*applied sciences*

MDPI

*Article*

# Biases in AI-Supported Industry 4.0 Research: A Systematic Review, Taxonomy, and Mitigation Strategies

**Javier Arévalo-Royo [1], Francisco-Javier Flor-Montalvo [2], Juan-Ignacio Latorre-Biel [1], Emilio Jiménez-Macías [3], Eduardo Martínez-Cámara [4,*] and Julio Blanco-Fernández [4]**

[1] Institute of Smart Cities (ISC), Public University of Navarre, 31006 Pamplona, Spain; arevalo.158423@e.unavarra.es (J.A.-R.); juanignacio.latorre@unavarra.es (J.-I.L.-B.)
[2] Higher School of Engineering and Technology, International University of La Rioja (UNIR), 26004 Logroño, Spain; franciscojavier.flor@unir.net
[3] Department of Electrical Engineering, University of La Rioja, 26004 Logroño, Spain; emilio.jimenez@unirioja.es
[4] Department of Mechanical Engineering, University of La Rioja, 26004 Logroño, Spain; julio.blanco@unirioja.es
* Correspondence: eduardo.martinezc@unirioja.es; Tel.: +34-941-299-533

**Abstract**

Industrial engineering research has been reshaped by the integration of artificial intelligence (AI) within the framework of Industry 4.0, characterized by the interplay between cyber-physical systems (CPS), advanced automation, and the Industrial Internet of Things (IIoT). While this integration opens new opportunities, it also introduces biases that undermine the reliability and robustness of scientific and industrial outcomes. This article presents a systematic literature review (SLR), supported by natural language processing techniques, aimed at identifying and classifying biases in AI-driven research within industrial contexts. Based on this meta-research approach, a taxonomy is proposed that maps biases across the stages of the scientific method as well as the operational layers of intelligent production systems. Statistical analysis confirms that biases are unevenly distributed, with a higher incidence in hypothesis formulation and results dissemination. The study also identifies emergent AI-related biases specific to industrial applications such as predictive maintenance, quality control, and digital twin management. Practical implications include stronger reliability in predictive analytics for manufacturers, improved accuracy in monitoring and rescue operations through transparent AI pipelines, and enhanced reproducibility for researchers across stages. Mitigation strategies are then discussed to safeguard research integrity and support trustworthy, bias-aware decision-making in Industry 4.0.

**Keywords:** industrial AI; CPS; IIoT; digital twin; predictive maintenance; quality control

check for updates

Academic Editors: Filipe Pereira, Carlos Felgueiras, Vítor Carvalho and Pedro M. B. Torres

Received: 16 September 2025
Revised: 6 October 2025
Accepted: 9 October 2025
Published: 11 October 2025

**Citation:** Arévalo-Royo, J.; Flor-Montalvo, F.-J.; Latorre-Biel, J.-I.; Jiménez-Macías, E.; Martínez-Cámara, E.; Blanco-Fernández, J. Biases in AI-Supported Industry 4.0 Research: A Systematic Review, Taxonomy, and Mitigation Strategies. *Appl. Sci.* **2025**, *15*, 10913. https://doi.org/10.3390/app152010913

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

The practice of industrial engineering rests on scientific principles, whose understanding forms the foundation for technological development in production environments [1]. The scientific method, grounded in observation, hypothesis formulation, experimentation, and analysis of results, remains the core mechanism for knowledge generation in engineering [2]. Yet, the current immersion in Industry 4.0 has substantially altered the framework in which these principles are applied. The availability of massive datasets produced by cyber-physical systems and IIoT platforms, together with computational simulation, machine learning (ML), and other AI techniques, has become central to research,

development, and industrial practice. These capabilities enable the anticipation of complex process behavior and the optimization of experimental design, while simultaneously reducing the costs associated with empirical validation [3]. Moreover, the automation of data acquisition through smart sensors and real-time analytics has become essential for production management in smart factories, enhancing accuracy and reliability in applications such as equipment monitoring and adaptive decision-making [4]. The present study is not intended as a product design, process simulation, or industrial case analysis. It is conceived as meta-research, drawing on a systematic literature review enriched with natural language processing techniques. This design makes it possible to organize scattered findings, detect recurrent and novel forms of bias, and propose practical measures to reduce their impact on Industry 4.0 research. A systematic literature review supported by natural language processing and co-occurrence analysis was employed. This approach enables the synthesis of dispersed findings and the detection of bias patterns across a broad corpus of studies, capacities not attainable through simulation or isolated experimental trials. The combination of SLR and NLP has been successfully applied in recent meta-research to map methodological trends [5]

The analysis of industrial data has likewise been profoundly shaped by advanced AI models capable of identifying latent patterns within large-scale information streams. Such techniques have expanded inference capabilities and facilitated the detection of complex correlations that would remain unnoticed through conventional analysis [6]. Nevertheless, the inherent opacity of certain AI models raises concerns regarding the interpretability of outcomes and the potential propagation of biases introduced during algorithmic training [7].

Within the scope of Industry 4.0, these systems are employed not only to synthesize information from multiple sensors but also to generate automated reports supporting real-time decision-making. In parallel, AI enables the automatic production of technical summaries [8] and the identification of usage patterns among operators and production managers, tasks managed by recommendation algorithms similar to those used in scientific communication [9]. However, the growing automation of these functions raises questions about the authenticity of result interpretation and the potential homogenization of decision-making within complex industrial contexts [10].

This gives rise to several central questions: What are the defined stages of the engineering research process, and in which phases does the application of AI become especially decisive in Industry 4.0 environments? To what extent does the integration of these systems represent a genuine scientific and technological paradigm shift? Under what specific conditions does their use affect the reliability and reproducibility of results obtained in smart factories or cyber-physical systems? What measures can be implemented to counteract adverse effects stemming from algorithmic opacity or the propagation of biases? While relevant contributions exist that partially address these issues, comprehensive studies offering a global and systematized view of the state of the art—particularly in the context of Industry 4.0—remain limited.

Table 1 summarizes the main constructs examined in prior work, the representative sources addressing them, and the specific gaps that this study aims to fill.

Mehdiyev et al. [11] argue that AI-based methods, though useful for detecting correlations in the large volumes of data generated by industrial systems, do not necessarily yield conceptual knowledge about the underlying processes. This view aligns with the conclusions of Astleitner [12], who stresses that a considerable share of reported findings in scientific literature lacks robust validity. Complementing this perspective, Popper [13] warned that many studies merely reproduce biases already embedded in earlier research

on the same topic, with the value of their conclusions largely depending on the proportion of genuine versus spurious relationships within each field.

**Table 1.** Research constructs reviewed in prior studies and corresponding gaps addressed in this work.

| Construct | Gap Addressed in This Study |
| --- | --- |
| Bias typologies and definitions | Provides a unified taxonomy structured across research stages in Industry 4.0. |
| Phase-specific distribution of biases | Delivers quantitative mapping of bias occurrence across all stages. |
| Emergent AI-related biases in CPS/IIoT | Identifies and formalizes ten novel biases specific to industrial AI contexts. |
| Methodological transparency and explainability | Links explainability challenges to phase-specific manifestations of bias. |
| Human oversight and governance | Outlines practical oversight mechanisms aligned with each stage. |
| Mitigation strategies | Consolidates phase-tailored strategies into an operational framework. |
| Paradigm and epistemic framing | Connects claims of paradigm shift to observed bias patterns. |

Similarly, Lamata [14] cautioned that excessive reliance on Big Data without a supporting theoretical framework can lead to spurious correlations, flawed conclusions, and inflated expectations regarding the predictive capacity of AI. The mere accumulation of data cannot substitute the traditional scientific method based on hypothesis formulation and mathematical modeling, since correlation does not imply causation. Furthermore, data quality issues in engineering research tend to be amplified rather than corrected when the volume of information is increased [15]. This purported new paradigm, in which AI supported by Big Data shapes the dynamics of research and development in industrial engineering, offers significant opportunities but also entails risks if accepted without critical scrutiny [16].

In relation to the hypothesis that the incorporation of AI into research may represent a scientific paradigm shift, Kuhn [17] defined a paradigm as the set of assumptions, values, and techniques that guide the work of a scientific community during a given period. Paradigm shifts are not cumulative; they arise when the prevailing framework accumulates unresolved anomalies and become consolidated when new methodologies or technologies prove more fruitful. From this perspective, if the application of AI in industrial engineering indeed constitutes such a transformation, one must ask which mechanisms prompt researchers to misinterpret results produced by algorithmic systems [18].

Horvitz et al. [19] emphasize that the development of AI requires the integration of principles ensuring comprehensibility, reliability, and auditability, thereby preserving human oversight in critical stages of research and industrial decision-making while minimizing bias. AI may automate specific tasks, but its outputs must remain subject to human scrutiny, as they are not exempt from either inherent or inherited biases.

Within this framework, cognitive biases can be defined as systematic tendencies in information processing that distort perception, judgment, and decision-making, often without the researcher's awareness or under the influence of emotional factors [20].

In this study, the term bias is understood in its broad scientific sense as a systematic deviation that affects either the research method itself or the interpretation of the data derived from it. This includes both cognitive biases inherent to human reasoning and methodological biases introduced by the design, training, or deployment of AI systems. Importantly, biases are not treated here as isolated errors or random noise, but as structural distortions that can persist throughout the research process, shaping problem formulation,

experimental design, data analysis, and interpretation of results. This definition establishes the conceptual foundation upon which the taxonomy and mitigation strategies proposed in this work are built. Such biases affect industrial engineering research by undermining objectivity in hypothesis formulation, data analysis, and result interpretation. Their study in the scientific literature allows the identification of recurrent patterns that shape the construction of knowledge [21]. Recognizing and defining them is indispensable for designing effective corrective strategies.

The present investigation is based on the hypothesis that the use of AI in engineering research within Industry 4.0 constitutes a new paradigm, and that researchers employing these tools are exposed both to common cognitive biases and to those stemming from the very nature of AI [22]. Moreover, it is argued that the appearance of such biases across the various stages of the research process does not follow a homogeneous pattern. The interaction between human and AI-related biases may give rise to new forms of distortion which, although generally acknowledged, have not yet been sufficiently identified [23]. Analyzing and understanding these emerging biases is thus a necessary step toward the design of effective mitigation strategies aimed at reducing their impact and ensuring the integrity and quality of the research process in the field under study.

To guide the reader through the forthcoming sections, the remainder of the article is structured as follows: Section 2 details the methodological framework and the systematic literature review process adopted in this study. Section 3 presents the results of the analysis, including the taxonomy of 100 biases and the identification of emerging AI-related biases in CPS and IIoT contexts. Section 4 discusses the findings, situating them within the broader body of related research, and outlines their practical implications, limitations, and avenues for future inquiry. Finally, Section 5 summarizes the main conclusions and highlights the contribution of this work to advancing the understanding and mitigation of biases in Industry 4.0 research.

## 2. Materials and Methods

In line with well-established contributions in the field, this study assumes that the engineering research process can be structured into seven incremental stages (Figure 1). This sequential scheme has been adopted in the present meta-research and also guides the organization of the article. The first stage corresponds to the identification of the research problem or question, which marks the starting point through the formulation of a clear and precise inquiry. Within the context of Industry 4.0, this initial stage is often linked to challenges arising in cyber-physical systems, the Industrial Internet of Things (IIoT), or the integration of digital twins—domains in which a proper formulation of the research question is decisive for directing both the collection and processing of large-scale data.

In addition to identifying the main stages of the research process, it is essential to consider the methodological contexts in which cognitive and AI-related biases are most frequently observed. The reviewed literature encompasses a wide spectrum of study designs, including experimental research in industrial environments, simulation-based investigations focused on cyber-physical production systems, large-scale data analytics on IIoT platforms, and systematic literature reviews addressing theoretical and methodological aspects. Each of these contexts exhibits distinct exposure profiles to bias: for example, experimental studies tend to manifest experimenter and confirmation biases, simulations often amplify algorithmic opacity and automation biases, while meta-analyses and literature reviews are particularly sensitive to database filtering and knowledge homogenization biases. Recognizing this diversity of methodological scenarios is fundamental to correctly interpreting the origin, distribution, and impact of the biases identified in the subsequent analysis.

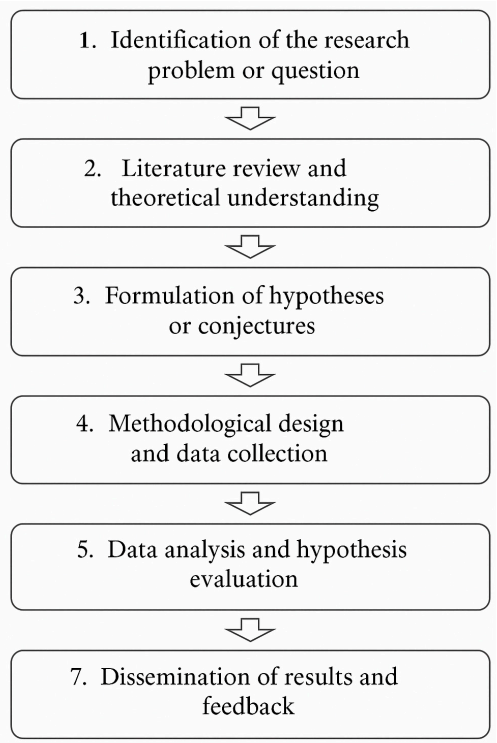


**Figure 1.** Stages of scientific research.

To conduct a systematic and comprehensive analysis of cognitive biases in engineering research, the premise was established that such biases, in their initial manifestation, are comparable to those observed in other scientific domains. Consequently, a broad literature review was carried out using the SCOPUS database. The search strategy in SCOPUS was designed to retrieve peer-reviewed scientific articles addressing different dimensions of research bias. To ensure extensive coverage and avoid premature restrictions in study selection, a deliberately wide-ranging query was used, incorporating terms such as "cognitive biases," "confirmation bias," "selection bias," "experimenter bias," "ethical issues in research," "scientific misconduct," "replication crisis," "research integrity," and "peer review bias," with results restricted to academic journal publications. This search was executed on 10 February 2025, yielding 132,961 documents, which constitute the initial corpus underlying the analysis of bias and research integrity.

To summarize the content of such a large corpus and identify biases recurrently addressed within it, traditional review and critical reading techniques were combined with natural language processing (NLP) methods, thereby extending the limitations of human reading capacity [24]. Specifically, the graph-based TextRank algorithm was adapted using the pytextrank v3.3.0 library, integrated with spaCy [25]. This approach enabled the extraction of representative sentences from each document and the generation of synthetic summaries capturing the semantic essence of the articles.

To ensure the reliability of the keyword extraction and summarization process, a manual validation procedure was conducted on a representative sample of the retrieved corpus. Specifically, 200 documents (approximately 0.15% of the total) were randomly selected and independently reviewed by two researchers. The extracted keywords and representative sentences identified by the NLP pipeline were compared against manually annotated references, obtaining an agreement rate of 91.4%. Discrepancies were discussed and resolved jointly, and the resulting validation guided the fine-tuning of the TextRank parameters and the filtering thresholds applied in the subsequent analysis.

For the identification of recurring biases, a lexical co-occurrence analysis was performed, assessing the frequency with which terms associated with biases and fallacies appeared across different sections of the texts. For this purpose, custom developments in Python were combined with the open-source library Gensim v4.3.3, which supports the automated detection of relationships through a similarity graph between sentences based on the co-occurrence of key terms. All computational outputs were continuously supervised and validated through the qualitative and critical analysis of the human researcher.

Drawing on both the quantitative interpretation of these data and the manual review of the most cited papers, the taxonomy corresponding to the defined research method was populated. The identified biases were grouped according to the conceptual similarity of their descriptions in the analyzed articles, thereby establishing their potential incidence across the various stages of the research process.

## 3. Results

Two sets of results can be distinguished:

1. Those obtained through a quantitative approach based on data processing with advanced computational tools.
2. Those derived from a detailed manual review of articles specifically addressing AI within the analyzed corpus.

### *3.1. Data Analysis and Hypothesis Evaluation in Industry 4.0*

A total of 100 recurrent biases were identified across the different stages of the defined research process. The identification process combined automated text-mining and manual review. A corpus of over 130,000 documents retrieved from SCOPUS was processed with TextRank to extract candidate terms, followed by co-occurrence analysis using Gensim to group related concepts. Two independent reviewers manually validated a representative subset of 200 articles, reaching agreement above 90%. This hybrid pipeline ensured both scalability and reliability, consistent with recent work employing NLP to systematize bias detection in scientific research.

The corresponding table (Table 2) lists them in alphabetical order, indicating the stage or stages of the investigative process in which they tend to manifest with greater intensity, and providing in each case a bibliographic reference that elaborates on their definition and impact. This classification forms the basis for examining how biases influence both the formulation of hypotheses and the interpretation of results—dimensions that are particularly sensitive in Industry 4.0 environments, where decision-making relies on cyber-physical systems and the automated processing of large volumes of data.

**Table 2.** Bias, phase, reference.

| Name | Phase | Ref. |
|---|---|---|
| Actor-observer bias | 1,3,5,7 | [26] |
| Ad hominem | 7 | [27] |
| Ambiguity effect | 3,4,5 | [28] |
| Anchoring effect | 2,3,5 | [29] |
| Argument from ignorance | 3,5,6 | [30] |
| Attentional bias | 2,4,5 | [31] |
| Authority bias | 2,5,6,7 | [32] |
| Availability cascade | 2,3,7 | [33] |
| Availability heuristic | 2,3,5 | [34] |
| Backfire effect | 6,7 | [35] |
| Bandwagon effect | 1,2,3,7 | [36] |
| Base rate fallacy | 3,5 | [37] |

**Table 2.** *Cont.*

| Name | Phase | Ref. |
|---|---|---|
| Base rate neglect | 2,3,5 | [38] |
| Belief bias | 2,3,4,7 | [39] |
| Black sheep effect | 6,7 | [40] |
| Blind spot | 2,7 | [41] |
| Bystander effect | 7 | [42] |
| Cherry Picking | 1,2,3,4,5,6,7 | [43] |
| Clustering illusion | 2,3,5 | [44] |
| Cognitive dissonance | 3, 5, 6 | [45] |
| Cognitive fluency bias | 7 | [46] |
| Confirmation bias | 1,2,3,4,5,6,7 | [47] |
| Conservatism bias | 2,5,6 | [48] |
| Context effect | 4,5,7 | [49] |
| Contrast effect | 2,5,7 | [50] |
| Correlation-causation | 5,6 | [51] |
| Cryptomnesia or false memories | 2,3,7 | [52] |
| Cultural | 2,3,5,6,7 | [53] |
| Curse of knowledge | 7 | [54] |
| Declinism | 2,6,7 | [55] |
| Defensive attribution | 4,5,6 | [56] |
| Distinction bias | 3,5 | [57] |
| Dunning-Kruger effect | 2,3,5 | [58] |
| Endowment effect | 3,4,7 | [59] |
| Escalation of commitment | 4,5,6 | [60] |
| Essentialism fallacy | 2,3,6 | [61] |
| Experimenter bias | 4,5,6 | [62] |
| False consensus effect | 3,6,7 | [63] |
| False memory effect | 3,6,7 | [64] |
| False uniqueness effect | 1,3,7 | [65] |
| Focus effect | 4,5,7 | [66] |
| Framing asymmetry | 3,7 | [67] |
| Framing effect | 5,7 | [68] |
| Funding bias | 1,3,4,7 | |
| Gambler's fallacy | 3,4,5 | [69] |
| Google effect | 2,3 | [70] |
| Groupthink | 3,4,6,7 | [71] |
| Halo effect | 2,6,7 | [72] |
| Hindsight bias | 6,7 | [73] |
| Horn Effect | 5,7 | [74] |
| Hostile attribution bias | 7 | [75] |
| IKEA effect | 4,7 | [76] |
| Illusion of control | 4,5 | [77] |
| Illusion of neutrality | 4,6,7 | [78] |
| Information bias | 2,4,7 | [79] |
| Information overload bias | 2,3,6,7 | [80] |
| Ingroup favoritism | 7 | [81] |
| Internal validity bias | 4,5,6 | [82] |
| Justification bias | 5,6,7 | [83] |
| Just-world hypothesis | 2,3,5,6,7 | [84] |
| Labeling effect | 4 | [85] |
| Loss aversion | 4 | [86] |
| Matilda effect | 7 | [87] |
| Matthew effect | 2,3,5,6,7 | [88] |
| Mere exposure effect | 2,3,5,7 | [89] |
| Missing data bias | 4,5,6,7 | [90] |
| Moral luck | 6,7 | [91] |

**Table 2.** *Cont.*

| Name | Phase | Ref. |
|---|---|---|
| Naïve cynicism | 2,7 | [92] |
| Naïve realism | 2,3,5,6,7 | [93] |
| Negativity bias | 2,5,6 | [94] |
| Omission bias | 4,5,7 | [95] |
| Optimism bias | 3,4,5,6,7 | [96] |
| Outcome bias | 5,6,7 | [97] |
| Outgroup homogeneity effect | 2,3,5,6,7 | [98] |
| Overconfidence effect | 1,3,5,7 | [99] |
| Parkinson's law of triviality | 7 | [100] |
| Pessimism bias | 3,4,5,6,7 | [101] |
| Placebo effect | 4,5,6 | [102] |
| Pollyanna effect | 6,7 | [103] |
| Pro-innovation bias | 3,4,7 | [104] |
| Pseudocertainty effect | 3,4,5 | [105] |
| Pygmalion effect or self-fulfilling prophecy | 3,4,5 | [106] |
| Ranking bias | 2,3 | [107] |
| Reactance | 4,7 | [108] |
| Recency effect | 2,7 | [109] |
| Regression fallacy | 5,6 | [110] |
| Replication crisis | 4,5,7 | [111] |
| Reverse Matilda effect | 7 | [112] |
| Risk compensation effect | 4,5,7 | [113] |
| Selective perception bias | 2,3,5,6,7 | [114] |
| Self-serving bias | 5,6,7 | [115] |
| Source homogeneity bias | 2,3,5,6,7 | [116] |
| Status quo bias | 3,4,6,7 | [117] |
| Suggestibility | 4,5,7 | [118] |
| Sunk cost fallacy | 3,4,5,6 | [119] |
| Survivorship bias | 4,5,6,7 | [120] |
| Tachypsychia | 7 | [121] |
| Temporal framing effect | 7 | [122] |
| Zeigarnik effect | 4,5,7 | [123] |
| Zero-risk bias | 4,5 | [124] |

The analysis of the absolute frequency of each bias revealed notable differences depending on the stage of the process. Table 3 shows the number of distinct biases that can potentially occur in each phase of research.

**Table 3.** Phases and frequency of bias occurrence.

| Phase | Frequency |
|---|---|
| Identification of the research problem or question | 6 |
| Literature review and theoretical understanding | 44 |
| Hypothesis formulation | 50 |
| Methodological design and data collection | 38 |
| Data analysis and hypothesis evaluation | 48 |
| Conclusions and paradigm comparison | 36 |
| Results dissemination and feedback | 58 |

To provide a solid theoretical basis for these observations, a classical chi-square $\chi^2$ test was applied. A total of 100 distinct biases were identified, distributed across 280 instances throughout the stages of the research process. This served as the starting point for testing the null hypothesis that the incidence of biases is evenly distributed across all

phases. The computed chi-square statistic was $\chi^2 = 42.0$ with df = 6. The corresponding the *p*-value < 0.001, which leads to the rejection of the null hypothesis of homogeneous distribution. This result confirms that the occurrence of biases is not evenly spread across the phases of the research process but instead shows significant variation.

1. Under the assumption of a homogeneous distribution, the expected frequency for each phase is E = 40
2. The $\chi^2$ statistic was calculated as the sum of the terms $\frac{(0-E)^2}{E}$, where O represents the observed frequency (Table 4).

**Table 4.** Results of the calculations of $\chi^2$.

| Phase | O | E | $\chi^2$ |
|---|---|---|---|
| 1 | 6 | 40 | 28.90 |
| 2 | 44 | 40 | 0.40 |
| 3 | 50 | 40 | 2.50 |
| 4 | 38 | 40 | 0.10 |
| 5 | 48 | 40 | 1.60 |
| 6 | 36 | 40 | 0.40 |
| 7 | 58 | 40 | 8.10 |

3. The test statistic was therefore computed as $\Sigma\frac{(0-E)^2}{E} = 42.0$
4. This value was compared against the critical $\chi^2$ distribution with $k - 1 = 6$ degrees of freedom and a significance level of $\alpha = 0.05$.
5. If the computed $\chi^2$ exceeds the critical value, the null hypothesis must be rejected, indicating that the incidence of biases varies significantly across the stages of the research process, as hypothesized. Otherwise, the hypothesis of homogeneous distribution could not be discarded.

Figure 2 illustrates graphically how the numerical analysis of the results, supported by a consolidated statistical method, confirms that the distribution is not homogeneous.

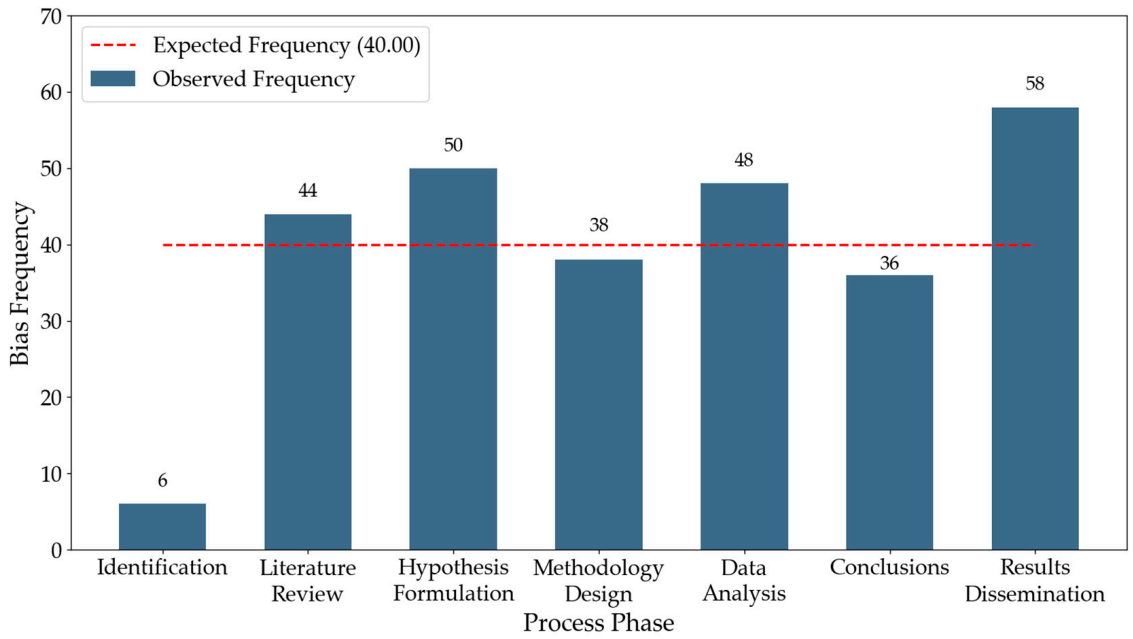


**Figure 2.** Bias distribution in the research process phases.

The results show that the stage in which the researcher tends to exercise the least control over the process is precisely where the greatest likelihood of bias is observed. Thus,

the phase of results dissemination and feedback exhibited the highest concentration, accounting for 20.71% of occurrences. Intermediate stages such as hypothesis formulation (17.86%) and data analysis (17.14%) also reflected a high incidence of cognitive biases during data interpretation and conjecture development. Literature review and theoretical understanding (15.71%), together with methodological design and data collection (13.57%), presented a moderate level of bias. By contrast, the stages of conclusions and paradigm comparison (12.86%) and the identification of the research problem (2.14%) recorded the lowest percentages. This distribution highlights the need for researchers to adopt a proactive attitude, remaining alert to the emergence of biases throughout their investigative process.

Beyond the overall distribution, the analysis also reveals that certain types of biases are particularly prevalent at specific stages of the research process. During the literature review phase, authority bias, information overload bias, and database filtering bias are among the most frequent, reflecting the influence of dominant authors and selective retrieval systems on knowledge construction. Hypothesis formulation is strongly affected by confirmation bias, anchoring effect, and cognitive hyperparameterization bias, which can skew the direction and scope of proposed research questions. In the methodological design and data collection phase, experimenter bias and technological dependence in scientific inference are recurrent, often shaping how data are gathered and interpreted. The data analysis stage shows high incidence of scientific automation bias, correlation–causation confusion, and cumulative bias in AI models, revealing a tendency to accept algorithmic outputs without critical scrutiny. Finally, results dissemination and feedback are dominated by selective dissemination bias, Matilda and Matthew effects, and framing effects, which collectively distort how findings are presented and perceived by the broader community. This stage-specific pattern underscores the importance of tailoring mitigation strategies to the dominant biases present in each phase.

With regard to the five most recurrent biases, cherry picking, present across all seven stages, entails the intentional selection of data that support or refute a pre-existing hypothesis, while disregarding information that could contradict or validate it. This is closely linked to the persistent risk of confirmation bias, which translates into a tendency to privilege information that reinforces prior beliefs, often justifying the spurious data selection mentioned above. Added to this, the Matthew effect, observed in five instances, reflects the accumulation of advantages by already established individuals or theories, to the detriment of emerging contributions. The just-world hypothesis underscores the inclination to believe that events occur fairly and in accordance with ethical assumptions, thereby fostering biased interpretations of causality in research results. Finally, naïve realism complements this set of recurrent biases, manifesting as the belief that one's own perception of reality is objective and that those who question it are mistaken.

### *3.2. AI-Related Emergent Biases Identified in CPS/IIoT*

From the analysis conducted, and in contrast to the traditional scientific paradigm, ten emergent biases were identified in the reviewed literature. These biases are directly associated with the use of AI in engineering research applied to cyber-physical systems (CPS) and the Industrial Internet of Things (IIoT). Emergent biases were distinguished from established categories by detecting terms absent in prior taxonomies but recurrent in CPS and IIoT publications. Each candidate was cross-checked across multiple industrial domains and confirmed through manual validation. This step mirrors hybrid approaches in the recent literature reviews that integrate automated extraction with expert assessment to capture novel trends [125]:

1. Scientific automation bias (Phase: Data analysis and hypothesis evaluation): The tendency to accept AI-generated results without critical examination, neglecting to

verify the validity of those results or the quality of input data. In engineering, this is evident when predictive models (e.g., structural behavior or fault analysis) are employed uncritically, leading to flawed decisions in the design or maintenance of infrastructures [126].

2. Algorithmic opacity bias (Phase: Literature review and theoretical understanding): Stemming from the lack of transparency in AI models, this bias hampers the detection of inherent errors or distortions in their functioning. In engineering research, automated monitoring systems may exclude or filter critical information without explanation, undermining the reliability of conclusions. In digitalized production environments, where CPS integrate large volumes of real-time data, this opacity compromises decision traceability and may generate failures that are difficult to diagnose [127].
3. Knowledge homogenization bias (Phase: Literature review and theoretical understanding): AI models trained on large datasets tend to reinforce established theories and approaches, which may limit the exploration of innovative ideas in engineering. For example, recommendation systems may consistently return the same sources or well-recognized authors, disregarding emerging lines of research. In highly automated industrial contexts, this reduces the diversity of considered solutions and may hinder the adoption of disruptive approaches in areas such as process optimization, materials design, or predictive maintenance [128].
4. Cognitive hyperparameterization bias (Phase: Hypothesis formulation): This bias manifests when AI-based methodologies are prioritized over traditional empirical methods, creating problems in engineering fields where physical experimentation is essential to validate models in critical performance scenarios. Excessive reliance on simulations or automated tools may compromise the validity of results in the absence of complementary experimental testing. In advanced production environments, this can lead to design or control decisions based on overfitted models with low generalization capacity, thereby increasing the risk of failures in production systems and infrastructure operation [129].
5. Technological dependence in scientific inference (Phase: Methodological design and data collection): Occurs when experimental planning becomes overly conditioned by AI-based tools, sidelining indispensable empirical validation practices. In engineering, this is seen in the uncritical adoption of smart sensors or automated acquisition systems, which, although optimizing data collection, may limit the researcher's ability to detect anomalies unforeseen by the algorithms. In industrial contexts, such dependence may result in predictive maintenance or quality control processes that reproduce technological system limitations rather than overcoming them, thus compromising decision reliability [130].
6. AI-assisted confirmation bias (Phase: Hypothesis formulation): This bias arises when AI tools are employed to search and filter information that validates the researcher's pre-existing hypotheses, rather than exposing them to contradictory evidence. For instance, an AI-driven academic search engine may prioritize studies aligned with the researcher's initial belief, thereby reinforcing conviction instead of challenging it [131].
7. Database filtering bias (Phase: Literature review and theoretical understanding): The use of AI-based systems for automated literature selection can lead to the exclusion of relevant studies due to biased indexing or filtering criteria, producing a partial view of the state of the art. This is particularly critical in engineering, where diversity of perspectives drives innovation. A search system omitting certain types of publications can thus distort the research landscape [132]. Although both knowledge homogenization bias and database filtering bias may appear related, they originate

at different stages and operate through distinct mechanisms. Database filtering bias emerges earlier in the research pipeline, during the automated retrieval and selection of literature, when indexing rules or algorithmic filters exclude relevant studies and thereby constrain the diversity of the knowledge base from the outset. Knowledge homogenization bias, by contrast, manifests later, during the processing and modeling of information, when AI systems trained on large datasets disproportionately reinforce prevailing theories, canonical sources, or widely accepted methodologies. As a result, database filtering bias restricts what information enters the analysis, while knowledge homogenization bias shapes how that information is weighted, interpreted, and reproduced, limiting the exploration of novel hypotheses or disruptive approaches in engineering research [133].

8. Cumulative bias in AI models (Phase: Data analysis and hypothesis evaluation): This occurs when AI models are trained on datasets that already contain historical biases, thereby reinforcing prior errors and distorting scientific inference. In engineering, such cumulative bias may negatively affect predictive failure systems for critical infrastructures, causing them to perpetuate error patterns instead of correcting them.
9. Algorithmic optimization bias in experimentation (Phase: Methodological design and data collection): Refers to the adjustment of models or experimental parameters to maximize computational efficiency at the expense of fidelity and precision in representing complex phenomena. For instance, numerical simulations might be oversimplified to reduce computation times, thereby compromising the validity of engineering simulations by sacrificing detail or realism [134].
10. Selective dissemination bias of scientific findings (Phase: Results dissemination and feedback): Describes the tendency of certain AI systems (e.g., publication platforms or automated dissemination networks) to favor and amplify positive results or those aligned with dominant trends, while neglecting rigorous studies with null or negative results. In engineering, this distorts the perception of technological development success, potentially limiting innovation by rendering invisible findings that could be critical for scientific progress but do not fit prevailing narratives [135].

## 4. Discussion

The role of the contemporary engineering researcher, particularly in the context of Industry 4.0, involves not only the design and application of advanced digital tools but also the capacity to identify and mitigate biases that threaten the validity and reliability of research outcomes. Based on the taxonomy and phase-specific distribution of biases presented in Section 3, the following strategies are proposed as operational guidelines. They define the object of management—the cognitive and AI-related biases identified in this study—and the subject of management—the actors and systems responsible for addressing them, including researchers, AI tools, and industrial processes. These strategies are designed to align mitigation actions with the specific stages of the research workflow where biases are most likely to occur.

1. Education and awareness (Phases 1–7, Subjects: researchers): Researchers must be systematically trained to recognize the cognitive mechanisms and algorithmic distortions that generate biases, particularly confirmation bias, automation bias, and cumulative bias. This training should address how these distortions emerge at each stage of the research process—from problem definition to dissemination—and equip researchers with methodological literacy to critically interpret AI outputs and ensure transparency and interpretability in industrial contexts [136].
2. Process transparency (Phases 3–6, Subjects: researchers and AI systems): AI models used in cyber-physical systems and IIoT environments should integrate explainabil-

ity components that make their internal decision logic interpretable. Transparent pipelines enable the detection of distortions such as correlation–causation confusion, cumulative bias, and automation bias, facilitating auditing and validation of results during hypothesis formulation, analysis, and dissemination stages [137].

3. Data diversification (Phases 2–5, Subjects: researchers and data engineers): Expanding the diversity and representativeness of training and operational datasets mitigates database filtering bias and knowledge homogenization bias. Integrating heterogeneous data sources from multiple industrial contexts reduces structural distortions and enhances robustness and generalizability, improving the reliability of results in data-intensive stages [138].
4. Human supervision (Phases 4–6, Subjects: researchers and domain experts): Implementing human-in-the-loop approaches ensures that human expertise validates critical outputs of automated systems, counteracting scientific automation bias and technological dependence in inference. Human oversight is particularly crucial in high-stakes CPS applications, where algorithmic recommendations must be critically assessed before deployment [139].
5. Continuous auditing and monitoring (Phases 5–7, Subjects: researchers, institutions, and AI governance bodies): Governance frameworks should define clear responsibilities between human agents and AI systems, ensuring continuous evaluation of model behavior and early detection of bias re-emergence during operation. Periodic bias audits are essential in dynamic industrial environments, where input distributions and operational conditions evolve over time [140].
6. Iterative model adjustment (Phases 5–6, Subjects: AI developers and researchers): Reinforcement learning with human feedback (RLHF) and other adaptive techniques should be applied to continuously align model outputs with operational and ethical requirements, reducing the persistence of cumulative and hyperparameterization biases. This iterative refinement improves both performance and reliability in real-world CPS scenarios [141].
7. Multi-phase mitigation techniques (Phases 3–6, Subjects: AI developers and data scientists): Bias control must be implemented throughout the research lifecycle—before, during, and after model training. Pre-processing (e.g., data rebalancing), in-process (e.g., adversarial debiasing), and post-processing (e.g., output calibration) interventions work together to reduce the impact of biases across methodological design and analysis stages, enhancing reproducibility [142].
8. Multidimensional evaluation (Phases 5–6, Subjects: researchers and system evaluators): Model evaluation should go beyond conventional accuracy metrics to include fairness, robustness, explainability, and resilience. Adopting this broader evaluation framework prevents the amplification of subtle biases, such as automation bias or cumulative bias, and supports trustworthy deployment of AI systems in industrial contexts [143].
9. Cross-validation methods (Phases 5–6, Subjects: researchers and data scientists): Employing rigorous validation strategies, such as k-fold or stratified cross-validation, mitigates sampling and selection biases while improving error estimation. These techniques enhance the reliability and generalizability of results across data subsets and reduce the risk of overfitting driven by biased data partitions [144].
10. Interdisciplinary collaboration (Phases 1–7, Subjects: researchers, data scientists, ethicists, and domain experts): Collaboration across disciplines is essential to define boundaries for AI intervention, contextualize findings, and maintain human oversight in decision-making. This integrative approach aligns technical solutions with societal

and industrial values, mitigating risks associated with unexamined bias propagation and reinforcing accountability throughout the research process [145].

### 4.1. Practical Implications

Bias does not affect all industrial domains in the same way. In manufacturing, reliance on autonomous decision-making often leads to automation bias and a tendency to trust systems without sufficient verification. Here, routine human checks and independent audits become essential safeguards. The energy sector faces different difficulties: predictive models may be trained on incomplete records and then struggle when demand patterns or environmental conditions change. Without diverse datasets and regular model updates, reliability quickly drops. Logistics and supply chains bring yet another profile, where the selective communication of results and the confusion between correlation and causation can mislead planning or inventory decisions. These examples illustrate that mitigation cannot be handled with a universal recipe. What is needed are solutions that take into account the conditions of each sector. The taxonomy and framework developed in this study are meant to guide such tailoring, helping practitioners link specific types of bias with concrete operational settings and choose responses that work in practice [146].

### 4.2. Comparative Insights, Limitations, and Future Directions

Recent analyses confirm that the presence of bias in industrial AI research is closely linked to data quality, representativeness, and interpretability [147]. The taxonomy developed here extends this perspective by consolidating cognitive and algorithmic biases into a phase-oriented framework aligned with the research workflow in Industry 4.0.

A sensitivity check was performed by adjusting thresholds in TextRank and window sizes in co-occurrence analysis. While absolute frequencies of detected terms varied, the relative distribution across research phases remained stable, indicating robustness [148]. The study has limitations: the review corpus was limited to SCOPUS and Web of Science, and manual validation was carried out on a finite subset. Future work should broaden the database coverage, incorporate domain-specific corpora, and integrate explainable AI techniques to refine classification and validation [149].

In summary, the discussion confirms that the research question has been addressed and that the identified gap has been filled. The taxonomy and mitigation framework provide a comprehensive response to the lack of systematized bias analysis in Industry 4.0 research, reinforcing both conceptual clarity and practical applicability.

## 5. Conclusions

The reviewed literature establishes clear relationships between the presence of biases in AI-supported research and their direct impact on the robustness of scientific outcomes. These relationships have been addressed statistically through a quantitative examination of existing publications, which has allowed for mathematical verification that biases are not distributed homogeneously across the different stages of the research process. This confirms one of the initial hypotheses formulated.

Through a complementary qualitative analysis, specific biases linked to the use of AI technologies in engineering have been identified and compiled. Among these, the uncritical acceptance of automated results and the difficulty of accessing the internal functioning of algorithms are particularly prominent. These circumstances, together with the others analyzed in this study, demonstrate that the use of AI is not a neutral process; rather, it reinforces pre-existing biases and generates new potential sources of error, with a negative impact on the objectivity and reliability of the scientific knowledge produced.

In order to adequately address these difficulties, it is imperative that researchers explicitly acknowledge them. In this regard, the contribution of this work lies in the systematic compilation of information—previously dispersed—on the different types of identified biases, as well as in the presentation of strategies proposed by the specialized literature for their mitigation. These include systematic human oversight, algorithmic transparency, diversification of datasets, and the regular implementation of critical audits.

From this, it follows that machines do not replace humans but rather extend their capabilities in environments characterized by cyber-physical systems, smart manufacturing, and IIoT infrastructures, which define the ongoing industrial digital transformation. Within this framework, the proper management of biases not only helps ensure the validity of scientific outcomes but also strengthens the reliability of automated production processes, optimizes real-time decision-making, and sustains the technological competitiveness of the so-called fourth industrial revolution, while orienting its progress towards the envisioned Industry 5.0, in which the human factor regains prominence.

**Author Contributions:** Conceptualization, J.A.-R. and J.B.-F.; methodology, J.A.-R. and F.-J.F.-M.; software, J.A.-R.; validation, E.M.-C., E.J.-M. and J.-I.L.-B.; formal analysis, J.-I.L.-B. and E.J.-M.; investigation, J.A.-R. and F.-J.F.-M.; resources, E.M.-C.; data curation, J.A.-R.; writing—original draft preparation, J.A.-R., F.-J.F.-M. and J.B.-F.; writing—review and editing, J.B.-F., E.J.-M. and J.-I.L.-B.; visualization, E.M.-C.; supervision, J.B.-F.; project administration, E.M.-C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Portillo-Blanco, A.; Guisasola, J.; Zuza, K. Integrated STEM Education: Addressing Theoretical Ambiguities and Practical Applications. *Front. Educ.* **2025**, *10*, 1568885. [CrossRef]
2. Correa, J. Science and Scientific Method. *Int. J. Sci. Res.* **2022**, *11*, 621–633. [CrossRef]
3. Wang, R. Active Learning-Based Optimization of Scientific Experimental Design. In Proceedings of the 2021 2nd International Conference on Artificial Intelligence and Computer Engineering, ICAICE 2021, Hangzhou, China, 5–7 November 2021; pp. 268–274. [CrossRef]
4. Obioha-Val, O.; Oladeji, O.O.; Selesi-Aina, O.; Olayinka, M.; Kolade, T.M. Machine Learning-Enabled Smart Sensors for Real-Time Industrial Monitoring: Revolutionizing Predictive Analytics and Decision-Making in Diverse Sector. *Asian J. Res. Comput. Sci.* **2024**, *17*, 92–113. [CrossRef]
5. Necula, S.C.; Dumitriu, F.; Greavu-Șerban, V. A Systematic Literature Review on Using Natural Language Processing in Software Requirements Engineering. *Electronics* **2024**, *13*, 2055. [CrossRef]
6. Kong, X.; Jiang, X.; Zhang, B.; Yuan, J.; Ge, Z. Latent Variable Models in the Era of Industrial Big Data: Extension and Beyond. *Annu. Rev. Control* **2022**, *54*, 167–199. [CrossRef]
7. Mienye, I.D.; Swart, T.G. A Comprehensive Review of Deep Learning: Architectures, Recent Advances, and Applications. *Information* **2024**, *15*, 755. [CrossRef]
8. Cachola, I.; Lo, K.; Cohan, A.; Weld, D.S. TLDR: Extreme Summarization of Scientific Documents. In Proceedings of the Findings of the Association for Computational Linguistics Findings of ACL: EMNLP 2020, Seattle, WA, USA, 5–10 July 2020; pp. 4766–4777. [CrossRef]
9. Lee, Y.L.; Zhou, T.; Yang, K.; Du, Y.; Pan, L. Personalized recommender systems based on social relationships and historical behaviors. *Appl. Math. Comput.* **2022**, *437*, 127549. [CrossRef]
10. Werdiningsih, I.; Marzuki; Rusdin, D. Balancing AI and Authenticity: EFL Students' Experiences with ChatGPT in Academic Writing. *Cogent. Arts Humanit.* **2024**, *11*, 2392388. [CrossRef]
11. Mehdiyev, N.; Majlatow, M.; Fettke, P. Interpretable and Explainable Machine Learning Methods for Predictive Process Monitoring: A Systematic Literature Review. *arXiv* **2023**, arXiv:2312.17584. [CrossRef]

12. Astleitner, H. We Have Big Data, But Do We Need Big Theory? Review-Based Remarks on an Emerging Problem in the Social Sciences. *Philos. Soc. Sci.* **2024**, *54*, 69–92. [CrossRef]
13. Lee, Y.; Pawitan, Y. Popper's Falsification and Corroboration from the Statistical Perspectives. In *Karl Popper's Science and Philosophy*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 121–147. [CrossRef]
14. Lamata, P. Avoiding Big Data Pitfalls. *Heart Metab.* **2020**, *82*, 33–35. [CrossRef] [PubMed]
15. Bertrand, Y.; Van Belle, R.; De Weerdt, J.; Serral, E. Defining Data Quality Issues in Process Mining with IoT Data. In *Lecture Notes in Business Information Processing*; Springer: Berlin/Heidelberg, Germany, 2023; Volume 468, pp. 422–434. [CrossRef]
16. He, J.; Feng, W.; Min, Y.; Yi, J.; Tang, K.; Li, S.; Zhang, J.; Chen, K.; Zhou, W.; Xie, X.; et al. Control Risk for Potential Misuse of Artificial Intelligence in Science. *arXiv* **2023**, arXiv:2312.06632. [CrossRef]
17. Anand, G.; Larson, E.C.; Mahoney, J.T. Thomas Kuhn on Paradigms. *Prod. Oper. Manag.* **2020**, *29*, 1650–1657. [CrossRef]
18. Buyl, M.; De Bie, T. Inherent Limitations of AI Fairness. *Commun. ACM* **2022**, *67*, 48–55. [CrossRef]
19. Horvitz, E.; Young, J.; Elluru, R.G.; Howell, C. Key Considerations for the Responsible Development and Fielding of Artificial Intelligence. *arXiv* **2021**, arXiv:2108.12289. [CrossRef]
20. Da Silva, S.; Gupta, R.; Monzani, D. Editorial: Highlights in Psychology: Cognitive Bias. *Front. Psychol.* **2023**, *14*, 1242809. [CrossRef] [PubMed]
21. Aini, R.Q.; Sya'bandari, Y.; Rusmana, A.N.; Ha, M. Addressing Challenges to a Systematic Thinking Pattern of Scientists: A Literature Review of Cognitive Bias in Scientific Work. *Brain Digit. Learn.* **2021**, *11*, 417–430. [CrossRef]
22. Krenn, M.; Pollice, R.; Guo, S.Y.; Aldeghi, M.; Cervera-Lierta, A.; Friederich, P.; dos Passos Gomes, G.; Häse, F.; Jinich, A.; Nigam, A.; et al. On Scientific Understanding with Artificial Intelligence. *Nat. Rev. Phys.* **2022**, *4*, 761–769. [CrossRef]
23. Vicente, L.; Matute, H. Humans Inherit Artificial Intelligence Biases. *Sci. Rep.* **2023**, *13*, 15737. [CrossRef]
24. Morita, R.; Watanabe, K.; Zhou, J.; Dengel, A.; Ishimaru, S. GenAIReading: Augmenting Human Cognition with Interactive Digital Textbooks Using Large Language Models and Image Generation Models. *arXiv* **2025**, arXiv:2503.07463. [CrossRef]
25. Jugran, S.; Kumar, A.; Tyagi, B.S.; Anand, V. Extractive Automatic Text Summarization Using SpaCy in Python NLP. In Proceedings of the 2021 International Conference on Advance Computing and Innovative Technologies in Engineering, ICACITE 2021, Greater Noida, India, 4–5 March 2021; pp. 582–585. [CrossRef]
26. Edwards, A.; Edwards, C. Does the Correspondence Bias Apply to Social Robots?: Dispositional and Situational Attributions of Human Versus Robot Behavior. *Front. Robot. AI* **2022**, *8*, 788242. [CrossRef]
27. Samoilenko, S.A.; Cook, J. Developing an Ad Hominem Typology for Classifying Climate Misinformation. *Clim. Policy* **2024**, *24*, 138–151. [CrossRef]
28. Yoo, S. LLMs as Deceptive Agents: How Role-Based Prompting Induces Semantic Ambiguity in Puzzle Tasks. *arXiv* **2025**, arXiv:2504.02254. [CrossRef]
29. Peretz-Lange, R.; Gonzalez, G.D.S.; Hess, Y.D. My Circumstances, Their Circumstances: An Actor-Observer Distinction in the Consequences of External Attributions. *Soc. Pers. Psychol. Compass.* **2024**, *18*, e12993. [CrossRef]
30. Wehrli, S.; Hertweck, C.; Amirian, M.; Glüge, S.; Stadelmann, T. Awareness, and Ignorance in Deep-Learning-Based Face Recognition. *AI Ethics* **2021**, *2*, 509–522. [CrossRef]
31. Rastogi, C.; Zhang, Y.; Wei, D.; Varshney, K.R.; Dhurandhar, A.; Tomsett, R. Deciding Fast and Slow: The Role of Cognitive Biases in AI-Assisted Decision-Making. *Proc. ACM Hum. Comput. Interact.* **2022**, *6*, 3512930. [CrossRef]
32. González-Sendino, R.; Serrano, E.; Bajo, J.; Novais, P. A Review of Bias and Fairness in Artificial Intelligence. *Int. J. Interact. Multimed. Artif. Intell.* **2024**, *9*, 5–17. [CrossRef]
33. Tump, A.N.; Pleskac, T.J.; Kurvers, R.H.J.M. Wise or Mad Crowds? The Cognitive Mechanisms Underlying Information Cascades. *Sci. Adv.* **2020**, *6*, eabb0266. [CrossRef] [PubMed]
34. Suri, G.; Slater, L.R.; Ziaee, A.; Nguyen, M. Do Large Language Models Show Decision Heuristics Similar to Humans? A Case Study Using GPT-3.5. *J. Exp. Psychol. Gen.* **2023**, *153*, 1066. [CrossRef]
35. Cau, F.M.; Tintarev, N. Navigating the Thin Line: Examining User Behavior in Search to Detect Engagement and Backfire Effects. *arXiv* **2024**, arXiv:2401.11201. [CrossRef]
36. Knyazev, N.; Oosterhuis, H. The Bandwagon Effect: Not Just Another Bias. In Proceedings of the ICTIR 2022—Proceedings of the 2022 ACM SIGIR International Conference on the Theory of Information Retrieval, Madrid, Spain, 11–12 July 2022; pp. 243–253. [CrossRef]
37. Erbacher, R.F. Base-Rate Fallacy Redux and a Deep Dive Review in Cybersecurity. *arXiv* **2022**, arXiv:2203.08801. [CrossRef]
38. Kovačević, I.; Manojlović, M. Base Rate Neglect Bias: Can It Be Observed in HRM Decisions and Can It Be Decreased by Visually Presenting the Base Rates in HRM Decisions? *Int. J. Cogn. Res. Sci. Eng. Educ.* **2024**, *12*, 119–132. [CrossRef]
39. Ashinoff, B.K.; Buck, J.; Woodford, M.; Horga, G. The Effects of Base Rate Neglect on Sequential Belief Updating and Real-World Beliefs. *PLOS Comput. Biol.* **2022**, *18*, e1010796. [CrossRef] [PubMed]
40. Fraune, M.R. Our Robots, Our Team: Robot Anthropomorphism Moderates Group Effects in Human–Robot Teams. *Front. Psychol.* **2020**, *11*, 540167. [CrossRef]

41. Gautam, S.; Srinath, M. Blind Spots and Biases: Exploring the Role of Annotator Cognitive Biases in NLP. In Proceedings of the HCI+NLP 2024—3rd Workshop on Bridging Human-Computer Interaction and Natural Language Processing, Mexico City, Mexico, 21 June 2024; pp. 82–88. [CrossRef]
42. Bremers, A.; Parreira, M.T.; Fang, X.; Friedman, N.; Ramirez-Aristizabal, A.; Pabst, A.; Spasojevic, M.; Kuniavsky, M.; Ju, W. The Bystander Affect Detection (BAD) Dataset for Failure Detection in HRI. *arXiv* **2023**, arXiv:2303.04835. [CrossRef]
43. Zimring, J.C. Bias with a Cherry on Top: Cherry-Picking the Data. In *Partial Truths*; Columbia University Press: New York City, NY, USA, 2022; pp. 52–60. [CrossRef]
44. Kobayashi, T.; Kitaoka, A.; Kosaka, M.; Tanaka, K.; Watanabe, E. Motion Illusion-like Patterns Extracted from Photo and Art Images Using Predictive Deep Neural Networks. *Sci. Rep.* **2022**, *12*, 3893. [CrossRef]
45. Seran, C.E.; Tan, M.J.T.; Karim, H.A.; Aldahoul, N.; Joshua, M.; Tan, T.; Hezerul; Karim, A. A Conceptual Exploration of Generative AI-Induced Cognitive Dissonance and Its Emergence in University-Level Academic Writing. *Front. Artif. Intell.* **2025**, *8*, 1573368. [CrossRef]
46. Zwaan, L. Cognitive Bias in Large Language Models: Implications for Research and Practice. *NEJM AI* **2024**, *1*, e2400961. [CrossRef]
47. Bashkirova, A.; Krpan, D. Confirmation Bias in AI-Assisted Decision-Making: AI Triage Recommendations Congruent with Expert Judgments Increase Psychologist Trust and Recommendation Acceptance. *Comput. Hum. Behav. Artif. Hum.* **2024**, *2*, 100066. [CrossRef]
48. Peters, U. Algorithmic Political Bias in Artificial Intelligence Systems. *Philos. Technol.* **2022**, *35*, 25. [CrossRef]
49. Westberg, M.; Främling, K. Cognitive Perspectives on Context-Based Decisions and Explanations. *NEJM AI* **2021**, *1*, e2400961. [CrossRef]
50. Kliegr, T.; Bahník, Š.; Fürnkranz, J. A Review of Possible Effects of Cognitive Biases on Interpretation of Rule-Based Machine Learning Models. *Artif. Intell.* **2021**, *295*, 103458. [CrossRef]
51. Xiao, Y.; Wang, S.; Liu, S.; Xue, D.; Zhan, X.; Liu, Y. FITNESS: A Causal De-Correlation Approach for Mitigating Bias in Machine Learning Software. *arXiv* **2023**, arXiv:2305.14396. [CrossRef]
52. Philip, J.; Ωθ, W.; Ruas, T.; Abdalla, M.; Gipp, B.; Mohammad, S.M. Citation Amnesia: On The Recency Bias of NLP and Other Academic Fields. *arXiv* **2024**, arXiv:2402.12046. [CrossRef]
53. Tao, Y.; Viberg, O.; Baker, R.S.; Kizilcec, R.F. Cultural Bias and Cultural Alignment of Large Language Models. *Proc. Natl. Acad. Sci. USA Nexus* **2024**, *3*, 346. [CrossRef] [PubMed]
54. Cao, Y.; Shui, R.; Pan, L.; Kan, M.Y.; Liu, Z.; Chua, T.S. Expertise Style Transfer: A New Task Towards Better Communication between Experts and Laymen. In Proceedings of the Annual Meeting of the Association for Computational Linguistics, Virtual, 7–10 July 2020; pp. 1061–1071. [CrossRef]
55. Schedl, M.; Lesota, O.; Brandl, S.; Lotfi, M.; Ticona, G.J.E.; Masoudian, S. The Importance of Cognitive Biases in the Recommendation Ecosystem. *arXiv* **2024**, arXiv:2408.12492. [CrossRef]
56. Leo, X.; Huh, Y.E. Who Gets the Blame for Service Failures? Attribution of Responsibility toward Robot versus Human Service Providers and Service Firms. *Comput. Hum. Behav.* **2020**, *113*, 106520. [CrossRef]
57. Tavares, S.; Ferrara, E. Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies. *Sci* **2023**, *6*, 3. [CrossRef]
58. Horowitz, M.C.; Kahn, L. Bending the Automation Bias Curve: A Study of Human and AI-Based Decision Making in National Security Contexts. *Int. Stud. Q.* **2023**, *68*, sqae020. [CrossRef]
59. Gong, Q. Machine Endowment Cost Model: Task Assignment between Humans and Machines. *Humanit. Soc. Sci. Commun.* **2023**, *10*, 129. [CrossRef]
60. Barkett, E.; Long, O.; Kröger, P. Getting out of the Big-Muddy: Escalation of Commitment in LLMs. *arXiv* **2025**, arXiv:2508.01545. [CrossRef]
61. Hufendiek, R. Beyond Essentialist Fallacies: Fine-Tuning Ideology Critique of Appeals to Biological Sex Differences. *J. Soc. Philos.* **2022**, *53*, 494–511. [CrossRef]
62. Maneuvrier, A. Experimenter Bias: Exploring the Interaction between Participant's and Investigator's Gender/Sex in VR. *Virtual Real.* **2024**, *28*, 96. [CrossRef]
63. Choi, J.; Hong, Y.; Kim, B. People Will Agree What I Think: Investigating LLM's False Consensus Effect. In Proceedings of the Findings of the Association for Computational Linguistics: NAACL 2025, Albuquerque, NM, USA, 29 April–4 May 2025; pp. 95–126. [CrossRef]
64. Pataranutaporn, P.; Archiwaranguprok, C.; Chan, S.W.T.; Loftus, E.; Maes, P. Synthetic Human Memories: AI-Edited Images and Videos Can Implant False Memories and Distort Recollection. In Proceedings of the Conference on Human Factors in Computing Systems—Proceedings, Honolulu, HI, USA, 11–16 May 2024; Volume 1. [CrossRef]
65. Mou, Y.; Xu, T.; Hu, Y. Uniqueness Neglect on Consumer Resistance to AI. *Mark. Intell. Plan.* **2023**, *41*, 669–689. [CrossRef]

66. Herrebrøden, H. Motor Performers Need Task-Relevant Information: Proposing an Alternative Mechanism for the Attentional Focus Effect. *J. Mot. Behav.* **2023**, *55*, 125–134. [CrossRef]
67. Marwala, T.; Hurwitz, E. Artificial Intelligence and Asymmetric Information Theory. *arXiv* **2015**, arXiv:1510.02867. [CrossRef]
68. Ulnicane, I.; Aden, A. Power and Politics in Framing Bias in Artificial Intelligence Policy. *Rev. Policy Res.* **2023**, *40*, 665–687. [CrossRef]
69. Saeedi, P.; Goodarzi, M.; Canbaz, M.A. Heuristics and Biases in AI Decision-Making: Implications for Responsible AGI. In Proceedings of the 2025 6th International Conference on Artificial Intelligence, Robotics and Control (AIRC), Savannah, GA, USA, 15 July 2025. [CrossRef]
70. Gong, C.; Yang, Y. Google Effects on Memory: A Meta-Analytical Review of the Media Effects of Intensive Internet Search Behavior. *Front. Public Health* **2024**, *12*, 1332030. [CrossRef]
71. Wiss, A.; Showstark, M.; Dobbeck, K.; Pattershall-Geide, J.; Zschaebitz, E.; Joosten-Hagye, D.; Potter, K.; Embry, E. Utilizing Generative AI to Counter Learner Groupthink by Introducing Controversy in Collaborative Problem-Based Learning Settings. *Online Learn.* **2025**, *29*, 39–65. [CrossRef]
72. Nicolau, J.L.; Mellinas, J.P.; Martín-Fuentes, E. The Halo Effect: A Longitudinal Approach. *Ann. Tour Res.* **2020**, *83*, 102938. [CrossRef]
73. Wang, J.; Redelmeier, D.A. Cognitive Biases and Artificial Intelligence. *NEJM AI* **2024**, *1*, 2400639. [CrossRef]
74. Noor, N.; Beram, S.; Huat, F.K.C.; Gengatharan, K.; Mohamad Rasidi, M.S. Bias, Halo Effect and Horn Effect: A Systematic Literature Review. *Int. J. Acad. Res. Bus. Soc. Sci.* **2023**, *13*, 1116–1140. [CrossRef] [PubMed]
75. Lyu, Y.; Combs, D.; Neumann, D.; Leong, Y.C. Automated Scoring of the Ambiguous Intentions Hostility Questionnaire Using Fine-Tuned Large Language Models. *arXiv* **2025**, arXiv:2508.10007. [CrossRef]
76. Vuculescu, O.; Beretta, M.; Bergenholtz, C. The IKEA Effect in Collective Problem-Solving: When Individuals Prioritize Their Own Solutions. *Creat. Innov. Manag.* **2021**, *30*, 116–128. [CrossRef]
77. Nowotny, H. AI and the Illusion of Control. In Proceedings of the Paris Institute for Advanced Study, Paris, France, 2 April 2024; Volume 1, p. 10. [CrossRef]
78. Xavier, B. Biases within AI: Challenging the Illusion of Neutrality. *AI Soc.* **2024**, *40*, 1545–1546. [CrossRef]
79. Ntoutsi, E.; Fafalios, P.; Gadiraju, U.; Iosifidis, V.; Nejdl, W.; Vidal, M.E.; Ruggieri, S.; Turini, F.; Papadopoulos, S.; Krasanakis, E.; et al. Bias in Data-Driven Artificial Intelligence Systems—An Introductory Survey. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2020**, *10*, e1356. [CrossRef]
80. Goette, L.; Han, H.-J.; Leung, B.T.K. Information Overload and Confirmation Bias. 2024. Available online: https://ssrn.com/abstract=4843939 (accessed on 1 October 2025).
81. Ciccarone, G.; Di Bartolomeo, G.; Papa, S. The Rationale of In-Group Favoritism: An Experimental Test of Three Explanations. *Games Econ. Behav.* **2020**, *124*, 554–568. [CrossRef]
82. Pacchiardi, L.; Teši'c, M.T.; Cheke, L.; Hernández-Orallo, J. Leaving the Barn Door Open for Clever Hans: Simple Features Predict LLM Benchmark Answers. *arXiv* **2024**, arXiv:2410.11672. [CrossRef]
83. Kartal, E. A Comprehensive Study on Bias in Artificial Intelligence Systems: Biased or Unbiased AI, That's the Question! *Int. J. Intell. Inf. Technol.* **2022**, *18*, 309582. [CrossRef]
84. Kim, S.; Sohn, Y.W. The Effect of Belief in a Just World on the Acceptance of AI Technology. *Korean J. Psychol. Gen.* **2020**, *39*, 517–542. [CrossRef]
85. Haliburton, L.; Ghebremedhin, S.; Welsch, R.; Schmidt, A.; Mayer, S. Investigating Labeler Bias in Face Annotation for Machine Learning. *Front. Artif. Intell. Appl.* **2023**, *386*, 145–161. [CrossRef]
86. Zhou, J. A Review of the Relationship Between Loss Aversion Bias and Investment Decision-Making Process. *Adv. Econ. Manag. Political Sci.* **2023**, *27*, 143–150. [CrossRef]
87. Fonseca, J. The Myth of Meritocracy and the Matilda Effect in STEM: Paper Acceptance and Paper Citation. *arXiv* **2023**, arXiv:2306.10807. [CrossRef]
88. Liao, C.H. The Matthew Effect and the Halo Effect in Research Funding. *J. Inf.* **2021**, *15*, 101108. [CrossRef]
89. Sguerra, B.; Tran, V.-A.; Hennequin, R. Ex2Vec: Characterizing Users and Items from the Mere Exposure Effect. In Proceedings of the 17th ACM Conference on Recommender Systems, RecSys 2023, Singapore, 18–22 September 2023; Volume 1, pp. 971–977. [CrossRef]
90. Bhatti, A.; Sandrock, T.; Nienkemper-Swanepoel, J. The Influence of Missing Data Mechanisms and Simple Missing Data Handling Techniques on Fairness. *arXiv* **2025**, arXiv:2503.07313. [CrossRef]
91. Kneer, M.; Skoczeń, I. Outcome Effects, Moral Luck and the Hindsight Bias. *Cognition* **2023**, *232*, 105258. [CrossRef]
92. Tsfati, Y.; Barnoy, A. Media Cynicism, Media Skepticism and Automatic Media Trust: Explicating Their Connection with News Processing and Exposure. *Communic. Res.* **2025**, 00936502251327717. [CrossRef]
93. Surden, H. Naïve Realism, Cognitive Bias, and the Benefits and Risks of AI. *SSRN Electron. J.* **2023**, *23*, 4393096. [CrossRef]

94. Chiarella, S.G.; Torromino, G.; Gagliardi, D.M.; Rossi, D.; Babiloni, F.; Cartocci, G. Investigating the Negative Bias towards Artificial Intelligence: Effects of Prior Assignment of AI-Authorship on the Aesthetic Appreciation of Abstract Paintings. *Comput. Human Behav.* **2022**, *137*, 107406. [CrossRef]
95. Cheung, V.; Maier, M.; Lieder, F. Large Language Models Show Amplified Cognitive Biases in Moral Decision-Making. *Proc. Natl. Acad. Sci. USA* **2025**, *122*, e2412015122. [CrossRef] [PubMed]
96. Owen, M.; Flowerday, S.V.; van der Schyff, K. Optimism Bias in Susceptibility to Phishing Attacks: An Empirical Study. *Inf. Comput. Secur.* **2024**, *32*, 656–675. [CrossRef]
97. Stone, J.C.; Gurunathan, U.; Aromataris, E.; Glass, K.; Tugwell, P.; Munn, Z.; Doi, S.A.R. Bias Assessment in Outcomes Research: The Role of Relative Versus Absolute Approaches. *Value Health* **2021**, *24*, 1145–1149. [CrossRef]
98. Li, W.; Zhou, X.; Yang, Q. Designing Medical Artificial Intelligence for In- and out-Groups. *Comput. Human Behav.* **2021**, *124*, 106929. [CrossRef]
99. Sihombing, Y.R.; Prameswary, R.S.A. The Effect of Overconfidence Bias and Representativeness Bias on Investment Decision with Risk Tolerance as Mediating Variable. *Indik. J. Ilm. Manaj. Dan Bisnis* **2023**, *7*, 1. [CrossRef]
100. Borowa, K.; Zalewski, A.; Kijas, S. The Influence of Cognitive Biases on Architectural Technical Debt. In Proceedings of the 2021 IEEE 18th International Conference on Software Architecture (ICSA), Stuttgart, Germany, 22–26 March 2021. [CrossRef]
101. Montag, C.; Schulz, P.J.; Zhang, H.; Li, B.J. On Pessimism Aversion in the Context of Artificial Intelligence and Locus of Control: Insights from an International Sample. *AI Soc.* **2025**, *40*, 3349–3356. [CrossRef]
102. Kosch, T.; Welsch, R.; Chuang, L.; Schmidt, A. The Placebo Effect of Artificial Intelligence in Human-Computer Interaction. *ACM Trans. Comput.-Hum. Interact.* **2022**, *29*, 32. [CrossRef]
103. Marineau, J.E.; Labianca, G. (Joe) Positive and Negative Tie Perceptual Accuracy: Pollyanna Principle vs. Negative Asymmetry Explanations. *Soc. Netw.* **2021**, *64*, 83–98. [CrossRef]
104. Obendiek, A.S.; Seidl, T. The (False) Promise of Solutionism: Ideational Business Power and the Construction of Epistemic Authority in Digital Security Governance. *J. Eur. Public Policy* **2023**, *30*, 1305–1329. [CrossRef]
105. Gulati, A.; Lozano, M.A.; Lepri, B.; Oliver, N. BIASeD: Bringing Irrationality into Automated System Design. *arXiv* **2022**, arXiv:2210.01122. [CrossRef]
106. De-Arteaga, M.; Elmer, J. Self-Fulfilling Prophecies and Machine Learning in Resuscitation Science. *Resuscitation* **2023**, *183*, 109622. [CrossRef]
107. Yang, T.; Han, C.; Luo, C.; Gupta, P.; Phillips, J.M.; Ai, Q. Mitigating Exploitation Bias in Learning to Rank with an Uncertainty-Aware Empirical Bayes Approach. In Proceedings of the WWW 2024—Proceedings of the ACM Web Conference, Madrid, Spain, 13–17 May 2024; Volume 1, pp. 1486–1496. [CrossRef]
108. Candrian, C.; Scherer, A. Reactance to Human versus Artificial Intelligence: Why Positive and Negative Information from Human and Artificial Agents Leads to Different Responses. *SSRN Electron. J.* **2023**, 4397618. [CrossRef]
109. Wang, P.; Yang, H.; Hou, J.; Li, Q. A Machine Learning Approach to Primacy-Peak-Recency Effect-Based Satisfaction Prediction. *Inf. Process. Manag.* **2023**, *60*, 103196. [CrossRef]
110. Del Giudice, M. The Prediction-Explanation Fallacy: A Pervasive Problem in ScientificApplications of Machine Learning. *Methodology* **2024**, *20*, 22–46. [CrossRef]
111. Gundersen, O.E.; Cappelen, O.; Mølnå, M.; Nilsen, N.G. The Unreasonable Effectiveness of Open Science in AI: A Replication Study. *arXiv* **2024**, arXiv:2412.17859. [CrossRef]
112. Malecki, W.P.; Kowal, M.; Krasnodębska, A.; Bruce, B.C.; Sorokowski, P. The Reverse Matilda Effect: Gender Bias and the Impact of Highlighting the Contributions of Women to a STEM Field on Its Perceived Attractiveness. *Sci. Educ.* **2024**, *108*, 1474–1491. [CrossRef]
113. Vellinga, N.E. Rethinking Compensation in Light of the Development of AI. *Int. Rev. Law Comput. Technol.* **2024**, *38*, 391–412. [CrossRef]
114. Kim, S. Perceptions of Discriminatory Decisions of Artificial Intelligence: Unpacking the Role of Individual Characteristics. *Int. J. Hum. Comput. Stud.* **2025**, *194*, 103387. [CrossRef]
115. Wu, M.; Li, Z.; Yuen, K.F. Effect of Anthropomorphic Design and Hierarchical Status on Balancing Self-Serving Bias: Accounting for Education, Ethnicity, and Experience. *Comput. Hum. Behav.* **2024**, *158*, 108299. [CrossRef]
116. Lee, M.H.J. Examining the Robustness of Homogeneity Bias to Hyperparameter Adjustments in GPT-4. *arXiv* **2025**, arXiv:2501.02211. [CrossRef]
117. Oschinsky, F.M.; Stelter, A.; Niehaves, B. Cognitive Biases in the Digital Age—How Resolving the Status Quo Bias Enables Public-Sector Employees to Overcome Restraint. *Gov. Inf. Q.* **2021**, *38*, 101611. [CrossRef]
118. Fabi, S.; Hagendorff, T. Why We Need Biased AI How Including Cognitive and Ethical Machine Biases Can Enhance AI Systems. *arXiv* **2022**, arXiv:2203.09911. [CrossRef]
119. Kleinberg, J.; Oren, S.; Raghavan, M.; Sklar, N. Stochastic Model for Sunk Cost Bias. *PMLR* **2021**, *161*, 1279–1288. [CrossRef]

120. Gupta, P.; MacAvaney, S. On Survivorship Bias in MS MARCO. In Proceedings of the SIGIR 2022—Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, 11–15 July 2022; pp. 2214–2219. [CrossRef]
121. Phillips, I.; Upadhyayula, A.; Flombaum, J. Tachypsychia—The Subjective Expansion of Time—Happens in Immediate Memory, Not Perceptual Experience. *J. Vis.* **2020**, *20*, 1466. [CrossRef]
122. Choi, S. Temporal Framing in Balanced News Coverage of Artificial Intelligence and Public Attitudes. *Mass Commun. Soc.* **2024**, *27*, 384–405. [CrossRef]
123. Kodden, B. The Art of Sustainable Performance: The Zeigarnik Effect. In *The Art of Sustainable Performance*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 67–73. [CrossRef]
124. Korteling, J.E.; Paradies, G.L.; Sassen-van Meer, J.P. Cognitive Bias and How to Improve Sustainable Decision Making. *Front. Psychol.* **2023**, *14*, 1129835. [CrossRef] [PubMed]
125. Ye, A.; Maiti, A.; Schmidt, M.; Pedersen, S.J. A Hybrid Semi-Automated Workflow for Systematic and Literature Review Processes with Large Language Model Analysis. *Future Int.* **2024**, *16*, 167. [CrossRef]
126. Kücking, F.; Hübner, U.; Przysucha, M.; Hannemann, N.; Kutza, J.O.; Moelleken, M.; Erfurt-Berge, C.; Dissemond, J.; Babitsch, B.; Busch, D. *Automation Bias in AI-Decision Support: Results from an Empirical Study*; IOS Press: Amsterdam, The Netherlands, 2024. [CrossRef]
127. Chuan, C.H.; Sun, R.; Tian, S.; Tsai, W.H.S. EXplainable Artificial Intelligence (XAI) for Facilitating Recognition of Algorithmic Bias: An Experiment from Imposed Users' Perspectives. *Telemat. Inform.* **2024**, *91*, 102135. [CrossRef]
128. Daniil, S.; Slokom, M.; Cuper, M.; Liem, C.C.S.; van Ossenbruggen, J.; Hollink, L. On the Challenges of Studying Bias in Recommender Systems: A UserKNN Case Study. *arXiv* **2024**, arXiv:2409.08046. [CrossRef]
129. Roth, B.; de Araujo, P.H.L.; Xia, Y.; Kaltenbrunner, S.; Korab, C. Specification Overfitting in Artificial Intelligence. *Artif. Intell. Rev.* **2024**, *58*, 35. [CrossRef]
130. Li, S. Computational and Experimental Simulations in Engineering. In Proceedings of the ICCES 2023, Shenzhen, China, 26–29 May 2023; Volume 145. [CrossRef]
131. Wang, B.; Liu, J. Cognitively Biased Users Interacting with Algorithmically Biased Results in Whole-Session Search on Debated Topics. In Proceedings of the ICTIR 2024—Proceedings of the 2024 ACM SIGIR International Conference on the Theory of Information Retrieval, Washington, DC, USA, 13 July 2024; Volume 1, pp. 227–237. [CrossRef]
132. Kacperski, C.; Bielig, M.; Makhortykh, M.; Sydorova, M.; Ulloa, R. Examining Bias Perpetuation in Academic Search Engines: An Algorithm Audit of Google and Semantic Scholar. *First Monday* **2023**, *29*, 11. [CrossRef]
133. Suresh, H.; Guttag, J. A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle. In Proceedings of the ACM International Conference Proceeding Series, New York City, NY, USA, 20–24 October 2021. [CrossRef]
134. van Stein, N.; Thomson, S.L.; Kononova, A.V. A Deep Dive into Effects of Structural Bias on CMA-ES Performance along Affine Trajectories. *arXiv* **2024**, arXiv:2404.17323. [CrossRef]
135. Soleymani, H.; Saeidnia, H.R.; Ausloos, M.; Hassanzadeh, M. Selective Dissemination of Information (SDI) in the Age of Artificial Intelligence (AI). *Library Hi Tech News* 2023, *ahead-of-print*. [CrossRef]
136. Beer, P.; Mulder, R.H. The Effects of Technological Developments on Work and Their Implications for Continuous Vocational Education and Training: A Systematic Review. *Front. Psychol.* **2020**, *11*, 535119. [CrossRef]
137. Marcinkevičs, R.; Vogt, J.E. Interpretable and Explainable Machine Learning: A Methods-Centric Overview with Concrete Examples. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2023**, *13*, e1493. [CrossRef]
138. Hort, M.; Chen, Z.; Zhang, J.M.; Harman, M.; Sarro, F. Bias Mitigation for Machine Learning Classifiers: A Comprehensive Survey. *ACM J. Responsible Comput.* **2022**, *1*, 11. [CrossRef]
139. Mosqueira-Rey, E.; Hernández-Pereira, E.; Alonso-Ríos, D.; Bobes-Bascarán, J.; Fernández-Leal, Á. Human-in-the-Loop Machine Learning: A State of the Art. *Artif. Intell. Rev.* **2023**, *56*, 3005–3054. [CrossRef]
140. Minkkinen, M.; Laine, J.; Mäntymäki, M. Continuous Auditing of Artificial Intelligence: A Conceptualization and Assessment of Tools and Frameworks. *Digit. Soc.* **2022**, *1*, 21. [CrossRef]
141. Casper, S.; Davies, X.; Shi, C.; Gilbert, T.K.; Scheurer, J.; Rando, J.; Freedman, R.; Korbak, T.; Lindner, D.; Freire, P.; et al. Open Problems and Fundamental Limitations of Reinforcement Learning from Human Feedback. *arXiv* **2023**, arXiv:2307.15217. [CrossRef]
142. Feldman, T.; Peake, A. End-To-End Bias Mitigation: Removing Gender Bias in Deep Learning. *arXiv* **2021**, arXiv:2104.02532. [CrossRef]
143. Khakurel, U.; Abdelmoumin, G.; Rawat, D.B. Performance Evaluation for Detecting and Alleviating Biases in Predictive Machine Learning Models. *ACM Trans. Probabilistic Mach. Learn.* **2025**, *1*, 1–34. [CrossRef]
144. Demircioğlu, A. Applying Oversampling before Cross-Validation Will Lead to High Bias in Radiomics. *Sci. Rep.* **2024**, *14*, 11563. [CrossRef] [PubMed]

145. Zhang, C.; Kim, J.; Jeon, J.H.; Xing, J.; Ahn, C.; Tang, P.; Cai, H. Toward Integrated Human-Machine Intelligence for Civil Engineering: An Interdisciplinary Perspective. In Proceedings of the Computing in Civil Engineering 2021—Selected Papers from the ASCE International Conference on Computing in Civil Engineering 2021, Orlando, FL, USA, 12–14 September 2021; pp. 279–286. [CrossRef]
146. Li, X.; Yang, C.; Møller, C.; Lee, J. Data Issues in Industrial AI System: A Meta-Review and Research Strategy. *arXiv* **2024**, arXiv:2406.15784. [CrossRef]
147. Rosado Gomez, A.A.; Calderón Benavides, M.L. Framework for Bias Detection in Machine Learning Models: A Fairness Approach. In Proceedings of the WSDM 2024—Proceedings of the 17th ACM International Conference on Web Search and Data Mining, Merida, Mexico, 4–8 March 2024; pp. 1152–1154. [CrossRef]
148. Razavi, S.; Jakeman, A.; Saltelli, A.; Prieur, C.; Iooss, B.; Borgonovo, E.; Plischke, E.; Lo Piano, S.; Iwanaga, T.; Becker, W.; et al. The Future of Sensitivity Analysis: An Essential Discipline for Systems Modeling and Policy Support. *Environ. Model. Softw.* **2021**, *137*, 104954. [CrossRef]
149. Siddique, S.; Haque, M.A.; George, R.; Gupta, K.D.; Gupta, D.; Faruk, M.J.H. Survey on Machine Learning Biases and Mitigation Techniques. *Digital* **2024**, *4*, 1. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.