

Challenges and Opportunities in Information Manipulation Detection: An Examination of Wartime Russian Media

Chan Young Park*
Carnegie Mellon University
chanyoun@cs.cmu.edu

Julia Mendelsohn*
University of Michigan
juliame@umich.edu

Anjalie Field*
Stanford University
anjaliief@stanford.edu

Yulia Tsvetkov
University of Washington
yuliats@cs.washington.edu

Can we use existing NLP tools to understand information manipulation in emerging crises?

Kind of! Our dataset, VoynaSlov, and analyses can help. But we expose many open challenges.

Motivation

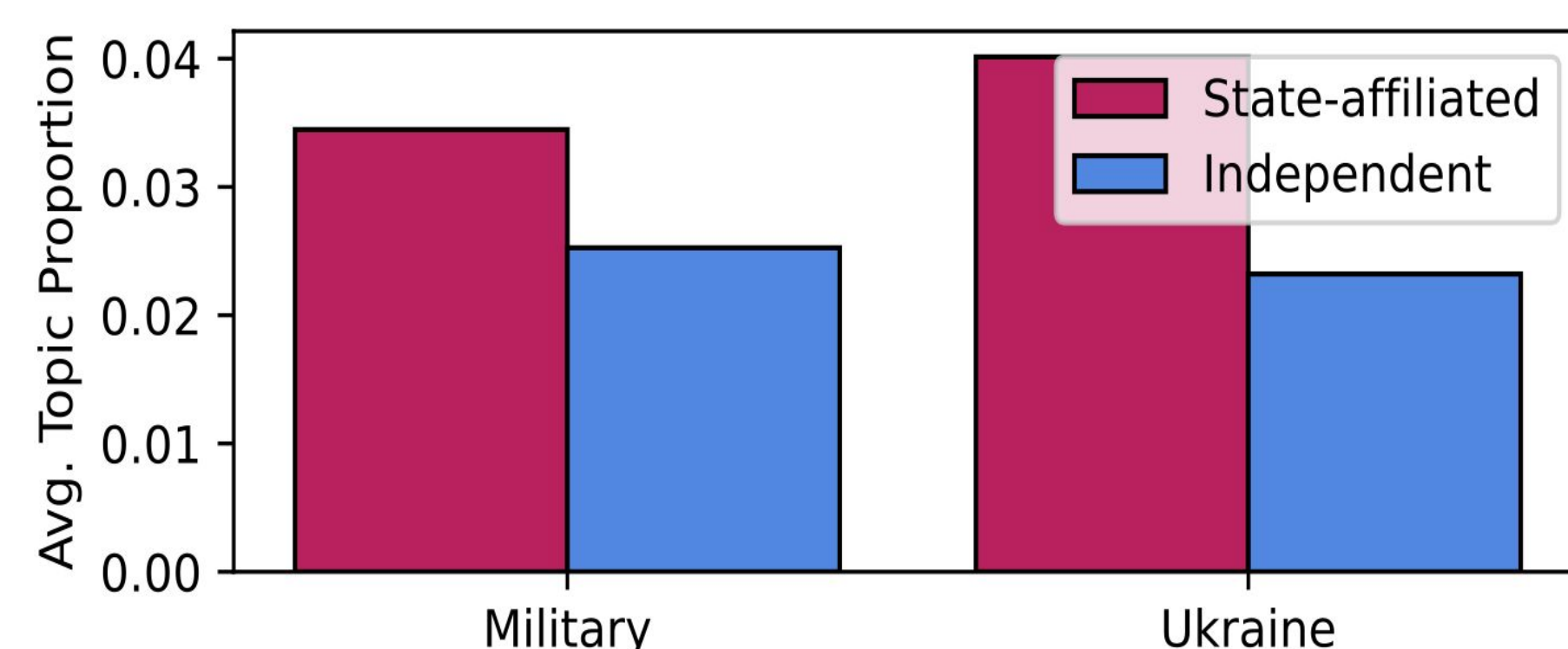
- 2022 Russia-Ukraine War is manifested in ongoing **information warfare**
- The **manipulation of public opinion** over social media is “a critical threat to democracy”
- Analysis through lens of **media effects** of agenda-setting, framing & priming compared across:
 - **Time** (before and after Feb 24 invasion)
 - **Platform** (VKontakte vs. Twitter)
 - **Media control** (state-affiliated vs. independent)

Agenda Setting

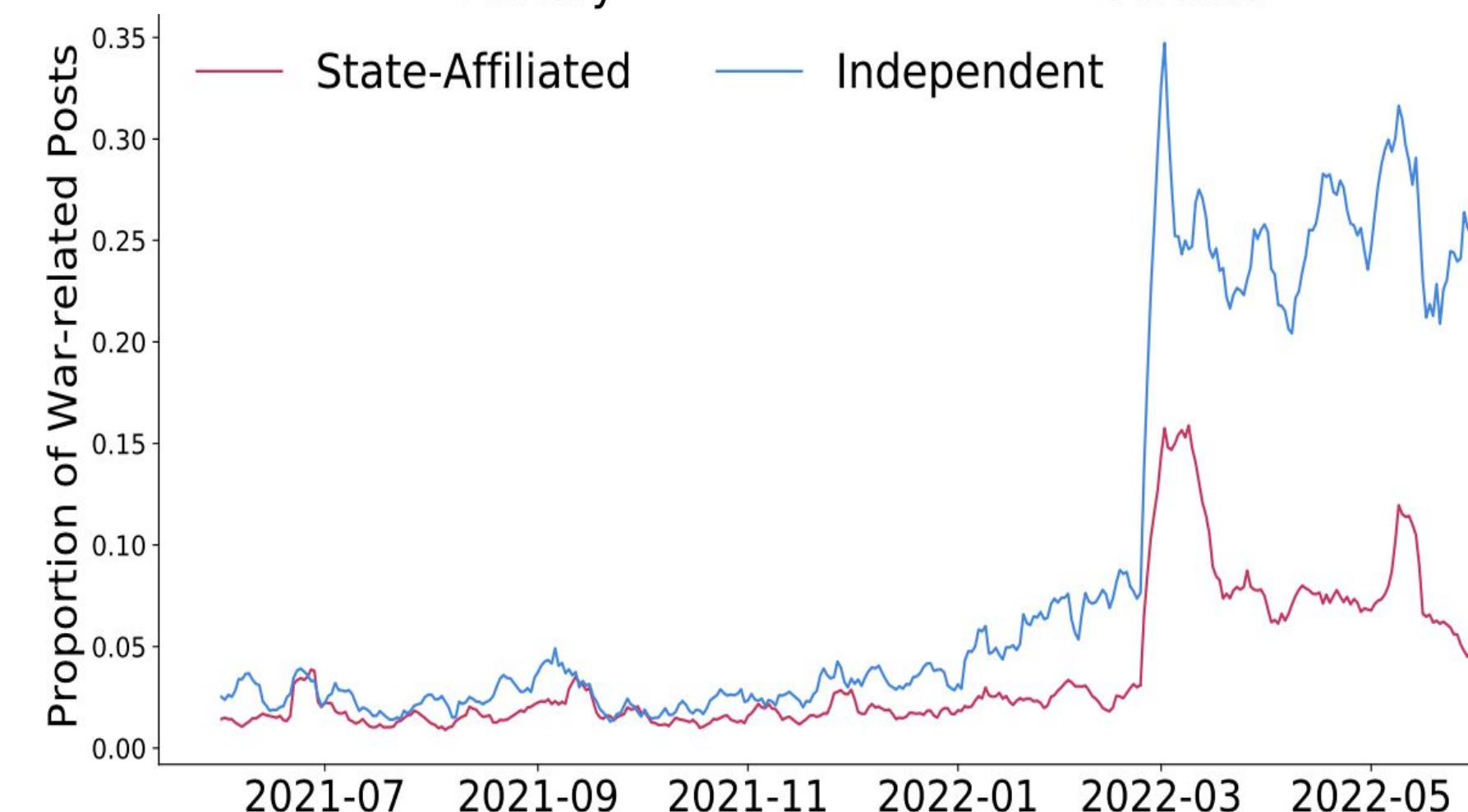
Definition: what topics are discussed influence the importance placed on the topics by audiences

Speculation: state-affiliated media avoids war coverage

NLP tools: Topic models, word stats



Topic models suggest state-affiliated media cover the war more



But independent media use war-related terms more frequently

VoynaSlov Dataset

Post
Лавров заявил о подталкивании Украины со стороны США к провокациям против России
(*Lavrov says US nudging Ukraine toward provocations against Russia*)

Metadata

Time 26 Jan 2022 (pre-war or during war)

Platform (VKontakte or Twitter)

Source (state-affiliated or independent)

72

57

1

8026

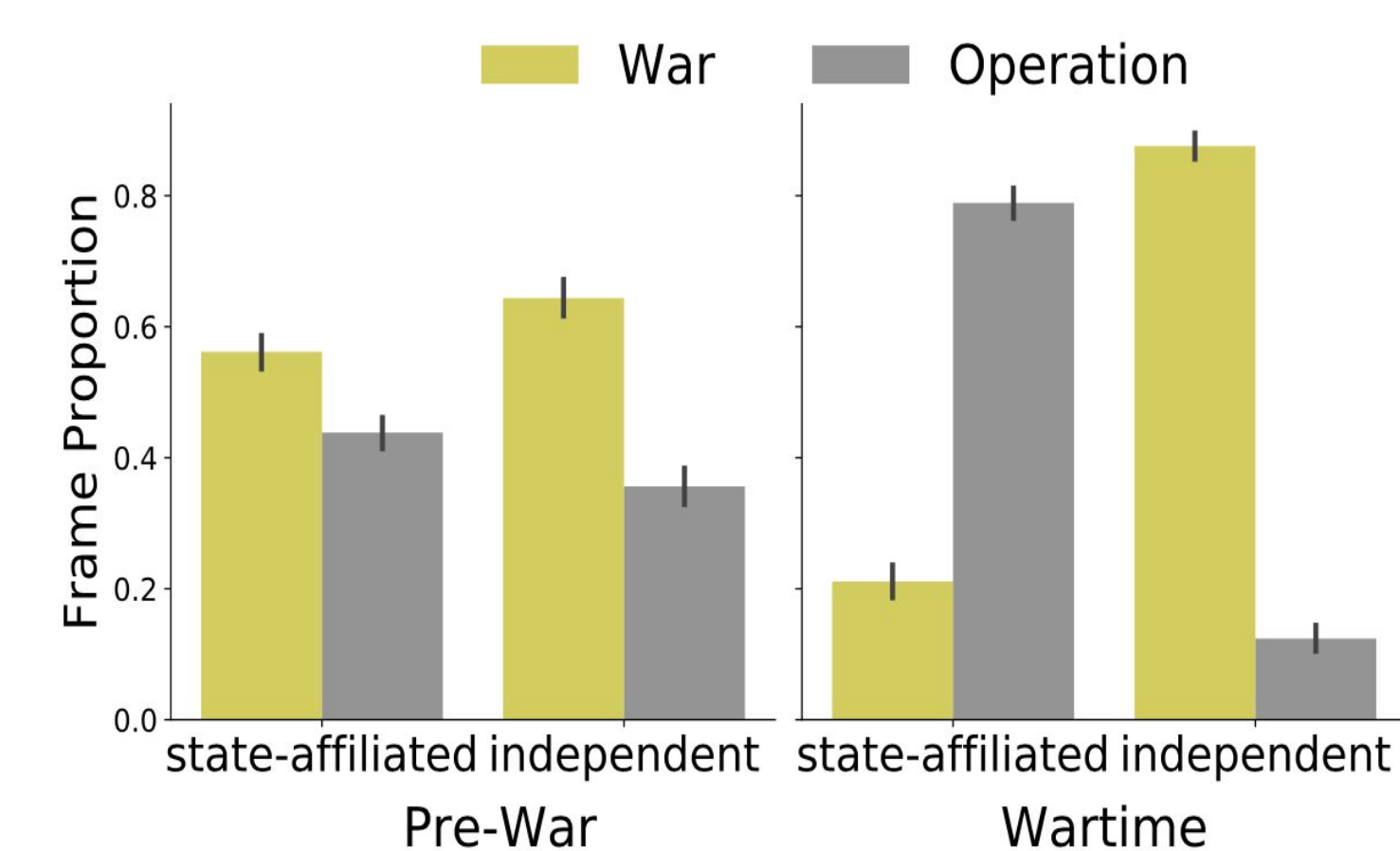
Media effects analysis:
agenda setting, framing, priming

Future directions for NLP research in information manipulation crises

Framing

Definition: how those topics are discussed can influence the way audiences understand them.

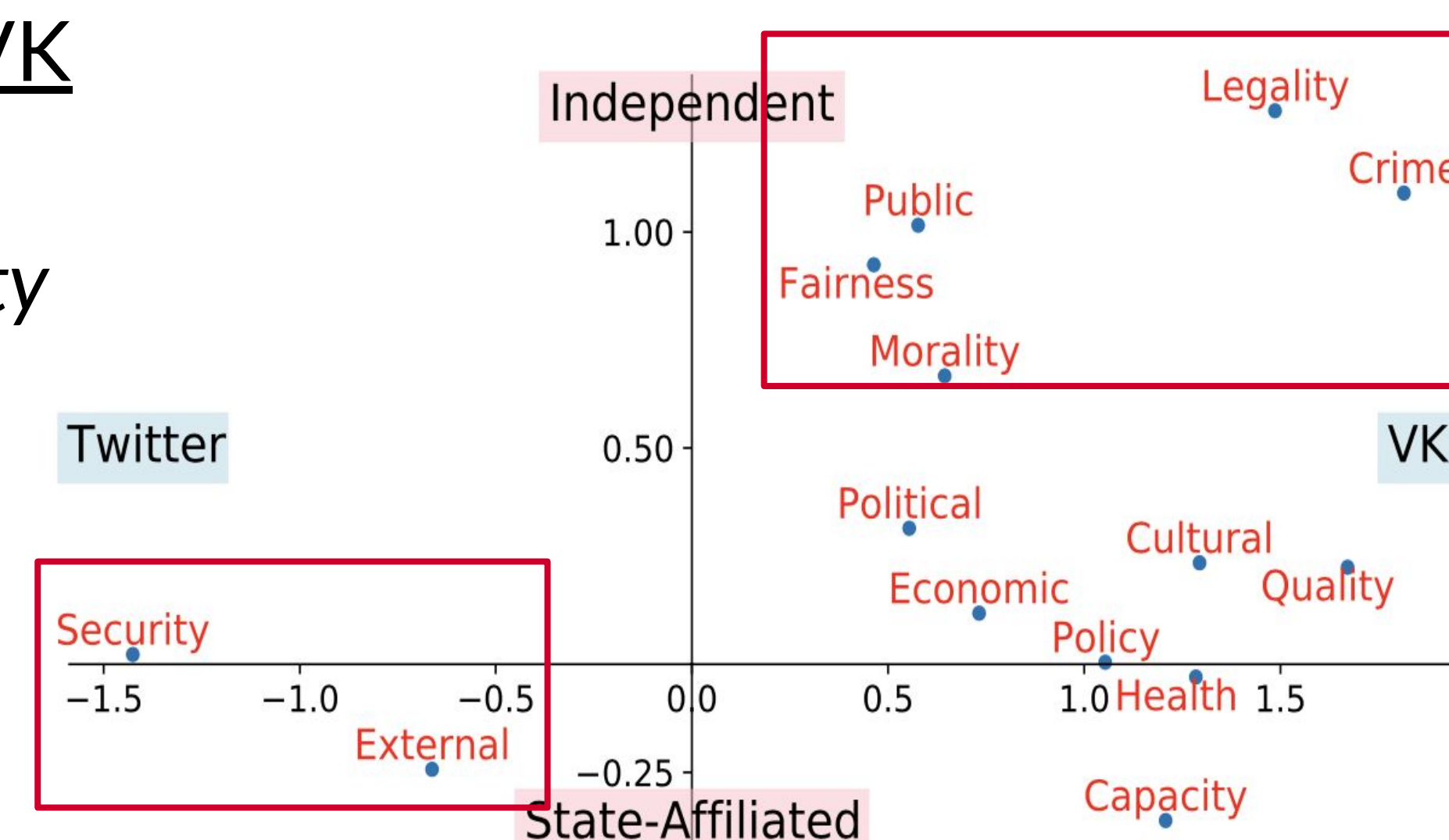
NLP tools: Classifier trained on Media Frames Corpus (MFC), issue-specific frames (“War” vs. “Military Operation”)



Since war began, Independent prefer War, state-affiliated prefer Operation frame

Independent+VK
Legality, Crime, Fairness, Morality

Twitter
External, Security



Paper



Data

Priming

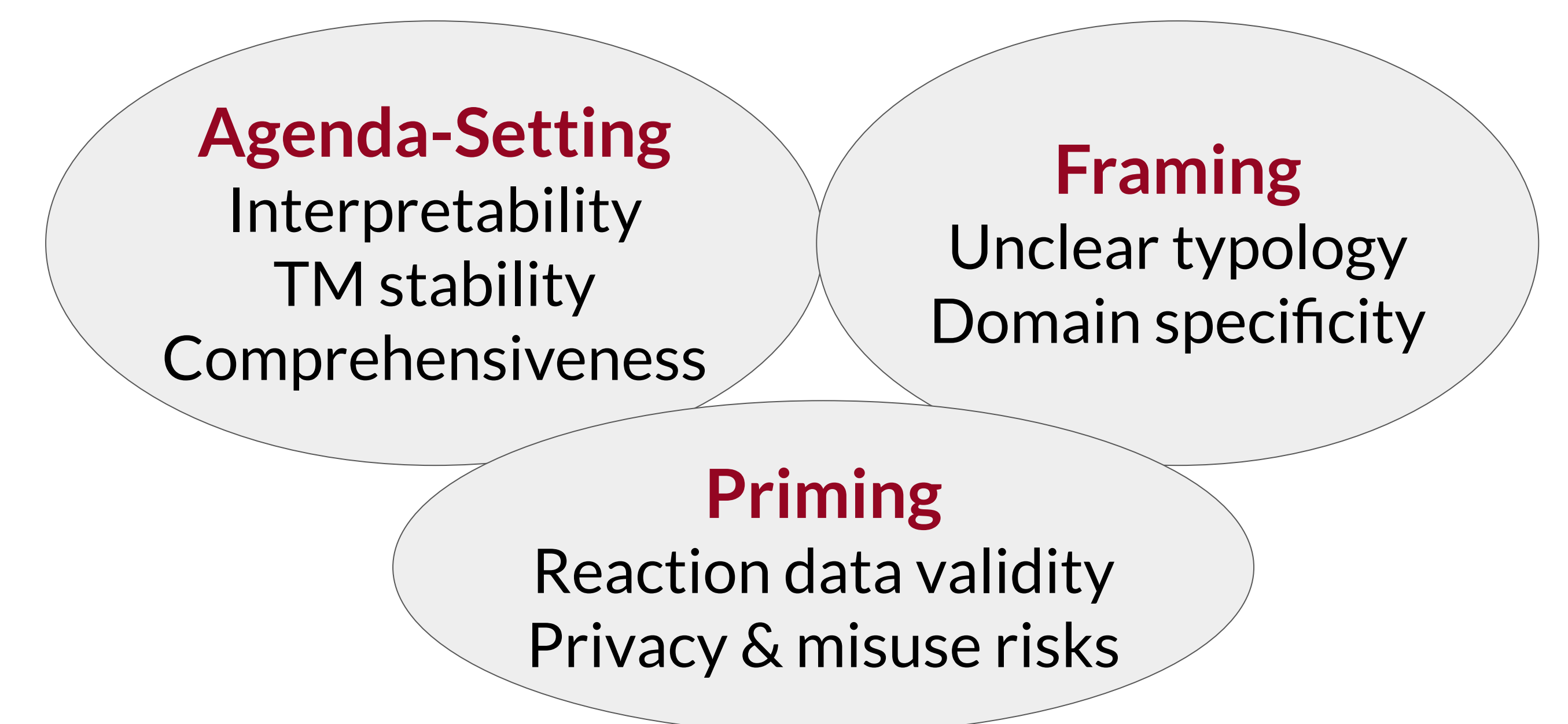
Definition: framing/agenda setting effects on users

NLP tools: user engagement as proxy

- **Civilian-focused frames** (Public Sentiment, Morality) and **multimedia** → higher engagement

Comments on state-affiliated media	Comments on independent media
Economic Quality of Life	Political Public Sentiment Fairness

Open Challenges



We hope our dataset, analysis, and discussion of open challenges for NLP can facilitate the identification and prevention of information manipulation