

# A Multidisciplinary Approach to Tackling Online Misinformation

Udo Kruschwitz  
University of Regensburg  
`udo.kruschwitz@ur.de`

2 April 2023



Universität Regensburg

# COURAGE: A Multidisciplinary Approach to Tackling Online Misinformation

Udo Kruschwitz

(joint work with many others)

<https://www.upf.edu/web/courage>

# **COURAGE: A Multidisciplinary Approach to Tackling Online Misinformation ... and other Toxic Content**

Udo Kruschwitz

(joint work with many others)

<https://www.upf.edu/web/courage>

# Where to start?

**... let's ask ChatGPT ...**



# Today's Talk



I have been invited to give a keynote talk at the ROMCIR workshop on Reducing Online Misinformation through Credible Information Retrieval. I want to present an educational approach that we follow in the COURAGE project. How should I structure my talk?

# Today's Talk



Congratulations on being invited to give a keynote talk at the ROMCIR workshop! Here are some suggestions for structuring your talk:  

1. Introduction: Start by introducing yourself and your background in the field of reducing online misinformation. Briefly explain the focus of the ROMCIR workshop and how your presentation relates to it.
2. Overview of the COURAGE project: Provide a brief overview of the COURAGE project, including its goals and objectives. Explain how the project aims to reduce online misinformation through education.
3. Educational approach: Describe the educational approach that the COURAGE project follows in detail. Explain how this approach helps to reduce online misinformation by promoting critical thinking and digital literacy skills.
4. Examples: Provide some examples of how the COURAGE project has implemented its educational approach. This could include case studies, examples of successful initiatives, or examples of specific educational tools or resources.
5. Results: Share any results or findings from the COURAGE project that support the effectiveness of its educational approach. This could include data on changes in user behavior or attitudes towards online misinformation.
6. Conclusion: Summarize your key points and emphasize the importance of promoting digital literacy and critical thinking to reduce online misinformation. Encourage attendees to consider implementing similar approaches in their own work.
7. Q&A: Allow time for questions and discussion at the end of your presentation. Be prepared to answer questions about the educational approach, examples, and results you presented, as well as broader questions about reducing online misinformation.

# 1. Introduction



# Who am I?



# Some Activities I am Involved in

**Information Retrieval specialist group**

We're a community of IT professionals and enthusiasts who are passionate about information retrieval.

**Welcome to our group!**

If you work with search engines, develop or deploy search engines, have research interest in the field, or if you simply want to learn more, we will be happy to welcome you to our group!

We bring together industry, academia and end users with a focus on professional, intranet, enterprise and ecommerce search engines.

Udo Kruschwitz, Chair

**Information Retrieval specialist group: Search Solutions**

Search Solutions is an annual conference that is usually held...

**LingoTowns**

Help us with our research in Natural Language Processing. Your fun becomes our data as we uncover ambiguities in natural expression.

**PLAY**

**SIGNAL**

**Signal AI**  
@SignalHQ

Media monitoring, reputation management, regulatory compliance and market...

**Signal AI**  
@SignalHQ

We're so thrilled to be on the 'Best of the Best' list at @innovateuk after winning the 'Best of the Best' award!

**Innovate UK**

Best of the Best  
10 November 2015, London

11 November 2015 — News story  
**Media tracking firm wins Knowledge Transfer Partnership 2015**  
Signal Media and Essex University carry off 'Best of the Best' award after KTP associate joins company as head of research.

© 2019 Twitter About Help Centre Privacy policy Cookies Ads

**ECIR 2023**

6th April 2023  
Dublin, Ireland

# Selected Papers



Universität Regensburg

Research Article

## Interactive query expansion for professional search applications

**Tony Russell-Rose**   
Goldsmiths, University of London

**Philip Gooch**  
Scholarcy, UK

**Udo Kruschwitz**  
Universität Regensburg, Germany



## Exploring Fake News Detection with Heterogeneous Social Media Context Graphs

Gregor Donabauer<sup>(✉)</sup> and Udo Kruschwitz

Information Science, University of Regensburg, Regensburg, Germany  
{gregor.donabauer, udo.kruschwitz}@ur.de

Germany  
UK

## Aggregating Crowdsourced and Automatic Judgments to Scale Up a Corpus of Anaphoric Reference for Fiction and Wikipedia Texts

**Juntao Yu**<sup>1</sup>, **Silviu Paun**<sup>2\*</sup>, **Maris Camilleri**<sup>1</sup>, **Paloma Carretero Garcia**<sup>1</sup>,  
**Jon Chamberlain**<sup>1</sup>, **Udo Kruschwitz**<sup>3</sup> and **Massimo Poesio**<sup>4</sup>

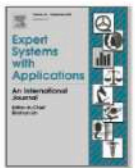
<sup>1</sup>University of Essex, UK; <sup>2</sup>Amazon Research, Romania;

<sup>3</sup>University of Regensburg, Germany; <sup>4</sup>Queen Mary University of London, UK;  
j.yu@essex.ac.uk; silviupn@amazon.com; mcamil@essex.ac.uk; pcarre@essex.ac.uk;  
jchamb@essex.ac.uk; udo.kruschwitz@ur.de; m.poesio@qmul.ac.uk;

## Signal Briefings: Monitoring News Beyond the Brand

James Brill<sup>1</sup>, Dyaa Albakour<sup>1</sup>, José Esquivel<sup>1</sup>, Udo Kruschwitz<sup>2</sup>, Miguel  
Martinez<sup>1</sup> and Jon Chamberlain<sup>3</sup>

Applications  
0, 113365



## Rights Monitoring

✉, Chris Fox <sup>a</sup> ✉, Massimo Poesio <sup>b</sup> ✉

Vivenhoe Park, Colchester CO4 3SQ, United Kingdom  
University of London, Mile End Rd, London E1 4NS, United Kingdom  
El Fayoum, Faiyum Governorate, Egypt  
Civilian Rights, 54 Commercial Street, London E1 6LT, United Kingdom

**... enough about me**

# ROMCIR 2023

## *ROMCIR 2023*

The 3rd Workshop on Reducing Online  
Misinformation through Credible  
Information Retrieval – April 2nd, Dublin,  
Ireland

Motivation and Relevance to  
ECIR

Aim and Themes

Call for Papers

Schedule (tentative)

Keynote Speakers

Organizers

For these reasons, the themes of interest include, but are not limited to, the following:

- Access to genuine information
- Bias detection
- Bot/spam/troll detection
- Computational fact-checking
- Crowdsourcing for information genuineness assessment
- Deep fakes
- Disinformation/misinformation detection
- Evaluation strategies to assess information genuineness
- Fake news/review detection
- Harassment/bullying/hate speech detection
- Information polarization in online communities, echo chambers
- Propaganda identification/analysis
- Retrieval of genuine information
- Security, privacy, and information genuineness
- Sentiment/emotional analysis
- Stance detection
- Trust and reputation
- Societal reaction to misinformation

# Context

Egypt

## Egypt to fine creators of weather misinformation

Egypt Independent • April 1, 2023

REUTERS® World Business Legal Markets Breakingviews Technology Investigation

Disrupted  
2 minute read · March 27, 2023 9:58 PM GMT+2 · Last Updated 5 days ago

### Europol sounds alarm about criminal use of ChatGPT, sees grim outlook

### Support the Guardian

Fearless, independent, reader-funded

Support us →

News Opinion Sport Culture Lifestyle More

The Guardian view Columnists Cartoons Opinion videos Letters

Opinion ChatGPT

A fake news frenzy: why ChatGPT could be disastrous for truth in journalism  
*Emily Bell*

CHATHAM HOUSE

### Statement from the listed authors of Stochastic Parrots on the "AI pause" letter

Timnit Gebru, Emily M. Bender, Angelina McMillan-Major, Margaret Mitchell

March 31, 2023

future of life INSTITUTE

### Pause Giant AI Experiments: An Open Letter

We call on all AI labs to immediately pause for at least 6 months the training of AI systems more powerful than GPT-4.

### ChatGPT has opened a new front in the fake news wars

Search engines with the latest 'generative AI' obscure the sources for their responses. The result is a breeding ground for disinformation, writes Jessica Cecil.

THE WORLD TODAY 31 MARCH 2023 — 4 MINUTE READ

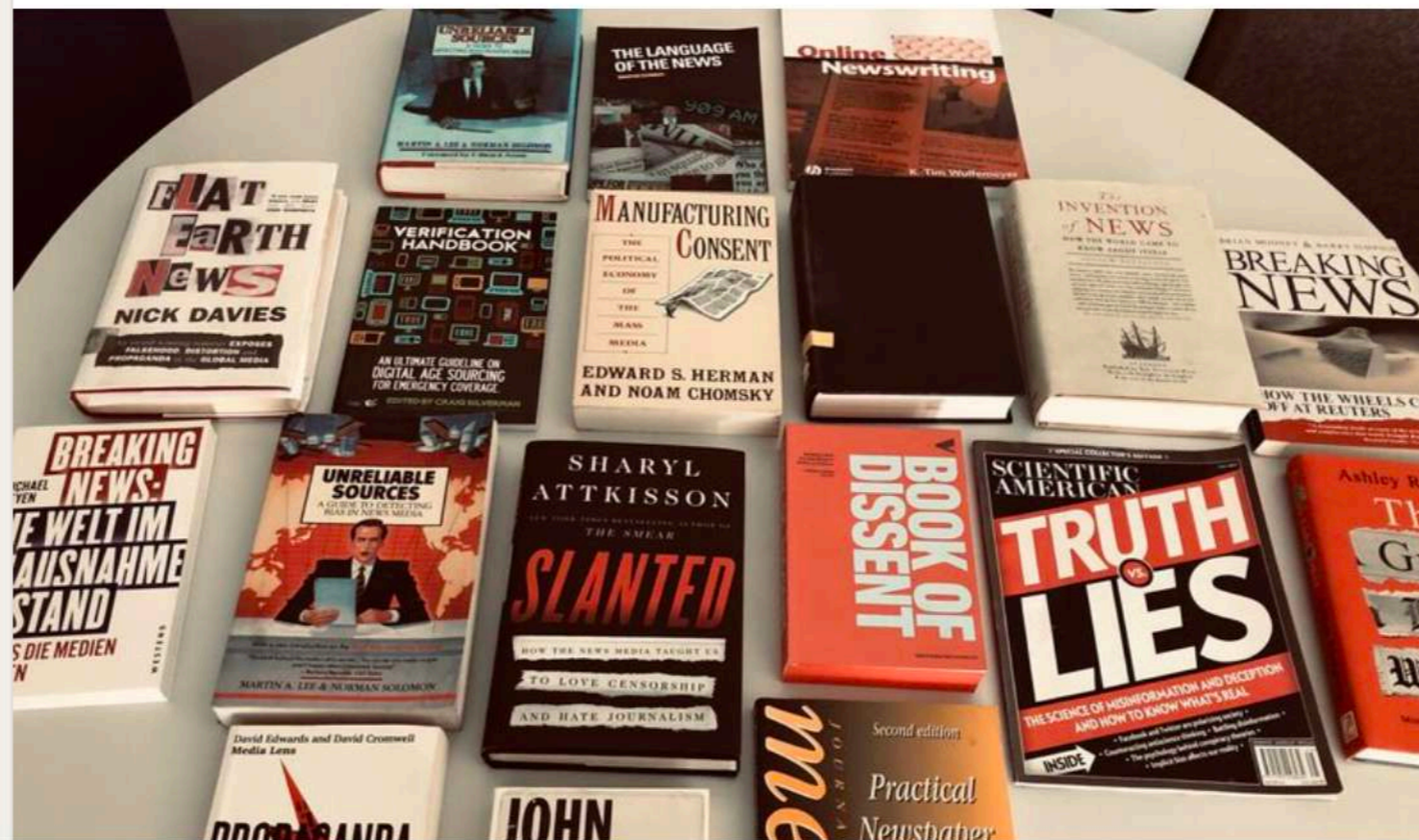
**... obviously this touches on much broader issues, so let's get back to our key concern here ...**

# Context



**Jochen L. Leidner (He/Him) • 1st**  
 AI Professor | Scientist-Engineer | Consultant | Advisor  
 2h • Edited •

Lecture "Media Manipulation, Propaganda & Fake News": attempting to fill the gap that journalists left... How does Fake News production and detection work? Who are the actors behind Brexit, Trump, Ukraine? Some familiar names, some unfamiliar ones... What are the motives and goals? let us go to consult the sources.





# Context

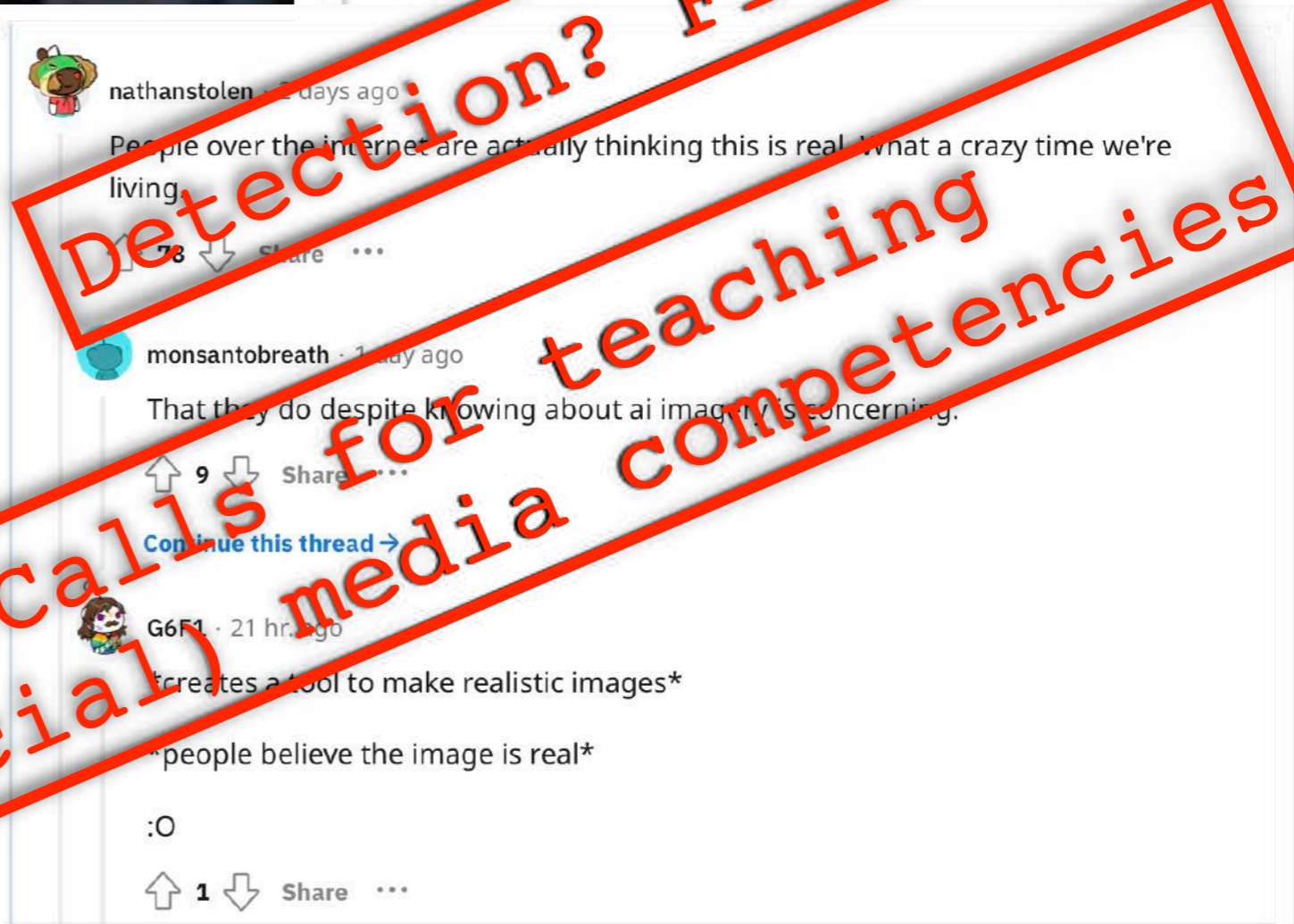


# Context



Detection? Filtering?

calls for teaching (social) media competencies








# The Problem

*Natural Language Engineering* (2022), 1–26  
doi:10.1017/S1351324922000481



EMERGING TRENDS

## Emerging trends: Unfair, biased, addictive, dangerous, deadly, and insanely profitable

Kenneth Church<sup>1,\*</sup> , Annika Schoene<sup>1</sup> , John E. Ortega<sup>1</sup> , Raman Chandrasekar<sup>1</sup>  and Valia Kordoni<sup>2</sup> 

<sup>1</sup>Institute for Experiential AI, Northeastern University, Boston, MA 02115, USA and <sup>2</sup>Humboldt-Universität zu Berlin, Berlin, Germany

\*Corresponding author. E-mail: [k.church@northeastern.edu](mailto:k.church@northeastern.edu)

(Received 15 November 2022; revised 15 November 2022; accepted 15 November 2022)

There is a growing concern that much of this criticism could also be applied to the NLP community. There is a real danger that the court of public opinion may not view our work on Risks 1.0 as part of the solution and might even see our work as part of the problem. We need to make progress on both Risks 1.0 as well as Risks 2.0.

The books mentioned above (Fisher 2022; Bergen 2022) have quite a bit to say about Risks 2.0. Many academics are mentioned (e.g., Chaslot, DiResta, Farid, Kaiser, Müller, Rauchfleisch, Schwarz), but there is relatively little discussion of our toxicity classifiers. Our classifiers may not have the impact we would hope because there are few incentives for social media companies to reduce toxicity, as will be discussed in Section 6.

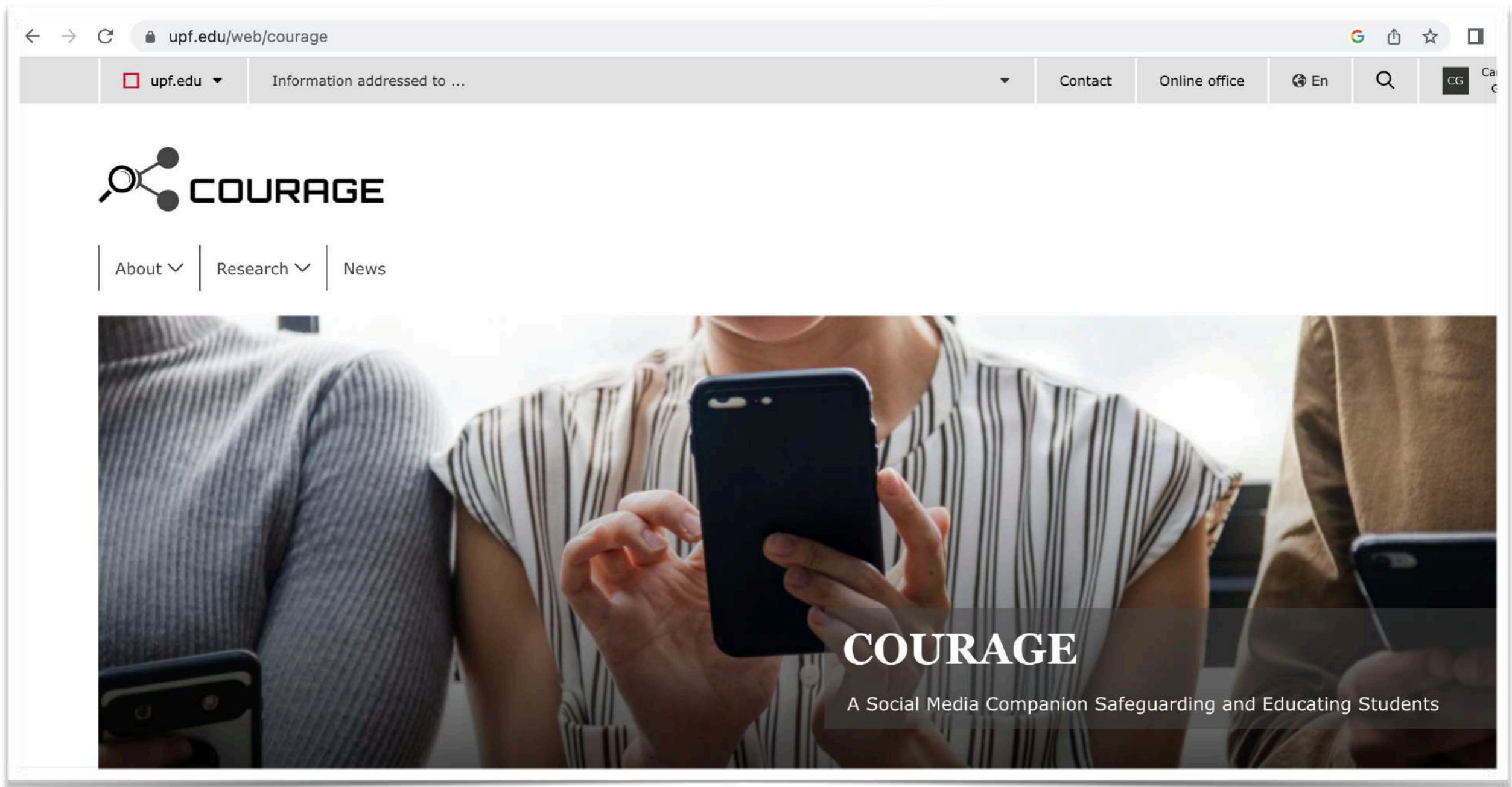
criticisms from other fields and focus on implications for computer science, we will not attempt to contribute to those other fields. There is quite a bit of work in computer science on these risks, especially on Risks 1.0 (bias and fairness), but more work is needed, especially on Risks 2.0 (addictive, dangerous, and deadly).

## 2. COURAGE Project

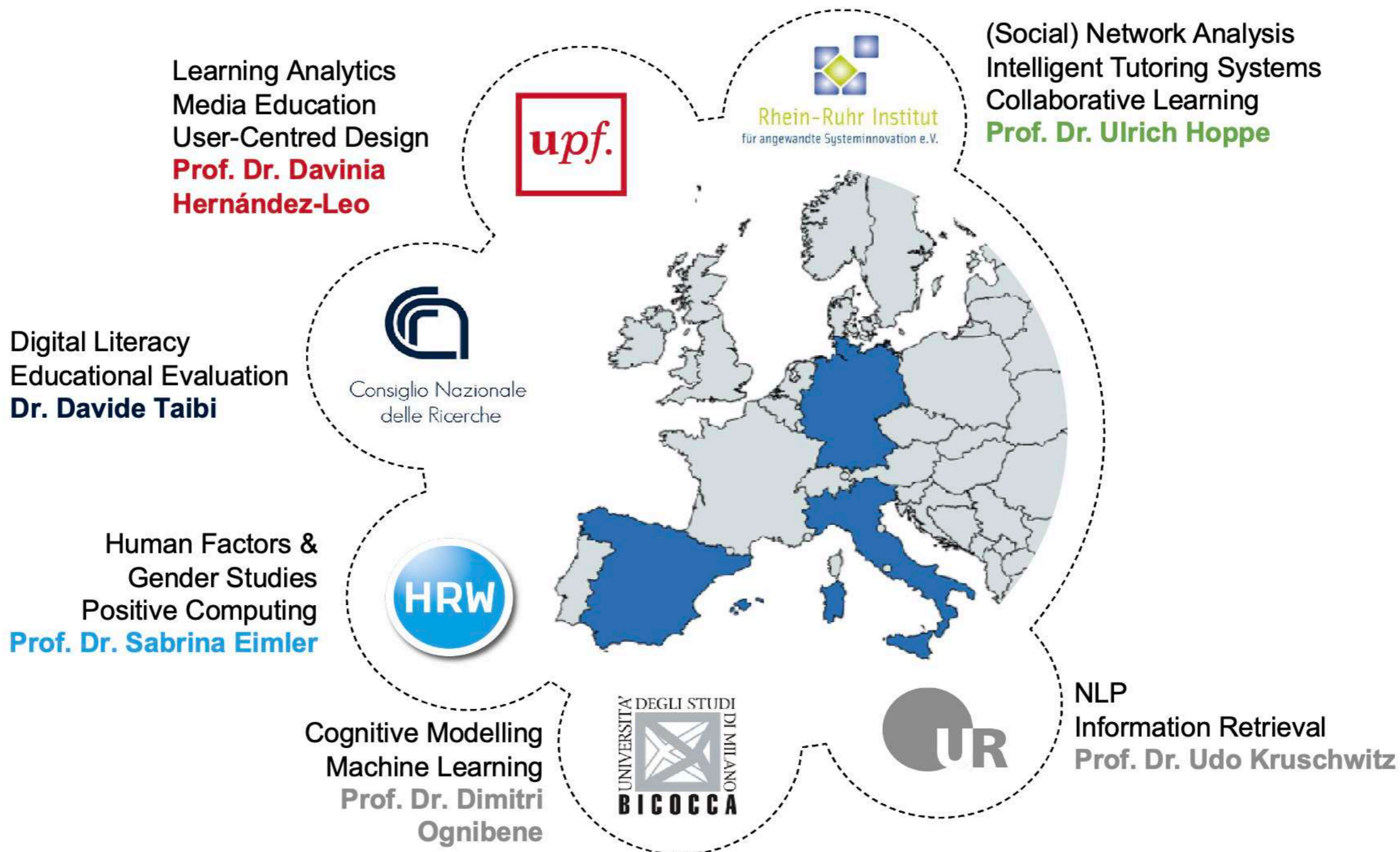
# COURAGE Project



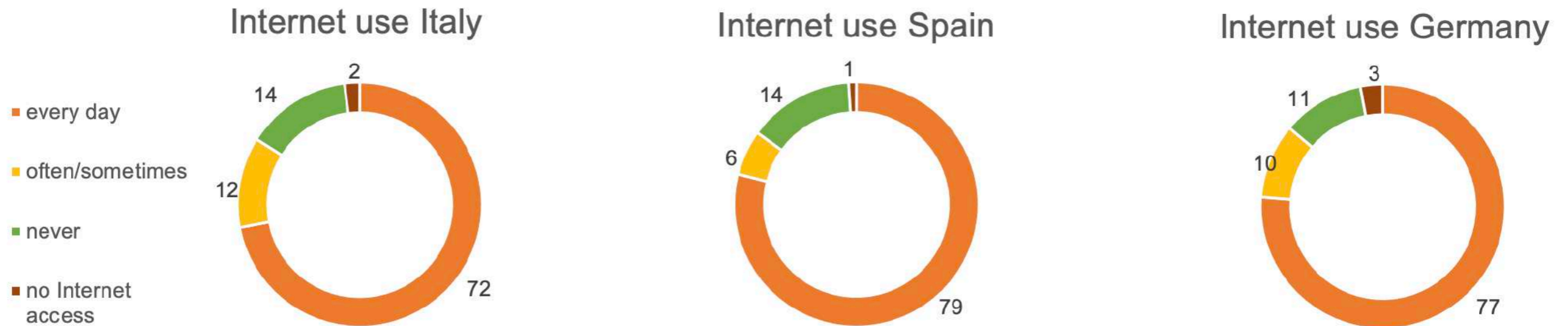
Universität Regensburg



# COURAGE: International & Multidisciplinary

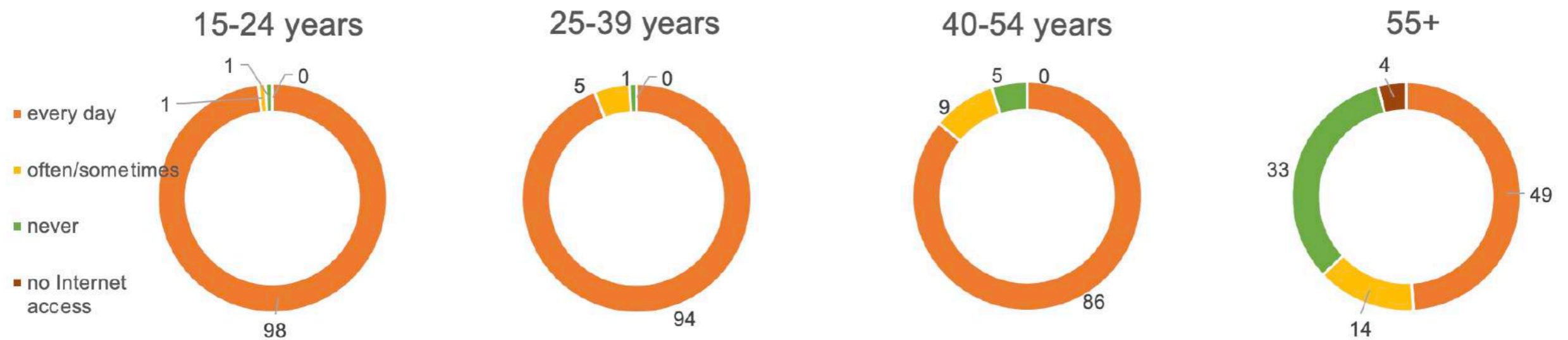


# Motivation: Internet Use per Country



- Overall **76%** of EU citizens use the **Internet daily** and further **9%** use it **often** or **sometimes**
- **Smartphones** are the main means of accessing the Internet (European Commission, 2020)

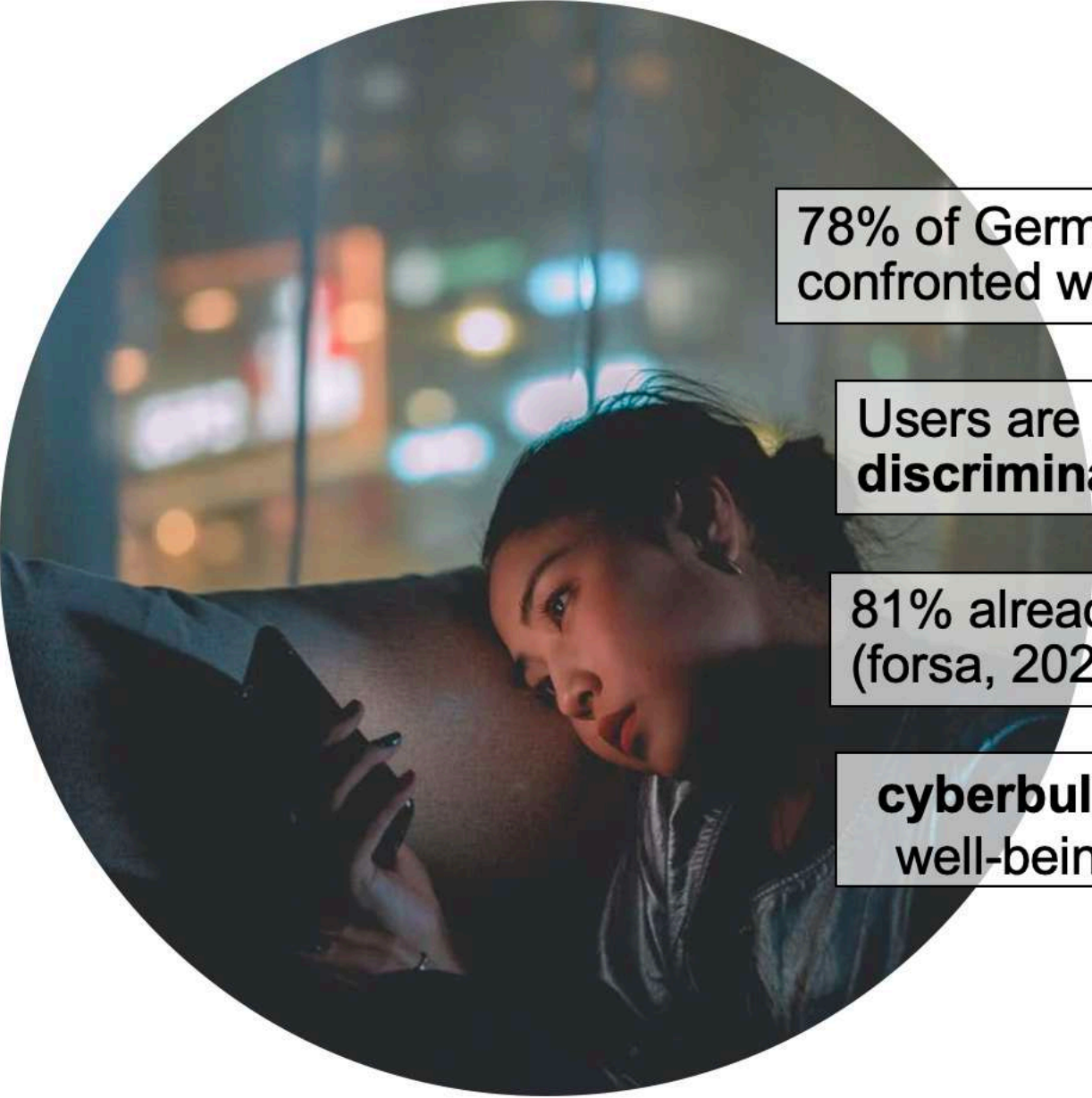
# Motivation: Internet Use in EU by Age



- Using **online social networks** is the most mentioned activity in six countries of the EU
- Overall, using **online social networks** is one of the three most mentioned activities in 17 countries (European Commission, 2020)



# Motivation: Downsides of Social Media



78% of German Internet users aged 14-24 were already confronted with **cyberbullying** on social media (Newall, 2018)

Users are confronted with **racism** and **discrimination** on social media (Wang et al., 2021)

81% already dealt with **hate comments** (forsa, 2022)

**cyberbullying** and **cyberhate** impact on adolescents' well-being and emotions (Fulantelli et al., 2022)

# Motivation: Consequences of Engaging with Toxic Content



Somatic complaints (Herge et al., 2016)

Depressive symptoms (Ybarra et al., 2011)

Suicidal symptoms (Zaborskis et al., 2018)

Emotional problems (Wigderson & Lynch, 2013)

Poorer academic performance (Wigderson & Lynch, 2013)

Less innovation, less creativity (Wigderson & Lynch, 2013)

Post-traumatic stress symptoms (Baldry et al., 2019)

## Some more Facts

- 91% of 16-24 year olds use the Internet for social networking
- But: inability to deal with threats and biases (e.g. 80% believe sponsored content) [Wineburg et al., 2016; Horowitz, 2018]
- **Social media literacy skills (SML) can improve** users' experience in social media platforms
- The inclusion of SML interventions in schools can lead to social media-literate individuals that feel in control

[1] Wei, L., Gong, J., Xu, J., Eeza Zainal Abidin, N., & Destiny Apuke, O. (2023). Do social media literacy skills help in combating fake news spread? Modelling the moderating role of social media literacy skills in the relationship between rational choice factors and fake news sharing behaviour. [2] Paxton, S. J., McLean, S. A., & Rodgers, R. F. (2022). "My critical filter buffers your app filter": Social media literacy as a protective factor for body image. [3] Tamplin, N. C., McLean, S. A., & Paxton, S. J. (2018). Social media literacy protects against the negative impact of exposure to appearance ideal social media images in young adult women but not men. [4] Mingoia, J., Hutchinson, A. D., Gleaves, D. H., & Wilson, C. (2019). The impact of a social media literacy intervention on positive attitudes to tanning: A pilot study. *Computers in Human Behavior*, 90, 188–195 [5] Schreurs, L., & Vandenbosch, L. (2022). Should I post my very best self? The within-person reciprocal associations between social media literacy, positivity-biased behaviors and adolescents' self-esteem. *Telematics and Informatics*, 73

# What is Needed?



Strengthen self-protection skills among students

Increase students' awareness of potential threats

Help students counter the spread of toxic content

Improve digital literacy

**Educational interventions**

**Enable educators  
to support students!**

# 3. Educational Approach

# Approach of COURAGE: Educational Interventions

The **COURAGE project** seeks to develop social media interventions that provide students with **hands-on experiences** to learn about the threats and dangers that can exist within such environments.

**COURAGE** proposes integrating educational opportunities within a **simulated social media platform**.

**Virtual learning companions (VLCs)** transfer learning with the help of computer-simulated characters through interactive chat interface (Chou et al., 2003).

To achieve this, we explored the use of **VLCs** that allow users to interact directly with social media scenarios and receive immediate support, rather than teaching them in isolation from dangerous situations.

# Approach of COURAGE: Educational Interventions II

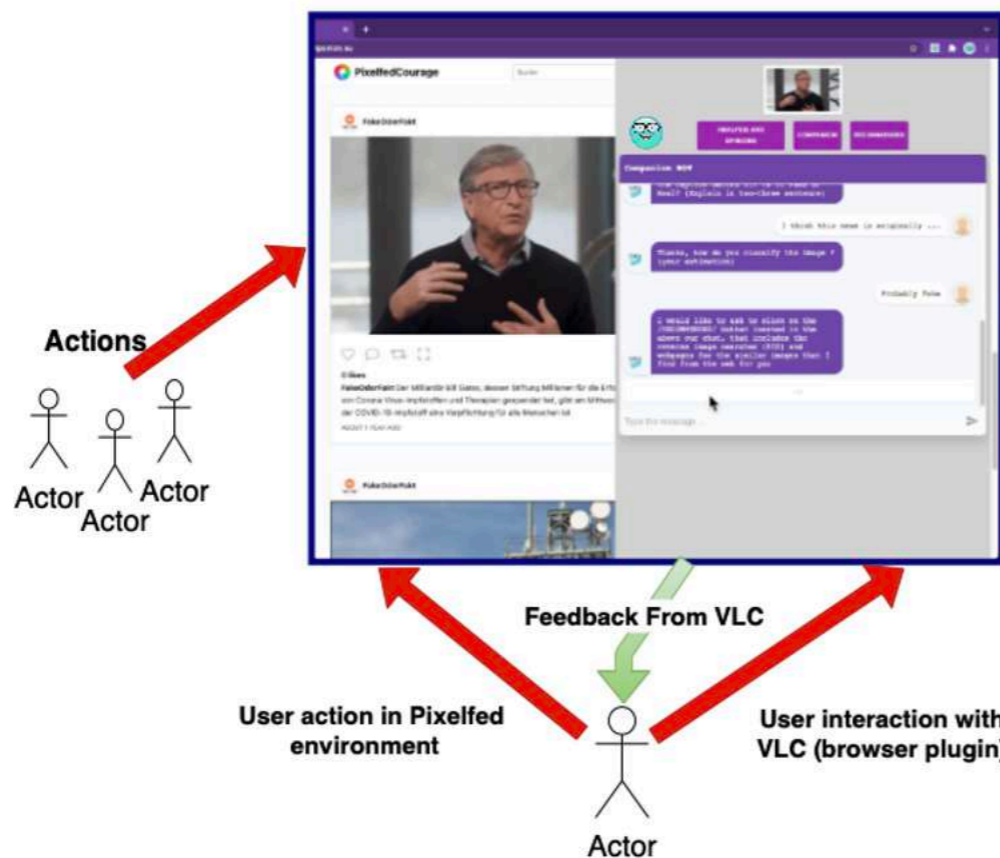
- COURAGE aims at **empowering adolescents** to confidently interact and utilize social media (addressing how to deal with **misinformation** as one aspect).
- A **multi-disciplinary** approach building on psychology and pedagogy combined with data science and AI-driven approaches
- Building on Intelligent Tutoring Systems: **Virtual Learning Companion** enabling learning and interaction support with the aim to raise awareness and resilience on the part of the learners

# Approach of COURAGE: Educational Interventions III

- Used in **classrooms** and informal settings providing a playful, adaptive and engaging setting
- Adolescents interact with a social media environment under restrictions of pedagogical responsibility and guidance
- Example component: **narrative scripts** implementing a collaborative learning flow pattern to **raise learners' awareness** and teach competencies

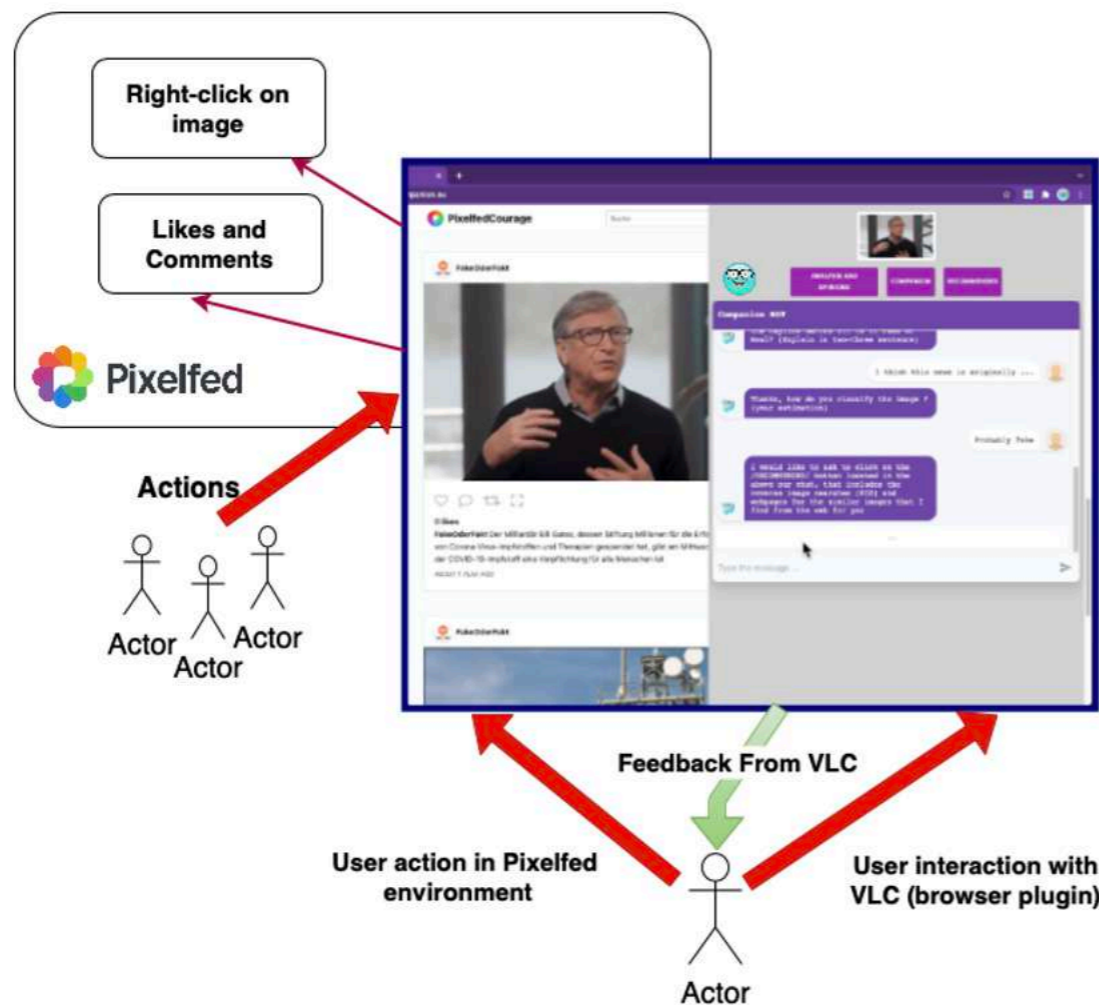


# COURAGE - Conceptual Architecture



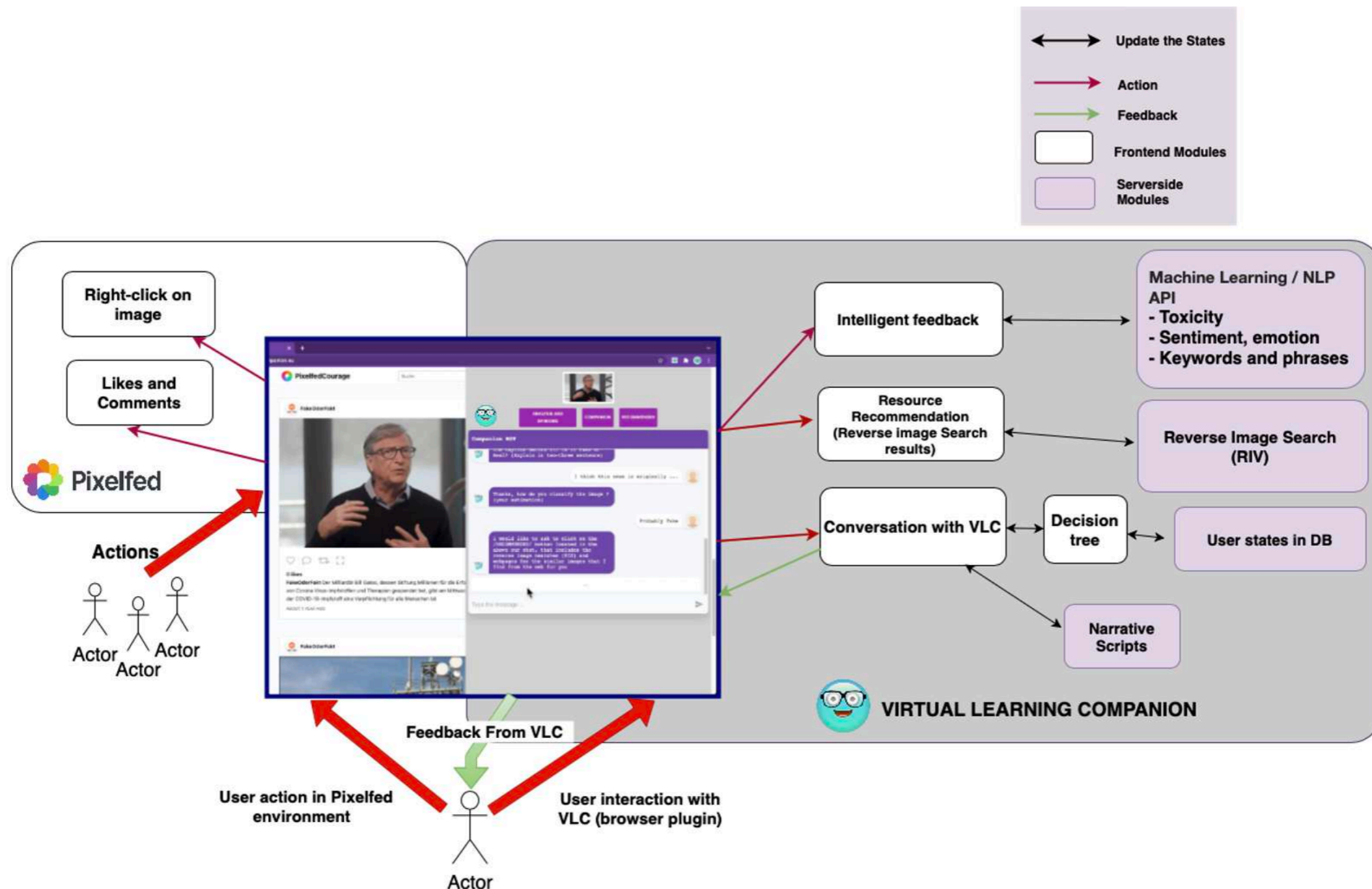
VIRTUAL LEARNING COMPANION

# COURAGE - Conceptual Architecture

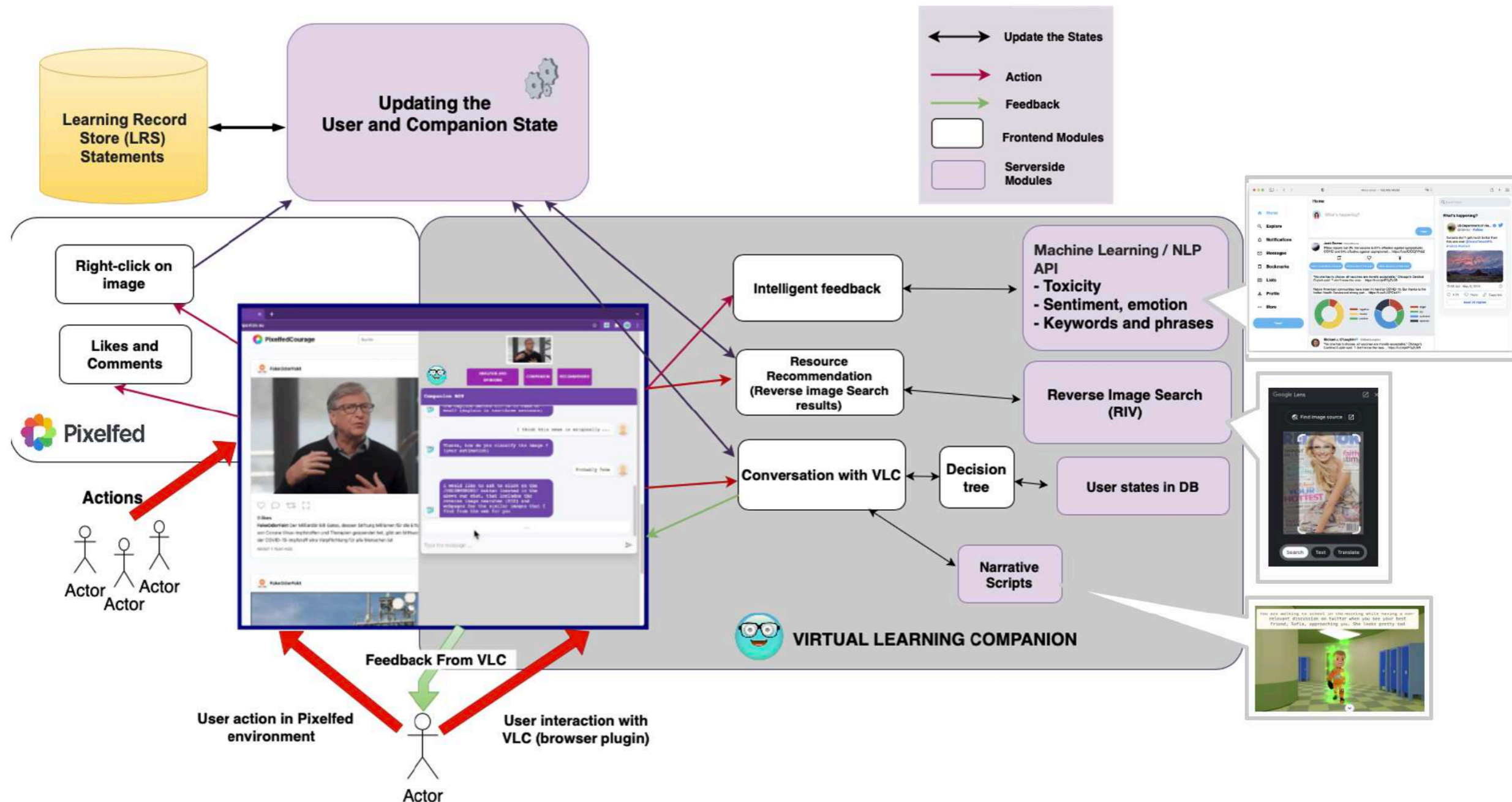


VIRTUAL LEARNING COMPANION

# COURAGE - Conceptual Architecture

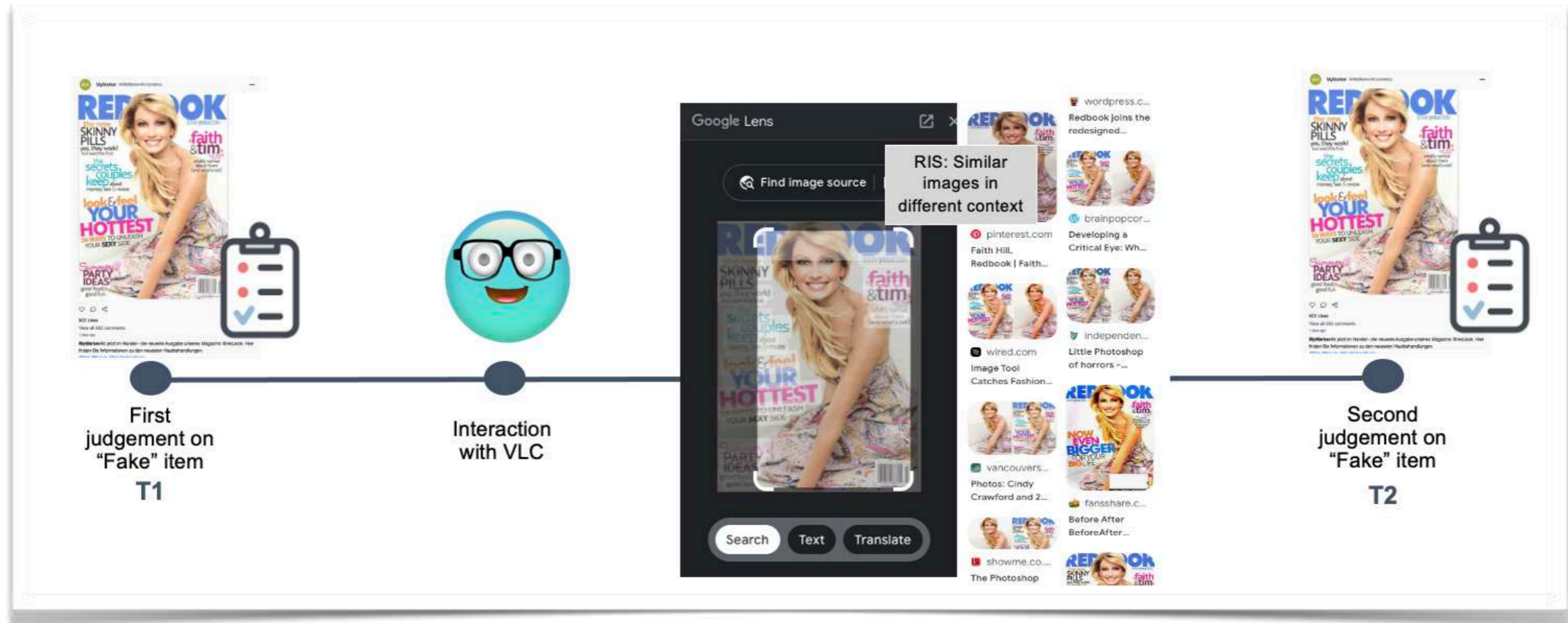


# COURAGE - Conceptual Architecture



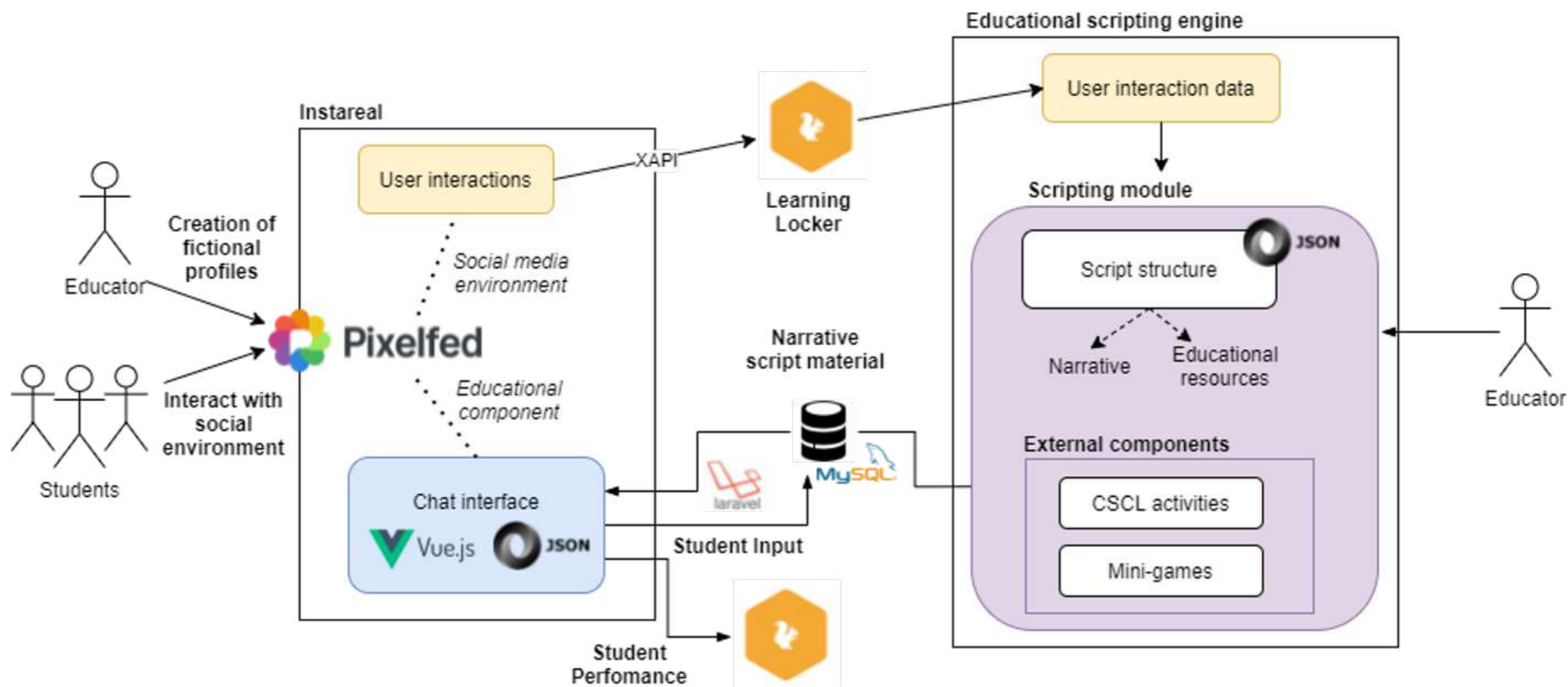
# 4. Examples

# 'Fake or Fact': Fake Image Detection

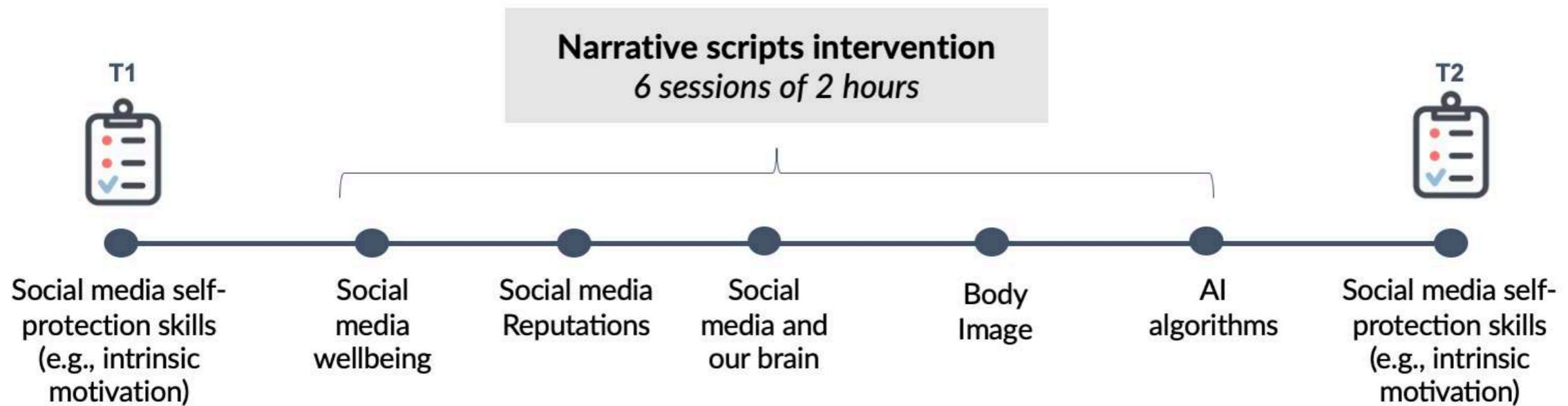


- Pilot study: aggregated distance of users ( $N = 22$ ) to expert judgements goes down (before/after VLC interaction)
- Access to alternative sources containing the same image supports correct judgement (RIS)

# Narrative Scripts



# Narrative Scripts II

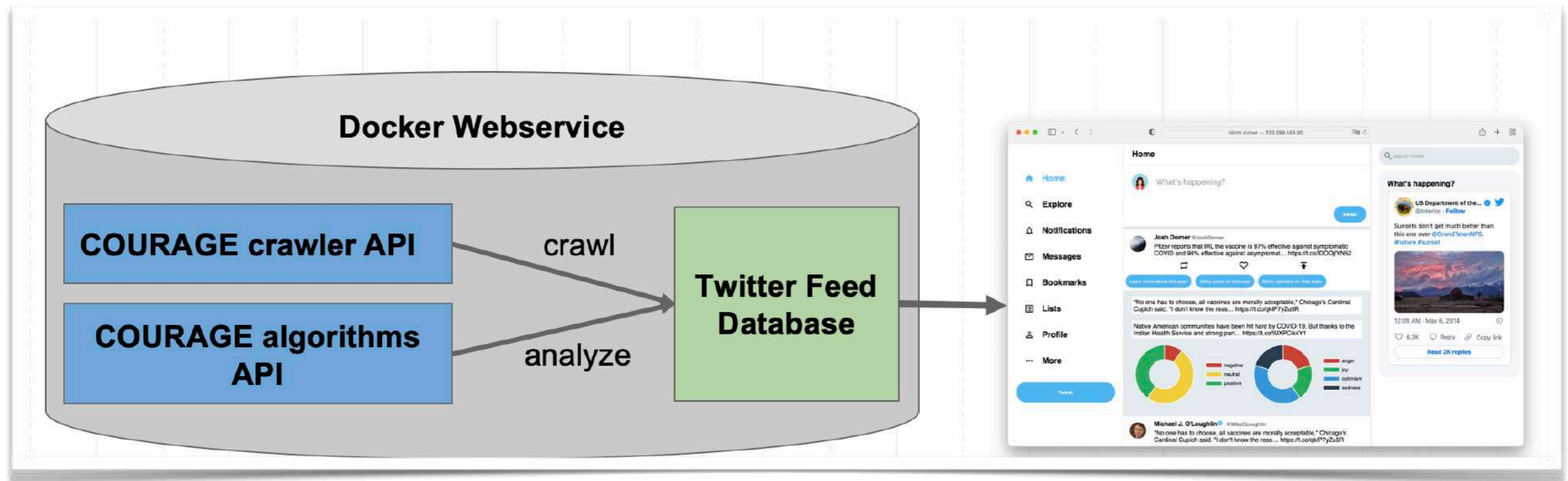


- Results for  $N = 125$  show that narrative scripts generate **intrinsic motivation**, especially in **interest** and **enjoyment** ( $M = 4.41$ ,  $SD = 1.25$ )
- Low perceived pressure/tension ( $M = 2.52$ ,  $SD = 1.21$ )

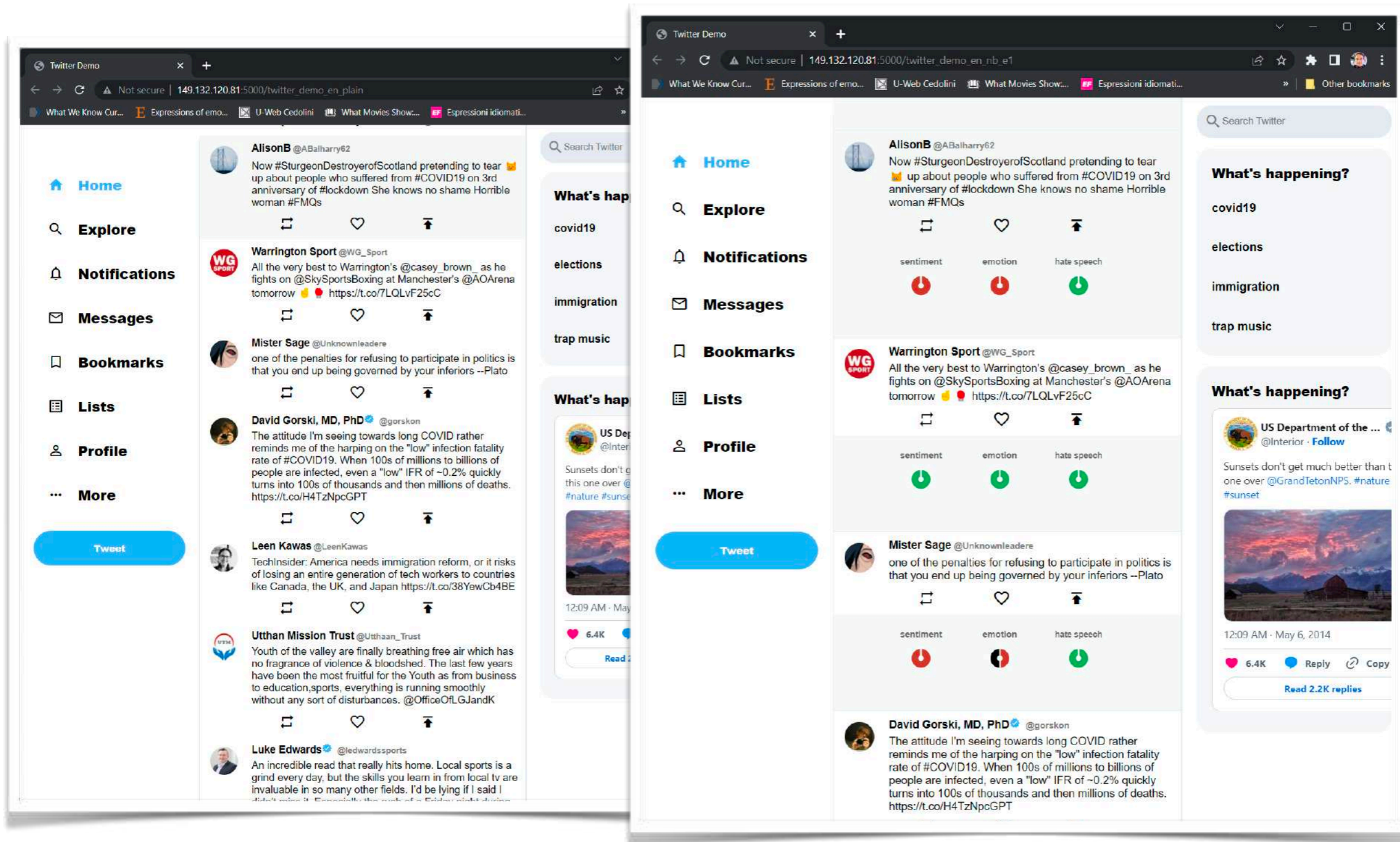


# Augmenting Social Media

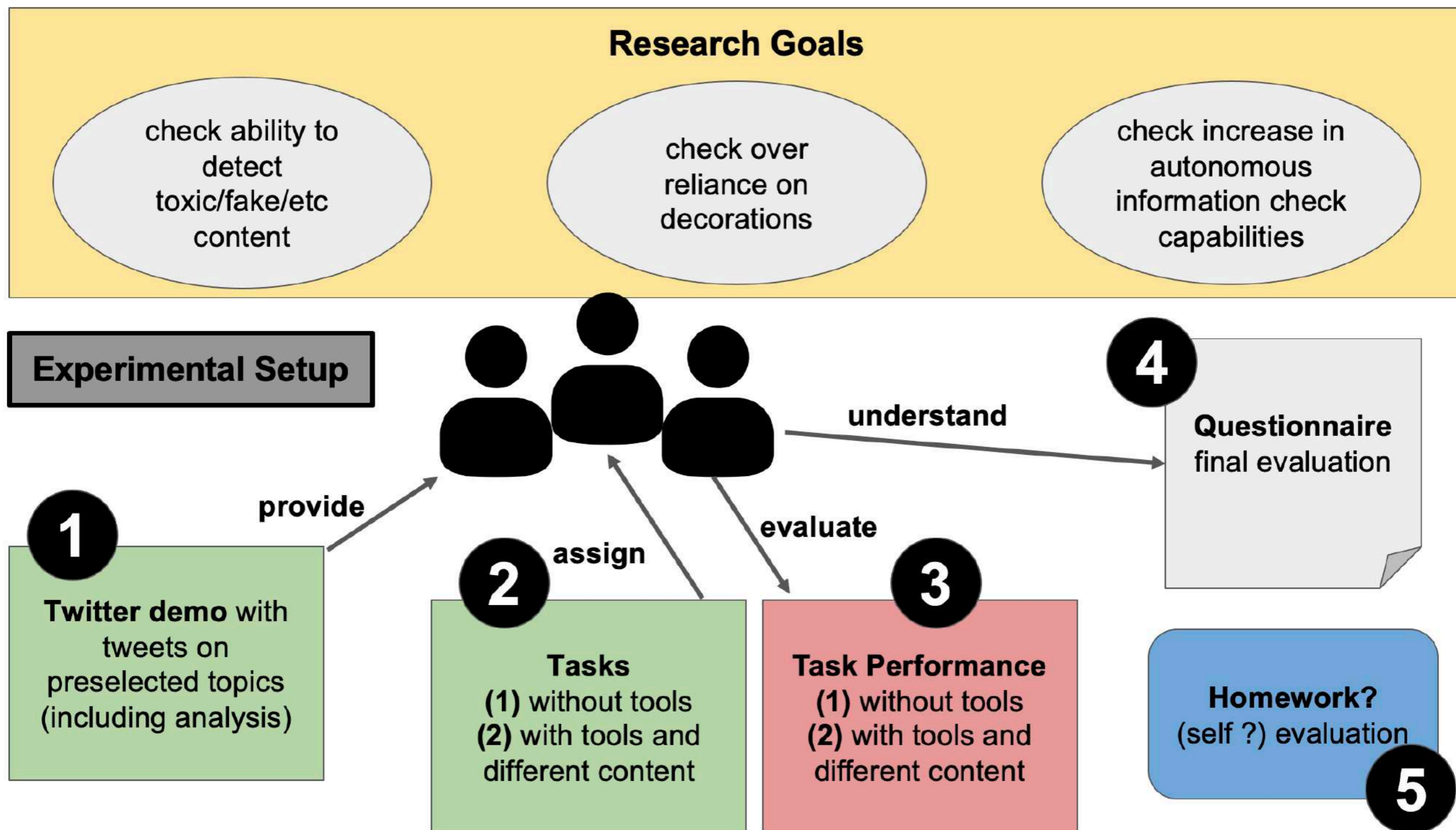
- Crawling new data from Twitter using Twitter API
- Analyzing crawled social media content with COURAGE API
- Providing analysis results in the social media feed



# Augmenting Social Media II



# Augmenting Social Media III



# 5. Future Directions

# Experimental Work

- School trials in Italy and Spain
- Crowdsourcing experiments (using Prolific)
- Explore **nudging and boosting** more systematically

# Nudging & Boosting

# Behavioural Economics

- Heuristics and biases → Nudging
- Adaptive rationality → Boosting

**P** Explain this

Privacy risk where  
Red = High privacy risk  
Yellow = Medium privacy risk  
Green = Low privacy risk

**MEDICAL QUESTION 2:** Does traction help low back pain?

**HEALTH ISSUE:** **back pain** - The spine is a column of bones (vertebrae) held together by muscles, tendons and ligaments and cushioned by shock-absorbing disks. A problem in any part of your spine can cause back pain. *Source: Mayo Clinic*

**TREATMENT:** **traction** - a pulling force exerted on a skeletal structure (as in a fracture) by means of a special device *Source: Merriam-Webster*

You can click on links below

[Make Decision](#)

*helpful:* The medical treatment **helps** if the treatment is effective and has a direct positive influence on the specified illness.

*inconclusive:* The effectiveness of a medical treatment is **inconclusive** if medical professionals are still unsure if the treatment will have a positive, negative or no influence on the specified illness.

*does not help:* The medical treatment **does not help** if the treatment is ineffective and either has no effect or has a direct negative influence on the specified illness.

24 results returned:

**P** **Lumbar Traction Offers No Benefit for Back Pain**  
<https://www.verywellhealth.com/>  
 Does Traction Really Work for Low Back Pain? A study confirms that using lumbar traction with exercise for low back pain does not offer improved outcomes when compared to physical therapy exercises alone.

**P** **Does Spinal Decompression Really Work in Treating Low Back Pain?**  
<https://www.verywellhealth.com/>  
 A Questionable Treatment Spinal decompression may help treat low back pain, but this popular treatment isn't a sure thing. Advertising for spinal decompression targets people with degenerative disc disease, bulging discs, herniated discs, or spinal stenosis.

**P** **Traction Therapy for Chronic Low Back Pain**  
<http://www.barclayphysicaltherapy.com/>  
 The cost of health care is rising every year in the United States. And part of that economic burden is the management of chronic low back pain (CLBP).

**P** **Traction for Low Back Pain With or Without Sciatica: An Updated Systematic Review Within the Framework of the Cochrane Collaboration**  
<https://www.researchgate.net/>  
 Systematic review. To determine if traction is more effective than reference treatments, placebo/sham traction, or no treatment for low back pain (LBP).

**Out of 100 websites visited for health & medical issues in a search engine:**

How many out of 100:	.gov/.org websites	other websites
<b>Benefits</b>		
will share your data with ≤ 2 3rd party companies?	59	14
<b>Harms</b>		
will share your data with ≥ 8 3rd party companies?	11	33

# Nudge vs. Boost

- **Nudge:** “A nudge, as we will use the term, is any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives. To count as a mere nudge, the intervention must be easy and cheap to avoid.” Thaler and Sunstein 2008
- **Boost:** “Goal of expanding (boosting) the decision maker's set of competences and thus helping them to reach their objectives” Grüne-Yanoff and Hertwig 2016



# Nudge vs. Boost

- **Nudges** apply libertarian paternalism for desired outcomes
- **Example:** in the UK there is no need to pay pension contributions but default is that you do, i.e. opt-out (similar example: opt-in vs. opt-out of organ donation)
- **Key point:** Low cost but problems returns after removal of Nudge

# Nudge vs. Boost

- **Boosts foster competencies** to produce desirable outcomes
- **Example:** vary sense of connection to future self, e.g.:
  - ▶ Your current pension contribution → No trips abroad during retirement
  - ▶ Increase your pension by X → 2 trips abroad per year during retirement
- **Key point:** Higher cost but competencies remain after removal of Boost

## How does it fit **COURAGE?**

- Boosts a particularly promising paradigm to strengthen online users' competencies and counteract the challenges of the digital world
- Seems the perfect scenario for addressing misinformation and false information, manipulative choice architectures etc.
- Educating the online users (i.e. teenagers) rather than imposing rules, restrictions or suggestions on them

# Boosting and Nudging in COURAGE:

## Some Research Questions

- **Compared to a control environment**, to what extent are (**boosting / nudging**) effective at increasing consumption of correct information?
- **Compared to a control environment**, what impacts do (**boosting / nudging**) have on the teaching media competencies?
- To what extent are the **claims about boosting compared to nudging** evident in our scenario?

# How do we test this?

- Between-group user studies using Latin square design
- Expose users to different systems (e.g. nudge vs. baseline/control) in a task-based evaluation
- Check for differences, e.g. by testing competencies (at different time intervals)

# 6. Conclusions

# Conclusions

- Misinformation comes in many shapes and colours
- The problem is here to stay (only getting worse)
- One way of addressing this: detection, filtering, labelling ... (main focus of ROMCIR for the rest of the day)
- Different (*complementary*) approach: live with it, deal with it (analogy: bias in LLMs)
- That requires a multidisciplinary team (expertise in NLP, ML, education, design ...)
- I took a high-level view on the second approach

# Some References to COURAGE

frontiers | Frontiers in Artificial Intelligence

TYPE Hypothesis and Theory  
PUBLISHED 09 January 2023  
DOI 10.3389/frai.2022.654930

Check for updates

## OPEN ACCESS

EDITED BY  
Julita Vassileva,  
University of Saskatchewan, Canada

REVIEWED BY  
Chien-Sing Lee,  
Sunway University, Malaysia  
Li Liu,  
Chongqing University, China  
Jorge Luis Bacca Acosta,  
Konrad Lorenz University Foundation,  
Colombia

\*CORRESPONDENCE  
Dimitri Ognibene  
✉ dimitri.ognibene@unimib.it  
Davide Taibi  
✉ davide.taibi@itd.cnr.it

SPECIALTY SECTION  
This article was submitted to  
AI for Human Learning and Behavior  
Change,  
a section of the journal  
Frontiers in Artificial Intelligence


## Challenging social media threats using collective well-being-aware recommendation algorithms and an educational virtual companion

Dimitri Ognibene<sup>1,2\*</sup>, Rodrigo Wilkens<sup>3</sup>, Davide Taibi<sup>4\*</sup>,  
Davinia Hernández-Leo<sup>5</sup>, Udo Kruschwitz<sup>6</sup>,  
Gregor Donabauer<sup>6</sup>, Emily Theophilou<sup>5</sup>,  
Francesco Lomonaco<sup>1</sup>, Sathya Buršić<sup>1</sup>, Rene Alejandro Lobo<sup>5</sup>,  
J. Roberto Sánchez-Reina<sup>5</sup>, Lidia Sánchez-Reina<sup>5</sup>,  
Johanna Börsting<sup>7</sup>, Ulrich Hoppe<sup>8</sup>,  
Nils Malzahn<sup>8</sup> and Sabrina Eimler<sup>7</sup>

Datenbank Spektrum  
<https://doi.org/10.1007/s13222-023-00436-3>

## SCHWERPUNKTBEITRAG

## Moving Beyond Benchmarks and Competitions: Towards Addressing Social Media Challenges in an Educational Context

Dimitri Ognibene<sup>1</sup>  · Gregor Donabauer<sup>1,2</sup> · Emily Theophilou<sup>3</sup> · Sathya Buršić<sup>1</sup> · Francesco Lomonaco<sup>1</sup> · Rodrigo Wilkens<sup>4</sup> · Davinia Hernández-Leo<sup>3</sup> · Udo Kruschwitz<sup>2</sup>

Received: 17 October 2022 / Accepted: 25 January 2023  
© The Author(s) 2023



## Some References to COURAGE II

- Aprin, F. et al. (2021). Is This Fake or Credible? A Virtual Learning Companion Supporting the Judgment of Young Learners Facing Social Media Content. In: International Conference on Web-Based Learning. Springer.
- Hernández-Leo, D. et al. (2021). Narrative scripts embedded in social media towards empowering digital and self-protection skills. In European Conference on Technology Enhanced Learning. Springer.
- Lobo Quintero et al. (2022). Analyzing the intrinsic motivation in narrative scripts to enhance social media awareness. International Conference on Higher Education Learning Methodologies and Technologies Online.

# References to Nudging / Boosting



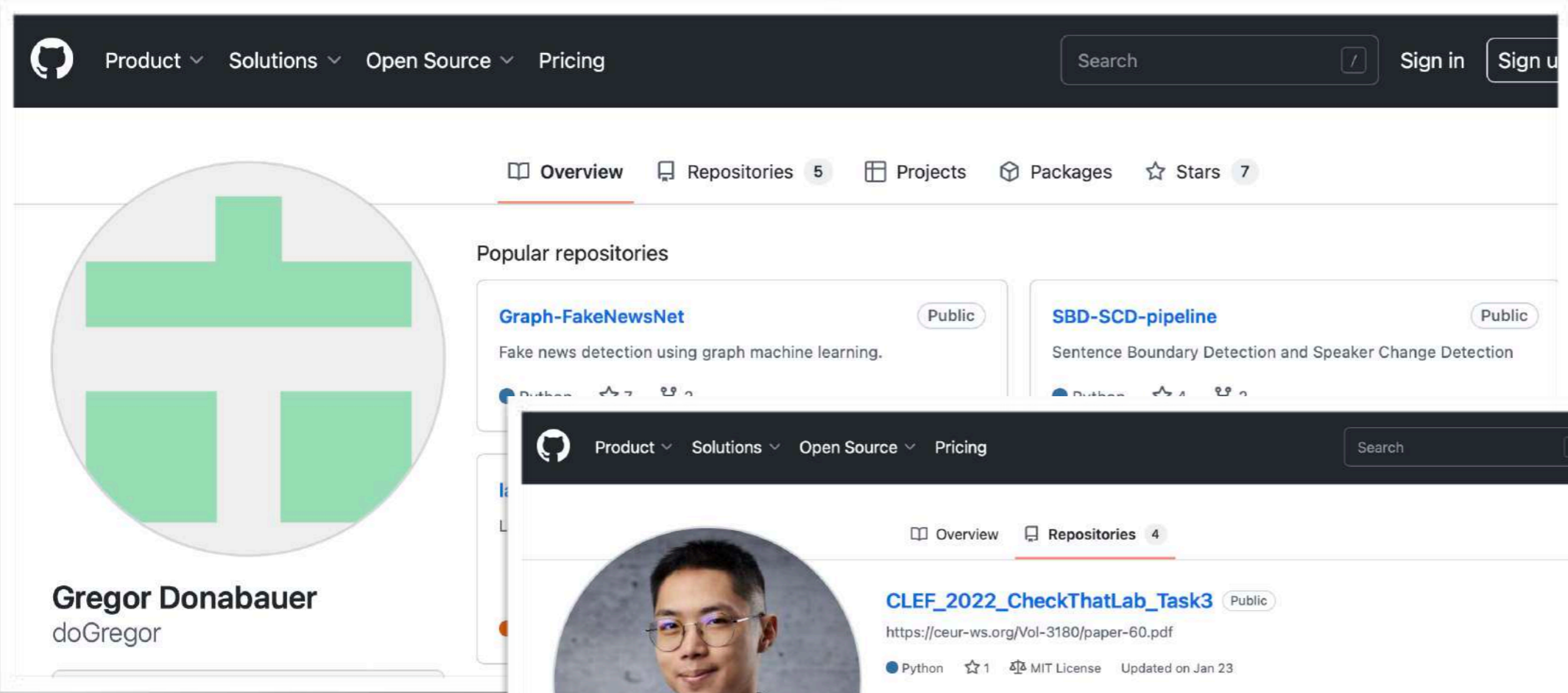
Universität Regensburg

- Hertwig, R., T. Grüne-Yanoff (2017). *Nudging and Boosting: Steering or Empowering Good Decisions*. *Perspectives on Psychological Science*, 12(6), 973–986.
- Kozyreva, A., S. Lewandowsky, R. Hertwig (2019). *Citizens Versus the Internet: Confronting Digital Challenges With Cognitive Tools*.
- Lorenz-Spreen, P., Lewandowsky, S., Sunstein, C. R., Hertwig, R. (2020). How behavioural sciences can promote truth, autonomy and democratic discourse online. *Nature Human Behaviour*, 1-8.
- Lorenz-Spreen, P., Geers, M., Pachur, T., Hertwig, R., Lewandowsky, S., Herzog, S. M. (2021). Boosting people's ability to detect microtargeted advertising. *Scientific Reports*, 11(1), 1-9.

# Applying Nudging / Boosting to Search

- Zimmerman, S., A. Thorpe, J. Chamberlain and U. Kruschwitz (2020) *Towards Search Strategies for Better Privacy and Information*, CHIIR 2020.
- Zimmerman, S., A. Thorpe, C. Fox, U. Kruschwitz (2019) *Privacy Nudging in Search: Investigating Potential Impacts*, CHIIR 2019.
- Zimmerman, S., A. Thorpe, C. Fox, U. Kruschwitz (2020) *Investigating the Interplay between Searchers' Privacy Concerns and their Search Behaviour*, SIGIR 2019.
- Zimmerman, S., S. Herzog, J. Chamberlain, D. Elweiler, U. Kruschwitz (2020). Towards a Framework for Harm Prevention in Web Search. BIRDS Workshop @ SIGIR.

# Sample Resources



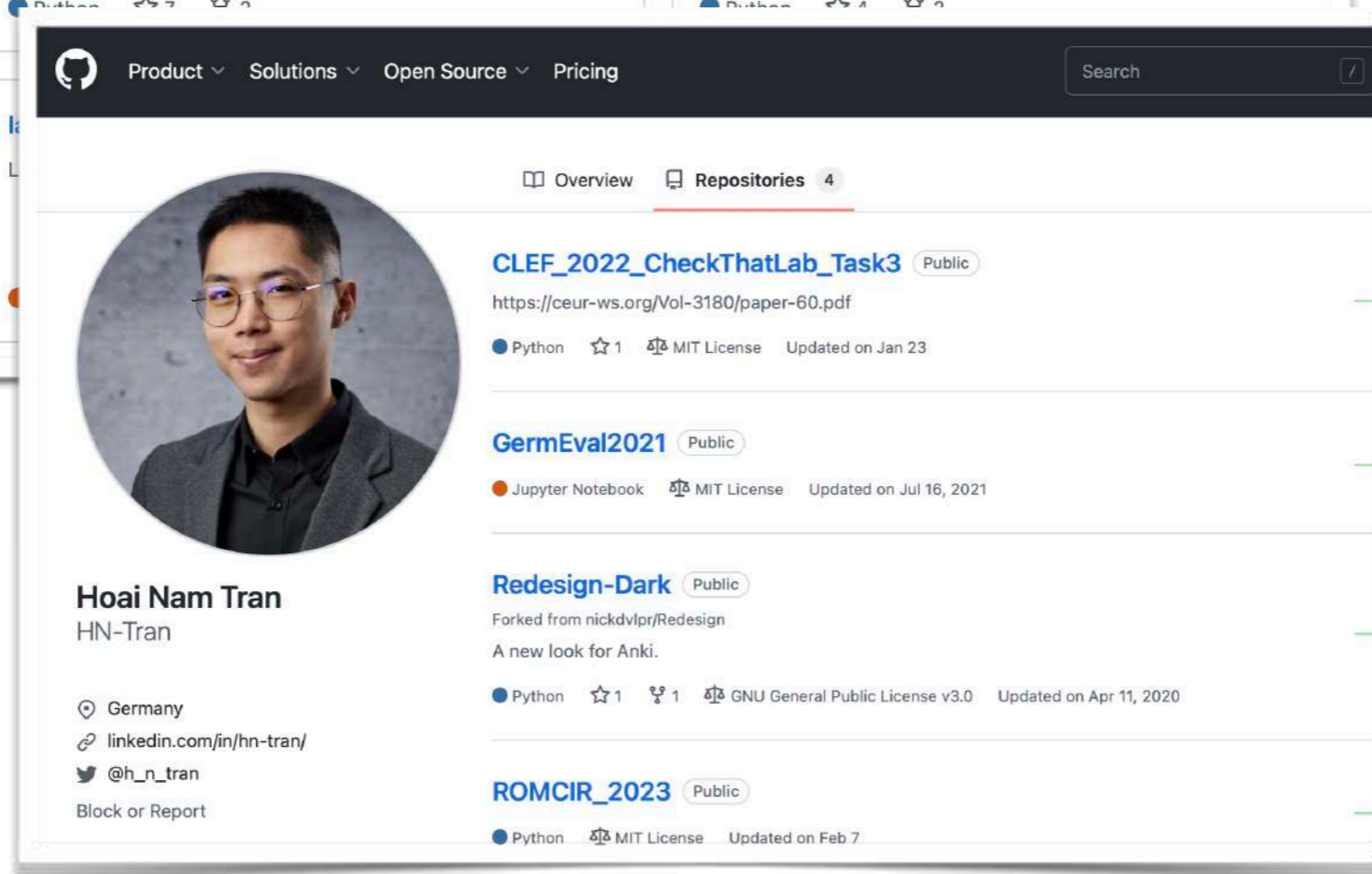
This screenshot shows the GitHub profile of Gregor Donabauer. The profile header includes navigation links for Product, Solutions, Open Source, and Pricing, along with a search bar and sign-in options. The main content area displays 'Popular repositories' with two items: 'Graph-FakeNewsNet' (Public) and 'SBD-SCD-pipeline' (Public). A large green cross icon is overlaid on the left side of the profile.

**Gregor Donabauer**  
doGregor

Overview | Repositories 5 | Projects | Packages | Stars 7

**Graph-FakeNewsNet** (Public)  
Fake news detection using graph machine learning.

**SBD-SCD-pipeline** (Public)  
Sentence Boundary Detection and Speaker Change Detection



This screenshot shows the GitHub profile of Hoai Nam Tran. The profile header includes navigation links for Product, Solutions, Open Source, and Pricing, along with a search bar and sign-in options. The main content area displays 'Repositories' with four items: 'CLEF\_2022\_CheckThatLab\_Task3' (Public), 'GermEval2021' (Public), 'Redesign-Dark' (Public), and 'ROMCIR\_2023' (Public). A circular profile picture of Hoai Nam Tran is visible on the left.

**Hoai Nam Tran**  
HN-Tran

Germany  
linkedin.com/in/hn-tran/  
@h\_n\_tran  
Block or Report

Overview | Repositories 4

**CLEF\_2022\_CheckThatLab\_Task3** (Public)  
https://ceur-ws.org/Vol-3180/paper-60.pdf  
Python 1 star MIT License Updated on Jan 23

**GermEval2021** (Public)  
Jupyter Notebook MIT License Updated on Jul 16, 2021

**Redesign-Dark** (Public)  
Forked from nickdvpr/Redesign  
A new look for Anki.  
Python 1 star 1 fork GNU General Public License v3.0 Updated on Apr 11, 2020

**ROMCIR\_2023** (Public)  
Python MIT License Updated on Feb 7

# Acknowledgements

Thanks to ... Dimitri Ognibene, Francesco Lomonaco, Sathya Buršić (Milano-Bicocca); Sabrina Eimler, Veronica Schwarze, Johanna Schäwel (HRW); Davide Taibi, Lidia Scifo (CNR); Davinia Hernández-Leo, Emily Theophilou, Rene Alejandro Lobo (UPF); Ulrich Hoppe, Farbod Aprin, Nils Malzahn (RIAS); Gregor Donabauer, Hoai-Nam Tran (Regensburg) ... as well as all other contributors to the COURAGE project (including many slides used here), also thanks to Steve Zimmerman (Essex) (e.g., slides on boosting and nudging)

# Acknowledgements

This work is part of the project *COURAGE — A social media companion safeguarding and educating students* and funded by the Volkswagen Foundation in the topic Artificial Intelligence and the Society of the Future (Nos. 95563, 95564, 95565, 95566, 95567, and 9B145).

Thanks ...







# Egypt to fine creators of weather misinformation

 Egypt Independent · April 1, 2023



 **Harrie Oosterhuis** @HarrieOos · 23h



If you're coming to Dublin for @ecir2023 this to week:

We are super lucky with the weather! ☀️

Don't forget to bring sunscreen. 😎🧴

#ECIR2023

🔍  H

	Today	Tomorrow	10 days
Today Mostly sunny			32° 21°
Sunday, 2 Apr Partly cloudy			32° 19°
Monday, 3 Apr Mostly cloudy			33° 19°
Tuesday, 4 Apr Mostly sunny			34° 21°
Wednesday, 5 Apr Partly cloudy			34° 22°
Thursday, 6 Apr			34°