AI Competitions as Infrastructures: Examining Power Relations on Kaggle and Grand Challenge in AI-driven Medical Imaging

Luitse, D.; Blanke, T.; Poell, T.

Publication date
2022
Document Version
Final published version
Published in
Selected Papers of #AoIR2022

Selected Papers of #AoIR2022:
The 23rd Annual Conference of the
Association of Internet Researchers
Dublin, Ireland / 2-5 Nov 2022

AI COMPETITIONS AS INFRASTRUCTURES: EXAMINING POWER RELATIONS ON KAGGLE AND GRAND CHALLENGE IN AI-DRIVEN MEDICAL IMAGING

Dieuwertje Luitse
University of Amsterdam

Tobias Blanke
University of Amsterdam

Thomas Poell
University of Amsterdam

Artificial Intelligence (AI) is quickly being taken-up across scientific disciplines, medical imaging is no exception. To stimulate development and facilitate the scientific evaluation of new approaches, AI-based research in medical imaging is increasingly organised in a competitive manner through machine-learning development platforms. These platform-based AI competitions are typically organised by consortia that issue dedicated research problems, focusing on specific tasks, such as breast cancer detection. Participating teams can compete to develop and evaluate their machine learning models using a given dataset and achieve the highest performance on a set of predefined tasks. Subsequently, teams are ranked on a platform-controlled central leaderboard by the calculated score of their (best) submission. This paper examines how such platform-based competitions shape power relations in medical image AI research around the globe.

Performance competitions follow a long-standing tradition in AI subfields like Computer Vision or Natural Language Processing. A prime example of how they play a vital role in AI research is the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) (Russakovsky et al., 2015). Associated with the ImageNet dataset (Deng et al., 2009), this competition ran between 2010 and 2017. In 2012, a team from the University of Toronto won the Challenge with a deep neural network-driven model that outperformed all competing teams by a significant margin using GPUs (Krizhevsky et al., 2012). This moment is often marked as the turning point in AI research, shifting the field towards the

Suggested Citation (APA): Luitse, D.M.R., Blanke T. & Poell, T. (2022, November). *AI Competitions as Infrastructures: Examining Power Relations on Kaggle and Grand Challenge in AI-driven Medical Imaging*. Paper (or panel) presented at AoIR 2022: The 23rd Annual Conference of the Association of Internet Researchers. Dublin, Ireland: AoIR. Retrieved from http://spir.aoir.org.

currently dominant deep learning paradigm, which is heavily data-driven and requires large amounts of computing power (Dotan and Milli, 2020).

While many competitions have been organised independently or in conjunction with prestigious AI conferences, the last decade has seen an increase of competitions organised on machine-learning development platforms such as Kaggle and Grand Challenge—two of the leading platforms in AI-based medical imaging. Such platforms provide the infrastructure to run competitions and the necessary tools for teams to participate. In addition, teams are commonly awarded prizes that can be over $1 million.

For medical image analysis, platform-hosted competitions constitute an important research infrastructure, steering global research and development in this dedicated AI subfield. These research infrastructures have been celebrated as “accelerators for medical imaging innovations” (Bulten et al. 2022: 154) and models that perform well in these competitions are more likely to be adapted for clinical practice. Furthermore, they further steer the individual careers of winning participants through the publication of competition research papers (Reinke et al., 2021).

Yet, little is known about how these platform-infrastructures shape the conditions for medical image-AI research, including model production and evaluation. It is, however, vital to gain such insights, as there are growing concerns over the power dynamics in (medical) AI research and the development of machine-learning applications that amplify existing inequalities towards historically disadvantaged communities (Benjamin 2019; Thomas, 2022). So far, critical work has primarily focused on power imbalances in the construction of single research components like medical imaging datasets (e.g., Oakden-Rayner, 2020; Rostamzadeh et al., 2022). But while such individual elements are significant, platform-hosted medical imaging competitions operate across the medical AI research pipeline. They structure the definition of research problems and the construction of datasets, as well as how models are produced and evaluated by introducing tasks and metrics. It is therefore critical to understand the organisation of these competitions in their entirety, and the position key actors—platforms, organising institutions, and competition participants—occupy within them.

This paper addresses this issue through a critical empirical case study of 120 medical imaging competitions on Kaggle and Grand Challenge between 2017 and 2022. Combining a critical AI and platform studies research approach, we investigate how competitions on Kaggle and Grand Challenge are organised—under what infrastructural conditions and by whom—and how this shapes processes of model production and evaluation in AI-driven medical imaging.

Studying Competitions on Kaggle and Grand Challenge

To understand how competitions on Kaggle and Grand Challenge shape production and evaluation processes in AI-based medical image research, we developed a digital methods approach that draws from various primary materials. The empirical analysis proceeded along three main lines of inquiry. First, we analysed Kaggle and Grand Challenge’s platform documentation to identify the ways in which both platforms regulate the organisation of competitions through their set up guidelines and their

several infrastructural components. Second, we draw from a set of collected competition descriptions to analyse the institutional backgrounds of competition organisers, the medical imaging datasets they provide and the ways in which these actors define research problems and set metrics of model evaluation. Lastly, we analysed information from leaderboards to map the distribution of competition participants in terms of geographical location and institutional background. This mapping provided insight into the geographical context submitted—well-performing—models have been produced, as well as the types of intellectual backgrounds they build upon.

Examining AI competitions as research infrastructures, we show that Kaggle and Grand Challenge as well as competition organisers shape power relations in medical AI research at the level of data and task design, model production and evaluation in several distinct ways. For example, the platforms exercise infrastructural power by pre-structuring task design, metrics for model evaluation and use of computer power. Organising institutions, on the other hand, can be seen to use their powerful position to direct research focus towards segmentation as they mainly design competitions towards this task. In addition, the datasets underlying these competitions are mainly constructed with data from patients in the United States, Western Europe and China, while dataset diversity is approached technically, and descriptions rarely include information on patient distribution in terms of gender and race.

Because competitions play such an important role in steering medical image AI research directions and influence the types of models that are implemented into clinical settings, these findings highlight the impact these powerful actors ultimately have on these issues. Therefore, we stress the importance of further considering platform-based competitions as key AI research infrastructures to locate the power differentials in the development of machine-learning models, not just in subfields of AI-based medicine, but also beyond.

References

Benjamin, R. (2019). *Race After Technology: Abolitionist Tools for the New Jim Code*. Cambridge: Polity Press

Bulten, W., Kartasalo, K., Chen, P.-H. C., et al. (2022). Artificial intelligence for diagnosis and Gleason grading of prostate cancer: The PANDA challenge. *Nature Medicine*, 28(1), 154–163.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A Large-Scale Hierarchical Image Database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255.

Dotan, R. & Milli, S. (2020). Value-laden disciplinary shifts in machine learning. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, Barcelona, Spain.

Krizhevsky A., Sutskever I. & Hinton G.E. (2012). Imagenet classification with deep convolutional neural networks. *25th International Conference on Neural Information Processing Systems—Volume 1* (NIPS'12), 1097–1105.

Oakden-Rayner, L. (2019). *Exploring large scale public medical image datasets* (arXiv:1907.12720). arXiv. http://arxiv.org/abs/1907.12720

Reinke, A., Tizabi, M. D., Eisenmann, M., & Maier-Hein, L. (2021). Common Pitfalls and Recommendations for Grand Challenges in Medical Artificial Intelligence. *European Urology Focus*, *7*(4), 710–712.

Rostamzadeh, N., Mincu, D., Roy, S., Smart, A., Wilcox, L., Pushkarna, M., Schrouff, J., Amironesei, R., Moorosi, N., & Heller, K. (2022). Healthsheet: Development of a Transparency Artifact for Health Datasets. *2022 ACM Conference on Fairness, Accountability, and Transparency*, 1943–1961.

Russakovsky, O., Deng, J., Su, H., et al. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115(3): 211–252.

Thomas, R. (2021, January). *Medicine's Machine Learning Problem*. Boston Review. Retrieved September 29, 2022, from https://bostonreview.net/articles/rachel-thomas-medicines-machine-learning-problem/