This is the submitted (author version) of an article accepted for publication in *BioScience*.

Fundamental errors of data collection & validation undermine claims of 'Ideological Intensification' in STEM

Emilio M. Bruna[1,2*]

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

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

[*]embruna@ufl.edu

> "@arizonalumni: Good luck to former #UofA student and @NASCAR champ @KurtBusch as he attempts to race in both the Indy 500 and Coke 600. #BearDown!"

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 (Diep 2023, Kelderman 2023, Kumar 2023). Many of the DEI programs now under fire were actually mandated and implemented decades ago by congress with broad bipartisan support (Watts et al. 2015) in response to the dramatic lack of racial, ethnic, and gender parity in STEM disciplines (Palid et al. 2023). More recent ones have been motivated by increasing evidence that diverse teams are more creative or have a competitive advantage (Hong and Page 2004, Fenster 2014, Hundschell et al. 2022), as well as employer demands for a diverse and culturally competent STEM workforce. Despite this long history and the demonstrable impact of many DEI programs, however, individuals and organizations critical of DEI programs often claim that these initiatives have become increasingly pervasive and ideological (Iyer 2022). However, this assertion is rarely supported with empirical evidence.

The National Association of Scholars (i.e., NAS) recently published a report by Mason Goad and Bruce R. Chartwell (Goad and Chartwell 2022) which the authors 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, funding agency databases, and repositories for scientific literature for instances of "DEI-related terminology" (e.g.,"diversity", "equity", "justice", "race"). They claim to have found a dramatic increase in the use of these terms in university communications and the scientific literature since 2010, which they conclude is unambiguous empirical evidence of "ideological intensification" in the academic and scientific arenas (Goad and Chartwell 2022). They also conclude that if these trends continue, "the future of STEM, along with

the rest of the academy, is almost certainly imperiled" (see Goad and Chartwell (2022), p. 47), and encourage others to use their data-mining tools and database to conduct similar research. Since the report's release in December 2022, it has been widely hailed and distributed by prominent DEI critics such as Jordan Peterson and Christopher Rufo.

Readers of the NAS report, especially those familiar with scientometric research, will quickly identify some glaring analytical shortcomings. These include the absence of any formal statistical tests, the use of a single (and questionable) "control" term in literature searches, and using the absolute number of DEI-related tweets or scholarly publications emerging from universities as the foundation of their analyses and graphs (Fig. 1). This last issue is particularly egregious — the trends they purport to have documented, and which they attribute to institutions increasingly emphasizing "DEI ideology" over science, are simply artefacts of both Twitter use and publication numbers increasing dramatically since 2010. Put another way *one would expect to see increases like those they report even if the proportional effort made by institutions remained unchanged*, which is why it is essential to conduct analyses such as these with 'relativized' rather than absolute values.

That said, none of this actually matters in light of what I discovered when accepting a challenge made by the report's authors in their *Technical Appendix* (p. 48–50).

Goad and Chartwell made the laudable decision to make their code publicly available (National Association of Scholars 2022a), along with the 'clean' data on which they base their conclusions (National Association of Scholars 2022b), "so that other analysts can scrutinize the methods and replicate them" (Goad and Chartwell (2022), p. 48). When I did so, I found that they failed to conduct even the most rudimentary data validation procedures prior to text-mining. Using standard tools and simple methods, I found that their "clean" data sets contain thousands of irrelevant records and duplications [*Supplementary Materials and Methods*]. Notable examples include the tweet that opened this *Letter* — one of over 12000 about topics ranging from sporting events ("race") to

members of the Supreme Court ("justice") to hedge funds ("equity") — along with more than 2000 NSF grants for ecological and evolutionary research on species "diversity". Others can be found in their dataset of "DEI articles in STEM journals", which included at least 20537 duplicated records (inflating their estimate of DEI-related publications in Google Scholar and PubMed by 18.74% and 26.7%, respectively), hundreds of articles published in cultural studies, humanities, and legal journals such as *Critical Sociology*, *The Medical Law Review*, and *The Annual Review of Law and Social Science*, and thousands of non-DEI articles on topics ranging from palliative care for cancer patients to transcatheter aortic valve replacements (see *Supplementary Materials*).

Research from think tanks and advocacy organizations heavily influences policy, legislation, and contemporary debates related to scientific research and higher education (Gándara and Ness 2019, Baig et al. 2020). Computational approaches can greatly expand the scope and impact of this research, but only if the conclusions are based on robust methods and reliable data. Furthermore, methodological transparency by organizations publishing outside of the traditional scholarly literature are commendable, but only when accompanied by self-accountability. Because the conclusions in Goad and Chartwell's report were based solely on datasets that are clearly of questionable quality, the NAS should adhere to its principles and retract the report. Failure to do so would be an ironic example of what they claim has become pervasive in university settings: the prioritizing of ideology over intellectual rigor.

References

Baig MI, Shuib L, Yadegaridehkordi E. 2020. Big data in education: A state of the art, limitations, and future research directions. International Journal of Educational Technology in Higher Education 17: 44.

Diep F. 2023. Florida governor asked all public universities for spending data on diversity and critical race theory. The Chronicle of Higher Education.

Fenster CB. 2014. Broader Impacts Come of Age. BioScience 64: 645–646.

Gándara D, Ness EC. 2019. Ideological Think Tanks and the Politics of College Affordability in the States. The Journal of Higher Education 90: 717–743.

Goad M, Chartwell BR. 2022. Ideological intensification: A quantitative study of diversity, equity, and inclusion in STEM subjects at American universities. National Association of Scholars (https://www.nas.org/reports/ideological-intensification/full-report).

Hong L, Page SE. 2004. Groups of diverse problem solvers can outperform groups of high-ability problem solvers. Proceedings of the National Academy of Sciences 101: 16385–16389.

Hundschell A, Razinskas S, Backmann J, Hoegl M. 2022. The effects of diversity on creativity: A literature review and synthesis. Applied Psychology 71: 1598–1634.

Iyer A. 2022. Understanding advantaged groups' opposition to diversity, equity, and inclusion (DEI) policies: The role of perceived threat. Social and Personality Psychology Compass 16: e12666.

Kelderman E. 2023. The Plan to Dismantle DEI. The Chronicle of Higher Education.

Kumar D. 2023. Florida bills target 'political loyalty tests' in college diversity efforts. Tampa Bay Times.

National Association of Scholars. 2022a. Code for "Quantitative study of diversity, equity and inclusion in STEM subjects in United States universities" (https://github.com/NASorg/quantdei).

National Association of Scholars. 2022b. Data for "Quantitative study of diversity, equity

and inclusion in STEM subjects in US universities"
(https://zenodo.org/record/6360904).

Palid O, Cashdollar S, Deangelo S, Chu C, Bates M. 2023. Inclusion in practice: A systematic review of diversity-focused STEM programming in the United States. International Journal of STEM Education 10: 2.

Watts SM, George MD, Levey DJ. 2015. Achieving Broader Impacts in the National Science Foundation, Division of Environmental Biology. BioScience 65: 397–407.

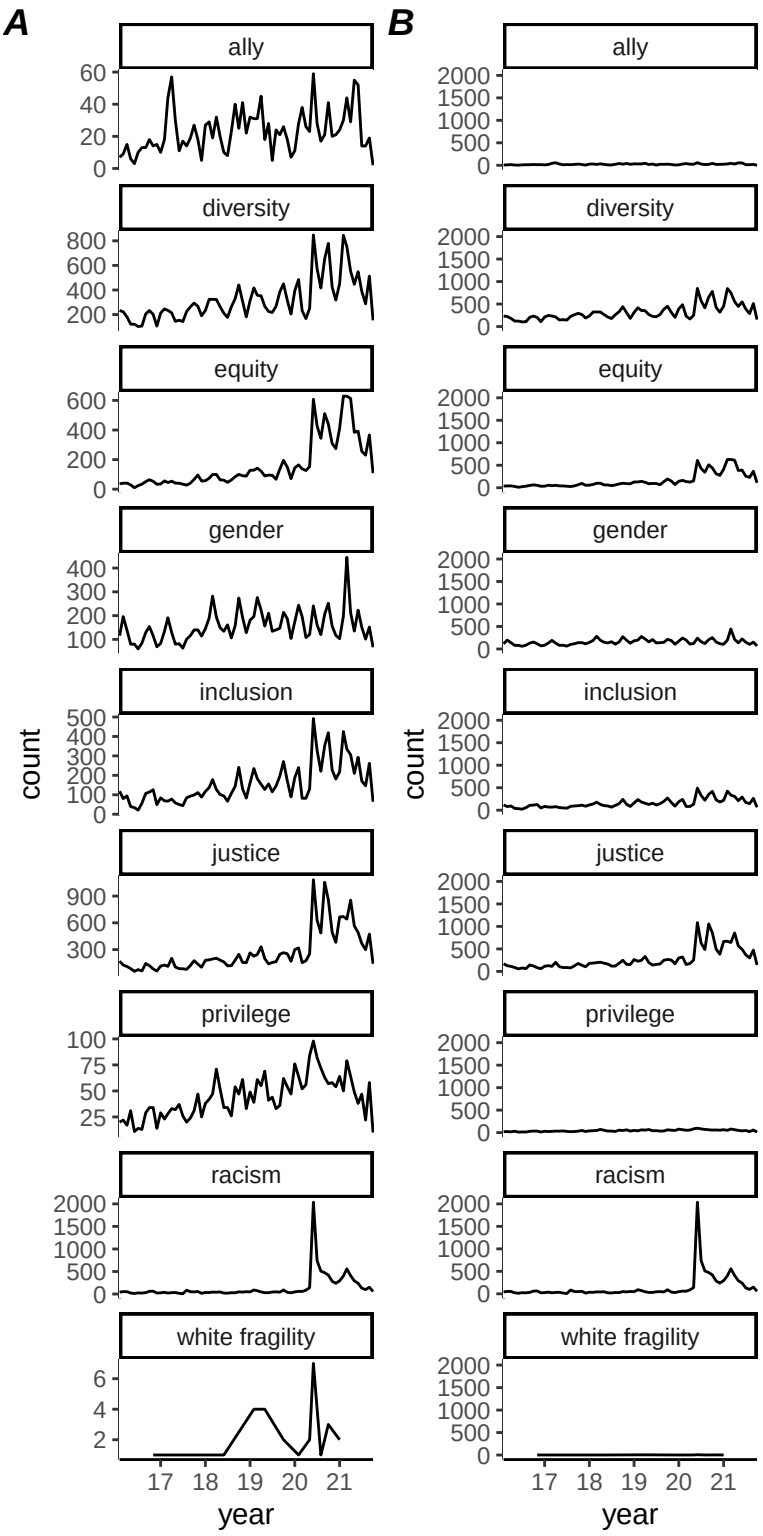

Figure 1: (A) Subset of Figure 8 from the NAS report (*'Fig 8: DEI-related Tweets from all school-related accounts by DEI term'*); the floating y-axes accentuate negligible increases in very rare terms. (B) The same panels but with identical y-axes scaled by the frequency of the most common term. Note that both sets of figures were made with the original, uncorrected NAS data, so the actual number of tweets for each term is much lower.