NORDIS

# Activity 4: INNOVATION & TECHNOLOGY
## Led by University of Bergen

@NORDISHUB

This project has received funding from the European Union

# University of Bergen Team

- Carl-Gustav Lindén, Professor
- Andreas Lothe Opdahl, Professor
- Duc Tien Dang Nguyen, Associate Professor
- Laurence Dierickx, PostDoc
- In collaboration with: Sohail Ahmed Khan, PhD Candidate (MediaFutures)



This project has received funding from the European Union

# Focus 1: State of the art in fact-checking technology



This project has received funding from the European Union

# Identifying practical applications and tools

- **134 fact-checking and service tools** encompassing a diverse range of technologies and methodologies to discover, verify, and classify facts. https://docs.google.com/spreadsheets/d/1oIFvwR8b_7v9osdJwueNo4KSBb2p7PuqOQD4XKhO7k8/edit#gid=0
- Majority of tools focus on mining and analysing textual content
- Limited tools available for verifying visual content (Copy/Move Forgeries, Image Splicing Forgeries, Image Cropping, Deepfake Media, Cheap Fake Media/Image Re-contextualization)
- Smaller subset of tools designed for educational purposes

*Lindén, C. G., Dang-Nguyen, D. T., Salas-Gulliksen, C., Khan, S. A., Amelie, M., & Dierickx, L. (2022). State of the art in fact-checking technology. NORDIS Project Report: University of Bergen, Task 4*



This project has received funding from the European Union

# Online fact-checking tools database

Merging of the NORDIS database with the database of the Oslo Metropolitan University (SCAM Project)

**249 tools** covering identification, production, distribution

Hosted on NORDIS website

10 Show entries

Search:

| | Tool | URL | Description | Technology | Service_1 | Service_2 | Type | Text | Image | Video | Audio | Site_pa | Social_ | Bot |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | AccountAnalysis | https://accountan | Tool for gathering information on Twittter users: when and what type of contect they publish, what accounts they interact with, and what websites they share. (No AI or machine learning.) | NA | identification | NA | Commercial, Personal | FALSE | FALSE | FALSE | FALSE | FALSE | TRUE | FALSE |
| 2 | Adblock Plus | https://adblockplu | Adblock Plus is a browser extension and app that blocks advertisements and websites through the use of filtering lists. The tool automatically whitelists ads that meet the 'acceptable ads' standards — so ads can avoid being blocked on this platform by adhering to those standards. Users can opt out of even these ads by adjusting the default settings. The tool, while initially an adblocker, has increasingly seen itself as a way for users to protect themselves not only from harmful ads, but also from other harmful sites, including those that spread disinformation. | Filtering lists | identification | ads | NA | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE |
| 3 | Adobe | NA | Described in interview with IFCN as a 'tool that they deliver to their members', and it's 'for video, for image, and PDF content'. | NA | distribution | verification | NA | TRUE | TRUE | TRUE | FALSE | FALSE | FALSE | FALSE |
| 4 | Adverif.ai | https://www.adve | AdVerif.ai – Fight Fake News with AI. Fake News is proliferating and creating real losses for publishers and media companies.AdVerif.ai is developing the FakeRank AI to automatically identify disinformation, build trustworthiness into online media and advertising networks, and support the news ecosystem by empowering humans to fight fake news at scale. FakeRank is built upon state-of-the-art deep technology, demonstrated in a series of research publications, and is utilizing proprietary data through partnerships with IFCN organizations. It was recently named among CB Insights' 2019 International Game Changers – startups with | AI | identification | NA | NA | TRUE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE |

https://nordishub.eu/2022/10/13/fact-checking-tools-database/

@NORDISHUB

This project has received funding from the European Union

# Focus 2: Identifying the fact-checkers users' needs



This project has received funding from the European Union

# Interviews with Nordic fact-checkers

- 14 interviews with professional fact-checkers and newsrooms managers in Denmark, Finland, Norway and Sweden, with various professional backgrounds and professional profiles (from young to experimented or tech-oriented)
- Tools are means, not ends
- Use the same limited number of tools (Google Search Engine, CrowdTangle, TinEye)
- Open to innovation but no time/skills to discover or test new tools
- AI: enabler rather than a complete solution
- Need for ethics to enable trustworthiness: reliability, accuracy, transparency
- Need for understanding how AI-based tools work (explainability) + human-in-the-loop

*Dierickx, L., Sheikhi, G., Nguyen, D. T. D., Lindén, C. G. (2022). Report on the user needs of fact-checkers. NORDIS Project Report: University of Bergen, Task 4.*



This project has received funding from the European Union

# More specific needs

- Monitoring of social networks (TikTok, Telegram, YouTube, the most difficult)
- Monitoring of political debates broadcast in audiovisual media
- Claim collection and detection on social networks and audiovisual news media (radio and TV), eventually providing links to fact-checks already published
- Verifying with context, especially for YouTube and video fact-checks
- Resources for archiving problematic content published on the web, already published fact-checks, or links to trustworthy websites
- **Resources and tools more adapted to the Nordic specificities and languages**



This project has received funding from the European Union

# Use case: fact-checking the Ukraine war

*(part of an international comparative study)*

- Distance + language do not allow a full understandability
- Unreliable sources (both sides) and can take longer to fact-check
- Need for collaborations between fact-checkers
- Blurred images are the most challenging
- Biggest challenge: AI generated content
- Exposure to violent content

- Propaganda on both sides, but cannot be compared (soft vs hard)
- The young generation of fact-checkers need to know more about propaganda
- Need for tools to monitor disinformation in the Russian-speaking population in the country
- Could fact-check more but lack of human resources (micro-structures, national priorities)



This project has received funding from the European Union

# The gap between the design and the needs

## Systematic literature review on automated fact-checking (367 papers)

- Most research in AFC focuses on providing technological solutions to solve the social problem of information disorder
- Moving from a technological point of view toward a social one is needed: involving developers and fact-checkers/journalists + keeping the human in the loop
- Although the conditions of use of a technological artifact are multifaceted, the reliability of the results that it provides also participates in building a trust-based relationship

*Dierickx, L., Lindén, C. G., & Opdahl, A. L. (2023). Automated Fact-Checking to Support Professional Practices: Systematic Literature Review and Meta-Analysis. International Journal of Communication, 17(2023)*



This project has received funding from the European Union

# Focus 3: Meeting the users' needs throughout responsible technology



This project has received funding from the European Union

# Multimedia Forensics Suite

## Development of a set of prototypes and tools for visual content verification

- FotoVerifier, a tool for image/video tampering detection exploiting digital image forensics (DIF) techniques
- CLI, a collection of Python scripts that offer a command-line interface for accessing the functionality of FotoVerifier
- NameSleuth, an online tool that analyses social media platform traces from uploaded images
- DivNoise, a tool and database for identifying the acquisition source of media data

https://github.com/fotoverifier/ https://dedigi.fotoverifier.eu/

https://namesleuth.fotoverifier.eu/ https://demo.divnoise.fotoverifier.eu/

**Dang-Nguyen, D.T.**, Sjøen, V.V., Le, D.H., Dao, T.P., Tran, A.D. and Tran, M.T., 2023, January. Practical Analyses of How Common Social Media Platforms and Photo Storage Services Handle Uploaded Images. In *MMM 2023*

Tran, C.H., Tran, Q.T., Long-Vu, Q.C., Nguyen, H.S., Tran, A.D. and **Dang-Nguyen, D.T.**, 2022, June. Dedigi: a privacy-by-design platform for image forensics. In *MMM 2022*.

Casagrande, A., Belli, A., Pasquini, C., **Dang-Nguyen, D.T.**, 2023, September. A Data Collection for Source Identification on Diverse Camera Sensors. In *ECML PKDD 2023*.

@NORDISHUB

This project has received funding from the European Union

NORDIS

## Analyze file name

Analyze image source from the file name

Uploade the image to extract the filename

Browse... No file selected.

Or input the file name manually

Input the filename...

Analyze Analyzing process may take up to a few minutes

**Select Image:**

Browse... No file selected.

**Or drag and drop image here:**

Drag and Drop Image Here

Submit

## Or try with our sample images!

**Apple iPhone 13 Rear**

**Apple iPad mini 5 Frontal**

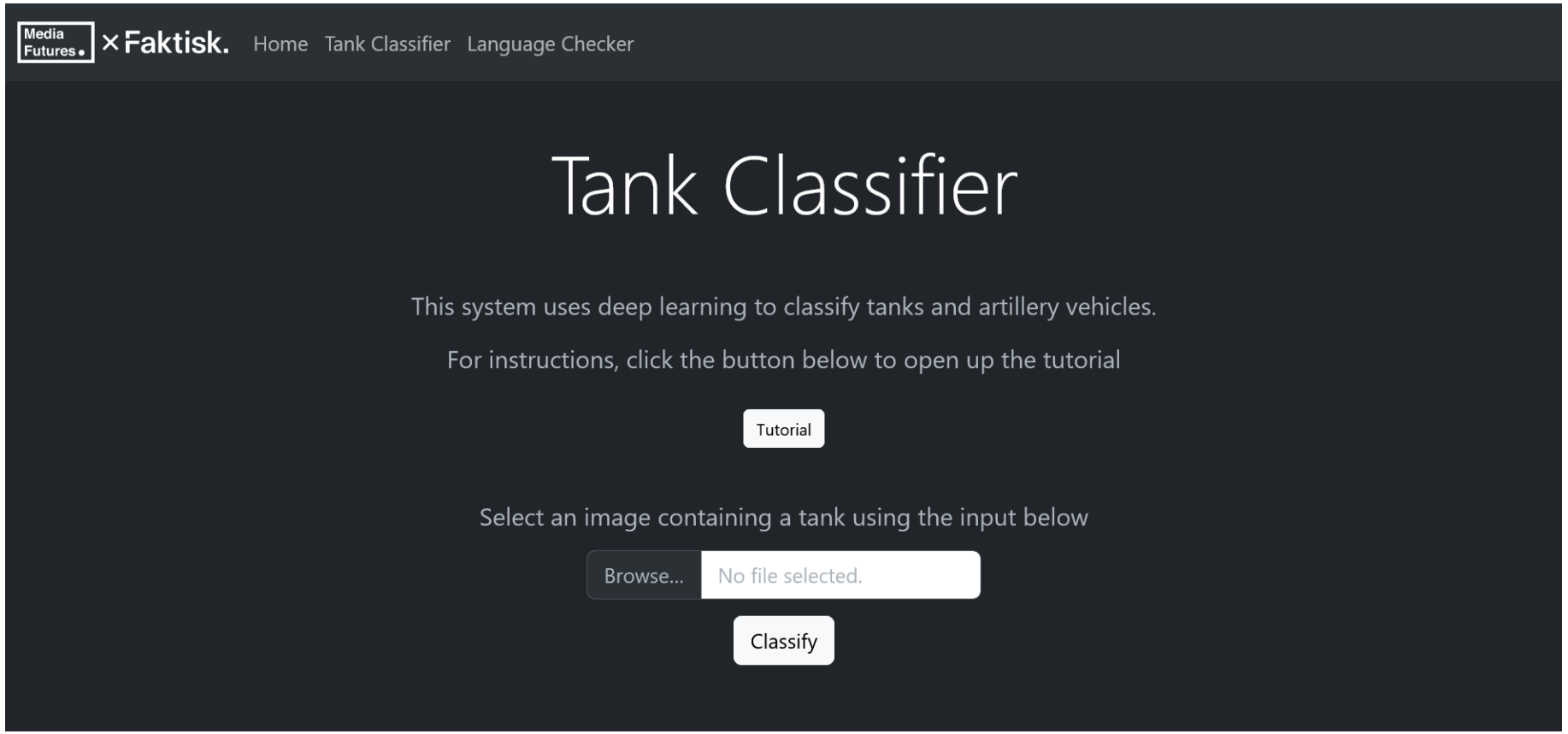
**Canon EOSR Rear**

# Development of the Tank Classifier

MediaFutures

X

Faktisk




https://faktisk.demo.mediafutures.no/tank/

@NORDISHUB

This project has received funding from the European Union

# Paths for improving AI-based tools

## Considering the user needs (accuracy, reliability, trustworthiness)

Development of a data quality assessment framework to assess the quality of datasets used in automated fact-checking:

- "The Ethical Dimensions of Data Quality for Automated Fact-Checking" (presented at C+J / European Data Journalism Joint Conference, Zürich)
- Extension of the framework to AI-based tools in journalism (CARMA Conference, Sevilla)

## Encouraging interdisciplinarity = developing common language + epistemology

Dierickx, L., & Lindén, C. G. (2023). **Fine-Tuning Languages: Epistemological Foundations for Ethical AI in Journalism**. In Proceedings 10th IEEE Swiss Conference on Data Science (SDS) (pp. 42-49). IEEE.



This project has received funding from the European Union

# Focus 4: The challenges of generative artificial intelligence



This project has received funding from the European Union

# Detecting deep-and-cheapfakes

## Establishing a research community and formulating a baseline for detecting out-of-context <image, caption> pairs

- Organising grand research challenges on cheapfake detection at top venues in Multimedia (ACM MMSys, ACM MM, IEEE ICME)
- Proposing baseline methods on cheapfake detection

Developing techniques for evaluating the quality of generated content, such as images, videos, or sentences, by generative AI


Pham, K.L., Nguyen, M.T., Tran, A.D., Dao, M.S. and **Dang-Nguyen, D.T.**, 2023, June. Detecting Cheapfakes using Self-Query Adaptive-Context Learning. In *ACM ICMR 2023*

Tran, Q.T., Tran, T.P., Dao, M.S., La, T.V., Tran, A.D. and **Dang Nguyen, D.T.**, 2022, October. A Textual-Visual-Entailment-based Unsupervised Algorithm for Cheapfake Detection. In *ACM MM 2022*.

Nguyen, T.S., Dang, V., Tran, M.T. and **Dang-Nguyen, D.T.**, 2023. Leveraging Cross-Modals for Cheapfakes Detection. In *ACM ICMR 2023*

La, T.V., Dao, M.S., Tran, Q.T., Tran, T.P., Tran, A.D. and **Dang-Nguyen, D.T.**, 2022, October. A Combination of Visual-Semantic Reasoning and Text Entailment-based Boosting Algorithm for Cheapfake Detection. In *ACM MM 2022*.

Moholdt, E., **Khan, S.A.**, and **Dang-Nguyen, D.T.**, 2023, September. Detecting OOC Image Caption Pair in News: A Counter-Intuitive Method. In *ACM CBMI 2023*.





This project has received funding from the European Union

# Addressing biases and “hallucinations”

NORDIS

The rapid rise of generative artificial intelligence (GAI) systems among newsrooms has brought to the fore the potential of technology to complement or augment journalism

Resurgence of ethical discussions since these systems are prone to errors and bias: risks of amplifying existing biases or spreading misinformation

How are European news media and self-regulatory bodies addressing these ethical challenges?

Corpus analysis of 33 guidelines and recommendations

**Need for human oversight and approval of any (G)AI-generated content to mitigate biases and prevent the proliferation of fabricated content**

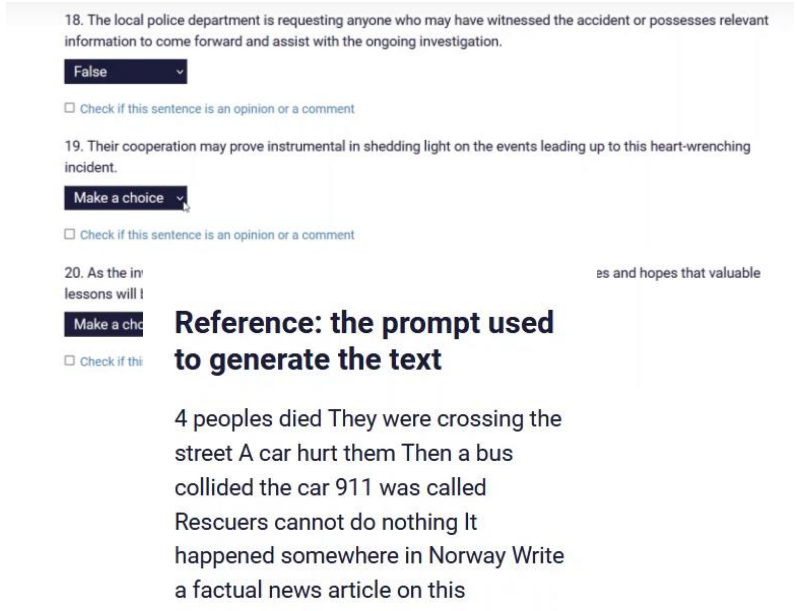
**Responsible uses require responsible tools**

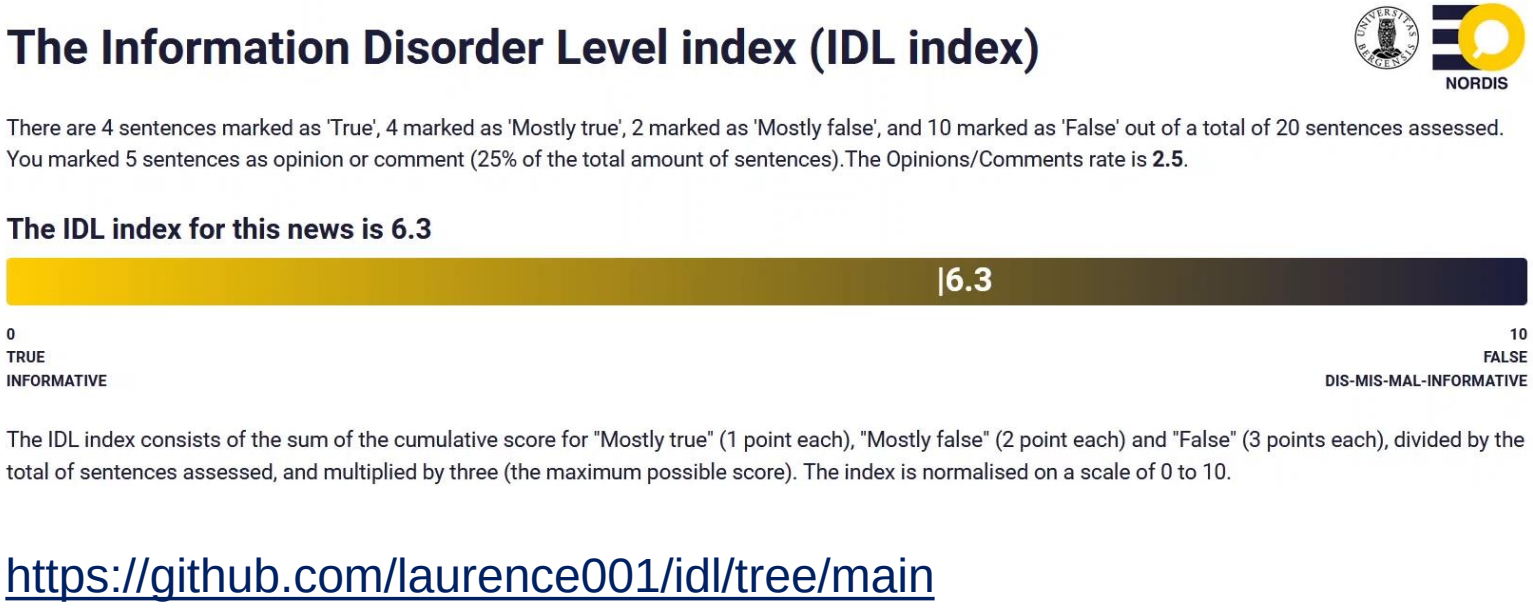


This project has received funding from the European Union

# Quantifying factuality: the IDL index

The Information Disorder Level (IDL) index is a human-based judgement metric, grounded in NLP human-based assessment tradition. It is language-independent and was developed to evaluate the factuality of machine-generated content.




https://github.com/laurence001/idl/tree/main

@NORDISHUB

This project has received funding from the European Union

# Thank you!

**Visit the NORDIS website!**
www.nordishub.eu

@NORDISHUB

This project has received funding from the European Union