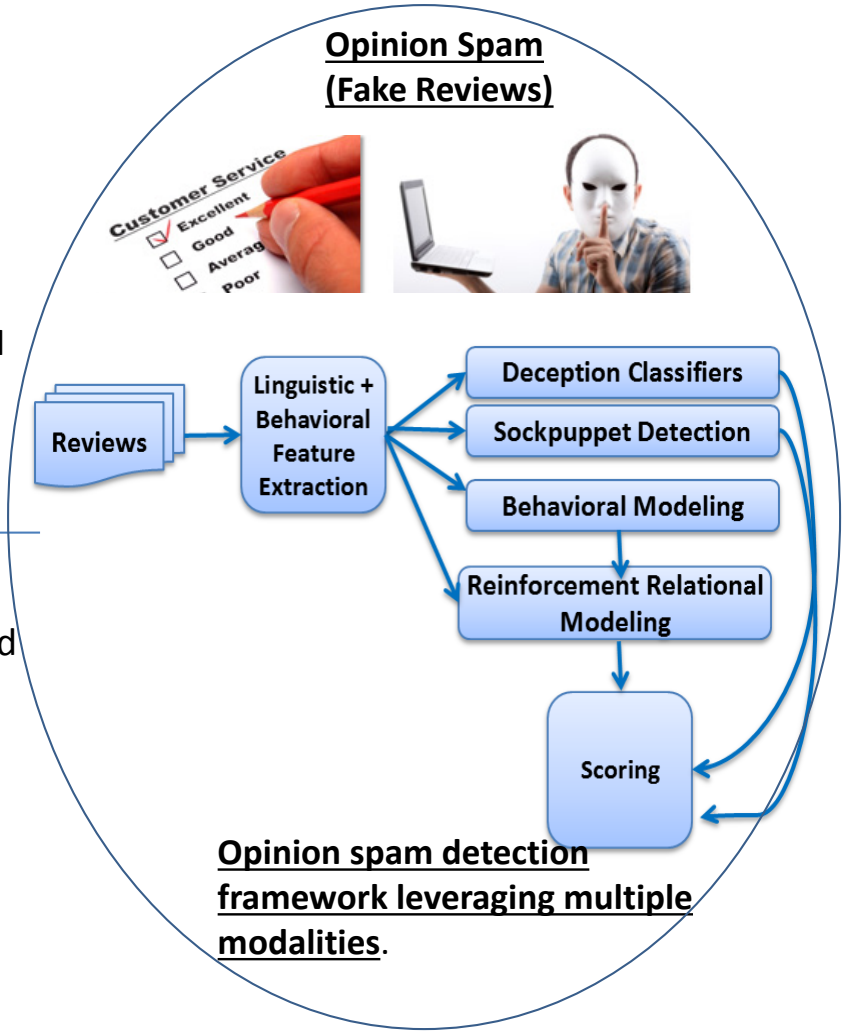
Statistical Models for Opinion Spam Detection Leveraging Linguistic and Behavioral Cues

Challenge:

- Detecting fake reviews (opinion spam) from limited ground truth data for learning.
- Unsupervised learning and evaluation

Solution:

- Learning from noisy and gold standard data
- exploit reinforcement relations between (1) fraudsters, (2) fake reviews, (3) deceptive language, and (4) fraud behavior



Scientific Impact:

- Focus on learning and evaluation on unlabeled data
- Strategies in detecting opinion spam can be applied to sockpuppets, viral hoaxes, forensic linguistics, and other abuse

Broader Impact:

- Improve trustworthiness of web content and reduce social implausibility
- Course on security analytics leveraging NLP, DM, and statistical methods
- Characterizes several behavioral modalities of deception posing a marketing, consumer, and economic risk

PI: Arjun Mukherjee (NSF 1527364 , Univ of Houston, arjun@uh.edu