STANFORD UNIVERSITY STANFORD, CALIFORNIA 94305



SCHOOL OF EDUCATION TELEPHONE: (650) 723-1761

December 17, 2013

To the NSF Review Panel,

I am writing to voice my strong commitment to collaborate on the proposed research titled, "Interactive Machine Learning for Text Analysis", submitted by PI Jeffrey Heer.

I am the director of Stanford's Center for Computational Social Science, and I led the NSFfunded interdisciplinary collaboration (CDI grant) called "Mimir". In these roles and in my research, I have tried to advance computational data analysis techniques and extend their use in the social sciences. Much of my own work concerns social network dynamics, dynamic topic modeling, and their interrelation. I am familiar with data visualization as well, having created one of the first dynamic social network imaging tools (SoNIA). I have collaborated extensively with computer scientists in natural language processing and data visualization to write a series of papers on language flows in academe across disciplines, on language leads and lags between funding agencies and publishing outlets, the nature of language from interdisciplinary mixing, and so on. I am familiar with the research at the intersection of these fields and I consider Prof. Heer's work on large scale text analysis and visualization to be exemplary and incredibly useful.

In much of my research, I must explore data and identify where language communities and forms of argumentation arise and test whether network mechanisms underlie them. I frequently find myself using NLP software for topic modeling and text visualization, such as Stanford's Topic Modeling Toolkit and "Termite" for topic visualization (developed by proposal Senior Personnel Jason Chuang). While these tools are quite helpful, sometimes topic models are too coarse-grained to resolve desired units of analysis; for example, they are insufficient for identifying specific scientific methods and tracking their dissemination. Tools for fine-grained interactive text analysis, as proposed by Prof. Heer, are sorely needed. Instead of believing there is a "one" ground truth, we believe there are just "samples" of them and they can be biased. As such, we also constantly work hard to validate the results found by natural language processing and compare them to our intuitions and samples of classification. This frequently relies on inductive forms of data exploration and model verification via visualization. As such, the tools developed by Prof. Heer and his students are immensely useful to me in my research efforts.

I have collaborated with Prof. Heer and his students in the past (see *Poetics*, 2013 for review). They have always been conscientious collaborators who work hard to solve problems. I struggle to find better colleagues and am eager to sustain a working relation with him for years to come. The field of data visualization is increasingly useful for natural language processing, and collaborations across those fields and with domain scholars like myself, will advance these methods in a way that brings them greater applicability, validity and generalizability.

Sincerely Yours,

Daniel A. McFarland Associate Professor of Education, (by courtesy) Sociology and Organizational Behavior