



Defining, Detecting and Countering Disinformation

Bilyana Lilly

Why Should Western Governments Care?



Presentation Roadmap



Defining disinformation

Disinformation in Russia's and China's view of warfare

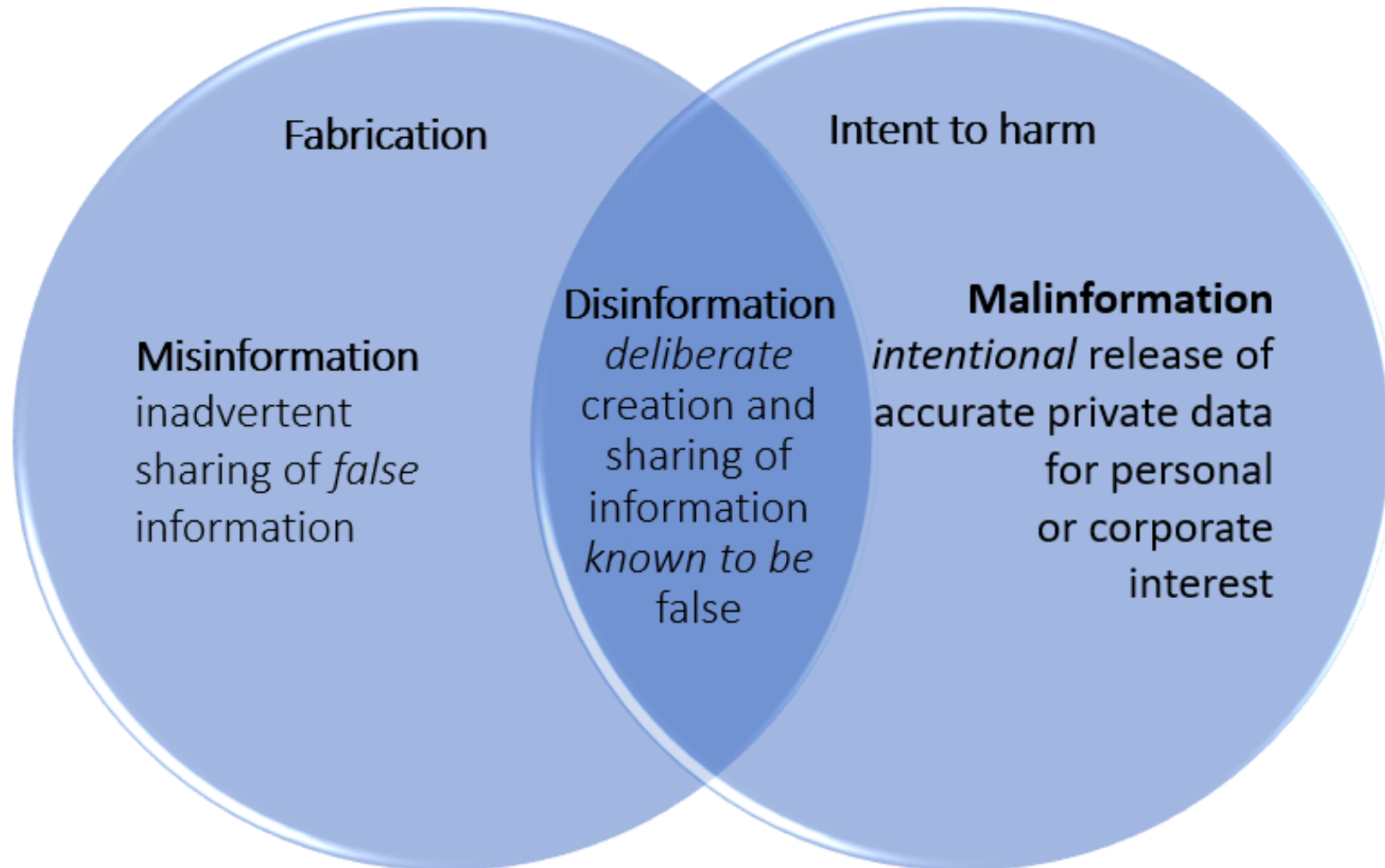
Disinformation case studies

Detecting and countering disinformation: best practices

Exercise: Detecting Disinformation

Defining Disinformation

- The related conceptual space
→
- Component of state-sponsored influence activities and information warfare



Disinformation in Russia's View of Warfare

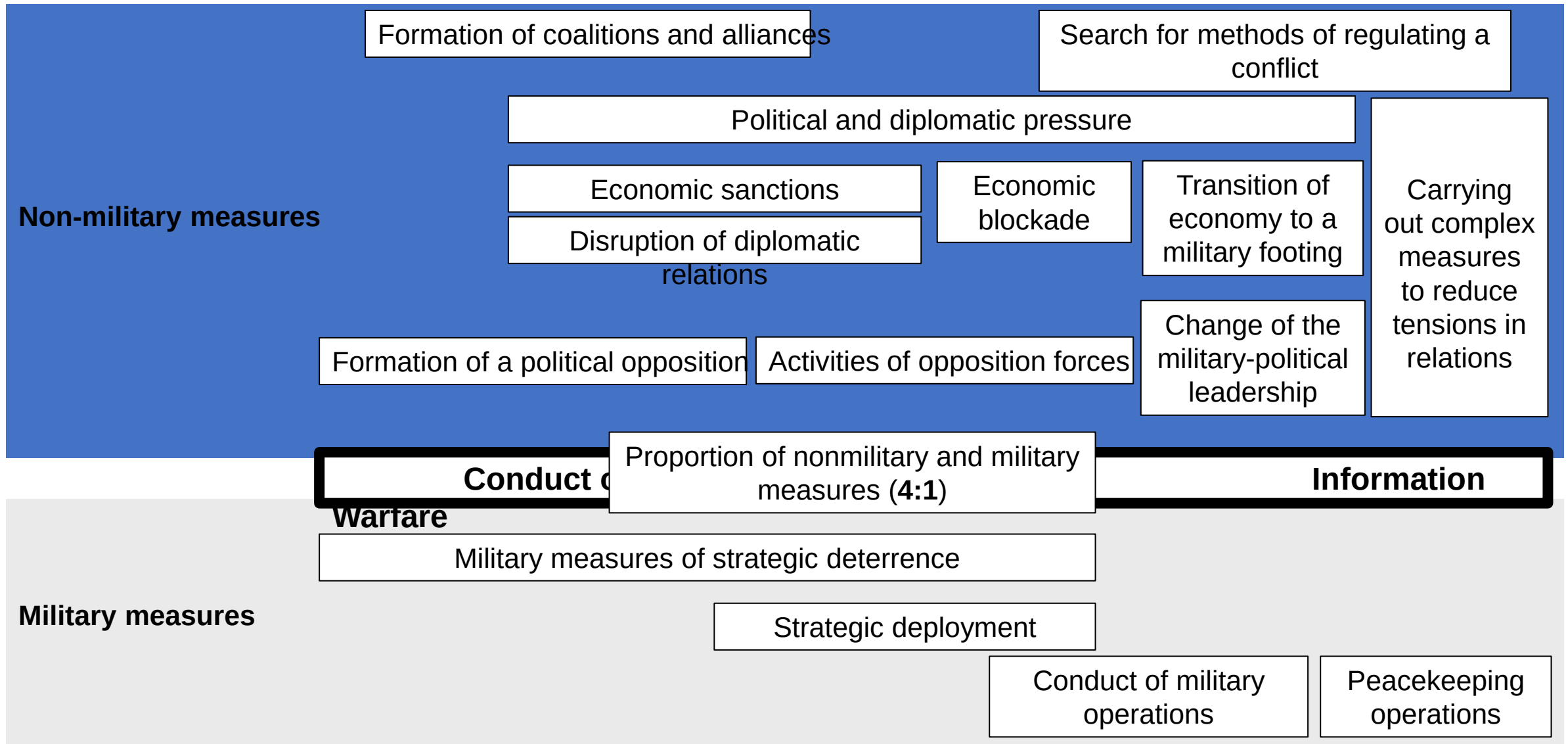
Psychological operations
(including disinformation)

Cyber operations

Information
warfare

Modern warfare

Disinformation in Russia's View of Warfare



Disinformation in China's View of Warfare

Psychological warfare
(including disinformation)



Public opinion warfare



Legal warfare



"Three warfares"
concept

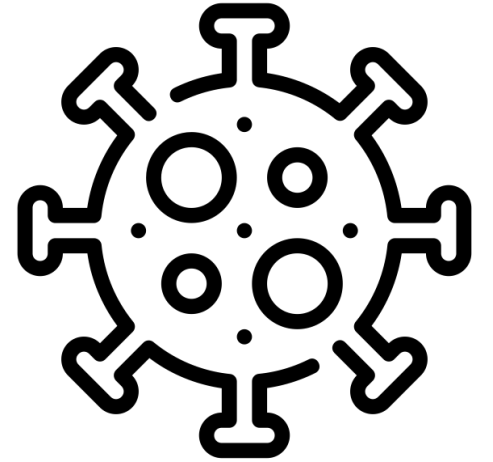
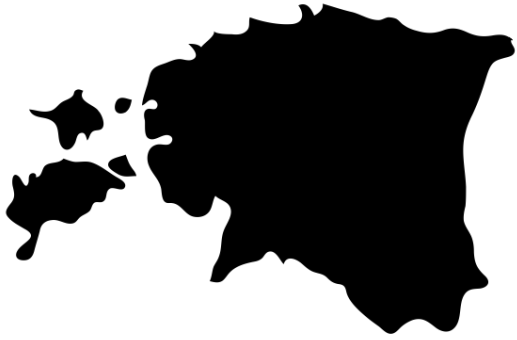




Main Takeaways

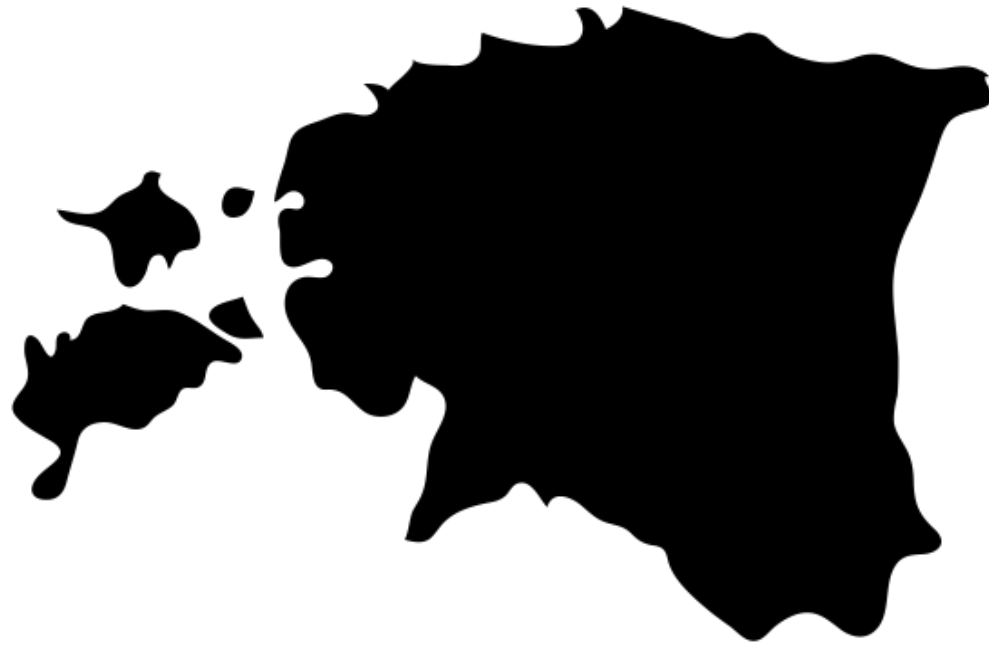


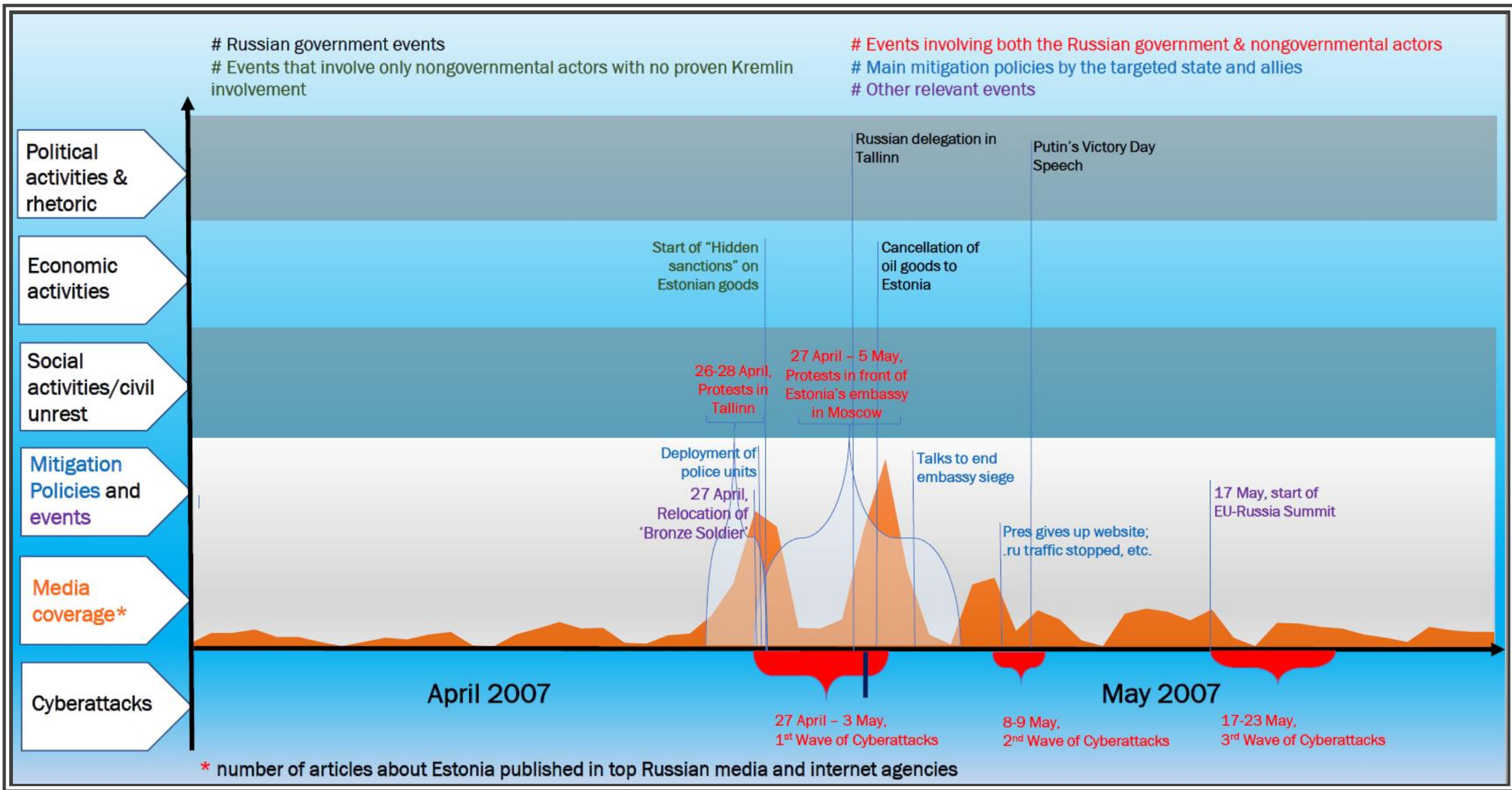
- Disinformation: *fabricated* narratives, *intentionally* constructed or disseminated with *state* support
- Present in Russian and Chinese warfare strategy



Disinformation Case Studies

Estonia 2007





U.S. 2016 & Pizzagate



U.S. 2016 & Pizzagate

Disinformation linked to protests and cyber operations

- Intelligence gathering
- Exploitation of US social media platforms (Twitter, Facebook & Instagram)
 - Spread anti-Hilary and pro-Trump propaganda
 - Spread divisive messages
 - Sponsored political rallies in the US
- Distribution of disinformation through RT and Sputnik
- Cyberattacks on political IT infrastructure and exfiltration of subsequently leaked data, circulated by Russian and Western media



U.S. 2016 & Pizzagate



- Leaked emails from the Clinton campaign adviser John Podesta
- Disinformation: Clinton-run child abuse ring in Comet Ping Pong pizza
- Disseminated through Alex Jones, Twitter, Facebook, Instagram, Yelp
- December 2016, Edgar Welch, convinced from online forums that he had to find the children himself, entered the pizzeria with an rifle

Germany's 2017 Elections and the Lisa Case



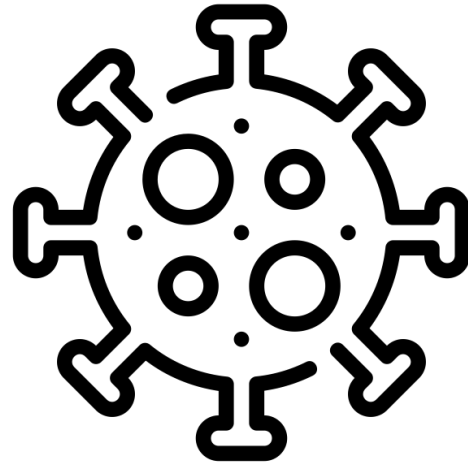
Germany's 2017 Elections and the Lisa case

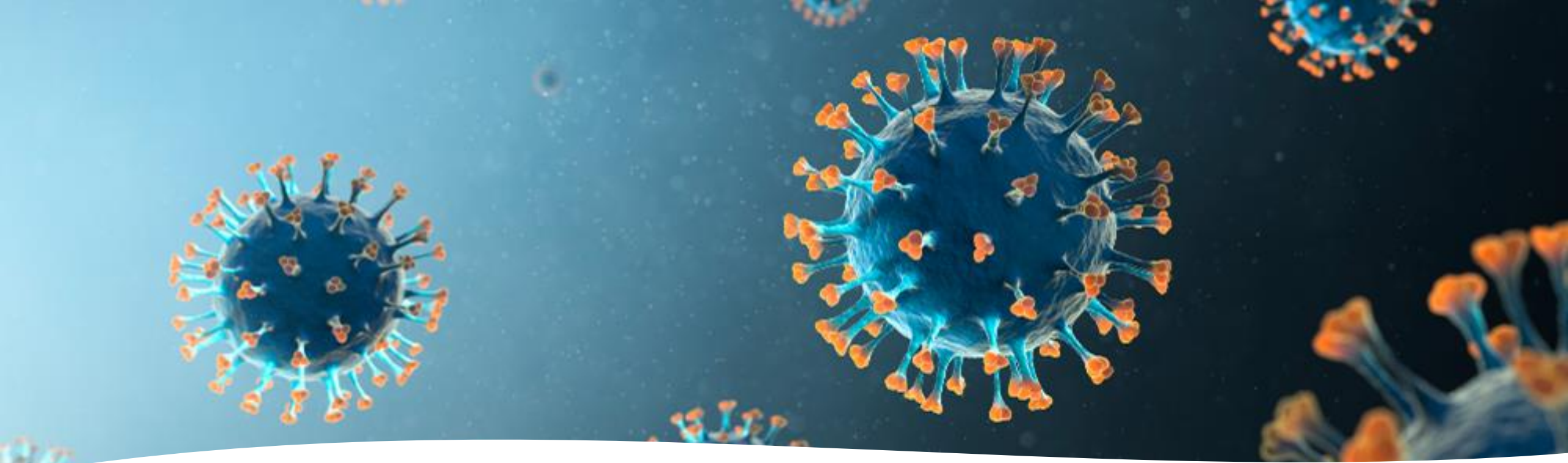
Link between Russian disinformation, protests and Russia's political leadership

- 13-year old Russian-German girl allegedly violated
- German police exposed the story as fake
- Russian-state channels broadcast the story and Russia's senior diplomatic elite supported it
- Demonstrations via Facebook involving representatives of the German-Russian minority and neo-Nazi groups



Covid-19





Disinformation in Practice: Covid-19

- In June 2020, EU publicly accused China and Russia of spreading disinformation
- Disinformation topics: health care hoaxes, false claims, online scams, hate speech and COVID-19 coronavirus conspiracy theories
- Channels: RT, Sputnik, Twitter, Facebook

Disinformation: Channels, Actors and Amplifiers

Channels



- Media:
 - Russia: Russia Today (RT), Sputnik, Russian domestic media, radio, news agencies
 - China: CCTV International rebranded in 2016 as China Global Television (CGTV), China Radio International (CRI), Xinhua News Service, China Daily
- Social media:
 - Russia: VKontakte, Telegram
 - China: Wechat, Weibo
 - International: Twitter, Facebook, Instagram

Actors and amplifiers:



- Security services and military: FSB, GRU, PLA
- Government officials
- Unattributed groups: Internet Research Agency (IRA), trolls, bots
- Unwitting actors: social media users, journalists



Main Takeaways



- Disinformation:
 - Used for various purposes against various countries
 - Disseminated through a variety of channels, actors and amplifiers
- Objectives:
 - exploit existing issues to erode social cohesion
 - undermine trust in public institutions and erode democratic governance
 - undermine decision-making capacity

Disinformation: Best Practices

- Government and international organizations initiatives:
 - U.S.
 - Countering Foreign Influence Task Force (Department of Homeland Security)
 - Global Engagement Center (Department of State)
 - EU initiatives
 - Joint Framework on countering hybrid threats
 - Code of Practice signed by Facebook, Google, Twitter and I
- Non-profit organizations:
 - Teaching media literacy
 - Eurasia Foundation, Balkan Security Network
 - Fact-checking websites
 - Factcheck.org, factchecker, Stop Fake, EUvsFisinfo, Correctiv
- Social media companies
 - Twitter, Facebook, Youtube suspend accounts

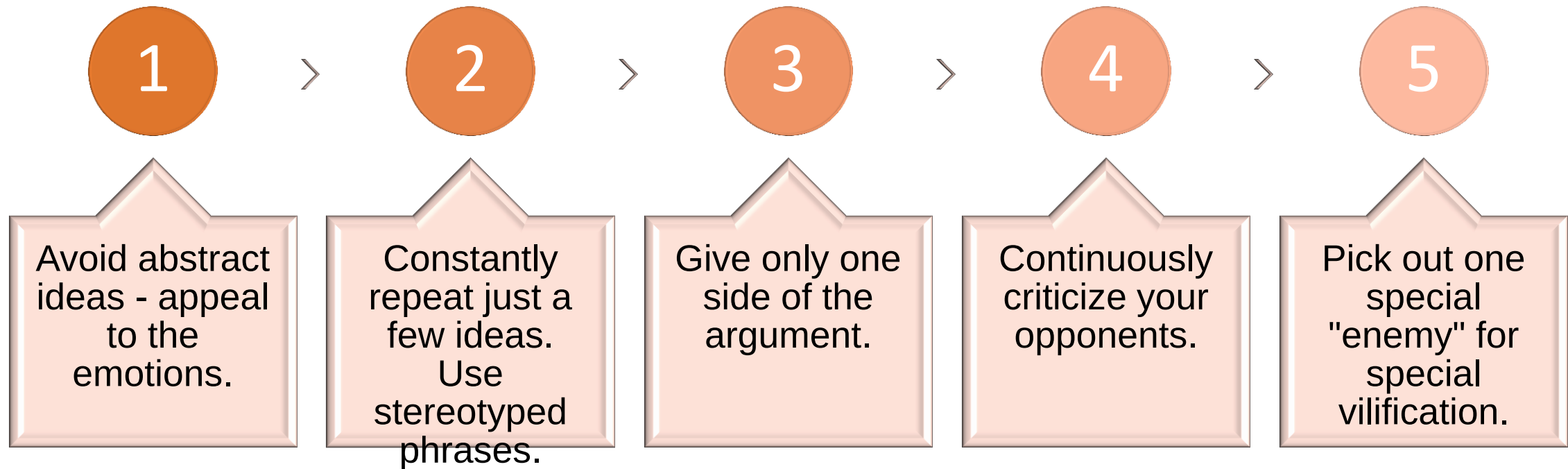




Practical Exercise: Detect Disinformation

- **Headline:** Does it match the content of the article?
- **URL:** Is the source legitimate?
- **Content:** Is it informative and fact-based or emotional?
- **Sources:** Are sources referenced in the text?
- **Images:** Does the image reflect the story? What does it aim to convey?
- **Author:** Is the author reputable?
- **Engagement:** Shares do not equate to truth.

What makes disinformation effective?



Your Turn: Disinformation or Not?



- Link to the article:
<https://pl.sputniknews.com/opinie/2020060912551003-swiat-oczami-polaka-z-przedmiesc-paryza-cenzura-prewencyjna-mediow-spolecznościowych-Sputnik/>

Practical Exercise: Detect Bots

Profile picture: Is it genuine?

Activity: Is the account highly active?

Name: Is the username genuine?

Creation date: Was the account recently created?

Language: Are there translation errors?

Information: Is there personal information?

Engagement: Does the account have real followers?



Your Turn: Bot or Not?

Profile picture: Is it genuine?

Activity: Is the account highly active?

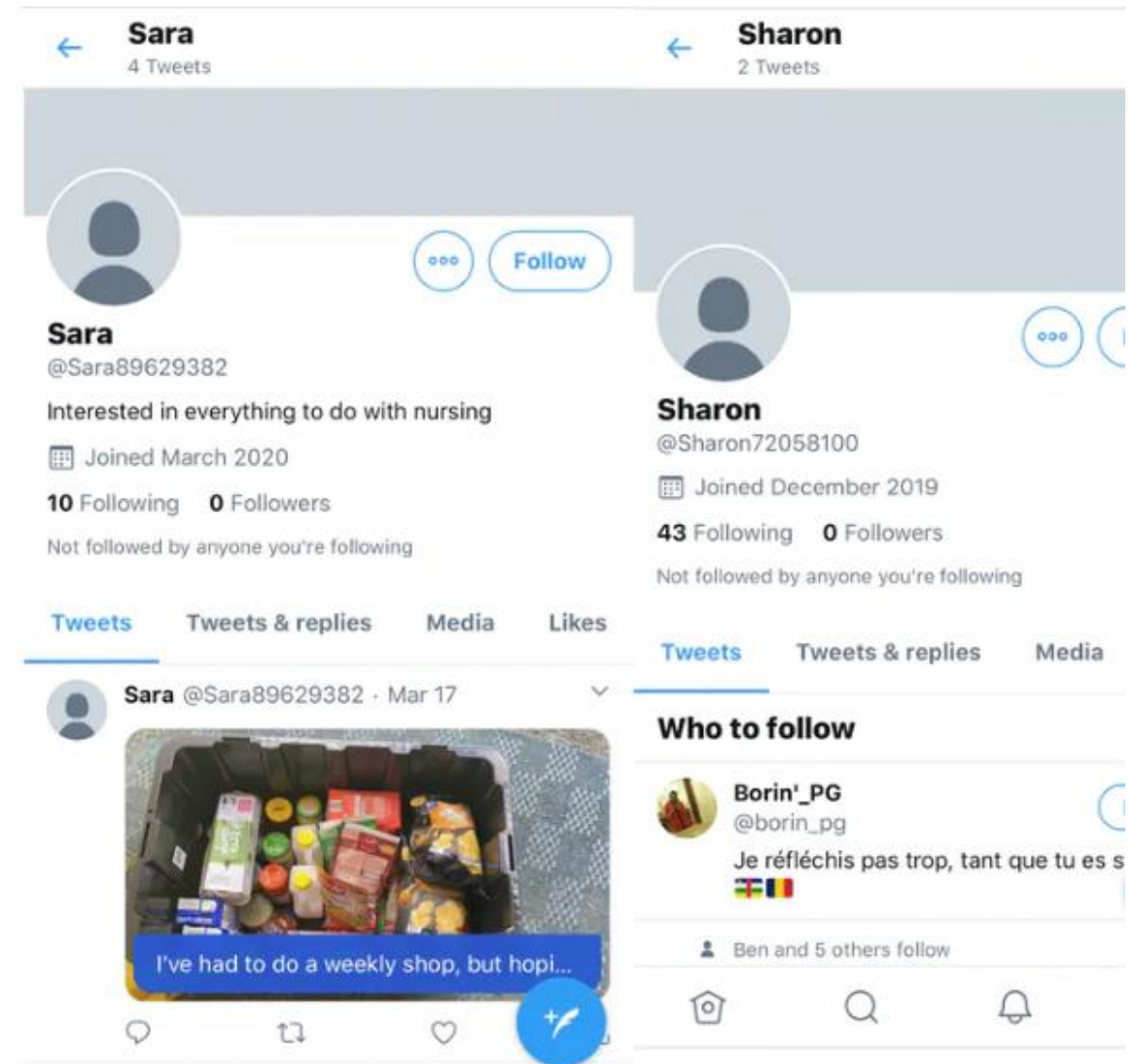
Name: Is the username genuine?

Creation date: Was the account recently created?

Language: Are there translation errors?

Information: Is there personal information?

Engagement: Does the account have real followers?



Detecting Russian Trolls on Social Media

- Existing research:
 - Factors: content of tweets (themes, stance, emotion, tweet length), changes in accounts' behavior, time zone, geolocation
 - Techniques: human coding, machine learning techniques, text analysis
- Our approach:
 - Linguistic-based machine learning model
 - Databases: IRA trolls dataset, 2015 – 2017, ~2mil tweets
 - 98% accuracy in distinguishing Russian trolls from non-trolls



Thank you for your attention

E-mail: bilyanalilly@gmail.com

Twitter: [@bilyanalilly](https://twitter.com/bilyanalilly)



BACKUP SLIDES

DON'T GET TRICKED BY ONLINE MISINFORMATION

Remember these checks when browsing social media

Source

Look at what lies beneath. Check the about page of a website or account, look at any account info and search for names or usernames.

History

Does this source have an agenda? Find out what subjects it regularly covers or if it promotes only one perspective.

Evidence

Explore the details of a claim or meme and find out if it is backed up by reliable evidence from elsewhere.

Emotion

Does the source rely on emotion to make a point? Check for sensational, inflammatory and divisive language.

Pictures

Pictures paint a thousand words. Identify what message an image is portraying and whether the source is using images to get attention.

Think **SHEEP** before you share