Bias Amplification Chains in ML-based Systems with an Application to Credit Scoring*

Alessandro G. Buda[1,2,†], Greta Coraglia[1,2], Francesco A. Genco[1,2], Chiara Manganini[1,2] and Giuseppe Primiero[1,2]

[1]LUCI Lab, Department of Philosophy, University of Milan, via Festa del Perdono 7, 20122 Milan, Italy
[2]MIRAI S.r.l.

Abstract

Machine Learning (ML) systems, whether predictive or generative, not only reproduce biases and stereotypes but, even more worryingly, amplify them. Strategies for bias detection and mitigation typically focus on either *ex post* or *ex ante* approaches, but are always limited to two steps analyses. In this paper, we introduce the notion of Bias Amplification Chain (BAC) as a series of steps in which bias may be amplified during the design, development and deployment phases of trained models. We provide an application to such notion in the credit scoring setting and a quantitative analysis through the BRIO tool.

Keywords

ML Fairness, Bias Amplification, Responsible AI,

1. Introduction

In recent years, research in the field of artificial intelligence has seen its impact grow in the lives of many people. In particular, algorithmic fairness, and its negative counterpart, algorithmic bias, have become hot topics: many experts, as well as ordinary people, have realized that systems based on *Machine Learning* (ML), whether predictive or generative, not only reproduce biases and stereotypes but, even more worryingly, amplify and reinforce them. This phenomenon, known as bias amplification, urgently requires addressing to start relying on automatic systems.

In academic research, risks of bias have been outlined for a decade now and solutions are being sought, following two main approaches to algorithmic fairness:

- an *ex post* approach, in which fairness metrics are defined to identify and mitigate the presence of bias [1, 2], and
- an *ex ante* approach, which sees algorithmic bias "as *evidence* of the underlying social and technical conditions that (re)produce it" [3, p.2], focusing on eXplainable Artificial Intelligence (XAI).

Both of these approaches, despite their great differences, are characterized by identifying an ideal, *abstract, probabilistic model* to which an *empirical, non-deterministic result* must come as close as possible. In the former approach, the evaluation of distance between the two is performed without assuming any knowledge of the underlying model. In the latter approach, the analysis seeks to identify a transparent underlying structure which minimizes distance from a given justification. Standard theoretical computer science and philosophy of computing terminology [4, 5] have called these two terms of comparison *Levels of Abstractions* (LoAs); the complex structure of ML systems has already

3rd Workshop on Bias, Ethical AI, Explainability and the Role of Logic and Logic Programming (BEWARE24), co-located with AIxIA 2024, November 25-28, 2024, Bolzano, Italy

*Work funded by the PRIN project n.2020SSKZ7R BRIO (Bias, Risk and Opacity in AI), PRIN Project n. 20223E8Y4X SMARTEST (Simulation of Probabilistic Systems for the Age of the Digital Twin), and through the Project "Departments of Excellence 2023-2027" awarded to the Department of Philosophy "Piero Martinetti" of the University of Milan.

†Corresponding author.

alessandro.buda@iusspavia.it (A. G. Buda); greta.coraglia@unimi.it (G. Coraglia); francesco.genco@unimi.it (F. A. Genco); chiara.manganini@unimi.it (C. Manganini); giuseppe.primiero@unimi.it (G. Primiero)

required a reconsideration of such terminology in terms of a User Level approach [6], which in turn motivates a future redesign of the whole ontology of such systems.

However, when dealing with complex AI systems, and especially in the case of generative AI, there are several levels to consider, and at each of these levels bias amplification phenomena can occur. We will refer to the bias amplification phenomenon in a system with more than one level as *Bias Amplification Chains* (BAC). Consider the case of text-to-image generation: one should first take into account that "[f]airness [...] is not a purely technical construct, having social, political, philosophical and legal facets" [7, p.1], and therefore consider how bias is reproduced and *pre-amplified* from society to training datasets, *via* image captioning [8]; secondly, one should take into consideration how bias present in training datasets is amplified in ML models due to model accuracy, model capacity, model overconfidence, amount of training data, and how such bias can vary throughout the training process [9]; then, one should focus on bias amplification occurring in generated images, primarily caused by discrepancies between training captions and text prompts [10]; lastly, one should consider the surprising fact that bias mitigation operations may themselves unexpectedly *boost* bias [8]. This behavior can also be seen in action in the predictive context, where a ML model is trained to predict properties that are relevant for taking decisions, especially in critical fields like finance, healthcare, or social justice.

A perspective similar to ours on the way bias propagates and amplifies throughout the life cycle of machine learning was given by [11], although the authors did not offer any detailed methodology to quantify the divergence between the bias in input and the one in output (respectively, the first and the last links of the BAC). On the other hand, our point of view differs from that of [12] and others, who argue for a *directional* approach, meaning one where causality of the amplification is taken into account: our analysis is causal in the sense that one can use the methods presented here to further investigate conditioning, but it should be noted that a compositional approach – meaning one allowing us to propagate that information through the various levels presented in Section 6 – is still lacking.

In the present work we describe and exemplify the case of bias amplification in the sense of [11] in a credit scoring setting, conducting our analysis by means of a recent bias detection tool called BRIO[1], which is a model-agnostic tool designed to assess the bias and related risk of unfairness of prediction tasks on tabular data. For a technical presentation of its basic features and a discussion on validation, we refer to [1]. The most recent developments of BRIO in the direction of assessing bias and risk can be instead found in [2]. This proprietary software has been developed on the basis of a family of logics for trustworthiness assessment [13, 14, 15, 16], and has already been used to assess fairness in credit scoring models [2].

To show our methodology, we examine a simple amplification chain consisting of just three links. The first link identifies the amplification of bias occurring in sampling the training dataset from the real population, thus identifying the input bias of our ML pipeline. The second link corresponds to the divergence between the distribution on which the model is trained and the one to which it is applied (test set). Lastly, the third link identifies the amplification of bias outputted by our model, that is, how the bias is amplified from the dataset used to perform the analysis to the predictions produced by the model. Our goal is achieved through the following functionalities:

- *FreqVsRef*: a functionality of BRIO which considers the first and the second link of the chain, by investigating how the dataset population distribution (Freq) amplifies bias on a property of interest as observed in the real population (Ref);
- *FreqVsFreq*: a functionality of BRIO which considers the second and third links of the chain, by investigating how the distribution divergence between the sensitive groups increases or decreases in the predicted outcome, with respect to the true distribution;
- an in-depth final analysis to inspect the entire chain, quantifying the contribution of each link to amplification, and identifying which of them most urgently requires mitigation interventions.

In practice, we conduct our analysis in the paradigmatic context of credit scoring. We use the UCI German Credit Dataset [17], despite several reported limitations [18, 19], because our purpose is not to

[1]The open source code is available at https://github.com/DLBD-Department/BRIO_x_Alkemy.

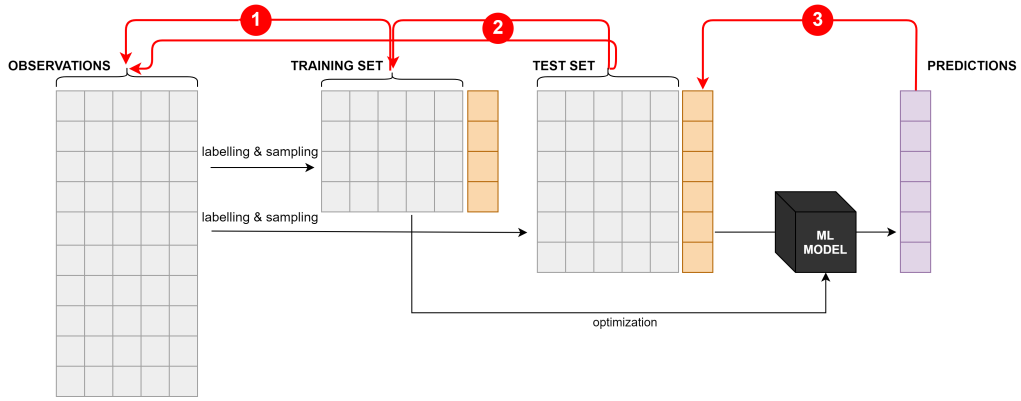

Figure 1: Illustration of three possible links of a Bias Amplification Chain (BAC). **Link 1**: a sensitive feature's distribution in the training set is skewed with respect to the observed distribution in the real population, generally due to a biased labelling or sampling process. **Link 2**: the inputs constituting the test set show a significantly different distribution from the one learnt by the ML model during training. **Link 3**: the model's predictions further amplify the bias as the predicted distribution of the outcomes diverges more from the desired one, compared to the ground truth distribution.

provide a comprehensive data analysis, but rather to show a novel methodology. The full outputs of the tool will remain available in the open `repository` of this project.

The paper is structured as follows. First, we introduce BRIO's main functionalities in Section 2. These will be used in subsequent Sections 3, 4, and 5 – each of which focuses on a specific link of the BAC – to compute the corresponding bias amplification. Finally, some final insights on the possible refinements of the notion of BAC will be given in Section 6.

2. The BRIO Method

BRIO's bias detection method takes in input

- the predictions of a ML model encoded as a set of datapoints with relative features,
- a set of parameters, including the designation of one or more *sensitive features*, also called *protected attributes*,
- and a distribution of reference, which might be automatically computed on the input dataset of the model, or externally provided,

On this input, it returns an evaluation of the possibility that the model under consideration is unfair with respect to the designated features when comparing the predictions and the reference distribution. To do so, the system allows conducting two kinds of analyses, consisting in:

1. *FreqVsRef*: comparing the behaviour of the AI system against a desirable one;
2. *FreqVsFreq*: comparing the behaviour of the AI system with respect to a sensitive group c_1 against another sensitive group c_2 related to the same feature F, with $F = c_1, c_2$.

If the second analysis alerts of a possibly biased behaviour, one can conduct a subsequent check on some (or all the) subclasses of the considered sensitive classes. This second check is meant to verify whether the bias encountered at the level of the classes can be explained away by non-sensitive features of the individuals that it is morally acceptable to use for taking decisions.

One can use the result of analyses 1 and 2 and, keeping track of the (sub)classes where they fail, compute:

3. *Hazard*: an aggregate measure of the number of datapoints where a test of type 1 or 2 fails, of how far it is from *not* failing on some datapoints (when so), and of how difficult was it for it to fail.

Age	Total population	Training set
0-24 years old	24.15%	15.00%
25-39 y.o.	19.17%	54.00%
40-59 y.o.	26.81%	26.00%
Over 60 y.o.	29.87%	5.00%

Gender	Total population	Training set
Female	50.65%	30.00%
Male	49.35%	70.00%

Marital status	Total population	Training set
Single	45.48%	54.00%
Married/separated/divorced/widowed	54.52%	46.00%

Table 1
Distributions of age, gender, and marital status in the German population and in the training set.

Of this latter measure, fully detailed in [2, §5], two options are provided, one focusing on *group fairness* and one on *individual fairness*. In the present work, we only consider the one involving group fairness[2].

Finally, it should be noted that, since the tool is model-agnostic, the same analyses 1, 2, and 3 described above can be applied to study the distribution of the *ground truth* (as it was the prediction of a perfect classifier). This feature will be used in Section 5.

3. Bias in Sampling the Training Set

In this section, we apply the FreqVsRef bias detection module of the BRIO tool in order to evaluate whether the process of labelling and sampling from the general population[3] to produce the training set is *biased* in the sense that it inaccurately represents the real distribution Bias is in fact mathematically defined as a systematic deviation from the true estimation. Although in the ML literature this term is typically used in relation to the decisions of a model, when it comes to the training set, we are interested in checking for bias contained in the *features*, rather than in the *labels* (or decisions). In fact, as Figure 1 illustrates, the reference distribution against which we want to compare the training set contains no labels.

For our experiments, we use the UCI German Credit Dataset [17], which offers a comprehensive compilation of attributes relevant for creditworthiness evaluation. The dataset comprises $1,000$ instances, each characterised by a set of 20 input variables and an associated binary label representing the occurrence or not of the default event. Our reference distributions are provided by statistics concerning the German society in 2024, see Table 1. It should be noted that this will eventually bring us to compare data from 2024 (our reference distribution) and data from the 1970s[4] (our training set based on UCI German Credit). While this is unfortunate, as obviously the two distribution differ in many sensible ways, we could not gather high quality data about general demographics in the desired period of time. Since the present work is expository of a general methodology, we shelf this issue as accidental.

[2]This is merely a choice in perspective, as we aim to describe possible discrimination of groups of people, instead of single individuals. From a technical point of view, BRIO allows for both.

[3]References for general demographic data come from

- https://www.statista.com/statistics/454349/population-by-age-group-germany/ (age)
- https://www.statista.com/statistics/454338/population-by-gender-germany/ (gender)
- https://www.destatis.de/EN/Themes/Society-Environment/Population/Current-Population/Tables/population-by-marital-status.html (marital status)

as of September 2024.

[4]One of the many problems with this dataset, is that it is usually attributed to the 1990s, while in reality its data was collected between 1973 and 1975 [18, §3.1] – when there were *two* Germanies!

Sensitive feature	Result of the test	Value of the divergence
Gender	violation	0.1005
Age	violation	0.3812
Marital status	violation	0.0300

Table 2
Computed divergences between [17] and Table 1.

The first step of the analysis consists in running the FreqVsRef bias detection module of the BRIO tool in order to evaluate whether the distribution of individuals in the different classes of the training set diverges too much from the distribution indicated by the statistics about German society. The method returns the respective Kullback-Leibler divergences[5] [20] with Laplacian smoothing, and compares it with the BRIO automated threshold set to *high* sensitivity, see Table 2.

The results show that the distributions of both gender and age in the training set do not match the corresponding distribution in real population: when further looking at the training set, in fact, one finds that males are extremely over-represented, and that people in age groups 40-59 and over 60 are greatly under-represented. The respective distributions of marital statuses are almost swapped.

4. Bias in the Test Set

The sampling to produce a test set for the prediction task is the next link of our BAC. By test set we mean the set of unseen data used to evaluate how the final model performs on unseen data, as opposed to the validation set which can be used to fine-tune the model itself. A major assumption of the ML approach is that the training and the test data share the same statistical distribution, being Independent and Identically Distributed (*i.i.d.*). However, in many practical applications, the *i.i.d.* assumption fails due to unforeseen distributional changes occurring once the ML system is deployed and used on new inputs (test set). This becomes relevant for our discussion because, in addition to a possible degradation of performance, the failure of *i.i.d.* might also amplify bias. We here distinguish two different perspectives from which the bias amplification occurring at the test set level can be thought of.

In a first sense, sampling unseen inputs for the test set may be amplifying bias in the properties of interest as they occur in the general population. In this case, an evaluation of this bias can be computed in the same vein as done for bias in the sampling of the training set.

In a second sense, the test set may amplify or compensate bias with respect to the training set. This is evidence of the fact that the sampling process of the test set was different from that producing the training set. Put differently, a model trained on a given distribution is then used to predict on a dataset which sensibly differs for some property of interest. This relates to the Out-of-Distribution problem, which occurs when the ML system is used on a distribution that significantly diverges from the one learnt during the training phase [21]. Evaluating such bias difference can be done through the BRIO tool in the same vein as done in the next phase of our analysis (Section 5).

Note that:

a. assuming a neutral bias amplification in sampling the training set, the bias input of the chain at the next stage will be ascribed entirely to the first sense of bias in the test set;
b. assuming a neutral bias difference in the second sense for the test set, the bias input of the chain at the next stage will be ascribed entirely to the second sense of bias in the test set.

[5]The Kullback–Leibler divergence D_{KL} is mathematically expressed by the following equation:

$$D_{\mathrm{KL}}(P \parallel Q) = \sum_{x \in X} P(x) \cdot \log(\frac{P(x)}{Q(x)})$$

and quantifies the discrepancy of a considered probability distribution Q with respect to a reference probability distribution P.

Sensitive feature	Hazard w/out conditioning	Cumulative hazard
Gender	0	0
Age	0.0202	0.0512
Marital status	0	0

Table 3
Target: ground truth

5. Bias in Model Output

To determine the extent to which the model's predictions amplify bias relative to the actual distribution of the sensitive groups of interest, we run the bias detection module of the BRIO tool on both the model's predictions and the ground truth and compare the former result against the latter. For the present example, we run experiments for two different sensitive features: *age* and *gender*. Our predictive model is obtained through a standard credit scorecard modelling process based on the methodology of OptBinning[6].

The bias detection module returns – along with a list of fairness violations alerts – a hazard value for each dataset and each sensitive feature that indicates how dangerous the violation is. This is the value that we will employ to compute the bias amplification between the ground truth and the result of passing it through the model.

The bias detection tests have been conducted with the following arguments.

- Target feature: ground truth labels (Table 3), and model predictions (Table 4).
- Sensitive features:
  - gender (male and female);
  - age (four age groups: 0-24, 25-39, 40-59, over 60);
  - marital status (single and married/separated/divorced/widowed).
- Divergence used to measure the discrepancy between frequencies: Jensen–Shannon [22].
- Threshold: automatically computed with sensitivity set to *high*.
- Function to aggregate divergences between pairs of sensitive classes (only for the test on age): arithmetical mean.
- Conditioning variables for double-checks on subclasses:
  - Attribute 3, "credit history" (no credits taken / all credits paid back duly, all credits at this bank paid back duly, existing credits paid back duly till now, delay in paying off in the past, critical account / other credits existing but not at this bank);
  - Attribute 10, "other debtors / guarantors" (none, co-applicant, guarantor).

Our results are collected in Table 3 and in Table 4. We display hazard values, in particular cumulative hazard is the sum of all estimated hazard values including without conditioning. The aim is to assess whether looking at subclasses increases the difference in behaviour or not. Recall that a full list of violations with their respective estimated hazards is available on Git[7].

First, it should be noted that the hazard measure for gender and marital status is quite similar – it is not actually *identical*, see the full report in the Git `repo`, since it differs from the 15^{th} decimal point on. This is partly explained by the fact that *all* women in our dataset are married.[8]

Now we move on to the significance of our analysis. The most notable difference in comparing our analysis on the ground truth with that on the model predictions occurs with gender: while the ground truth appears to be non-biased with respect to gender,[9] this is not the case for the model which

[6] http://gnpalencia.org/optbinning/scorecard.html
[7] https://github.com/DLBD-Department/BRIO_x_Alkemy/tree/main/BEWARE_2024
[8] This is perhaps another argument in favour of the thesis of [18].
[9] Notice that having an hazard value be equal to 0 does *not* mean that the ground truth is perfectly balanced, but that it is unbalanced within acceptable limits according to the selected threshold. More on this is detailed in [2, §5].

Sensitive feature	Hazard w/out conditioning	Cumulative hazard
Gender	0.0695	0.1673
Age	0.0613	0.1480
Marital status	0.0695	0.1673

Table 4
Target: prediction of the model

Sensitive feature	Amplification
Gender	`undef.`
Age	189.0625%
Marital status	`undef.`

Table 5
Bias amplification percentage

introduces "new" bias even while conditioning with possible predictors of good credit behaviour. While less strinkingly so, the model seems to introduce new bias with respect to age as well. In order to compute *by how much* this is the case, we represent bias amplification by a percentage change in hazard:

$$\texttt{amplification} = \frac{(\texttt{predictions\_hazard} - \texttt{ground\_truth\_hazard}) \cdot 100}{\texttt{ground\_truth\_hazard}}$$

whose results are collected in Table 5.

Of course, as for the hazard values related to gender and marital status, we would need to divide 0.1673 by 0, so we cannot compute an actual percentage. When this is the case, in general, we have a situation in which the model introduces a bias which is not at all present in the ground truth. While we cannot compute a percentage value, we can still evaluate the seriousness of the situation by considering the hazard values computed by the BRIO tool relatively to the predictions of the model. For instance, as far as gender and marital status are concerned, the bias introduced is comparable: the tool yielded for both sensitive features an hazard value (without conditioning) of 0.0695 and a cumulative hazard value of 0.1673. As already mentioned, this is probably due to the extreme correlation between the gender value associated to women and the marital status value associated to married people. We leave as future work the task of defining a mathematical *measure* of bias amplification that enables us to homogeneously treat all these cases.

6. Conclusion

Bias amplification describes the process by which bias in the data propagates throughout the ML pipeline, ultimately leading to unfair outcomes. By building on this notion, in this paper, we introduced the concept of the Bias Amplification Chain (BAC), illustrated three links thereof, and used the BRIO framework to quantify each of these.

An interesting avenue for future research involves exploring more complex configurations of the Bias Amplification Chain itself. In particular, refining our BAC with additional links to the chain could provide a more fine-grained and realistic analysis of bias amplification. For instance, we want to investigate how bias is amplified during the phase of data curation, when in order to optimize the training certain variables are discarded (feature selection), and synthetic features are automatically crafted by the system from the initial predictors (feature engineering).

More generally, the notion of BAC provides a model to quantitatively reason about where and when unfairness is produced along the ML life cycle. Breaking down the phenomenon of ML unfairness in different links of a chain representing the ML pipeline can potentially yield valuable insights on the effectiveness of certain mitigation methods (e.g., increasing the data quality, increasing the sample

size, learning fair representations, etc.). Finally, a major open question remains how to meaningfully aggregate the preliminary results here obtained in order to gain a holistic insight on the phenomenon of ML unfairness.

References

[1] G. Coraglia, F. A. D'Asaro, F. A. Genco, D. Giannuzzi, D. Posillipo, G. Primiero, C. Quaggio, Brioxalkemy: a bias detecting tool, in: BEWARE@AI*IA, 2023. URL: https://api.semanticscholar.org/CorpusID:267200510.

[2] G. Coraglia, F. A. Genco, P. Piantadosi, E. Bagli, P. Giuffrida, D. Posillipo, G. Primiero, Evaluating ai fairness in credit scoring with the brio tool, 2024. arXiv:2406.03292.

[3] M. Ziosi, D. Watson, L. Floridi, A genealogical approach to algorithmic bias, Minds and Machines 34 (2024) 1–17. doi:10.1007/s11023-024-09672-2.

[4] L. Floridi, The method of levels of abstraction, Minds Mach. 18 (2008) 303–329. URL: https://doi.org/10.1007/s11023-008-9113-7. doi:10.1007/s11023-008-9113-7.

[5] G. Primiero, Information in the philosophy of computer science, Routledge, 2016. URL: https://www.routledgehandbooks.com/doi/10.4324/9781315757544.ch10. doi:10.4324/9781315757544.ch10.

[6] A. G. Buda, G. Primiero, A pragmatic theory of computational artefacts, Minds Mach. 34 (2024) 139–170. URL: https://doi.org/10.1007/s11023-023-09650-0. doi:10.1007/S11023-023-09650-0.

[7] J. Foulds, R. Islam, K. N. Keya, S. Pan, An intersectional definition of fairness, 2019. URL: https://arxiv.org/abs/1807.08362. arXiv:1807.08362.

[8] Y. Hirota, Y. Nakashima, N. Garcia, Quantifying societal bias amplification in image captioning, in: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 13440–13449. doi:10.1109/CVPR52688.2022.01309.

[9] M. Hall, L. van der Maaten, L. Gustafson, M. Jones, A. Adcock, A systematic study of bias amplification, 2022. URL: https://arxiv.org/abs/2201.11706. arXiv:2201.11706.

[10] P. Seshadri, S. Singh, Y. Elazar, The bias amplification paradox in text-to-image generation, 2023. URL: https://arxiv.org/abs/2308.00755. arXiv:2308.00755.

[11] H. Suresh, J. V. Guttag, A framework for understanding sources of harm throughout the machine learning life cycle, Equity and Access in Algorithms, Mechanisms, and Optimization (2019).

[12] A. Wang, O. Russakovsky, Directional bias amplification, in: M. Meila, T. Zhang (Eds.), Proceedings of the 38th International Conference on Machine Learning, volume 139 of *Proceedings of Machine Learning Research*, PMLR, 2021, pp. 10882–10893. URL: https://proceedings.mlr.press/v139/wang21t.html.

[13] F. A. D'Asaro, G. Primiero, Probabilistic typed natural deduction for trustworthy computations, in: TRUST@AAMAS, 2021. URL: https://api.semanticscholar.org/CorpusID:245423393.

[14] F. A. Genco, G. Primiero, A typed lambda-calculus for establishing trust in probabilistic programs, 2023. arXiv:2302.00958.

[15] F. A. D'Asaro, F. Genco, G. Primiero, Checking trustworthiness of probabilistic computations in a typed natural deduction system, 2024. arXiv:2206.12934.

[16] E. Kubyshkina, G. Primiero, A possible worlds semantics for trustworthy non-deterministic computations, International Journal of Approximate Reasoning 172 (2024) 109212. URL: https://www.sciencedirect.com/science/article/pii/S0888613X24000999. doi:https://doi.org/10.1016/j.ijar.2024.109212.

[17] H. Hofmann, Statlog (German Credit Data), UCI Machine Learning Repository DOI: https://doi.org/10.24432/C5NC77 (1994).

[18] U. Grömping, South German Credit Data: Correcting a Widely Used Data Set., Department II, Beuth University of Applied Sciences, Berlin, 2019. URL: https://www1.beuth-hochschule.de/FB_II/reports/Report-2019-004.pdf.

[19] J. Simson, A. Fabris, C. Kern, Lazy data practices harm fairness research, in: Proceedings of the

2024 ACM Conference on Fairness, Accountability, and Transparency, FAccT '24, Association for Computing Machinery, New York, NY, USA, 2024, p. 642–659. URL: https://doi.org/10.1145/3630106.3658931. doi:10.1145/3630106.3658931.

[20] S. Kullback, R. A. Leibler, On Information and Sufficiency, The Annals of Mathematical Statistics 22 (1951) 79 – 86. URL: https://doi.org/10.1214/aoms/1177729694. doi:10.1214/aoms/1177729694.

[21] J. Liu, Z. Shen, Y. He, X. Zhang, R. Xu, H. Yu, P. Cui, Towards out-of-distribution generalization: A survey, 2023. URL: https://arxiv.org/abs/2108.13624. arXiv:2108.13624.

[22] J. Lin, Divergence measures based on the shannon entropy, IEEE Transactions on Information Theory 37 (1991) 145–151. doi:10.1109/18.61115.