



# Workshop Agenda

- 09:00 – 10:50 Part I
  - 09:00 – 09:05 Welcome
  - 09:05 – 09:40 Introduction into Visual Disinformation and Internet Memes
  - 09:40 – 10:50 Hands-on: classification of IWT memes
- 10:50 – 11:10 Break
- 11:10 – 13:00 Part II



Schweizerische Eidgenossenschaft  
Confédération suisse  
Confederazione Svizzera  
Confederaziun svizra

Eidgenössisches Departement für Verteidigung,  
Bevölkerungsschutz und Sport VBS  
**armasuisse**  
Wissenschaft und Technologie



# PERSPEKTIVEN

## Introduction into Visual Disinformation and Internet Memes

### Raphael Meier, Marco Willi, Michael Graber

# Lausanne, 27th March 2022

# Disinformation is not a new phenomena

INSS  
INSTITUTE FOR NATIONAL STRATEGIC STUDIES

STRATEGIC PERSPECTIVES 11

**Deception, Disinformation, and Strategic Communications: How One Interagency Group Made a Major Difference**

by Fletcher Schoen and Christopher J. Lamb



Center for Strategic Research  
Institute for National Strategic Studies  
National Defense University

Source: [Operation INFEKTION - Wikipedia](#)



# Conflict between Ukraine and Russia (2022)

## Ukraine: Russia Mobilises Disinfo Alongside Military

Could a sudden increase in fake news be a precursor to real-world escalation?

Source: [Ukraine: Russia Mobilises Disinfo Alongside Military | Institute for War and Peace Reporting \(iowpr.net\)](#)

17 Feb, 2022 12:24 / Home / Russia & FSU

**Western media destroyed their reputation with fake Ukraine news – Moscow**

Source: [Western media destroyed their reputation with fake Ukraine news – Moscow — RT Russia & Former Soviet Union](#)

**cnn politics** The Biden Presidency Facts First US Elections

**US cranks up the information warfare against Putin**



Analysis by [Stephen Collinson](#), CNN

Updated 1420 GMT (2220 HKT) February 17, 2022

### Information warfare expert says the U.S. is finally countering Russia at its own game

BY HALEY OTT

UPDATED ON: FEBRUARY 17, 2022 / 11:52 AM / CBS NEWS



Source: [Information warfare expert in Ukraine says the U.S. is finally countering Russia at its own game - CBS News](#)



# Conflict between Ukraine and Russia (2022)

## Ukraine: Russia Mobilises Disinfo Alongside Military

Could a sudden increase in fake news be a precursor to real-world escalation?

Source: [Ukraine: Russia Mobilises Disinfo Alongside Military | Institute for War and Peace Reporting \(iwr.net\)](#)

17 Feb, 2022 12:24 / Home / Russia & FSU

**Western media destroyed their reputation with fake Ukraine news – Moscow**

Source: [Western media destroyed their reputation with fake Ukraine news – Moscow — RT Russia & Former Soviet Union](#)

**cnn politics** The Biden Presidency Facts First US Elections

**US cranks up the information warfare against Putin**



Analysis by [Stephen Collinson](#), CNN

Updated 1420 GMT (2220 HKT) February 17, 2022

**Information warfare expert says the U.S. is finally countering Russia at its own game**

BY HALEY OTT

UPDATED ON: FEBRUARY 17, 2022 / 11:52 AM / CBS NEWS



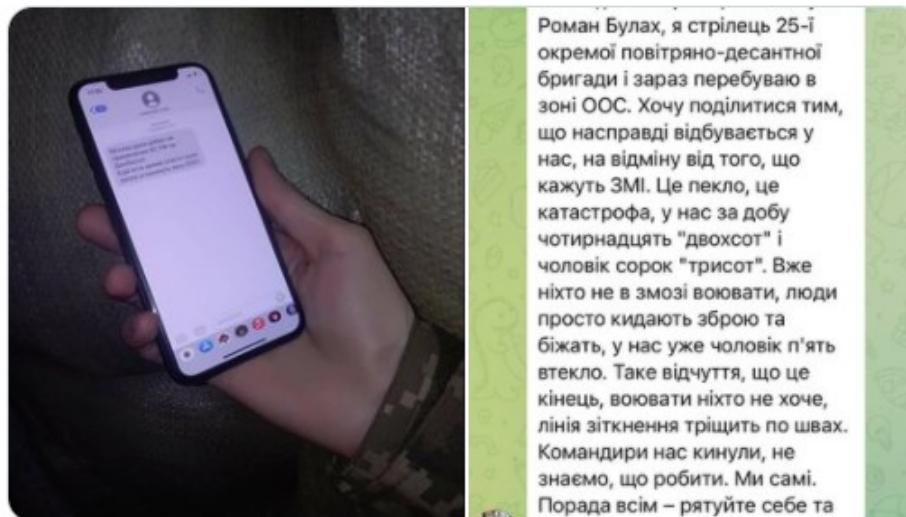
Source: [Information warfare expert in Ukraine says the U.S. is finally countering Russia at its own game - CBS News](#)

# Conflict between Ukraine and Russia (2022)



Olga Tokariuk   
@olgatokariuk

Psychological warfare in action: Ukrainian soldiers on the frontline start getting text messages telling them that Russian troops have an order to advance and advising them to flee/surrender. On Telegram, fake 'Ukrainian soldiers' write in broken Ukrainian about demoralized army



2:13 nachm. · 23. Feb. 2022 · Twitter Web App

520 Retweets 62 Zitierte Tweets 974 „Gefällt mir“-Angaben

Source: [Olga Tokariuk auf Twitter: "Psychological warfare in action: Ukrainian soldiers on the frontline start getting text messages telling them that Russian troops have an order to advance and advising them to flee/surrender. On Telegram, fake 'Ukrainian soldiers' write in broken Ukrainian about demoralized army https://t.co/GwCdgEZWrn" / Twitter](https://twitter.com/olgatokariuk/status/1562500000000000000)



Ukraine / Україна   
@Ukraine

## Types of Headaches

### Migraine



### Hypertension



### Stress



### Living next to Russia



2:09 nachm. · 7. Dez. 2021 · Twitter for iPhone

144.072 Retweets 42.950 Zitierte Tweets 697.316 „Gefällt mir“-Angaben

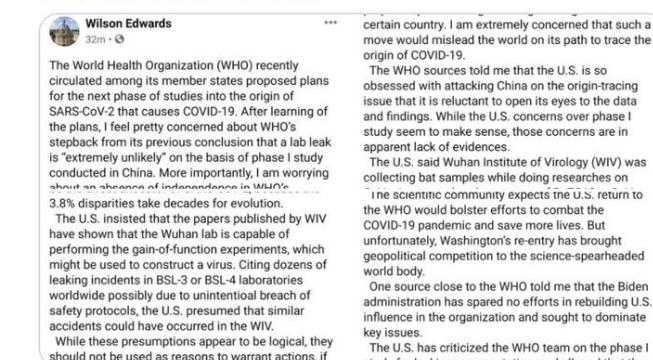
# «Wilson Edwards» (Chinese Info Op)

 WilsonEdwards  
7 Tweets

**WilsonEdwards**  
@WilsonE84957402  
⌚ Switzerland 📅 Joined July 2021  
**12 Following 23 Followers**  
Followed by Zhang Heqing张和清

[Tweets](#) [Tweets & replies](#) [Media](#) [Likes](#)

**WilsonEdwards** @WilsonE84957402 · Jul 25  
#COVID19 #COVID19nsw @WHO @DrTedros @FauciFan As scientific issues continue to be politicized, I have no doubt that this advisory group would be succumbed to a political tool.  
[m.facebook.com/story.php?stor...](https://m.facebook.com/story.php?stor...)

  
The World Health Organization (WHO) recently circulated among its member states proposed plans for the next phase of studies into the origin of SARS-CoV-2 that causes COVID-19. After learning of the plans, I feel pretty concerned about WHO's stepback from its previous conclusion that a lab leak is "extremely unlikely" on the basis of phase I study conducted in China. More importantly, I am worrying about an absence of independence in WHO's 3.8% disparities take decades for evolution.  
The U.S. insisted that the papers published by WIV have shown that the Wuhan lab is capable of performing the gain-of-function experiments, which might be used to construct a virus. Citing dozens of leaking incidents in BSL-3 or BSL-4 laboratories worldwide possibly due to unintentional breach of safety protocols, the U.S. presumed that similar accidents could have occurred in the WIV.  
While these presumptions appear to be logical, they should not be used as reasons to warrant actions, if

• 10 116 215

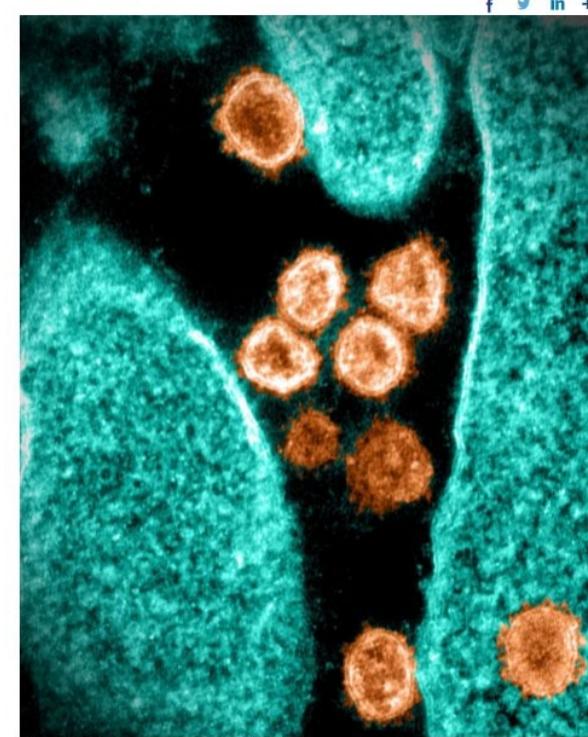
 CHINADAILY.com.cn | Global Edition | Aug 10, 2021

HOME OPINION VIDEO WORLD CHINA TECHNOLOGY BUSINESS CULTURE TRAVEL

• Newsmakers

COVID-19 origin tracing: Claim emerges of 'intimidation' from the US

CGTN | Updated: 2021-07-30 14:42



SARS-CoV-2, the virus that causes COVID-19. [Photo/Agencies]

A European biologist has come to the fore with a startling claim that a World Health Organization (WHO) advisory group on tracing the origins of pathogens, including the virus responsible for the COVID-19 pandemic, would become a "political tool."

Bearing the name **Wilson Edwards**, the person took to social media, including Facebook and Twitter, on July 24 to describe how much he was worried about

# «Wilson Edwards» (Chinese Info Op)



Embassy of Switzerland in Beijing   
@SwissEmbChina

...

Looking for Wilson Edwards, alleged  biologist, cited in press and social media in China over the last several days. If you exist, we would like to meet you! But it is more likely that this is a fake news, and we call on the Chinese press and netizens to take down the posts.

In the last several days, a large number of press articles and social media posts citing an alleged Swiss biologist have been published in China. While we appreciate the attention on our country, the Embassy of Switzerland must unfortunately inform the Chinese public that this news is false.

1. There is no registry of any Swiss citizen with the name "Wilson Edwards".
2. There are no academic articles in the biology field cited under this name.
3. The Facebook account cited as having published his commentary was only opened on 24 July, 2021, and has only posted this one post so far. It only has 3 friends. It is likely that this Facebook account was not opened for social networking purposes.

While we assume that the spreading of this story was done in good faith by the media and netizens, we kindly ask that anyone having published this story take it down and publish a corrigendum.

过去几天内，多家媒体发布了有关一位所谓的瑞士生物学家的新闻。我们感谢大家对我们国家的关注，但瑞士驻华大使馆必须遗憾地指出，这是一条错误的新闻。

1. 瑞士没有任何登记姓名为“Wilson Edwards”的公民。
2. 在生物学界没有以该名字署名的学术文章。
3. 发表评论的脸书账户在2021年7月24日刚刚开通，至今仅发帖一条，账户好友只有三位。该账户可能并非以网络社交而开设。

媒体和网民或许是无意间转发了该新闻，但我们恳请您即刻删除并刊登更正声明。

12:34 nachm. - 10. Aug. 2021 - Twitter Web App

4.253 Retweets 1.135 Zitierte Tweets 10.125 „Gefällt mir“-Angaben



Emilie  
@paintsandsings

...

Antwort an @SwissEmbChina

When you need a biologist , just find one and use his name in reverse



en.m.wikipedia.org  
E. O. Wilson - Wikipedia

2:23 nachm. - 10. Aug. 2021 - Twitter for iPad

4 Retweets 3 Zitierte Tweets 99 „Gefällt mir“-Angaben





# «Wilson Edwards» (Chinese Info Op)

## IN DEPTH RESEARCH & ANALYSIS

THE SWISS BIOLOGIST THAT NEVER WAS

A Chinese influence operation focused on COVID-19

*By Ben Nimmo, Global IO Threat Intelligence Lead and the IO Threat Intelligence Team*

### EXECUTIVE SUMMARY:

On July 24, 2021, someone posing as a Swiss biologist named Wilson Edwards claimed, on Facebook and Twitter, that the United States was putting pressure on World Health Organization scientists studying the origins of COVID-19 in an attempt to blame the virus on China. Within 48 hours, hundreds of social media accounts around the world had picked up on the story. Within a week, Chinese state media including the [Global Times](#) and [People's Daily](#) were running headlines about the alleged US “intimidation.”

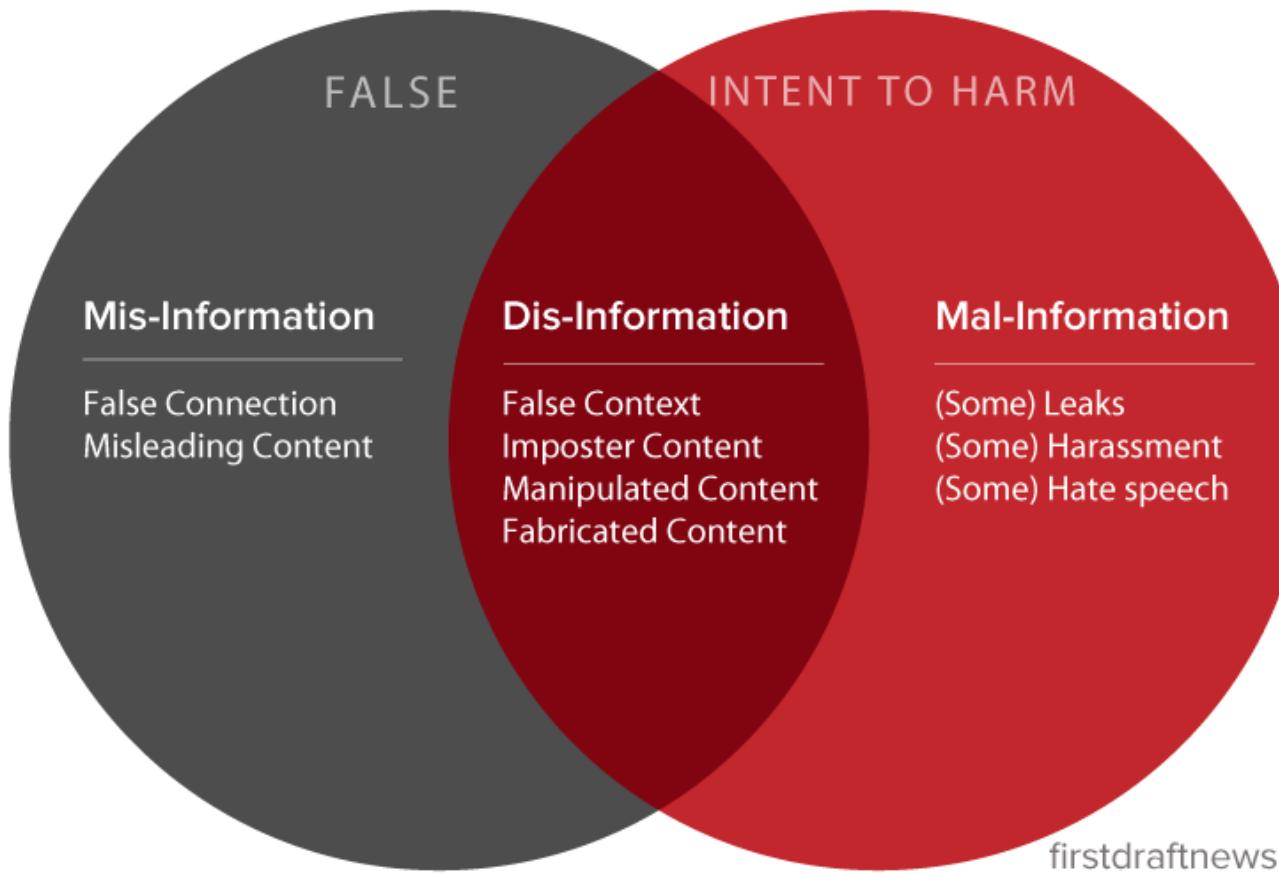
On August 10, the Swiss Embassy in Beijing [announced](#) that there was no record of any Swiss citizen by that name. The same day, we investigated and removed the Facebook account as fake. It had been created on July 24, less than 12 hours before it started posting about the coronavirus pandemic.

In essence, this campaign was a hall of mirrors, endlessly reflecting a single fake persona. Our investigation uncovered that almost the entire initial spread of the “Wilson Edwards” story on our platform was inauthentic — the work of a multi-pronged, largely unsuccessful influence operation that originated in China. The operation brought together the original fake account, several hundred additional inauthentic accounts and a cluster of authentic accounts, including those that belonged to employees of Chinese state infrastructure companies across four continents. Outside these clusters, only a handful of real people engaged with the operation’s content.

Source: [November-2021-CIB-Report.pdf \(fb.com\)](#)

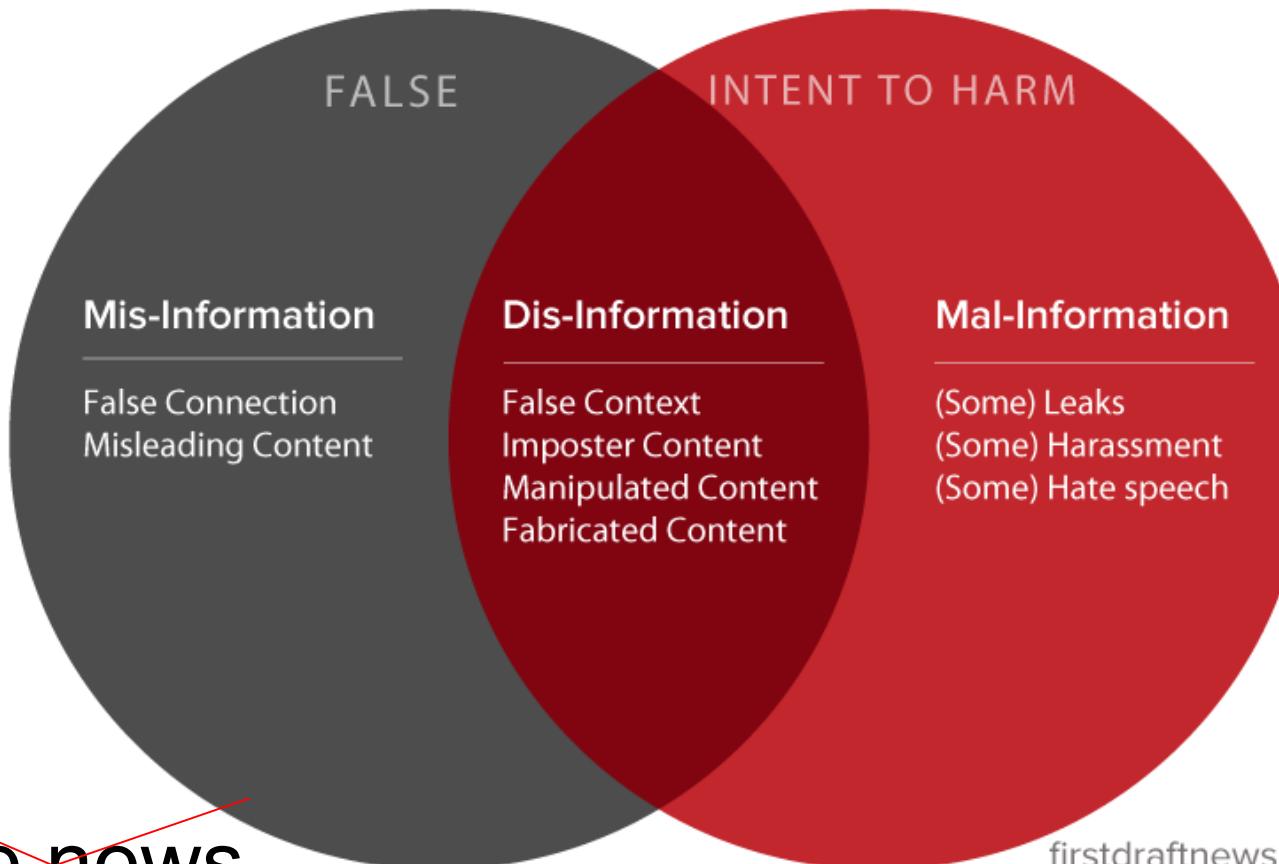
# Disinformation

## INFORMATION DISORDER



# Disinformation

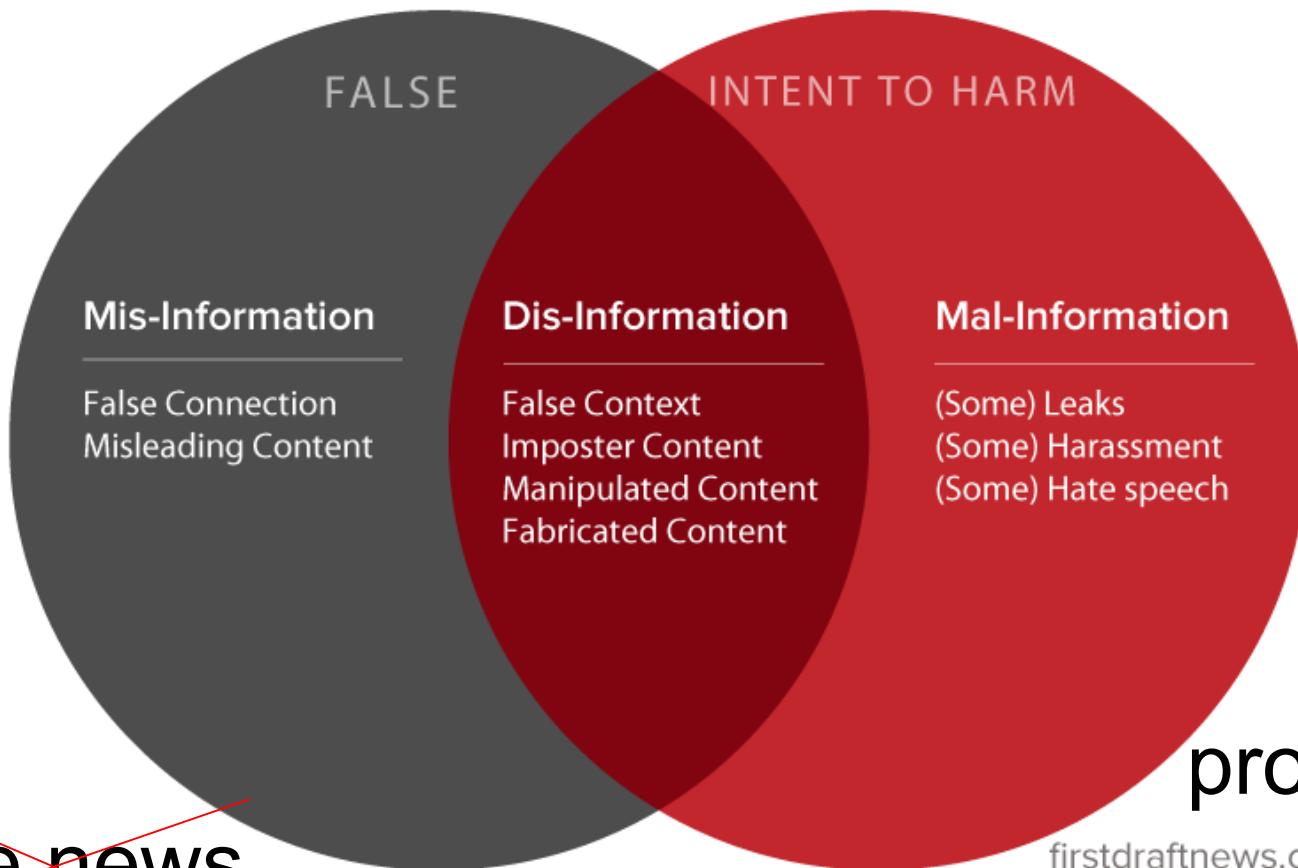
## INFORMATION DISORDER



~~fake news~~

[firstdraftnews.org](http://firstdraftnews.org)

## INFORMATION DISORDER



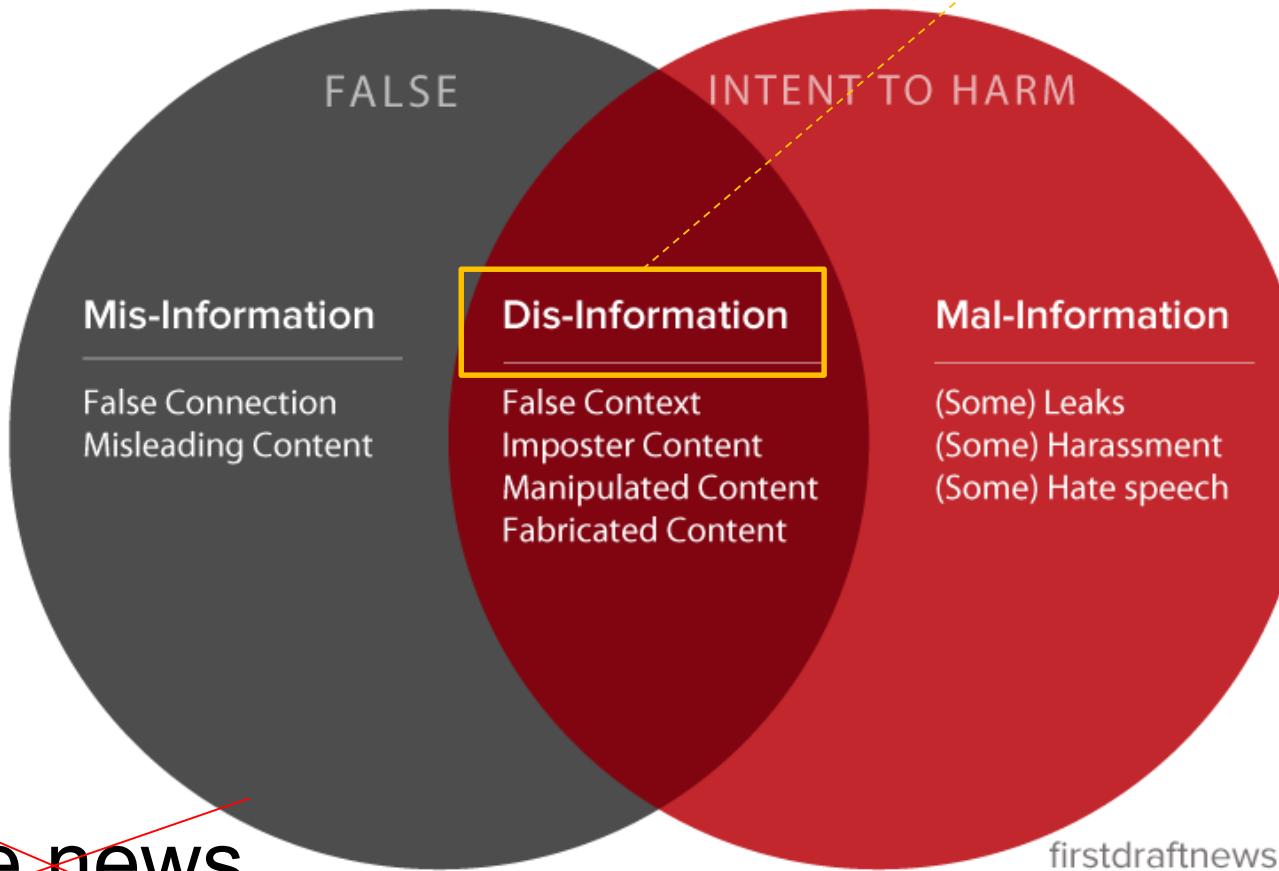
~~fake news~~

[firstdraftnews.org](http://firstdraftnews.org)

# Disinformation

- Half-truths
- Spread by «unwitting crowds»

## INFORMATION DISORDER



[firstdraftnews.org](http://firstdraftnews.org)



# Target group of Disinformation?



Population living in Burmese countryside

Source: [Burmese govt fake images to frame Rohingya for violence \(france24.com\)](https://factcheck.afp.com/burmese-govt-fake-images-to-frame-rohingya-for-violence-france24-com)



Supporter of president Buhari  
Source: <https://factcheck.afp.com/>



# Why is disinformation so prevalent right now? (our hypothesis)

## «Ideal environment»

### Technological Factors

### Social Factors

### Economic Factors

- Internet
  - Mobile devices
  - Social media
- Destabilization
  - Profilicity
  - Post-truth society
- Attention economy
  - Surveillance capitalism

«Democratization»

«Collective sensemaking»

«Emotion > Evidence»

McIntyre, Lee. *Post-truth*. Mit Press, 2018.

Georg Franck: Ökonomie der Aufmerksamkeit. Ein Entwurf. Dtv, München 2007, ISBN 978-3-423-34401-2 (EA München 1998).

Moeller, Hans-Georg, and Paul J. D'Ambrosio. "Sincerity, authenticity and profilicity: Notes on the problem, a vocabulary and a history of identity." *Philosophy & Social Criticism* 45.5 (2019): 575-596.

Zuboff, Shoshana (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: PublicAffairs.

# Information Operation?

- Definition conundrum (Brangetto et al. 2016)
  - *[...] a lack of consensus when it comes to defining all the elements that make up the strategic application of power in the information domain."*
- Conceptual and definitional conundrum (Cordey 2019)
  - *The first main conclusion of this study is that the conceptual framework around influence-related activities, and particularly around related terms in cyberspace [...], remains unclear and to some extent even incoherent with a certain amount of overlap."*
- **information = influence (used as synonym)**

Sources: [CyCon\\_2016\\_book\\_sisu.indd \(ccdcoe.org\)](http://ccdcoe.org), [Cyber-Reports-2019-10-CyberInfluence.pdf \(ethz.ch\)](https://www.ethz.ch/content/dam/ethz/ethz-website-prod/content/uni/bulletin/2019/10/cyber-influence-report-2019.pdf)



# Information Operation

- Meta 2021
  - “**Coordinated efforts to manipulate or corrupt public debate for a strategic goal.**”
- → Purpose of information operations
  - **Short-term goal** (e.g. influence public opinion during election)
  - **Long-term goal** (e.g. distort information space, paralyze decision-making, undermine public discourse and democratic institutions, make people loose their understanding of what is true and false)

# Information operations: components

Intent  
(e.g. create strategic confusion  
to undermine democracy)

+

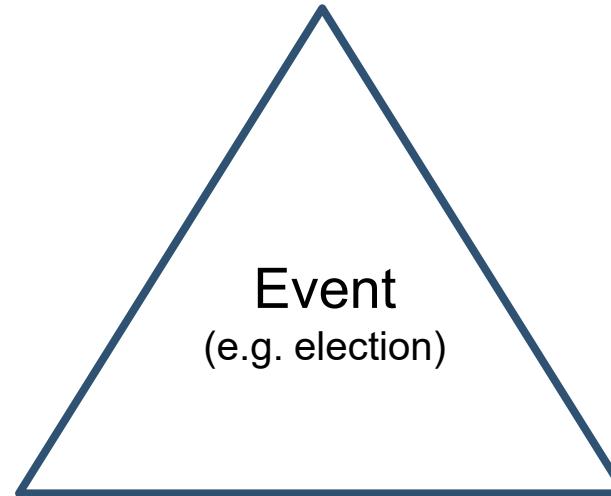
Strategic narrative  
(e.g. 'X are the protectors of  
conservative values')

# Information operations: components

Intent  
(e.g. create strategic confusion to undermine democracy)

+

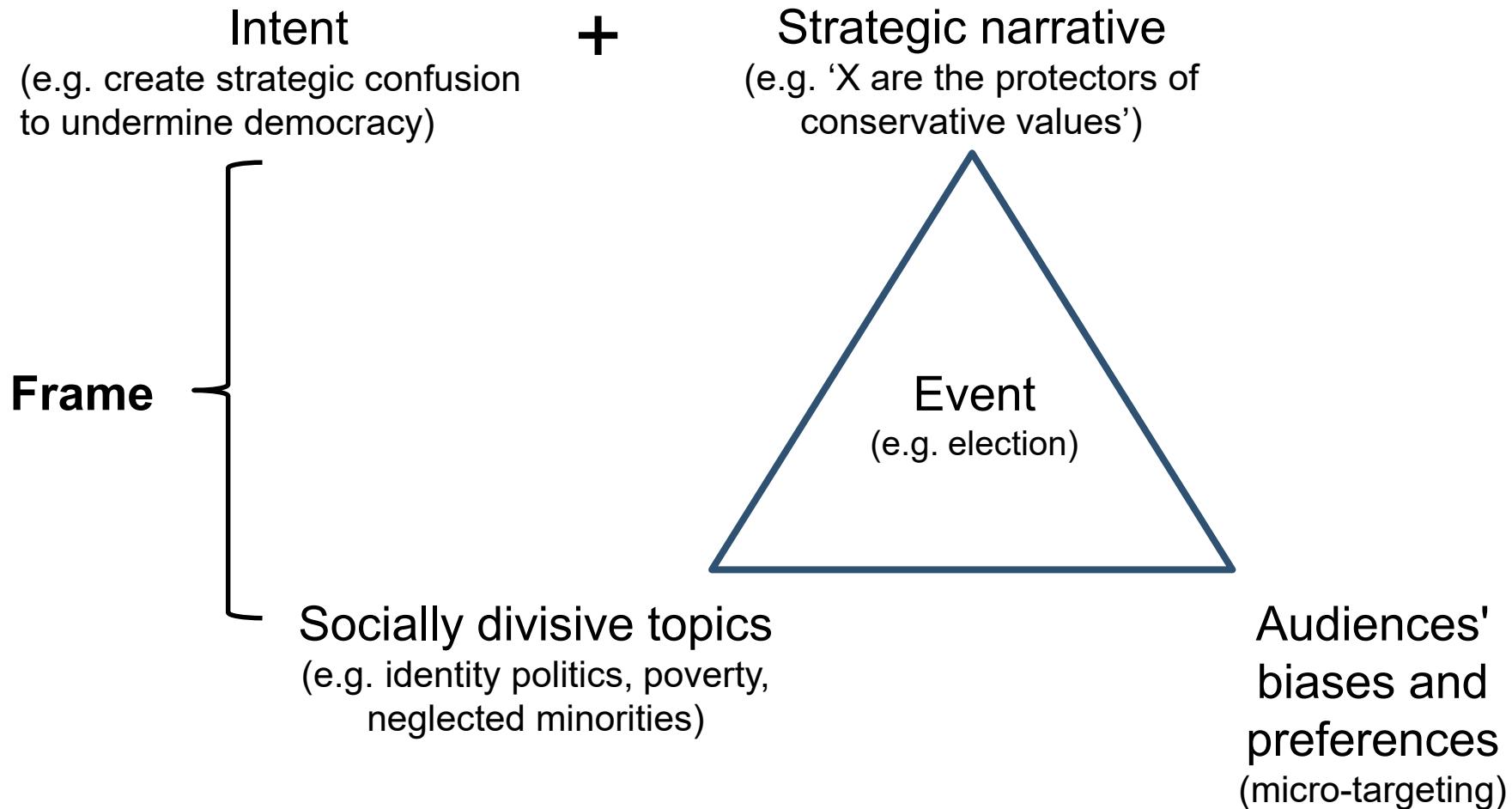
Strategic narrative  
(e.g. 'X are the protectors of conservative values')



Socially divisive topics  
(e.g. identity politics, poverty, neglected minorities)

Audiences' biases and preferences  
(micro-targeting)

# Information operations: components



→ Visual content will be tailored to fit purpose of a particular frame

Pijpers BMJ, Ducheine PAL. *Influence Operations in Cyberspace-How They Really Work*. Amsterdam; 2020. <https://ssrn.com/abstract=3698642>.



# Framing theory

- In essence, framing theory suggests that **how something is presented to the audience (called “the frame”)** influences the choices people make about how to process that information. **Frames are abstractions that work to organize or structure message meaning.** The most common use of frames is in terms of the frame the news or media place on the information they convey. They are thought to influence the perception of the news by the audience, in this way it could be construed as a form of second level agenda-setting – they not only tell the audience **what to think** about (agenda-setting theory), but also **how to think** about that issue (second level agenda setting, framing theory).

Source: [Framing Theory | Mass Communication Theory \(masscommtheory.com\)](http://masscommtheory.com)



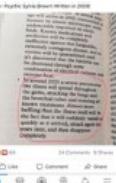
# Visual Framing

- «Coleman (2010) defines visual framing as a form of framing analysis. She draws on Entman's (1993: 52) well-known definition of framing to define **visual framing** as the “selection of one view, scene, or angle when making the image, cropping, editing or selecting it” (Coleman 2010: 237) that “**selects some aspects of a perceived reality and makes them more salient in a communicating text**, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation” (p. 235).»

Brennen, J.S., Simon, F.M., & Nielsen, R.K. (2021). Beyond (Mis)Representation: Visuals in COVID-19 Misinformation. *The International Journal of Press/Politics*, 26, 277 - 299.

# Visual Framing

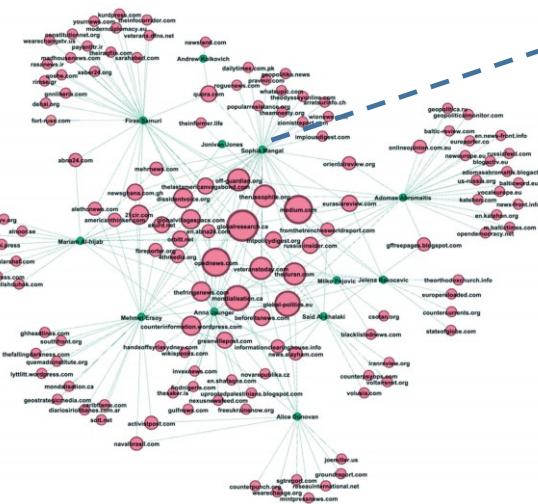
Table 2. The Most Common Frames.

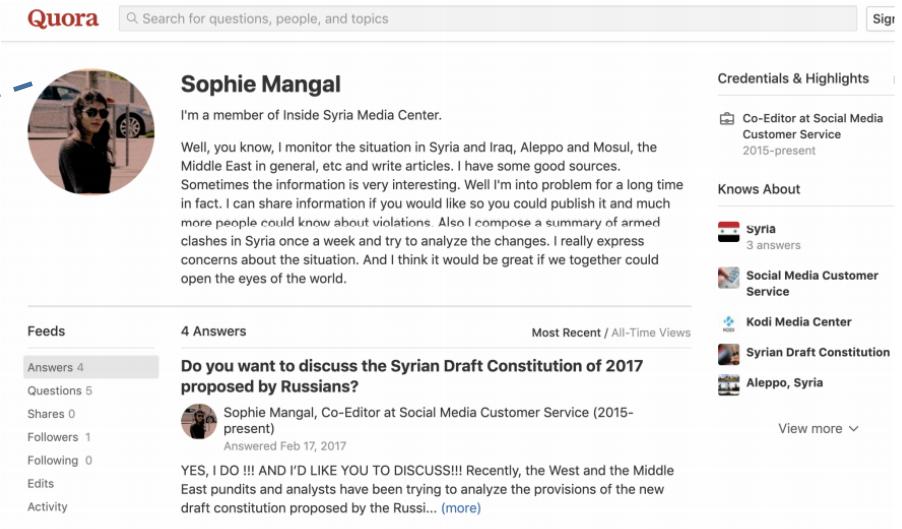
	Frame	Description	Example	Prevalence (%)
1	Authoritative Agency	Valanced claims about actions of public authorities		40
2	Virulence	Over- or understates the spread of the virus, or claims that the disease is not real		33
3	Medical Efficacy	Offers medical information, highlights tests, vaccines, equipment, suggests that there exist cures, treatments, or preventatives for the virus		29
4	Intolerance	Expresses racism, xenophobia, sexism, etc.		15
5	Prophecy	Suggests virus was previously predicted		10
6	Satire	Satirical or humorous content		6

- 96 online articles/posts including misinformation
- Frame of text and visual usually align
- Article/post can transport multiple frames

Brennen, J.S., Simon, F.M., & Nielsen, R.K. (2021). Beyond (Mis)Representation: Visuals in COVID-19 Misinformation. *The International Journal of Press/Politics*, 26, 277 - 299.

# Visual Disinformation Inventory

- Miscontextualized images/videos
  - Shallowfakes/cheapfakes
  - **Internet memes**
  - Deepfakes
  - Staged images/videos
- often integrated in firehoses of falsehood (e.g. via information laundering, bot/cyborg army)
- 
- ↓
- Increased requirements in time & expertise



**Sophie Mangal**  
I'm a member of Inside Syria Media Center.  
  
Well, you know, I monitor the situation in Syria and Iraq, Aleppo and Mosul, the Middle East in general, etc and write articles. I have some good sources. Sometimes the information is very interesting. Well I'm into problem for a long time in fact. I can share information if you would like so you could publish it and much more people could know about violations. Also I compose a summary of armed clashes in Syria once a week and try to analyze the changes. I really express concerns about the situation. And I think it would be great if we together could open the eyes of the world.

**Feeds**

Answers 4      Questions 5      Shares 0      Followers 1      Following 0      Edits      Activity

4 Answers      Most Recent / All-Time Views

**Do you want to discuss the Syrian Draft Constitution of 2017 proposed by Russians?**

Sophie Mangal, Co-Editor at Social Media Customer Service (2015-present)  
Answered Feb 17, 2017

YES, I DO !!! AND I'D LIKE YOU TO DISCUSS!!! Recently, the West and the Middle East pundits and analysts have been trying to analyze the provisions of the new draft constitution proposed by the Russi... [\(more\)](#)

Credentials & Highlights

Co-Editor at Social Media Customer Service 2015-present

Knows About

Syria 3 answers

Social Media Customer Service

Kodi Media Center

Syrian Draft Constitution

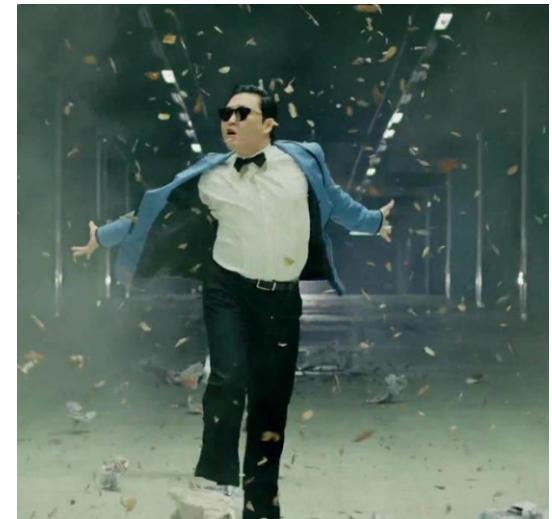
Aleppo, Syria

View more ▾

Source: Potemkin Pages & Personas: Assessing GRU Online Operations, 2014-2019

# Internet memes

- Definition (Shifman 2014):
  - "(a) a group of digital items sharing **common characteristics** of content, form, and/or stance, which (b) were created with **awareness of each other**, and (c) were circulated, imitated, and/or **transformed** via the Internet by **many users**"
- Examples:
  - Success kid (2007)
  - Gangnam Style (2012)
  - #MeToo (2017)





# Why are memes so interesting?

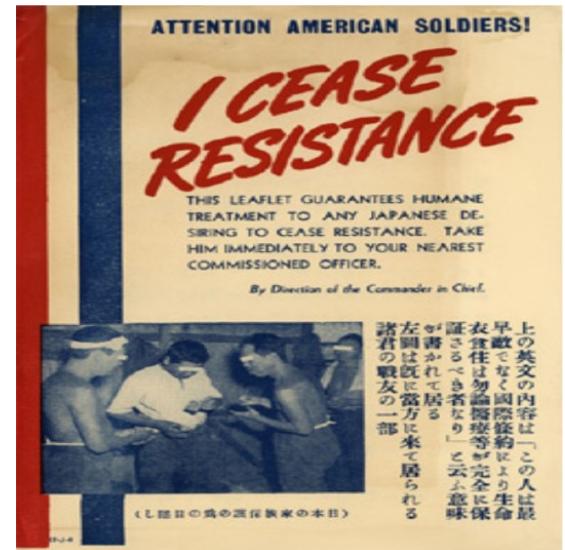
- Four qualities that make them an attractive medium to communicate (Dr. Joan Donovan, Harvard):
  - Memes are authorless, no attached authorship
  - Promote social in-groups and out-groups (require in-group context to understand them)
  - Promotes a form of participation (by altering them)
  - They are sticky (are shared a lot)
- → ideal instrument to plant a (polarizing) idea
- Memes forbidden in Russia, China

# Memes as digital leaflet propaganda

- Memes can act as **agents of legitimization/deligitimization**
  - Legitimization: shine positive light on idea/person by highlighting good qualities inherent in subject
  - Deligitimization: focuses on negative components and also aids in «Othering» of groups who may find said ideas to be a positive thing



Source: [Iraqi Air Force leaflets urge citizens to keep away from ISIS sites - Iraqi News](#)



# Image With Text (IWT) Meme



Potemkin Pages & Personas: Assessing  
**GRU** Online Operations, 2014-2019



DiResta et al. 2018, The tactics and tropes of the **Internet Research Agency**

# ✚ Pepe the frog (~2008)



Russian Embassy, UK  
@RussianEmbassy  
Russia government organization

In today's papers: pundits call on @Theresa\_May to disrupt possible Russia-US thaw. No trust in Britain's best friend and ally?



FT World News und 5 weitere Personen

Donald J. Trump  
@realDonaldTrump

"@codyave: @drudgereport @BreitbartNews  
@WriteinTrump "You Can't Stump the Trump"  
[youtube.com/watch?v=MKH6PA...](https://youtube.com/watch?v=MKH6PA...)"



RETWEETS 7,631 LIKES 10,266



1:53 am - 13 Oct 2015

4 7.6K 10K ...



@MalangKhostay  
(suspended)

# Taliban



Malang

@MalangKhostay

How western Media sees the world:  
**PROGRESS:** **REGRESS:**



16:40 · 02 sept. 21 · Twitter for Android

Visual disinformation

VBS / armasuisse / W+T / WTC / Dr. Raphael Meier



Malang

@MalangKhostay



Malang

@MalangKhostay

I have your back



Taliban



Anti-ZOG American

And you have mine

8:53 AM · Sep 2, 2021 · Twitter for Android





# IWT Meme datasets

Name	Task	Content	Annotations
<b>MultiOFF dataset</b>	Offensive meme detection	IWT memes (n=743), OCR'd	Offensive/not-offensive (binary)
2016 Presidential Election Memes	Descriptive statistics for presidential candidates	IWT memes (n>30'000), OCR'd	Grouped by candidate
Tumblr dataset (2018)	Sentiment analysis	Tumblr posts (n>200'000, text+image)	Automated emotion annotation (15 classes)
Meme generator dataset (2018)	-	IWT memes (n=86'310), OCR'd	-
Reddit memes 2018 dataset ( <a href="#">link</a> )	Unique meme classification	IWT memes (n=5716)	Unique meme label (115 classes)
<b>Facebook hateful meme dataset (2020)</b>	Hate speech classification	IWT memes (n=10'000)	Type of hate (5 classes)
SemEval Challenge Memotion Analysis (2020)	Sentiment analysis	IWT memes (n=9871)	Type of humour (5 classes)
ImgFlip575k Meme dataset	Image captioning	IWT memes (n=575948, 100 most popular meme templates)	-

1. Suryawanshi S, Chakravarthi BR, Arcan M, Buitelaar P. Multimodal Meme Dataset (MultiOFF) for Identifying Offensive Content in Image and Text.; 2020.
2. Hu A, Flaxman S. Multimodal Sentiment Analysis To Explore the Structure of Emotions. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York, NY, USA: ACM; 2018. doi:10.1145/3219819.3219853.
3. Beskow DM, Kumar S, Carley KM. The evolution of political memes: Detecting and characterizing internet memes with multi-modal deep learning. Inf Process Manag. 2020;57(2). doi:10.1016/j.ipm.2019.102170.
4. Sharma C, Bhageria D, Scott W, PYKL S, Das A, et al. SemEval-2020 Task 8: Memotion Analysis -- The Visuo-Lingual Metaphor! arXiv. August 2020. <http://arxiv.org/abs/2008.03781>.

# «Strong brother frame»

Russia is stronger than  
NATO and helps fighting  
terrorists in Syria

Source: [Disclosing networks of state-linked information operations \(twitter.com\)](#)

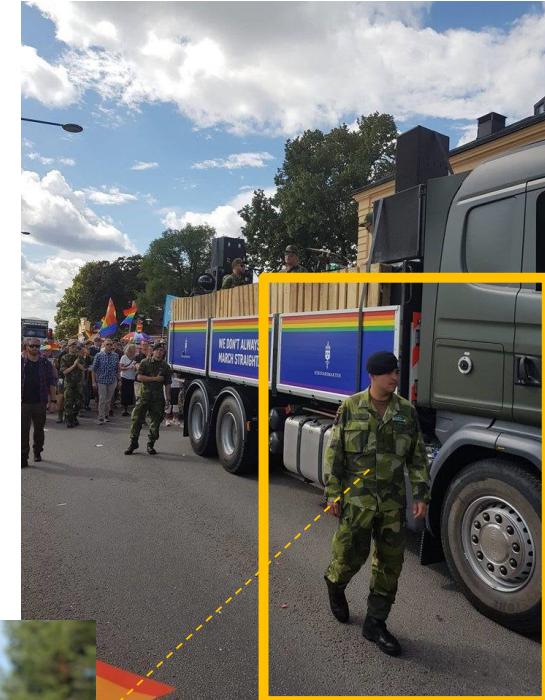
# Russia, GRU, 2020-2021, Syria/NATO

Russian President Vladimir Putin called for helping the Syrian regions, in which refugees from neighboring countries can return. The corresponding statement was made by the Russian head of state before the meeting with German Chancellor Angela Merkel. <https://t.co/n8A8OhsH4B>



# Russia, GRU, 2020-2021, Syria/NATO

**Swedish military forced to attend a gay pride parade!** Steadily marching into NATO (through the back entrance)!  
Everything you need to know about tolerance) # Sweden # gayparade # NATO # gay # armed forces #Sweden #gaypropaganda <https://t.co/ww1arcabEd>



Swedish military uniform

# Russia, GRU, 2020-2021, Syria/NATO

# The **US** may block the delivery of #  
**Turkey** F-35 because of the contract with #  
**Russia** for the supply of S-400. #Russia  
#USA #Turkey #Syria #NATO # S400 #  
F35 # Russia # Turkey # ABM # Air  
Defense # Policy <https://t.co/dfo8JsGcyl>



**Western values** can bring to **Armenia** the same **extremists** who are now raging in Germany. And about the **parades**, I generally keep quiet. Vaughn Ukraine is already on the verge of holding them ...  
<https://t.co/lcxihCY6Z5>



# Russia, GRU, 2020-2021, Syria/NATO

The #JebhatNusra militants are preparing provocations with the use of chemical weapons for the subsequent accusation of the #Syrian government.  
<https://t.co/LziOdbJzYh>



#сирия #белые\_каски #терроризм #война  
#фальшивые\_каски #syria #WhiteHelmets  
#nusra #EasternGhouta  
<https://t.co/Nxs1VLBuJp>



# Russia, GRU, 2020-2021, Syria/NATO

RT @publikatsii\_ru: Something like this  
<https://t.co/fO51wAKSaR>



@Voltzzzzzz36 @MichaelKillian1  
@guyverhofstadt <https://t.co/bxzCUU7f2f>





# Countermeasures

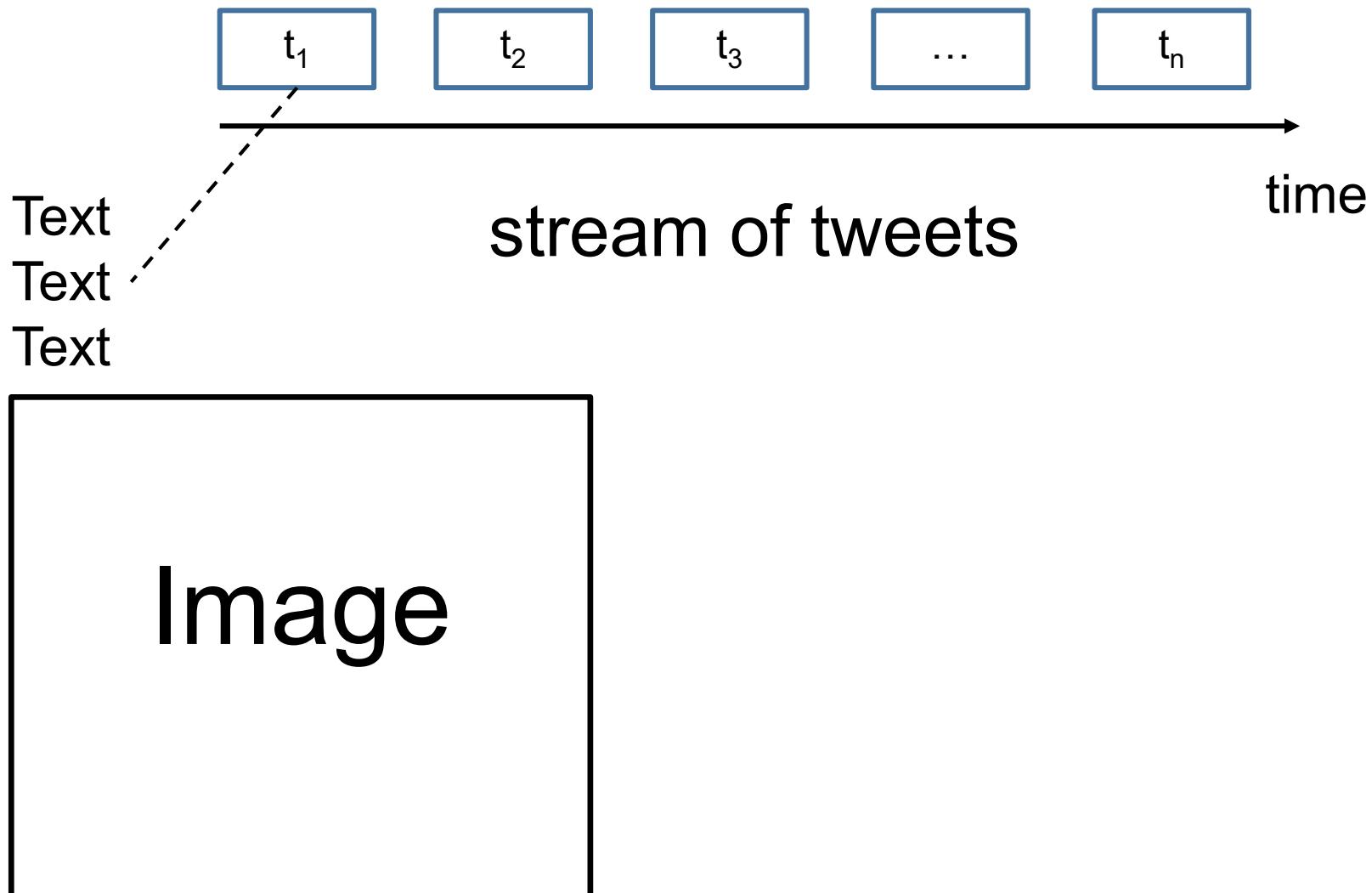


# Potential countermeasures

- Media literacy
- Fact-checking organizations, supporting journalism
- Preventative strategic communication
- Detection of «coordinated inauthentic behavior» by social media platforms (→ Machine Learning)
- **Characterization of uni-/multimodal content on social media to support human analysts (→ Machine Learning)**
  - →federal police, intelligence services, and armed forces

[Combating Information Manipulation: A Playbook for Elections and Beyond | National Democratic Institute \(ndi.org\)](#)  
[RESIST 2 Counter Disinformation Toolkit - GCS \(civilservice.gov.uk\)](#)

# Level of Analysis



# Level of Analysis

t<sub>1</sub>

t<sub>2</sub>

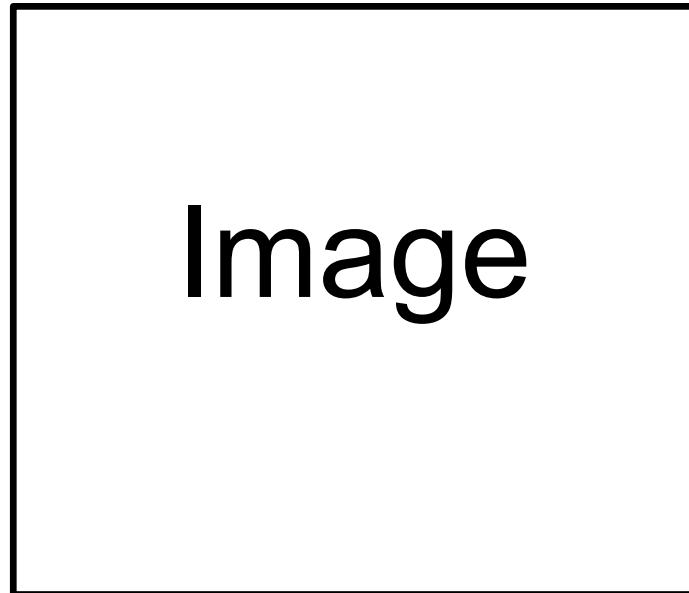
t<sub>3</sub>

...

t<sub>n</sub>

Text  
Text  
Text

content



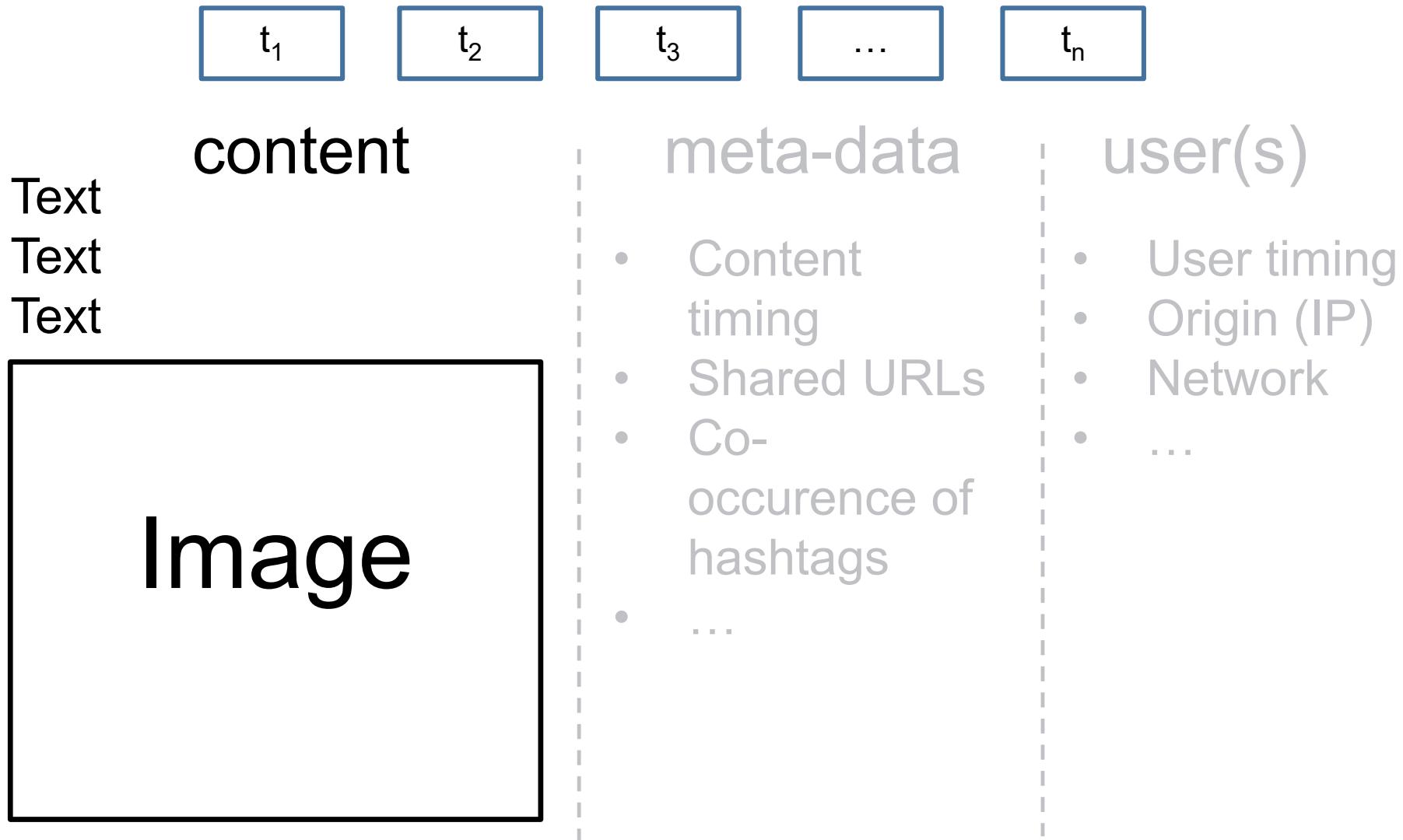
meta-data

- Content timing
- Shared URLs
- Co-occurrence of hashtags
- ...

user(s)

- User timing
- Origin (IP)
- Network
- ...

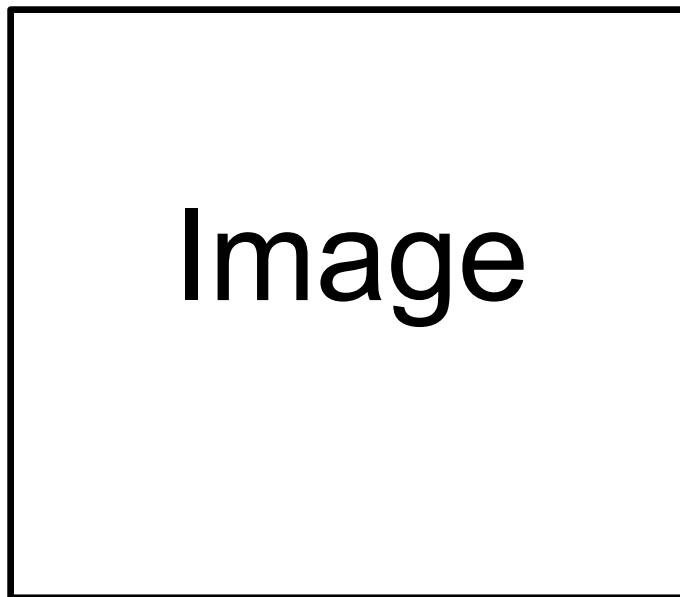
# Level of Analysis



# Level of Analysis

content

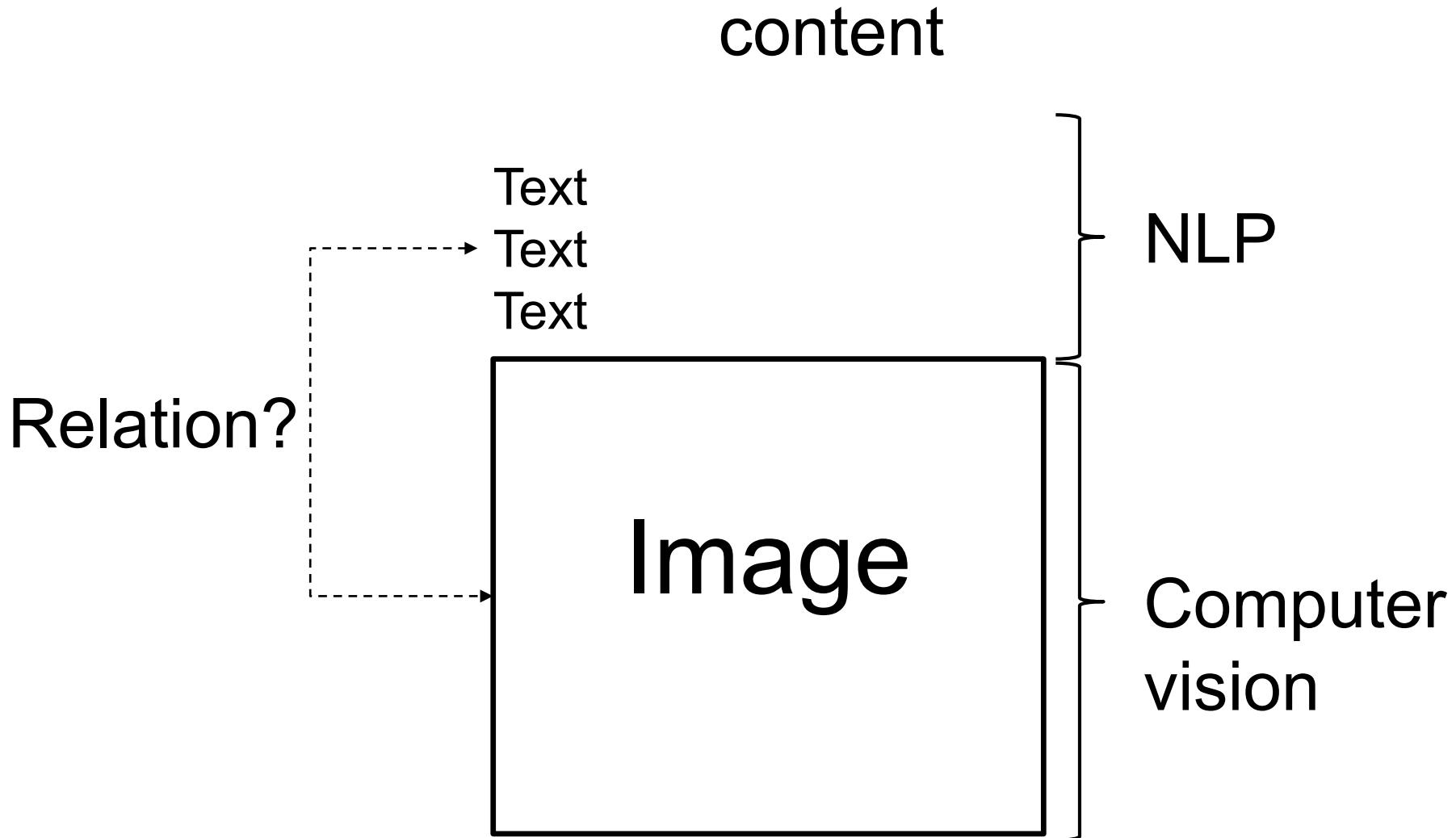
Text  
Text  
Text



NLP (e.g. frequency analysis, Transformers, ...)

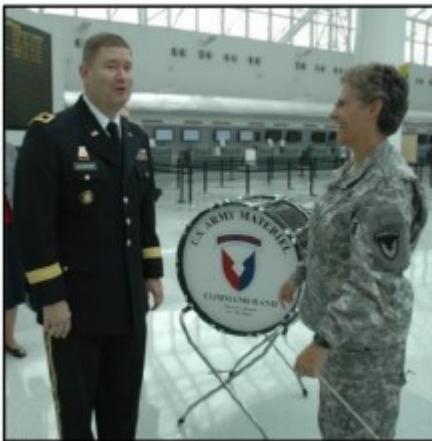
Computer vision (e.g. Convolutional Neural Networks)

# Level of Analysis



Otto, C., Springstein, M., Anand, A. et al. Characterization and classification of semantic image-text relations. Int J Multimed Info Retr 9, 31–45 (2020)

# Image-text relation



A man in dress uniform talks with a woman wearing camouflage.



A turnbuckle, stretching screw or bottlescrew is a device for adjusting the tension or length of ropes, cables, tie rods, and other tensioning systems. It normally consists of two threaded eye bolts, [...].

**Fig. 6** Examples for the *Anchorage* (left) and *Illustration* (right) classes. (Sources: see Sect. 4.1)

Otto, C., Springstein, M., Anand, A. et al. Characterization and classification of semantic image-text relations. Int J Multimed Info Retr 9, 31–45 (2020)



# Content: interesting aspects to analyze

- Image type (e.g. IWT meme vs. non IWT-meme image)
- Topic/theme of content
- Stance of author
- Hateful or offensive content
- Sentiment (negative/positive valence)
- Veracity of content (automated fact-checking)
- Image-text similarity
- Semantic image-text relation
- ...

→Classification problems

Alam, Firoj, et al. "A survey on multimodal disinformation detection." *arXiv preprint arXiv:2103.12541* (2021).

# Info Ops Detection

SCIENCE ADVANCES | RESEARCH ARTICLE

SOCIAL SCIENCES

## Content-based features predict social media influence operations

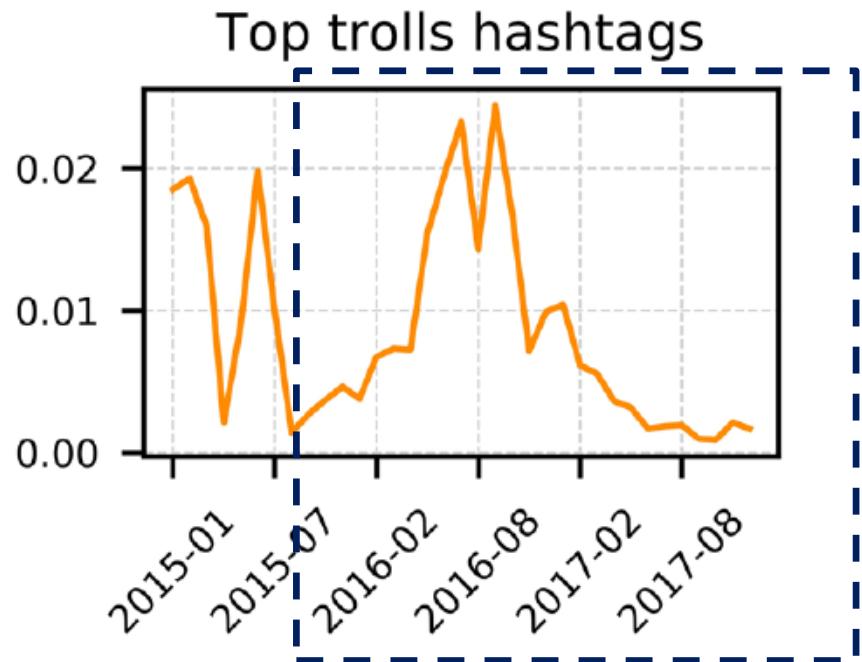
Meysam Alizadeh<sup>1\*</sup>, Jacob N. Shapiro<sup>1</sup>, Cody Buntain<sup>2</sup>, Joshua A. Tucker<sup>3</sup>

We study how easy it is to distinguish influence operations from organic social media activity by assessing the performance of a platform-agnostic machine learning approach. Our method uses public activity to detect content that is part of coordinated influence operations based on human-interpretable features derived solely from content. We test this method on publicly available Twitter data on Chinese, Russian, and Venezuelan troll activity targeting the United States, as well as the Reddit dataset of Russian influence efforts. To assess how well content-based features distinguish these influence operations from random samples of general and political American users, we train and test classifiers on a monthly basis for each campaign across five prediction tasks. Content-based features perform well across period, country, platform, and prediction task. Industrialized production of influence campaign content leaves a distinctive signal in user-generated content that allows tracking of campaigns from month to month and across different accounts.

- text only, **meta-content** really helpful
- impact probably very limited (see next slide)

Copyright © 2020  
The Authors, some  
rights reserved;  
exclusive licensee  
American Association  
for the Advancement  
of Science. No claim to  
original U.S. Government  
Works. Distributed  
under a Creative  
Commons Attribution  
NonCommercial  
License 4.0 (CC BY-NC).

# ✚ Trend: increased operational security



*"In response to increased efforts at stopping them, the more sophisticated threat actors — including from Russia and China — have improved their operational security ("OpSec"). They are showing more discipline to avoid careless mistakes like logging into purportedly American accounts from St. Petersburg in Russia. Some are also getting better at avoiding language discrepancies and similar inauthenticity clues by re-appropriating authentic content from domestic communities, rather than creating their own."*

Source: [Threat Report: Combating Influence Operations | Meta \(fb.com\)](#)

Source: [Content-based features predict social media influence operations \(science.org\)](#)

→ Better methods to characterize multimodal content needed



# Take-Home Messages

- Combination of technical, social, and economic factors which provide basis for spread of disinformation
- Disinformation is usually part of a larger information operation which has short and long-term goals
- Disinformation typically contains half-truths and may be shared by an «unwitting crowd»
- Internet memes are a powerful tool to influence online narratives (→ digital leaflet propaganda)
- Actors are increasing their operational security rendering information operations more difficult to detect
- Role of Machine Learning as potential countermeasure still being developed



# Hands-on part

- Objectives:
  - Familiarize yourself with IWT memes
  - Perform experimental tasks
  - Reflect on topics and answer questions
  - Develop a first understanding of problems and potential solutions
  - **Spark interest in subject**
- Datasets:
  - Facebook Hateful Memes Challenge 2020 + COCO-Text dataset
  - MultiOFF dataset (offensive/non-offensive memes)



# Questions?

Thank you very much for your attention!

→ Hands-on part: <https://github.com/i4Ds/AMLD-2022-Visual-Disinformation>