

Your Truth \neq My Truth

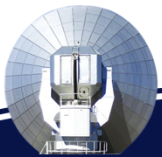
Technical and Social Elements of Computational Propaganda

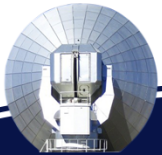
Gerhard Backfried

SAIL LABS

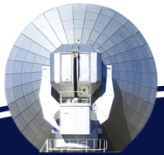
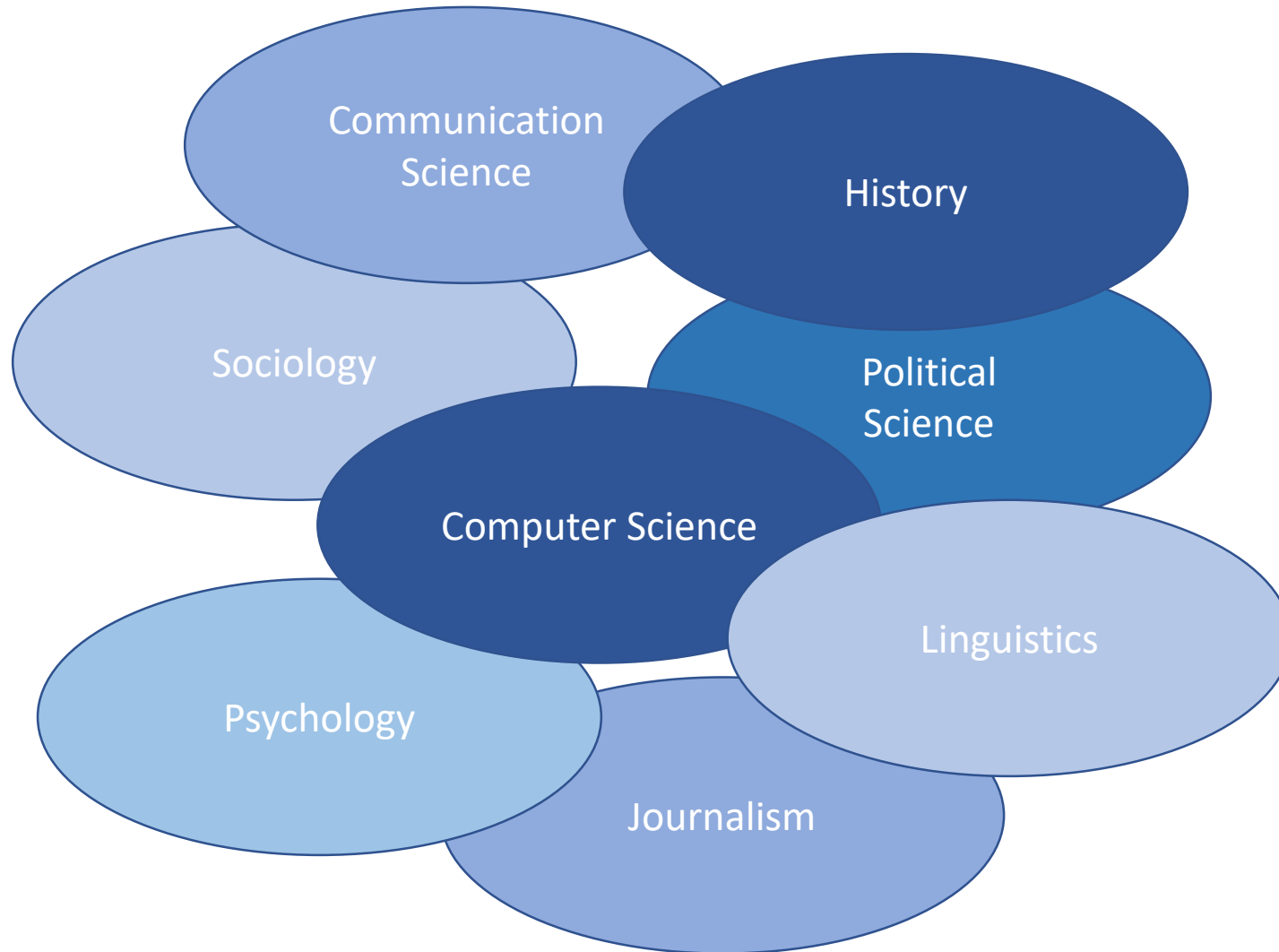
University of Vienna

Fake News Conference,
Vienna, Nov 30, 2018





Interdisciplinarity



Terminology

..intentionally and verifiable false news ...

...false stories that appear to be news...

..cognitive manipulation ...
confirming biases...

- Propaganda
- Misinformation / Disinformation / Malinformation
- Manipulation
- Fake News
- Deep Fakes

...perceived and deliberate distortion of news with the
intention to affect the political landscape and to
exacerbate divisions in society ...

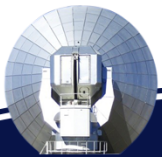
...includes all forms of false, inaccurate or
misleading information designed, presented
and promoted intentionally to cause public
harm or for profit ...

..deliberate presentation of false
or misleading claims as news...

...false, often sensational stories ...

..cheap to produce..

..promote ideologies, confuse, sow discontent and create polarization...



Terminology

..intentionally and verifiable false news ...

...false stories that appear to be news...

..cognitive manipulation ...
confirming biases...

- Propaganda
- Misinformation / Disinformation / Malinformation
- Manipulation
- Fake News
- Deep Fakes

...perceived and deliberate distortion of news with the intention to affect the political landscape and to exacerbate divisions in society ...

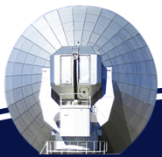
...includes all forms of false, inaccurate or misleading information designed, presented and promoted intentionally to cause public harm or for profit ...

..deliberate presentation of false or misleading claims as news...

...false, often sensational stories ...

..cheap to produce..

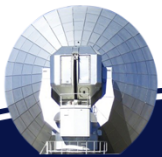
..promote ideologies, confuse, sow discontent and create polarization...



Terminology II

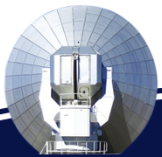
- Propaganda
- Misinformation / Disinformation / Malinformation
- Manipulation
- Fake News
- Deep Fakes

Fake news is a type of **yellow journalism** or **propaganda** that consists of **deliberate disinformation** or **hoaxes** spread via **traditional print and broadcast news media or online social media**. The term is also at times used to **cast doubt** upon legitimate news from an opposing political standpoint, a tactic known as the **lying press**. The **false information** is then often reverberated as **misinformation** in **social media**, but occasionally finds its way to the **mainstream media** as well. Fake news is written and published with the **intent to mislead in order to damage** an agency, entity, or person, and/or gain **financially or politically**, often using **sensationalist, dishonest, or outright fabricated headlines** to increase readership, online sharing, and Internet click revenue. In the latter case, it is similar to sensational online **clickbait headlines** and relies on advertising revenue generated from this activity, regardless of the veracity of the published stories.



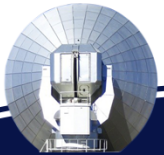
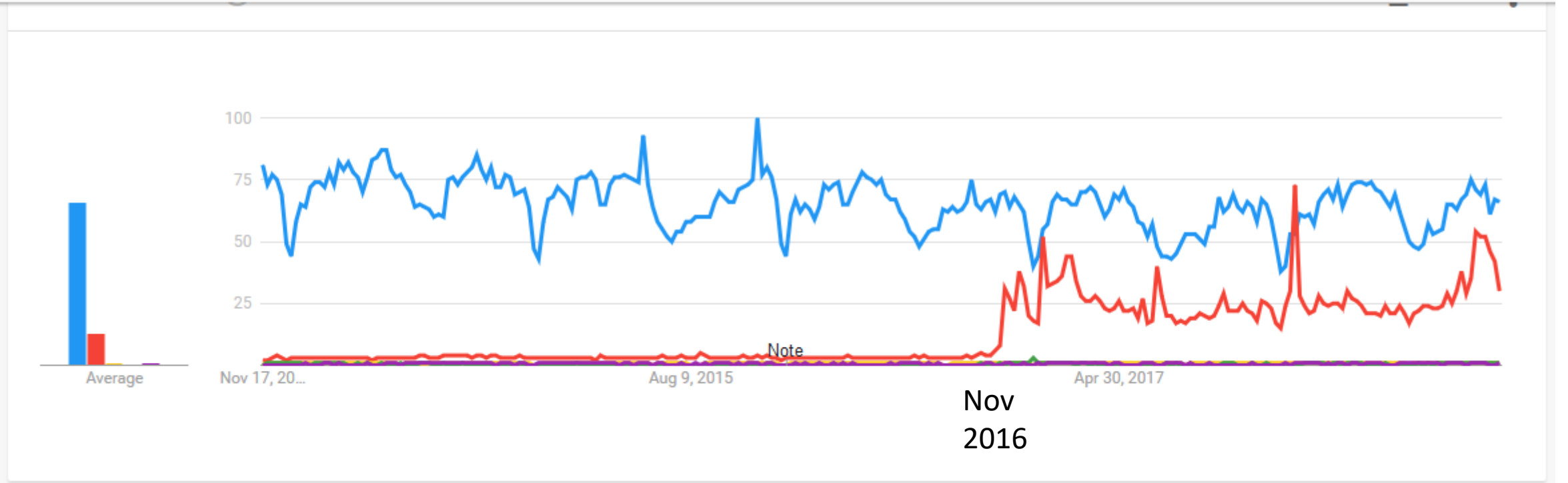
Terminology III

- Computational propaganda / disinformation
 - The use of algorithms, automation, human curation to purposefully create, manage and distribute misleading information (on all media)
 - Combines social and technical aspects
 - Technical: algorithms, agents, platforms, big-data, statistics, ML,...
 - Social: human actors and motivations behind, agenda, social interaction,...
 - BOTH aspects need to be addressed to handle/counter it!
 - Automation/scalability/anonymity, especially with SM
 - Digital astroturfing / trolling / PsyOps
- Goals
 - Fabricated illusion of large-scale consensus and popularity
 - Drown-out and harass opposition



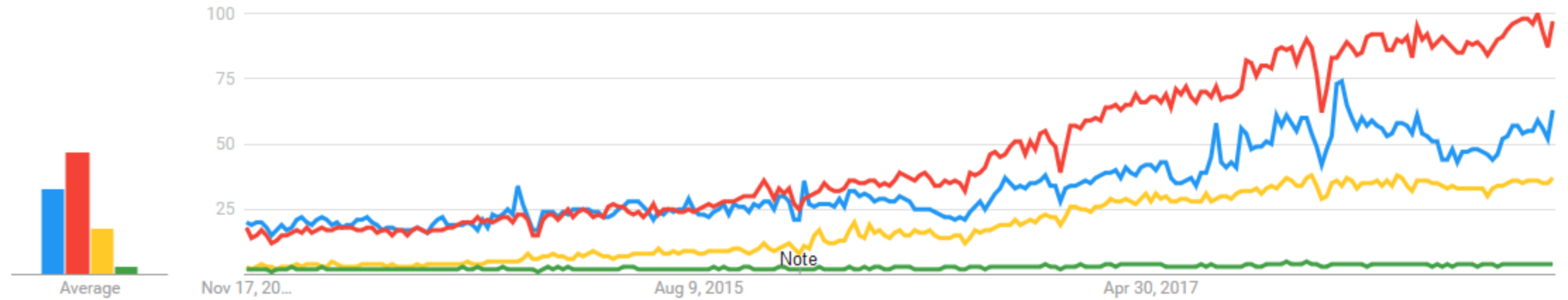
● Propaganda ● Fake News ● Misinformation ● Disinformation ● Lying Press

Worldwide, Past 5 years



● Artificial Intelligence ● Machine Learning ● Deep Learning ● Natural Language Processing

Worldwide, Past 5 years



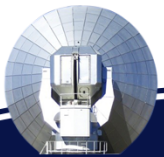
Nov 17, 20...

Aug 9, 2015

Apr 30, 2017

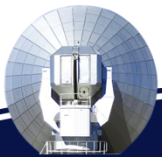
Note

Nov
2016



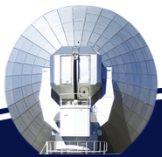
Content

- Emotional content / activation
 - You can be part of it!, This is your fight!
 - Create feelings/emotions which stay after/beyond the message
 - Bypass rational thought
- Short, simple messages/narratives
 - Fear, anxiety, insecurity, intimidation, defamation,...
 - Get core-terminology into reader's heads
- Language determines view of the World
 - Framing
 - Frequency
 - Dominion of terminology = dominion of perspective
 - Attack democracy by attacking language!



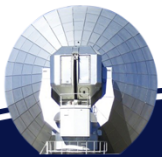
Language Boundaries

- What is accepted? / who determines that?
 - Society/law
 - Shades of gray / areas to *explore* but not get into conflict with the law
- Propaganda recipees
 - Adding negative connotations
 - Democracy, Liberal
 - Re-using historical terminology
 - *Umvolkung*
 - Euphemisms
 - *Remigration*
 - Neologisms
 - *Ethnomasochism*, White Racism, Rapefugees, *Staatsversagen* (state failure)
 - Tailored foreign-language content
 - Turkish media, Russian, ISIS,...

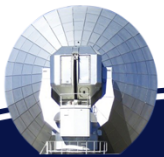
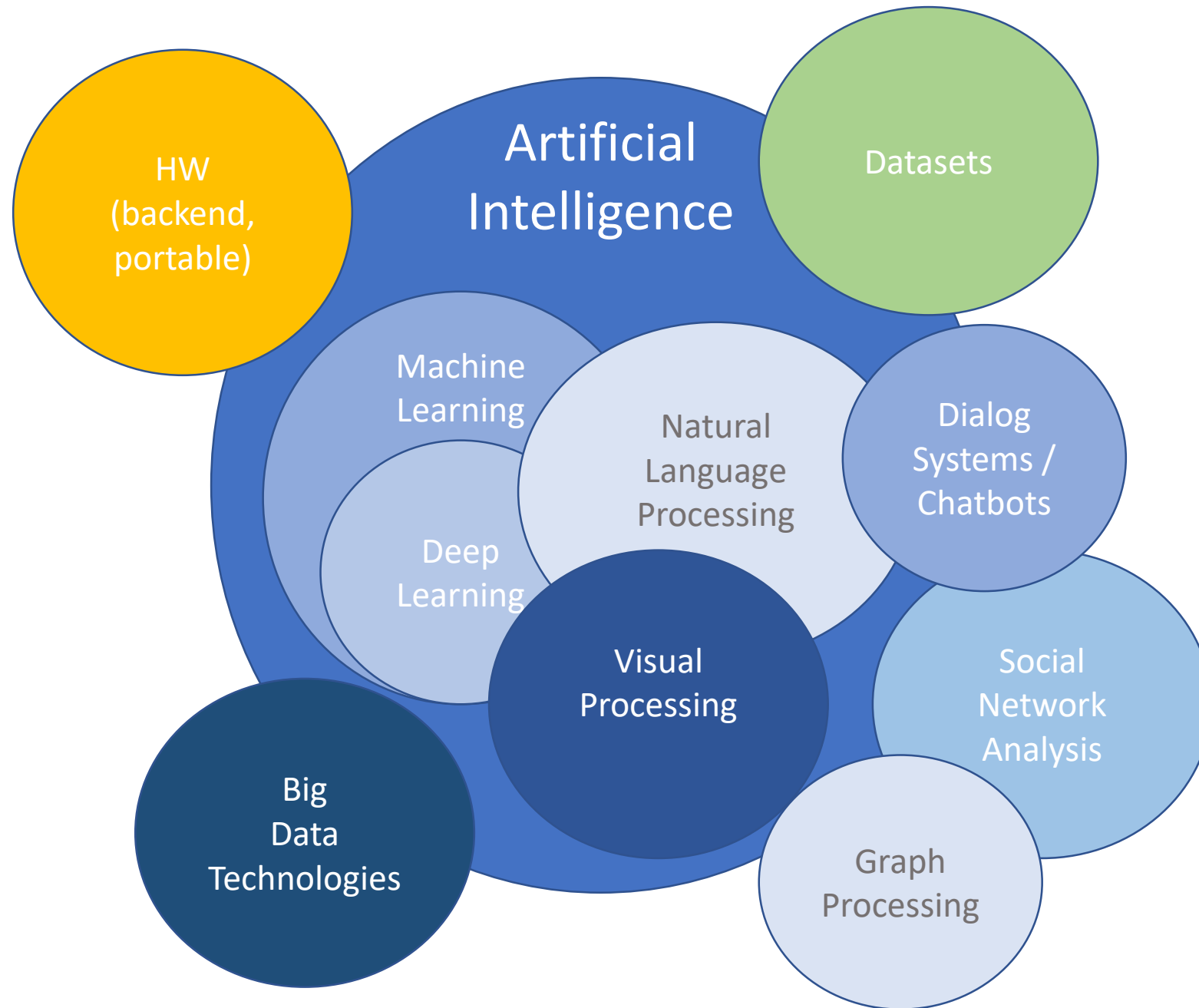


Media Landscape

- Echo chambers
 - High % SM, sharing, forwarding
 - Specific target-groups
 - Connections not possible in other ways/real life
 - Personalized programming – strengthens echo-chambers
- Proprietary media (AfD TV, Rumiya, Sputnik,...)
 - Interviews, conferences, live reports,...
- Counter-public to mainstream media
 - counter/alternative public to general, public discourse

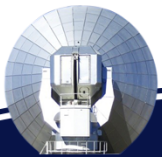


AI&IT



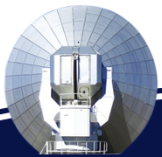
Remedies

- Inter disciplinary field – also should best be tackled that way!
 - IT, NLP, Linguistics, Psychology, Communication Science,...
 - Technical + Social
- Social
 - Media literacy
 - Educational programs
 - Factual reporting / journalism 3.0 / Hippocratic Oath
 - Active communication and arguments/disputes
 - Know echo-chambers and interfere
 - Law vs censure

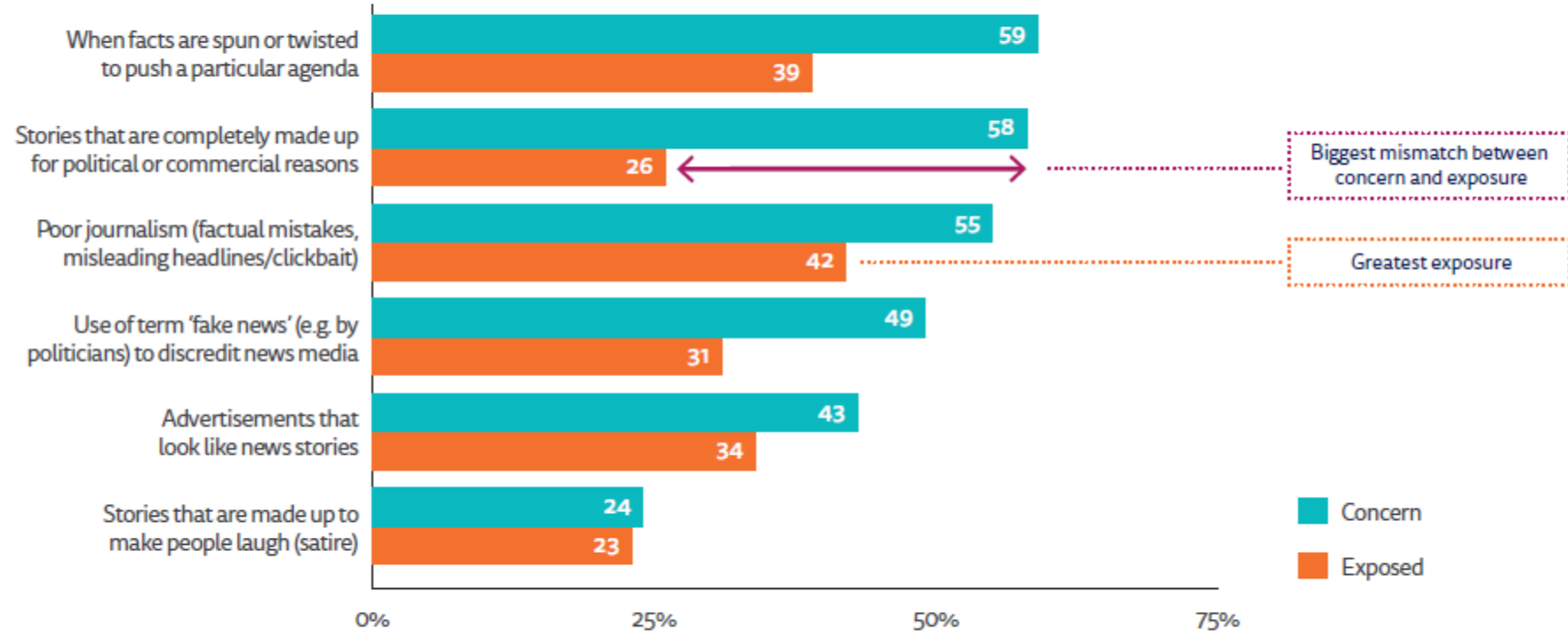


Remedies

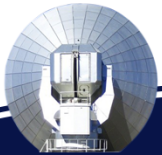
- Technical
 - Professional/crowd-sourced fact-checking
 - Automation
 - Source & author identification
 - Author, publisher credentials, web-presence
 - Known bias / allegiance
 - Origin / propagation
 - Can it be traced back?
 - What are others writing about the same subject?
 - Graph-analytics / distribution patterns
 - Content
 - Stylometric measures
 - Clickbait
 - Sentiment
 - References (unnamed sources)
 - Reused content/images
 - Intention?



PROPORTION WHO SAY THEY ARE VERY OR EXTREMELY CONCERNED ABOUT EACH, AND PROPORTION WHO SAY THEY SAW EACH IN THE LAST WEEK – ALL MARKETS



Q_FAKE_NEWS_2. To what extent, if at all, are you concerned about the following. Q_FAKE_NEWS_3. In the last week which of the following have you personally come across? Base: Total sample in all markets.



Thank you – Merci – Gracias – Terima kasih – Dziękuję – Спасибо – شكراً – Takk – Tack – Tack – Grazie – Gràcies – 感謝 – Ευχαριστώ – תודה

SAILLABS

TECHNOLOGY

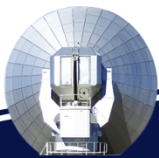
Gerhard Backfried

Mariannengasse 14 · 1090 Vienna · Austria

Ph +43-1-580 95- 0 · Fax +43-1-580 95-580

www.sail-labs.com

SAILLABS



References

- Slide 2: <http://www.usbdata.co/hot-air-balloons-in-the-sky.html>
- Slide 3:
 - https://en.wikipedia.org/wiki/Fake_news
 - JRC Digital Economy Working Paper 2018-2, Bertin Martens et al 04/2018
https://ec.europa.eu/jrc/communities/sites/jrccties/files/dewp_201802_digital_transformation_of_news_media_and_the_rise_of_fake_news_final_180418.pdf
- Background
 - Ute Schaeffer, Fake Statt Fakt, dtv 2018
 - George Bara, Gerhard Backfried, Dorothea Aniola, “Fake or Fact? Theoretical and Practical Aspects of Fake News” in Eloi Bosse and Galina Rogova (Eds) Information Quality in Information Fusion and Decision Making, Springer, due March 2019
- Deep Fakes -> Xinhuas English AI anchor: <https://youtu.be/GAfiATTQufk>

