National Science Foundation

Where Discoveries Begn

Anti-Harassment Study: Technical Appendix

Table of Contents

I. Introduction

The National Science Foundation's (NSF) anti-harassment study responds to questions in the FY 2020 Learning Agenda related to NSF's recent anti-harassment policies to promote safe, harassment-free environments for the practice of science. This technical memo accompanies the main report from the study and describes the data, samples, and analysis methods used in greater detail.

This study has three major components:

1. **Communication analysis:** a qualitative review of NSF's communication activities, including communication materials and dissemination strategy.

2. **Term and condition analysis:** a quantitative, descriptive analysis of awardee organizations' public web content on sexual harassment, other harassment, and sexual assault policies, with particular attention to whether the organizations include references or links to NSF's anti-harassment policy webpages, especially the NSF Term & Condition (T&C) on harassment.[1] This consists of two parts:

   a) Identifying links to NSF's T&C-related webpages using Ahrefs, a search engine optimization tool

   b) Analyzing webpage content to identify whether institutions have anti-harassment policies on their webpages and whether their webpages reference NSF's anti-harassment policies, including the T&C

3. **Conference policy analysis:** a quantitative and qualitative analysis of NSF proposals that measures how often they include references to anti-harassment policies and practices before and after NSF's new policy.

II. Data

This section describes the data used in each analysis.

A. Communication Analysis

The communication analysis relied on two types of data: (1) artifacts related to communications around NSF's anti-harassment policies and (2) interviews with NSF staff who played key roles in the development and dissemination of those policies.

1. Communication Artifacts

The analysis included communication materials around NSF's anti-harassment policies: disseminating and publicizing the need for them; their development, implementation, and public reaction; and their role within broader anti-harassment efforts. These included policy language and related guidance, applicable government reports, NSF website materials, Federal Register materials, formal communications plans, internal communications, organizational journals and websites, Congressional inquires and testimony, press statements, social media events and mentions, etc. These materials included both original/primary documents (such as policy guidance or journal publications) as well as references and secondary artifacts (such as an internal email distributing a social media post for discussion or a list of attendees at an in-person event). The sample excluded some other types of communication, such as presentations at conferences and speeches at grantee convenings.

---

[1] The T&C requires any organizations receiving new awards or funding amendments to existing awards after October 22, 2018, to "notify NSF of any findings/determinations of sexual harassment, other forms of harassment, or sexual assault regarding an NSF funded Principal Investigator (PI) or co-PI, or of the placement of the PI or co-PI on administrative leave, or the imposition of any administrative action relating to harassment or sexual assault finding or investigation." It also triggers NSF engagement with institutions filing reports.

NSF staff from the Evaluation and Assessment Capability (EAC) section and the Office of Equity and Civil Rights (OECR, formerly the Office of Diversity and Inclusion) identified the initial set of materials determined to be most relevant for the analysis. This sample was supplemented by additional suggestions of relevant artifacts identified during the interviews with staff from OECR, the Office of Legislative and Public Affairs (OLPA), the Office of General Counsel (OGC), and the Office of Budget, Finance and Award Management (BFA).

The final sample for analysis includes 110 artifacts spanning January 2016 through December 2020. Due to the nature of the sample of materials and exclusion of some forms of communication, the analysis likely underestimates the extent of the approach.

2. Interviews

As part of a broader effort to obtain input from NSF on the plans for this study, interviews were conducted with seven key NSF stakeholders, including individuals from OECR and OGC who were architects of the new policy. A semi-structured protocol was used for these interviews, for which one objective was to gauge policy implementation relevant to Questions 4 (the T&C analysis) and 5 (the conference policy analysis). Interview topics included the following:

- Whether there is/was a formal communication plan for either policy
- How each policy has been communicated:
  - Within NSF
  - To review panels, including any instructions that were given in FY 2018, before the policy was instituted (may apply only to conference policy)
  - To the research community outside NSF
  - To other government entities
- What communication activities are perceived to be most successful in educating the intended audience on the policies
- What the ongoing and planned efforts are to communicate the policies

The information provided via these interviews was designed to provide context supporting the analysis of communcation materials.

B. Term and Condition Analysis

For the term and condition (T&C) analysis, the first step was collecting data from NSF systems to identify institutions subject to the T&C. This included all institutions that were on new awards or funding amendments to existing awards after the T&C went into effect. Specifically, the study included all institutions on these awards that were listed as the awardee institution or as a principal investigator (PI) or co-PI institution. The second step was collecting data on the institutions' characteristics using several sources within and outside of NSF, such as the Integrated Postsecondary Education Data System (IPEDS). The third step was gathering data from Ahrefs, a search engine optimization tool that allows us to identify webpages linking to NSF's webpages related to the T&C. The last step was collecting text data from institutions' webpages to use for an analysis of webpage content.

1. NSF Data Systems

The study used the following NSF data systems to identify all institutions that were subject to the T&C.

(i) NSF Solr API: NSF's Solr Application Programming Interface (API) was used to identify institutions subject to the anti-harassment T&C. The first step was identifying all new awards that had effective dates between when the T&C went into effect (Oct. 22, 2018) and the start of the study (Feb. 1, 2021). All institutions that received awards during this time were added to the list of institutions of interest. Because the award data did not include PI or co-PI institution information, Solr was used to look up all proposals submitted from FY 2016 onward. Awards were then matched to

proposals to pull in the PI and co-PI institutions from the proposal data for all awards that went into effect during the window of interest.

(ii) NSF Fastlane Database: The T&C applies to all funding amendments that occurred after Oct. 22, 2018. NSF staff used the Fastlane database to identify any funding amendments from Oct. 22, 2018, to Feb. 1, 2021 (the start of the study) and shared this with the study team.[2] Solr was then used to identify the original proposals for the funding amendments to identify their PI and co-PI.

To help us match the list of institutions subject to the T&C with external data sources (e.g., IPEDS), additional institution information was pulled from the Fastlane database, specifically, the "cover sheet" data for each awardee institution for all awards granted from Oct. 22, 2018 to Feb. 1, 2021. This included the institution's DUNS number, ID, city, and state.

2. Institutional Characteristics Data

Information about the institutions subject to the T&C was collected from the following sources:

(i) IPEDS: From the Integrated Postsecondary Education Data System (IPEDS) for 2019–20, an indicator for whether an institution is an Historically Black College or University (HBCU) and an indicator for whether an institution is a Tribal College or University was collected.

(ii) Carnegie Classifications: From the Carnegie Classification of Institutions of Higher Education for 2018, the basic Carnegie classifications were used to classify institutions into three categories of research intensity: very high research activity (R1, doctoral universities); high research activity (R2, doctoral universities); and companies, nonprofits, and IHEs outside the U.S.. Indicators for whether an institution is a minority-serving institution or a Hispanic-serving institution were also collected.

(iii) U.S. Department of Education, Office of Postsecondary Education: To identify Alaska Native- and Native Hawaiian-serving institutions, the Office of Postsecondary Education's eligibility matrix for Title III, Part A funding was used. Institutions are eligible for this funding if they meet the following conditions: "An Alaska Native-serving institution may receive a grant under section 317 of the Higher Education Act (HEA) if, at the time of application, it has an enrollment of undergraduate students that is at least 20 percent Alaska Native students. A Native Hawaiian-serving institution may receive a grant authorized under section 317 of the HEA if, at the time of application, it has an enrollment of undergraduate students that is at least 10 percent Native Hawaiian students."[3] Institutions have to apply for and receive a designation to be eligible.

(iv) NSF Solr API: To group institutions by how much funding they receive from NSF, Solr was used to collect all awards with an effective date in FY 2016 through FY 2020. The total intended award amount was summed for each awardee institution on these awards. Institutions were placed into funding categories that roughly distributed institutions subject to the T&C evenly: $1–250,000; $250,001–1,000,000; $1,000,001–2,500,000; and $2,500,001 or more. (Institutions that did not receive funding directly as an awardee instittuion during this period of time were not placed into any funding category and are not included in the analyses that report results by funding category.)

3. Assembling the Institution Dataset for the Analysis

Data from NSF's Solr API and the funding amendment data from Fastlane were first combined to create one dataset with all awards or funding amendments that began after the T&C went into effect and before the start of this study. In total, there were 35,114 unique awards or funding amendments during the window of interest. PI or co-PI institutions from the proposal data were identified for 99% of these. From this a list of institutions was constructed that included

---

[2] Funding amendments were defined as any amendment to an existing award that changed the amount of funding for an award.
[3] https://www2.ed.gov/programs/iduesannh/eligibility.html

all institutions that either received an award or funding amendment or had a PI or co-PI on an award or funding amendment during this period of time.

Many of the institutions had multiple variations of their names. To help de-duplicate institutions, names were normalized to remove common phrases (e.g. "and") and punctuation (e.g. hyphens, apostrophes). When available, DUNS numbers and Institution IDs from Fastlane award cover sheets were also used to de-duplicate institutions. For cases where the institution's name was a research foundation, corporation, or center affiliated with a college, the name was changed to be the college they are associated with (e.g., "West Virginia University Research Corporation" became "West Virginia University"). Some institution names referred to systems (e.g., University of California System) or did not specify a campus (e.g., University of New Mexico). In these cases, institution names were set to the main flagship campus in the system, unless the PI or co-PI institution specified a different campus in the system.

To match the list of institutions to IPEDS, DUNS number was used whenever possible. (Approximately 60% of higher education institutions found in IPEDS were matched with the DUNS number). For the remaining institutions, repeated rounds of matching were conducted using name, state, and city (except for PI and co-PI institutions, which did not have state and city and could only use name). Exact matches were done first, followed by fuzzy matches on name within state. All fuzzy matches were reviewed before the match was accepted, and manual checks were conducted for any unmatched institutions using city and state to assist. All higher education institutions in the U.S. were matched to IPEDS except one that was not listed in IPEDS.

To merge in institutional characteristics for all higher education institutions, the UnitID field from IPEDS was used to merge with Carnegie Classification data to bring in special designation data (minority-serving institutions (MSIs) and Hispanic-serving institutions (HSIs)) and research intensity. Data from the Office of Postsecondary Education on Alaska Native- and Native Hawaiian-serving institutions were merged in by hand. Because there were so few institutions that had these designations, it was possible to limit the pool of eligible institutions using state and then hand match on name.

To identify the total funding awarded to each institution from FY 2016 to FY 2020, a dataset was pulled from Solr that listed all institutions that received any award from FY 2016 to FY 2020 and the award amount. For every award and its associated awardee institution, the cleaned awardee institution name from the cleaning process was merged in. For each cleaned institution name, the award amounts were summed and then this total funding amount was merged back into the list of eligible institutions. Some institutions that were subject to the T&C did not have any awarded funding from FY 2016 to FY 2020 because they were exclusively PI or co-PI institutions.

Last, the domain names were identified for all institutions. To identify domains for U.S. higher education institutions, the domain for the web address listed in IPEDS for an institution was extracted and checked to ensure that it was valid. For any invalid domains and all other institutions not in IPEDS (such as companies and nonprofits), the following process was used to determine the correct domains:

1. Conducted a Google search for the institution. Collected the top 3 webpages from the search results.
2. Filtered out common domains not linked to institutions (e.g., Wikipedia).
3. Took the first webpage with an exact match between the domain and institution name.
4. Among the top three results, identified if the same domain appears more than once. If so, used that domain for any institutions that did not yet have a domain.
5. Manually reviewed all other institution domains.

In this cleaned dataset (the *institution dataset*), 4,108 unique institutions were identified that either received an award or funding amendments or had a PI or co-PI on an award or funding amendment during this period of time. In total, 30 percent were higher education institutions, 12 percent nonprofits, 38 percent companies, 16 percent individuals, and 4 percent government institutions. For the analysis sample (discussed more in the Sample section (III.A) below),

individuals or government institutions were not included, while higher education institutions, nonprofits, and companies were included.

4. Ahrefs

Data were collected from Ahrefs, a search engine optimization company that can identify all webpages linking to a webpage of interest on the internet. Ahrefs was used to identify all webpages linking to following webpages, which all provide information about NSF's new award T&C regarding sexual harassment, other forms of harassment, and sexual assault:[4]

- NSF's T&C webpage[5]
- NSF's T&C News Release, 10-082[6]
- NSF's webpage on stopping harassment, which mentions the T&C several times[7]

The query used in Ahrefs identified all links to these webpages that were active within the 24-month period prior to February 25, 2021, the date the search was conducted. The Ahrefs search returned 203,911 URLs "linking" to the NSF webpages. All instances where URLs were linking to the same webpage were de-duplicated, yielding 50,816 unique URLs.

For each URL, the domain name associated with the URL was identified.[8] In some cases, the URLs collected from Ahrefs contained the IP addresses rather than associated domain names (e.g., the IP 160.10.5.68 resolves to westga.edu). For these, all webpages with IP addresses were mapped to their appropriate domain. Domains were then cleaned, normalized, and de-duplicated, resulting in 249 unique domains.

The Ahrefs data was then merged with the *institution dataset,* merging on domain. Out of 249 domains in the Ahrefs data, 83 domains merged with the list of institutions subject to the T&C, indicating that 83 institutions out of 3,296 subject to the T&C in the analysis sample had at least one webpage linking to the NSF webpages. The final dataset for the Ahrefs analysis (the *Ahrefs analysis dataset*) contains the 3,297 institutions subject to the T&C and includes an indicator variable for each institution, indicating if the institution had a link to at least one of the three NSF webpages referencing the T&C.

5. Institutional Webpages

Text was collected from institutions' webpages for analysis. Webpages of interest were identfied using automated web searches via Google's search engine and then their text was collected for analysis. The specific web searches used and the process of testing and choosing the optimal web searches is discussed further in the Methods section (IV.B) below.

C. Proposal Analysis

The proposal analysis combined data from NSF systems to identify conference proposals and bring in PI institution attributes. The analysis also uses data on institutions' characteristics using several sources within and outside of NSF, such as the Integrated Postsecondary Education Data System (IPEDS).

---

[4] The T&C requires any organizations receiving new awards or funding amendments to existing awards after October 22, 2018, to "notify NSF of any findings/determinations of sexual harassment, other forms of harassment, or sexual assault regarding an NSF funded Principal Investigator (PI) or co-PI, or of the placement of the PI or co-PI on administrative leave, or the imposition of any administrative action relating to harassment or sexual assault finding or investigation." It also triggers NSF engagement with institutions filing reports.

[5] https://www.nsf.gov/od/odi/term_and_condition.jsp

[6] https://www.nsf.gov/news/news_summ.jsp?cntn_id=296610

[7] https://www.nsf.gov/od/odi/harassment.jsp

[8] For example, within the following URL, the domain is the bolded portion:
https://www.**westga.edu**/academics/cosm/geosciences/profile.php?emp_id=23740

1. NSF Data Systems

The analysis used data from the following NSF data systems to identify all conference proposals before and after NSF's new policy:

(i) NSF Solr API: The analysis team used NSF's Solr API to pull text content and metadata for proposals that requested funding to convene a conference, workshop, symposium, meeting, or summit, or requested travel funding to attend a specific one of those events. The analysis sample included proposals received between Oct. 1, 2015, and March 31, 2021.

2. Institutional Characteristics Data

This analysis used the same approach outlined in the Terms and Conditions Analysis data section (II.B.2) to identify PI institutions' characteristics from IPEDS and Carnegie Classifications. This analysis also used the following sources for additional data on institutions' characteristics:

(i) NSF Solr API: The analysis team used NSF's Solr API "Institution Attributes" field to identify Alaska Native- and Native Hawaiian-serving institutions. The Alaska Native PI institutions reported in the Solr API for the proposal set is consistent with those identified using the U.S. Department of Education, Office of Postsecondary Education 2021 Eligibility Matrix data (as specified under the Terms and Conditions Analysis data section, II.B.2). However, the U.S. Department of Education's 2021 Eligibility Matrix classifies one institution as a Native Hawaiian institution that is not classified as such in the proposal dataset, and this analysis used NSF's categorization instead of the U.S. Department of Education's.

3. Assembling the Proposal Dataset for the Analysis

For a description on the iterative process used for identifying proposals with conference funding, refer to the Methods section (IV.C.1). The analysis team systematically leveraged NSF's Solr API to create customized searches based on specific content criteria on fields like the proposal's title and proposal reception date ranges.

To merge PI institution characteristics onto the proposal dataset, the analysis followed the same name normalization approach as the one outlined in the Terms and Conditions Analysis Data section (II.B.3), but for PI institution name. This resulted in PI institutions having normalized institution names as well as a UnitID, which allowed us to bring in IPEDS (HBCUs and Tribal colleges) and Carnegie Classification data (research intensity, MSI, HSI). All higher education PI institutions in the U.S. were matched to IPEDS except one that was not listed in IPEDS. To bring in institutional characteristics reported in the Carnegie Classification data for all U.S. higher education PI institutions, the analysis used the UnitID field from IPEDS to merge with Carnegie Classification data.

To merge in Alaska Native- and Native Hawaiian-serving institution status at the PI institution level, the analysis used a crosswalk that mapped NSF Solr API "Institution Attributes" from the proposal institution level to the PI institution level. There was no variation at the proposal attribute level for these two indicators in the proposal dataset. The analysis team then merged the attributes using the normalized institution and PI institution names.

To create the anti-harassment indicator (which indicates whether a proposal references an anti-harassment policy or practice) described in the Methods section (IV.C.2), the description field was pre-processed using the following steps: (1) separating the text into sentences; (2) removing numbers, Roman numerals, and stopwords (articles, prepositions, conjunctions, and pronouns) from the text; (3) changing hyphenated terms associated with anti-harassment to single terms (e.g., "code of conduct" or "code-of-conduct" got converted to "codeofconduct"); (4) replacing URLs with a general URL indicator; and (5) lemmatizing terms[9] except for anti-harassment-associated terms.

---

[9] Lemmatizing groups similar words that have different forms (e.g., "organize" and "organizing") and reduces noise in the data.

The main interrupted time series model (ITS) analysis outlined in the Methods section (IV.C.3) used three datasets: one at the annual level, another at the quarterly level, and one at the monthly level. The latter two refer to fiscal periods. In each dataset, a time trend variable (t) reflects the number of periods since the announcement/effective date. This variable was centered by calculating the number of days between the start of the fiscal year (FY) or quarter and the start of the announcement/effective periods, divided by the number of days in the year/quarter.

III. Sample

A. Term and Condition Analysis

For both parts of the term and condition analysis, the *institution dataset* was used for the analysis sample. However, this did not include institutions that are individuals or government institutions, as they receive a very small portion of total funding.[10] Table 1 shows the characteristics of institutions that are subject to the T&C that were included in the analysis sample.

TABLE 1. CHARACTERISTICS OF INSTITUTIONS THAT ARE SUBJECT TO THE T&C

Institution Characteristics	Number of Institutions	Percentage of Institutions	Percentage of Total Funding (FY 2016–FY 2020)
All Institutions	3,296		
All			
Higher education institution	1,241	37.6	94.3
Company	1,578	47.9	4.9
Nonprofit	477	14.5	0.8
By Funding Level			
$0	743	22.5	0.0
$1 to $250,000	747	22.7	0.5
$250,001–1,000,000	585	17.7	1.2
$1,000,001–2,500,000	609	18.5	3.0
$2,500,001 or more	612	18.6	95.3
Higher Education Institutions	1,241		
By Research Intensity			
Very high research activity (R1, Doctoral University)	129	10.4	78.3
High research activity (R2, Doctoral University)	129	10.4	10.8
All others	882	71.1	9.9
Research intensity unknown	101	8.1	1.0
By Minority-Serving Status			
Non-minority-serving	862	69.5	87.2
Minority-serving	278	22.4	11.9
Hispanic-serving	137	11.0	5.7
HBCU	68	5.5	1.7
Tribal college	20	1.6	0.3
Native Hawaiian	4	0.3	0.0

---

[10] 662 individuals and 150 government institutions were identified as subject to the T&C. However, they received a very small portion of total funding: 0.1 percent for individuals and 2.2 percent for government institutions.

Institution Characteristics	Number of Institutions	Percentage of Institutions	Percentage of Total Funding (FY 2016– FY 2020)
Alaska Native	4	0.3	0.7
Minority-serving status unknown	101	8.1	1.0
By Funding Level			
$0	202	16.3	0.0
$1 to $250,000	97	7.8	0.1
$250,001–1,000,000	234	18.9	0.5
$1,000,001–2,500,000	225	18.1	1.5
$2,500,001 or more	483	38.9	97.9

HBCU = Historically Black College or University

Note: Research intensity is from the Carnegie Classification of Institutions of Higher Education's (2018) basic classifications. HSI (Hispanic-serving institution), HBCU (Historically Black College or University), Native Hawaiian institution, Alaska Native institution, and Tribal college are subsets of MSI (minority-serving institution). HBCU and Tribal college designations were retrieved from IPEDS, MSI and HSI from Carnegie Classifications of Institutions of Higher Education (2018), and Alaska Native and Native Hawaiian from U.S. Department of Education's Office of Postsecondary Education's 2021 eligibility lists for Title III, part A grants. The overwhelming majority of institutions with Carnegie Classifications or MSI status unknown (n=101) were outside the U.S.

B. Proposal Analysis

The proposal identification analysis included all proposals in Solr that were received between Oct. 1, 2015, and March 31, 2021, that have both a summary and a description. The sample included only lead proposals (among the collaborative proposals). The anti-harassment analysis only included proposals that were identified as requesting conference funding (based on the conference proposal identification analysis). Table 2, Table 3, and Table 4 show the characteristics of proposals that request funding by fiscal year for all proposals, for funded proposals, and for not funded proposals, respectively.

TABLE 2. CHARACTERISTICS OF PROPOSALS THAT REQUEST CONFERENCE FUNDING BY FISCAL YEAR (ALL PROPOSALS)

	Before New Policy			After New Policy		
Characteristics (percentages)	**2016**	**2017**	**2018**	**2019**	**2020**	**2021***
Proposal Funding Status						
Funded	79.9	75.1	78.1	79.8	62.6	44.4
Not funded	20.1	24.9	21.9	20.2	37.4	55.6
PI Institutional Characteristics						
Research intensity						
Very high research activity institution (R1)	66.3	65.0	67.2	63.4	63.2	53.8
High research activity institution (R2)	11.0	11.5	10.8	11.7	11.5	15.8
Non-R1 and non-R2 institution	8.7	11.6	11.3	10.8	11.2	8.2
Companies, nonprofits, and IHEs outside the U.S.	14.1	12.0	10.7	14.0	14.0	22.2
Minority-serving status						
Non-minority-serving institution	91.4	87.9	90.1	88.7	87.7	90.6
Minority-serving institution	8.6	12.1	9.9	11.3	12.3	9.4
Hispanic-serving institution	3.8	5.4	3.9	5.7	5.4	7.0
HBCU	0.8	1.3	1.3	1.2	2.4	0.6
Tribal college	0.1	0.0	0.0	0.1	0.0	0.0

Characteristics (percentages)	Before New Policy			After New Policy		
	2016	2017	2018	2019	2020	2021*
Native Hawaiian institution	0.1	0.1	0.0	0.0	0.0	0.0
Alaska Native institution	0.0	0.1	0.0	0.0	0.0	0.0
PI Characteristics						
Gender						
Female	28.3	26.9	28.1	31.4	28.0	26.3
Male	61.0	60.8	57.3	49.6	51.2	53.2
Did not report	10.7	12.3	14.7	18.9	20.8	20.5
Ethnicity						
Hispanic or Latino	4.8	4.6	4.6	5.1	4.6	4.7
Not Hispanic or Latino	83.7	83.1	82.9	75.2	71.8	70.2
Did not report	11.4	12.3	12.5	19.7	23.6	25.1
Race						
Asian	15.7	16.5	17.3	15.8	15.1	17.0
Black	3.1	4.7	3.6	4.7	5.0	8.2
White	67.4	63.7	61.7	58.7	55.0	48.0
Other	1.0	2.0	1.5	1.8	1.9	3.5
Did not report	12.8	13.0	15.9	19.0	23.0	23.4
Disability status						
Disability	1.1	1.9	0.7	1.1	1.2	1.8
No disability	75.0	79.3	84.7	74.2	66.1	62.6
Did not report	23.9	18.8	14.7	24.8	32.7	35.7
Number of Proposals	1224	1211	1350	1126	840	171

IHE = institution of higher education; HBCU = Historically Black College or University; PI = principal investigator

Note: FY 2018 includes proposals received through November 1, 2018, before the new policy was announced. Minority-serving institution, Hispanic-serving institution, and Research intensity come from the Carnegie Classification of Institutions of Higher Education. Tribal college and HBCU were retrieved from IPEDS. Native Hawaiian institution and Alaska Native institution were retrieved from NSF's Solr search engine. The U.S. Department of Education's 2021 Eligibility Matrix classifies one institution as a Native Hawaiian institution that is not classified as such in NSF's Solr data.

*FY 2021 only includes the first two quarters of the fiscal year.

TABLE 3. CHARACTERISTICS OF PROPOSALS THAT REQUEST CONFERENCE FUNDING BY FISCAL YEAR (FUNDED PROPOSALS)

Characteristics (percentages)	Before New Policy			After New Policy		
	2016	2017	2018	2019	2020	2021*
Proposal Funding Status						
Funded	100.0	100.0	100.0	100.0	100.0	100.0
Not funded	0.0	0.0	0.0	0.0	0.0	0.0
PI Institutional Characteristics						
Research intensity						
Very high research activity institution (R1)	69.4	66.4	67.9	65.5	64.4	52.6
High research activity institution (R2)	10.1	11.1	10.1	10.8	11.0	18.4
Non-R1 and non-R2 institution	7.7	11.0	11.5	10.2	10.3	7.9
Companies, nonprofits, and IHEs outside the U.S.	12.8	11.4	10.5	13.5	14.3	21.1
Minority-serving status						
Non-minority-serving institution	92.0	88.1	90.7	89.2	87.5	85.5

Characteristics (percentages)	Before New Policy			After New Policy		
	2016	2017	2018	2019	2020	2021*
Minority-serving institution	8.0	11.9	9.3	10.8	12.5	14.5
Hispanic-serving institution	3.6	5.4	3.8	5.1	5.5	10.5
HBCU	0.5	1.4	0.9	1.1	2.1	1.3
Tribal college	0.1	0.0	0.0	0.1	0.0	0.0
Native Hawaiian institution	0.1	0.0	0.0	0.0	0.0	0.0
Alaska Native institution	0.0	0.1	0.0	0.0	0.0	0.0
PI Characteristics						
Gender						
Female	27.8	27.3	29.3	31.8	29.8	32.9
Male	61.6	60.6	57.7	51.2	51.0	50.0
Did not report	10.6	12.1	13.0	17.0	19.2	17.1
Ethnicity						
Hispanic or Latino	3.6	3.7	4.4	4.7	4.0	5.3
Not Hispanic or Latino	85.4	84.2	83.9	77.4	73.2	73.7
Did not report	11.0	12.1	11.8	17.9	22.8	21.1
Race						
Asian	15.2	17.4	17.7	16.8	16.0	22.4
Black	3.3	5.2	2.8	3.7	4.9	9.2
White	68.3	63.5	63.4	60.7	55.5	47.4
Other	1.0	2.0	1.2	1.6	2.3	1.3
Did not report	12.2	12.0	14.8	17.2	21.3	19.7
Disability status						
Disability	0.9	1.4	0.6	1.0	1.5	3.9
No disability	75.5	79.6	84.5	73.9	65.6	64.5
Did not report	23.6	18.9	15.0	25.1	32.9	31.6
Number of Proposals	978	909	1055	899	526	76

IHE = institution of higher education; HBCU = Historically Black College or University; PI = principal investigator

Note: FY 2018 includes proposals received through November 1, 2018, before the new policy was announced. Minority-serving institution, Hispanic-serving institution, and research intensity come from the Carnegie Classification of Institutions of Higher Education. Tribal college and HBCU were retrieved from IPEDS. Native Hawaiian institution and Alaska Native institution were retrieved from NSF's Solr search engine. The U.S. Department of Education's 2021 Eligibility Matrix classifies one institution as a Native Hawaiian institution that is not classified as such in NSF's Solr data.

*FY 2021 only includes the first two quarters of the fiscal year.

TABLE 4. CHARACTERISTICS OF PROPOSALS THAT REQUEST CONFERENCE FUNDING BY FISCAL YEAR (NOT FUNDED PROPOSALS)

Characteristics (percentages)	Before New Policy			After New Policy		
	2016	2017	2018	2019	2020	2021*
Proposal Funding Status						
Funded	0.0	0.0	0.0	0.0	0.0	0.0
Not funded	100.0	100.0	100.0	100.0	100.0	100.0
PI Institutional Characteristics						
Research intensity						
Very high research activity institution (R1)	53.7	60.6	64.7	55.1	61.1	54.7
High research activity institution (R2)	14.6	12.6	13.2	15.4	12.4	13.7

Characteristics (percentages)	Before New Policy			After New Policy		
	2016	2017	2018	2019	2020	2021*
Non-R1 and non-R2 institution	12.6	13.2	10.5	13.2	12.7	8.4
Companies, nonprofits, and IHEs outside the U.S.	19.1	13.6	11.5	16.3	13.7	23.2
Minority-serving status						
Non-minority-serving institution	89.0	87.1	88.1	86.8	88.2	94.7
Minority-serving institution	11.0	12.9	11.9	13.2	11.8	5.3
Hispanic-serving institution	4.9	5.3	4.4	7.9	5.1	4.2
HBCU	2.0	1.0	3.1	1.3	2.9	0.0
Tribal college	0.0	0.0	0.0	0.0	0.0	0.0
Native Hawaiian institution	0.0	0.3	0.0	0.0	0.0	0.0
Alaska Native institution	0.0	0.0	0.0	0.0	0.0	0.0
PI Characteristics						
Gender						
Female	30.1	25.8	23.7	30.0	24.8	21.1
Male	58.9	61.3	55.6	43.6	51.6	55.8
Did not report	11.0	12.9	20.7	26.4	23.6	23.2
Ethnicity						
Hispanic or Latino	9.8	7.3	5.4	6.6	5.7	4.2
Not Hispanic or Latino	77.2	79.8	79.3	66.5	69.4	67.4
Did not report	13.0	12.9	15.3	26.9	24.8	28.4
Race						
Asian	17.5	13.9	15.9	11.9	13.7	12.6
Black	2.4	3.3	6.1	8.8	5.1	7.4
White	63.8	64.6	55.6	50.7	54.1	48.4
Other	0.8	2.0	2.4	2.6	1.3	5.3
Did not report	15.4	16.2	20.0	26.0	25.8	26.3
Disability status						
Disability	2.0	3.3	1.0	1.3	0.6	0.0
No disability	73.2	78.1	85.4	75.3	66.9	61.1
Did not report	24.8	18.5	13.6	23.3	32.5	38.9
Number of Proposals	246	302	295	227	314	95

IHE = institution of higher education; HBCU = Historically Black College or University; PI = principal investigator

*FY 2021 only includes the first two quarters of the fiscal year.

Note: FY 2018 includes proposals received through November 1, 2018, before the new policy was announced. Minority-serving institution, Hispanic-serving institution, and research intensity come from the Carnegie Classification of Institutions of Higher Education. Tribal college and HBCU were retrieved from IPEDS. Native Hawaiian institution and Alaska Native institution were retrieved from NSF's Solr search engine. The U.S. Department of Education's 2021 Eligibility Matrix classifies one institution as a Native Hawaiian institution that is not classified as such in NSF's Solr data.

Table 5 shows differences between funded and not funded proposals for FY 2018, the baseline year before the policy was implemented. This table provides additional context for understanding how the funded and not funded proposals included in the interrupted time series analysis differ across proposal characteristics.

TABLE 5. CHARACTERISTICS OF PROPOSALS THAT REQUEST CONFERENCE FUNDING, BY PROPOSAL FUNDING STATUS (FY 2018)

Characteristics (percentages)	Funded	Not funded	Difference
PI Institutional Characteristics			
Research intensity			
Very high research activity institution (R1)	67.9	64.7	3.1*
High research activity institution (R2)	10.1	13.2	-3.1*
Non-R1 and non-R2 institution	11.5	10.5	1.0*
Companies, nonprofits, and IHEs outside the U.S.	10.5	11.5	-1.0*
Minority-serving status			
Non-minority-serving institution	90.7	88.1	2.6*
Minority-serving institution	9.3	11.9	-2.6*
Hispanic-serving institution	3.8	4.4	-0.6*
HBCU	0.9	3.1	-2.2*
Tribal college	0.0	0.0	0.0
Native Hawaiian institution	0.0	0.0	0.0
Alaska Native institution	0.0	0.0	0.0
PI Characteristics			
Gender			
Female	29.3	23.7	5.6*
Male	57.7	55.6	2.1*
Did not report	13.0	20.7	-7.7*
Ethnicity			
Hispanic or Latino	4.4	5.4	-1.1*
Not Hispanic or Latino	83.9	79.3	4.6*
Did not report	11.8	15.3	-3.5*
Race			
Asian	17.7	15.9	1.8*
Black	2.8	6.1	-3.3*
White	63.4	55.6	7.8*
Other	1.2	2.4	-1.1*
Did not report	14.8	20.0	-5.2*
Disability status			
Disability	0.6	1.0	-0.4*
No disability	84.5	85.4	-1.0*
Did not report	15.0	13.6	1.4*
Number of Proposals	1055	295	1350

IHE = institution of higher education; HBCU = Historically Black College or University; PI = principal investigator

Note: FY 2018 includes proposals received through November 1, 2018, before the new policy was announced. Minority-serving institution, Hispanic-serving institution, and research intensity come from the Carnegie Classification of Institutions of Higher Education. Tribal college and HBCU were retrieved from IPEDS. Native Hawaiian institution and Alaska Native institution were retrieved from NSF's Solr search engine. The U.S. Department of Education's 2021 Eligibility Matrix classifies one institution as a Native Hawaiian institution that is not classified as such in NSF's Solr data.

* Significantly different from zero at the .05 level, two-tailed test.

IV. Methods

A. Communication Analysis

To analyze the sample of communication materials, each artifact was manually coded along four dimensions:

1. Timing:
    a. Prior to 2018: Historical NSF communication on sexual harassment
    b. January–July 2018: Policy development and public comment
    c. August–December 2018: Term and Condition release and rollout
    d. January 2019–December 2020: Ongoing communications and updates
2. Mode/communication type:
    a. Formal policy or guidance
    b. Internal planning
    c. Press release or public statement
    d. In-person or social media live event
    e. Social media post or reference
    f. Press or publication mention
3. Primary audience:
    a. NSF staff and stakeholders (including panelists and grantees)
    b. Broader scientific community
4. Policy focus/content:[11]
    a. Need for updated/explicit anti-harassment policy at NSF
    b. Term and Condition
    c. Conference policy
    d. Broader anti-harassment efforts

Based on final coding, conducted descriptive analysis was conducted, including cross-tabulations based on timing and primary audience (see Table 6).

To supplement the analyses of communication materials described above, information provided during the interviews was synthesized to describe perceptions of what was most effective for communicating policy internally to NSF and externally to the research field and other stakeholders.

---

[11] Note that this is the only dimension where the categories are not mutually exclusive; one artifact could be coded as including a policy focus in multiple categories.

TABLE 6. COUNTS AND CHARACTERISTICS OF COMMUNICATION ARTIFACTS

Characteristic	Prior to 2018 NSF Audience	Broader Community	Total (pct.)	Jan.–Jul. 2018 NSF Audience	Broader Community	Total (pct.)	Aug.–Dec. 2018 NSF Audience	Broader Community	Total (pct.)	Jan. 2019–Dec. 2020 NSF Audience	Broader Community	Total (pct.)	All Time Periods NSF Audience	Broader Community	Total (pct.)
Primary Audience (percentage)	**2** (40%)	**3** (60%)	**5**	**14** (56%)	**11** (44%)	**25**	**50** (70%)	**21** (30%)	**71**	**2** (22%)	**7** (78%)	**9**	**68** (62%)	**42** (38%)	**110**
Mode															
Formal policy or guidance	1	--	**1** (20%)	1	2	**3** (12%)	--	3	**3** (4%)	1	2	**3** (33%)	3	7	**10** (9%)
Internal planning	--	--	**--**	6	--	**6** (24%)	27	--	**27** (38%)	1	--	**1** (11%)	34	--	**34** (31%)
Press release or public statement	--	2	**2** (40%)	1	5	**6** (24%)	1	3	**4** (6%)	--	3	**3** (33%)	2	13	**15** (14%)
In-person or social media live event	--	--	**--**	1	1	**2** (8%)	4	3	**7** (10%)	--	--	**--**	5	4	**9** (8%)
Social media post or reference	--	--	**--**	2	--	**2** (8%)	9	5	**14** (20%)	--	--	**--**	11	5	**16** (15%)
Press or publication mention	1	1	**2** (40%)	3	3	**6** (24%)	9	7	**16** (23%)	--	2	**2** (22%)	13	13	**26** (24%)
Content															
Need			**3** (60%)			**5** (20%)			**8** (11%)			**3** (33%)			19 (17%)
T&C			**--**			**9** (36%)			**47** (66%)			**2** (22%)			58 (53%)
Conference policy			**--**			**--**			**--**			**4** (44%)			4 (4%)
Broader efforts			**2** (40%)			**12** (48%)			**17** (24%)			**2** (22%)			33 (30%)

Note: The content categories are not mutually exclusive and will sum to more than 110 artifacts or greater than 100 percent.

B. Term and Condition Analysis

1. Identify Links to the T&C Using Ahrefs

For the Ahrefs analysis, the analysis sample was the *Ahrefs analysis dataset* discussed previously in the Data section (II.B). This is an institution-level file that contains the 3,297 higher education institutions, companies, and nonprofits subject to the T&C and includes each institution's characteristics and an indicator for whether the institution linked to NSF's T&C-related webpages.

For this analysis, the share of institutions subject to the T&C that link to NSF's T&C-related webpages was reported in aggregate and then separately by the following types of institutions: higher education institutions, companies, or nonprofits. For the higher education institutions, the findings were also disaggregated by institutions' direct award funding, research intensity, and minority-serving status.

2. Analyze Webpage Content

(i) Step 1: Create a codebook to identify key constructs to search for on institutions' webpages

A codebook was developed for the webpage content analysis which describes the key constructs (called codes) the study is aiming to identify when reviewing webpages (see Table 7). The codebook was used to manually code webpages for a sample of institutions. These manually coded webpages were then used to train a supervised Natural Language Processing (NLP) algorithm. Codes 1a and 1b align with the following research question: Have institutions funded by NSF established their own anti-harassment policies and publicized these via their website? Code 2 aligns with the following research question: Do NSF-funded institutions reference the NSF's anti-harassment T&C on their websites? Code 3 is a more specific version of code 1a, which identifies whether an institution has a policy specifically against harassment at conferences or convenings. (Note: Code 3 was included in model development but later dropped it because the number of positive cases (webpages that would be coded "yes" for code 3) identified was insufficient to train the model on. This is discussed further in Step 3 below).

TABLE 7. CODEBOOK FOR WEBPAGE CLASSIFICATION

Code	Definition
1a. Institution has a policy against harassment.	A statement of an institution's policy regarding sexual harassment or other harassment. The policy must apply to the institution as a whole, rather than a subset of it, such as a policy specific to a department or a single research center within an institution.
1b. Institution has a detailed policy, including (1) a clear definition of harassment or (2) policies and procedures for reporting.	A statement of an institution's policy regarding sexual harassment or other harassment, which includes (1) a definition of harassment or (2) policies and procedures for handling reported cases, such as reporting mechanisms. The policy must apply to the institution as a whole, rather than a subset of it, such as a policy specific to a department or research center.
2. Institution references NSF's harassment policies.	A reference or link to NSF's policies against harassment with the goal of disseminating information to an institution's community about NSF's harassment policies.
3. Institution has a conference policy against sexual harassment.	A statement of an institution's policies or code of conduct regarding sexual harassment, other harassment, and sexual assault at conferences or convenings.

(ii) Step 2: Test and identify web search parameters to use

To determine which web search approach to use, the performance of different combinations of parameters was tested. Specifically, the following parameters were varied:

1. Search terms: Three different search terms were used to capture institutions' policies against sexual harassment or other harassment:
   - Harassment
   - Harassment sexual assault
   - Harassment OR (sexual assault)

Each of these three search terms were also tested with the following addition: *(NSF OR "National Science Foundation").* For example, one search term set was *harassment AND (NSF OR "National Science Foundation").*

2. Within domain or not: The web search was either limited to webpages within an institution's domain or, as an alternative, the web search was not restricted to an institution's domain but did include the institution name in the search terms. For example, if one search approach was *harassment site:georgetown.edu,* then the alternative search would have been *harassment "Georgetown University."*

To test the different combinations of search parameters, a random sample of institutions (n=19) was selected.[12] The sample was stratified by institution category: companies or nonprofits, very high intensity research institutions, high intensity research institutions, minority-serving postsecondary institutions, and all other postsecondary institutions.

After running all search approach combinations for all institutions in the sample and collecting the top 20 webpages results of each search, approximately 670 unique webpages were collected and manually coded. For each web search approach combination and each code, the percent of institutions identified via manual coding was calculated as having a "yes" for a given code for which the web search approach returned at least one "yes." This performance measure is a modified version of "recall" (or sensitivity) calculated at the institution level. Recall was emphasized over other performance measures due to the interest in ensuring the web search would not miss a relevant webpage. To determine how many webpages would need to be collected to not miss relevant "yeses," institution-level recall was calculated, varying the number of webpage results from each search. Two top-performing web search approaches were identified (see Table 8). The first was targeted at identifying webpages relevant for code 1a and 1b and had 97 percent recall at the institution level for code 1a and 90 percent for code 1b. The second was targeted at code 2 and had 87 percent recall at the institution level. For code 3, none of the manually coded webpages were a "yes," so it was not possible to calculate recall. After choosing these two web search approaches, these searches were conducted for all institutions subject to the T&C and collected text from all of the webpages.

TABLE 8. OPTIMAL WEB SEARCH APPROACHES

Number	Search Term	Within Domain or Not	Number of Webpage Results
1	Harassment	Within domain	12
2	Harassment AND (NSF OR "National Science Foundation")	Within domain	5

(iii) Step 3: Develop and test Natural Language Processing model
To develop the NLP model, a sample of 50 institutions was selected and institutions were split into a training and validation set (40 institutions) and a test set (10 institutions) to evaluate the final model of performance.[13] This yielded

---

[12] The initial analysis included 15 institutions but 4 more were added to ensure that at least two institutions from the following minority-serving institution categories were included: HBCU, Tribal college, and Hispanic-serving institutions. A relatively small number of institutions were chosen to pilot test the web search approaches because the number of webpages that would need to be manually coded to test all combinations of the web search approaches for an institution is fairly large high (On average, collected 35 *unique* webpages were collected for each institution in the test sample.) The web searches that had optimal performance in the pilot test sample were consistently the best performing across institutions. This implied that there did not appear to be a lot of variation in web search performance across institutions, so expanding the pilot sample to include additional institutions was unlikely to change the web search approach that would ultimately be chosen as having optimal performance.
[13] The analysis included the set of 19 institutions that were used for the web search optimization within this set of 50 because they had already been manually coded. Because these institutions were part of the web search optimization process, which tested multiple web search approaches, there were more manually coded webpages for these institutions than for the 31 new

a total of 549 webpages in the training and validation sets and 136 webpages in the test set.[14] As with the search parameter optimization, the sample was stratified to ensure that it included a proportional number of companies or nonprofits, very high intensity research institutions, high-intensity research institutions, minority-serving institutions, and all others.

Using the manually-coded webpages in the training and validation set, two separate transformer-based deep learning models were trained, one for codes 1a and 1b and another for code 2, at the webpage level to predict whether a page was relevant. The model for codes 1a and 1b was a multi-label classifier, while the code 2 model was a single-label classifier.[15]

As part of the development process for the 1a/1b model, additional webpages were manually coded to improve performance. A random sample of 29 additional institutions was selected and the first three webpage results were manually coded for the 1a/1b-focused web search (Search 1 in Table 8). The institutions were then randomized into the training, validation, and test sets, and the model was re-trained incorporating the new content.[16] In total for the 1a/1b model, there were 1,143 webpages from 46 institutions in the training and validation sets and 360 webpages from 33 institutions in the test set.

For code 2, there were initially too few "yes" cases in the webpages collected for the 50 institutions to adequately train the model. To supplement the "yes" cases, 133 additional webpages were manually coded from 66 institutions that were identified by Ahrefs as linking to NSF's T&C-related harassment webpages (100 of these webpages were "yes" for code 2). Institutions from these pages were split into the training, validation, and test sets and the model was retrained to incorporate the additional data.

When developing the model for code 2, the definition of the code was also expanded. Previously, the webpage was required to reference to NSF's T&C to be a "yes" for code 2. However, there were many webpages that referenced NSF's harassment policies, and the model was struggling to distinguish these cases from T&C-specific cases. As a result, the definition was broadened for a "yes" on code 2 to be a reference to any of NSF's harassment policies.[17] (To identify T&C-specific references, a word token search[18] of all "yes" cases the algorithm identified was used. This

---

institutions that were manually coded for model development. For institutions in the training and validation sets, the analysis used any additional webpages that had been coded as part of the web search optimization process from web searches that were within the institutions' domain to help with model development. But, for all institutions in the test set, the analysis only included webpages that came from the two optimal searches. This was intended to make sure performance was tested on the exact set of searches that would be used to scale up for all institutions.

[14] For the second web search approach listed in Table 7, the Google searches sometimes yielded fewer than 5 relevant webpages.

[15] A multi-label classifier allows each input to have multiple labels assigned to it, non-exclusively. For example, a webpage fed into our multi-label model can be assigned code1a, 1b, 1a and 1b, or neither. A single-label classifier focuses on only one label at a time, although this single label does not have to be binary, as in our problem. The single-label classifier described here determines whether a webpage should be assigned code 2.

[16] The analysis used the top three web search results because most institutions have at least one relevant webpage for codes 1a and 1b within the top three search results and this focused resources on bringing in representative content from more institutions, rather than collecting more "yeses" for a smaller set of institutions.

[17] Specifically, the code was the following: A reference or link to NSF's new T&C. Does not need to include the phrase "term and condition," but must include references to the main provision of the T&C (that is, notifying NSF of "any findings/determinations of sexual harassment, other forms of harassment, or sexual assault regarding an NSF funded Principal Investigator (PI) or co-PI, or of the placement of the PI or co-PI on administrative leave, or the imposition of any administrative action relating to harassment or sexual assault finding or investigation").

[18] By splitting the text into individual word tokens, the phrases of interest could be more easily identified without having to account for differences in whitespace or word boundaries that may lead to misses in a simple text search.

is discussed in greater detail below.) All of the manually coded webpages were recoded and the model was retrained with this new code 2 definition.

Another refinement to the model for code 2 involved imposing a restriction that a webpage must have the text "NSF" or "National Science Foundation" to count as a relevant "yes" webpage. This filtering helped improve performance slightly.

For code 3, there were zero "yes" cases in the 50 manually coded institutions' webpages. There were two webpages for specific conferences that mentioned anti-harassment policies and at least five webpages mentioning NSF's new conference policy. No institutions were found that specifically mentioned their harassment policies applied to conferences. As a result, no NLP model was trained to identify code 3.

After developing the code 1a/1b model and the code 2 model, the test set was used to evaluate the final performance before applying the model to the full set of institutions (see Table 9, Table 10). Performance was reported at the webpage level and the institution level.

For model 1a/1b, performance was primarily evaluated at the institution level. This is because institutions typically have 1–3 main webpages that are focused on describing their harassment policy but have many other webpages in their domain that mention their harassment policies without focusing on them. (For example, a webpage discussing bystander intervention strategies may reference the institutions' harassment policy in passing.) These cases were much harder for the model to accurately classify, leading to a lower recall rate at the webpage level. However, because the main goal of this project was to identify whether institutions have a harassment policy on their webpages or not, performance at the institution level—where the model performed substantially better—was given primary consideration.

At the institution level, the model achieved balanced accuracy of 78 percent for code 1a and 88 percent for code 1b; recall of 84 percent for code 1a and 94 percent for code 1b; and specificity of 71 percent for code 1a and 81 percent for code 1b (see Table 8).

TABLE 9. ALGORITHM PERFORMANCE FOR CODE 1A/1B

| Performance Metric | Institution Level | | Webpage Level | |
	Code 1a: Institution has policy against harassment	Code 1b: Institution has detailed policy against harassment	Code 1a: Institution has policy against harassment	Code 1b: Institution has detailed policy against harassment
Balanced Accuracy	78%	88%	72%	71%
Sensitivity/Recall	84%	94%	67%	55%
Specificity	71%	81%	77%	86%
Number of Institutions or Webpages	33	33	360	360

Note: This table includes the random sample of institutions (and their webpages) included in the test set for code 1a/1b. Balanced accuracy is the arithmetic mean of sensitivity and specificity; sensitivity is the percentage of "yes" cases for a given code that were identified as a "yes" by the algorithm; and specificity is the percentage of "no" cases for a code that were correctly identified as being a "no" by the algorithm.

For code 2, institutions typically did not reference NSF's harassment policies on more than 1-2 webpages, so there was less reason to emphasize institution versus webpage level performance. For code 2, the model achieved balanced accuracy of 83 percent, recall of 82 percent, and specificity of 83 percent at the institution level. At the

webpage level, the model achieved 87 percent balanced accuracy, 76 percent recall, and 97 percent specificity (Table 10).[19]

TABLE 10. ALGORITHM PERFORMANCE FOR CODE 2 (INSTITUTION REFERENCES NSF'S ANTI-HARASSMENT POLICIES)

Performance Metric	Institution Level	Webpage Level
Balanced Accuracy	83%	87%
Sensitivity/Recall	82%	76%
Specificity	83%	97%
Number of Institutions or Webpages	23	331

Note: This table includes the random sample of institutions (and their webpages) included in the test set for code 2. Balanced accuracy is the arithmetic mean of sensitivity and specificity; sensitivity is the percentage of "yes" cases for a given code that were identified as a "yes" by the algorithm; and specificity is the percentage of "no" cases for a code that were correctly identified as being a "no" by the algorithm.

After testing the algorithm, the two optimal web searches were run for all the remaining institutions and applied the algorithms for code 1a and 1b and for code 2 to the text collected from the webpages. For code 2, the text of all pages identified as "yes" for code 2 was also searched for the following phrases, separated into individual word tokens: "term and condition" (not case sensitive) and "T&C" (case sensitive). This was intended to provide an estimate of how many NSF harassment references were explicitly references to the new term and condition. However, these results should be viewed as a lower bound estimate. Some institutions' webpages reference the T&C in less direct language (e.g., "National Science Foundation announces new policy related to reporting harassment"), and these would not be captured by the word token searches as T&C references.

(iv) Step 4: Conduct descriptive analysis

The analysis dataset was assembled using the *institution dataset* previously created and adding in the indicator variables derived from the NLP algorithms that indicate if an institution has at least one webpage coded as a "yes" for code 1a, code 1b, and code 2. An indicator was also created for whether an institution had a code 2 reference that was explicitly mentioning the T&C, based on the word token searches. The share of institutions with at least one webpage coded "yes" for code 1a, code 1b, code 2, and a subset of code 2 that contains a T&C reference was reported. This information was also separately reported by institution type (higher education, company, nonprofit) and institution characteristics (research intensity, minority-serving status, and funding level).

C. Proposal Analysis

1. Step 1: Identifying Proposals with Conference Funding

The proposal analysis began with a manual review of a random sample of proposals and identification of whether each proposal requested funding for a convening. Convenings include conferences, workshops, symposia, meetings, and summits. Both online and in-person events are included. Training courses, summer schools, and professional development programs are excluded. A proposal was considered to request funding for a convening if it requested funding for conference organizers or speakers, conference facilities, conference materials or supplies, travel to a conference, or meals for attendees (as discussed in budget justification sections in proposals).

---

[19] Performance measures depend on the complexity of the classification task. In this case, because all of the webpages collected through the web search were related to NSF and "harassment", it was challenging to train a model to differentiate when those terms were referencing an NSF anti-harassment policy or something related but different.

A conference indicator was created that categorized a proposal as requesting conference funding if it contained particular search terms in the title or if it had an object class code of 4160. Conference search terms included "conference," "workshop," "symposium," "meeting," and "summit," as well as their respective plurals, and the title could not contain the term "summer."

The performance of this indicator was tested on the test set. This indicator achieved a balanced accuracy rate of 99 percent (see Table 11).

TABLE 11. PROPOSAL CATEGORIZATION ALGORITHM PERFORMANCE

Performance Metric	Value
Balanced Accuracy	99%
Sensitivity/Recall	100%
Specificity	99%
Number of Proposals	127

Note: This table includes the random sample of proposals included in the test set. Balanced accuracy is the arithmetic mean of sensitivity and specificity; sensitivity is the percentage of proposals in that category correctly identified as being in that category; and specificity is the percentage of proposals not in that category that were correctly identified as not being in that category. Performance is determined by comparing the algorithm to manual coding.

2. Step 2: Identifying References to Anti-Harassment Policies and Practices

To measure whether the proposals that request conference funding include references to anti-harassment policies and practices, the analysis used a text mining algorithm. The analysis began with the proposals that were identified as requesting conference funding in Step 1. Of proposals that requested conference funding, a random subset that included anti-harassment keywords (including "harassment" and "anti-discrimination") was manually reviewed and coded. Unigram and bigram analyses were then conducted to determine the types of terms that are most likely to occur in proposals that contain anti-harassment references. Based on this analysis, an initial indicator was created that contained the unigrams and bigrams that were most likely to be found in proposals that referenced anti-harassment and not proposals that did not reference anti-harassment. This indicator was tested on the validation set and found several false positives. The algorithm was refined to reduced the number of false positives found in the validation set.

The final indicator identified proposals where their description contained: (1) "policy/policies," "harassment-free," or variations of "free of harassment" in the same sentence as the key anti-harassment terms (terms containing the "harass" stem or "anti-discrimination"); or (2) "code of conduct" within four sentences of key anti-harassment terms.

The performance of this final indicator was tested on a final hold-out test set (see Table 12). This approach attained a final balanced accuracy of 94 percent. The indicator correctly identified 96 percent of proposals that contained anti-harassment references (sensitivity/recall) and correctly identified 93 percent of proposals that did not contain anti-harassment references (specificity).

Each subgroup achieved a balanced accuracy of at least 80 percent. There is no established threshold in the literature for a minimum acceptable accuracy levels, and the range of accuracy levels often vary by task. However, the overall accuracy level of over 90 percent provides high confidence that the algorithm is highly aligned with the intended measure. Each subgroup of interest was reviewed to ensure at least 10 observations in the validation and test set. For subgroups of interest with too few observations to include at least 10 observations in both the validation and test sets, every proposal in which a key anti-harassment keyword occurred was manually coded to ensure balanced accuracy rates of 100 percent. This included HBCU, Tribal college, Native Hawaiian institution, Alaska Native institution, PIs of other races, and PIs with reported disabilities.

TABLE 12. ANTI-HARASSMENT ALGORITHM PERFORMANCE

Characteristics (percentages)	Balanced Accuracy	Sensitivity/ Recall	Specificity
All Proposals	94	96	93
Proposal Funding Status			
Funded	96	96	95
Not funded	92	95	88
PI Institutional Characteristics			
Research intensity			
Very high research activity institution (R1)	97	96	97
High research activity institution (R2)	83	75	91
Non-R1 and non-R2 institution	92	100	83
Companies, nonprofits, and IHEs outside the U.S.	88	100	75
Minority-serving status			
Non-minority-serving institution	94	97	91
Minority-serving institution	94	88	100
Non-Hispanic-serving institution	95	97	92
Hispanic-serving institution	90	80	100
Non-HBCU	94	96	93
HBCU	n.a.	n.a.	n.a.
Tribal college	n.a.	n.a.	n.a.
Native Hawaiian institution	n.a.	n.a.	n.a.
Alaska Native institution	n.a.	n.a.	n.a.
PI Characteristics			
Gender			
Female	91	95	88
Male	97	94	100
Did not report	94	100	88
Ethnicity			
Hispanic or Latino	90	100	80
Not Hispanic or Latino	94	93	96
Did not report	92	100	83
Race			
Asian	100	100	100
Black	88	100	75
White	93	93	93
Other	n.a.	n.a.	n.a.
Did not report	95	100	90
Disability status			
Disability	n.a.	n.a.	n.a.
No disability	93	94	92
Did not report	97	100	94
Number of Proposals	128	128	128

Source: NSF's Solr database

Note: This table includes the random sample of proposals included in the anti-harassment test set. Balanced accuracy is the arithmetic mean of sensitivity and specificity; sensitivity is the percentage of proposals that refer to anti-harassment policies that were correctly identified as such; and specificity is the percentage of proposals that do not refer to anti-harassment policies that were correctly identified as such.

Performance is determined by comparing the algorithm to manual coding. There were too few proposals from HBCUs, Tribal colleges, Native Hawaiian institutions, Alaska Native institutions, PIs with disabilities, and PIs in the other race group to estimate an accuracy rate. For these proposals, every proposal that contained a key anti-harassment keyword was manually coded to ensure an accuracy rate of 100%.

3. Step 3: Interrupted Time Series Model

An interrupted time series model was used to estimate the extent to which proposals were more likely to include anti-harassment references after the new policy. An analytic dataset was created that only included proposals identified as requesting funding for a convening (based on Step 1) and including the anti-harassment indicator created in Step 2. The average characteristics of the proposals in the sample were estimated at the annual level, quarterly level, and monthly level (in three separate analytic datasets). The main model estimated the change in anti-harassment policy after the announcement date and after the effective dates:

$$Y_t = \beta_0 + \beta_1 AfterAnn_t + \beta_2 AfterEff_t + e_t$$

The Y_t represents the proportion of proposals in period t that refer to an anti-harassment policy, $AfterAnn_t$ represents whether period t is after the policy announcement date (Nov. 2, 2018), and $AfterEff_t$ represents whether period t is after the policy effective date (of Feb. 25, 2019). The sum of β_1 and β_2 represents the change in the percent of proposals that reference an anti-harassment policy after the new policy. There were also versions of this model that include proposal characteristics as covariates.

A secondary model was estimated that included a time trend, to account for any increases in references in anti-harassment that may have already been occurring before the new policy:

$$Y_t = \beta_0 + \beta_1 t + \beta_2 AfterAnn_t + \beta_3 AfterEff_t + \beta_4(t * AfterEff_t) + e_t$$

The t indicates the number of periods after the policy announcement date (periods before the policy are negative).

4. Sensitivity Analysis for Principal Investigators

A sensitivity analysis was also performed that only included PIs who submitted at least one proposal before and at least one proposal after the implementation of NSF's new policy. The advantage of this model is that it removes any concerns about unobservable characteristics of PIs changing over time, but this subset of PIs may not generalize to all PIs.

An analytic dataset was created at the PI-level, which includes the percent of proposals the PI submitted with anti-harassment references before the policy and after the policy. Only proposals submitted before the policy was announced (for the pre-period, before Nov. 2, 2018) and after the policy was implemented (for the post-period, after Feb. 25, 2019) were included. The following simplified model was used for this analysis:

$$Y_t = \beta_0 + \beta_1 AfterPolicy_t + e_t$$

The β_1 represents the change in the percent of anti-harassment references after the new policy relative to before.

V. Supplementary Findings

A. Proposal Analysis

1. Interrupted Time Series Results

Table 13 shows the results of the interrupted time series models at the annual, quarterly, and monthly level, with and without time trends and with and without covariates. The estimates vary between 12 and 18 percentage points, with most estimates between 16 and 18 percentage points. Table 14 shows the estimates by subgroup (for the quarterly model without a time trend or covariates). The quarterly model with a time trend and without covariates was chosen as the "main model" for a few reasons. The annual model included few observations (7) and thus included less detailed information than the quarterly model. The monthly model included many more observations (66) but many of the months included very few proposals and thus the noise was higher in this model. The quarterly model thus balanced these two considerations (enough data points and minimal noise in the data). The time trend was included in the main model so that the main coefficient reflects the difference at the time of the policy. There is a slight downward trend in anti-harassment references during the COVID-19 pandemic (although this was not statistically significant), so including the time trend means that the main coefficient represents the change at the time of the policy (instead of an average over the full post-period from FY 2019 to 2021). Lastly, including covariates in the models did not improve precision given the relatively few degrees of freedom in these models so the model without covariates is preferred.

TABLE 13. CHANGE IN PERCENTAGE OF PROPOSALS WITH REFERENCES TO ANTI-HARASSMENT POLICIES AFTER NEW POLICY

Model	Annual	Annual	Quarterly	Quarterly	Monthly	Monthly
With Time Trend	17.87* (2.42)	n.a.	17.98* (2.79)	16.9 (44.66)	17.26* (2.35)	17.32* (2.53)
Without Time Trend	15.81* (1.94)	X	17.11* (1.35)	12.48 (7.18)	16.83* (1.04)	18.29* (1.63)
Covariates	No	Yes	No	Yes	No	Yes
Number of Observations	7	X	24	24	66	66

Note: This table includes coefficients and standard errors from the interrupted time series model. Each column represents a separate regression. Standard errors are in parentheses. The annual model with covariates could not be estimated due to an insufficient number of degrees of freedom in the model. The change is the combination of the change after the announcement date and the effective date. Each observation represents a time period.

* Significantly different from zero at the .05 level, two-tailed test.

TABLE 14. PERCENTAGE OF PROPOSALS WITH REFERENCES TO ANTI-HARASSMENT POLICIES BEFORE AND AFTER NEW POLICY, AND THE CHANGE, BY PROPOSAL CHARACTERISTICS

Characteristics	Before Policy	After Policy	Change
All Proposals	0.82	17.93	17.11* (1.35)
Proposal Funding Status			
Funded	0.94	18.82	17.88* (1.8)
Not funded	0.37	16.26	15.9* (0.67)
PI Institutional Characteristics			
Research intensity			
Very high research activity institution (R1)	0.87	19.07	18.2* (1.94)
High research activity institution (R2)	0.00	17.36	17.36* (2.33)
Non-R1 and non-R2 institution	0.28	14.31	14.02* (2.66)
Companies, nonprofits, and IHEs outside the U.S.	1.39	14.5	13.11* (1.56)

Characteristics	Before Policy	After Policy	Change
Minority-serving status			
Non-minority-serving institution	0.88	18.87	17.99* (1.51)
Minority-serving institution	0.25	10.34	10.09* (2.64)
Non-Hispanic-serving institution	0.85	18.06	17.21* (1.35)
Hispanic-serving institution	0.00	16.30	16.3* (3.95)
Non-HBCU	0.83	18.29	17.46* (1.44)
HBCU	0.00	1.56	1.56 (1.26)
Tribal college	0.00	0.00	0.00 (n.a.)
Native Hawaiian institution	0.00	0.00	0.00 (n.a.)
Alaska Native institution	0.00	0.00	0.00 (n.a.)
PI Characteristics			
Gender			
Female	0.46	20.12	19.66* (1.68)
Male	0.98	17.26	16.28* (1.64)
Did not report	0.88	16.89	16.01* (1.78)
Ethnicity			
Hispanic or Latino	0	24.93	24.93* (3.91)
Not Hispanic or Latino	0.77	17.54	16.77* (1.35)
Did not report	1.34	18.21	16.87* (1.89)
Race			
Asian	1.07	11.35	10.28* (1.69)
Black	2.1	12.02	9.92* (4.75)
White	0.64	20.32	19.67* (1.21)
Other	0	14.06	14.06* (6.01)
Did not report	0.95	19.0	18.04* (2.25)
Disability status			
Disability	0	6.25	6.25 (4.9)
No disability	0.84	18.56	17.72* (1.2)
Did not report	0.66	17.46	16.8* (2.01)
Number of Proposals	4110	1812	5922

Source: NSF's Solr database

HBCU = Historically Black College or University; PI = principal investigator; R1 institution = an institution classified as having "very high research activity" in the Carnegie Classification of Institutions of Higher Education

Note: The change column includes coefficients and standard errors from the interrupted time series model. Each cell represents a separate regression, and the coefficients represent the estimated change in the percentage of proposals with references to anti-harassment policies that occurred after the implementation of NSF's new policy for that proposal characteristic (e.g., funded proposals). Standard errors will be in parentheses. This comes from the quarterly model and includes no covariates or pre/post time trends. The change is the combination of the change after the announcement date and the effective date.

* Significantly different from zero at the .05 level, two-tailed test

Table 15 shows that PIs who submitted proposals before and after the policy change increased their anti-harassment references by 13.1 percentage points after the new policy became effective. These results suggest that PIs who submitted prior proposals have changed the information they include in their proposals in response to the policy. This is a somewhat lower estimate than the estimate from the main model, which suggests that some of the increase in the main model may also be from new PIs submitting proposals with anti-harassment references.

TABLE 15. CHANGE IN PERCENTAGE OF PROPOSALS WITH REFERENCES TO ANTI-HARASSMENT POLICIES AFTER NEW POLICY, FOR PIS WHO SUBMITTED AT LEAST ONE PROPOSAL BEFORE AND AT LEAST ONE PROPOSAL AFTER THE POLICY CHANGE

Variable	Value
Change	13.05* (1.97)
Number of Principal Investigators	346
Number of Observations	692

Note: This table includes the coefficient and standard error from a model that only includes the subset of principal investigators who submitted at least one proposal before and one proposal after the policy change (346 principal investigators). The coefficient represents the change between the average percent of references after the change relative to before the change. Standard errors are in parentheses. This analysis was not conducted at the annual, quarterly, or monthly level because very few PIs would have submitted a conference proposal in each one of those periods (e.g., submitting a proposal each quarter during the time period). The intercept for this model was 17.82 (standard error 1.4), indicating that approximately 2 percent of PIs referenced anti-harassment before the policy.

* Significantly different from zero at the .05 level, two-tailed test