Columbia University in the City of New York | *New York, NY 10027*

KATHLEEN McKEOWN
Director, Institute for Data Sciences and Engineering
Henry and Gertrude Rothschild Professor
(212) 939-7118
kathy@cs.columbia.edu
http://www.cs.columbia.edu/~kathy/

Department of Computer Science
450 Computer Science Building
(212) 939-7000
Fax: (212) 666-0140

8 March 2015

To Magic Grant Selection Committee,

I am writing to express my enthusiastic support for the project "Emotional Impact: Discovering the Process of Generating Humor and Novels." This project is a combination of the expert talents of Stanford's Human-Computer Interaction (HCI) group with Columbia's Natural Language Processing (NLP) group. Stanford's leadership and expertise in crowdsourcing complex tasks is a perfect complement to our work in story understanding and sentiment analysis. This collaboration will push our understanding of the most human elements of storytelling: emotional impact.

With the rise of social media, sentiment analysis is an important NLP problem with applications to businesses that want to participate in online conversations about their brands. A problem of equal importance is determining which speakers are the influencers – the ones that through a complex combination of traits can change people's minds. Both sentiment analysis and influencer detection are problems that must address the social and emotional factors of language. NLP has been successful in both areas. This is strong evidence that the complexities of emotional impact are possible to understand in the near future, if the right approaches are taken.

The early work our team has done on humor has shown that the problem of emotional impact can be scoped to a level where we can easily study it in a data-driven way. Similar to studying 140-character tweets and their sentiment, humor is short, there are millions of data points, and it is cheap and easy for humans to label it for quality. Data alone is not enough, but paired with strong models of emotional impact taken from the literature, we can test and determine where to refine the theories.

After humor, the next important step in generating emotional impact is to test it on a broader, more challenging domain: novels. Novels represent an achievement in communication that many humans aspire to, but never manage to complete because of the length and complexity of the process. To crowdsource a novel would demonstrate an impressive step in computer-mediated communication and collaboration.

Our group is already expert in the mechanics of story writing such as coherence and plot structure. However, a novel requires more than good mechanics. Understanding emotional impact will bring us much closer to allowing us to generate stories with the narrative drive that compels readers to turn pages. Decomposing the elements of emotional impact and gathering data on the process people go through to generate it will be invaluable to the process of teaching

computers how to write novels and how to communicate with us over long periods of time in a way that is most comfortable to us. I have a first-year student, Jessica Ouyang, committed to the project who has already done groundbreaking work in understanding narrative in NLP. Particularly, she has applied narrative theories of suspense to short narratives to validate the theories. Being able to generate them in a next step. In order to leverage the crowd to help generate suspenseful narratives, she will be paired with Lydia Chilton – a senior student at Stanford who will be a post-doc next year splitting her time between Stanford and Columbia in order to ensure that the collaboration is seamless.

I think this is an important project to the future of NLP. I think its chance of success within a year are high. Although it is a hard computer science problem, it uses the flexibility of crowds. Additionally, it is on the trajectory of modern NLP research. I think this is the ideal team to do it.

Sincerely,

Kathleen McKeown
Director, Institute for Data Sciences and Engineering
Henry and Gertrude Rothschild Professor of Computer Science