Article

Understanding the Role of Objectivity in Machine Learning and Research Evaluation

Saleha Javed *,† **, Tosin P. Adewumi** † **, Foteini Simistira Liwicki and Marcus Liwicki**

Machine Learning Group, Department of Computer Science, Electrical and Space Engineering, EISLAB, Luleå University of Technology, 97187 Luleå, Sweden; oluwatosin.adewumi@ltu.se (T.P.A.); foteini.liwicki@ltu.se (F.S.L.); marcus.liwicki@ltu.se (M.L.)
* Correspondence: saleha.javed@ltu.se
† These authors contributed equally to this work.

Abstract: This article makes the case for more objectivity in Machine Learning (ML) research. Any research work that claims to hold benefits has to be scrutinized based on many parameters, such as the methodology employed, ethical considerations and its theoretical or technical contribution. We approach this discussion from a Naturalist philosophical outlook. Although every analysis may be subjective, it is important for the research community to keep vetting the research for continuous growth and to produce even better work. We suggest standardizing some of the steps in ML research in an objective way and being aware of various biases threatening objectivity. The ideal of objectivity keeps research rational since objectivity requires beliefs to be based on facts. We discuss some of the current challenges, the role of objectivity in the two elements (product and process) that are up for consideration in ML and make recommendations to support the research community.

1. Introduction

The subject of objectivity has long been considered by philosophers and researchers over time [1,2]. Objectivity is fundamental in the philosophy of science. The subject is very important if researchers are to ascertain the facts of the real world or ML researchers are to be more credible in a field sometimes viewed with suspicion for "blackbox" models. Indeed, some fundamental problems in ML are identified as philosophical [3]. These are:

1. What is knowledge?
2. Can knowledge be acquired from data?
3. What is a good teacher?
4. How to distinguish true scientific theory from false one?
5. How to form good inductive theories?

ML involves the use of non-explicit, computer algorithms trained on sample data to iteratively perform better at a given task [4,5]. It makes an inductive inference from samples of data [6]. More objectivity will help in answering the identified questions.

In this article, we use the standpoint of Naturalist philosophical theory to evaluate how ML, as a discipline, operates. Naturalism is a theory of knowledge containing imagination, belief, knowledge, and uncertainty [4,7,8]. Many of the ideas in Naturalism are testable and corroborated by science for their usefulness. The untested assumptions central to Naturalism give legitimacy to scientific systems, defining boundaries of investigation [7]. For example, a basic assumption is that random sampling is representative for a given population [9]. Other ML assumptions (which are not directly bound to naturalism, but may be limitations, in some cases) are: the future is similar to the past, there's a cost of error called the loss function (describing the difference between the ground truth and the predicted output) and, finite (past) data is available for training [3,10].

The contributions of this article include the identification of weaknesses, discussed later, in current ML practices and recommendations towards a more objective approach. The following sections include a brief exposition on objectivity, the methodology in ML, the discussion about the two elements, (1) product objectivity and (2) process objectivity, under consideration, and the conclusion. The brief exposition on objectivity discusses, at a higher level, some views of what objectivity is, including Longino's; the methodology in ML highlights methodology trends and some metrics used in ML; and the discussion section argues for an objective, standard approach in ML.

2. Brief Exposition on Objectivity

Product objectivity and process objectivity form two key parts of objectivity [1,11,12]. The objectivity of both product (entities of accurate representations of the world) and process (methods devoid of subjective views) are critical to the overall objectivity of science [1]. Process objectivity is the basis of the scientific method utilised in ML [1]. The scientific method involves developing and accepting or rejecting a scientific view through non-whimsical and non-subjective criteria [11]. Experiments based on rapidly-evolving assumptions have no objectivity, according to Longino [11]. Objectivity eliminates biases in its many forms, like cognitive or sampling biases, as much as possible. 'Let nature speak for itself' of the nineteenth century [13] may be equivalent to 'let the model speak for itself' of today in ML. This is because the evaluation of the performance of a given model will show if it's biased or not, as it has been witnessed with some of the state-of-the-art (SotA) models [14]. Objectivity should make us act in certain ways, just as subjectivity will have us behave in certain ways.

What we know by experience is usually not detached from how we got to know it or is relayed or learned through existing natural means [1]. Hence, the extent to which one can be objective is often in relative terms. Haraway [15] is of the view that the pursuit of objectivity, in which the subject is completely cut off from the object, is an illusion. She, therefore, calls for a balance where there will be faithful accounts of the object and the subject's perspective, in order for the subject to take responsibility. This is more so that the sociological factors and apparatus of science are perspectival [1]. It has been argued in the literature that this perspectival attribute of the facts of science, which makes absolute (process) objectivity unlikely, makes it crucial to have an active community (including its peer-review mechanism) that guides on how best to achieve objectivity [4,11]. With regards to product objectivity, when the fact (data), and only the fact, determines our belief, we are objective. Longino, however, reports that careful consideration should be given to the reliability of what actually makes up this fact [11].

Objectivity cannot be unconnected to ethics. It is important to be objective in order to say one has acted ethically, at least, from the view of deontological ethics [16]. Research ethics, which studies the ethical problems and issues that arise from conducting research [17], makes it possible for the research community to provide guidelines for conducting research ethically. This is important for reproducibility of experimental results, where it is the facts that should matter.

Longino's Viewpoint on Dissent Bringing Objectivity

Helen Elizabeth Longino is one of those philosophers of science who emphasized the significance of the social and moral values of research. For her, getting process objectivity right is essential. She places importance on criticism when it comes to validating or even understanding any contemporary research. Whilst analyzing any theory, the key idea is to keep in mind that the production of knowledge is a social property that will have an impact on society. The products of research have to go across continuous, cooperative and critical inquiries from all points of view, after which it may be termed objective, having managed to hold its ground against criticisms [18].

According to Longino, two shifts of perspective make it possible to see how scientific knowledge or method is objective [19,20], as indicated in Table 1. The first shift is to return

to the idea of science as practice. Refocusing on science as practice makes the second shift possible, which regards scientific method as something practised, not primarily by individuals, but by social groups. Longino pointed out tools to dissent to research work, which may lead it to be practically objective. These tools are briefly described below:

Table 1. Longino's Two Shifts of Perspective To See How Scientific Method is Objective.

$\overrightarrow{FirstShift}$	Focus on Science As Practice
$\overleftarrow{SecondShift}$	Refocus on Scientific Method Practiced by Social Group

1. **Social Knowledge (Accept/Reject):**
   Objectivity neither belongs to one single person nor can it be ensured by an individual: it is a community's practice. Any knowledge conceived or presented, by an individual, moulds into social knowledge when it passes through evaluation by others. Such evaluations often improve and reshape the knowledge into even more valuable and productive work. Even a rejected hypothesis serves a role in future research, as the research community becomes more and more aware of the shortcomings of certain biased results as well.

2. **Criticism**
   It is crucial for objectivity to reduce the influence of subjective preference of individual background beliefs or assumptions. Although criticism may not completely eliminate the influence of subjective preferences, especially in fields like ML where certain parameters/settings have to be optimised and controlled, it provides the means for evaluating how much influence they have over the formation of the scientific contribution. Criticism should cover every relevant aspect of research.

3. **Shared Standards**
   Relevant criticism appeals to what is accepted by those concerned. Standards held on to by a community make members of such a community responsible for the goals of the community.

4. **Recognised Avenues**
   Research communities form different platforms, based on collective interests, and these platforms are responsible for verifying and validating their members' work. This helps to maintain worthy standards. Platforms, like public forums, conferences, peer-review journals, also compete among one another based on the credibility and social benefits they bring [11]. Part of the purpose of all these activities is to extract effective criticism from community members and prevent idiosyncratic values from shaping knowledge.

5. **Role of the Community**
   The role of the community is quite significant as they are the receiver of the outcome of research and may be affected by it. As difficult as determining community response may be due to its qualitative nature, some of the ways through which we can quantify the response and evaluate research may be through grants scored, number of publications, contribution to textbook contents and awards by a scientific avenue. Such quantitative measures provide relatively simple ways of comparing factors [21]. However, there are limitations to quantitative measures [22]. For example, the number of publications does not tell the whole story, as there are other important factors, like the quality of the medium/journal.

3. Methodology in ML

The system of methods in ML is founded on similar principles that guide science as a whole. Central to the ML research lifecycle is the relationship between input data and output [5]. This relationship is modelled by a function, which may be expressed as an algorithm. The astronomer Johannes Kepler, born in the 16th century, put forward their 3 laws of motion by adopting the simplest model that could fit the data he obtained from

Tycho Brahe [5]. Kepler's scientific approach, approximately 400 years ago, is similar to what obtains in ML [5,23]. This can be represented in the following algorithm:

1.　Collect data and analyse,
2.　Split the data so as to have a final set to confirm predictions,
3.　Apply a suitable model to the data,
4.　Iterate over attributes of the model that gives best fit
5.　Then validate the model on the held-out split data

This lifecycle is, more or less, the practice for the different tasks in the different areas of ML, with some minor differences, including in computer vision (CV) and natural language processing (NLP).

Various metrics are used to measure performance, depending on the task at hand. Some of the metrics include accuracy, bilingual evaluation understudy (BLEU) [24], Precision (how many selected items are relevant), Recall (how many relevant items are selected), F1 (harmonic mean of Precision and Recall), normalised mutual information (NMI) and mean average precision (mAP) [25]. Statistical significance tests are very important because some of the theoretical foundations and practical implementations of ML are dependent on statistics and randomness [23]. Some researchers, like Reimers and Gurevych [26], argue that significance tests have weaknesses and we agree, however, conducting no significance test (or statistical analysis) is worse and should not be an option. Indeed, they also admit there's no perfect evaluation system [26].

4. Discussion

As discussed above, the two distinct elements under consideration are (1) product objectivity and (2) process objectivity when speaking of objectivity in ML, and they have their own questions with different scope. There are the input, models and outputs of ML, on the one hand. On the other hand, we have the researchers, who develop ML systems, and their research program. Machines have no gender and are not social beings, so machine learning holds out the possibility of the sort of objectivity science seeks for. However, investigations have shown that ANNs display biases seen in the data they are trained with, just as humans display biases dependent on their experiences [5]. Hence, biases can slip in through the "back door". Hope is not lost, however. The solution will require a version of Longino's community-based process to make the results of ML more objective than we could get with humans. Longino observed that "scientists speak of the objectivity of data" ([11], p. 171), [27]. According to her, checking to see that the data seem right is one of the functions of peer-review. She further suggests that a method is not necessarily objective just because it is empirical [11]. Furthermore, according to [12], process objectivity influences product objectivity such that, adherence to a process that is designed to produce objectivity gives objective results. The converse is also true. The ANN, data and their results (or outputs) will be discussed under product objectivity in the following paragraph while the research processes and the ML community will be discussed under process objectivity in the latter paragraph.

An important question regarding the first element is, "Do inductive inferences that do not require a human brain avoid the sorts of cognitive biases that accompany those made by social beings?" We have pointed out that investigations have revealed that this is not the case, as biases are noticeable in models and the data used to train them [5,14]. This makes Longino's argument about what constitutes the "fact" in ML pungent [11], given that they form representations of the world. Furthermore, her first tool of social knowledge in the earlier section is important here for evaluating experimental data, presented results and conclusions made. Criticism, the second tool, ensures scrutiny of the "facts", based on shared standards through recognised avenues for doing so. Hence, publishing codes of models/algorithms and datasets used in the experiments goes a long way in securing credibility, an essential prerequisite to objectivity [28,29].

For the second element, process objectivity, an important question will be, "Is the research community associated with ML guilty of gender-based biases (or other biases)

and if so, how do we make a less biased research community?" Again, Longino's tools for dissent, mentioned earlier, are helpful here [19]. The tools are interwoven such that they complement one another in accomplishing the goal of objectivity in science, and in this case, ML. For credibility and reproducibility, it is crucial that the details of methods of research are published, including the hardware involved. Shared standards, the third tool, make it possible to vet the methods employed to establish if the experiments and their results are reproducible. Furthermore, key practices, some of which were mentioned in the earlier section should be adhered to, especially with regards to the held-out split data for validation to avoid test set feedback on training [30]. There have been many problems of reproducibility in the field, though this is not limited to ML alone [29]. Unfair comparison is another common problem in the field. As rightly pointed out by Musgrave et al. [30], when comparing two or more entities or models (or algorithms), it is best to keep as many related factors as possible constant.

In addition, statistical analysis, such as significance test or confidence interval (C.I.), should be carried out. C.I. gives the range of values where the parameter in question (such as the mean or variance) would be [27]. Although statistical analyses do not show proof of anything experimentally, they provide measurement of the likely error in a conclusion or the level of confidence attached to a statement [27]. As an example, consider the recent results of two models from word embeddings experiments, where the test set F1 scores for a default model *A* and another model *B* are 0.661 and 0.676, respectively. In this example, the scores reported are averages of several observations, though some research have reported single observations [26]. A speedy conclusion can be made in certain quarters that *B* outperforms *A* with such a small difference without any statistical analysis, even though there's randomness in the nature of these experiments. Even a large difference still warrants a statistical significance test, though the difference is more persuasive than when there's a small difference.

As one might observe, it is a mistake to think that ML automatically gives us objectivity, but if the prescriptions are followed, ML can get us much closer to the desired objectivity. Pushing the boundaries of ML objectivity would be transparent or explainable artificial intelligence (AI), as some researchers have been doing, by publishing qualitative or insightful pieces of their research and their limitations [31,32].

Scrutinizing the Role of Bias against Objectivity

As discussed above, objectivity holds vital significance in science. Thus, a critical question concerns whether there is a rule of thumb for someone to be objective, and if not, whether we can device one. This rule of thumb or model can assist researchers or evaluators of research work in ascertaining the degree of its objectivity. A possible model applicable to ML from the available literature is to use the common Likert scale, say, from 1 to 5. The objective will be to determine which of the qualities have the strongest presence or relevance in a piece of work, where 1 indicates the low relevance and 5 the high relevance.

Bias, in this context, refers to a partial view that prevents objectivity [33]. Cognitive bias, for example, comes in many different forms [34]. While it is out of the scope of this work to discuss the extensive subject of bias, Lavesson and Davidsson, including other researchers, provide helpful surveys on the subject [35]. Bias can be present in the data, the algorithm or the methods involved in a research work. Below is a list of a few biases and Tables 2 and 3 give examples of the two Likert scales with a few qualities. The qualities represented on the scale may be selected based on relevance for, say, the researcher or the funding agency, who will use it at the relevant stage in the research process (say, the award of grants).

1. **Creative Bias:**
   This takes a negative view of creative ideas and projects, relative to those that are more practical [36,37].
2. **Qualitative Bias:**
   Deliberate or erroneous effort to influence the result of a piece of research work for a

desired outcome may be termed qualitative bias. Such bias severely affects the whole research community and society and is strictly prohibited [38,39].

3. **Procedural Bias:**
   This is a type of inductive bias that refers to the influence of the order of a set of steps defined in the presented model of any work [40].

4. **Ground Truth Bias:**
   Ground truth refers to the held-out data that is often used as the basis for comparison in ML tasks. Custom tailoring of the designed model or even the training data to optimise results for a given ground truth is a bias that is simply unacceptable because of the potential dangers during application [41].

Meanwhile, a few examples of desirable qualities are also listed below, depending on what part of the research process is involved.

1. **Socially Ethical:**
   Ethics plays a major role in scientific work and various guidelines for research ethics are specified by research communities [42].

2. **Credible References:**
   Use of credible citations and references, such as from peer-reviewed journals, raises the quality of any research publication. Credibility may be determined by the h-index or number of citations achieved by a research work [43,44], however, care must be taken as this is not always the case.

3. **Peer Review:**
   Peer review holds immense value when it comes to evaluating any research work. Although it has its short-comings, it is still a useful tool for maintaining standards in research communities [45,46].

Table 2. The 'Desirable Measures' of Evaluation for Degree of Objectivity.

	Degree				
Evaluation Key	**1—Least Applicable**	**2**	**3**	**4**	**5—Strongly Applicable**
Socially Ethical					
Credible References					

Table 3. The 'Undesirable Measures' of Evaluation for Degree of Objectivity.

	Degree				
Evaluation Key	**1—Least Applicable**	**2**	**3**	**4**	**5—Strongly Applicable**
Creative Bias					
Qualitative Bias					

5. Conclusions

As ML researchers, pushing the degree of objectivity higher in our research should be a priority. The facts of the data should shape our belief. Naturalism affords us certain assumptions upon which we can build knowledge of the systems around us. These assumptions and the scientific knowledge in the ML field cannot be antithetical to each other. Reproducible and transparent ML will be essential to its future success, besides being ethical. Therefore, more commitment on the part of the ML community (including the peer-review mechanism) is required to achieve the level of objectivity that will be satisfactory to all, or at least, most practitioners.

Abbreviations

The following abbreviations are used in this manuscript:

ANN	Artificial Neural Network
BLEU	Bilingual Evaluation Understudy
ML	Machine Learning
NLP	Natural Language Processing

References

1. Reiss, J.; Sprenger, J. Scientific Objectivity. In *The Stanford Encyclopedia of Philosophy*, Winter 2017 ed.; Zalta, E.N., Ed.; Metaphysics Research Lab, Stanford University: Stanford, CA, USA, 2017.
2. Parry, W.T.; Hacker, E.A. *Aristotelian Logic*; Suny Press: Albany, NY, USA, 1991.
3. Cherkassky, V.; Mulier, F.M. *Learning from Data: Concepts, Theory, and Methods*; John Wiley & Sons: Hoboken, NJ, USA, 2007.
4. Adewumi, T.P.; Liwicki, F.; Liwicki, M. Conversational Systems in Machine Learning from the Point of View of the Philosophy of Science—Using Alime Chat and Related Studies. *Philosophies* **2019**, *4*, 41. [CrossRef]
5. Stevens, E.; Antiga, L. *Deep Learning with PyTorch: Essential Excerpts*; Manning Publications: Shelter Island, NY, USA, 2019.
6. Thagard, P. Philosophy and machine learning. *Can. J. Philos.* **1990**, *20*, 261–276. [CrossRef]
7. Creath, R. Logical Empiricism. In *The Stanford Encyclopedia of Philosophy*, Summer 2020 ed.; Zalta, E.N., Ed.; Metaphysics Research Lab, Stanford University: Stanford, CA, USA, 2020.
8. Hartman, J. *Nature and Naturalism: A Philosophy Primer*; lulu.com: Morrisville, NC, USA, 2013.
9. Kazmier, L.J. *Schaum's Outline of Theory and Problems of Business Statistics*; McGraw-Hill: New York, NY, USA, 1976.
10. Cherkassky, V.; Friedman, J.H.; Wechsler, H. *From Statistics to Neural Networks: Theory and Pattern Recognition Applications*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012; Volume 136.
11. Longino, H.E. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*; Princeton University Press: Princeton, NJ, USA, 1990.
12. Shamoun, D.; Calabrese, E. On Objective Risk. 2015. Available online: https://www.mercatus.org/system/files/Shamoun-On-Objective-Risk_0.pdf (accessed on 15 March 2021) .
13. Daston, L.; Galison, P. The image of objectivity. *Representations* **1992**, *40*, 81–128. [CrossRef]
14. Zhang, Y.; Sun, S.; Galley, M.; Chen, Y.C.; Brockett, C.; Gao, X.; Gao, J.; Liu, J.; Dolan, B. Dialogpt: Large-scale generative pre-training for conversational response generation. *arXiv* **2019**, arXiv:1911.00536.
15. Haraway, D. Situated knowledges: The science question in feminism and the privilege of partial perspective. *Fem. Stud.* **1988**, *14*, 575–599. [CrossRef]
16. White, M.D. 40 Immanuel Kant. In *Handbook of Economics and Ethics*; Edward Elgar Pub: Cheltenham, UK, 2009; p. 301. ISBN 9780198793991.
17. Shamoo, A.E.; Resnik, D.B. *Responsible Conduct of Research*; Oxford University Press: Oxford, UK, 2009.
18. Jukola, S. Longino's theory of objectivity and commercialized research. In *Empirical Philosophy of Science*; Springer: Berlin, Germany, 2015; pp. 127–143.
19. Longino, H.E. *The Fate of Knowledge*; Princeton University Press: Princeton, NJ, USA, 2002.
20. Segerstrale, U.; Longino, H. Book review: Science as Social Knowledge: Values and Objectivity in Scientific Inquiry. *Contemp. Sociol.* **1992**, *21*, 138. [CrossRef]
21. Bornmann, L.; Mutz, R. Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *J. Assoc. Inf. Sci. Technol.* **2015**, *66*, 2215–2222. [CrossRef]

22. Zahedi, Z.; Costas, R.; Wouters, P. How well developed are altmetrics? A cross-disciplinary analysis of the presence of 'alternative metrics' in scientific publications. *Scientometrics* **2014**, *101*, 1491–1513. [CrossRef]
23. Hagan, M.T.; Demuth, H.B.; Beale, M.; De Jesus, O. *Neural Network Design*, 2nd ed.; PWS Publishing Co.: Boston, MA, USA, 2000.
24. Papineni, K.; Roukos, S.; Ward, T.; Zhu, W.J. BLEU: A method for automatic evaluation of machine translation. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, Stroudsburg, PA, USA, 7–12 July 2002; pp. 311–318.
25. Indurkhya, N.; Damerau, F.J. *Handbook of Natural Language Processing*; CRC Press: Boca Raton, FL, USA, 2010; Volume 2.
26. Reimers, N.; Gurevych, I. Why comparing single performance scores does not allow to draw conclusions about machine learning approaches. *arXiv* **2018**, arXiv:1803.09578.
27. Douglas, C.M. *Design and Analysis of Experiments*; Wiley: Hoboken, NJ, USA, 2001; ISBN 978-1-119-63542-0.
28. McDermott, M.; Wang, S.; Marinsek, N.; Ranganath, R.; Ghassemi, M.; Foschini, L. Reproducibility in machine learning for health. *arXiv* **2019**, arXiv:1907.01463.
29. Olorisade, B.K.; Brereton, P.; Andras, P. Reproducibility in Machine Learning-Based Studies: An Example of Text Mining. In Proceedings of the MLWorkshop at the 34th International Conference on Machine Learning (ICML2017), Sydney, Australia, 6–11 August 2017. Available online: https://www.semanticscholar.org/paper/Reproducibility-in-Machine-Learning-Based-Studies%3A-Olorisade-Brereton/2502db705985d6736b1cc9c8e3f13f617a6a9a98 (accessed on 15 March 2021).
30. Musgrave, K.; Belongie, S.; Lim, S.N. A metric learning reality check. *arXiv* **2020**, arXiv:2003.08505.
31. Adewumi, T.P.; Liwicki, F.; Liwicki, M. Exploring Swedish & English fastText Embeddings with the Transformer. *arXiv* **2020**, arXiv:2007.16007.
32. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient estimation of word representations in vector space. *arXiv* **2013**, arXiv:1301.3781.
33. Dictionary, C. Cambridge advanced learner's dictionary. In *PONS-Worterbucher, Klett Ernst Verlag GmbH*; Cambridge University Press: Cambridge, UK, 2008.
34. Blackwell, S.E.; Woud, M.L.; MacLeod, C. A question of control? Examining the role of control conditions in experimental psychopathology using the example of cognitive bias modification research. *Span. J. Psychol.* **2017**, *20*, E54. [CrossRef]
35. Lavesson, N.; Davidsson, P. Evaluating learning algorithms and classifiers. *Int. J. Intell. Inf. Database Syst.* **2007**, *1*, 37–52. [CrossRef]
36. de Jesus, S.N.; Rus, C.L.; Lens, W.; Imaginário, S. Intrinsic motivation and creativity related to product: A meta-analysis of the studies published between 1990–2010. *Creat. Res. J.* **2013**, *25*, 80–84. [CrossRef]
37. Hennessey, B.A. *The Creativity—Motivation Connection*; Cambridge University Press: Cambridge, UK, 2010. [CrossRef]
38. Morse, J.M. Critical analysis of strategies for determining rigor in qualitative inquiry. *Qual. Health Res.* **2015**, *25*, 1212–1222. [CrossRef]
39. Onwuegbuzie, A.J.; Leech, N.L. Validity and qualitative research: An oxymoron? *Qual. Quant.* **2007**, *41*, 233–249. [CrossRef]
40. Gordon, D.F.; Desjardins, M. Evaluation and selection of biases in machine learning. *Mach. Learn.* **1995**, *20*, 5–22. [CrossRef]
41. Lei, Y.; Bezdek, J.C.; Romano, S.; Vinh, N.X.; Chan, J.; Bailey, J. Ground truth bias in external cluster validity indices. *Pattern Recognit.* **2017**, *65*, 58–70. [CrossRef]
42. Adler, E.; Clark, R. *How It's Done: An Invitation to Social Research*; Cengage Learning: Boston, MA, USA, 2007.
43. Rubin, H.R.; Redelmeier, D.A.; Wu, A.W.; Steinberg, E.P. How reliable is peer review of scientific abstracts? *J. Gen. Intern. Med.* **1993**, *8*, 255–258. [CrossRef] [PubMed]
44. Ware, M. *Peer Review: Benefits, Perceptions and Alternatives*; Citeseer: Princeton, NJ, USA, 2008.
45. Balci, O. Verification, validation, and accreditation. In Proceedings of the 1998 Winter Simulation Conference, Proceedings (Cat. No.98CH36274), Washington, DC, USA, 13–16 December 1998; Volume 1, pp. 41–48.
46. Anney, V.N. Ensuring the quality of the findings of qualitative research: Looking at trustworthiness criteria. *J. Emerg. Trends Educ. Res. Policy Stud.* **2014**, *5*, 272–281.