



The Impact of Anonymization on Sentiment Analysis of Social Media Postings

Prof. Dr. Thomas J. Lampoltshammer

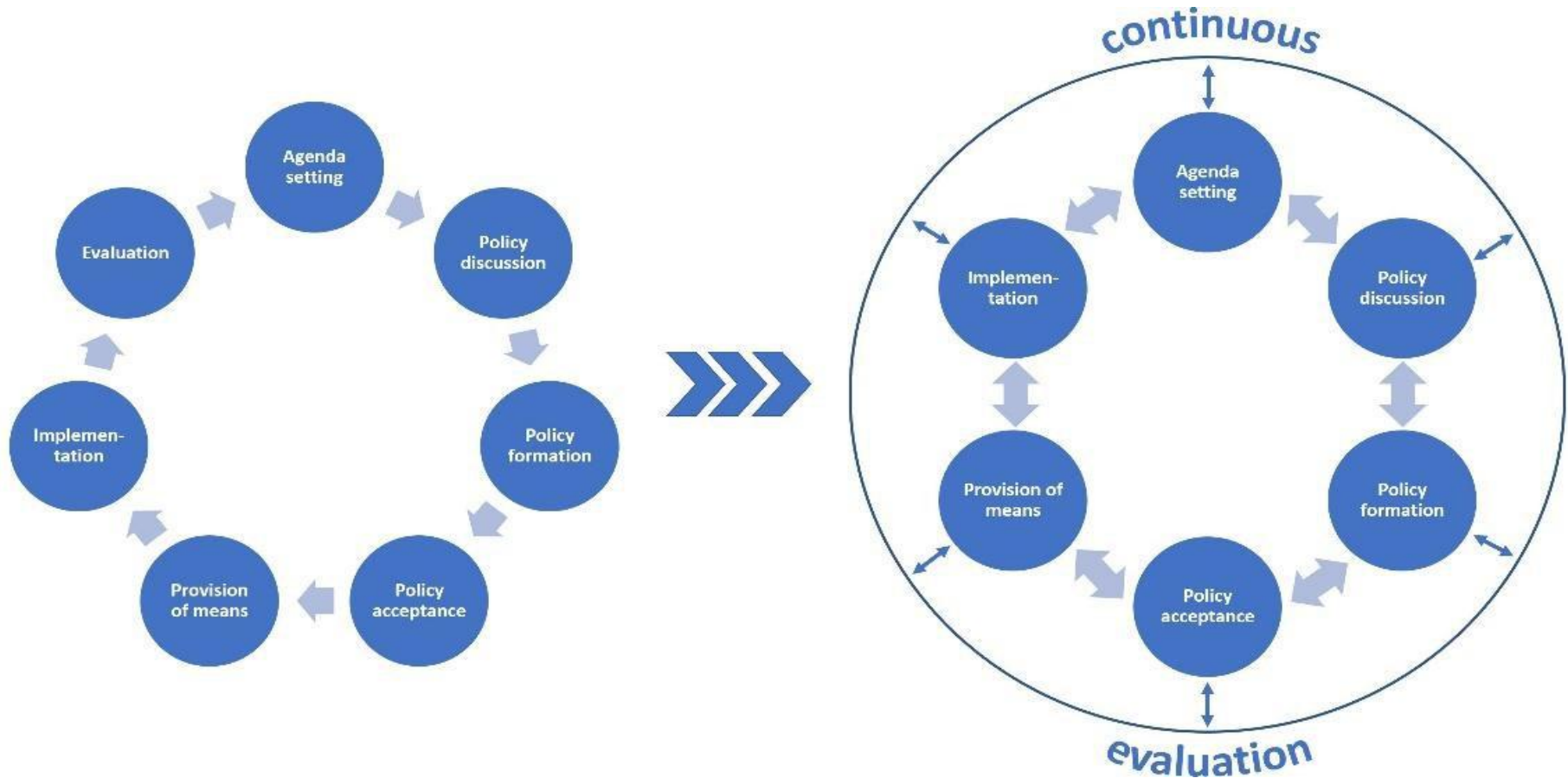




Αλέξης Τσίπρας Πρωθυπουργός της Ελλάδας



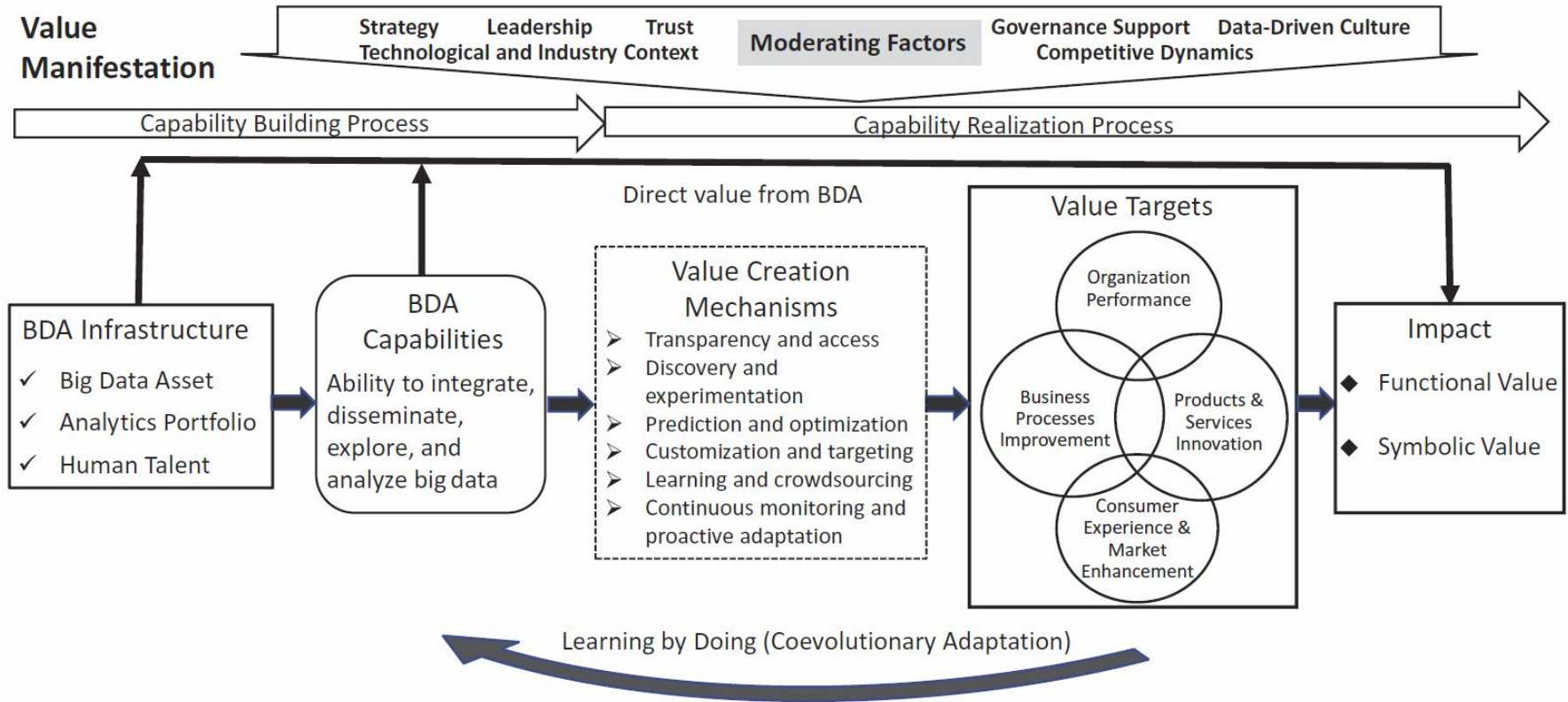
Use of Big Data in the Policy Cycle



Höchtl et al., 2016



Use of Big Data in the Policy Cycle

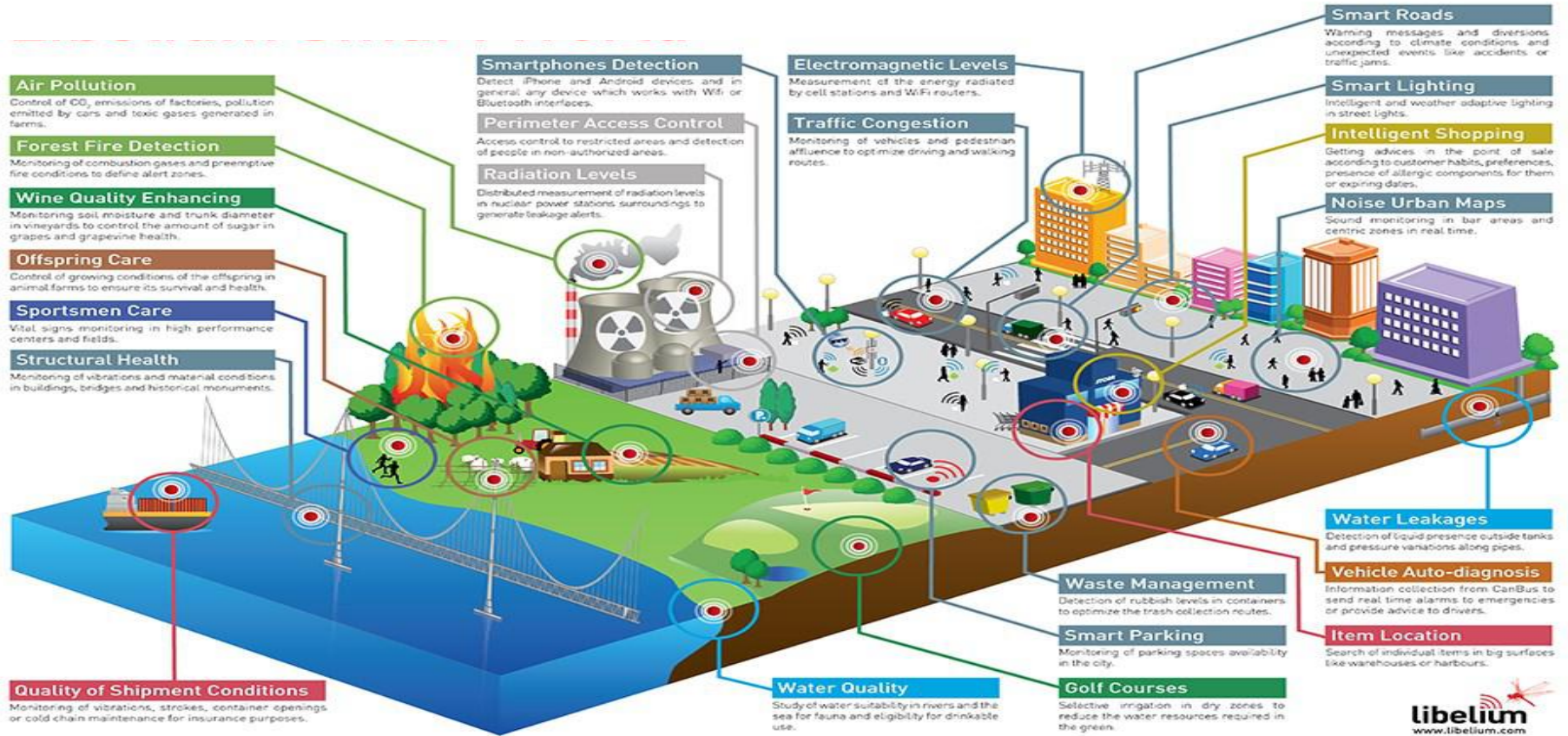


Investments --- Assets ----- Capabilities ----- Applications ----- Targets ----- Impacts ----- Value

Grover et al., 2018

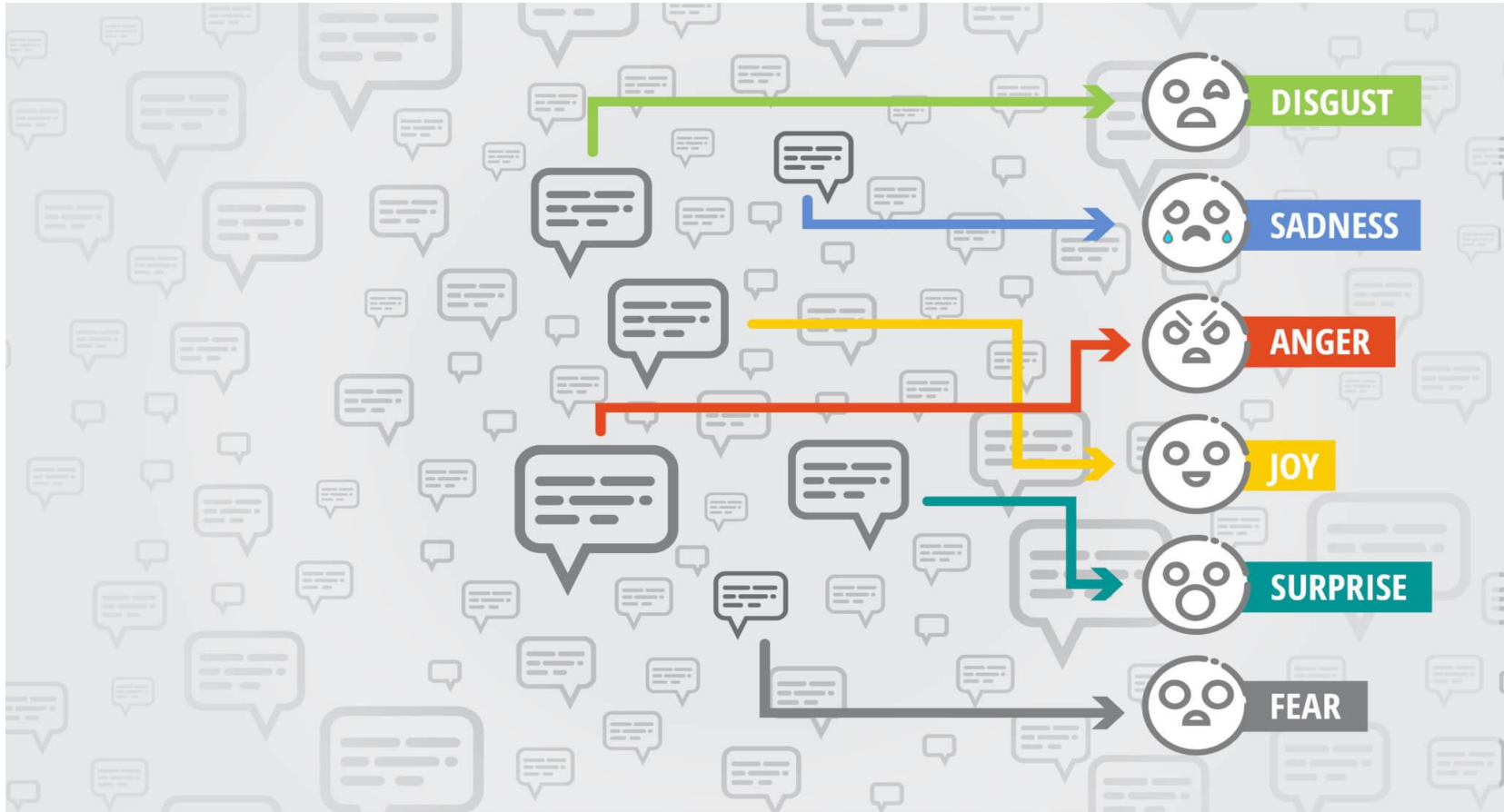


Data Sources in Smart Cities



<https://visualmethodculture.wordpress.com/2015/03/20/smart-cities-and-why-they-need-a-lot-more-social-scientists-to-get-involved/>

What is Sentiment Analysis

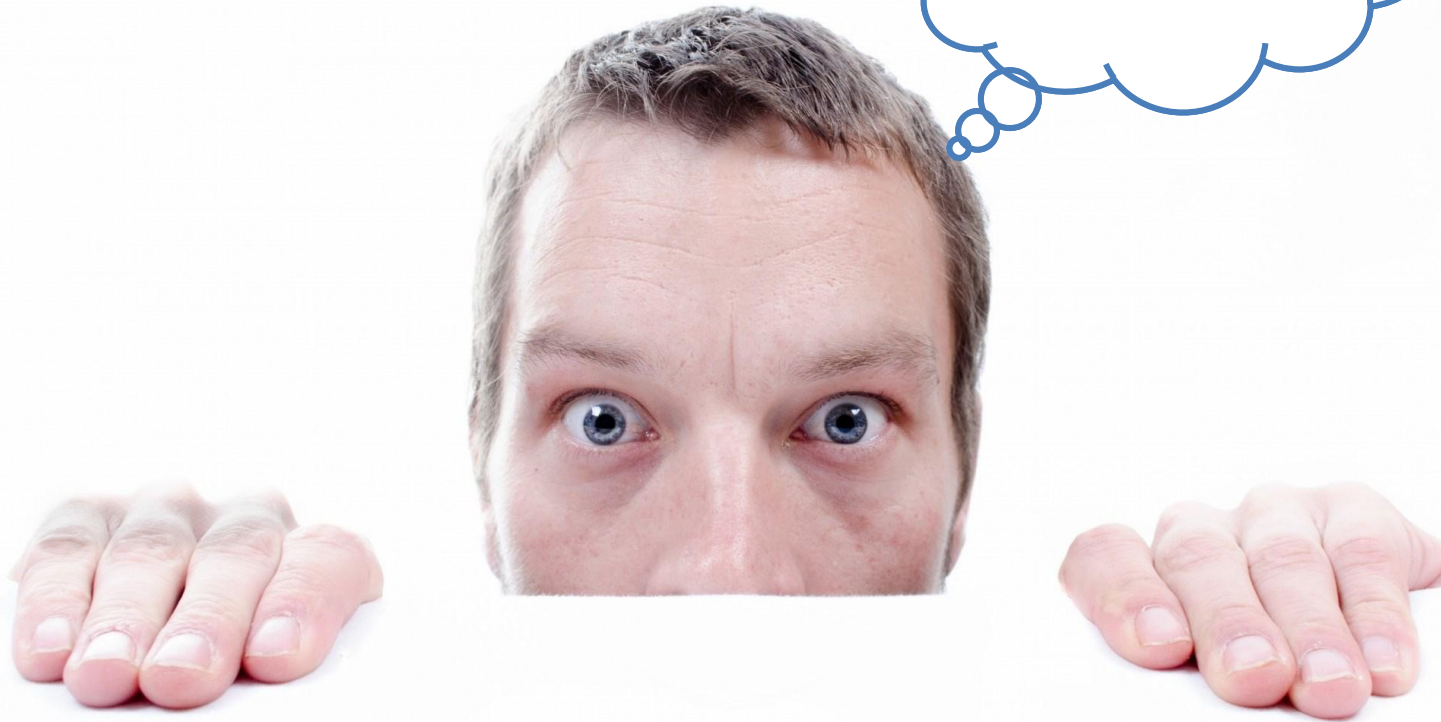


<https://www.crimsonhexagon.com/blog/what-is-sentiment-analysis/>



Remember, remember ...

25.05.2018



GDPR - why should I care ?

*"Infringements [...] shall [...] be subject to administrative fines up to **20 000 000 EUR, or** in the case of an undertaking, **up to 4 % of the total worldwide annual turnover** of the preceding financial year, whichever is higher."*

GDPR Article 83 (5)



Social Media Analysis is over?!



- **Naïve approach towards the implementation of anonymization**
- **Impact analysis of anonymization on sentiment analysis of social media**

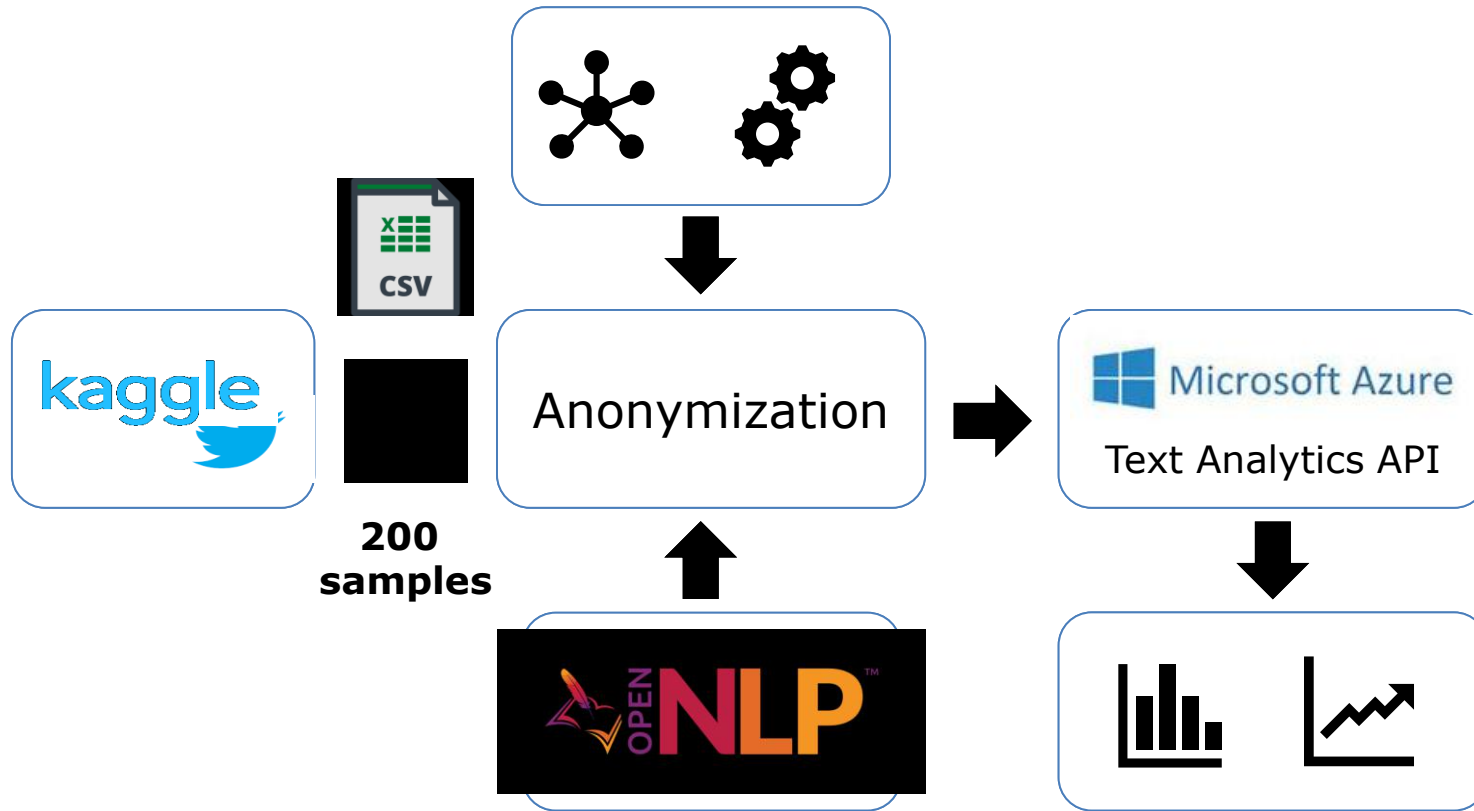




kaggle



Processing Pipeline



POS as the Foundation

POS tagging

[about Parts-of-speech.Info](#)

Enter a **complete sentence** (no single words!) and click at "POS-tag!". The tagging works better when grammar and orthography are correct.

Text:

What is the job title of Frank Gallagher ?

▲
Pronoun, WP

✎ Edit text

Adjective

Adverb

Conjunction

Determiner

Noun

Number

Preposition

Pronoun

Verb

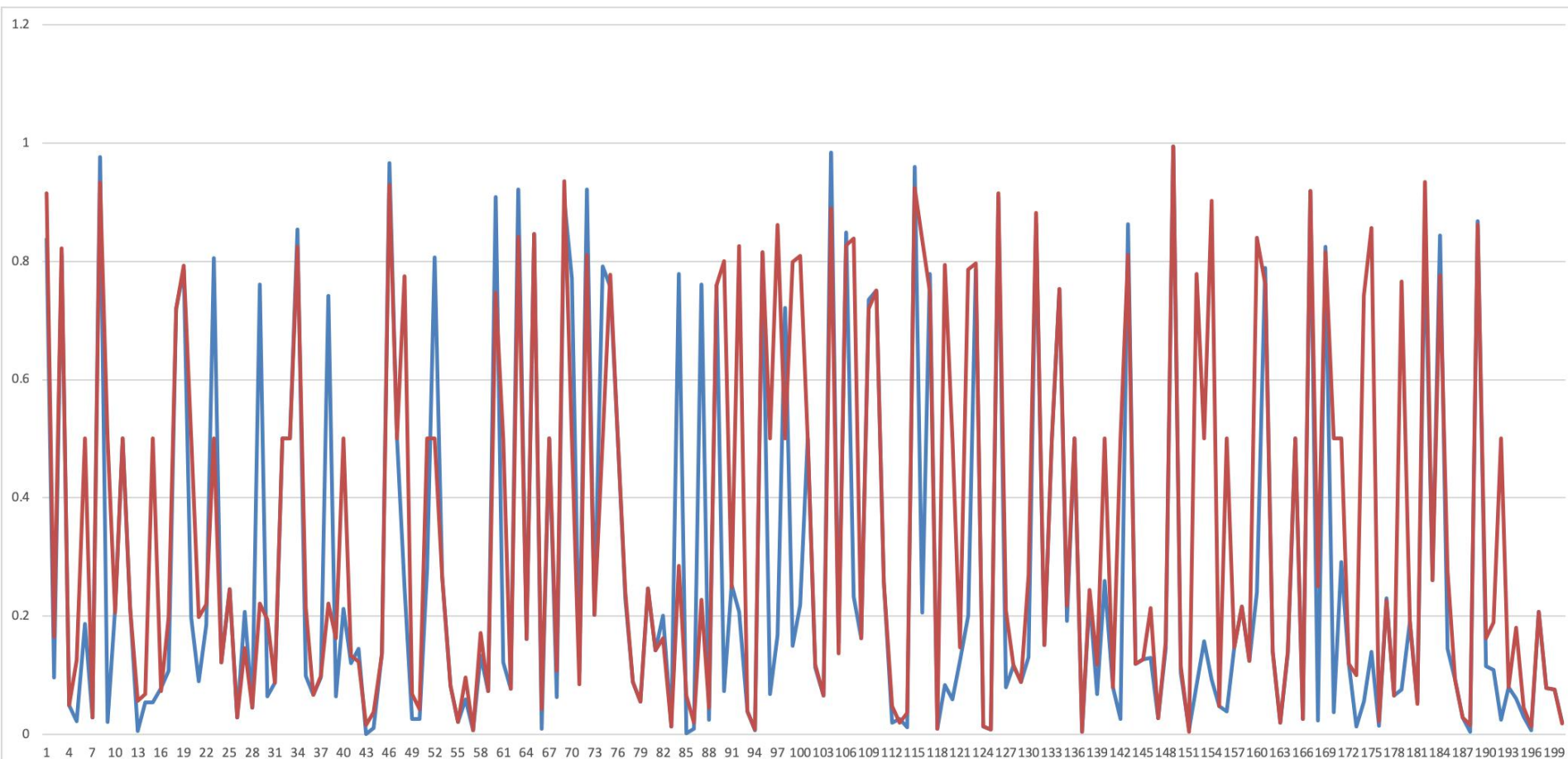
<https://miningbusinessdata.com/tips-choosing-correct-entities-dialogflow/>



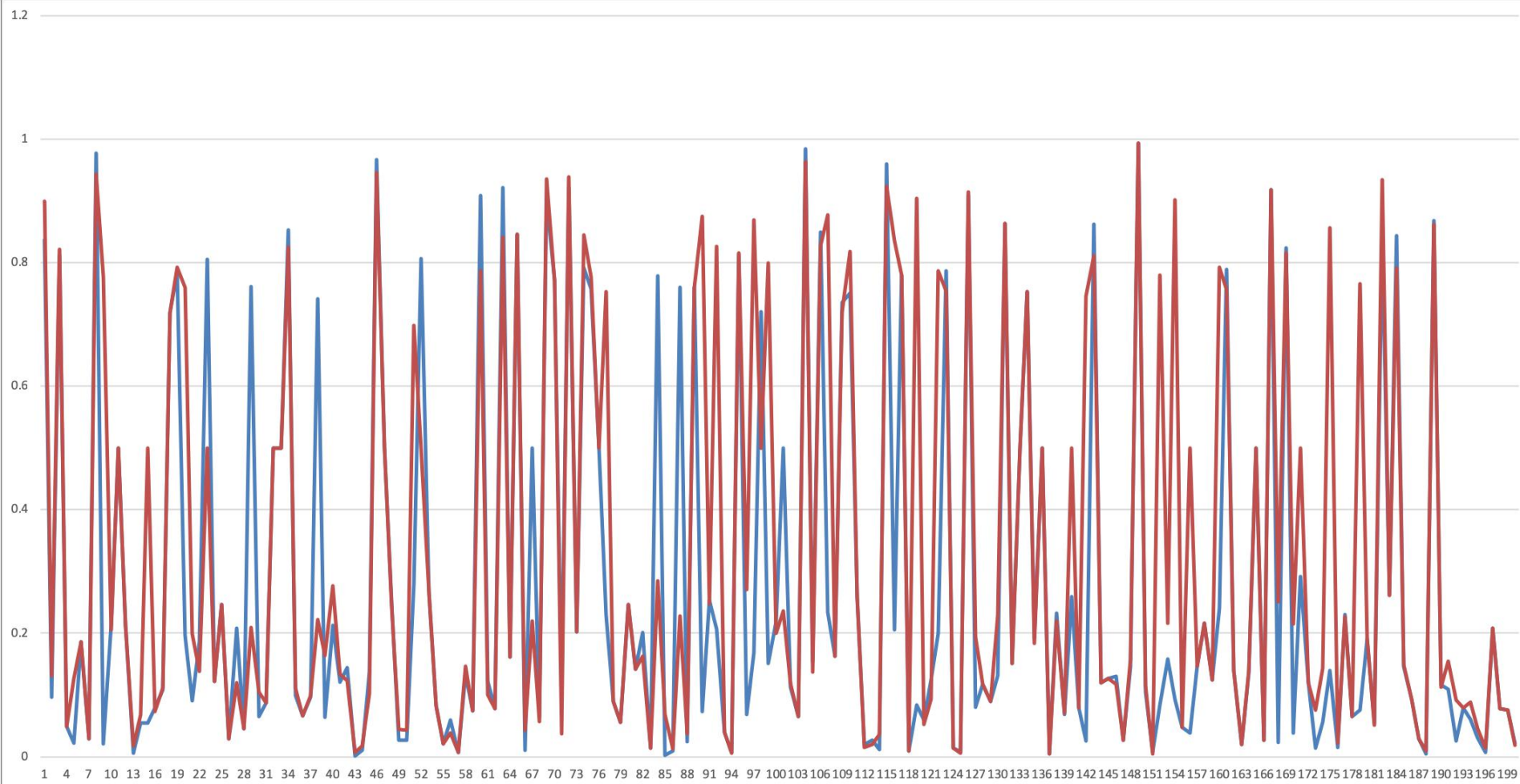
1. Replace NNP with pseudonym
2. Remove NNP completely
3. Bag-of-words w/o fixed negations



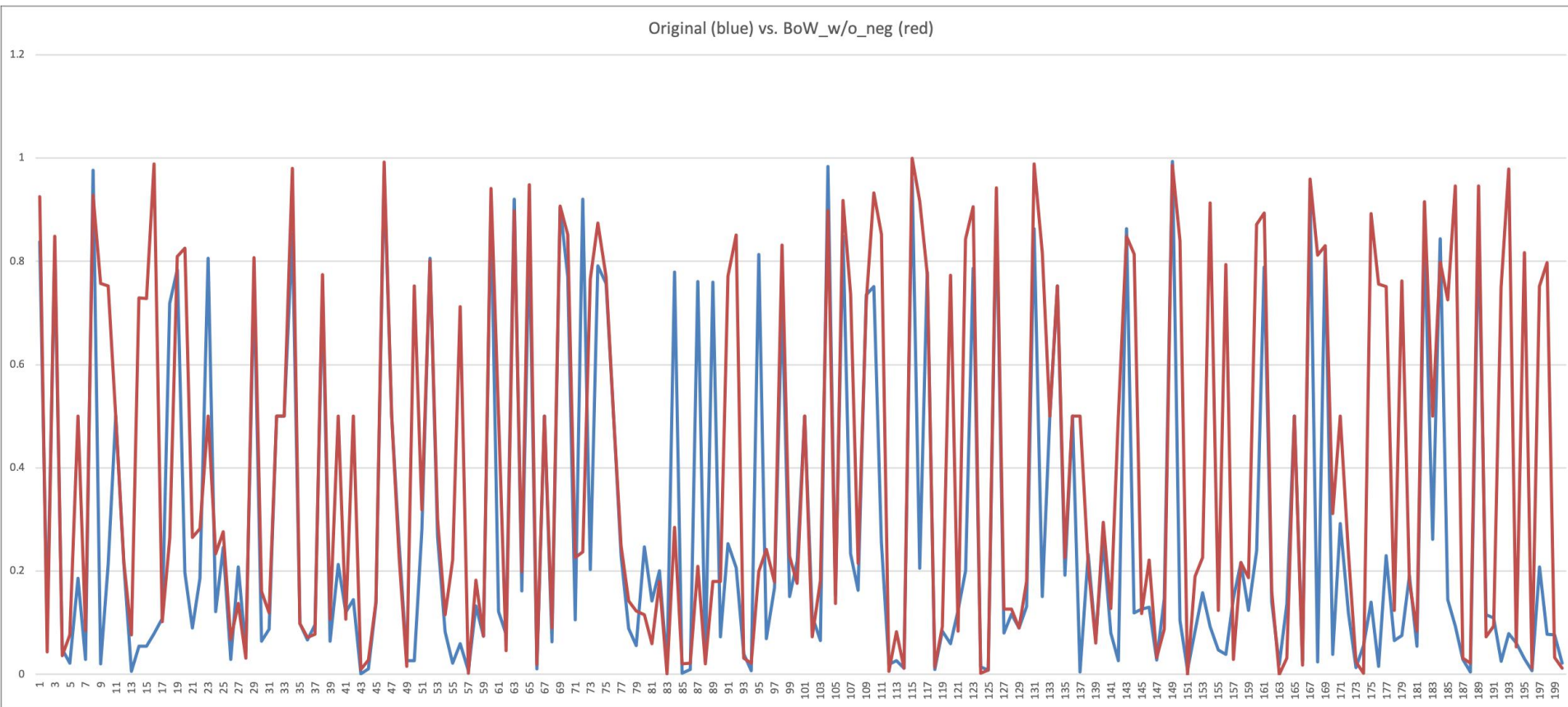
Results Strategy 1 vs. Original



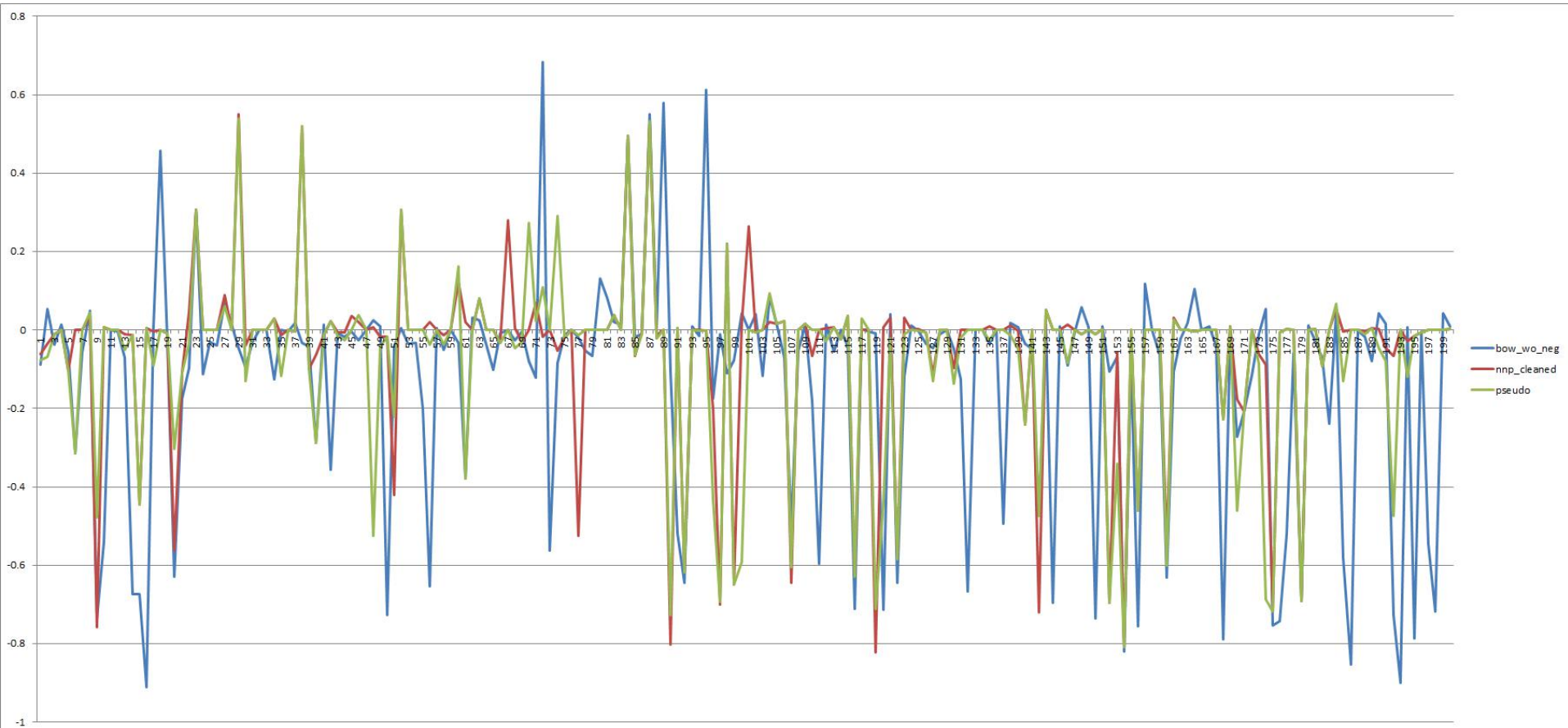
Results Strategy 2 vs. Original



Results Strategy 3 vs. Original



Results Strategy 1 vs. 2 vs. 3



- Different kinds of social media and user contents
- Additional sentiment analysis tools
- Additional anonymization techniques
- Legal analysis of implications of the individual anonymization levels



Thank You!

