Fundamental errors of data collection & validation undermine claims of 'Ideological Intensification' made by the National Association of Scholars

Emilio M. Bruna[1,2]

[1] Department of Wildlife Ecology and Conservation, University of Florida, PO Box 110430, Gainesville, FL 32611-0430, USA

[2] Center for Latin American Studies, University of Florida, PO Box 115530, Gainesville, FL 32611-5530, USA

Author Note

Address for correspondence: University of Florida, WEC, PO Box 110430, Gainesville, FL 32611-0430, USA (embruna@ufl.edu)

The code and data used in this analysis are available for download and improvement at https://github.com/embruna/quantdei_nas.

Introduction

Efforts to advance Diversity, Equity, and Inclusion (hereafter, DEI) at universities in the United States have emerged as another contentious issue in an increasingly polarized political climate (*1*). While individuals and organizations critical of DEI often claim that these programs have become increasingly pervasive and ideological, this assertion is rarely supported with empirical evidence.

The National Association of Scholars (NAS) recently published a report by Mason Goad and Bruce R. Chartwell (*2*) which they claim is "the largest quantitative study of the growth of DEI-related language in the sciences" published to date. Goad and Chartwell searched university web pages and Twitter accounts, data bases of funding agencies, and repositories of the scientific literature to quantify changes over time in the use of DEI-related terminology (e.g., "advocacy", "ally", "diversity", "equity", "justice", "privilege", "race"). They conclude that the use of these terms in university communications and the scientific literature has increased dramatically since 2010, and that this is unambiguous empirical evidence of "ideological intensification" in the academic and scientific arenas. Goad and Chartwell conclude by encouraging others to use their data-mining tools and database in similar studies, because if the trends they observed continue "the future of STEM, along with the rest of the academy, is almost certainly imperiled" ((*2*), p. 47). Since its release in December 2022, the NAS report has been widely hailed and distributed by prominent DEI critics such as Jordan Peterson, Colin Wright, and Christopher Rufo (*3–5*).

Goad and Chartwell made the laudable decision to post the code used to harvest and process the data used in their report (*6*), along with the data with which they graphed the trends on which they base their conclusions (*7*), "so that other analysts can scrutinize the methods and replicate them" ((*2*), p. 48). Having done so, it is obvious that they failed to carry out even the most rudimentary data validation procedures: their "clean" data sets are riddled with thousands of records unrelated to their study as well as duplicates that vastly inflate their sample sizes. Notable examples include more than 5000 tweets about sporting

events ("race"), members of the state or federal bench ("justice"), banking ("equity"), and other non-DEI topics, hundreds of grants, publications, and tweets about ecological research ("diversity"), thousands of duplicated grants, and articles published in journals such as journals the Annual Review of Sthat were treated as publications in "STEM journals" (see *Appendix*).

Because Goad and Chartwell's conclusions were based entirely on visualizations of these data, the National Association of Scholars should retract their report immediately. Failure to do so will undermine both the institutional integrity and intellectual rigor that they claim to espouse as fundamental principles.

References

1. The Plan to Dismantle DEI. *The Chronicle of Higher Education* (2023), (available at https://www.chronicle.com/article/the-plan-to-dismantle-dei).

2. M. Goad, B. R. Chartwell, "Ideological intensification: A quantitative study of diversity, equity, and inclusion in STEM subjects at american universities." (National Association of Scholars, New York, 2022), p. 50.

3. J. Peterson, [@Jordanbpeterson] retweet of @NASorg: Diversity, Equity, and Inclusion (DEI) ideology undermines the norms of open discourse, objectivity, devotion to evidence, and intellectual independence...[Tweet] (2022), (available at https://twitter.com/NASorg/status/1600157129167577088).

4. C. Rufo, [@Realchrisrufo]: Retweet of @GoadMason: Visualizing the rise of DEI in Florida's Universities {Tweet]. *Twitter* (2023), (available at https://twitter.com/GoadMason/status/1623804389252030468).

5. M. Goad, The Ideological Intensification of DEI in STEM (2023), (available at https://www.realityslaststand.com/p/the-ideological-intensification-of).

6. National Association of Scholars, Quantitative study of diversity, equity and inclusion in STEM subjects in US universities (2022), doi:10.5281/zenodo.6360904.

7. National Association of Scholars, Data for: Quantitative Study of Diversity, Equity and Inclusion in STEM Subjects in US Universities (2022), doi:10.5281/zenodo.6360904.

8. R Core Team, "R: A language and environment for statistical computing" (manual, Vienna, Austria, 2020), (available at https://www.R-project.org/).

9. H. Wickham, M. Averick, J. Bryan, W. Chang, L. D. McGowan, R. François, G. Grolemund, A. Hayes, L. Henry, J. Hester, M. Kuhn, T. L. Pedersen, E. Miller, S. M. Bache, K. Müller, J. Ooms, D. Robinson, D. P. Seidel, V. Spinu, K. Takahashi, D. Vaughan, C. Wilke, K. Woo, H. Yutani, Welcome to the tidyverse. *Journal of Open Source Software*. **4**, 1686 (2019).

10. S. Firke, "Janitor: Simple tools for examining and cleaning dirty data" (manual, 2021), (available at https://CRAN.R-project.org/package=janitor).

11. BrunaLab/quantdei_nas, (available at https://github.com/BrunaLab/quantdei_nas).

Appendix A

Data Review and Validation

Below I present a brief overview of the methods used to review the contents of 5 data sets used by Goad and Chartwell to visualize trends in DEI-language use. These data sets can be found in the `'out/twitter'`,`'out/grants'`, and `'out/scholarship'` folders of the NAS Report's Github repository (*5*).

1. University Twitter accounts: `tweets_clean.csv`
2. National Science Foundation (i.e., NSF) grants: `nsf_all_grants_summary_data.csv`
3. National Institutes of Health (i.e., NIH) grants: `nih_parsed_all.fst`
4. Scientific publications indexed in Google Scholar: `google_scholar.fst`
5. Scientific publications indexed in PubMed: `pubmed.fst`

Although many of these errors are immediately evident by simply scanning their data sets, I also reviewed their data with code written in the R statistical programming language (*8*). This code, which included functions from the `tidyverse` (*9*) and `janitor` (*10*) libraries for filtering, de-duplicating, and summarizing data frames, is available at (*11*) along with `.csv` files of the results. Below I provide summaries and representative examples of the errors revealed by the validation procedures . It is important to emphasize that the error estimates presented are conservative, as the procedures described here are merely a "first pass" using relatively simple methods; more robust validation efforts, for example using keyword co-associations, will almost certainly identify additional errors.

University Twitter accounts

Goad and Chartwell searched 895 university accounts for over 20 terms they define as DEI-related (*2*). They used the resulting dataset of N = 151284 tweets (`'tweets_clean.csv'`) to graph the use of the DEI-terms over time. Many of the terms for which they searched, however, have uses and meanings beyond DEI. For instance, "race"

could refer to competitions or athletic events, “ally” is a common nickname for “Allison”, “justice” is the title used by members of federal or state bench, and introductions are often prefaced by the phrase “it is my privilege to. . . ”.

I reviewed Goad and Chartwell’s twitter dataset for tweets that might be using seven of their DEI-related search terms in a non-DEI context. These terms were: “advocacy”, “ally”, “diversity”, “equity”, “justice”, “privilege”, and “race”. I first filtered `'tweets_clean.csv'` for all tweets they assigned to a terms (e.g., “race”), then searched this subset of tweets for strings related to non-DEI uses of that term (e.g., “5K”, “nascar”, “sailing”). To ensure that the resulting tweets were not related to DEI, I eliminated any that included the entire suite of DEI-terms with which Goad and Chartwell conducted their searches (e.g., “racism”, “equality”, “gender”, “social justice”, “blm”, “equity”, see `"validation code"` in (*11*)). Note that this method provides a conservative estimate of any non-DEI tweets included by Goad and Chartwell in there analyses, as it will only capture tweets using the limited suite of non-DEI terms I included. The complete list of filtering strings for each of the 7 DEI-terms I reviewed is in `'twitter_errors.R'`; the complete collection of non-DEI tweets is in `'twitter_notdei.csv'`.

The seven search terms reviewed comprise 64.34% of Goad and Chartwell’s twitter dataset (N = 97337 tweets). Based on the conservative validation method described above, at least 11.38% of the tweets for the seven focal terms I reviewed are not DEI-related, with the percentage of irrelevant tweets for a given term ranging from 1.93 - 36.48% (Tables 1,2). If there were no additional errors in these or the remaining 14 terms, the overall error rate for the entire data set would be 7.32%.

NIH and NSF grants

Goad and Chartwell also failed to screen for alternative uses of their focal terms when reviewing the grants awarded by NSF and NIH. For example, N = 2936 of the NSF grants they identify as being DEI-focused when searching with the term “diversity” are actually

grants for ecological or evolutionary research on genetic, phylogenetic, or species diversity (see Table 3, `'grants_nsf_diversity.csv'`). They also inflated their sample sizes for the total number of DEI-related grants awarded by NSF and NIH because they failed to account for the mechanism by which agencies transfer funds for to collaborators on successful proposals. A single funded proposal will often be represented in award databases by multiple records because the researchers based at different institutions will be allocated their respective portions of the grant as separate awards. While calculating the total support for DEI-related activities by NSF and NIH requires adding the amount of the individual awards, by not consolidating the different awards for the same project they have inflated their estimates of NSF and NIH grants by 12.13% and 66.67%, respectively.

Scientific publications in Google Scholar

Finally, Goad and Chartwell sought to identify DEI-related publications in the scientific literature. To do so they searched the repositories Google Scholar, arXiv, Web of Science, and PubMed for DEI-related articles in science, technology, engineering, and mathematics (STEM) journals by using search strings including a STEM-term and one of their DEI-related terms (e.g., "biology diversity"). I reviewed their data from Google Scholar and Pubmed.

Goad and Chartwell again failed to search their results for duplicate records. The 20537 duplicates that remained in these datasets inflated their estimate of DEI-related publications in Google Scholar and PubMed by 15.78% and 21.07%. They also failed to exclude hundreds of articles that were published in cultural studies, humanities, and legal journals (Table 4), as well as non-DEI articles on topics ranging from impact of the COVID-19 pandemic on the food rationing in Rwanda to workplace interventions for facilitating breastfeeding by working women (see Table 5,`'gs_neurology_examples.csv'`, `'pm_nondei_examples.csv'`.

Appendix B

References

1. The Plan to Dismantle DEI. *The Chronicle of Higher Education* (2023), (available at https://www.chronicle.com/article/the-plan-to-dismantle-dei).

2. M. Goad, B. R. Chartwell, "Ideological intensification: A quantitative study of diversity, equity, and inclusion in STEM subjects at american universities." (National Association of Scholars, New York, 2022), p. 50.

3. J. Peterson, [@Jordanbpeterson] retweet of @NASorg: Diversity, Equity, and Inclusion (DEI) ideology undermines the norms of open discourse, objectivity, devotion to evidence, and intellectual independence...[Tweet] (2022), (available at https://twitter.com/NASorg/status/1600157129167577088).

4. C. Rufo, [@Realchrisrufo]: Retweet of @GoadMason: Visualizing the rise of DEI in Florida's Universities {Tweet]. *Twitter* (2023), (available at https://twitter.com/GoadMason/status/1623804389252030468).

5. M. Goad, The Ideological Intensification of DEI in STEM (2023), (available at https://www.realityslaststand.com/p/the-ideological-intensification-of).

6. National Association of Scholars, Quantitative study of diversity, equity and inclusion in STEM subjects in US universities (2022), doi:10.5281/zenodo.6360904.

7. National Association of Scholars, Data for: Quantitative Study of Diversity, Equity and Inclusion in STEM Subjects in US Universities (2022), doi:10.5281/zenodo.6360904.

8. R Core Team, "R: A language and environment for statistical computing" (manual, Vienna, Austria, 2020), (available at https://www.R-project.org/).

9. H. Wickham, M. Averick, J. Bryan, W. Chang, L. D. McGowan, R. François, G. Grolemund, A. Hayes, L. Henry, J. Hester, M. Kuhn, T. L. Pedersen, E. Miller, S. M. Bache, K. Müller, J. Ooms, D. Robinson, D. P. Seidel, V. Spinu, K. Takahashi, D. Vaughan, C. Wilke, K. Woo, H. Yutani, Welcome to the tidyverse. *Journal of Open Source Software*. **4**, 1686 (2019).

10. S. Firke, "Janitor: Simple tools for examining and cleaning dirty data" (manual, 2021), (available at https://CRAN.R-project.org/package=janitor).

11. BrunaLab/quantdei_nas, (available at https://github.com/BrunaLab/quantdei_nas).

Table B1

Irrelevant tweets attributed to seven different DEI terms and the total number of tweets for each term in the original dataset. Note that this percentage is a conservative estimate, as it is based on a preliminary review.

DEI Term	Irrelevant Tweets (N)	Total Tweets (N)	% Irrelevant
advocacy	574	6311	9.10
ally	365	2074	17.60
diversity	511	26499	1.93
equity	347	11883	2.92
justice	2393	21707	11.02
privilege	1341	3676	36.48
race	5543	25187	22.01

Table B2
Sample tweets incorrectly attributed to different DEI terms.

DEI Term	sample irrelevant tweet
advocacy	a passionate physician and educator commited ot medical education, patient advocacy and community medicine, @u
	work with #uofa state relations to advocate for the university at the state legislature. become an #advocat: h
	the basic trial advocacy class at the @uarizonalaw school argued their case in a mock trial on saturday, nov.
	students in the basic trial advocacy class at the @uarizonalaw had their mock trials on saturday, nov. 18. the
	rt @uarizonalaw: icymi: read how @uarizonalaw students help veteran's in the community through the veterans' a
ally	@ally_mahoney11 y grades can be given for students if the faculty member decides and student approves in writi
	@ally_rae12508 congratulations!
	@ally_rae12508 welcome to the sun devil family!
	@ally_roy @mdwhite121 congratulations!
	@marissahhhill @asusportslawbiz @ally_ausborn congrats on your accomplishments, #asugrad! we're proud to celeb
diversity	#asuyearinreview: arizona has the greatest diversity of rattlesnakes anywhere in the world.
	@asusols snake
	rt @drbiology: 66mya, an asteroid hit the yucatn, killing ~75% of earth's life diversity. tonight learn the s
	@asuorigins there are 14,000 known species of ants. actual number is probably 2x that. diversity of their soci
	a new university of arizona-led study uses big data to assess why the diversity of species varies across the g
	a new study co-authored by university of arizona researchers provides the first quantitative assessment of how
equity	rt @squaredawaybc: haven't seen our #retirement and #personfinance blog @bostoncollege @retirementrsrch? pleas
	""highly speculative:"" prof. renee jones talks to @businessinsider about private equity ""#unicorn"" start up
	rt @hnbayld: congrats!! alex mancebo, @jonesday boston office, focuses his practice on private equity, m&a
	rt @adambsterling: thanks to benjamin clinger of @kirkland_ellis for his crash course on private equity m&
	rt @bold_francesco: can't wait to read professor @ianhaneylopez's new book merge left! every class i've taken
justice	@arizonapbs will honor the legacy of supreme court justice sandra day oconnor with the national premiere of
	.@asucrimjustice researchers have found that there is a higher likelihood of receiving a false guilty plea dur
	two weeks before her first year at asu, carson swisher changed her major, and it changed her life.
	now the #
	neal katyal has argued more cases before the u.s. supreme court than any other minority attorney, breaking the
	.@asu's home in washington, d.c., is the first building in the nation's capital named for two remarkable women
privilege	rt @azathletics: 90 years ago today, #beardown was born. it is a privilege to recognize the legacy of john byr
	rt @uapolicechief: thank u @asuatoday & @uaaa for this very special honor. it's a privilege to work with a
	rt @wendelldneal: two of the greatest guys i have ever had the privilege of working with over the years. grea
	rt @phillywill11: i had the privilege of popping up on 2 kids a couple days ago.. #blessed https://t.co/czayno
	rt @azathletics: 90 years ago today, #beardown was born. it is a privilege to recognize the legacy of john byr
race	join the @beantowncats for the jeff coombs memorial virtual road race and boston marathon celebration. https:/
	ronald a. wilson, ua title ix director and a former presiding judge for the city of south tucson, will speak a
	join the @beantowncats in the jeff coombs memorial road race on sept. 9. register here: https://t.co/
	rt @bio5: #artificialintelligence wont be spawning supercomputers or #robots programmed to end the human race
	good luck to former #uofa student and @nascar champ @kurtbusch as he attempts to race in both the indy 500 and

Table B3
Sample non-DEI grants awarded by the NSF included by searching for the term 'Diversity'.

Sample	Grant Title
1	estimation & observation of stochastic biochemical networks
2	workshop proposal for deep time earth-life observatories (detelos)
3	doctoral dissertation research: applying bathymetric lidar to advance marine landscape ecology in the third dimension
4	reu site: achieving heightened goals: undergraduate research in ecology at the mountain research station
5	reu site: integrative biology and ecology of marine organisms
6	workshop/collaborative research: vision 2020: an open space technology workshop on the future of earthquake engineering; st. louis, missouri; january 2010
7	reu site: summer fellowships in biogeochemistry and climate change
8	reu site: network for earthquake engineering simulation - reducing seismic vulnerability
9	reu site: undergraduate research experiences in tropical conservation science
10	the cepob3b young cluster: a new laboratory for studying the role of environment in planet formation and cluster evolution
11	dissertation research: diversification and evolution of major trophic modes in the xylariaceae: exploring the role of previously unknown symbiotrophic and saprotrophic fungi
12	doctoral dissertation improvement grant: plant use and domestic economy among eurasian mobile pastoralists: semirech'ye, kazakhstan during the bronze and iron age interface
13	dissertation research: plant-herbivore community assembly and the problem of specificity: do insect herbivores specialize among sympatric, congeneric plants in tropical forests?
14	factors that influence the amount and pattern of genetic diversity in zymv
15	dissertation research: the consequences of global events on vertebrate biodiversity: the paleozoic actinopterygian radiation
16	dissertation research: the latitudinal gradient in plant diversity: evidence from the sedges.
17	dissertation research: integrating morphology, molecules and ecology to understand diversification and species coexistence within the madagascar olive, noronhia (oleaceae)
18	dissertation research: characterization of foliar fungal endophyte communities of sequoia sempervirens and investigation of their symbiotic relationship
19	dissertation research: plant chemical defenses and nectar traits mediating floral competition
20	dissertation research: hydrological controls of riverine ecosystems of the napo river (amazon basin): implications for the management and conservation of biodiversity
21	dissertation research: spatial and temporal variation in an ant-plant interaction
22	dissertation research: the biotic environment and the context-dependent nature of plant-microbial symbiosis
23	dissertation research: determining the effect of hybridization on the evolvability of phenotypic traits using genomic markers
24	dissertation research: the relative importance of species, genotype, and trait diversity on ecosystem function of the tall grass prairie under varying environmental conditions
25	dissertation research: effects of the aquatic:terrestrial habitat ratio on an amphibian predator and its prey
26	dissertation research: phenotypic and niche evolution in the antbirds (aves, thamnophilidae)
27	dissertation research: regional trophic diversity dynamics in north american eocene mammals
28	dissertation research: a phylogenetic characterization of the lichen microbiome
29	dissertation research: rad phylogenetics: harnessing next-generation sequencing for molecular systematics
30	dissertation research: population genetic structure and the evolution of specialization within a multi-species ant-microbe symbiosis
31	dissertation research: genetics of colorful pigmentation in anolis lizards
32	dissertation research: a multilocus, multi-species study of community assembly in an isolated fauna
33	dissertation research: trait evolution and speciation in encelia
34	dissertation research: the evolution of self-incompatibility loci in the invasive ascidian genus, ciona
35	dissertation research: from two dimensions to three: origins and evolutionary development of the sepal crest in iris
36	rig: genome size and karyotype evolution in esociform fishes
37	collaborative research: the community ecology of viral pathogens - causes and consequences of coinfection in hosts and vectors
38	collaborative research: the community ecology of viral pathogens - causes and consequences of coinfection in hosts and vectors
39	collaborative research: the community ecology of viral pathogens - causes and consequences of coinfection in hosts and vectors
40	collaborative research: the community ecology of viral pathogens - causes and consequences of confection in hosts and vectors
41	collaborative research: the community ecology of viral pathogens - causes and consequences of coinfection in hosts and vectors
42	nets: small: the packet-scale paradigm: realizing end-to-end congestion-control for terabit networks
43	workshop on dna origami
44	international research fellowship program: towards conservation and sustainable use: genetic diversity and reproductive mode of the endangered himalayan fungus o. sinensis
45	collaborative research: innovation in social networks
46	collaborative research: innovation in social networks
47	collaborative research: assembling the euteleost tree of life - addressing the major unresolved problem in vertebrate phylogeny
48	rui: biodiversity inventory: gregarines parasitizing north american cockroaches
49	ltreb: long term studies of social behavior in a colonial bird
50	collaborative research: biotic surveys of central saharan oases

Table B4

Sample non-STEM journals and the number of articles from each included in NAS analysis of DEI-focused publications in STEM outlets.

Repository	Source	N
Google Scholar	race ethnicity and education	48
	race & class	25
	science education	25
	educational studies in mathematics	24
	journal of chemical education	19
	cbelife sciences education	17
	physics teacher	17
	educational researcher	14
	cultural studies of science education	13
	physical review physics education research	13
	annual review of law and social science	12
	journal of mathematics teacher education	12
	race, gender & class	12
	teachers college record	12
	teaching race and anti-racism in contemporary	12
	urban education	12
	cambridge journal of education	11
	critical sociology	11
	journal for research in mathematics education	11
	journal of negro education	11
PubMed	j law med ethics	142
	int j law psychiatry	122
	j urban health	108
	j health polit policy law	92
	hosp law newsl	89
	law hum behav	86
	behav sci law	80
	j am acad psychiatry law	66
	med law	64
	j law med	49
	contraception	45
	am j law med	40
	j contemp health law policy	38
	health law vigil	33
	annu rev popul law	31
	med sci law	31
	j health hosp law	30
	med law rev	30
	law med health care	27
	aids policy law	23

Table B5
Sample non-DEI publications harvested from PubMed and Google Scholar.

Repository	Title	Year	Source
Google Scholar	translating the biology of aging into novel therapeutics for alzheimer disease	2019	neurology
	revisiting protein aggregation as pathogenic in sporadic parkinson and alzheimer diseases	2019	neurology
	revised airlie house consensus guidelines for design and implementation of als clinical trials	2019	neurology
	novel biomarker signatures for idiopathic rem sleep behavior disorder a proteomic and system biology approach	2018	neurology
	the biology of cutaneous neurofibromas consensus recommendations for setting research priorities	2018	neurology
	serum neurofilament light in familial alzheimer disease a marker of early neurodegeneration	2017	neurology
	the autism epidemic ethical legal and social issues in a developmental spectrum disorder	2017	neurology
	biological tumor volume in 18fetpet before radiochemotherapy correlates with survival in gbm	2015	neurology
	dystrophin quantification biological and translational research implications	2014	neurology
	defining the clinical course of multiple sclerosis the 2013 revisions	2014	neurology
PubMed	exploring us shifts in antiasian sentiment with the emergence of covid19	2020	int j environ res public health
	a critical review of theory in breast cancer screening promotion across cultures	2008	annu rev public health
	navigating uncertainty employment and womens safety during covid19 reflections of sexual assault resistance educators	2020	gend work organ
	chronographic theory of development aging and origin of cancer role of chronomeres and printomeres	2015	curr aging sci
	there is a balm in gilead black social workers spiritual counterstory on the covid19 crisis	2020	soc work public health
	like i have no choice a qualitative exploration of hiv diagnosis and medical care experiences while incarcerated and their effects	2019	behav med
	interventions that retain african americans in hivaids treatment implications for social work practice and research	2015	soc work
	hivaids a minority health issue	2005	med clin north am
	lower hiv prevalence among asianpacific islander men who have sex with men a critical review for possible reasons	2011	aids behav
	culture in cancer survivorship interventions for asian americans a systematic review and critical analyses	2021	asian am j psychol