

The Case Against Registered Reports

Odd Erik Gundersen

■ *Registered reports have been proposed as a way to move from eye-catching and surprising results and toward methodologically sound practices and interesting research questions. However, none of the top-twenty artificial intelligence journals support registered reports, and no traces of registered reports can be found in the field of artificial intelligence. Is this because they do not provide value for the type of research that is conducted in the field of artificial intelligence?*

Registered reports have been touted as one of the solutions to the problems surrounding the reproducibility crisis. They promote good research practices and combat data dredging.¹ What is not to like? Still, although registered reports have been embraced by many different fields, it is not the case for artificial intelligence (AI). It is hard to find traces of registered reports in the field of AI. Is this because they are not needed in AI? Is there not a case for registered reports in AI research?

What Are Registered Reports?

Registered reports are research protocols that are peer-reviewed ahead of the experiments being conducted. The idea is that by provisionally committing to publish papers before the results are known, we move away from eye-catching and surprising results and toward methodologically sound practices and interesting research questions.

Although similar, a *registered report* is not the same as *preregistration*. The former is peer-reviewed, while the latter is not; instead, a research protocol is made available in a publicly accessible repository ahead of the experiments being conducted.

Rigorous experimental methods are encouraged by requiring that the research questions, hypotheses, experiments, and analytical approach be described ahead of time. Stating the research questions and hypotheses explicitly before conducting the experiments prevent researchers from changing these after having seen the results. Also, committing to the analysis method before seeing the results will prevent researchers from changing the analysis until an acceptable result is achieved.

The journal *Psychologic Science* even requires the reporting of effect sizes and confidence intervals ahead of time, while the *European Journal of Personality* encourages authors to include pilot data as part of the registered report.

Peer Review

Registered reports are typically peer-reviewed in two different ways: One way is that the registered report could be considered a first version of the final paper containing the background, related research, the research question, hypotheses, data collection, and the analytical methods. The reviewers review the information provided and give input to it. If accepted, the publisher guarantees to publish the paper as long as the methodology is followed, disregarding the outcome of the experiments. After the experiment has been conducted, the reviewers review the final paper, and it will be published — provided that the authors follow the methodology described in the first version of the paper already having been reviewed and accepted. The other way to implement the peer-review process is to publish the registered report itself after the first round of reviews. Then, a separate paper describing the research and its results is guaranteed publication if the authors follow the research protocol described in the published registered report.

The Search for Registered Reports in the Field of AI

The Center for Open Science² maintains a list of journals that support registered reports.³ At the time of writing, it counts more than 250 journals, and only a small handful of them focus on computer science. None of the journals are dedicated to AI.

Being curious about whether this list reflected the reality for the field of AI, I did a small experiment. I visited the homepages of the top-twenty AI journals as ranked by Scimago Journal & Country Rank⁴ and surveyed how well they foster open and transparent research by registering whether they encourage authors to share code and data, provide badges that promote open and transparent research, and support registered reports. The findings are illustrated in figure 1.

On the positive side, eighteen of the twenty journals encourage the authors to publish both code and data. Some of the publishers even provide their own tools for this exact purpose. One of the journals focuses on tutorials and surveys, so they can be excused for not promoting open and transparent science in this form as neither tutorials nor surveys document experimental results. The other journal is more surprising, as it is published by the Association for Computing Machinery. The Association for Computing Machinery promotes reproducibility through artifact review and badges,⁵ just to give an example, so they are well aware of the advantages of publishing code and data used to conduct experiments. I might of course have missed something, as I only reviewed the submission guidelines and did not try to go through the submission process. While the number of journals supporting code and data being shared is good, it is hard to find which papers in these journals do share code and data. None of the journals provide easily accessible badges for this. Badges are used by some of the journals to indicate supporting materials; however, supporting materials are not restricted to code and data. It could as well include videos, images and even podcasts. Although some of the publishers promote reproducibility badges, they are not used by the AI journals they do publish. Finally, and maybe not directly surprising as the list of journals supporting registered reports maintained by the Center for Open Science did hint at this, none of the twenty surveyed journals support registered reports. Not being one who gives up easily, I also surveyed the AI papers and registered reports published in the journal *Royal Society Open Science*, but I did not find any traces of registered reports being published in the field of AI there, either.

For Which Types of Research Are Registered Reports Appropriate?

Are registered reports applicable for all types of research? The short answer is *no*.

First, registered reports document empirical research, so it makes no sense to register theoretical research.

Second, registering a report only makes sense when you have a hypothesis to test. Given that empirical research can be divided into research that tests hypotheses, so-called confirmatory experiments, and research that helps design experiments and identify hypotheses (Cohen 1995), registered reports do not provide any value to the latter. Instead, such research

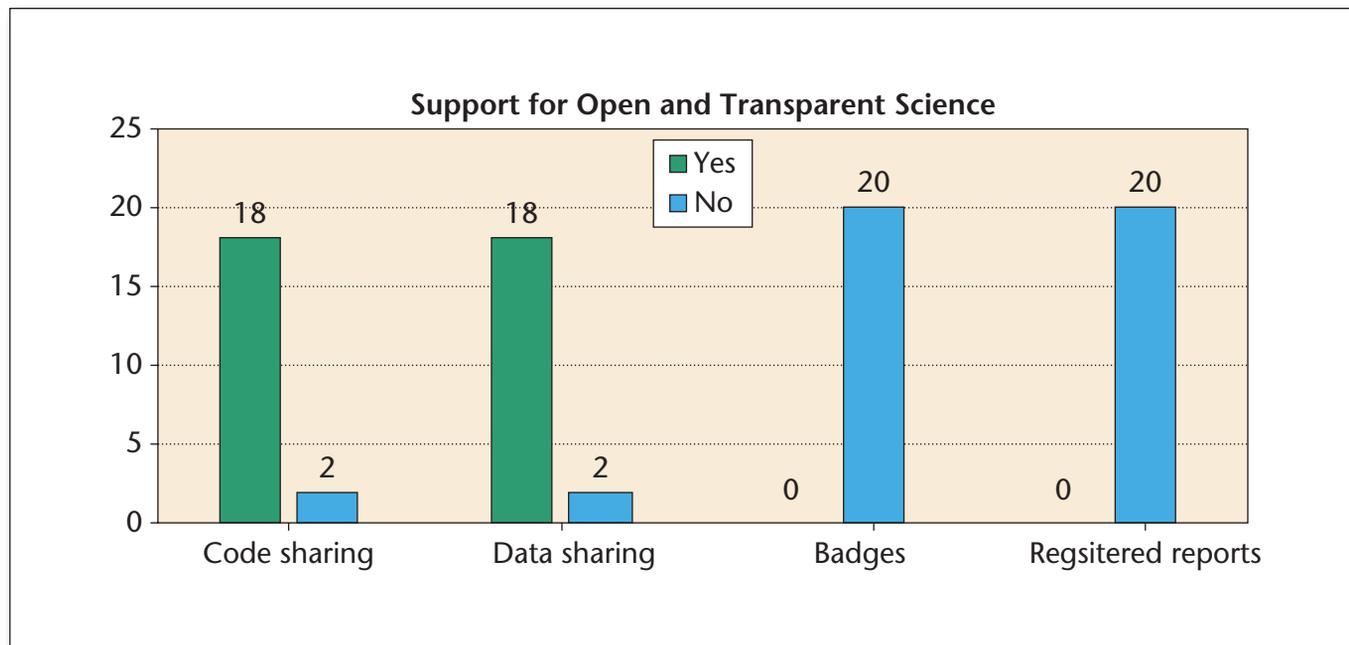


Figure 1. Results from Surveying the Top-Twenty AI Journals.
(As ranked by Scimago Journal & Country.)

is necessary to develop the research projects that are candidates for registered reports. So, only confirmatory experiments are candidates for submitting registered research, and these include manipulation experiments that try to establish causal relationships between factors and observation experiments designed to observe the association between factors and variables. Hence, more than anything else, registered reports are appropriate when the experiments are conducted to establish truths about the world with high confidence.

Third, the extra peer-review step of registered reports slows down the research process. If one is to take the reviews into account when conducting an experiment, one has to wait on these before starting to conduct the experiment. Hence, registered reports might not be suitable when time is of the essence, such as when proposing and testing novel ideas. Novel ideas are only novel *if no one else proposes them*, and hence getting them published quickly is more important than rigorous evaluation.

Fourth, the process with two review steps is not compatible with how most conferences currently select research for publication. Conferences have a single review cycle with one deadline that everyone must meet. However, this could be alleviated, for an example, by conferences that support more than one review cycle. For such conferences, registered reports could be submitted to one cycle and if accepted, the full paper containing the results as well, could be reviewed in a later one. However, there would be

pressure to complete the experiment before the last review cycle. What would happen in cases where registered reports were provisionally accepted, but where the researchers were not able to complete the experiments and submit the final paper before the last deadline? One could argue that conferences could just publish the registered reports lacking the results, but that would result in loose ends where only the description of the experiment and not the results were published. It would be most confusing.

Fifth and finally, most top conferences in AI are organized annually and are thus optimized for spreading research results quickly. This is important in research fields that develop quickly, such as the fields of computer science and AI. In most other research fields this is not the case, and journals are the main source for spreading research results. Hence, registered reports might belong in journals and not in conferences, and thus research studies that are to be submitted to journals are candidates for registered reports.

Conjectures and Refutations

Not only have few, if any, registered reports been published in the field of AI, but the discussion about them has also been missing. There is at least one notable example (Mannarswamy and Roy 2018). This poses the question: What causes this lack of interest in registered reports in the AI community? Below, I propose some conjectures followed by my refutations.

Most Research in AI Is Theoretical, So Registered Reports Are Not Needed. Registered reports are only needed for experimental research, and as most research in AI is theoretical, there is no need to introduce registered reports in AI research. Pat Langley (Langley 1988) noted that machine learning, which is an important part of AI research, has both theoretical and experimental aspects and that most learning algorithms are too complex for a theoretical analysis. Thus, the field has a significant experimental component. Other areas of AI, such as knowledge representation and reasoning, have a less significant experimental component and rely to a larger degree on theory. Still, most AI research is not theoretical. According to a survey we did, eighty-one percent of research published at the top AI conferences is experimental (Gundersen and Kjensmo 2018). So, the proposition that most AI research is theoretical is clearly wrong.

The Increased Amount of Research in AI Leads to More Trustworthy Results, So Introducing Registered Reports Provide Little Value. There is a clear increase in submissions to the most prominent conferences in AI including the Association for Advancements in Artificial Intelligence Conference on AI, the International Joint Conference on Artificial Intelligence, the International Conference on Machine Learning, and the Neural Information Processing Systems conference, which means that many research teams are involved in the field. The field of AI is what John Ioannidis (Ioannidis 2005) calls a *hot scientific field*, and one of the surprising side effects of this hotness is that it is less likely that research results are true. One reason is that research teams might pursue and disseminate the most impressive positive results, leading to publication bias and less rigor in finding the objective truth.

Registered Reports Are Mostly Valuable for Young Researchers Only, But Most Research in AI Is Done by Seasoned Researchers, So No Need for Registered Reports. Several of the researchers that Pain (2015) interviews state that registered reports are most valuable for young researchers, as it ensures that their research follows good research practices from feedback on the research protocols before conducting the research. Also, they do not have to conduct full research projects before getting good publications on their curriculum vitae. Given the growth of research papers published in AI, there must be an influx of new researchers to the field that are not experienced in the field of AI. Hence, all published research, even at top AI conferences, cannot be done by seasoned AI researchers, so registered reports could, indeed, help improve the rigorousness of AI research.

AI Research Is Already Rigorous, So the Value of Registered Reports Is Negligible. All the large conferences in AI have introduced reproducibility checklists and guidelines, and there is a focus on open and transparent research. However, according to Ioannidis (2005), AI is a

young field where analytical methods are still being developed; the greater the flexibility in designs, definitions, and analytical modes in a scientific field, the less likely that the research finding is true. And there are some systematic studies that indicate that Ioannidis might be right. Dacrema et al. (2019) are not able to reproduce the results presented in many recent papers on recommender systems, Henderson et al. (2018) give ample proof that many results in deep reinforcement learning cannot be trusted, and Lucic et al. (2018) show that several developments of the seminal generative adversarial network model are not the improvements they were presented as. There are many reasons why evaluating AI research is hard, but these recent examples do indicate that there is still at least some potential for improving the analytical methods used by AI researchers.

Registered Reports Were Proposed by Psychologists to Solve a Problem in Psychology That Does Not Exist in AI. Registered reports were indeed proposed to solve methodological problems in psychology. However, AI is not just a hard science where experiments are conducted on deterministic computers; it is, to a larger degree, a science conducted on the borders of computers, humans, and society. Hence, the AI community must adopt the lessons learned from the communities that have conducted such research before.

There Is No Reproducibility Crisis in AI. I clearly disagree, as you would know if you read my previous column in *AI Magazine*. Publishing biases, hotness of the field, poorly documented research, research methodology and analytics methods still being in development, small effect sizes, and stochasticity of algorithms and environments are some of the issues that lead to poor reproducibility. There is still work to do to fully understand all issues related to reproducibility.

The Case Against Registered Reports

In conclusion, I am not able to build *any* case against registered reports. They would clearly provide value in the field of AI. However, not for all research, and not even for all empirical research. Maybe they could be supported by conferences, but I am not arguing for it. I clearly think journals should support registered reports, but as my study shows, none of the top-twenty AI research journals do. I was not even able to find any registered reports being published on the topic of AI.

So, if someone wants to do novel AI research, it could be to submit the first registered report in our field. The question is: *Where?*

Notes

1. The practice of mining data for statistically significant relationships between variables without evaluating the relationships on separate data.



Fourth Conference on AI, Ethics, and Society

May 19–21

www.aies-conference.com

2. www.cos.io/
3. www.cos.io/our-services/registered-reports
4. www.scimagojr.com/journalrank.php?area=1700&category=1702
5. www.acm.org/publications/policies/artifact-review-and-badging-current

References

- Cohen, P. R. 1995. *Experimental Methods for Artificial Intelligence*. Cambridge, MA: The MIT Press
- Dacrema, M. F.; Cremonesi, P.; and Jannach, D. 2019. Are We Really Making Much Progress? A Worrying Analysis of Recent Neural Recommendation Approaches. In *Proceedings of the 13th Association for Computing Machinery (ACM) Conference on Recommender Systems*, 101–109. New York: ACM. doi.org/10.1145/3298689.3347058.
- Gundersen, O. E., and Kjensmo, S. 2018. State of the Art: Reproducibility in Artificial Intelligence. In *Proceedings of the Thirty-Second Association for the Advancement of Artificial Intelligence (AAAI) Conference on Artificial Intelligence*, 1644. Palo Alto, CA: AAAI Press.
- Henderson, P.; Islam, R.; Bachman, P.; Pineau, J.; Precup, D.; and Meger, D. 2018. Deep Reinforcement Learning that Matters. In *Proceedings of the Thirty-Second Association for the Advancement of Artificial Intelligence (AAAI) Conference on Artificial Intelligence*, 3207. Palo Alto, CA: AAAI Press.
- Ioannidis, J. P. A. 2005. Why Most Published Research Findings are False. *PLoS Medicine* e124. doi.org/10.1371/journal.pmed.0020124
- Langley, P. 1988. Machine Learning as an Experimental Science. *Machine Learning* 3: 5–8. doi.org/10.1023/A:1022623814640
- Lucic, M.; Kurach, K.; Michalski, M.; Gelly, S.; and Bousquet, O. 2018. Are GANs Created Equal? A Large-Scale Study. *arXiv.org* arXiv:1711.10337v4
- Mannarswamy, S., and Roy, S. 2018. Evolving AI from Research to Real Life — Some Challenges and Suggestions. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI)*. IJCAI, Inc. doi.org/10.24963/ijcai.2018/717.
- Pain, E. 2015. Register Your Study as a New Publication Option. *Science* 350: 130. www.sciencemag.org/careers/2015/12/register-your-study-new-publication-option
- Odd Erik Gundersen** (PhD, Norwegian University of Science and Technology) is the Chief AI Officer at the renewable energy company TrønderEnergi AS and is an adjunct associate professor at the Department of Computer Science at the Norwegian University of Science and Technology. Gundersen has applied AI in the industry, mostly for startups, since 2006. Currently, he is investigating how AI can be applied in the renewable energy sector and for driver training, and how AI can be made reproducible.