of the algorithm called *Raven Roost Optimization* RRO. Their analysis reaches a conclusion similar to that of previous studies dealing with other solvers. Raven Roost Optimization "*contains no real novelty compared to the well-established Particle Swarm Optimization, and in fact, will perform worse in many situations due to the inherent bias towards its starting point*".

We consider that the combination of critical analysis of existing proposals, in conjunction with stricter criteria for the acceptance in the literature of new proposals, giving special attention to their contribution, could address the lack of algorithm innovation. We provide further details in this regard in Section 3.

2.2. Deficiencies in experimental validation and evaluation practices

Facing a concern about an excessive focus on experimental results, which is not always a good indicator of a proposal's merit [21], a significant event occurred in 1997, deeply affecting the view on the relationship between an optimization algorithm and its experimental benefits. Wolpert and Macready's proposal, known as the Non-Free Lunch theorem, NFL [40], suggests that "*any two optimization algorithms are equivalent when their performance is averaged across all possible problems*" [41]. This implies that a metaheuristic solver might

outperform others in a specific problem or benchmark only at the cost of performing worse on others. Consequently, there would be no point in searching for a metaheuristic algorithm that outperforms all others, as it could not be improved in general compared to existing ones. This raises serious doubts about the future and interest of this research line.

Several studies have explored the NFL, including its impact on different binary representations [42] and non-uniform distributions [43]. When the NFL holds, summary statistics cannot be used to guide search [44]. For more information, we refer to various reviews on NFL such as [45,46]. Whitley [45] and more recently, McDermott [47] describe situations in with the NFL does not limit the metaheuristics:

- **Focusing on real-world problems:** While Whitley [45] is more skeptical, McDermott [47] argues that subsets of real-world problems allow for scenarios in which one algorithm could be consistently better than random search across all real-world problems within such subsets. As an example, the study in [48] advocated comparing metaheuristics on large sets of problems and proposed a methodology for evaluating their performance on extremely large problem sets. Their empirical study not only found performance differences among five metaheuristics on a set of binary real-world applications with more than 2^{1000} problem instances, but also demonstrated the effects of the NFL theorems on the closed-under-permutation problem set. Additionally, Piotrowski et al. [49] recommended testing the algorithm in large sets of benchmark problems to better simulate realistic experimental environments.
- **Specializing an algorithm with domain-specific or problem subset knowledge:** It is now widely accepted in the literature that including domain-specific knowledge can greatly improve the algorithm in the problem to be tackled [50], in particular in memetic algorithms [51], not being affected by NFL.
- **Proving always that the proposal is better for the particular problem at hand:** It is crucial that researchers "consider more formally whether the methods they develop for particular classes of problems actually are better than other algorithms" [45], including the random search algorithm [47].

Another significant concern in bioinspired and evolutionary computation is the quality of the experimental benchmarks. A substantial number of proposals in the literature conduct insufficient comparative studies to assess the performance of the developed algorithms. The primary issue is their failure to compare with the state-of-the-art methods and similar solvers. However, this is not the case in many of such studies. This is remarked in [45]:

"Too many experimental papers (especially conference papers) include no comparative evaluation; researchers may present a hard problem (perhaps newly minted) and then present an algorithm to solve the problem. The question as to whether some other algorithm could have done just as well (or better!) is ignored".

and remarked in [19]:

"when new algorithms are proposed, unfortunately, many of them are only compared to very basic and classical algorithms (such as Genetic Algorithms or Particle Swarm Optimization). These algorithms have been widely surpassed by more advanced versions over the years, so obtaining better performance than the naive version of classical algorithms is relatively easy to achieve, and it does not imply competitive performance".

To demonstrate any improvements, it is essential to perform comparisons against competitive algorithms and validate that the individual components of the proposals genuinely contribute to enhancing overall performance Barr1995.

Furthermore, the set of competitive algorithms should be chosen considering the state-of-the-art of each benchmark [52], as the behavior of different bioinspired and evolutionary algorithms can largely depend on the characteristics of the benchmark under consideration [53]. In this regard, a recent study by Piotrowski et al. [54] concludes that "*in the present paper we show that the choice of the set of benchmarks used for the comparison may greatly affect the ranking of optimizers*".

Another common criticism expressed against many modern bioinspired algorithms is that the algorithm itself is too complex. A reduction in the number of algorithmic components would facilitate a more comprehensive understanding and analysis of the algorithm's constituent parts. In this line of reasoning, the extensive experiments carried out in [55] showed that "*some meta-heuristics have to be simplified because they contain operators that structurally bias their search by favoring sampling from some parts of the decision space*". Moreover, this study analyzed the L-SHADE-EpSin algorithm Awad2016, which is a step-by-step designed algorithm developed from different variants of Differential Evolution Storn1997. In their experimentation, the authors proved that "*the proposed simplified variant of L-SHADE-EpSin is highly competitive in a wide collection of artificial benchmarks and real-world problems*" compared to other bioinspired algorithms and the non-simplified version of this algorithm.

In [21] it was already mentioned that the overemphasis on publishing based on experimental results poses various more issues, which may even raise doubts about the results' usefulness when compared to appropriate algorithms. For instance, excessive adaptation to new problems, as opposed to previous ones, may not justify the contribution of the detected improvement. Moreover, the most widely considered problems are often the ones that can be solved by new algorithms, not the most adequate for studying and improving the proposals. We will address the influence of both the type of benchmarks and their effects in Section 2.3.

2.3. Benchmark biases and the misrepresentation of algorithm performance

Many studies lack an appropriate and unbiased benchmark, using their own comparison functions, which limits the generalizability of their findings to real world problems. This practice may introduce biases that distort the accurate assessment of the performance of any new optimization algorithm, leading to misleading conclusions.

In this context, the work in [56] examines the different types of biases present in common metaheuristic algorithms. Among them, the study concludes that origin bias and center bias are the most recurrent ones, due to the center bias operators that appear in some bioinspired and evolutionary algorithms. This phenomenon has been thoroughly studied in [57] by systematically analyzing this behavior between several bioinspired algorithms. The reported experiments show a significant tendency in most modern bioinspired solvers to explore mainly the center of domain search. Thus, the majority of these algorithms generate misleading performance results because they are compared considering objective functions whose global optima are located at the center of the solution space. Most of these biases can be avoided by resorting to specific benchmarks proposed over these years, as explained in Section 4.

One reflection we cannot lose sight of within this matter is that benchmark performance cannot be a goal in itself. Benchmarks are a valuable tool for the design and evaluation of algorithms that can potentially perform well in real-world problems, but, as it is said in [45]: "*Don't assume that a search method that does well on classic benchmarks will work equally well on real-world problems. Sometimes algorithms are overly tuned to do well on benchmarks and in fact don't work well on real-world applications*". This idea was also highlighted in Del Ser et al. [6],LaTorre et al. [28]. Similarly, in [58] more detailed issues of relevance are given for real-world optimization problems. We will describe them in Section 4.1.

Similarly to [26], Ceberio and Calvo [59] discuss "*the role of experimentation in the two approaches of conducting research: engineering vs. scientific*", with thoughtful remarks about benchmarks and experimentation details, and present their understanding of the fundamental

principles of both approaches. For this disparity, Kudela proposes the incorporation of a greater number of real world problems into the benchmark [57], by "*the construction of a large set of challenging real-world benchmark problems*". In this way, the author advocates for the establishment of a repository that includes the following:

1. "*several heterogeneous benchmark sets with a unified way of calling the test problems*";
2. "*trusted implementations (source codes) of both standard EC methods and up-to-date state-of-the-art techniques*";
3. "*data obtained from running the algorithms (from (2)) on the benchmarks (from (1))*".

Alternatively, another methodological path that can be followed when benchmarking bioinspired solvers is to continue using synthetic functions. However, as argued in [60], "*the design of benchmarks would benefit from a better understanding of which properties appear in real-world problems, how common these properties are and in which combinations they are found. In addition, it is of particular interest to identify characteristics of real-world problems that are not yet represented in artificial benchmarks and to identify real-world problems that might be usable as part of a benchmark suite. All these aspects would then provide better guidance to the development of algorithms that perform well in the real world*".

Therefore, it becomes crucial to assess the characteristics of the benchmarks in use, as done in [61–63], and to analyze whether these characteristics match those of real world problems. An interesting work on this matter is [60], in which, through a questionnaire, several important properties are identified, such as the presence of noise, solution constraints, multiple but reduced objectives, or unknown optima, among others. Such characteristics are not always effectively represented in synthetic benchmarks.

Here, comes a major cornerstone of benchmarking in metaheuristic optimization: their *representativeness*. It refers to the extent to which a benchmark function accurately reflects the characteristics and challenges of real-world optimization problems. A *representative* benchmark ensures that the performance of metaheuristic algorithms is assessed in a way that is both meaningful and transferable to real-world scenarios. In this regard, Chen et al. [64] studies the importance of benchmark *representativeness* in evaluating optimization algorithms. To this end, it defines three levels of this quality depending on the set of problems to be represented, and introduces a quantitative metric based on several features of benchmark problems, including separability, elementary functions, or dispersion metric, among others.

Another fundamental aspect that is sometimes ignored is to keep in mind that, for real-world problems, there are other qualities of the proposals, apart from the performance. These qualities, like their complexity, hardware requirements, for distributed models the achieved speedup and robustness, or the required time to obtain a reasonably good solution, can be even more relevant than the performance when choosing the right optimizer in many practical applications [58]. Moreover, it is crucial to note that while the stopping criterion is predetermined in most benchmarks, in practical applications this aspect is often more flexible and subject to external constraints. Therefore, it is necessary to evaluate the metrics for a range of function evaluations to gain a more comprehensive understanding of the algorithms' behavior in different scenarios [65].

All these experimental problems will be further examined in Section 4 and illustrated with best practices methodologies for researchers in Section 5. We believe that if these guidelines are applied to new proposals and adherence to them is encouraged by reviewers in editorial processes, these noted experimental issues can be effectively avoided.

3. Separation of wheat from chaff: Distinction of weak proposals

As noted previously, the field of bioinspired algorithms remains problematic due to malpractices and ineffective metaphorical designs, which often result in algorithms that provide no scientific value for the field, for lacking sufficient innovation or flawed experimental designs. We refer to these algorithms as *weak proposals*. In Section 3.1, we describe several key studies that expose these *weak proposals* that do not incorporate innovation. In Section 3.2, we propose several pathways to carry out studies to identify these *weak proposals* in the literature, or to avoid these problems when designing new proposals.

3.1. Critical analyses and identification of algorithmic redundancy in bioinspired optimization

Fortunately, in recent years, several authors have recognized the need for a more rigorous development of new proposals along with a robust experimental framework, and identify in the literature several well-known proposals whose contribution in the field could be considered questionable, or *weak proposals*. This is particularly significant because, unfortunately, the scientific community tends to overlook and undervalue meticulous studies and evaluations of previous algorithms when compared to novel proposals. However, such critical work is crucial and should be prioritized in new proposals yet to be validated by peers.

Analyzing and identifying lack of novelty in bioinspired optimization algorithms has been a focus since early works (cf. [31,32]). These early studies built a foundation for evaluating particular solvers in the field, and their influence has persisted since their publication. An example is Teaching-Learning-Based-Optimization Rao2011, which was analyzed in [66]. This study found that the algorithm has a bias towards the origin during teaching, which increases as the population converges. In addition, the Gravitational Search Algorithm Rashedi2009 was analyzed in [67], finding that the force model in the algorithm depends only on agents' masses, not distances, contradicting the law of gravity.

Years after these studies, Camacho Villalón et al. [68] provided evidence that the Grey Wolf Optimizer Mirjalili2014, Firefly Algorithm Yang2009, and Bat Algorithm Yang2010 are not novel, but rather reformulations of ideas initially introduced for Particle Swarm Optimization Kennedy1995 and subsequently refurnished. In order to prove this claim, the search operators of such solver were rewritten and mathematically compared, concluding that they can all be regarded as existing variants of the Particle Swarm Optimization algorithms. In the studies mentioned in [69–71], several bioinspired algorithms, including the Intelligent Water Drops algorithm Shan2009 and Cuckoo algorithm Yang2009c, are critically examined. The authors argue that the metaphors used to justify these algorithms are often oversimplified or misapplied, leading to misleading or even deceptive claims regarding their novelty and effectiveness. In the case of Intelligent Water Drops, in particular, another study [69] shows that this algorithm can be seen as a simplification of Ant Colony Optimization Dorigo2004. Another example is [39], analyzing the Raven Roost Optimization RRO. The methodology of this paper involved a detailed examination of both the mathematical formulation and an analysis of the behavior, exposing a serious weakness (namely, an inherent search bias towards the starting point).

In the last couple of years, a notable increase in studies critically analyzing certain bioinspired algorithms by comparing them to others by examining their individual components or operators has been observed. Camacho-Villalón et al. [72] contend that the well-known Grey Wolf Optimizer Mirjalili2014, Moth-flame Algorithm Mirjalili2015f, Whale Optimization Algorithm Mirjalili2016, Firefly Algorithm Yang2009, Bat Algorithm Yang2010, and Ant Lion Optimizer Mirjalili2015a are not actually novel. A comprehensive component-based analysis of each algorithm to substantiate this assertion, these algorithms are identified as variants of Particle Swarm Optimization and Evolution Strategies. In this context, the work of Molina et al. [19] categorizes various bioinspired algorithms based on their closest classical solver in terms of algorithmic structure. This analysis highlights

that the distinctions between many bioinspired algorithms and their classical counterparts are often insufficient at the algorithmic level.

The Salp Swarm Optimization Algorithm Mirjalili2017 is a further example of a bioinspired algorithm that has recently been subject to criticism. Castelli et al. [73] identify several problems with this algorithm, which can be broadly classified according to its update rule, its physical motivation, and the discrepancy between their description and its available implementation. In addition, the study demonstrates that the original algorithm exhibits a bias towards the center, a point that was previously discussed. This bias toward the center was also detected in the Arithmetic Optimization Algorithm AOA, analyzed in [38] in terms of its mathematical formulation. As with other solvers, the lack of shifted versions of standard functions hinders the detection of this bias. The Whale Optimization Algorithm Mirjalili2016 has been recently added to the list of weak algorithms, as demonstrated by the inclusion of a center-bias operator at its core design [74]. In [75], several algorithms are examined to discern whether their mechanisms produce desirable qualities in their respective domains. However, the primary findings of this study reject this hypothesis, as recent algorithms do not consistently align with the behavior or phenomenon on which they are based.

Another way to identify weak proposals is through experiments that unfairly enthrone algorithms with non-desired behaviors. An example is the work in [76], which reveals the center bias present in many bioinspired algorithms by analyzing more than 100 bioinspired and evolutionary algorithms. The proposed methodology is to conduct experimental studies on functions whose optimal solution is situated at the center of the feasible set of solutions, and modifying them with specific displacements to avoid that. Under this methodology, it lists several algorithms that should not be used due to their strong center bias, including well-known proposals such as the Grey Wolf Optimizer Mirjalili2014, Dragonfly Algorithm MirjaliliD2016, Whale Optimization Algorithm Mirjalili2016, and Bird Swarm Algorithm Meng2016. This work also recommends avoiding other algorithms for their lack of algorithmic novelty, including Harmony Search and Cuckoo Search, among others. Another interesting work is [56] which, besides classifying different types of bias, proposes several techniques to detect structural bias in the exploration of an evolutionary algorithm. In addition, it provides an extensive list of different algorithms with a strong structural bias identified in the literature. The most common bias is toward the center, but there are also biases towards the boundaries and/or edges of the domain search, among others.

3.2. Pathways to detect weak proposals: Identifying redundant or ineffective algorithms

Through these reviews, it becomes evident that there is a growing interest in adequately comparing new proposals with existing ones in the literature. Based on previous examples, we have examined the methodologies used for this type of *straw-grain* discrimination studies, and we propose pathways that researchers can embrace to enforce a greater rigor in the proposal of new algorithms, detecting their lack of novelty:

• *Equation-level equivalence*, either at the operator or at the component level. This method involves comparing the algorithmic and mathematical descriptions of the search operators of various algorithms towards detecting a possible lack of diversity among them. Techniques for this first pathway include:

  – Conducting homologous component studies by analyzing general expressions of previously established algorithms to identify similarities and differences in their structure and behavior [27].
  – Utilizing formal verification methods to rigorously evaluate algorithmic redundancy [77]. The process begins by formally representing the algorithms to model their structure and behavior. Then,

automated tools such as model checkers or theorem provers are employed to identify equivalences or redundancies by systematically comparing algorithmic components and behaviors.

  – Determining equivalence through operator simplification. This case implies analyzing whether certain components or operators in a bioinspired algorithm contribute substantially to its performance. To assess the impact of individual components, ablation tests should be employed. These tests systematically remove or modify parts of the algorithm to evaluate their specific contribution to the search performance (or to any other quality aspect that is relevant for the problem(s) at hand).
  – Examining the standard definitions of operators to ensure that differences in terminology are not mistakenly interpreted as evidence of differing algorithmic behavior. By systematically analyzing and standardizing operator definitions, the community can more effectively compare meta-heuristic algorithms [78,79]. In addition, it can be applied to compare different implementations of the same algorithm across different software libraries.

• *Configuration-level equivalence*. This pathway refers to the equivalence between two algorithms when specific values of the parameters controlling their search operators can cause them to behave in a very similar manner. The challenge lies in determining when the behavior of one algorithm can be reduced to or generalized by another. Several results can produce different scenarios:

  – Non-innovative algorithm: If the performance of an acclaimed novel solver is essentially identical to an existing algorithm under a certain set of parameter values.
  – Generalization of an existing solver: If the new bioinspired solver introduces additional flexibility in its parameters or expands its search capabilities while maintaining equivalence with an earlier algorithm for specific parametric settings, it must be considered a more general version of the original algorithm, making the original method redundant.
  – Genuinely innovative algorithm: If the novel algorithm introduces significant differences in performance or behavior that cannot be replicated by adjusting the parameter values of an existing algorithm.

The pathways outlined above allow for the identification of weak proposals based exclusively on their definition. However, the final judgment of the worth of an algorithm must also consider the experimental comparison that demonstrates its practical value. In the following section, we introduce methodologies to guarantee reliable experimental comparisons between bioinspired algorithms.

4. Fair and right comparisons in bioinspired and evolutionary optimization: Replicability and benchmarks

To ensure that a proposal for a new bioinspired algorithm advances in the field with originality and practical impact, it is crucial to establish a well-defined experimental section that enables an impartial evaluation of its performance. This is typically associated with improving (or at least matching) the performance of state-of-the-art algorithms. However, the potential contribution of a new proposal is not necessarily limited to an improvement of its search performance (convergence speed or quality of solutions) with respect to the considered counterparts in the benchmark.

In Section 4.1, we outline several considerations that must be made to ensure that the final aim of bioinspired and evolutionary computation is to successfully solve real-world problems. In Section 4.2, we propose several pathways to guarantee fair and right comparisons. In Section 4.3, we underscore the importance of replicability in experimentation. Finally, Section 4.4 emphasizes the importance of using appropriate benchmarks.

4.1. Considerations in real-world problems

When using optimizers to address real-world problems, certain factors are of particular importance. In this regard, Osaba et al. [58] outline the distinct requirements at various stages of the process, including design, development, experimentation, and deployment, to effectively address real-world optimization problems. They identified multiple challenges inherent to the design of new bioinspired algorithms and proposed a series of steps to be followed throughout the algorithm's development. These steps range from the initial phase of problem modeling to the final validation of the algorithm:

- *Problem modeling and mathematical formulation*: It is recommended to begin with a thorough understanding of the real-world context and conceptualization of the problem, which should end in a formal mathematical formulation of the objectives, decision variables, and constraints involved in the problem. This mathematical formulation is essential for determining the structure of the problem and guiding the selection or development of suitable search algorithms. An important outcome of this step is a list of collected functional and non-functional requirements for the solver to be developed, which must be considered towards the design or selection of the solver and accounted for during the rest of the steps.
- *Algorithmic design, solution encoding and search operators*: Once the problem is clearly defined, the next step is to design an algorithm that can efficiently search the solution space. This involves determining how to encode potential solutions, which is critical in bioinspired algorithms, as it directly affects how the algorithm explores and exploits the search space. In addition, it is important to ensure that the algorithm design aligns with the specific constraints and requirements of the problem being solved. This step may involve adapting existing metaheuristics or developing new search operators that better fit the problem's characteristics. The design must consider the functional and non-functional requirements collected during the first step of the process, including factors such as speed, simplicity, deployability in specialized software/hardware, or the possibility of paralleling the implementation of the designed search operators, among others.
- *Performance assessment, comparison and reproducibility*: This phase is dedicated to evaluating the effectiveness of the proposed algorithm. The performance of the algorithm should be rigorously assessed using standard benchmarks or real-world data to determine if it meets the desired requirements. Comparisons with other state-of-the-art algorithms help contextualize their performance. Furthermore, replicability is a crucial factor, ensuring that the results are consistent and can be reproduced in different contexts. This assessment should also account for the specific needs and priorities of the real-world problem, such as computational efficiency, scalability, and robustness, which may be just as important as the raw performance of the optimization algorithm.
- *Algorithmic deployment for real-world applications*: The final step involves implementing the algorithm in the real-world scenario. This phase tests the algorithm's ability to deliver practical solutions under actual operational conditions, which may differ significantly from the observed and modeled experimental environments. Issues such as scalability, robustness, and ease of integration into existing systems become paramount, considering their computational cost and adaptability to dynamic real-world conditions. It is at this step that the algorithm proves its true value for practical impact.

Although the above process can summarize the overarching goal of optimization research (*to efficiently and effectively solve real-world optimization process*), bioinspired and evolutionary optimization research often assumes that the research objective is to design an algorithm that performs better than the state of the art on a certain set of benchmark problems. Although this is a reasonable target for early stage research aimed at producing innovative proposals, it should be subject to the same methodological standards in terms of experimentation and comparison between algorithmic choices.

4.2. Pathways to ensure fair and right experimental comparisons

The community is increasingly aware of the significant flaws exhibited by many new proposals, particularly in the experimental phase of their research. Researchers, like LaTorre et al. [28], have proposed methodological approaches to support more solid and principled experimentation in bioinspired and evolutionary computation. In this specific work, such guidelines can be summarized as follows:

- *Choice of benchmarks and algorithms for comparison*: Benchmarks are a fundamental aspect of algorithm evaluation. As discussed in Section 2.2, custom optimization functions can produce misleading conclusions. A more reliable analysis could be achieved by switching to a better, more standard benchmark or expanding to a larger set of problems [80]. LaTorre et al. [28] advocate for the use of well-designed standard benchmarks that include diverse test functions, representing a broad range of real-world problems. Comparisons with state-of-the-art algorithms should extend beyond just solution quality, incorporating aspects such as efficiency, simplicity, and convergence speed. This kind of analysis can guide future research and improvements in the algorithm's design.
- *Validation of the results*: In the context of metaheuristic algorithms, it is insufficient to present only the raw results of the algorithm's performance; statistical validation is a critical step of the validation of such results. Meta-heuristic algorithms, by their nature, are stochastic and produce variable results, making statistical tests [81,82] essential to determine whether observed differences in performance are statistically significant or simply due to the stochastic nature of the operators inside the solver. Fortunately, statistical validation has become a standard practice in the meta-heuristics literature. Additionally, visualization techniques are extremely valuable, as they help condense complex experimental information, even during the run of algorithms, into more interpretable forms that support and favor the quick assessment of the algorithm's behavior.
- *Components analysis and parameter tuning of the proposal*: The hypotheses of the proposal should be clearly outlined at the beginning of the paper and revisited during the validation of results. Furthermore, authors should conduct a comprehensive analysis of results, covering key aspects such as the search phase, the component analysis, the algorithm parameter tuning, and the statistical comparison with state-of-the-art algorithms. This comprehensive examination guarantees a robust evaluation of the proposed method and its performance in comparison to existing approaches.
- *Why is my algorithm useful?*: Prospective contributors should provide a clear and detailed explanation of why their proposed algorithm is worthy of attention and consideration within the research community. They should demonstrate that their algorithm is competitive with existing, state-of-the-art methods or present a compelling case for the methodological contributions that their work makes to the field. This clarity of communication is essential for conveying the significance of the proposed algorithm and its potential impact on the field.

In recent times, the community has taken up this role by proactively proposing different methodologies, tools, and frameworks. One of the most relevant directions followed in this regard is the proposal of standardized benchmarks and experimental conditions. An important advantage of them is that they facilitate algorithm comparison since researchers can directly compare their results in a transparent, externally verifiable, and replicable fashion.

4.3. Pathways to support the replicability of experimental studies

This is particularly relevant due to the questionable reproducibility or replicability of the results reported for many bioinspired algorithms in the literature. Replicability is one of the motivations at the heart of

the framework proposed in [79] to support the verifiable comparison of new meta-heuristic approaches. This framework hinges on several components:

- *Extensible and re-usable templates*: These templates are designed to offer flexibility by allowing researchers to configure their behavior through an extensible palette of components. This modular approach eliminates the need for altering existing code, making it easier to adapt, extend, and repurpose algorithmic components for a variety of problems.
- *White-box problem descriptions*: A description of the problem with analytic information that could be used to guide the algorithm's selection or construction. Providing detailed transparent descriptions of optimization problems, including analytic information such as objective landscape characteristics or known constraints, can guide the selection or design of operators that are well-suited to the specific problem under consideration.
- *Remotely accessible frameworks, components, and problems*: Hosting algorithm frameworks, optimization components, and a curated set of benchmark problems on a remotely accessible platform can foster reproducibility and collaborative research. Researchers could directly access them, enabling widespread reuse, replicability, and shared knowledge discovery.

4.4. Pathways to guarantee proper benchmarks

The limited variety of test functions commonly used in benchmarks within the community is another issue that has garnered research attention. A recent advancement in this direction is presented in [83], which provides extensive information on practical scenarios for the development of novel optimization algorithms. The review covers more than two hundred mathematical functions and more than fifty real-world engineering design problems. Recognizing the critical need for robust experimental evaluations to accompany the design of bioinspired algorithms, this work offers a comprehensive array of options for assessing the quality and effectiveness of new proposals.

Concerns about design biases in optimization research have increasingly drawn attention in recent years. As highlighted in [57], the center bias can distort the results and misrepresent the true performance of algorithms. To address this issue, Kudela [76] proposed a systematic procedure to identify methods that exhibit such biases, accompanied by a comprehensive analysis of bioinspired and evolutionary algorithms up to 2022. Although the adoption of diverse and well-designed benchmark functions, such as those reviewed in [83], can mitigate the occurrence of center bias, other forms of bias may still persist. This underscores the critical need for new methodologies to effectively detect and counteract these biases in algorithmic evaluations, ensuring fairness and rigor in optimization research.

In response to this need, Walden and Buzdalov [84] propose a novel mechanism based on statistical tests to identify potential biases in algorithms. Their benchmark comprises two identical and symmetrically located global optima, one positioned at the origin. By employing non-parametric statistical tests to analyze the best solutions across independent runs, the authors can detect significant behavioral differences around the optima. If such differences exist, the algorithm under evaluation is flagged as potentially origin-biased. Notably, their study identifies several algorithms that fail this test, including well-known algorithms. This work not only highlights specific instances of bias but also sets a precedent for systematic bias detection, paving the way for more reliable evaluations of bioinspired and evolutionary solvers.

As a final contribution to the advancement in this field, it is essential to ensure that algorithms are effective across a wide range of problems. In this context, Kumar et al. [30,85] have developed two benchmarks that cover nearly 50 diverse problems, aimed at facilitating more meaningful comparisons between algorithms and providing insights into their performance in real-world scenarios. Furthermore, the researchers conducted a comprehensive comparative analysis of state-of-the-art algorithms, culminating in a ranking system that serves as a valuable tool for evaluating the efficacy of new algorithmic proposals.

5. Methodologies and pathways for enhancing and refining existing metaheuristics

In the comprehensive review by Velasco et al. [18], it is revealed that 65% of the studies analyzed focus on improved versions of established algorithms, signaling a shift away from creating bioinspired algorithms based on novel analogies. This review examines over a hundred recent studies containing terms such as "new", "hybrid", or "improved optimization" in their titles. The vast majority of these proposals are actually refinements of existing solvers.

Unfortunately, the study in [18] underscores a relevant issue: most of these algorithms address engineering problems that lack direct applicability to real-world scenarios. Addressing this gap is the focus of the current section: to explore potential pathways for enhancing an existing bioinspired solver, making it more suitable for addressing specific problems or families of problems with greater relevance and utility. Such pathways are enumerated and described in what follows:

- *Automatic Equivalence between algorithms*: We begin with a common methodological approach, to add mechanisms or change existing ones for improved versions, thereby increasing the complexity of the model. When this is the case, ablation tests are highly recommended to assess the relative contribution of such added mechanisms and/or modifications to the acclaimed performance improvement.
  Given that ablation tests can be very time-consuming, several methodologies have recently emerged to automate this process. Among them, we highlight [86], which discusses how to automatically detect equivalence between stochastic population-based nature-inspired algorithms considering both a conceptual level – using the objective function value – and an operational level, expressed in terms of a measure of the population diversity. Their notion of equivalence requires that the average fitness values and the diversity of the population must not be significantly different in each generation for two algorithms under comparison to be equivalent to each other. Through the use of Markov chains as a representation of population and fitness, this methodology can assess the similarity of bioinspired algorithms.
- *Decomposition into their canonical components*: This is the approach followed in [87], which presented a methodology for the decomposition of bioinspired algorithms. Specifically, the methodology considers a set of templates that act as a framework to decompose and analyze bioinspired algorithms based on their characteristics and in a structured manner. In this pool of templates, we find different categories of components, such as the method of generating initial solutions, the solution pool itself, the solution archive, update mechanisms, input and output functions of the solution pool, and others. With all this, a similarity analysis between bioinspired algorithms can be performed by comparing their components within the pool of templates, facilitating not only the evaluation of algorithmic equivalence between them, but also a substrate of components for the automated adaptation and/or construction of meta-heuristic algorithms [88,89].
- *Evaluation of whether a metaheuristic algorithm is derived from an existing one*: Hu et al. [27] have recently introduced a novel methodology to evaluate the originality and research aim of bioinspired algorithms. This approach aims to discern whether a metaheuristic algorithm is a derivation of an existing method, based on the mathematical characterization of its compounding operators. Central to this methodology is the distinction between *root algorithm* and *homologous algorithm*, categorized by their search operators. A *root algorithm* represents a genuine innovation in the design of search operators, whereas a *homologous algorithm* is recognized for its practical utility or academic significance, even if it does not introduce fundamental algorithmic changes. The methodology proposed in this work defines two key constructs: perturbation mapping and difference mapping, which serve as formal representations of search operators. If a given operator of a bioinspired algorithm can be expressed as a simple

(especially linear) combination of these mappings derived from a *root algorithm*, the new algorithm is classified as *homologous*. This distinction allows researchers to determine whether an adaptation of a metaheuristic algorithm represents a true contribution or is an incremental variation without substantive novelty. To demonstrate its utility, the authors have applied this framework to several widely used algorithms, comparing them with their root counterparts.

- *Mathematically proven convergence improvements*: Choi et al. [90] introduces a mechanism to ensure the global convergence of swarm intelligence algorithms. This method involves applying stochastic perturbations to half of the swarm agents, followed by projecting all agents back into the feasible solution space to respect problem constraints. This projection is particularly critical in optimization problems with bounded search spaces. By incorporating controlled randomness, the perturbation mechanism helps maintain diversity within the swarm and prevents premature convergence by encouraging exploration in overly exploited areas. Remarkably, [90] provides mathematical proof that this approach, when integrated into a broad class of swarm algorithms, ensures global convergence under specific conditions.

- *Dynamic recalibration of metaheuristic algorithms*: Jia and Lu [91] proposes a novel update mechanism designed to dynamically recalibrate metaheuristic algorithms. The mechanism adjusts the balance between exploration and exploitation based on the current population distribution. It employs statistical analysis, calculating the standard deviation of agents' historical positions over recent generations, to measure population dispersion. This information is used to drive the algorithm, iteratively adapting its search behavior. The versatility and efficacy of this adaptation strategy are demonstrated with more than 50 different algorithms, demonstrating significant performance improvements across a diversity of solvers.

6. Prospects and cautions in automated design of metaheuristics

Advancing towards the automated design of metaheuristic algorithms has marked a significant leap in bioinspired and evolutionary optimization. This involves refining individual algorithms and automating their creation. The automated design of metaheuristic algorithms can provide breakthroughs in various scientific and engineering domains, making the automated design of metaheuristic algorithms the norm when addressing real-world optimization problems. However, it is important to note that while this approach is valid for creating well-functioning metaheuristic algorithms for a closed set of problems, it is indeed true that substantial additional experimentation and comparison against the previously mentioned baselines are required to validate its benefits in each real-world application.

In this section, we describe works that tackle this automated design process. First, in Section 6.1, we revisit several notable proposals to optimize the design. Next, Section 6.2 pauses at several recent contributions that incorporate Large Language Models (LLMs) for the design of new metaheuristic algorithms.

6.1. Automating the design of metaheuristic algorithms

Camacho-Villalón et al. [29] explore whether manually designing or automating the construction of metaheuristic algorithms can be a superior approach, yielding the following main observations:

- The manual design process often involves seeking inspiration from other fields of knowledge and iteratively refining algorithms based on trial and error. This approach, while effective, can be resource-intensive and time-consuming, as it requires substantial human involvement in identifying the optimal algorithmic design from a multitude of possibilities.

- In contrast, the automated design approach emphasizes the elimination of human intervention by leveraging recent advances in automatic algorithm configuration techniques. These techniques include hyper-heuristics, evolutionary programming, and reinforcement learning, which operate over the design space defined by algorithmic operators.

As further detailed in Zhao et al. [89], automated design methods can efficiently explore a vast space of algorithmic components, automatically selecting and combining operators in a way that optimizes performance. This not only accelerates the development of metaheuristic solvers but also offers the potential to discover novel algorithmic structures that may not be easily identified through manual design. By automating the creation process, researchers can focus more on applying these algorithms to real-world problems.

In general, the automatic development of metaheuristic algorithms involves two key steps:

1. The use of an automatic configuration tool to identify the optimal composition of operators and processing directives for a given algorithm. In this phase, the algorithm's performance is evaluated across various problem instances, and its configuration is adjusted iteratively until a predefined computational budget is met. Once this process is complete, the algorithm is considered to be fine-tuned with the best possible configuration.

2. The second step focuses on enabling the algorithm to be used for comparisons across different problem instances and configurations. To achieve this, a metaheuristic software framework is employed to define and implement the design space of the metaheuristic. This framework provides the infrastructure necessary to systematically explore various algorithmic components and their interactions. Subsequently, the framework is iterated with the previous step to automatically generate a fully functional algorithm capable of being tested against other problem configurations.

The development of software tools capable of supporting the creation, automatic generation, and evaluation of new algorithms represents one of the most promising advances in metaheuristic optimization research. Such tools can evaluate potential configurations and identify the best fit for a given problem. In [92], the authors introduce a novel software framework for creating new algorithms. Their software is built upon a standardized theoretical framework for analyzing, modifying, and generating metaheuristic algorithms. It features three distinct panels, each dedicated to a specific functionality. These panels enable users to configure continuous optimization problems and assemble new algorithms in a modular and code-free manner. The software framework provides access to over 110 benchmarking domains and includes robust statistical analysis, reporting key metrics such as fitness and runtime, alongside convergence plots to track algorithm progress.

Tools alike, in general, can help boost and democratize the development of new proposals for optimization algorithms.

6.2. Using large language models to automate the design of metaheuristic algorithms

Large Language Models [93] are artificial intelligence systems trained on vast amounts of data and are capable of processing and generating human-like text. LLMs can summarize texts, answer questions, translate languages, and even generate creative content. Recently, LLMs have been used to design and/or modify different algorithms. Thus, they can be considered an important and promising approach to automate the design of new evolutionary solvers. We herein comment on several recent examples of this trend.

A first example of this automatic approach is presented in van Stein and Bäck [94], where evolutionary Large Language Models (LLaMEA) are utilized to generate novel algorithms. Given a set of criteria and a task definition (search space), LLaMEA iteratively generates, mutates,
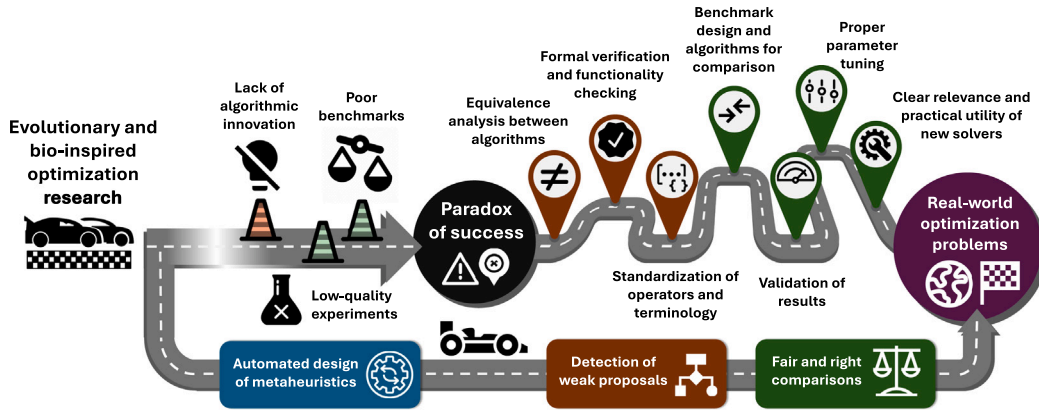
Fig. 1. Graphical summary of the issues, pathways and promising directions discussed in this manuscript.

and selects algorithms based on performance metrics and feedback from runtime evaluations. This methodology takes advantage of the capabilities of LLMs to create new metaheuristic algorithms that can then be evaluated and compared across a wide range of problem instances.

Another approach that harnesses contextual knowledge embedded in LLMs is FunSearch [95], which has represented a breakthrough in automated discovery of programs, hence having the potential to be applicable to generate and evolve metaheuristic solvers. By combining an LLM with an evolutionary framework, FunSearch systematically explores algorithmic design spaces, which can create novel optimization techniques that autonomously improve their performance through iterative refinement and runtime evaluation. Unlike traditional manual design, FunSearch can enable the discovery of unexpected and potentially superior bioinspired algorithms across different problem domains.

Liu et al. [96] have recently introduced the so-called Algorithm Evolution using the LLM (AEL) approach to designing optimization algorithms, with a specific focus on guided local search techniques. This framework leverages an LLM to generate evolutionary search operations, including initialization, crossover, and mutation, enabling the automated evolution of novel algorithmic solutions. AEL begins by prompting the LLM with a query that specifies the algorithm design task, information about the parent algorithms, and the expected format of the output. By harnessing the generative capabilities of the LLM, AEL reduces human involvement, and enables the discovery of new optimization methods.

The automated design of metaheuristic algorithms leveraging the rich knowledge embedded in LLMs represents a promising avenue for addressing the long-standing issue of solvers drawing inspiration from arbitrary or meaningless biological metaphors. However, to fully realize this potential, our pathways identified in previous sections are still required to ensure that the automated design of metaheuristic algorithms not only overcomes existing limitations, but also provides tangible practical value.

Injecting expert knowledge is crucial for solving real-world problems, and this is where LLMs can be especially useful by making domain expertise actionable. A promising application is their ability to translate textual descriptions of problem-specific features into concrete algorithmic improvements, enabling the design of memetic optimization algorithms tailored to the problem. This leads to two main scopes of application:

• Use LLMs to find a metaheuristic effective across diverse problems. However, results may not generalize beyond the studied set (due to the No Free Lunch theorem), so rigorous comparisons with well-tuned classical metaheuristics are essential.

• Use LLMs to design a metaheuristic suited to several instances of a particular problem. This requires (i) accounting for LLM inference costs in complexity comparisons, (ii) incorporating general-purpose and problem-specific algorithms as baselines, and (iii) evaluating generalization performance on unseen instances of the same problem.

Future research should rigorously evaluate if LLM-driven design of search algorithms provides a meaningful advantage over conventional approaches, considering solution quality, search efficiency, and the generalization capacity of the resulting algorithms. Such assessments must be grounded in principled methodological practices to ensure fair and informative comparisons with classical and ad-hoc solvers.

7. Conclusion and outlook

In this paper, which we graphically summarize in Fig. 1, we have critically examined challenges in evolutionary and bioinspired optimization, and outlined methodological ways to address them effectively. We have emphasized the need to differentiate truly innovative contributions (*the wheat in the chaff*), and have highlighted the importance of robust, non-biased experimental practices. Through recent examples in the literature, we have showcased strategies for the design, automated construction, improvement, and evaluation of bioinspired solvers, ensuring a more transparent assessment of the components of new proposals, and their fairer and more solid comparison to established metaheuristic methods.

The analyzed pathways are designed to refocus optimization research on its ultimate goal: developing algorithms capable of efficiently solving real-world problems. We strongly encourage the research community to adopt these recommendations as essential guidelines for achieving this objective. Researchers, editors, and reviewers must actively integrate the methodologies and best practices highlighted in this work to cultivate a more rigorous and innovative environment in evolutionary and bioinspired optimization.

Although the field has made significant progress, it still faces a long journey to realign itself and produce optimization and search tools capable of addressing major scientific and technological challenges across diverse disciplines. Ultimately, embracing these guidelines will significantly contribute to the advancement of knowledge and to the development of cutting-edge solutions to address real-world challenges and to redirect the field towards new metaheuristic solvers that deliver real practical value.

CRediT authorship contribution statement

Daniel Molina: Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Investigation, Conceptualization. **Javier Del Ser:** Writing – review & editing, Supervision,

Project administration, Methodology, Conceptualization. **Javier Poyatos:** Writing – original draft, Methodology, Investigation, Conceptualization. **Francisco Herrera:** Writing – review & editing, Supervision, Project administration, Methodology, Conceptualization.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used large-language models to improve the readability and language of the manuscript. After using this tool/service, the authors reviewed and edited the content as needed and assumed full responsibility for the content of the published article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Data availability

No data was used for the research described in the article.

References

[1] X.-S. Yang, Z. Cui, R. Xiao, A.H. Gandomi, M. Karamanoglu, Swarm Intelligence and Bio-Inspired Computation: Theory and Applications, Elsevier Science Publishers B. V., NLD, 2013, http://dx.doi.org/10.1016/B978-0-12-405163-8.00001-6.

[2] T. Back, D.B. Fogel, Z. Michalewicz, Handbook of Evolutionary Computation, IOP Publishing Ltd., New York, 1997, http://dx.doi.org/10.1201/9780367802486.

[3] A.K. Kar, Bio inspired computing – a review of algorithms and scope of applications, Expert. Syst. Appl. 59 (2016) 20–32, http://dx.doi.org/10.1016/j.ins.2014.01.026.

[4] X. Yao, Evolutionary Computation: Theory and Applications, World Scientific, 1999, http://dx.doi.org/10.1142/2792.

[5] D. Ashlock, Evolutionary Computation for Modeling and Optimization, vol. 571, Springer, New York, 2006, http://dx.doi.org/10.1007/0-387-31909-3.

[6] J. Del Ser, E. Osaba, D. Molina, X.-S. Yang, S. Salcedo-Sanz, D. Camacho, S. Das, P.N. Suganthan, C.A.C. Coello, F. Herrera, Bio-inspired computation: Where we stand and what's next, Swarm Evol. Comput. 48 (2019) 220–250, http://dx.doi.org/10.1016/j.swevo.2019.04.008.

[7] H. Al-Sahaf, Y. Bi, Q. Chen, A. Lensen, Y. Mei, Y. Sun, B. Tran, B. Xue, M. Zhang, A survey on evolutionary machine learning, J. R. Soc. N. Z. 49 (2) (2019) 205–228, http://dx.doi.org/10.1080/03036758.2019.1609052.

[8] Z.-H. Zhan, J.-Y. Li, J. Zhang, Evolutionary deep learning: A survey, Neurocomputing 483 (2022) 42–58, http://dx.doi.org/10.1016/j.neucom.2022.01.099.

[9] A.D. Martinez, J. Del Ser, E. Villar-Rodriguez, E. Osaba, J. Poyatos, S. Tabik, D. Molina, F. Herrera, Lights and shadows in evolutionary deep learning: Taxonomy, critical methodological analysis, cases of study, learned lessons, recommendations and challenges, Inf. Fusion 67 (2021) 161–194, http://dx.doi.org/10.1016/j.inffus.2020.10.014.

[10] I. Triguero, D. Molina, J. Poyatos, J. Del Ser, F. Herrera, General purpose artificial intelligence systems (GPAIS): properties, definition, taxonomy, societal implications and responsible governance, Inf. Fusion 103 (2024) 102135, http://dx.doi.org/10.1016/j.inffus.2023.102135.

[11] S. Chen, S. Chen, W. Hou, W. Ding, X. You, EGANS: Evolutionary generative adversarial network search for zero-shot learning, IEEE Trans. Evol. Comput. 28 (3) (2024) 582–596, http://dx.doi.org/10.1109/TEVC.2023.3307245.

[12] Q. Guo, et al., Connecting large language models with evolutionary algorithms yields powerful prompt optimizers, in: The Twelfth International Conference on Learning Representations, 2024, pp. 1–24, https://openreview.net/forum?id=ZG3RaNIsO8.

[13] J. Poyatos, D. Molina, A. Martínez-Seras, J. Del Ser, F. Herrera, Multiobjective evolutionary pruning of deep neural networks with transfer learning for improving their performance and robustness, Appl. Soft Comput. 147 (2023) 110757, http://dx.doi.org/10.1109/10.1016/j.asoc.2023.110757.

[14] J. Stork, A. Eiben, T. Bartz-Beielstein, A new taxonomy of global optimization algorithms, Nat. Comput. 21 (2020) 219–242, http://dx.doi.org/10.1007/s11047-020-09820-4.

[15] K. Rajwar, K. Deep, S. Das, An exhaustive review of the metaheuristic algorithms for search and optimization: Taxonomy, applications, and open challenges, Artif. Intell. Rev. 56 (11) (2023) 13187–13257, http://dx.doi.org/10.1007/s10462-023-10470-y.

[16] J. Ferrer, P. Delgado-Pérez, Metaheuristics in a nutshell, in: Optimising the Software Development Process with Artificial Intelligence, Springer, 2023, pp. 279–307, http://dx.doi.org/10.1007/978-981-19-9948-2_10.

[17] R. Martí, M. Sevaux, K. Sörensen, Fifty years of metaheuristics, In Press Eur. J. Oper. Res. (2024) 1–18, http://dx.doi.org/10.1016/j.ejor.2024.04.004.

[18] L. Velasco, H. Guerrero, A. Hospitaler, A literature review and critical analysis of metaheuristics recently developed, Arch. Comput. Methods Eng. 31 (2024) 125–146, http://dx.doi.org/10.1007/s11831-023-09975-0.

[19] D. Molina, J. Poyatos, J.D. Ser, S. García, A. Hussain, F. Herrera, Comprehensive taxonomies of nature-and bio-inspired optimization: Inspiration versus algorithmic behavior, critical analysis recommendations, Cogn. Comput. 12 (2020) 897–939, http://dx.doi.org/10.1007/s12559-020-09730-8.

[20] D. Molina, J. Poyatos, J.D. Ser, S. García, A. Hussain, F. Herrera, Comprehensive taxonomies of nature- and bio-inspired optimization: Inspiration versus algorithmic behavior, critical analysis recommendations, 2020, Last Update in 2024 at https://arxiv.org/abs/2002.08136.

[21] J.N. Hooker, Testing heuristics: we have it all wrong, J. Heuristics 1 (1) (1995) 33–42, http://dx.doi.org/10.1007/BF02430364.

[22] R.S. Barr, B.L. Golden, J.P. Kelly, M.G.C. Resende, W.R. Stewart, Designing and reporting on computational experiments with heuristic methods, J. Heuristics 1 (1) (1995) 9–32, http://dx.doi.org/10.1007/BF02430363.

[23] K. Sörensen, Metaheuristics—the metaphor exposed, Int. Trans. Oper. Res. 22 (1) (2015) 3–18, http://dx.doi.org/10.1111/itor.12001.

[24] A. Tzanetos, G. Dounias, Nature inspired optimization algorithms or simply variations of metaheuristics? Artif. Intell. Rev. 54 (3) (2021) 1841–1862, http://dx.doi.org/10.1007/s10462-020-09893-8.

[25] F. Campelo, C. Aranha, Lessons from the evolutionary computation bestiary, Artif. Life 29 (4) (2023) 421–432, http://dx.doi.org/10.1162/artl_a_00402.

[26] C. Aranha, C.L. Camacho Villalón, F. Campelo, M. Dorigo, R. Ruiz, M. Sevaux, K. Sörensen, T. Stützle, Metaphor-based metaheuristics, a call for action: the elephant in the room, Swarm Intell. 16 (2022) 1–6, http://dx.doi.org/10.1007/s11721-021-00202-9.

[27] Z. Hu, Q. Zhang, Y. Wang, Q. Su, Z. Xiong, Research orientation and novelty discriminant for new metaheuristic algorithms, Appl. Soft Comput. 157 (2024) 111521, http://dx.doi.org/10.1016/j.asoc.2024.111521.

[28] A. LaTorre, D. Molina, E. Osaba, J. Poyatos, J. Del Ser, F. Herrera, A prescription of methodological guidelines for comparing bio-inspired optimization algorithms, Swarm Evol. Comput. 67 (2021) 100973, http://dx.doi.org/10.1016/j.swevo.2021.100973.

[29] C.L. Camacho-Villalón, T. Stützle, M. Dorigo, Designing new metaheuristics: Manual versus automatic approaches, Intell. Comput. 2 (2023) 0048, http://dx.doi.org/10.34133/icomputing.0048.

[30] A. Kumar, G. Wu, M.Z. Ali, Q. Luo, R. Mallipeddi, P.N. Suganthan, S. Das, A benchmark-suite of real-world constrained multi-objective optimization problems and some baseline results, Swarm Evol. Comput. 67 (2021) 100961, http://dx.doi.org/10.1016/j.swevo.2021.100961.

[31] D. Weyland, A rigorous analysis of the harmony search algorithm: how the research community can be misled by a "novel" methodology, Int. J. Appl. Metaheuristic Comput. (IJAMC) 1 (2) (2010) 50–60, http://dx.doi.org/10.4018/jamc.2010040104.

[32] A.P. Piotrowski, J.J. Napiorkowski, P.M. Rowinski, How novel is the "novel" black hole optimization approach? Inf. Sci. 267 (2014) 191–200, http://dx.doi.org/10.1016/j.ins.2014.01.026.

[33] Z.W. Geem, Research commentary: Survival of the fittest algorithm or the novelest algorithm? The existence reason of the harmony search algorithm, in: Modeling, Analysis, and Applications in Metaheuristic Computing: Advancements and Trends, IGI Global, 2012, pp. 84–89, http://dx.doi.org/10.4018/jamc.2010010105.

[34] D. Weyland, A critical analysis of the harmony search algorithm—How not to solve sudoku, Oper. Res. Perspect. 2 (2015) 97–105, http://dx.doi.org/10.1016/j.orp.2015.04.001.

[35] O. Hasançebi, S. Ç.arbaş, E. Doğan, F. Erdal, M. Saka, Performance evaluation of metaheuristic search techniques in the optimum design of real size pin jointed structures, Comput. Struct. 87 (5) (2009) 284–302, http://dx.doi.org/10.1016/j.compstruc.2009.01.002.

[36] I. Fister Jr., U. Mlakar, J. Brest, I. Fister, A new population-based nature-inspired algorithm every month: is the current era coming to an end? in: Proceedings of the 3rd Student Computer Science Research Conference, 2016, pp. 33–37.

[37] L. Jiao, J. Zhao, C. Wang, X. Liu, F. Liu, L. Li, R. Shang, Y. Li, W. Ma, S. Yang, Nature-inspired intelligent computing: A comprehensive survey, Res. 7 (2024) 0442, http://dx.doi.org/10.34133/research.0442.

[38] L. Deng, S. Liu, Metaheuristics exposed: Unmasking the design pitfalls of arithmetic optimization algorithm in benchmarking, Appl. Soft Comput. 160 (2024) 111696, http://dx.doi.org/10.1016/j.asoc.2024.111696.

[39] M. Halsema, D. Vermetten, T. Bäck, N. Van Stein, A critical analysis of raven roost optimization, in: Proceedings of the Genetic and Evolutionary Computation Conference Companion, Association for Computing Machinery, 2024, pp. 1993–2001, https://dl.acm.org/doi/abs/10.1145/3638530.3664124.

[40] D. Wolpert, W. Macready, No free lunch theorems for optimization, IEEE Trans. Evol. Comput. 1 (1) (1997) 67–82, http://dx.doi.org/10.1109/4235.585893.

[41] D. Wolpert, W. Macready, Coevolutionary free lunches, IEEE Trans. Evol. Comput. 9 (6) (2005) 721–735, http://dx.doi.org/10.1109/TEVC.2005.856205.

[42] D. Whitley, A free lunch proof for gray versus binary encodings, in: Proceedings of the 1st Annual Conference on Genetic and Evolutionary Computation - Volume 1, in: GECCO'99, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1999, pp. 726–733.

[43] C. Igel, M. Toussaint, A no-free-lunch theorem for non-uniform distributions of target functions, J. Math. Model. Algorithms 3 (4) (2005) 313–322, http://dx.doi.org/10.1007/s10852-005-2586-y.

[44] D. Whitley, Functions as permutations: regarding no free lunch, walsh analysis and summary statistics, in: M. Schoenauer, K. Deb, G. Rudolph, X. Yao, E. Lutton, J.J. Merelo, H.-P. Schwefel (Eds.), Parallel Problem Solving from Nature PPSN VI, Springer, Berlin, Heidelberg, 2000, pp. 169–178, http://dx.doi.org/10.1007/3-540-45356-3_17.

[45] D. Whitley, Sharpened and focused no free lunch and complexity theory, in: E.K. Burke, G. Kendall (Eds.), Search Methodologies: Introductory Tutorials in Optimization and Decision Support Techniques, Springer US, Boston, MA, 2014, pp. 451–476, http://dx.doi.org/10.1007/978-1-4614-6940-7_16.

[46] C. Igel, No free lunch theorems: limitations and perspectives of metaheuristics, in: Y. Borenstein, A. Moraglio (Eds.), Theory and Principled Methods for the Design of Metaheuristics, Springer, Berlin, Heidelberg, 2014, pp. 1–23, http://dx.doi.org/10.1007/978-3-642-33206-7_1.

[47] J. McDermott, When and why metaheuristics researchers can ignore "no free lunch" theorems, SN Comput. Sci. 1 (1) (2020) 60, http://dx.doi.org/10.1007/s42979-020-0063-3.

[48] C. García-Martínez, F.J. Rodriguez, M. Lozano, Arbitrary function optimisation with metaheuristics, Soft Comput. 16 (12) (2012) 2115–2133, http://dx.doi.org/10.1007/s00500-012-0881-x.

[49] A.P. Piotrowski, J.J. Napiorkowski, A.E. Piotrowska, Metaheuristics should be tested on large benchmark set with various numbers of function evaluations, Swarm Evol. Comput. 92 (2025) 101807, http://dx.doi.org/10.1016/j.swevo.2024.101807.

[50] P. Bonissone, R. Subbu, N. Eklund, T. Kiehl, Evolutionary algorithms + domain knowledge=real-world evolutionary computation, IEEE Trans. Evol. Comput. 10 (3) (2006) 256–280, http://dx.doi.org/10.1109/TEVC.2005.857695.

[51] F. Neri, C. Cotta, Memetic algorithms and memetic computing optimization: A literature review, Swarm Evol. Comput. 2 (2012) 1–14, http://dx.doi.org/10.1016/j.swevo.2011.11.003.

[52] D. Molina, A. LaTorre, F. Herrera, An insight into bio-inspired and evolutionary algorithms for global optimization: Review, analysis, and lessons learnt over a decade of competitions, Cogn. Comput. 10 (4) (2018) 517–544, http://dx.doi.org/10.1007/s12559-018-9554-0.

[53] D. Molina, F. Moreno-García, F. Herrera, Analysis among winners of different IEEE CEC competitions on real-parameters optimization: is there always improvement? in: 2017 IEEE Congress on Evol. Comput. (CEC), 2017, pp. 805–812, http://dx.doi.org/10.1109/CEC.2017.7969392.

[54] A.P. Piotrowski, J.J. Napiorkowski, A.E. Piotrowska, Choice of benchmark optimization problems does matter, Swarm Evol. Comput. 83 (2023) 101378, http://dx.doi.org/10.1016/j.swevo.2023.101378.

[55] A.P. Piotrowski, J.J. Napiorkowski, Some metaheuristics should be simplified, Inf. Sci. 427 (2018) 32–62, http://dx.doi.org/10.1016/j.ins.2017.10.039.

[56] K. Rajwar, K. Deep, Structural bias in metaheuristic algorithms: insights, open problems, and future prospects, Swarm Evol. Comput. 92 (2025) 101812, http://dx.doi.org/10.1016/j.swevo.2024.101812.

[57] J. Kudela, A critical problem in benchmarking and analysis of evolutionary computation methods, Nat. Mach. Intell. 4 (12) (2022) 1238–1245, http://dx.doi.org/10.1038/s42256-022-00579-0.

[58] E. Osaba, E. Villar-Rodriguez, J. Del Ser, A.J. Nebro, D. Molina, A. LaTorre, P.N. Suganthan, C.A. Coello Coello, F. Herrera, A tutorial on the design, experimentation and application of metaheuristic algorithms to real-world optimization problems, Swarm Evol. Comput. 64 (2021) 100888, http://dx.doi.org/10.1016/j.swevo.2021.100888.

[59] J. Ceberio, B. Calvo, Time to stop and think: What kind of research do we want to do? 2024, https://arxiv.org/abs/2402.08298.

[60] K. van der Blom, T.M. Deist, V. Volz, M. Marchi, Y. Nojima, B. Naujoks, A. Oyama, T. Tušar, Identifying properties of real-world optimisation problems through a questionnaire, in: Many-Criteria Optimization and Decision Analysis: State-of-the-Art, Present Challenges, and Future Perspectives, Springer International Publishing, Cham, 2023, pp. 59–80, http://dx.doi.org/10.1007/978-3-031-25263-1_3.

[61] M. Hellwig, H.-G. Beyer, Benchmarking evolutionary algorithms for single objective real-valued constrained optimization – a critical review, Swarm Evol. Comput. 44 (2019) 927–944, http://dx.doi.org/10.1016/j.swevo.2018.10.002.

[62] O. Mersmann, M. Preuss, H. Trautmann, Benchmarking evolutionary algorithms: towards exploratory landscape analysis, in: Parallel Problem Solving from Nature, PPSN XI, 2010, pp. 73–82, http://dx.doi.org/10.1007/978-3-642-15844-5_8.

[63] H. Yin, D. Vermetten, F. Ye, T.H.W. Bäck, A.V. Kononova, Impact of spatial transformations on landscape features of CEC2022 basic benchmark problems, 2024, https://arxiv.org/abs/2402.07654.

[64] C. Chen, Q. Liu, Y. Jing, M. Zhang, S. Cheng, Y. Li, On the representativeness metric of benchmark problems in numerical optimization, Swarm Evol. Comput. 91 (2024) 101716, http://dx.doi.org/10.1016/j.swevo.2024.101716.

[65] A.P. Piotrowski, J.J. Napiorkowski, A.E. Piotrowska, Metaheuristics should be tested on large benchmark set with various numbers of function evaluations, Swarm Evol. Comput. 92 (2025) 101807, http://dx.doi.org/10.1016/j.swevo.2024.101807.

[66] J. Pickard, J. Carretero, V. Bhavsar, On the convergence and origin bias of the teaching-learning-based-optimization algorithm, Appl. Soft Comput. 46 (2016) 115–127, http://dx.doi.org/10.1016/j.asoc.2016.04.029.

[67] M. Gauci, T.J. Dodd, R. Groß, Why "GSA: a gravitational search algorithm" is not genuinely based on the law of gravity, Nat. Comput. 11 (2012) 719–720, http://dx.doi.org/10.1007/s11047-012-9322-0.

[68] C.L. Camacho Villalón, T. Stützle, M. Dorigo, Grey wolf, firefly and bat algorithms: Three widespread algorithms that do not contain any novelty, in: M. Dorigo, et al. (Eds.), International Conference on Swarm Intelligence, 12421, Springer, 2020, pp. 121–133, http://dx.doi.org/10.1007/978-3-030-60376-2_10.

[69] C.L. Camacho-Villalón, M. Dorigo, T. Stützle, Why the intelligent water drops cannot be considered as a novel algorithm, in: M. Dorigo, et al. (Eds.), International Conference on Swarm Intelligence, Springer, 2018, pp. 302–314, http://dx.doi.org/10.1007/978-3-030-00533-7_24.

[70] C.L. Camacho-Villalón, M. Dorigo, T. Stützle, The intelligent water drops algorithm: why it cannot be considered a novel algorithm: A brief discussion on the use of metaphors in optimization, Swarm Intell. 13 (2019) 173–192, http://dx.doi.org/10.1007/s11721-019-00165-y.

[71] C.L. Camacho-Villalón, M. Dorigo, T. Stützle, An analysis of why cuckoo search does not bring any novel ideas to optimization, Comput. Oper. Res. 142 (2022) 105747, http://dx.doi.org/10.1016/j.cor.2022.105747.

[72] C.L. Camacho-Villalón, M. Dorigo, T. Stützle, Exposing the grey wolf, moth-flame, whale, firefly, bat, and antlion algorithms: six misleading optimization techniques inspired by bestial metaphors, Int. Trans. Oper. Res. 30 (6) (2023) 2945–2971, http://dx.doi.org/10.1111/itor.13176.

[73] M. Castelli, L. Manzoni, L. Mariot, M.S. Nobile, A. Tangherloni, Salp swarm optimization: A critical review, Expert. Syst. Appl. 189 (2022) 116029, http://dx.doi.org/10.1016/j.eswa.2021.116029.

[74] L. Deng, S. Liu, Deficiencies of the whale optimization algorithm and its validation method, Expert. Syst. Appl. 237 (2024) 121544, http://dx.doi.org/10.1016/j.eswa.2023.121544.

[75] A. Tzanetos, Does the field of nature-inspired computing contribute to achieving lifelike features? Artif. Life 29 (4) (2023) 487–511, http://dx.doi.org/10.1162/artl_a_00407.

[76] J. Kudela, The evolutionary computation methods no one should use, 2023, https://arxiv.org/abs/2301.01984.

[77] C. Urban, A. Miné, A review of formal methods applied to machine learning, 2021, http://dx.doi.org/10.48550/arXiv.2104.02466, arXiv:2104.02466.

[78] J. Swan, et al., A research agenda for metaheuristic standardization, in: MIC 2015: The XI Metaheuristics International Conference, 2015.

[79] J. Swan, S. Adriaensen, A.E. Brownlee, K. Hammond, C.G. Johnson, A. Kheiri, F. Krawiec, J. Merelo, L.L. Minku, E. ozcan, G.L. Pappa, P. García-Sanchez, K. Sorensen, S. Vob, M. Wagner, D.R. White, Metaheuristics "in the large", European J. Oper. Res. 297 (2) (2022) 393–406, http://dx.doi.org/10.1016/j.ejor.2021.05.042.

[80] J. Del Ser, E. Osaba, A.D. Martinez, M.N. Bilbao, J. Poyatos, D. Molina, F. Herrera, More is not always better: insights from a massive comparison of meta-heuristic algorithms over real-parameter optimization problems, in: 2021 IEEE Symposium Series on Computational Intelligence (SSCI), 2021, pp. 1–7, http://dx.doi.org/10.1109/SSCI50451.2021.9660030.

[81] J. Derrac, S. García, D. Molina, F. Herrera, A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms, Swarm Evol. Comput. 1 (1) (2011) 3–18, http://dx.doi.org/10.1016/j.swevo.2011.02.002.

[82] J. Carrasco, S. García, M. Rueda, S. Das, F. Herrera, Recent trends in the use of statistical tests for comparing swarm and evolutionary computing algorithms: Practical guidelines and a critical review, Swarm Evol. Comput. 54 (2020) 100665, http://dx.doi.org/10.1016/j.swevo.2020.100665.

[83] P. Sharma, S. Raju, Metaheuristic optimization algorithms: A comprehensive overview and classification of benchmark test functions, Soft. Comput. 28 (4) (2024) 3123–3186, http://dx.doi.org/10.1007/s00500-023-09276-5.

[84] A. Walden, M. Buzdalov, A simple statistical test against origin-biased metaheuristics, in: S. Smith, J.a. Correia, C. Cintrano (Eds.), International Conference on the Applications of Evolutionary Computation, Springer Nature Switzerland, 2024, pp. 322–337, http://dx.doi.org/10.1007/978-3-031-56852-7_21.

[85] A. Kumar, G. Wu, M.Z. Ali, R. Mallipeddi, P.N. Suganthan, S. Das, A test-suite of non-convex constrained optimization problems from the real-world and some baseline results, Swarm Evol. Comput. 56 (2020) 100693, http://dx.doi.org/10.1016/j.swevo.2020.100693.

[86] I. Fister, I. Fister, A. Iglesias, A. Galvez, On detecting the novelties in metaphor-based algorithms, in: F. Chicano (Ed.), Proceedings of the Genetic and Evolutionary Computation Conference Companion, Association for Computing Machinery, 2021, pp. 71–72, http://dx.doi.org/10.1145/3449726.345941.

[87] J. de Armas, E. Lalla-Ruiz, S.L. Tilahun, S. Voß, Similarity in metaheuristics: A gentle step towards a comparison methodology, Nat. Comput. 21 (2) (2022) 265–287, http://dx.doi.org/10.1007/s11047-020-09837-9.

[88] T. Stützle, M. López-Ibáñez, Automated design of metaheuristic algorithms, in: Handbook of metaheuristics, Springer, 2019, pp. 541–579, http://dx.doi.org/10.1007/978-3-319-91086-4_17.

[89] Q. Zhao, Q. Duan, B. Yan, S. Cheng, Y. Shi, Automated design of metaheuristic algorithms: A survey, 2023, http://dx.doi.org/10.48550/arXiv.2303.06532, arXiv preprint arXiv:2303.06532.

[90] K.P. Choi, E.H.H. Kam, X.T. Tong, W.K. Wong, Appropriate noise addition to metaheuristic algorithms can enhance their performance, Sci. Rep. 13 (1) (2023) 5291, http://dx.doi.org/10.1038/s41598-023-29618-5.

[91] H. Jia, C. Lu, Guided learning strategy: A novel update mechanism for meta-heuristic algorithms design and improvement, Knowl. Based Syst. 286 (2024) 111402, http://dx.doi.org/10.1016/j.knosys.2024.111402.

[92] D. Acosta-Ugalde, J.M. Cruz-Duarte, S.E. Conant-Pablos, J.G. Falcon-Cardona, Beyond "novel"' metaphor-based metaheuristics: An interactive algorithm design software, in: 2024 IEEE Congress on Evolutionary Computation, CEC, 2024, pp. 1–8, http://dx.doi.org/10.1109/CEC60901.2024.10611912.

[93] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, in: Advances in Neural Information Processing Systems, vol. 26, 2013, pp. 3111–3119.

[94] N. van Stein, T. Bäck, LLaMEA: A large language model evolutionary algorithm for automatically generating metaheuristics, 2024, https://arxiv.org/abs/2405.20132.

[95] B. Romera-Paredes, M. Barekatain, A. Novikov, M. Balog, M.P. Kumar, E. Dupont, F.J. Ruiz, J.S. Ellenberg, P. Wang, O. Fawzi, et al., Mathematical discoveries from program search with large language models, Nature 625 (7995) (2024) 468–475, http://dx.doi.org/10.1038/s41586-023-06924-6.

[96] F. Liu, X. Tong, M. Yuan, X. Lin, F. Luo, Z. Wang, Z. Lu, Q. Zhang, An example of evolutionary computation + large language model beating human: Design of efficient guided local search, 2024, http://dx.doi.org/10.48550/arXiv.2401.02051, arXiv preprint arXiv:2401.02051.