Challenges of Enforcing Regulations in Artificial Intelligence Act — Analyzing Quantity Requirement in Data and Data Governance

Farhana Ferdousi Liza

School of Computing Sciences, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK.

Abstract

To make Artificial Intelligence (AI) systems and services accountable and regulated in the European Union market, in April 2021, the European Union Parliament published a proposal 'Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)', widely known as Artificial Intelligence Act (AI Act). Since then, many concerns have been raised in terms of compliance and whether the regulations are enforceable. However, to the best of our knowledge, none of them provided an explicit technical analysis of the challenges in enforcing the regulation. Among 85 Articles in the AI Act, we emphasize on the Article 10, the central regulatory requirement for data and data governance. In this paper, we have analyzed a specific requirement, the data quantity, to show the challenges of enforcing this requirement in a principled way. In our analysis, we have used deep learning modeling and machine learning generalization theory.

Keywords

Artificial Intelligence Act, Future Technologies, Generalization Theory, Deep Learning Modeling.

1. Introduction

In April 2021, the European Commission proposed regulations on Artificial Intelligence (AI) and published the Artificial Intelligence Act (AI Act) [1]. This landmark regulatory proposal aims to create a clear regulatory environment for AI providers and users to protect AI users from the harmful effects of AI deployments. The AI Act aims at facilitating the development of a single market for lawful and safe AI. Once it comes into force, this Act will impact internationally beyond the European Economic Area [2, 3], therefore a critical assessment is important for sustainable and fair regulation of AI systems.

AI Act Article 3(1) Annex I defines Artificial Intelligence (AI) which covers a wide range of computational methods designed to perform certain tasks, such as generating the content, assisting in decision-making about people's social security benefits, predicting an individual's risk of committing fraud or defaulting on a loan. The AI systems and approaches include machine learning approaches, including supervised, unsupervised, and reinforcement learning, using a

1st International Workshop on Imagining the AI Landscape After the AI Act (In conjunction with The first International Conference on Hybrid Human-Artificial Intelligence), Vrije Universiteit Amsterdam, Amsterdam, Netherlands, June 13, 2022

F.Liza@uea.ac.uk (F. F. Liza)

0000-0003-4854-5619 (F. F. Liza)

wide variety of methods including deep learning. AI is the capability of a computer system to mimic human cognitive functions such as learning and problem-solving. One way to train a computer system to mimic human learning and reasoning is to use a neural network, which is a series of algorithms taking inspiration from brains to model complex data representation. The neural network helps the computer system achieve AI through deep learning. In this paper, we have used this particular modeling for our analysis, however, the analysis applies to any AI system that is capable of learning and reasoning mimicking human intelligence.

AI systems have the potential to improve livelihood and overall economic and societal welfare. AI-powered systems may help to respond to key global challenges, such as emergency management and the novel coronavirus pandemic. For example, storm Eunice was predicted five days earlier than the storm hit the UK using numerical simulations techniques which provided the intervention opportunities to reduce the casualty. With AI techniques if we could predict the event earlier, it would be possible to prevent casualties. This makes Schultz et al. [4] wonder 'Can deep learning beat numerical weather prediction?'. Moreover, AI systems can contribute to better healthcare services [5], improve the supply chain [6], help in fast decision making [7] and service deployment [8, 9]. AI systems can also benefit the public sector in several ways, for example, by efficient resource allocation to personalized public services tailored to individual circumstances or early planning for resource allocation based on historical data.

While AI systems have unprecedented potential for society, they also come with substantial risks, raising a variety of legal and ethical challenges. AI systems have the potential to unpredictably harm people's life, health, and autonomy. They can also affect fundamental values of European society and around the world, leading to breaches of fundamental rights of people and assembly, non-discrimination, or the right to an effective and a fair trial, as well as consumer protection [10, 11, 12].

To mitigate the risks, the commission has proposed a risk-based regulatory approach. High-risk AI systems that use techniques involving the training of models based on training, validation, and testing data sets are required to meet the required criteria referred to in the AI Act Article 10. We particularly appreciate that the AI Act is imposing regulations on data and data governance to regulate AI, however, a closer analysis shows that the proposed AI Act may need improvement in many areas. This paper is a basis for a preliminary assessment of the AI Act for further discussion on data and data governance. In this paper, we will mainly concentrate on deep learning modeling (i.e., deep neural network models) as a representative AI system that is inspired by brain computation or human intelligence and the generalization theory of machine learning. Without loss of generality, the discussion applies to every AI model specifically those are capable of modeling complex nonlinear data representation.

2. Related works

Since the publication of the AI Act proposal, many critical assessments of the regulation and impact of the proposal are published [13, 14]. Fiazza [14] raised the concern of enforcing the AI Act Article 10 by stating that:

> To touch briefly on a topic worthy of a longer discussion, the Proposal mandates in Article 10 that all datasets used "shall be relevant, representative, free of errors

> and complete," "exhibiting the appropriate statistical properties," and that they be evaluated for possible biases. In so doing, the Proposal represents in the wider context of AI a problem already visible in the General Data Protection Regulation [5]: it is not presently known how to fulfill the bill. The problem is especially thorny in medical robotics in connection to accounting for rare events, anatomical variants and rare pathologies in the datasets.

The sentence 'it is not presently known how to fulfill the bill' defines the main challenge of enforcing the AI Act 2021. We have analyzed the data quantity requirement to elaborate further to show that it is not possible to precisely quantify the required quantity of data for training an AI system. Similarly, Ebers et al. [13] recognize the impossibility of enforcing the regulation in practice and raise the concern that the regulation in the current state might hamper innovation. They mainly considered the issues including data quality, bias, and the free-of-error requirement of the dataset. Human Right Watch (HRW) has done a critical evaluation of the AI Act and concluded that the regulation can endanger the Social Safety Net[1]. In our knowledge, there is no work has been done to show that the compliance and enforcement of the Article 10 are possible in a principled way.

We share the community's concerns. To be able to understand the challenges and to provide sustainable requirements, it is important to analyze the requirement from a technical perspective. Although existing works raise valid concerns, none of them provided an explicit technical analysis of the challenges in enforcing the regulation. Specifically, none of the studies have analyzed the data quantity requirement in AI Act Article 10. In this paper, we provide the technical analysis emphasizing the challenges of enforcing the regulation of the 'data quantity' requirement.

3. AI Act Article 10 and GDPR

In AI Act, the EU Commission proposes a risk-based categorization of AI systems with four levels of risk (e.g., unacceptable risk, high risk, low risk) and related regulatory obligations and restrictions. AI Act Article 10 outlines the central regulatory requirement for data and data governance. These requirements are for high-risk AI systems which are defined both by general characteristics (AI Act Article 6) and specifically targeted applications (AI Act Annex III). AI Act Article 6 defines 'high-risk' AI systems where the AI system is intended to be used as a safety component of a product, or itself a product, and this product is subject to an existing third-party conformity assessment. AI Act Annex III defines the targeted applications including Biometric identification, Law enforcement, and Employment. In addition, the commission has the power to directly designate an AI system as high-risk by adding it to AI Act Annex III (AI Act Article 84). Although analysis of the risk categorization is outside the scope of this paper, the proposed risk categorization seems like an unrealistic approach. If an AI system is developed outside the listed application area, later the AI system poses an unforeseen high-risk or unacceptable risk, the system may be penalized with high fees or the commission can not keep them accountable as they were not listed in Annex III at the time of AI system development.

[1] https://www.hrw.org/news/2021/11/10/how-eus-flawed-artificial-intelligence-regulation-endangers-social-safety-net$_f tn4$

If a high-risk AI system is trained with data; training, validation and test datasets must meet the quality criteria specified in paragraphs 2 to 5 in AI Act Article 10. Paragraph 2 defines the requirements such as appropriate data governance and data management procedures should apply, including design choices, data collection, relevant data preparation processes and prior assessment of the availability, quantity, and suitability of the required datasets, examination of bias, and identification of possible data gaps. Similarly, paragraph 3 outlines requirements such as the datasets must be relevant, representative, accurate and complete, and paragraph 4 requires that to be the extent necessary for the intended purpose, correspond to the characteristics or elements specific to the particular geographic, behavioral, or functional context in which the high-risk AI system is intended to be used. Although the Commission has outlined those requirements, the Commission does not define these characteristics or its understanding of 'data governance' at the level that can be used for designing relevant compliance and enforcement techniques. It, therefore, remains subjective where objective requirements are to be enforced concerning compliance.

This results in uncertainty in technical implementation on compliance, enforcement, and legal complexity. Like the General Data Protection Regulation (GDPR), the AI Act standardizes requirements for the handling of data. If an AI system uses personal data to develop high-risk AI systems, providers will be obliged to comply with the data handling requirements of the AI Act and the personal data processing requirements of the GDPR. The question arises, in the case of violating any requirements would they be double penalized? This question arises as the EU presents itself as a single interlocutor, not only in the management of personal data (GDPR) but now also for AI systems and services. Thus clarifications on how these laws will be applied in case of violating any requirements would encourage the sustainable development of AI systems and services within the EU's long-term vision with AI. Otherwise, from a techno-legal perspective, with the fear of penalization in the case of violation of these two related laws, specifically with the presence of non-explicit vague required criteria in the AI Act, the development of AI will be hindered by fear-driven discouragement of the potential investment.

4. Challenges to Enforce AI Act Article 10 Data Quantity Requirement

In this section, we will mainly analyze one specific requirement — data quantity — and two other related requirements — free of errors and complete data — in the AI Act Article 10 to emphasize the challenges to comply with the proposed regulation from the technical perspective. AI Act Article 10(2(e)) has the following prior assessment requirement:

> a prior assessment of the availability, **quantity** and suitability of the data sets that are needed;

To be able to pre-assess the data quantity for an AI system, we would need a complete understanding of a real process that we are aiming to model and a complete understanding of how a model makes a decision. When our understanding is limited for a real process, the modeling of the process would lead to a larger uncertainty. This is because we do not know

or understand the process and we can come up with our take on the process which might not be representative of the true process. For example, Saltelli and Funtowicz [15] have described climate change modeling where a better understanding of the process led to a larger uncertainty. In that scenario, the Intergovernmental Panel on Climate Change produced larger prediction uncertainty ranges, as opposed to smaller ones, as more and more processes, scenarios, and models are incorporated and cascading uncertainties made the lack of understanding effect felt in the final estimates. The fundamental problem is that the current state of statistics and nonconvex optimization theory can not provide probability distributions that represent a real-world complex data distribution and can not explain why an AI system (i.e., deep learning models) performs well in empirical settings for real-world problems [16]. With the current state of the computational learning theory, statistics, and mathematics, it is thus not possible to precisely quantify a dataset needed for AI systems including a deep learning model. To elaborate on that, we will use the generalization theory of machine learning. The quantity of the dataset is related to a concept called model generalization which is not yet understood for non-linear machine learning models including deep learning models.

To briefly describe generalization, when we train a model, we don't just want to train a model to perform well on the training data. We want it to generalize to data that the model hasn't seen in the training data, this unseen data is called held-out test data. For example, for human learning, if a human is told that Lion is dangerous (training), they can deductively reason that a Tiger might be dangerous even if they were never warned about the Tiger (held-out testing). To measure the model's generalization performance, we measure the model's performance on a held-out test dataset. If a model works well on the training set but fails to generalize on a held-out dataset, we conclude that the model has been overfitted. Overfitting is problematic for generalized learning. Improving generalization (or preventing overfitting) in AI models, especially for neural nets is still somewhat based on trial and error. The current approach is depending on the data distribution, with a change in data distribution (i.e., an increase in the dataset due to the availability of new data in the real scenario), the whole trial and error process would need to be repeated. Of course, there are a few simple strategies that can help reduce overfitting (e.g., Reducing model capacity, Early stopping, Regularization and weight decay, Ensembles, Data augmentation, and Stochastic regularization), but none of the techniques gives a guarantee of generalized performance, that means that we do not know when these models will fail to generalize. Therefore, we don't have learning guarantees, error bounds, science, and technology that shed light on the explicit evaluation of generalization. That means, it is not possible to provide a required quantity of a dataset for an AI system in real-world scenarios. The more realistic AI system would need to deal with out-of-distribution learning and generalization to arbitrary out-of-distribution is impossible [17] with the current state of the computational learning theory, statistics, and mathematics. AI Act does not specify the expectation on the generalization performance, thus it imposes difficulty in designing a prior assessment to be reasonably confident that the end-users of the high-risk system will not incur high-risk harm because of a data quantity issue.

In conventional learning theory, there is a relationship between the number of parameters and overfitting. When the number of parameters is larger than the number of training examples the model is called an over-parameterized model. From the optimization and computational learning theory's perspective, overparameterization should lead to overfitting. However, it

is not clear why overparameterization does not lead to overfitting in deep learning models [18]. The community has concluded that to understand the deep learning framework we might need to rethink conventional generalization theory [19] and is still working on the theoretical explanation of generalization [16]. For example, Simon et al. [20] concluded that 'The generalization of deep neural networks remains mysterious, with a full quantitative theory still out of reach'. Essentially Simon et al. [20] tried to formulate a principle theory of generalization and concluded that there is more work to be done to arrive at a general theory of deep learning generalization. Despite the lack of theoretical understanding, one of the benefits of using a deep neural network is that with domain expertise and heuristic strategies [21] deep learning modeling can perform empirically well with a high dimensional large dataset, whereas other models would become intractable. Therefore, it is common to perform studies on how algorithm performance scales with dataset size [22]. This means that algorithms performance improves in deep learning modeling with an increase in sample size whereas conventional models will even fail to model such a large dimensional dataset. However, domain expertise and heuristic strategies that make deep learning models perform superiorly compared to conventional models, do not give us a reasonable pre-assessment criteria on a sample size (i.e., data quantity) that could be used to make deep learning models perform with a acceptable generalization performance for a techno-legal framework.

Generally, for a machine learning approach, no theory can precisely quantify data for any type of real-world AI system modeling. Usually, the data size (i.e., sample size) is chosen based on different heuristics [23, 24] including a factor of the number of classes, a factor of the number of input features, a factor of the number of model parameters. However, these are all heuristics and therefore the model correctness and generalization rely on error analysis and different statistical significance tests.

When all these heuristics are followed, there is still a possibility of mistakes (e.g., with adversarial attacks) by well-trained machine learning models — conventional models [25] and deep learning models [26]. Whereas deep learning is remarkable in providing solutions to problems that were not possible using conventional machine learning techniques, adversarial attacks and the defenses against adversarial attacks are still under investigation. Do adversarial attacks related to data quantity used in the training? We do not know the answer. Therefore, with the presence of adversarial attacks, these heuristic approaches are not sufficiently reliable for providing good generalizations guarantee.

AI Act regulation, Article 10(2(e)) requires the following criteria for a dataset:

> Training, validation and testing data sets shall be relevant, representative, free of errors and complete.

As we have discussed that estimating the exact quantity (i.e., sample size) of the dataset for AI systems is impossible, a direct consequence is that it is not possible to evaluate the completeness of the data set. Moreover, with an unsupervised learning paradigm, it is difficult to evaluate the 'free of errors' criterion in the dataset as there is no explicit human labeling of the dataset and it is not clear what it means by 'error' in this context. One of the scenarios where the deep learning paradigm excels is unsupervised feature learning. For example, deep learning models are needed for high-dimensional domains (e.g., text, image, video) where manual feature engineering is

difficult [27]. In this context, defining an error-free dataset would be difficult given the size of the dataset used in the unsupervised learning framework. Overall, it is necessary to define quantifiable errors to design meaningful compliance and make regulation enforcement fair for an AI system. Such unsupervised learning features work well with the downstream application, however, there is still no theory to explain why automated feature learning using deep learning techniques works mostly well with downstream applications. At the same time, there is no theory on why adversarial examples [28, 29] can cause AI systems to make mistakes. Does this relate to the data-related error or the model-related error?

Machine learning has a close relation to statistics. "All models are wrong, but some are useful" is a famous quote often attributed to the British statistician George E. P. Box. Statisticians and researchers try to develop models aiming to predict the behavior of a certain process, for instance, the selling trend of a product or demand for a taxi in a city [30]. Thus, the idea of this quote is that every single model will be wrong, meaning that it will never represent the exact real process. This is true for most machine learning and AI systems, as no model can represent the exact real process.

Moreover, we don't have an understanding of the representation learned by an AI system (i.e, deep learning model). For example, adversarial examples are not distinguishable from the original examples in human eyes. It is possible that AI systems have a different interpretation of the learned representation or features than humans, and there are scopes of theoretical and empirical research to understand the deep learning models. It is difficult to enforce a regulation on completeness and free of errors of the dataset until we have a more in-depth understanding of the AI systems including deep learning models.

In attempts to regulate AI systems for high-risk application, the proposed data and data governance requirement criteria in AI Act would pose problematic circumstances from techno-legal perspective, especially when statistical, mathematical, and computational learning theory is not advanced enough to give evaluation metrics for those criteria. For example, anyone can claim that the used dataset for a AI system is in the right quantity for modeling and there is no principled evaluation framework is available to validate this claim. Without a principled quantitative or qualitative evaluation framework, these criteria are discursively meaningless [31], easy to manipulate, and not fit for purpose. When Ursula von der Leyen had pledged that, within 100 days of her election as President of the European Commission, she would have proposed new legislation on AI[2], Floridi [32] remarked that it was a reasonable strategy but an unrealistic timeline. Philosophically, the initiative was a starting point to ensure that the development of AI in the EU is ethically sound, legally acceptable, socially equitable, and environmentally sustainable. The underpinning vision is to have AI systems and services that seek to support the economy, society, and the environment. Fulfilling such a vision is not a simple one and, it will take time and effort to reach a final AI Act that can come close to fulfilling the vision. Yet, the vision, like von der Leyen's pledge, remains reasonable because the EU is set to deliver such a challenging philosophical framework. More clarity on expectations for the assessment of data and data governance from the legal perspective with a corresponding technical evaluation framework will make this unified, post-Westphalian approach [33] to have several positive effects. AI companies and vendors will have to deal with one EU regulation for

[2]https://ec.europa.eu/commission/presscorner/detail/en/ip_20_403

their AI systems and services, not with the individual Member States when they will have to prove that they comply with the new legislation.

5. Conclusion

In this paper, we have concentrated on a requirement in Article 10 which is related to data and data governance. There are other issues[3] such as application-specific risk categorization that need further consideration. The requirements of the AI Act would have an impact on society, future technologies, and innovation. For example, a global law firm Taylor Wessing has reported that the double regulatory compliance (i.e., GDPR and AI Act) on data governance might lead to financial consequences, the new AI Act imposes higher fines than the GDPR for violating the requirements under Article 10, namely up to EUR 30,000,000 or - in the case of companies - up to 6% of the total annual worldwide turnover of the previous financial year, whichever is higher[4]. This might limit innovation which would have economic consequence. To conclude, based on our analysis, we call for a technically implementable AI Act regulations which have long-term international, social, and economic implications. Particularly we call for more clarity from lawmakers on what are their expectations of the data and data governance.

Acknowledgement

We thank the three anonymous reviewers whose comments and suggestions helped improve and clarify this manuscript.

References

[1] E. Commission, Proposal for a Regulation Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) ; COM (2021) 206 final; European Commission: Brussels, Belgium, 2021.

[2] D. Svantesson, The european union artificial intelligence act: Potential implications for australia, Alternative Law Journal (2021) 1037969X211052339.

[3] G. Greenleaf, The 'brussels effect' of the EU's 'AI act' on data privacy outside europe (june 7, 2021), Privacy Laws Business International Report 1, 3-7, UNSW Law Research 171 (2021) 3–7. URL: https://ssrn.com/abstract=3898904.

[4] M. Schultz, C. Betancourt, B. Gong, F. Kleinert, M. Langguth, L. Leufen, A. Mozaffari, S. Stadtler, Can deep learning beat numerical weather prediction?, Philosophical Transactions of the Royal Society A 379 (2021) 20200097.

[5] A. Panesar, Machine learning and AI for healthcare, Springer, 2019.

[6] R. Toorajipour, V. Sohrabpour, A. Nazarpour, P. Oghazi, M. Fischl, Artificial intelligence in supply chain management: A systematic literature review, Journal of Business Research 122 (2021) 502–517.

[3] https://www.bbc.co.uk/news/technology-56745730

[4] https://www.taylorwessing.com/en/interface/2021/ai-act/fines-under-the-ai-act---a-bottomless-pit

[7] R. Vaishya, M. Javaid, I. H. Khan, A. Haleem, Artificial intelligence (ai) applications for covid-19 pandemic, Diabetes & Metabolic Syndrome: Clinical Research & Reviews 14 (2020) 337–339.
[8] H. J. Wilson, P. R. Daugherty, Collaborative intelligence: Humans and ai are joining forces, Harvard Business Review 96 (2018) 114–123.
[9] J. M. Corchado, P. Chamoso, G. Hernández, A. S. R. Gutierrez, A. R. Camacho, A. González-Briones, F. Pinto-Santos, E. Goyenechea, D. Garcia-Retuerta, M. Alonso-Miguel, B. B. Hernandez, D. V. Villaverde, M. Sanchez-Verdejo, P. Plaza-Martínez, M. López-Pérez, S. Manzano-García, R. S. Alonso, R. Casado-Vara, J. P. Tejedor, F. d. l. Prieta, S. Rodríguez-González, J. Parra-Domínguez, M. S. Mohamad, S. Trabelsi, E. Díaz-Plaza, J. A. Garcia-Coria, T. Yigitcanlar, P. Novais, S. Omatu, Deepint.net: A rapid deployment platform for smart territories, Sensors 21 (2021). URL: https://www.mdpi.com/1424-8220/21/1/236. doi:10.3390/s21010236.
[10] M. Ebers, S. Navas, Algorithms and law, Cambridge University Press, 2020.
[11] S. Gerke, T. Minssen, G. Cohen, Ethical and legal challenges of artificial intelligence-driven healthcare, in: Artificial intelligence in healthcare, Elsevier, 2020, pp. 295–336.
[12] T. Tzimas, Legal and Ethical Challenges of Artificial Intelligence from an International Law Perspective, volume 46, Springer Nature, 2021.
[13] M. Ebers, V. R. Hoch, F. Rosenkranz, H. Ruschemeier, B. Steinrötter, The european commission's proposal for an artificial intelligence act—a critical assessment by members of the robotics and ai law society (rails), J 4 (2021) 589–603.
[14] M.-C. Fiazza, The eu proposal for regulating ai: Foreseeable impact on medical robotics, in: 2021 20th International Conference on Advanced Robotics (ICAR), 2021, pp. 222–227. doi:10.1109/ICAR53236.2021.9659429.
[15] A. Saltelli, S. Funtowicz, When all models are wrong, Issues in Science and Technology 30 (2014) 79–85.
[16] T. J. Sejnowski, The unreasonable effectiveness of deep learning in artificial intelligence, Proceedings of the National Academy of Sciences 117 (2020) 30033–30038. URL: https://www.pnas.org/doi/abs/10.1073/pnas.1907373117. doi:10.1073/pnas.1907373117. arXiv:https://www.pnas.org/doi/pdf/10.1073/pnas.1907373117.
[17] H. Ye, C. Xie, T. Cai, R. Li, Z. Li, L. Wang, Towards a theoretical framework of out-of-distribution generalization, Advances in Neural Information Processing Systems 34 (2021).
[18] Z. Allen-Zhu, Y. Li, Y. Liang, Learning and generalization in overparameterized neural networks, going beyond two layers, in: H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, R. Garnett (Eds.), Advances in Neural Information Processing Systems, volume 32, Curran Associates, Inc., 2019. URL: https://proceedings.neurips.cc/paper/2019/file/62dad6e273d32235ae02b7d321578ee8-Paper.pdf.
[19] C. Zhang, S. Bengio, M. Hardt, B. Recht, O. Vinyals, Understanding deep learning requires rethinking generalization, International Conference on Learning Representations (2017).
[20] J. B. Simon, M. Dickens, M. R. DeWeese, Neural tangent kernel eigenvalues accurately predict generalization, CoRR abs/2110.03922 (2021). URL: https://arxiv.org/abs/2110.03922. arXiv:2110.03922.
[21] F. F. Liza, M. Grzes, Improving language modelling with noise contrastive estimation, in:

Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence, AAAI'18/IAAI'18/EAAI'18, AAAI Press, 2018.

[22] F. F. Liza, M. Grzes, Relating RNN layers with the spectral WFA ranks in sequence modelling, in: Proceedings of the Workshop on Deep Learning and Formal Languages: Building Bridges, Association for Computational Linguistics, Florence, 2019, pp. 24–33. URL: https://www.aclweb.org/anthology/W19-3903. doi:10.18653/v1/W19-3903.

[23] S. J. Raudys, A. K. Jain, et al., Small sample size effects in statistical pattern recognition: Recommendations for practitioners, IEEE Transactions on pattern analysis and machine intelligence 13 (1991) 252–264.

[24] R. Krishnaiah, L. Kanal, Dimensionality and sample size considerations in pattern recognition practice. handbook of statistics, 1982.

[25] H. Xiao, H. Xiao, C. Eckert, Adversarial label flips attack on support vector machines, in: Proceedings of the 20th European Conference on Artificial Intelligence, ECAI'12, IOS Press, NLD, 2012, p. 870–875.

[26] A. Chakraborty, M. Alam, V. Dey, A. Chattopadhyay, D. Mukhopadhyay, A survey on adversarial attacks and defences, CAAI Transactions on Intelligence Technology 6 (2021) 25–45.

[27] A. Halevy, P. Norvig, F. Pereira, The unreasonable effectiveness of data, IEEE intelligent systems 24 (2009) 8–12.

[28] I. J. Goodfellow, J. Shlens, C. Szegedy, Explaining and harnessing adversarial examples, 2014. URL: https://arxiv.org/abs/1412.6572. doi:10.48550/ARXIV.1412.6572.

[29] S. Garg, G. Ramakrishnan, Bae: Bert-based adversarial examples for text classification, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020, pp. 6174–6181.

[30] F. Rodrigues, I. Markou, F. C. Pereira, Combining time-series and textual data for taxi demand prediction in event areas: A deep learning approach, Information Fusion 49 (2019) 120–129.

[31] M. R. Olsson, Michel foucault: discourse, power/knowledge, and the battle for truth, Leckie, Gloria J (2010) 63–74.

[32] L. Floridi, The european legislation on ai: A brief analysis of its philosophical approach, Philosophy & Technology 34 (2021) 215–222.

[33] A. Linklater, Citizenship and sovereignty in the post-westphalian state, European Journal of International Relations 2 (1996) 77–103.