

DAY ONE PROJECT

Leveraging Machine Learning to Reduce
Cost and Burden of Reviewing Research
Proposals at Science and Technology
Agencies

Ilana Harrus
August 2020

Summary

With about \$130 billion USD, the United States leads the world in federal research and development (R&D) spending.¹ Most of this spending is distributed by science and technology agencies that use internal reviews to identify the best proposals submitted in response to competitive funding opportunities. As stewards of quality scientific research, part of each funding agency's mission is to ensure fairness, transparency, and integrity in the proposal-review process. The selection process is a crucial aspect of ensuring that federal dollars are invested in quality research.

Manual proposal review is time-consuming and expensive, costing an estimated \$2,000–\$10,000 per proposal. This equates to an estimated \$300 million spent annually on proposal review at the National Science Foundation alone. Yet at current proposal-success rates (between 5% and 20% for most funding opportunities), a substantial fraction of proposals reviewed are simply not competitive. We propose leveraging machine learning to accelerate the agency-review process without a loss in the quality of proposals selected and funded. By helping filter out non-competitive proposals early in the review process, machine learning could allow substantial financial and personnel resources to be repurposed for more valuable applications. Importantly, machine learning would not be used to evaluate scientific merit—it would only eliminate the poor or incomplete proposals that are immediately and unanimously rejected by manual reviewers.

The next administration should initiate and execute a pilot program that uses machine learning to triage scientific proposals. To demonstrate the reliability of a machine-learning-based approach, the pilot should be carried out in parallel with (and compared to) the traditional method of proposal selection. Following successful pilot implementation, the next administration should convene experts in machine learning and proposal review from funding agencies, universities, foundations, and grant offices for a day-long workshop to discuss how to scale the pilot across agencies. Our vision is that machine-learning will ultimately become a standard component of proposal review across science and technology agencies, and improving the efficiency of the funding process without compromising the quality of funded research.

Challenge and Opportunity

Allocating research funding is expensive, time-consuming, and inefficient for all stakeholders (funding agencies, proposers, reviewers, and universities). The actual cost of reviewing proposals (including employee salaries and administrative expenses) has never been published by any federal funding agency. Based on our experience with the process, we estimate the cost to be between \$2,000 and \$10,000 per proposal, with the variation reflecting the wide range of

¹ National Science Board: Science & Engineering Indicators (2020). *Research and Development: U.S. Trends and International Comparisons*. Available at <https://nces.nsf.gov/pubs/nsb20203/assets/nsb20203.pdf>.

proposals across programs and agencies. For the National Science Foundation (NSF), which reviews around 50,000 research proposals each year², this equates to an average of \$300 million spent annually on proposal review.

Multiple issues beyond cost plague the proposal-review process. These include the following:

(1) Decreasing proposal-success rates. This decline is attributable to a combination of an increase in the number of science, technology, engineering, and math (STEM) graduates in the United States³ and an increase in the size of average federal STEM funding awards (from \$110,000 to about \$130,000 in less than 10 years)⁴. Current success rates are low enough that the costs of applying for federal funding opportunities (i.e., from time spent on unsuccessful proposals) may outweigh the benefits (i.e., funding received for successful proposals).⁵

(2) Difficulties recruiting qualified reviewers.

(3) Delayed decisions. For example, NSF takes more than six months to reach a funding decision for about 30% of proposals reviewed.

(4) Increasing numbers of identical re-submissions. With proposal-success rates as low as 5%, the results of selection processes are often seen as representing “the luck of the draw” rather than reflective of fundamental proposal merit. Hence there is a growing tendency for principal investigators (PIs) to simply re-submit the same proposal year after year rather than invest the time to prepare new or updated proposals.

There is a consensus that the current state of proposal review is unsustainable. Most proposed solutions to problems summarized above are “outside” solutions involving either expanding available research funding or placing restrictions on the numbers of proposals submitted that may be submitted (by a PI or an institution). Neither option is attractive. Partisanship combined with the financial implications of COVID-19 render the possibility of an increased budget for S&T funding agencies vanishingly small. Restrictions on submissions are generally resented by scientists as a “penalty on excellence”. **Incorporating machine learning could improve the efficiency and effectiveness of proposal review at little cost and without limiting submissions.**

Incorporating machine learning would also align with multiple federal and agency objectives.

On January 4, 2011, President Obama signed the GPRA Modernization Act of 2010. One of the purposes of the GPRA Modernization Act was to “lead to more effective management of government agencies at a reduced cost”.⁶ One of NSF’s Evaluation and Assessment Capability (EAC) goals established in response to that directive is to “create innovative approaches to

² NSF alone funds almost one-fourth of all basic research proposals in the United States. As a result, most of the references in this memo are to NSF’s processes. Yet most—if not all—points made apply equally well to other S&T funding agencies.

³ See, for example: National Science Foundation (NSF). (2014). Who earns bachelor’s degrees in science and engineering? *Science and Engineering Indicators Report 2014*. Available at <https://www.nsf.gov/nsb/sei/edTool/data/college-14.html>.

⁴ NSF. (2019). *Report on the National Science Foundation’s Merit Review Process Fiscal Year (2017)*. NSB-2019-15. Available at <https://www.nsf.gov/nsb/publications/2018/nsb201915.pdf>.

⁵ von Hippel, T.; von Hippel, C. (2015). To Apply or Not to Apply: A Survey Analysis of Grant Writing Costs and Benefits. *PLOS ONE*, 10(3): e0118494. doi:10.1371/journal.pone.0118494.

⁶ NSF. (n.d.). GPRA Modernization Act of 2010. Available at https://www.nsf.gov/about/performance/gpra_modernization.jsp.

assessing and improving program investment performance”⁷. Indeed, two of the four key areas identified in NSF’s most recent Strategic Plan (2018) are to “make information technology work for us” and “Streamlining, standardizing and simplifying programs and processes.”⁸ In addition, NSF recognized the importance of reviewing its processes for efficiency and effectiveness in light of OMB memo M-17-26.⁹ NSF’s Strategic Plan includes a strong commitment to “work internally and with the Office of Management and Budget and other science agencies to find opportunities to reduce administrative burden.”¹⁰ These principles are also mentioned in NSF’s 2021 budget request to Congress¹¹, as a part of Strategic Goals (e.g., “Enhance NSF’s performance of its mission”) and Strategic Objectives (e.g., “Continually improve agency operations”). Finally, long-term goal outlined in the Strategic Plan is reducing the so-called “dwell time” for research proposals—i.e., the time between when a proposal is submitted and a funding decision is issued.¹²

Incorporating machine learning into proposal review would facilitate progress towards each of these goals. Using machine learning to limit the number of proposals subjected to manual review is a prime example of “making information technology work for us” and would certainly help streamline, standardize, and simplify proposal review. Limiting the number of proposals subjected to manual review would also reduce administrative burden as well as dwell time. In addition, money saved from using machine learning to weed out non-competitive proposals can be used to fund additional competitive proposals, thereby increasing return on investment (ROI) in research-funding programs. Additional benefits include an improved workload for expert reviewers—who will be able to focus on reviewing the scientific merit of competitive proposals instead of wasting time on non-competitive proposals—as well as the establishment of a strong disincentive for PIs to resubmit identical proposals years after years. The latter outcome in particular is expected to improve proposal quality in the long run.

Plan of Action

We propose the following steps to implement and test a machine-learning approach to proposal review:

(1) Initiate and execute a pilot program that uses machine learning to triage scientific proposals. To demonstrate the reliability of a machine-learning-based approach, the pilot should be carried out in parallel with (and compared to) the traditional method of proposal selection. The pilot would be deemed successful if the machine-learning

⁷ NSF. (n.d.). OIA’s Evaluation and Assessment Capability (EAC) Section. Available at <https://www.nsf.gov/od/oia/eac/index.jsp>.

⁸ NSF. (2018). *Building the Future: Investing in Discovery and Innovation*, p. 24.

⁹ White House Office of Management and Budget. (2017). Reducing the Burden for Federal Agencies by Rescinding or Modifying OMB Memoranda. M-17-26.

¹⁰ NSF. (2018). *Building the Future*, p. 26.

¹¹ NSF. (2020). *National Science Foundation: FY 2021 Budget Request to Congress, Performance & Management 3–4*.

¹² NSF. (2018). *Building the Future*, p. 29.

DAY ONE PROJECT

algorithm was able to reliably identify proposals ranked poorly by human reviewers,¹³ and/or proposals rejected unanimously by review panels. NSF—particularly the agency’s Science of Science and Innovation Policy (SciSIP) Program—would be a natural home for such a pilot.

(2) Showcase pilot results. Following a successful pilot, the next administration should convene experts in machine learning and proposal review from funding agencies, universities, foundations, and grant offices for a day-long workshop. The workshop would showcase pilot results and provide an opportunity for attendees to discuss how to scale the pilot across agencies.

(3) Scale pilot across federal government. We envision machine learning ultimately becoming a standard component of proposal review across science and technology agencies, improving the efficiency of the funding process without compromising the quality of funded research.

Reducing the numbers of scientific proposals handled by experts without jeopardizing the quality of science funded benefits everyone—high-quality proposals receive support, expert reviewers don’t waste time on non-competitive proposals, and the money saved on manual proposal review can be reallocated to fund additional proposals. Using machine learning to “triage” large submission pools is a promising strategy for achieving such objectives. Preliminary compliance checks are already almost fully automated.¹⁴ Machine learning would simply extend the automation stage one step further. We expect that initial costs of developing appropriate machine-learning algorithms and testing algorithms in pilots would ultimately be justified by greater long-run ROI in research-funding programs. We envision a pilot that could benefit the government but also foundations that are increasingly shouldering research funding. Ideally the pilot would be experimented in two different set-ups: a government funding agency and a Foundation.

¹³ At NSF, for instance, this could mean identifying proposals rated 3 or below. See, for example: NSF. (2020). *National Science Foundation’s Merit Review Process: Fiscal Year 2018 Digest*. NSB-2020-13.

¹⁴ NSF. (n.d.). Automated Proposal Compliance Checks Performed by Research.gov as of April 30th, 2018. Available at https://www.nsf.gov/bfa/dias/policy/autocheck/rgovcompliancechecks_april18.pdf.

Frequently Asked Questions

Is it possible technically to triage science proposals?

Yes. Funding agencies keep large databases of old, archived proposals. A machine-learning algorithm could be trained and validated on these proposals. Partnerships with universities could likely yield additional proposals for algorithm-training purposes.

Will machine learning be used to evaluate science?

Machine learning is not quite ready to “evaluate” science. Review by human experts is still needed to assess scientific merit. However, a substantial fraction of proposals all fail in specific ways that could be identified by an algorithm. In other words, we do not propose using machine learning to indicate which proposals *should* be funded, but rather to quickly identify proposals *that would never be funded at all*. Conducting a pilot in parallel with a conventional review process is a way to ensure that the two approaches yield the same results.

Would machine learning cause us to miss great but unconventional proposals?

It is very unlikely. The number of “great but unconventional” proposals is small to begin with. Those that do exist essentially never fall in the very bottom fraction (e.g., the bottom 20%) of ranked proposals, and are essentially never rejected by review panels as a whole. Hence a machine-learning algorithm would pass such proposals on for manual review instead of filtering them out.

How much will the pilot and program cost?

Expected costs are low. Expenses would be almost entirely limited to the costs of personnel needed to develop and administer the machine-learning algorithm. We propose supplying needed funds from the IT departments of implementing agencies (e.g., NSF and the National Institutes of Health (NIH) have dedicated groups for their reviews). The models would also become better and better as the algorithm learns from the growing statistics.

How much could agencies reallocate in costs by incorporating machine learning into pilot review?

The actual cost of reviewing proposals (including employee salaries and administrative expenses) has never been published by any federal funding agency, making it difficult to provide precise budget implications. That said, we estimate the average true cost of reviewing one proposal to about \$6,000 (between \$2,000 and \$10,000 depending on the program and the review scope). Considering the fact that NSF, NIH, and the National

DAY ONE PROJECT

Aeronautics and Space Administration (NASA) receive approximately 50,000, 60,000, and 10,000 research proposals annually, incorporating machine learning into proposal review could save as much as \$160 million per year.

Could machine learning be incorporated into proposal review at non-S&T agencies?

Yes! In fact, the potential of this approach goes beyond government entirely. Machine learning could be adopted by universities or foundations to help PIs and grant managers improve proposals before submission, for instance. This could yield better-quality proposals, save staff time dedicated to proposal revision, and reduce opportunity costs of proposal application.

What about machine-learning bias? Could machine learning disproportionately filter out proposals from certain applicant groups (e.g., underrepresented populations)?

This is unlikely. As stated above, we do not propose using machine learning to indicate which proposals *should* be funded, but rather to quickly identify proposals *that would never be funded at all*. Money saved from incorporating machine learning into proposal review could actually be used to fund special funding opportunities reserved for priority applicant groups (e.g., women, minorities, younger PIs, and PIs from less-well-known institutions).

Why would a pilot be the place to start, and what would be its evolution?

We want to demonstrate that a machine-learning-based approach works. Conducting a pilot in parallel with a conventional review process is a way to ensure that the two approaches yield the same results. Once proven, we expect that S&T funding agencies would work with other stakeholders to extend machine learning to proposal review in different programs. Our vision is that machine-learning-based proposal review would ultimately become standard across government.

About the Author



Dr. Ilana Harrus is the Senior Program Associate of the Artificial Intelligence: Applications/Implications Initiative hosted by the American Association for the Advancement of Science (AAAS) Scientific Responsibility, Human Rights and Law Program. An astrophysicist by training, Dr. Harrus spent more than 15 years working at NASA and NSF, including serving as Program Scientist for three NASA space missions and as the lead for the NASA Astrophysics R&D program. She is a PMI-certified Project Manager (PMP), and holds a Ph.D. in Physics from Columbia University as well as a Master in Information Systems from University of Maryland, Baltimore County.

About the Day One Project

The Day One Project is dedicated to democratizing the policymaking process by working with new and expert voices across the science and technology community, helping to develop actionable policies that can improve the lives of all Americans, and readying them for Day One of a future presidential term. For more about the Day One Project, visit dayoneproject.org.