DEPARTMENT OF POLITICAL SCIENCE
UNIVERSITY of WASHINGTON

January 5, 2014

To the NSF review panel:

I am writing in support of the proposal, CHS: Small: Interactive Machine Learning for Text Analysis, submitted by Jeffrey Heer. I am committed to collaborating on this timely and needed research.

I am a Professor of Political Science at the University of Washington. I have used machine learning methods in my research since 2007 and collaborated with computer scientists on several projects. This research includes classifying 400,000 legislative bill titles for policy topic, tracing the progress of policy ideas within the hundreds of thousands of pages of bill texts, and exploring explanations for the 2007-08 financial crisis across hundreds of legislative hearings.

Machine learning methods are just starting to take hold in the social sciences. The relatively rare scholars who do use them tend to be methodologists trained at a limited number of elite universities. This is unfortunate because text is such an important source of information about politics and other social phenomena. High startup costs associated with writing and using code for converting text to data, and for analyzing text as data have been major impediments to the use of machine learning methods in the social sciences. Social scientists are not computer scientists (most of them anyway). Those who try to be both risk being subpar at both.

One option is collaboration, something that we see with increasing frequency. However, cross-disciplinary collaborations can be professionally challenging given differing professional standards. Research questions of interest to political scientists may seem of little interest to computer scientists and vice versa.

Another option is to lower the costs of 'arming' social scientists, so that they can employ tools developed by experts, effectively and responsibly. The current proposal seeks to develop interactive tools that reflect how many social scientists go about their work. In the social sciences, classification is a very important research activity, whether issues, frames, tone, memes etc. But developing a valid and reliable classification scheme is often far from simple. In most cases a researcher starts with a rough framework, tests it on limited example text, makes modifications in response, tests it again, until path dependence sets in and the costs of additional modifications become prohibitive. When this manual process is applied to big data situations, 'lock in' tends to occur well before most of the information in the data can be examined. What is being proposed will lower the costs of this iterative process and radically improve the quality of social science research across many domains.

Box 353530101Gowen Hall Seattle, WA 98195-3530
206.543.2780 fax 206.685.2146www.polisci.washington.edu/

DEPARTMENT OF POLITICAL SCIENCE
UNIVERSITY of WASHINGTON

I am fairly familiar with commercial options available for someone with limited technical skills interested in using machine learning methods in their research. These resources are rapidly improving but the ones I know of essentially offer a selection of discrete tools wrapped in a GUI. They address the constraint of programming skills but do little to advance the interactive development of classification systems.

In my opinion, Jeffrey Heer and his graduate students are going about things in the right way. They are reaching out to researchers to better understand their goals and constraints, and working with them to develop general purpose tools that make valuable computer science innovations more broadly accessible. I am very excited to have the opportunity to collaborate with such capable scientists.

Sincerely,

John Wilkerson

Professor of Political Science

Director, Center for American Politics and Policy

Box 353530101Gowen Hall Seattle, WA 98195-3530
206.543.2780 fax 206.685.2146www.polisci.washington.edu/