DOMINOES
digital resilience to disinformation

4.1

Strategic, anticipatory and current analysis of disinformation and information-led hostile influencing

Rubén Arcos | URJC

doi.org/10.5281/zenodo.10064592

Co-funded by the European Union

2021-1-RO01-KA220-HED-000031158

ANIMV DARE. LEARN. INNOVATE.
Universidad Rey Juan Carlos
L-Università ta' Malta
NEW STRATEGY CENTER

DOMINOES
digital resilience to disinformation

Erasmus+

Rubén ARCOS | URJC

STRATEGIC, ANTICIPATORY AND CURRENT ANALYSIS OF DISINFORMATION AND INFORMATION-LED HOSTILE INFLUENCING

In this section 4.1. We are going to work on the following contents:

- Strategic analysis of disinformation and information-led hostile influencing
- Anticipatory analysis of disinformation and information-led hostile influencing
- Current analysis of disinformation and information-led hostile influencing

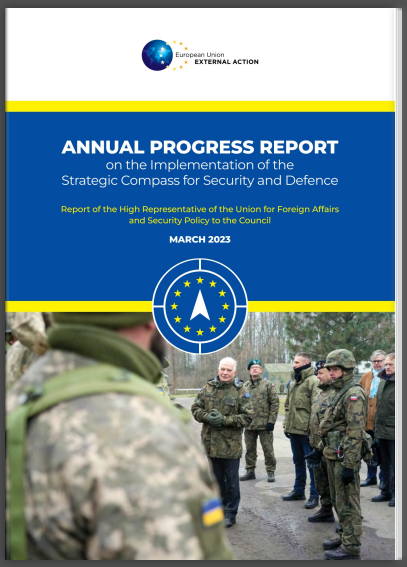
Rubén ARCOS | URJC

UNIT OBJECTIVES

- To provide an understanding of the different analytic approaches to disinformation
- To know existing frameworks and approaches employed by practitioners and academic experts when dealing with disinformation and manipulative content by foreign actors
- To understand the difference between current, strategic and anticipatory analysis of disinformation
- To know existing tools for the analysis of disinformation

Strategic analysis of disinformation and information-led hostile influencing

The March 2023 EU Annual Progress Report on the Implementation of the Strategic Compass for Security and Defence has highlighted the importance of situational awareness and analysis on disinformation and FIMI.



1/24

DOMINOES
digital resilience to disinformation

Strategic analysis of disinformation and information-led hostile influencing

"**Foreign Information Manipulation and Interference (FIMI)** is increasingly used as part of broader hybrid campaigns.

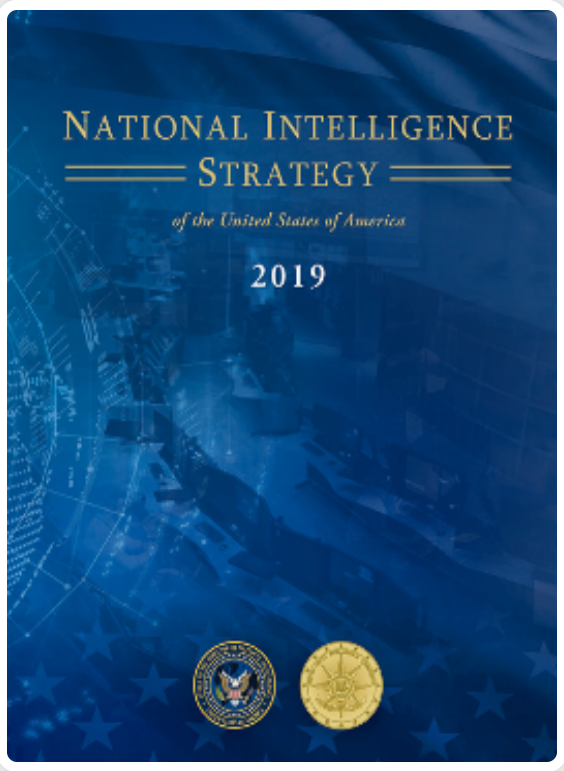
To better understand these threats, we are enhancing our situational awareness and analysis capabilities and have published a first report on these threats.

We are working with international partners, including the G7 and NATO, as well as stakeholders from civil society and private sector on establishing a new central FIMI data space for gathering information on threats stemming from disinformation and foreign information manipulation.

This will promote the sharing of information and analysis between all stakeholders about root causes, incidents and threats." (p.11)

Strategic analysis of disinformation and information-led hostile influencing

From an intelligence perspective, **strategic analysis** identifies and assesses "the capabilities, activities, and intentions of states and non-state entities to develop a deep understanding of the strategic environment, warn of future developments on issues of enduring interest" and supports policies and strategic decisions (ODNI 2019, p. 8)



INTEL - The NIS @ a Glance

Joomla! - the dynamic portal engine and content management...

Intelligence

Strategic analysis of disinformation and information-led hostile influencing

In the context of FIMI threat analysis, the **European External Action Service** has described its current analytical framework that consist of two elements:

1. An analysis cycle for strategically analyzing incidents
2. The DISARM framework

(See: Strategic Communications, Task Forces and Information Analysis 2023)

EEAS' FIMI Threats Analysis Cycle/Workflow

1. **Strategic monitoring:** Mapping the ecosystem of assets used by threat actors for manipulation and interference, and systematically monitoring them
2. **Prioritisation & Triage:** Filtering and classification of activities for prioritization
3. **Incident analysis and evidence collection:** gathering evidence though open-source collection and conducting analyses
4. **Knowledge pooling and sharing** (with key stakeholders and partners)
5. **Situational awareness:** provision of situational awareness and understanding, and feedback loop to inform further analytic efforts.

DISARM Framework

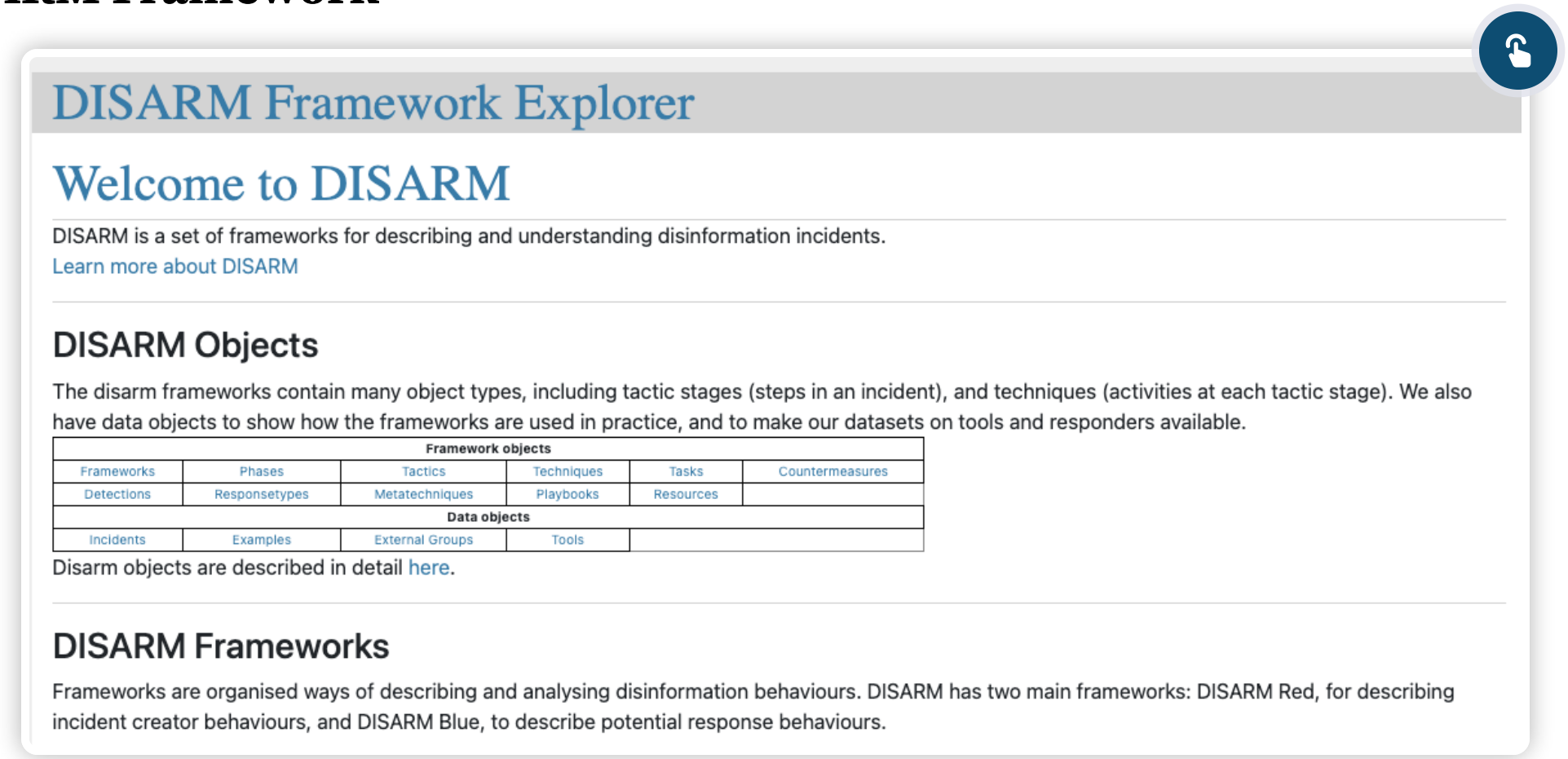
Regarding the DISARM framework, the DISARM foundation explains it as:

> "the open-source, master framework for fighting disinformation through sharing data & analysis, and coordinating effective action. The Framework has been developed, drawing on global cybersecurity best practices. It is used to help communicators, from whichever discipline or sector, to gain a clear shared understanding of disinformation incidents and to immediately identify defensive and mitigation actions that are available to them"
>
> (https://www.disarm.foundation/framework)

DISARM Frameworks "are organised ways of describing and analysing disinformation behaviours. DISARM has two main frameworks: DISARM Red, for describing incident creator behaviours, and DISARM Blue, to describe potential response behaviours." (https://disarmframework.herokuapp.com)

(https://disarmframework.herokuapp.com)

DISARM Framework



DISARM Framework Explorer

Welcome to DISARM

DISARM is a set of frameworks for describing and understanding disinformation incidents.
Learn more about DISARM

DISARM Objects

The disarm frameworks contain many object types, including tactic stages (steps in an incident), and techniques (activities at each tactic stage). We also have data objects to show how the frameworks are used in practice, and to make our datasets on tools and responders available.

Framework objects					
Frameworks	Phases	Tactics	Techniques	Tasks	Countermeasures
Detections	Responsetypes	Metatechniques	Playbooks	Resources	
Data objects					
Incidents	Examples	External Groups	Tools		

Disarm objects are described in detail here.

DISARM Frameworks

Frameworks are organised ways of describing and analysing disinformation behaviours. DISARM has two main frameworks: DISARM Red, for describing incident creator behaviours, and DISARM Blue, to describe potential response behaviours.

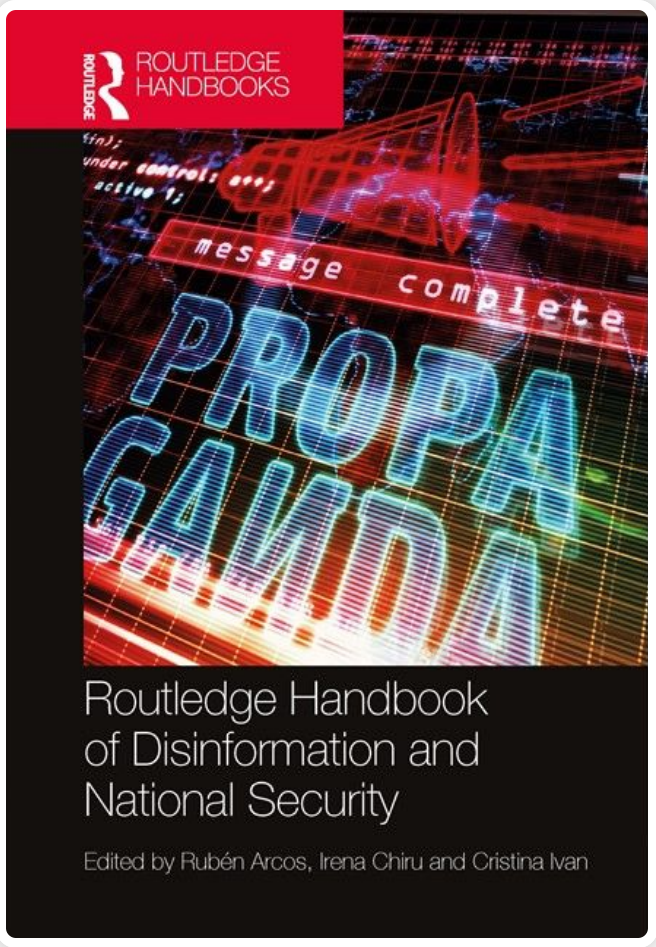
Anticipatory analysis of disinformation and information-led hostile influencing

- **Countering disinformation** is usually done reactively by conducting fact-checking and debunking manipulative content that has already been spread on social media platforms, private messaging apps and other channels.

- However, **proactive communication and anticipation of emerging issues** have always been the hallmark of successful strategic communication by industry practitioners.

  The **advantages** of an anticipatory approach are clear:

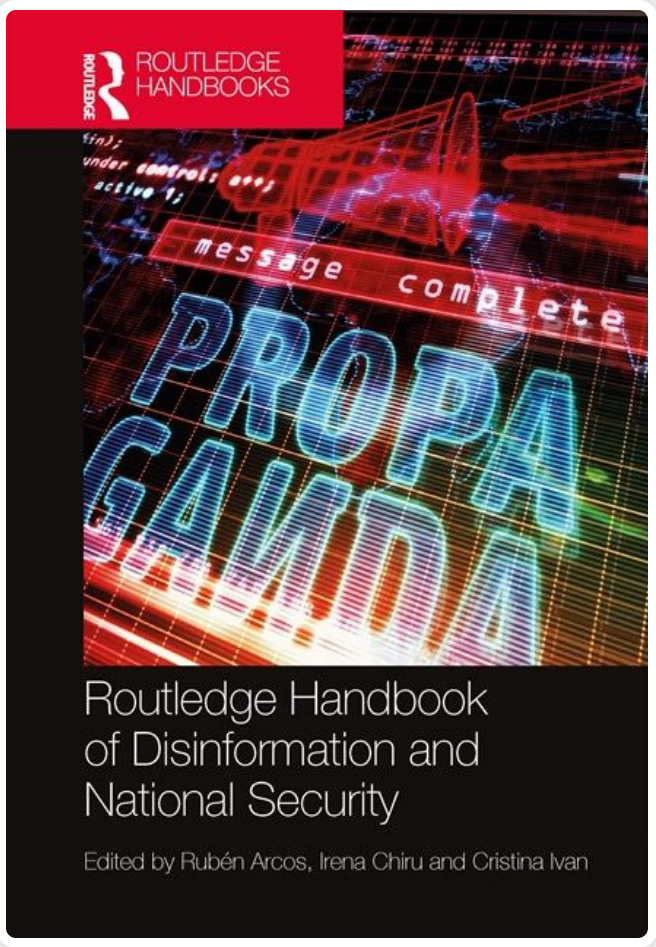
  - Companies can lead the discussion rather than react to information and opinions spread by others;
  - They can focus the discussion on specific public issues;
  - The organization's stance on key issues doesn't have to be articulated in response to claims, narratives and messages from others.



Anticipatory analysis of disinformation and information-led hostile influencing

- **Issues** are defined as “any development –usually in the public arena – which, if it continues, could have a SIGNIFICANT impact on the operation or future interests of the organization”
  (Jaques 2014, p. 323)

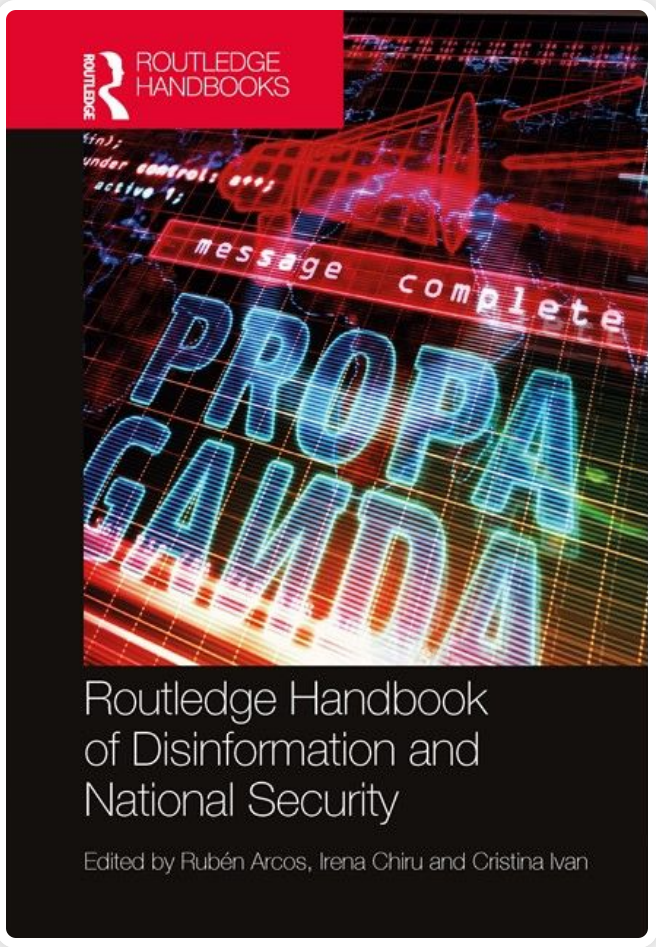
- When **hostile influence attempts target public opinion** abroad and shape part of the public debate on important issues, such as, for instance, climate change or public health, the need for early management of information manipulation on these issues becomes obvious for democracies



DOMINOES
digital resilience to disinformation

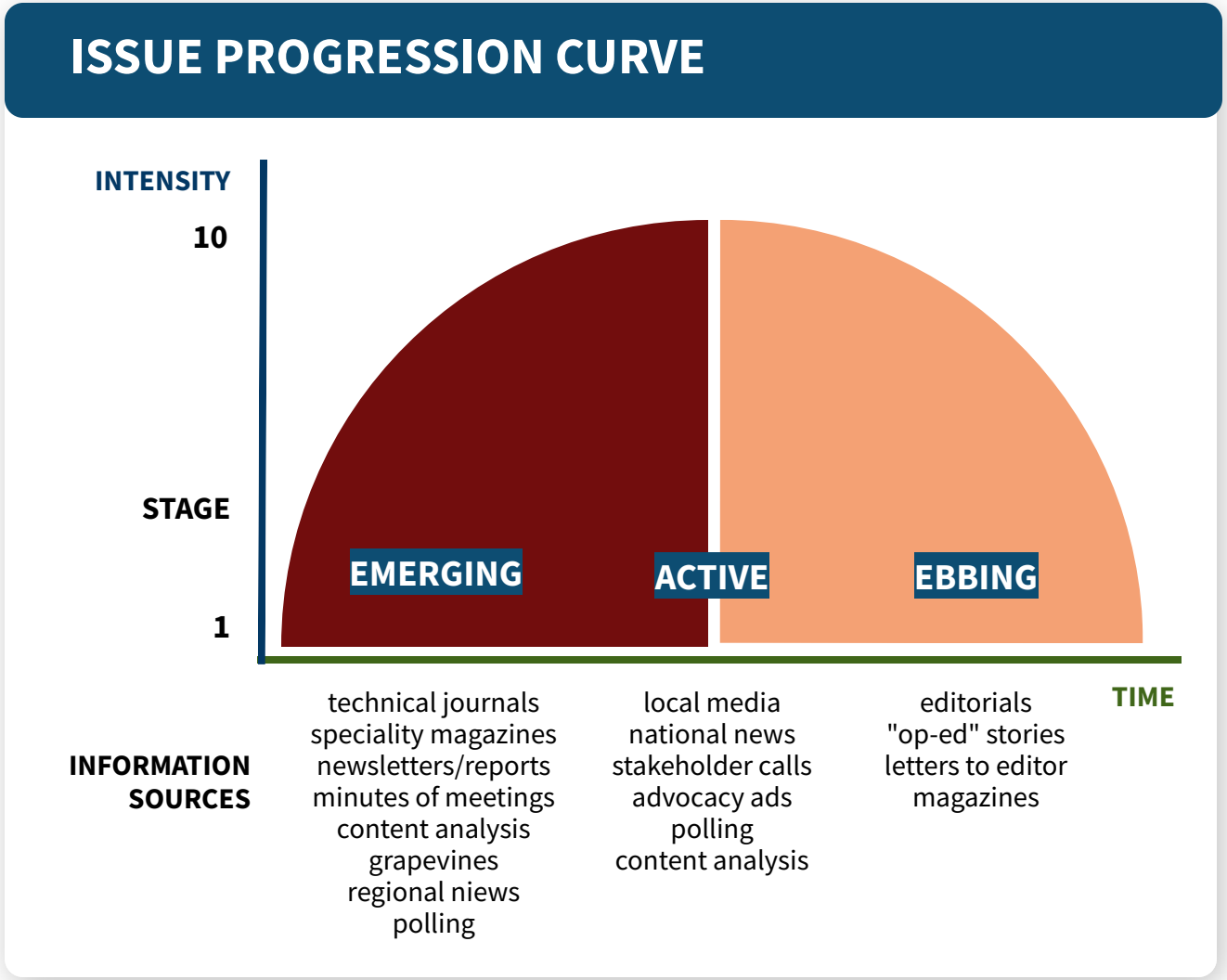
Anticipatory analysis of disinformation and information-led hostile influencing

- Crable and Vibbert introduced the idea that issues develop according to a certain pattern, a life cycle.
- In a simplified version of the life cycle of issues, three phases are distinguished: emerging, active, and ebbing.
- Depending on the stage in this progression curve, issues can be identified through different sources of information. When an issue is active, it receives more media attention, whereas in the emerging stage it has limited public attention.
- In the case of disinformation, the initial efforts of hostile actors may appear in the form of conspiracy theories in fringe media, internet forums and blogs, and then be revisited, adapted to the national or local context and amplified in mainstream social platforms.



DOMINOES
digital resilience to disinformation

Anticipatory analysis of disinformation and information-led hostile influencing



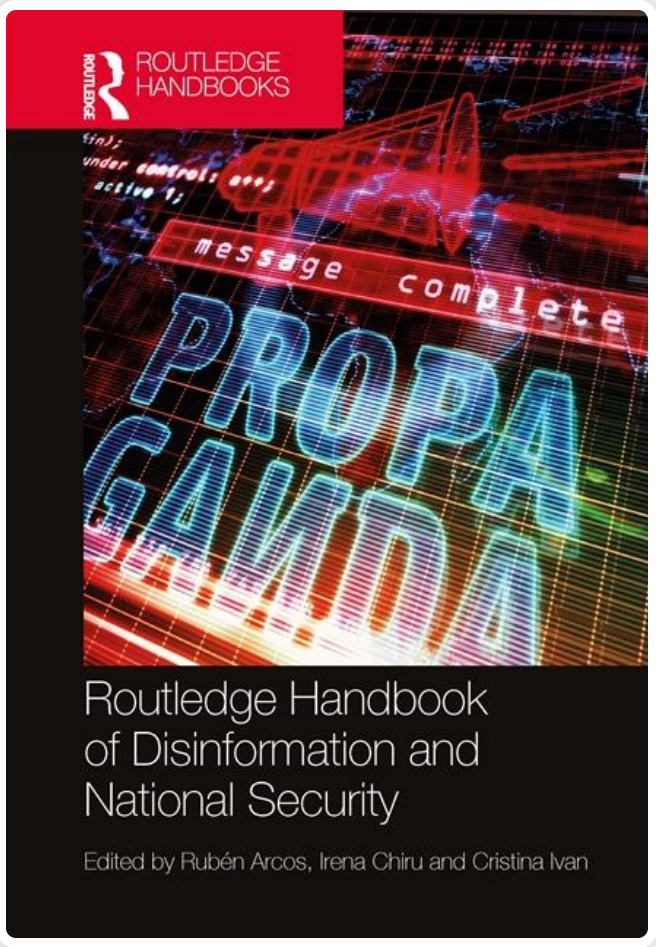
Source: Eli Sopow 1994

DOMINOES
digital resilience to disinformation

Anticipatory analysis of disinformation and information-led hostile influencing

- The anticipatory approach to countering disinformation aims to **create situational awareness and provide analyses and assessments** to facilitate decisions and actions by authorities and practitioners, including through preventive and positive communication strategies.

- Such an approach can **help in identifying risks and issues** that could be exploited by hostile actors at an early stage, plan strategic communication actions and avoid potential crises caused by disinformation and information manipulation.



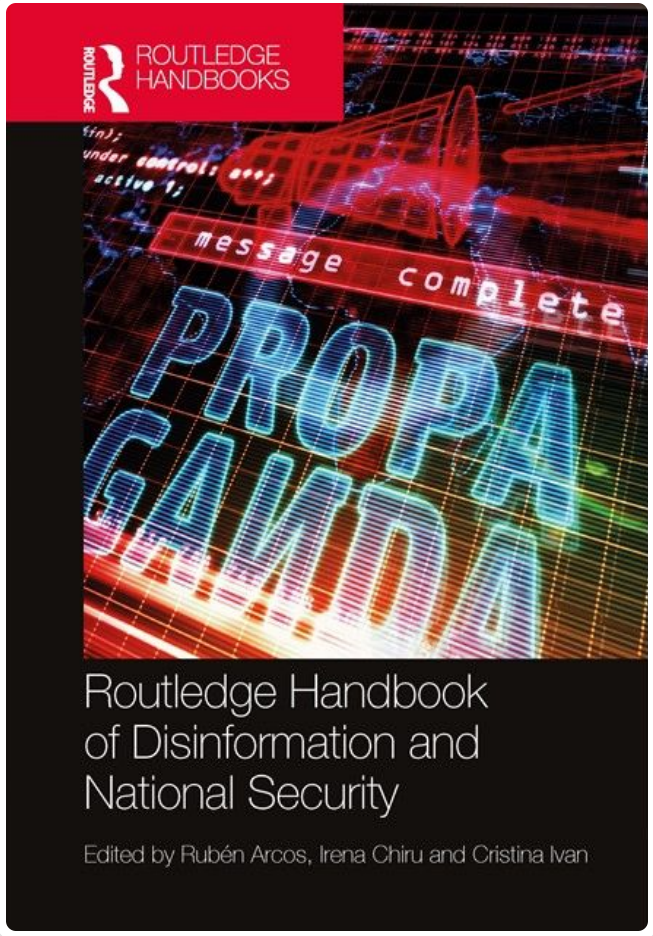
DOMINOES
digital resilience to disinformation

Anticipatory analysis of disinformation and information-led hostile influencing

- **Policymakers and practitioners can address FIMI** and coordinated disinformation activities with an anticipatory approach by assessing how likely it's that a hostile actor will exploit an existing vulnerability in a target democratic society and by what means.

- This means "anticipating the risk of an attack in the information environment by assessing the hostile entities' capabilities, intentions, and activities, and the target vulnerabilities".

- **Vulnerabilities** can arise from existing socio-political and historical conflicts, but other developments such as pandemics or economic crises can also present opportunities. Foreign state and non-state actors may activate latent issues and spread disinformation and hostile narratives.



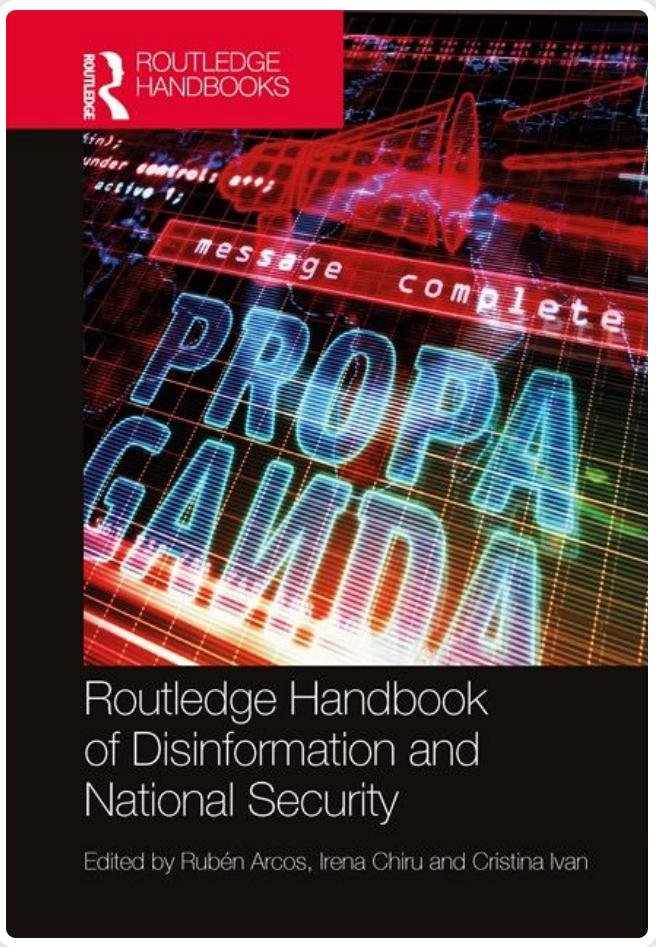
DOMINOES
digital resilience to disinformation

Anticipatory analysis of disinformation and information-led hostile influencing

- The development of **disinformation scenarios and associated indicators** (capability indicators, intention indicators, and vulnerability indicators) can help to identify impending threats in the information environment at an early stage.

  A systematic approach can lead to a display and warning function against disinformation.

- Let's look at some **examples of indicators** from the following plausible but fictitious scenario, so that you better understand what indicators are (what you can "theoretically" expect if such a scenario occurs) and what indications are: the actual changes in these indicators.



DOMINOES
digital resilience to disinformation

Erasmus+

Example: Scenario

On the way to an important parliamentary debate in country X and other European countries, where an **important decision** is to be made **that affects an authoritarian competitor state** Y, **this state interferes in the process** through a coordinated campaign in the information environment.

DOMINOES
digital resilience to disinformation

Erasmus+

Example: Vulnerability indicator

Indicator	Indication
Existing political divisions between mainstream political parties in the parliament of country X.	In recent months, the existing political polarisation in country X has increased from moderate to severe.

DOMINOES
digital resilience to disinformation

Erasmus+

Example: Capability indicator

Indicator	Indication
State-funded media of country Y broadcasting in the language of country X (different from the state of origin) and shares in private media companies in country X.	Country Y launched a news and entertainment channel in the language of country X last year.

DOMINOES
digital resilience to disinformation

Erasmus+

Example: Intentions indicator

Indicator	Indication
Public statements by leading politicians of country Y on the issue.	No public statement has been made on this political issue in recent weeks

Current analysis of disinformation and information-led hostile influencing

Current analysis of dis- and misinformation can be defined as the **descriptive analysis of information on both erroneous or manipulative content gathered on a daily basis that can potentially impact the understanding, attitudes and behaviors of specific audiences towards events, developments, persons, institutions and other objects**.

Existing analytic frameworks in the context of intelligence and security (See: Pherson and Pherson 2013; Margolis 2020), have identified different types of analyses:

- Epistemic analysis
- Descriptive analysis
- Evaluative analysis
- Estimative analysis
- Exploratory analysis

DOMINOES
digital resilience to disinformation

Current analysis of disinformation and information-led hostile influencing

Table 3: The Typology in Substantive Context

Source: Margolis 2020, pg. 5

	Foundational	Current		Anticipatory	
	Epistemic	Descriptive	Evaluative	Estimative	Exploratory
E.g., Nuclear weapons in China (1960s)[11]	What is the organization of China's nuclear weapons research effort? (1, 4) What is the size and makeup of China's nuclear arsenal? (1, 4)	What device did Beijing test yesterday? (2, 4) How did the region respond? (2, 4)	How capable is the new weapon design? (2, 4) Where is Beijing's nuclear weapons program going? (3, 5)	How would Beijing respond to a strike on its nuclear program? (3, 5) What arms control schemes would interest Beijing, if any? (3, 5)	How might the proliferation of this technology affect security dynamics elsewhere? (3, 6) What is the future of deterrence in East Asia? (3, 6)
E.g., Military reforms in China (1980s)[12]	Who leads China's military? (1, 4) What is the organization of the military after reforms? (1, 4)	What changes did Beijing just announce? (2, 4) How did the first post-reform exercise go? (2, 4)	What patterns are emerging in the reform effort? (2, 5) Why is Beijing reforming its military? (2, 5)	What are the prospects of the reform effort? (3, 5) How will the Soviet Union and Vietnam respond? (3, 5)	What is the future conventional military balance between Beijing and Moscow? (3, 6) How might China's civil-military relations evolve? (3, 6)
E.g., Handover of Hong Kong (1980s-90s)[13]	What are the provisions of the Basic Law? (1, 4) What international businesses operate in Hong Kong? (1, 4)	How did the region respond to the Joint Declaration? (2, 4) How are citizens reacting to accounts of Tiananmen? (2, 4)	What are Beijing's plans for Hong Kong? (2, 4) What are Beijing's redlines? (3, 5)	How stable will the transition be? (3, 5) What could trigger the flight of international businesses—and how would it unfold? (3, 5)	What is the future of "one country, two systems?" (3, 6) How might China's posture toward the West change? (3, 6)
Traditional Framework Types 1=Basic 2=Current 3=Estimative			Contemporary Framework Types 4=Current operational 5=Strategic 6=Anticipatory		

Current analysis of disinformation and information-led hostile influencing

In the theoretical framework develop by Margolis, current analysis includes **descriptive and evaluative analysis.**

Descriptive analysis enable situational awareness by disseminating analytic products such as summaries and updates; “they stay close to the information base and do not set a broader, interpretive analytic line” (Margolis 2020: 4).

In the context of disinformation, It can be said, that descriptive analysis is expected to bring the news about disinformation activities and ongoing developments. Descriptive analysis of disinformation will respond to questions such as What manipulative contents were disseminated yesterday by the state-sponsored media of an authoritarian state X?

On the other hand, **evaluative analysis**, similarly to interpretive journalism, “provides commentary, interpreting the news” thus enabling **critical reflection** and strategic awareness (Ibid.) Evaluative analysis will respond to questions such as **What the current increase of manipulative content from country X on issue Y is telling as about the policy and plans of Country X?**

Current analysis of disinformation and information-led hostile influencing

A **modified version of Harold Lasswell's communication model** of "Who says what, in which channel, to whom, with what effect?" provides an analytic framework for the current analysis of disinformation.

We will be doing descriptive analysis by answering the questions: **Who? What? When? Where? How?**

While our analysis will be more explanatory and evaluative in nature when responding to questions such as **What does it mean? and why?** (see: Pherson and Pherson 2013: 48)

Current analysis of disinformation and information-led hostile influencing

Who says what?

On behalf of whom?

(or on its own behalf)? With what intentions?

In what situations?

Producing what kind of effects?

With what assets?

Using what strategies?

(key messages and channels)

To which audiences?

(See Arcos 2018: 5)

Current analysis of disinformation and information-led hostile influencing

- On the other hand, in order to analyze disinformation content we can use existing tools for mapping the underlying argumentative structures and supporting elements of manipulative claims and narratives of hostile information-led influencing activities.

(See for example: https://www.rationaleonline.com)

DOMINOES
digital resilience to disinformation

Bibliography

and useful resources

2021-1-RO01-KA220-HED-000031158

DOMINOES
digital resilience to disinformation

References

- **Arcos, Rubén** (2018). Hybrid CoE Strategic Analysis 12: Post-event analysis of the hybrid threat security environment: assessment of influence communication operations. The European Centre of Excellence for Countering Hybrid Threats.
- **Arcos, Rubén** (2023). “Intelligence and awareness”, In Routledge Handbook of the Future of Warfare, edited by Artur Gruszczak and Sebastian Kaempf (Routledge): 272-283. https://doi.org/10.4324/9781003299011-29
- **Arcos, Rubén and Arribas, Cristina** (2023). “Anticipatory Approaches to Disinformation, Warning and Supporting Technologies” In Routledge Handbook of Disinformation and National Security, edited by Rubén Arcos, Irena Chiru and Cristina Ivan (Routledge)
- **Crabble, Richard E., and Vibbert, Steven L.** (1985) “Managing Issues and Influencing Public Policy,” Public Relations Review 11, no. 2,
- **EEAS.** (2023) 2023 Annual Progress Report on the Implementation of the Strategic Compass for Security and Defence. Report of the High Representative of the Union for Foreign Affairs and Security Policy to the Council.
- **Jaques Tony** (2014) Issue and Crisis Management: Exploring Issues, Crises, Risk and Reputation. (Oxford University Press Australia & New Zealand. Kindle edition, 2014), 323.
- **Margolis, J. Eli.** (2020) “Rethinking Analytic Disciplines, Reordering the Profession.” Studies in Intelligence 64, no. 4 (2020): 1-14
- **Pherson, Katherine H. and Randolph H. Pherson.**(2021) Critical thinking for strategic intelligence. Thousand Oaks: SAGE
- **Sopow, Eli.** (1994) The Critical Issues Audit, (Leesburg, VA: Issue Action Publications)
- **Strategic Communications, Task Forces and Information Analysis (STRAT.2).** (2023) 1st EEAS Report on Foreign Information Manipulation and Interference Threats. Towards a framework for networked defence, February 2023.

DOMINOES
digital resilience to disinformation

Digital cOMpetences INformatiOn EcoSystem

ANIMV DARE. LEARN. INNOVATE.
Universidad Rey Juan Carlos
L-Università ta' Malta
NEW STRATEGY CENTER

Co-funded by the European Union

2021-1-RO01-KA220-HED-000031158

Audiovisual and multimedia production: CIBERIMAGINARIO

DOMINOES
digital resilience to disinformation

Strategic analysis of disinformation and information-led hostile influencing

4.1.1

doi.org/10.5281/zenodo.10064639

Co-funded by the European Union

2021-1-RO01-KA220-HED-000031158

ANIMV DARE. LEARN. INNOVATE.
Universidad Rey Juan Carlos
L-Università ta' Malta
NEW STRATEGY CENTER

DOMINOES
digital resilience to disinformation

Strategic analysis of disinformation and information-led hostile influencing

The March 2023 EU Annual Progress Report on the Implementation of the Strategic Compass for Security and Defence has highlighted the importance of situational awareness and analysis on disinformation and FIMI,

"**Foreign Information Manipulation and Interference (FIMI)** is increasingly used as part of broader hybrid campaigns.

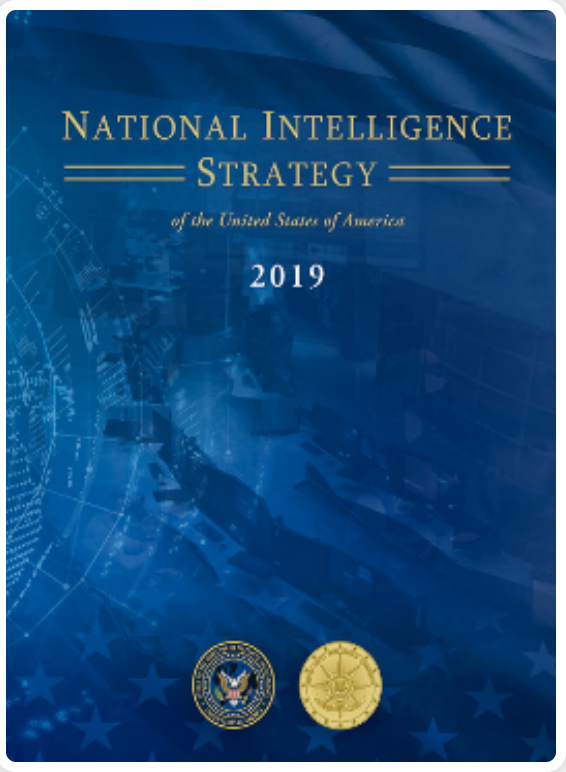
To better understand these threats, we are enhancing our situational awareness and analysis capabilities and have published a first report on these threats.

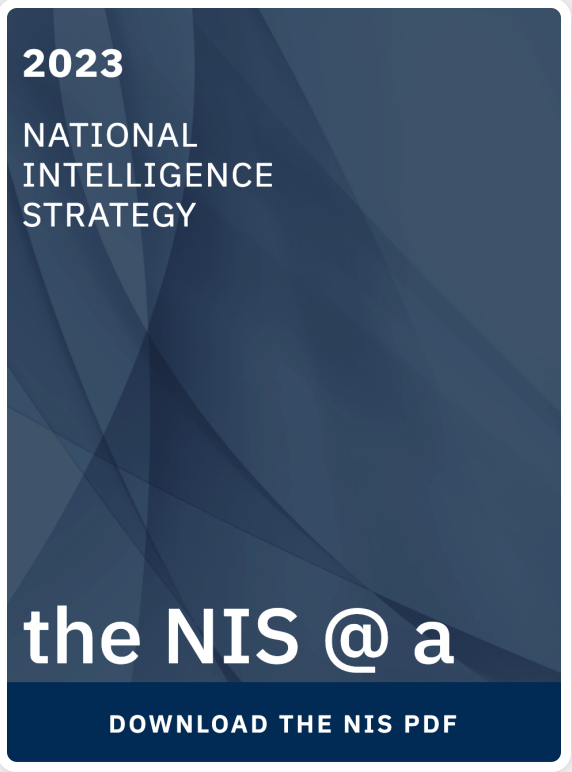
We are working with international partners, including the G7 and NATO, as well as stakeholders from civil society and private sector on establishing a new central FIMI data space for gathering information on threats stemming from disinformation and foreign information manipulation.

This will promote the sharing of information and analysis between all stakeholders about root causes, incidents and threats." (p.11)

Strategic analysis of disinformation and information-led hostile influencing

From an intelligence perspective, **strategic analysis** identifies and assesses "the capabilities, activities, and intentions of states and non-state entities to develop a deep understanding of the strategic environment, warn of future developments on issues of enduring interest" and supports policies and strategic decisions (ODNI 2019, p. 8)





DOMINOES
digital resilience to disinformation

Strategic analysis of disinformation and information-led hostile influencing

In the context of FIMI threat analysis, the **European External Action Service** has described its current analytical framework that consist of two elements:

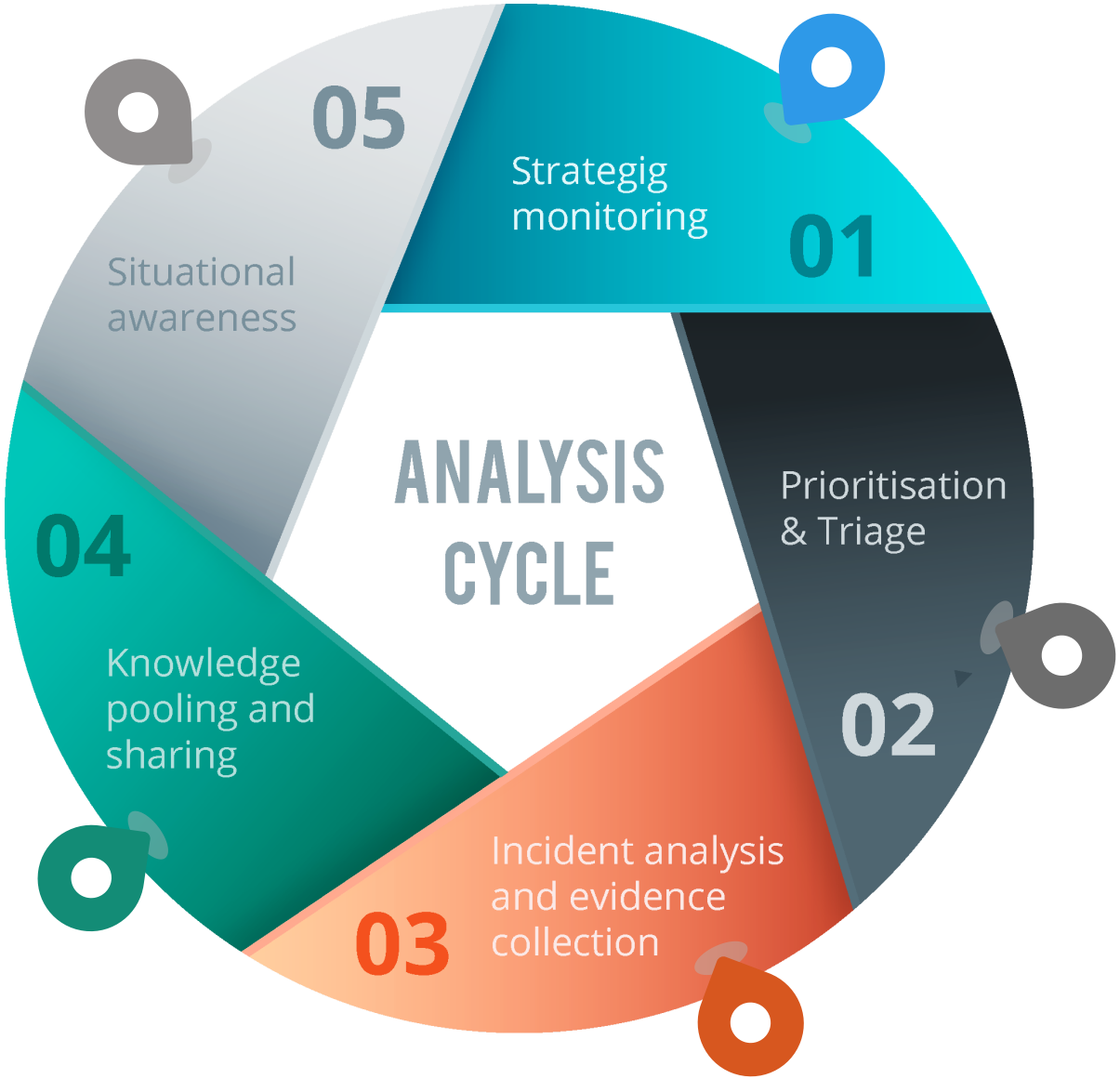
1. An analysis cycle for strategically analyzing incidents
2. The DISARM framework

(See: Strategic Communications, Task Forces and Information Analysis 2023)

DOMINOES
digital resilience to disinformation

EEAS current analytical framework for FIMI and DISARM Framework

EEAS current analytical framework for FIMI



DISARM Framework

Regarding the DISARM framework, the DISARM foundation explains it as:

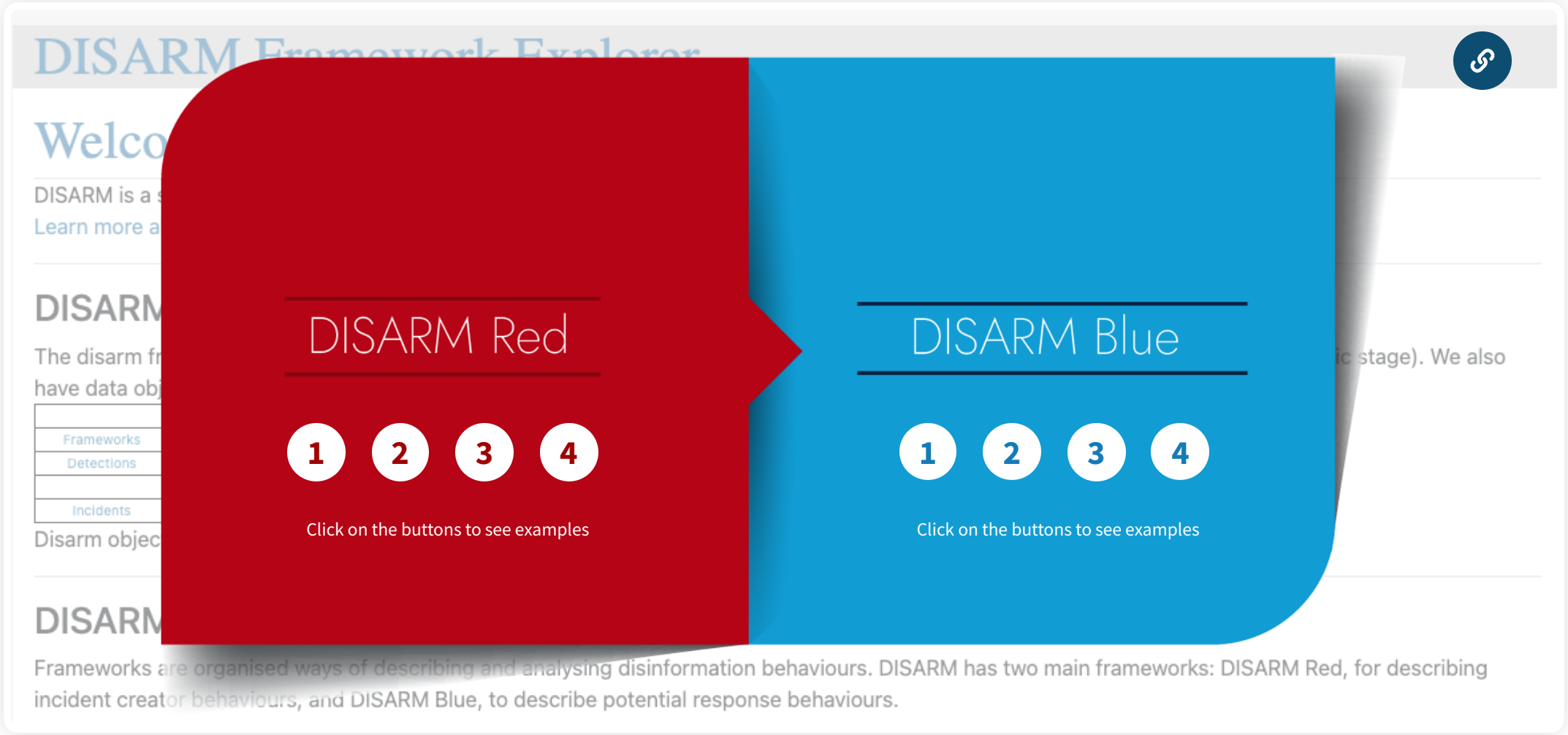
> "the open-source, master framework for fighting disinformation through sharing data & analysis, and coordinating effective action. The Framework has been developed, drawing on global cybersecurity best practices. It is used to help communicators, from whichever discipline or sector, to gain a clear shared understanding of disinformation incidents and to immediately identify defensive and mitigation actions that are available to them"
>
> (https://www.disarm.foundation/framework)

DISARM Frameworks "are organised ways of describing and analysing disinformation behaviours. DISARM has two main frameworks: DISARM Red, for describing incident creator behaviours, and DISARM Blue, to describe potential response behaviours." (https://disarmframework.herokuapp.com)

(https://disarmframework.herokuapp.com)

DOMINOES
digital resilience to disinformation

Erasmus+



C00159 Have a disinformation response plan

- e.g. Create a campaign plan and toolkit for competition short of armed conflict (this used to be called “the grey zone”).
- The campaign plan should account for own vulnerabilities and strengths, and not over-rely on any one tool of statecraft or line of effort.
- It will identify and employ a broad spectrum of national power to deter, compete, and counter (where necessary) other countries’ approaches, and will include understanding of own capabilities, capabilities of disinformation creators, and international standards of conduct to compete in, shrink the size, and ultimately deter use of competition short of armed conflict.

C0002

Innoculate. Positive campaign to promote feeling of safety

- Used to counter ability based and fear based attacks

T0088 Develop Audio-based Content

- Creating and editing false or misleading audio artifacts, often aligned with one or more specific narratives, for use in a disinformation campaign.
- This may include creating completely new audio content, repurposing existing audio artifacts (including cheap fakes), or using AI-generated audio creation and editing technologies (including deepfakes).

T0019 Generate information pollution

- Flood social channels; drive traffic/engagement to all assets; create aura/sense/perception of pervasiveness/consensus (for or against or both simultaneously) of an issue or topic.
- "Nothing is true, but everything is possible."
- Akin to astroturfing campaign.

T0086.001 Develop Memes

- Memes are one of the most important single artefact types in all of computational propaganda.
- Memes in this framework denotes the narrow image-based definition. But that naming is no accident, as these items have most of the important properties of Dawkins' original conception as a self-replicating unit of culture.
- Memes pull together reference and commentary; image and narrative; emotion and message.
- Memes are a powerful tool and the heart of modern influence campaigns.

T0019.001 Create fake research

- Create fake academic research.
- Example: fake social science research is often aimed at hot-button social issues such as gender, race and sexuality.
- Fake science research can target Climate Science debate or pseudoscience like anti-vaxx

C00019 Reduce effect of division-enablers

Includes

- Promote constructive communication by shaming division-enablers
- Promote playbooks to call out division-enablers

00073

Inoculate populations through media literacy training

- Use training to build the resilience of at-risk populations.
- Educate on how to handle info pollution.
- Push out targeted education on why it's pollution.
- Build cultural resistance to false content, e.g. cultural resistance to bullshit. Influence literacy training, to inoculate against “cult” recruiting.
- Media literacy training: leverage librarians / library for media literacy training. Inoculate at language.
- Strategic planning included as inoculating population has strategic value.
- Concepts of media literacy to a mass audience that authorities launch a public information campaign that teaches the program will take time to develop and establish impact, recommends curriculum-based training. Covers detect, deny, and degrade.

References

1. Arcos, Rubén (2023). “Intelligence and awareness”, In Routledge Handbook of the Future of Warfare, edited by Artur Gruszczak and Sebastian Kaempf (Routledge): 272-283. https://doi.org/10.4324/9781003299011-29

2. European Union External Action (2023). Annual Progress Report on the Implementation of the Strategic Compass for Security and Defence, Report of the High Representative of the Union for Foreign Affairs and Security Policy to the Council, March 2023. https://www.eeas.europa.eu/sites/default/files/documents/2023/StrategicCompass_1stYear_Report.pdf

3. ODNI (2019). The National Intelligence Strategy of the United States of America 2019, https://www.dni.gov/files/ODNI/documents/National_Intelligence_Strategy_2019.pdf

4. Strategic Communications, Task Forces and Information Analysis (2023). 1st EEAS Report on Foreign Information Manipulation and Interference Threats. Towards a framework for networked defence February 2023. https://www.eeas.europa.eu/sites/default/files/documents/2023/EEAS-DataTeam-ThreatReport-2023..pdf

DOMINOES
digital resilience to disinformation

Digital cOMpetences INformatiOn EcoSystem

ANIMV DARE. LEARN. INNOVATE.
Universidad Rey Juan Carlos
L-Università ta' Malta
NEW STRATEGY CENTER

Co-funded by the European Union
2021-1-RO01-KA220-HED-000031158

Author of contents: **Rubén Arcos (URJC)**
Audiovisual and multimedia production: CIBERIMAGINARIO

DOMINOES
digital resilience to disinformation

Anticipatory analysis of disinformation and information-led hostile influencing

4.1.2

doi.org/10.5281/zenodo.10064648

Co-funded by the European Union

2021-1-RO01-KA220-HED-000031158

ANIMV DARE. LEARN. INNOVATE.
Universidad Rey Juan Carlos
L-Università ta' Malta
NEW STRATEGY CENTER

(Arcos y Arribas 2023: 401-416)

Digital cOMpetences INformatiOn EcoSystem

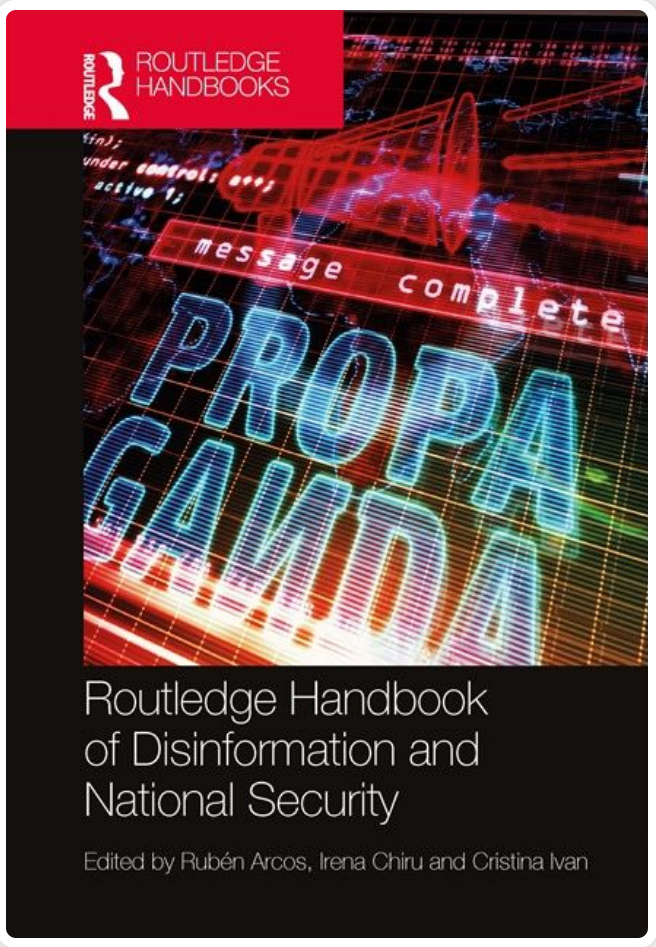
Anticipatory analysis of disinformation and information-led hostile influencing

- **Countering disinformation** is usually done reactively by conducting fact-checking and debunking manipulative content that has already been spread on social media platforms, private messaging apps and other channels.

- However, **proactive communication and anticipation of emerging issues** have always been the hallmark of successful strategic communication by industry practitioners.

  The **advantages** of an anticipatory approach are clear:

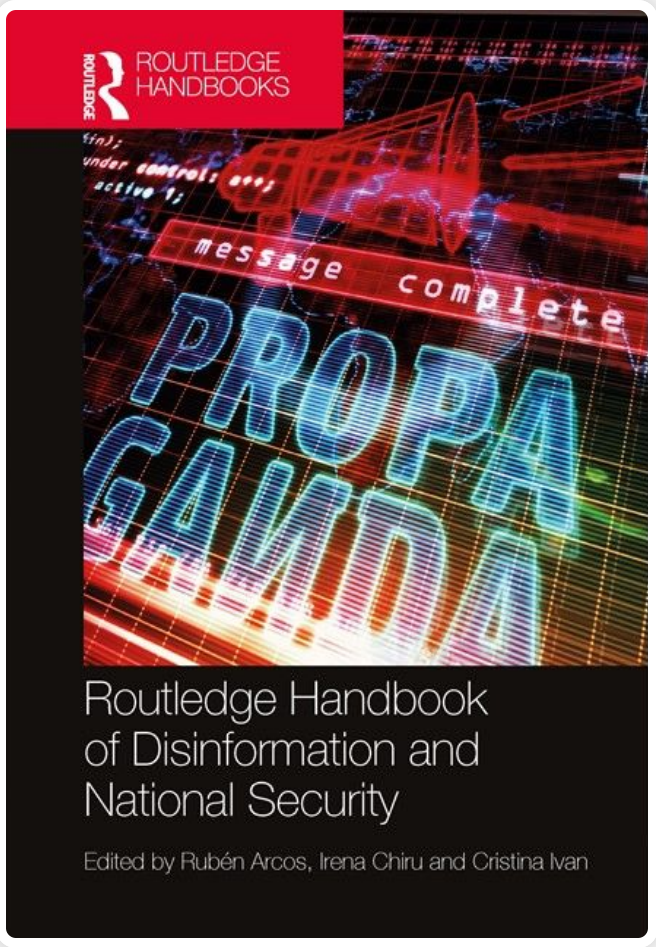
  - Companies can lead the discussion rather than react to information and opinions spread by others;
  - They can focus the discussion on specific public issues;
  - The organization's stance on key issues doesn't have to be articulated in response to claims, narratives and messages from others.



Anticipatory analysis of disinformation and information-led hostile influencing

- **Issues** are defined as “any development –usually in the public arena – which, if it continues, could have a SIGNIFICANT impact on the operation or future interests of the organization”
  (Jaques 2014, p. 323)

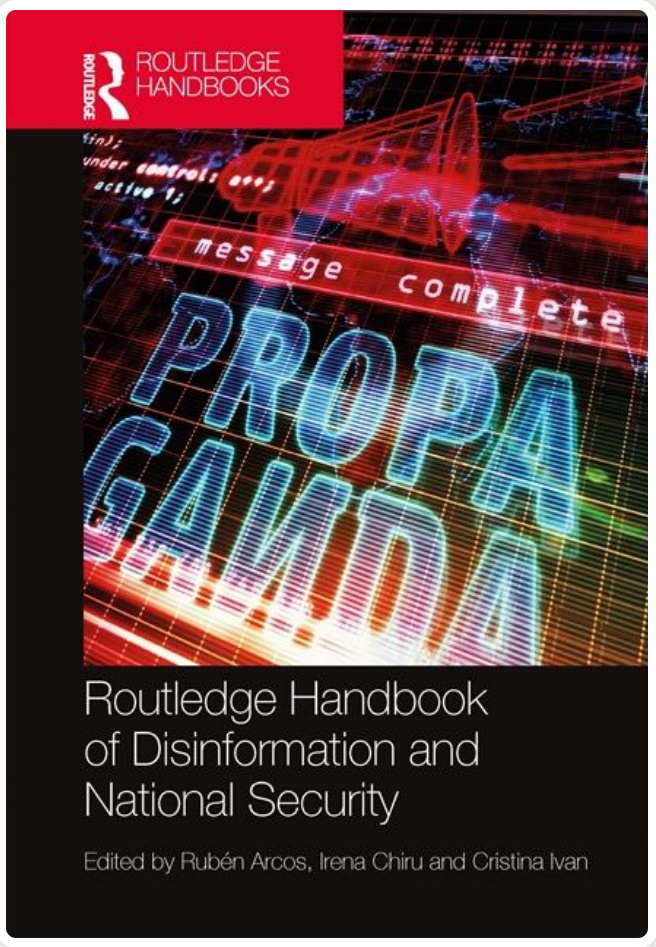
- When **hostile influence attempts target public opinion** abroad and shape part of the public debate on important issues, such as, for instance, climate change or public health, the need for early management of information manipulation on these issues becomes obvious for democracies



DOMINOES
digital resilience to disinformation

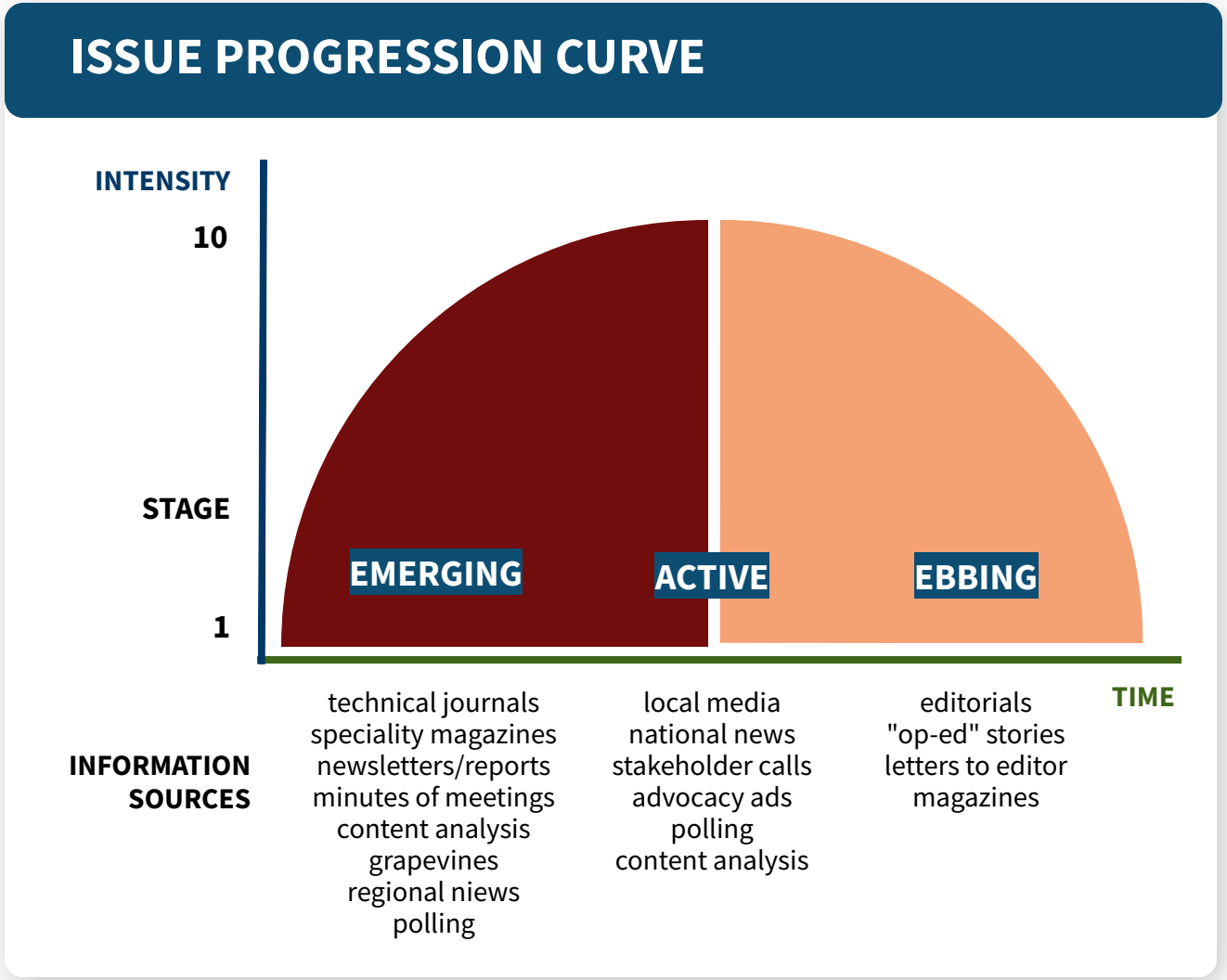
Anticipatory analysis of disinformation and information-led hostile influencing

- Crable and Vibbert introduced the idea that issues develop according to a certain pattern, a life cycle.
- In a simplified version of the life cycle of issues, three phases are distinguished: emerging, active, and ebbing.
- Depending on the stage in this progression curve, issues can be identified through different sources of information. When an issue is active, it receives more media attention, whereas in the emerging stage it has limited public attention.
- In the case of disinformation, the initial efforts of hostile actors may appear in the form of conspiracy theories in fringe media, internet forums and blogs, and then be revisited, adapted to the national or local context and amplified in mainstream social platforms.



DOMINOES
digital resilience to disinformation

Anticipatory analysis of disinformation and information-led hostile influencing



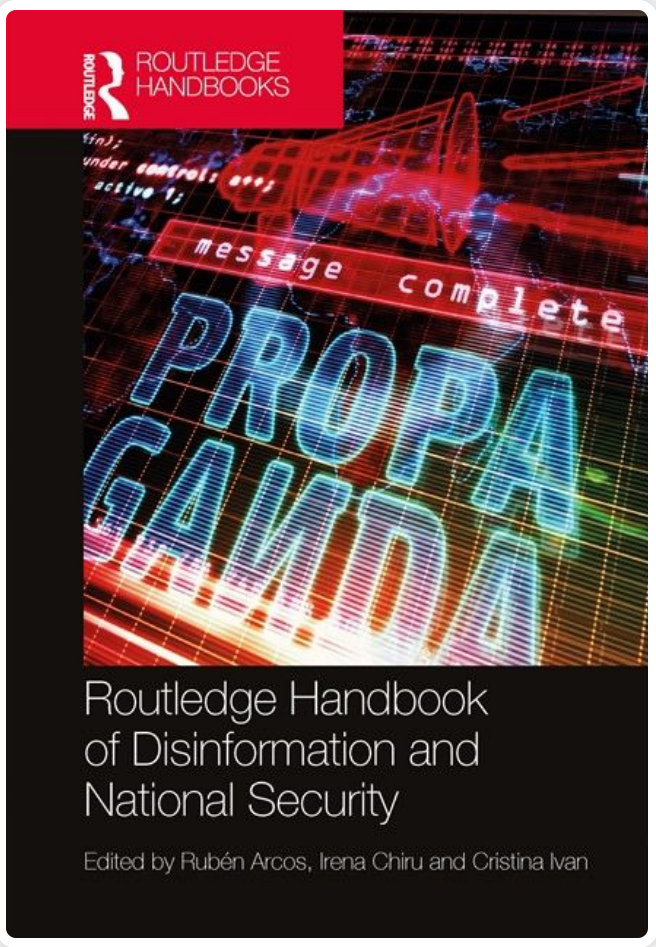
Source: Eli Sopow 1994

DOMINOES
digital resilience to disinformation

Anticipatory analysis of disinformation and information-led hostile influencing

- The anticipatory approach to countering disinformation aims to **create situational awareness and provide analyses and assessments** to facilitate decisions and actions by authorities and practitioners, including through preventive and positive communication strategies.

- Such an approach can **help in identifying risks and issues** that could be exploited by hostile actors at an early stage, plan strategic communication actions and avoid potential crises caused by disinformation and information manipulation.



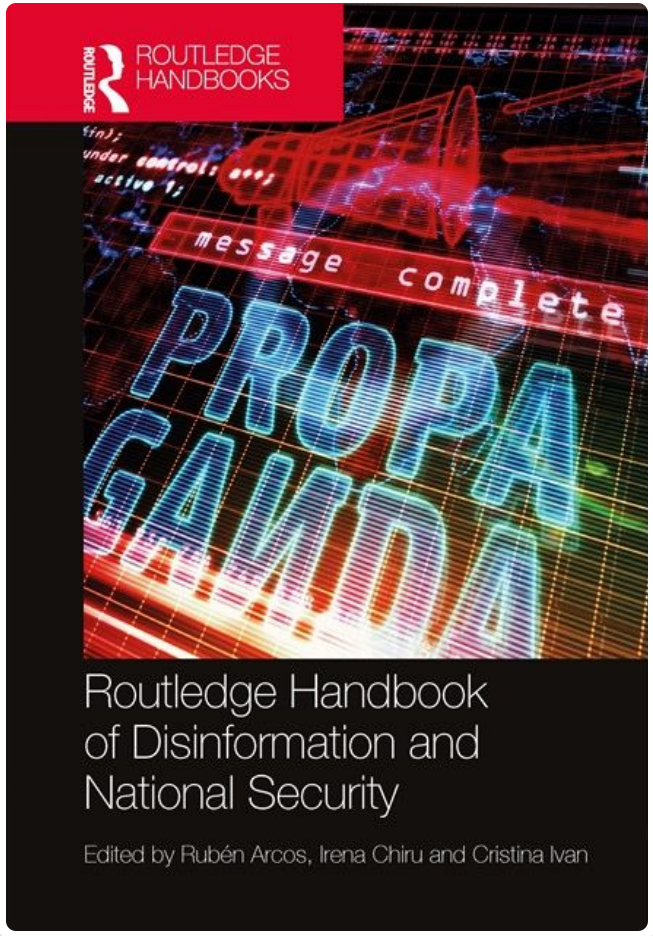
DOMINOES
digital resilience to disinformation

Anticipatory analysis of disinformation and information-led hostile influencing

- **Policymakers and practitioners can address FIMI** and coordinated disinformation activities with an anticipatory approach by assessing how likely it's that a hostile actor will exploit an existing vulnerability in a target democratic society and by what means.

- This means "anticipating the risk of an attack in the information environment by assessing the hostile entities' capabilities, intentions, and activities, and the target vulnerabilities".

- **Vulnerabilities** can arise from existing socio-political and historical conflicts, but other developments such as pandemics or economic crises can also present opportunities. Foreign state and non-state actors may activate latent issues and spread disinformation and hostile narratives.



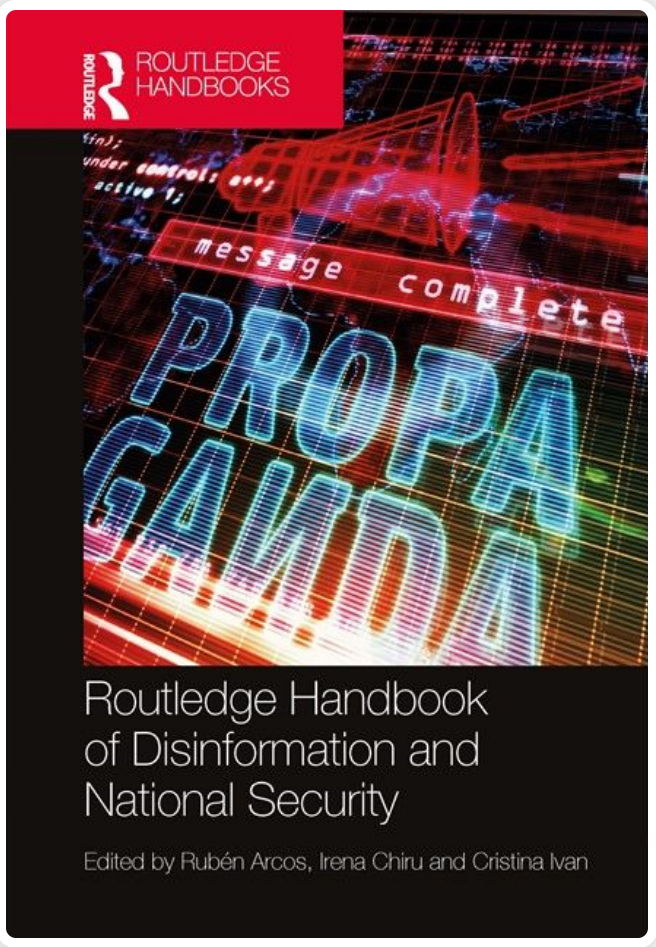
DOMINOES
digital resilience to disinformation

Anticipatory analysis of disinformation and information-led hostile influencing

- The development of **disinformation scenarios and associated indicators** (capability indicators, intention indicators, and vulnerability indicators) can help to identify impending threats in the information environment at an early stage.

  A systematic approach can lead to a display and warning function against disinformation.

- Let's look at some **examples of indicators** from the following plausible but fictitious scenario, so that you better understand what indicators are (what you can "theoretically" expect if such a scenario occurs) and what indications are: the actual changes in these indicators.



DOMINOES
digital resilience to disinformation

Erasmus+

Example: Scenario

On the way to an important parliamentary debate in country X and other European countries, where an **important decision** is to be made **that affects an authoritarian competitor state** Y, **this state interferes in the process** through a coordinated campaign in the information environment.

DOMINOES
digital resilience to disinformation

Erasmus+

Example: Vulnerability indicator

Indicator	Indication
Existing political divisions between mainstream political parties in the parliament of country X.	In recent months, the existing political polarisation in country X has increased from moderate to severe.

DOMINOES
digital resilience to disinformation

Erasmus+

Example: Capability indicator

Indicator	Indication
State-funded media of country Y broadcasting in the language of country X (different from the state of origin) and shares in private media companies in country X.	Country Y launched a news and entertainment channel in the language of country X last year.

DOMINOES
digital resilience to disinformation

Erasmus+

Example: Intentions indicator

Indicator	Indication
Public statements by leading politicians of country Y on the issue.	No public statement has been made on this political issue in recent weeks

DOMINOES
digital resilience to disinformation

Bibliography

and useful resources

2021-1-RO01-KA220-HED-000031158

DOMINOES
digital resilience to disinformation

References

- **Arcos, Rubén (2023).** “Intelligence and awareness”, In Routledge Handbook of the Future of Warfare, edited by Artur Gruszczak and Sebastian Kaempf (Routledge): 272-283. https://doi.org/10.4324/9781003299011-29
- **Arcos, Rubén and Arribas, Cristina (2023).** “Anticipatory Approaches to Disinformation, Warning and Supporting Technologies” In Routledge Handbook of Disinformation and National Security, edited by Rubén Arcos, Irena Chiru and Cristina Ivan (pp. 401-416) https://doi.org/10.4324/9781003190363-34
- **Crabble, Richard E., and Vibbert, Steven L. (1985)** “Managing Issues and Influencing Public Policy,” Public Relations Review 11, no. 2
- **Jaques,Tony (2014)** Issue and Crisis Management: Exploring Issues, Crises, Risk and Reputation. Oxford University Press Australia & New Zealand. Kindle edition, 323.
- **Sopow, Eli. (1994)** The Critical Issues Audit, Leesburg, VA: Issue Action Publications.

DOMINOES
digital resilience to disinformation

Digital cOMpetences INformatiOn EcoSystem

ANIMV DARE. LEARN. INNOVATE.
Universidad Rey Juan Carlos
L-Università ta' Malta
NEW STRATEGY CENTER

Co-funded by the European Union
2021-1-RO01-KA220-HED-000031158

Author of contents: **Rubén Arcos (URJC)**
Audiovisual and multimedia production: CIBERIMAGINARIO

DOMINOES
digital resilience to disinformation

Current analysis of disinformation and information-led hostile influencing

4.1.4

doi.org/10.5281/zenodo.10064651

Co-funded by the European Union

2021-1-RO01-KA220-HED-000031158

ANIMV DARE. LEARN. INNOVATE.
Universidad Rey Juan Carlos
L-Università ta' Malta
NEW STRATEGY CENTER

Current analysis of disinformation and information-led hostile influencing

Current analysis of disinformation and information-led hostile influencing

Existing analytic frameworks in the context of intelligence and security (See: Pherson and Pherson 2013; Margolis 2020), have identified different types of analyses:

- Epistemic analysis
- Descriptive analysis
- Evaluative analysis
- Estimative analysis
- Exploratory analysis

Current analysis of disinformation and information-led hostile influencing

In the theoretical framework develop by Margolis, current analysis includes **descriptive and evaluative analysis.**

Descriptive analysis enable situational awareness by disseminating analytic products such as summaries and updates; "they stay close to the information base and do not set a broader, interpretive analytic line" (Margolis 2020: 4).

In the context of disinformation, it can be said, that **descriptive analysis is expected to bring the news about disinformation** activities and ongoing developments. Descriptive analysis of disinformation will respond to questions such as What manipulative contents were disseminated yesterday by the state-sponsored media of an authoritarian state X?

On the other hand, **evaluative analysis**, similarly to interpretive journalism, "provides commentary, interpreting the news" thus enabling critical refection and strategic awareness (Ibid.)
Evaluative analysis will respond to questions such as What the current increase of manipulative content from country X on issue Y is telling as about the policy and plans of Country X?

DOMINOES
digital resilience to disinformation

Current analysis of disinformation and information-led hostile influencing

Table 3: The Typology in Substantive Context

Source: Margolis 2020, pg. 5

	Foundational	Current		Anticipatory	
	Epistemic	Descriptive	Evaluative	Estimative	Exploratory
E.g., Nuclear weapons in China (1960s)[11]	What is the organization of China's nuclear weapons research effort? (1, 4) What is the size and makeup of China's nuclear arsenal? (1, 4)	What device did Beijing test yesterday? (2, 4) How did the region respond? (2, 4)	How capable is the new weapon design? (2, 4) Where is Beijing's nuclear weapons program going? (3, 5)	How would Beijing respond to a strike on its nuclear program? (3, 5) What arms control schemes would interest Beijing, if any? (3, 5)	How might the proliferation of this technology affect security dynamics elsewhere? (3, 6) What is the future of deterrence in East Asia? (3, 6)
E.g., Military reforms in China (1980s)[12]	Who leads China's military? (1, 4) What is the organization of the military after reforms? (1, 4)	What changes did Beijing just announce? (2, 4) How did the first post-reform exercise go? (2, 4)	What patterns are emerging in the reform effort? (2, 5) Why is Beijing reforming its military? (2, 5)	What are the prospects of the reform effort? (3, 5) How will the Soviet Union and Vietnam respond? (3, 5)	What is the future conventional military balance between Beijing and Moscow? (3, 6) How might China's civil-military relations evolve? (3, 6)
E.g., Handover of Hong Kong (1980s-90s)[13]	What are the provisions of the Basic Law? (1, 4) What international businesses operate in Hong Kong? (1, 4)	How did the region respond to the Joint Declaration? (2, 4) How are citizens reacting to accounts of Tiananmen? (2, 4)	What are Beijing's plans for Hong Kong? (2, 4) What are Beijing's redlines? (3, 5)	How stable will the transition be? (3, 5) What could trigger the flight of international businesses—and how would it unfold? (3, 5)	What is the future of "one country, two systems?" (3, 6) How might China's posture toward the West change? (3, 6)
Traditional Framework Types 1=Basic 2=Current 3=Estimative			Contemporary Framework Types 4=Current operational 5=Strategic 6=Anticipatory		

Current analysis of disinformation and information-led hostile influencing

A **modified version of Harold Lasswell's communication model** of "Who says what, in which channel, to whom, with what effect?" provides an analytic framework for the current analysis of disinformation.

We will be doing descriptive analysis by answering the questions: **Who? What? When? Where? How?**

While our analysis will be more explanatory and evaluative in nature when responding to questions such as **What does it mean? and why?** (see: Pherson and Pherson 2013: 48)

DOMINOES
digital resilience to disinformation

Erasmus+

Current analysis of disinformation and information-led hostile influencing

Who says what?

On behalf of whom?

(or on its own behalf)? With what intentions?

In what situations?

Producing what kind of effects?

With what assets?

Using what strategies?

(key messages and channels)

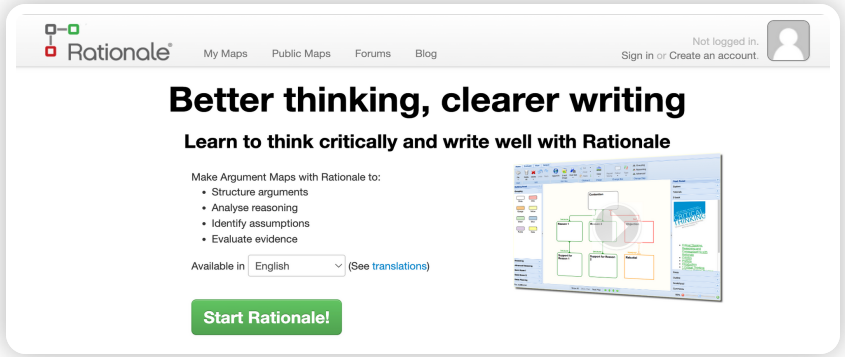
To which audiences?

(See Arcos 2018: 5)

DOMINOES
digital resilience to disinformation

Current analysis of disinformation and information-led hostile influencing

On the other hand, in order to analyze disinformation content we can use existing tools for mapping the underlying argumentative structures and supporting elements of manipulative claims and narratives of hostile information-led influencing activities.



(See for example: https://www.rationaleonline.com)

DOMINOES
digital resilience to disinformation

Digital cOMpetences INformatiOn EcoSystem

ANIMV DARE. LEARN. INNOVATE.
Universidad Rey Juan Carlos
L-Università ta' Malta
NEW STRATEGY CENTER

Co-funded by the European Union
2021-1-RO01-KA220-HED-000031158

Author of contents: **Rubén Arcos (URJC)**
Audiovisual and multimedia production: CIBERIMAGINARIO

DOMINOES
digital resilience to disinformation

4.2

Tech-driven solutions and emerging technologies to counter disinformation

Alexandra Anghel | ANIMV

doi.org/10.5281/zenodo.10064622

Co-funded by the European Union

2021-1-RO01-KA220-HED-000031158

ANIMV DARE. LEARN. INNOVATE.
Universidad Rey Juan Carlos
L-Università ta' Malta
NEW STRATEGY CENTER

DOMINOES
digital resilience to disinformation

Erasmus+

TECH-DRIVEN SOLUTIONS AND EMERGING TECHNOLOGIES TO COUNTER DISINFORMATION

The 4.2 module is dedicated to Tech-driven solutions and emerging technologies to counter disinformation.

Therefore, it will present the tricks and tips for delivering its educational content to any category of students, as well as the steps to be followed in order to use the support materials in a proper manner.

UNIT OBJETIVES

This particular module aims at presenting the main technological instruments and initiatives in the field of combating online disinformation, in addition to the previous sections that present the techniques employed by different disinformation processes. In the general architecture of the handbook, the module helps the target group to understand the phenomenon of disinformation in a comprehensive manner, by acknowledging which are the weapons one can use in order to avoid becoming a victim of false content, as well as developing their digital skills by exploring different technical solutions to combat online disinformation.

DOMINOES
digital resilience to disinformation

Combating the effects of disinformation in the online environment - setting the context

Technological developments and advances registered in the sector of the **Internet of Things** and social networks have created the premises for the expansion of the noxious effects of the disinformation phenomena, together with the rise of ubiquitous misinformation, disinformation, deep fakes, and post-truth.

Factors that influenced the rise of disinformation

- HYPER PARTISAN NEWS SITES
- TECHNOLOGICAL ADVANCEMENT
- POLITICIANS INCREASINGLY USING PROPAGANDA TERMS TO FRAME POLITICAL ISSUES

Technological developments and advances registered in the sector of Internet of Things and social networks have created the premises for the expansion of the noxious effects of the disinformation phenomena, together with the rise of **ubiquitous misinformation**, **disinformation**, **deep fakes**, and post-truth the rise of ubiquitous misinformation, disinformation, deep fakes, and post-truth.

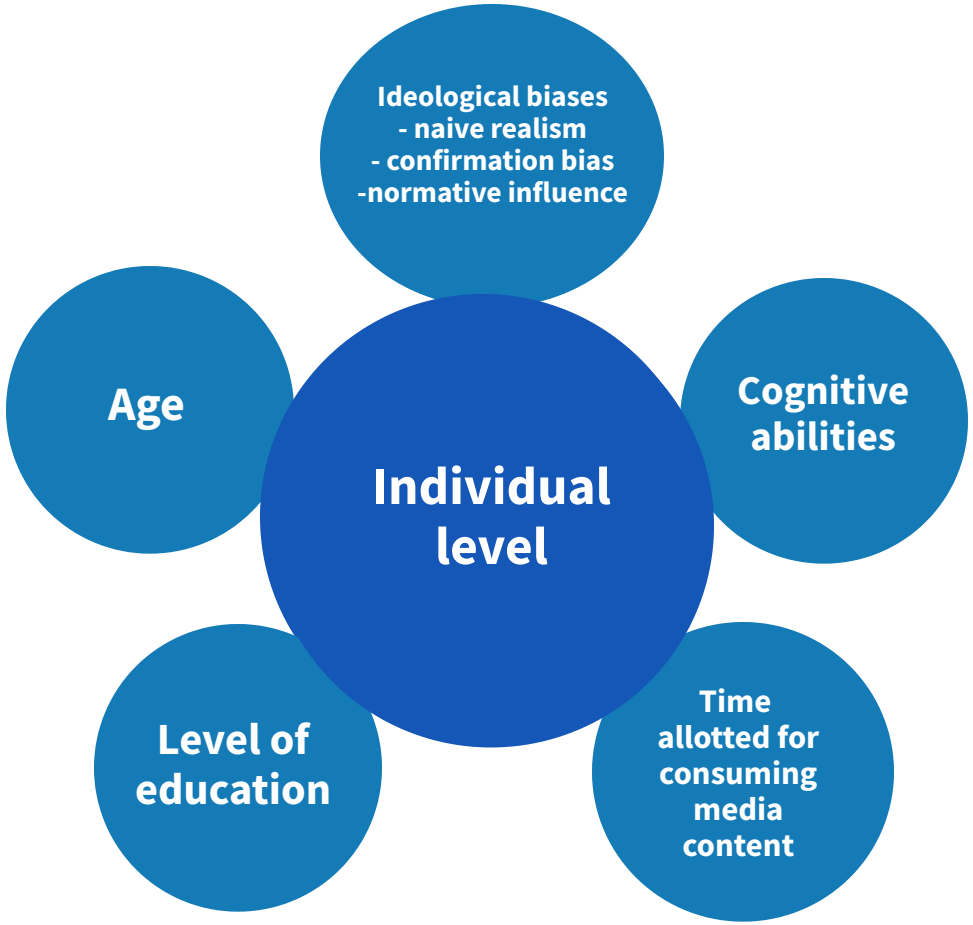
Technology has created the means for the expansion of the disinformation phenomenon, social media becoming one of the main sources of information for the population at large, as well as an important source of false content and digital deception. However, technology can also play an essential role in combating the effects of online disinformation and **propaganda** and in **containing** the expansion processes of these now defined security issues.

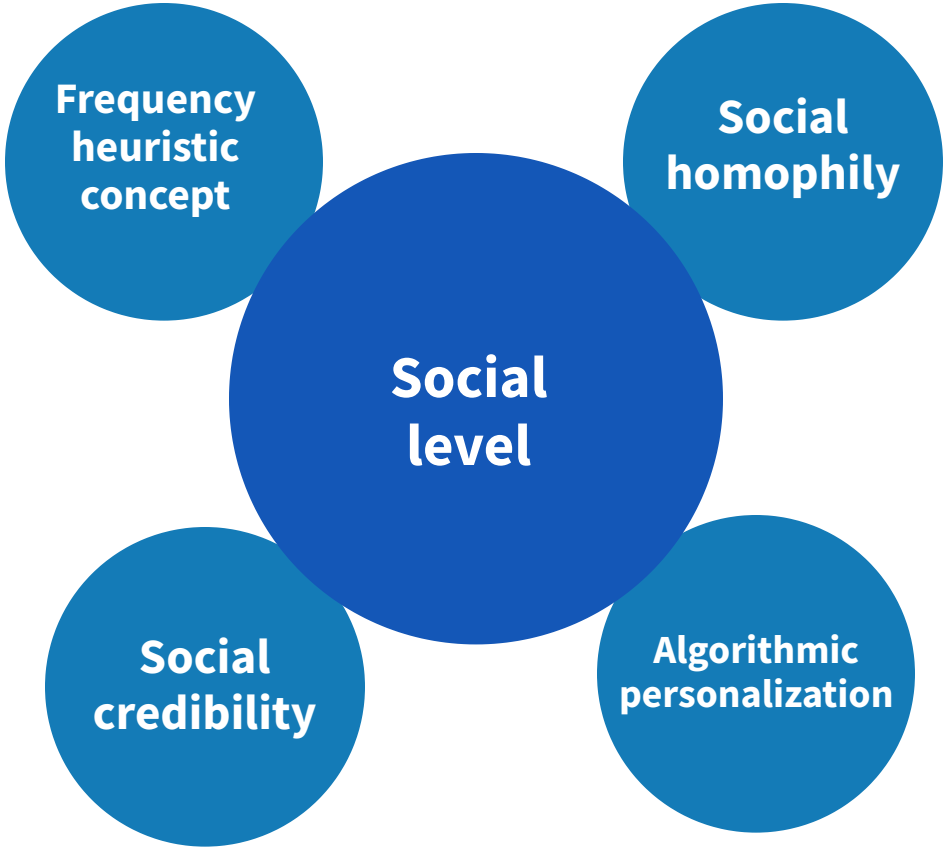
Factors that influenced the rise of **disinformation** across different media platforms:

- Hyper partisan news sites that use online propaganda as a business model for generating profit;
- Politicians increasingly using propaganda terms to frame political issues, instead of employing a fact-based approach;
- The technological advancement in the field of advertising algorithms and social media platforms that enabled the creation of partisan camps and polarized crowds

DOMINOES
digital resilience to disinformation

Factors that influence the spread of fake news and disinformation





Whereas technology can be used to amplify **disinformation** on social networks either through the creation and promotion of disinformation or through the use of social media bots, tech-driven solutions are also leading the way in the fight against disinformation. However, in order to better understand the way in which technology can be employed to counteract the negative effects of disinformation, it is important to acknowledge the factors that allow fake news and disinformation to spread at both individual and social level.

As far as individual level is concerned, factors can be categorized as follows:

Ideological biases – which includes (A) naive realism - refers to the individual tendency to trust more easily in information that is aligned with his/her own views, (B) confirmation bias – refers to the tendency of individuals to select and prefer to receive only that information which confirms their existing views, rejecting any piece of information that contravene their points of view and (C) normative influence – refers to the tendency of individuals to disseminate and consume socially safe options as a preference for social acceptance and affirmation (Shu, Sliva, Wang, Tang, & Liu, 2017, 24).

Cognitive abilities - Skills used on a daily basis for completing essential tasks, such as thinking, learning, working memory, listening, metacognition etc.

Time allotted for consuming media content – more time spent on social media shows a predisposition to believe the information available online, without checking its level of veracity

Level of education - helps individuals better understand the consequences of their online behavior.

Age - higher educated, older people tend to be more accurate in forming perceptions of information) (Allcott & Gentzkow, 2017, 228).

As the social level, the core of social media and collaborative information sharing on online platforms provides a supplementary dimension to disinformation and fake news, generally known as the echo chamber effect (Shu, Sliva, Wang, Tang, & Liu, 2017, 225).

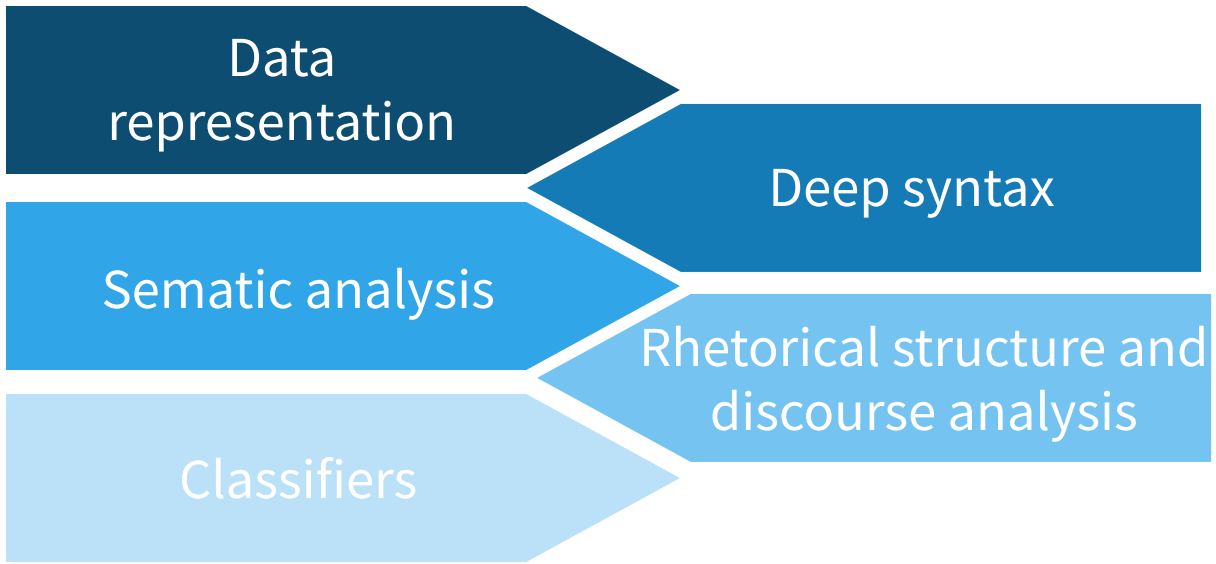
The **principles of naive realism**, confirmation bias, and normative influence theory stated above, describe the need of individuals to search, consume, and disseminate information that is in alignment with their own viewpoints and ideologies, developing, in consequence, the tendency to establish and develop connections with ideologically similar individuals (social homophily). Therefore, social media algorithms focus on customizing recommendations (algorithmic personalization) by suggesting content that better fits an individual's preferences, as well as by recommending connections to persons that share similar beliefs (Sharma, et al, 2019, 5).

Both social homophily and algorithmic personalization contribute to the development of echo chambers and filter bubbles, wherein individuals get less exposure to conflicting viewpoints and become isolated in their own information sphere (Garimella, Gionis, Parotsidis, & Tatti, 2017, 4663).

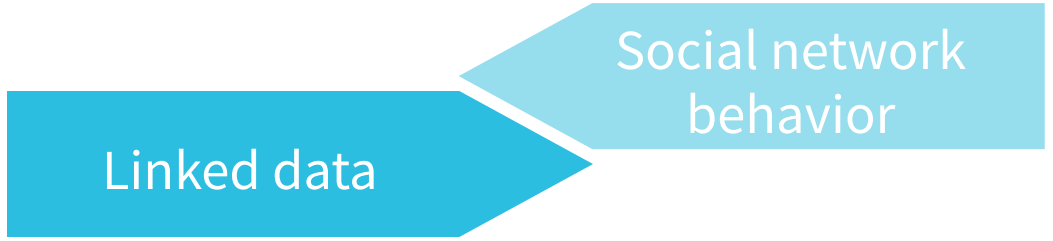
The existence of echo chambers can increase the chances of survival and continuous dissemination of fake news, aspect that can be explained by the phenomena of social credibility and frequency heuristic. The concept of social credibility indicates that people's perception of credibility of a piece of information tends to increase if others also perceive it as credible (especially in those cases when the credibility of the source of information cannot be tested), and the frequency heuristic concept defines the tendency to grant a higher level of credibility to a piece of information to which an individual is exposed multiple times (Shu, Sliva, Wang, Tang, & Liu, 2017, 25).

Taxonomy of technological methods to combat online disinformation

Linguistic approach



Network approach



Technology did not only create the premises for the expansion of online disinformation, but it also allowed the development of solutions to combat the negative effects of the above-mentioned phenomena. The majority of tech-driven solutions are relying on machine learning. The attractiveness of machine learning in the context of targeting and combating disinformation arises from the fact that machine learning models can recognize novel cases and react to them, based on prior learning. The possibility of continuous improvement of machine learning models, makes them seem like an effective tool to address the always-evolving world of disinformation.

In order to set the framework for a better understanding of the main solutions identified in the domain of combating the negative effects of online disinformation, it is essential to define the assessment methods that the most majority of technological solutions are based on. Therefore, the literature in the domain divides the methods, which emerged from various domains, using disparate techniques, into two main categories (see Conroy, Rubin, & Chen, 2015):

Linguistic approaches – focus on extracting and analyzing the content of deceptive messages in order to associate language patterns with deception. More specifically, this type of approach is aimed at identifying „leakages" in the content of the message analyzed by measuring the frequency and patterns of pronouns, conjunction, negative emotion word usage and so on (Feng & Hirst, 2013). The methods associated with this category are, as follows:

Data representation - One of the simplest methods of representing texts is the "bag of words" approach, which regards each word as a single, equally significant unit. In the bag of words approach, individual words or "n- grams" (multiword) frequencies are aggregated and analyzed to reveal cues of deception. However, by relying on isolated n-grams, often divorced from useful context information, any resolution of ambiguous word sense remains non-existent (Conroy, Rubin, & Chen, 2015, 2).

This method uses the principle of aligning profiles and the description of the writer's personal experience, in order to assess veracity based on compatibility scores: (1) compatibility with the existence of some distinct aspect (e.g. an art museum near a mentioned hotel); (2) compatibility with the description of some general aspect, such as location or service. In this case, the prediction of falsehood is shown to be at approximately 91% accurate (Conroy, Rubin, & Chen, 2015, 2).

Rhetorical Structure and Discourse Analysis - A method used to achieve the description of discourse, by identifying the instances of rhetoric relations between linguistic elements (Rubin & Lukoianova, 2014).

Classifiers -A mathematical model sufficiently trained from pre-coded examples in a specific category, that can predict instances of future deception on the basis of numeric clustering and distances (Conroy, Rubin, & Chen, 2015, 3).

Network approaches – focus on the usage of network properties and behavior to complement content-based approaches that rely on deceptive language and leakage cues to predict deception. The methods associated with this category are, as follows (Conroy, Rubin, & Chen, 2015, 3):

Linked data - An approach that leverages an existing body of collective human knowledge in order to assess the truth of new statements. The method is based on querying existing knowledge networks, or publicly available structured data (e.g. the Google Relation Extraction Corpus (GREC)). The structured data network of entities is connected through a predicate relationship. This particular method can help develop the applicability of fact-checking methods (Conroy, Rubin, & Chen, 2015, 3).

Social network behavior - besides content analysis, the use of metadata and telltale behavior of questionable sources can be examined. This method focuses on compiling the inclusion of hyperlinks or associated metadata to establish veracity assessments. As an example, centering resonance analysis (CRA), is a model of network-based text analysis, representing the content of large sets of texts by identifying the most important words that link other words in the network (Conroy, Rubin, & Chen, 2015, 3-4).

Machine Learning (ML) solutions to online disinformation

- *Linguistic features based methods*
- *Deception modelling based models*
- Clustering based models
- Predictive modelling based methods
- Content cues based models

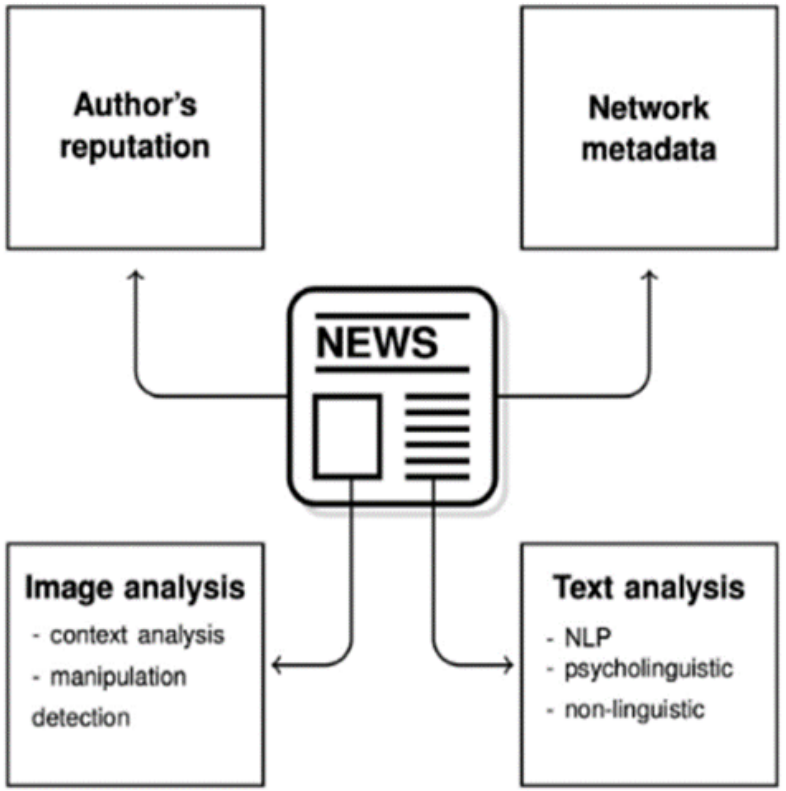


Image source: Choraś, Demestichas, Giełczyk, & Herrero, 2021, p. 2)

Techniques:

Rhetorical Structure Theory (RST): focuses on capturing the logic of a story in terms of functional relations created amongst different meaningful text units, describing, at the same time, a hierarchical structure for each story (Mann & Thompson, 1988). In accordance with the findings of the authors Victoria Rubin, Nadia Conroy and Yimin Chen (Rubin, Conroy, & Chen, 2015), empirical research confirmed in the last decades that writers tend to emphasise certain parts of their papers so as to express in a more evident manner the main ideas expressed in that article. In this context, the RST uses rhetorical connections to identify, in a systematic manner, the emphasized parts of a text (Parikh & Atrey, 2018, 439);

Vector Space Modeling (VSM): used to identify the rhetorical structure relations in the sets resulted after the application of RST. VSM helps at interpreting every news text as vectors in high dimensional space, aspect that requires for the extracted text to be modeled in an appropriate manner for the application of various computational algorithms (Rubin, Conroy, & Chen, 2015). In this context, each dimension of a certain vector space refers to the number of rhetorical relations in a complete set of news reports, representation which provides a simple explanation of a vector space, making it available for further analysis (Rubin & Lukoianova, 2014) (Parikh & Atrey, 2018, 439).

(3) Clustering based models – a known method to compare and contrast a large amount of data. For example, the gCLUTO package (Graphical CLUstering TOolkit) runs a large number of data set and sorts a small number of clusters using agglomerative clustering with the k-nearest neighbor approach, clustering similar news reports based on the normalized frequency of relations (Rubin, Conroy, & Chen, 2015);

(4) Predictive modelling based methods – develop the ability to make predictions about previously unseen news pieces on the results of a logistic regression process (Rubin, Conroy, & Chen, 2015);

(5) Content cues based models – a model based on the ideology of what journalists like to write for users and what are the preferences of users in terms of reading (choice gap), that leverages two different analyses: (I) lexical and semantic levels of analysis (automated methods can be used to extract stylometric features of the text, such as subjective terms, word length etc., further used to differentiate between journalistic formats) and (II) syntactic and pragmatic levels of analysis (the pragmatic function of headlines invokes reference to forthcoming parts in the discourse by making reference to forthcoming parts in the news story. This analysis also covers measuring news sites which have more share activity compared to sites that substantially produce more news content) (Parikh & Atrey, 2018, 440).

Complementary to the methods presented above, **the detection of false information and fake news can be performed by analyzing multiple types of digital content**, such as images, text data, network data, as well as the credibility degree of the author/source and its reputation (Choraś, Demestichas, Giełczyk, & Herrero, 2021, 1-2), as presented in Figure 1.

A survey conducted by a team of researchers from the University of Albany on the solutions to address fake news detection through text-analysis and mainstream fake news datasets showed that the state-of-the-art approaches for combating the effects of disinformation through detection can be clustered into five main categories, depending on the methods they use (Parikh & Atrey, 2018, 438):

(1) Linguistic features based methods – which extract key linguistic features from fake news and false information, as follows:

Ngrams: unigrams and bigrams are extracted from the matrix of words in a certain story. These are most often stored as TFIDF (Term Frequency Inverse Document Frequency) values for information retrieval. TFIDF refers to a numerical statistic that is intended to reflect how important a word is to the document that it is used in (Parikh & Atrey, 2018, 438);

Punctuation: using punctuation in an article can help the algorithms for fake news detection to make the difference between false and truthful texts, by collecting eleven types of punctuation, implemented through this detection (Parikh & Atrey, 2018, 438);

Punctuation: using punctuation in an article can help the algorithms for fake news detection to make the difference between false and truthful texts, by collecting eleven types of punctuation, implemented through this detection (Parikh & Atrey, 2018, 438);

Psycho-linguistic features: use the LIWC lexicon (Linguistic Inquiry and Word Count) in order to pick out appropriate proportions of words, allowing the system to determine the tone of the language (e.g. positive/negative emotions), statistics of the text (e.g. word counts), part-of-speech category (e.g. articles, nouns, verbs) (Pérez-Rosas, Kleinberg, Lefevre, & Mihalcea, 2018, 5);

Readability: includes the extraction of content features such as the number of characters, complex words, long words, number of syllables, word types, and number of paragraphs (Pérez-Rosas, Kleinberg, Lefevre, & Mihalcea, 2018, 5);

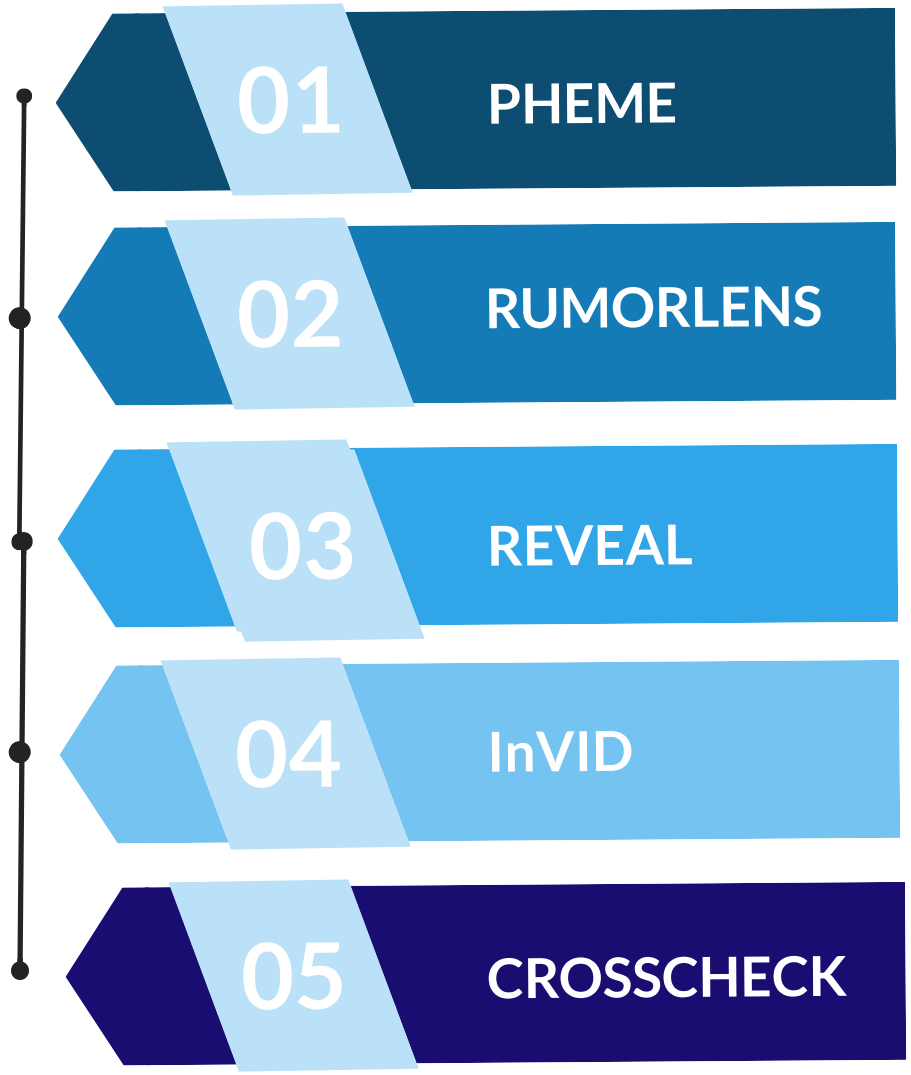
Syntax: focuses on extracting a set of features based on CFG (context-free grammar), which are heavily dependent on lexicalized production rules combined with their parent and grandparent nodes. Functions in this set are also encoded in TFIDF for information retrieval purposes (Parikh & Atrey, 2018, 439).

(2) Deception modelling based models – convert texts into a set of rhetorical relations connected in a hierarchical tree and identify the results of rhetorical structure relations by employing two theoretical

DOMINOES
digital resilience to disinformation

Projects

Inspiring practices, projects, interventions in the field - projects



In terms of fake news detection initiatives (with emphasis on rumors), both industry and the scientific community have registered efforts to identify and develop solutions, ranging from research projects (ongoing or already implemented) to fully-fledged applications. Examples of such initiatives in terms of projects are as follows:

PHEME - a 3-year research project funded by the European Commission, implemented during the period of 2014-2017, focusing on the study of natural language processing techniques for dealing with rumour detection and resolution

RumorLens - a 1-year research project that was implemented in 2014, funded by Google. Its main objective was to build a tool to aid journalists in finding posts that spread or correct a particular rumour on Twitter, by trying to identify the size of the audiences that those posts have reached

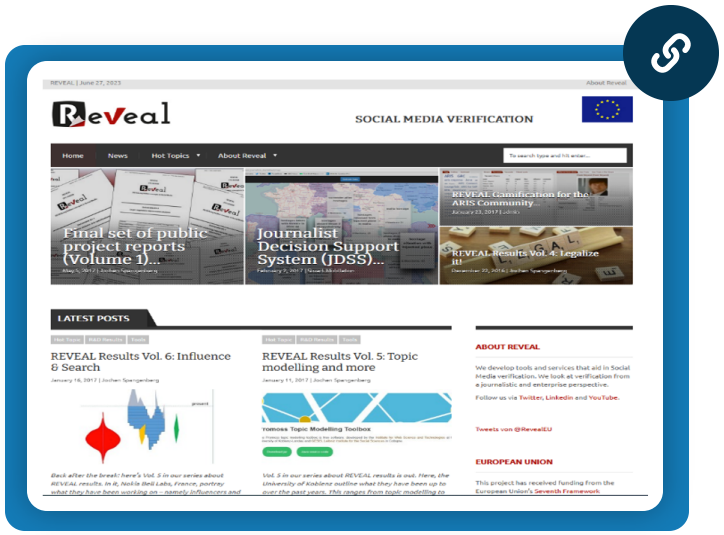
Reveal - a 3-year project funded by the European Commission, that aimed at verifying social media content from a journalistic and enterprise perspective, with a focus especially on image verification

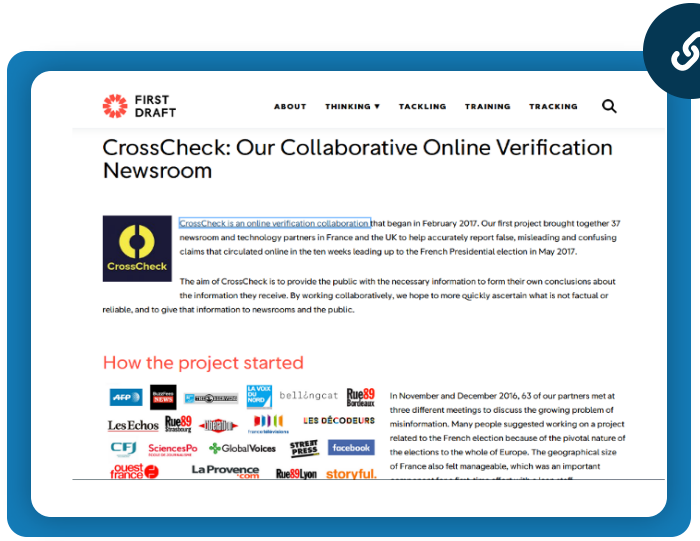
InVID - a Horizon 2020 project, funded by the European Commission with the target to build a platform providing services to detect, authenticate, and check the reliability and accuracy of newsworthy video files and video content spread via social media

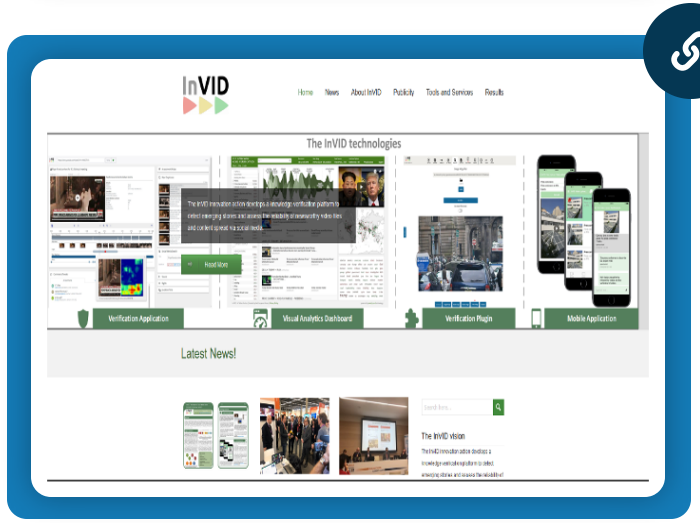
CrossCheck - collaborative verification project implemented by First Draft and Google News Lab, in collaboration with a number of newsrooms in France, with the objective to fight misinformation (mainly focusing on the French presidential election)

DOMINOES
digital resilience to disinformation

Inspiring practices, projects, interventions in the field - projects







In terms of fake news detection initiatives (with emphasis on rumors), both industry and the scientific community have registered efforts to identify and develop solutions, ranging from research projects (ongoing or already implemented) to fully-fledged applications. Examples of such initiatives in terms of projects are as follows:

PHEME - a 3-year research project funded by the European Commission, implemented during the period of 2014-2017, focusing on the study of natural language processing techniques for dealing with rumour detection and resolution

RumorLens - a 1-year research project that was implemented in 2014, funded by Google. Its main objective was to build a tool to aid journalists in finding posts that spread or correct a particular rumour on Twitter, by trying to identify the size of the audiences that those posts have reached

Reveal - a 3-year project funded by the European Commission, that aimed at verifying social media content from a journalistic and enterprise perspective, with a focus especially on image verification

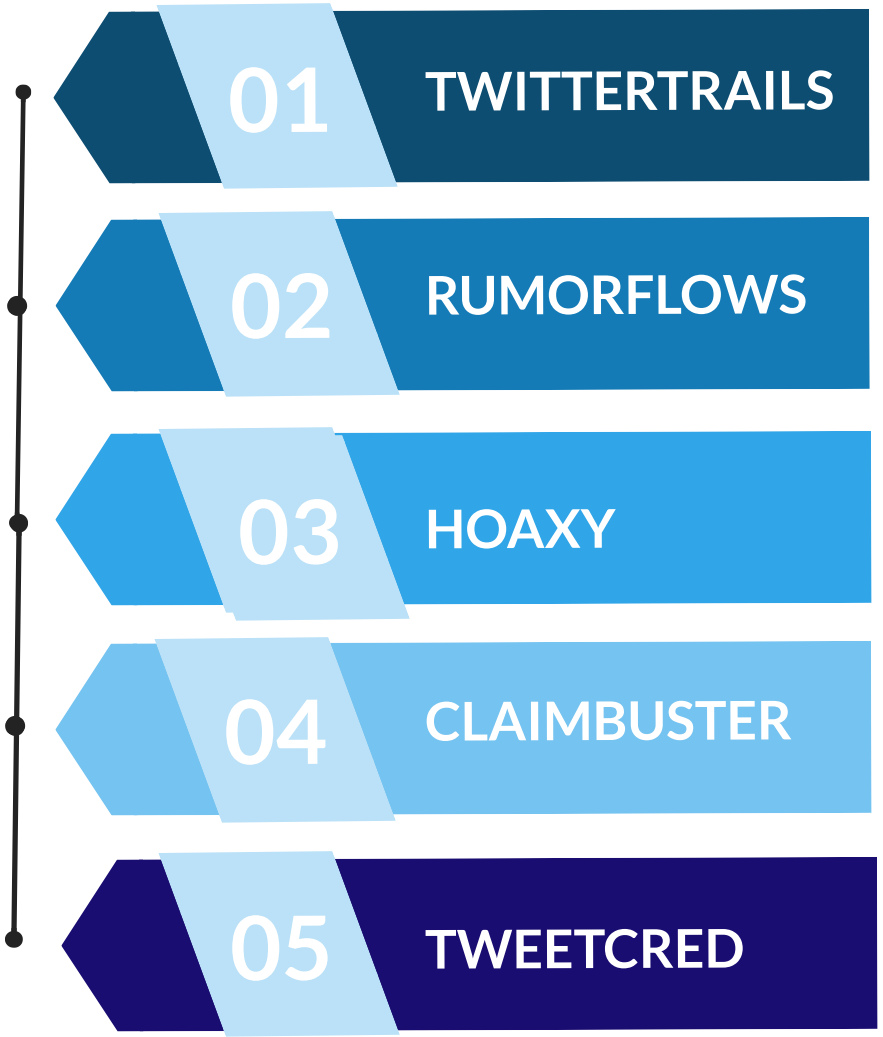
InVID - a Horizon 2020 project, funded by the European Commission with the target to build a platform providing services to detect, authenticate, and check the reliability and accuracy of newsworthy video files and video content spread via social media

CrossCheck - collaborative verification project implemented by First Draft and Google News Lab, in collaboration with a number of newsrooms in France, with the objective to fight misinformation (mainly focusing on the French presidential election)

DOMINOES
digital resilience to disinformation

Applications

Inspiring practices, projects, interventions in the field - applications



As far as applications are concerned, the following examples are worth mentioning:

TwitterTrails - an interactive, web-based tool that allows users to conduct an investigation on the origin and propagation characteristics of a rumor and its refutation, if applicable, on Twitter. It collects relevant tweets and automatically answers several important questions regarding a rumor: its originator, burst characteristics, propagators, and main actors according to the audience. In addition, this tool computes and reports the rumor's level of visibility.

RumorFlow - A framework that designs, adopts and implements multiple visualizations and modelling tools that can be mixed to identify rumor contents and analyze participant activity, either within a rumor, or across different rumors.

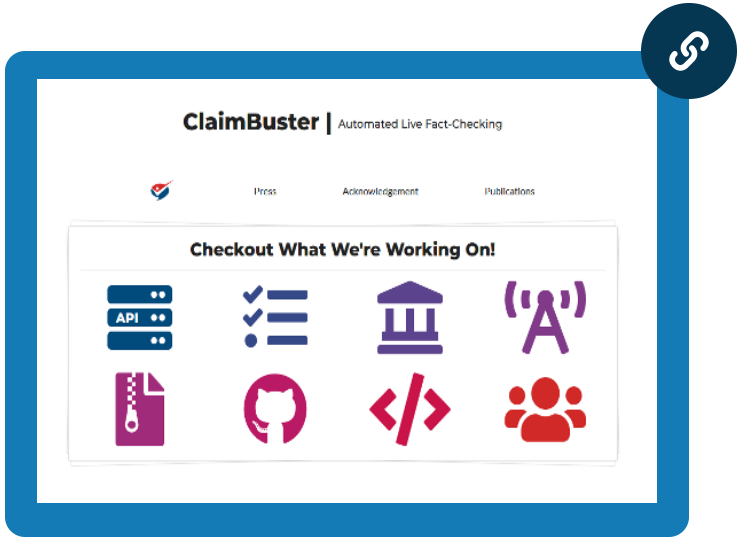
Hoaxy - a platform for the collection, detection and analysis of online misinformation and its related fact-checking efforts.

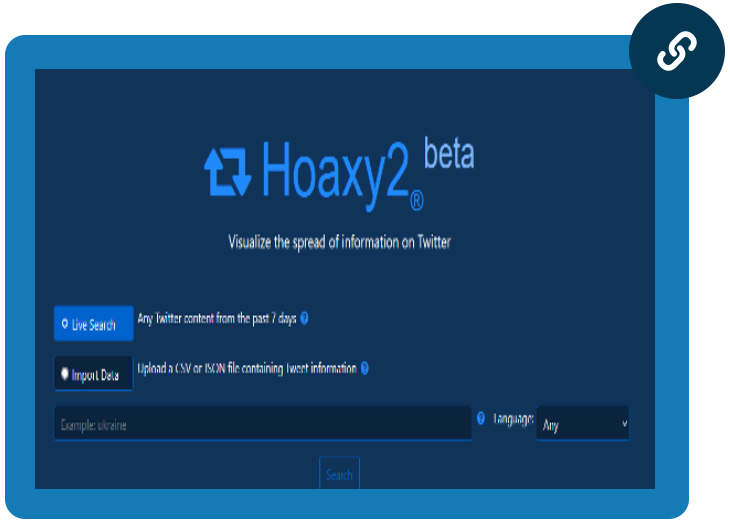
ClaimBuster - A project aiming to perform live fact-checking. The demo application shows check-worthy claims identified by the system for the 2016 U.S. election and it allows the user to input their own text to find factual claims.

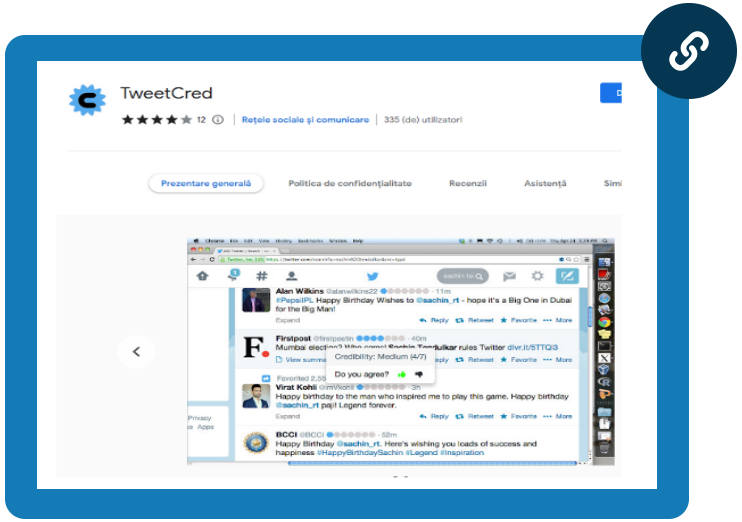
TweetCred - a real-time, web-based system developed to assess the credibility of content posted on Twitter. The system does not determine the veracity of stories, but it provides a credibility rating (scored 1 to 7) for each tweet in the Twitter timeline.

Therefore, we can conclude that the Machine Learning field allowed the development and expansion of applications and projects that can contribute to the active fight against the effects of online disinformation. However, the ML field is not the only one that provides inniatives in the counteracting online disinformation, another technology-based domain being represented by the serious games environment.

Inspiring practices, projects, interventions in the field - projects







As far as applications are concerned, the following examples are worth mentioning:

TwitterTrails - an interactive, web-based tool that allows users to conduct an investigation on the origin and propagation characteristics of a rumor and its refutation, if applicable, on Twitter. It collects relevant tweets and automatically answers several important questions regarding a rumor: its originator, burst characteristics, propagators, and main actors according to the audience. In addition, this tool computes and reports the rumor's level of visibility.

RumorFlow - A framework that designs, adopts and implements multiple visualizations and modelling tools that can be mixed to identify rumor contents and analyze participant activity, either within a rumor, or across different rumors.

Hoaxy - a platform for the collection, detection and analysis of online misinformation and its related fact-checking efforts.

ClaimBuster - A project aiming to perform live fact-checking. The demo application shows check-worthy claims identified by the system for the 2016 U.S. election and it allows the user to input their own text to find factual claims.

TweetCred - a real-time, web-based system developed to assess the credibility of content posted on Twitter. The system does not determine the veracity of stories, but it provides a credibility rating (scored 1 to 7) for each tweet in the Twitter timeline.

Therefore, we can conclude that the Machine Learning field allowed the development and expansion of applications and projects that can contribute to the active fight against the effects of online disinformation. However, the ML field is not the only one that provides inniatives in the counteracting online disinformation, another technology-based domain being represented by the serious games environment.

DOMINOES
digital resilience to disinformation

Tech-driven solutions to counter disinformation – serious games

The concept of serious game is considered to be an oxymoron, given the different domains that the term applies to, from entertainment to education, defense and even healthcare (Djaouti, Alvarez, Jessel, & Rampnoux, 2011, p. 26), receiving in the last years multiple definitions.

The concept of serious game is considered to be an oxymoron, given the different domains that the term applies to, from entertainment to education, defense and even healthcare (Djaouti, Alvarez, Jessel, & Rampnoux, 2011, p. 26). Even though the concept of serious game is being considered of recent history, the literature in the field shows that the first use of this term was tracked back in the Renaissance period, where Neo-Platonists used the syntagma of "serio ludere" to refer to the use of light-hearted humor in literature dealing with serious matters (Manning, 2004). During the past years, this concept has received multiple definitions, as follows:

"a free activity standing quite consciously outside 'ordinary' life as being 'not serious', but at the same time absorbing the player intensely and utterly" (Huizinga, 1951, 19);

"[...] serious games [...] have an explicit and carefully thought-out educational purpose and are not intended to be played primarily for amusement. This does not mean that serious games are not, or should not be, entertaining" (Djaouti, Alvarez, Jessel, & Rampnoux, 2011, 26);

"a game in which education (in its various forms) is the primary goal, rather than entertainment" (Michael & Chen, 2006, 17), "serious games have more than just story, art, and software, [they involve] pedagogy: activities that educate or instruct, thereby imparting knowledge or skill (Zyda, 2005, 26).

All these definitions were influenced by the vision of the author Ben Sawyer expressed in his paper "Serious Games: Improving Public Policy through Game-based Learning and Simulation", published in 2002. The main objective of this paper was to encourage the use of technology and knowledge from the entertainment video game industry to improve game-based simulations in public organizations (Sawyer & Rejeski, 2002).

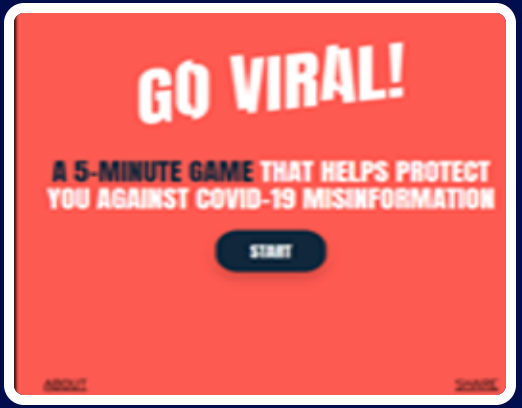
The above-mentioned paper was the promoter of the "Serious Games Initiative", an association that was founded in 2002 with the aim to promote the use of games for serious purposes (Djaouti, Alvarez, Jessel, & Rampnoux, 2011, 27). Therefore, this is the moment considered to be the "date of birth" of the oxymoron "serious games". In addition, 2002 was also the release date of America's Army, a game considered to be "[...] the first successful and well-executed serious game that gained total public awareness" (Djaouti, Alvarez, Jessel, & Rampnoux, 2011, 27), becoming, as a consequence, the starting point of the serious game current (with the current understanding and use of the concept). Michael Zyda, one of the members of the team that developed America's Army game proposed a definition that is referred to by various research papers: "a mental contest, played with a computer in accordance with specific rules, that uses entertainment, to further government or corporate training, education, health, public policy, and strategic communication objectives" (Zyda, 2005, 26).

Following 2002, most recent definitions of this concept tend to imply the use of digital games, instead of following the broader definition of "serious games" for both digital and non-digital games introduced in the 1970s. (Djaouti, Alvarez, Jessel, & Rampnoux, 2011, 27). One example is the definition used in 2011 by a team of researchers who studied the repurposing of games for educational objectives: "Serious games are very content-rich forms of educational media, often combining high fidelity visual and audio content with diverse pedagogical approaches" (Protopsaltis, et al., 2011, 37).

DOMINOES
digital resilience to disinformation

Erasmus+

Case studies – examples of serious games



For a better understanding of the role serious games play in the fight against online disinformation, we will present five such examples, which you can try by yourself, since they are available online without any costs:

The first one, called Bad News! Is a game that puts people in the position of a person who produces and disseminates fake news. The player's aim in the game is to obtain as many followers and shares as possible through creating and sharing different fake news items. Subjects exposed to the game tended to increase their willingness to engage in critical thinking and to take time to evaluate the accuracy of headlines that researchers exposed them to. Subjects exposed to the game tended to increase their willingness to engage in critical thinking and to take time to evaluate the accuracy of headlines that researchers exposed them to.

The second one, Harmony Square, is a game that places the player in the shoes of a candidate in elections who can be elected by creating political polarization. This game has also been shown to determine players to rate fake news as less accurate. As described on its website, “the goal of the game is to expose the tactics and manipulation techniques required in order to mislead people, build up a following, or exploit societal tensions for political purposes. Harmony Square works as a psychological “vaccine” against disinformation: playing it builds cognitive resistance against common forms of manipulation that the user may encounter online” (Harmony Square, 2021).

The Go Viral game is an online game based on misinformation spread during the COVID pandemic. In this particular game, players are asked to imagine that they controlled a social media profile and were asked to obtain as many likes and "credibility points", by sharing posts based on different argumentative fallacies.

Cracky Uncle is a game that explains different logical fallacies used by climate change deniers in the form of a cranky old man who issues pronouncements on the non-existence or alternative causes of climate change. Several experiments by researchers showed how playing the game increased the ability to identify and the knowledge of how to use logical fallacies by students in different study programs (Compton, van der Linden, Cook & Basol, 2021).

Another game developed in order to get the users familiarized with the principles of disseminating news in the online environment, as well as with the impact the way in which each piece of news is disseminated can have on the public opinion is the BBC Ireporter. In this game, users “play the role of a social media journalist who is faced with a major breaking story” (Cellan-Jones, 2018). The game is designed to be as realistic as possible, as well as immersive, including elements that involve chatting, having video calls with other journalists and so on. The players need to make decisions with tradeoffs, for example speed and accuracy, whether to publish a story as quickly as possible or to confirm first with a reliable source. The game educates the players more on the side of how good journalism is and what to consider before sharing a story (Bambang, 2020, 4).

In conclusion, serious games could prove a valuable asset in training citizens to detect disinformation attempts and build their resilience against them. Given their playful nature, they could capture the attention of all age groups and make them realise when they are targets of disinformation, thus preventing them from spreading those posts further and contributing to the viralisation of malicious content.

DOMINOES
digital resilience to disinformation

Exercise [1/3]

Which of the following are factors that allow fake nes and disinformation to spread at individual level?

You can select more than one answer

- social homophily
- naïve realism
- cognitives abilities
- age

SEND

Which of the following are included in the taxnomy of Machine Learning approaches for combating the effects of disinformation through detection ?

You can select more than one answer

- linguistic feature
- deception modelling
- clustering
- content cues

SEND

DOMINOES
digital resilience to disinformation

Exercise [2/3]

What does linguistic approach to combat online disinformation imply?

- focus on the usage of network properties and behavior to complement content-based approaches that rely on deceptive language and leakage cues to predict deception
- extracting and analyzing the content of deceptive messages in order to associate language patterns with deception
- prediction of instances of future deception on the basis of numeric clustering and distances

SEND

What does the Rhetorical Structure Theory refers to?

- extraction of content features such as the number of characters, complex words, long words, number of syllables, word types, and number of paragraphs
- capturing the logic of a story in terms of functional relations created amongst different meaningful text units, describing, at the same time, a hierarchical structure for each story
- collecting eleven types of punctuation, to make the difference between false and truthful texts

SEND

DOMINOES
digital resilience to disinformation

Exercise [3/3]

Complete the missing words in the sequence:

The concept of serious game is considered to be an _____ , given the different domains that the term applies to, from entertainment to education, defense and even healthcare (Djaouti, Alvarez, Jessel, & Rampnoux, 2011, p. 26), receiving in the last years the following definition: “a game in which _____ (in its various forms) is the primary goal, rather than _____” (Michael & Chen, 2006, 17), “serious games have more than just story, art, and software, [they involve] _____: activities that educate or instruct, thereby imparting _____ or skill (Zyda, 2005, 26).

Write your answer here.

0/1000

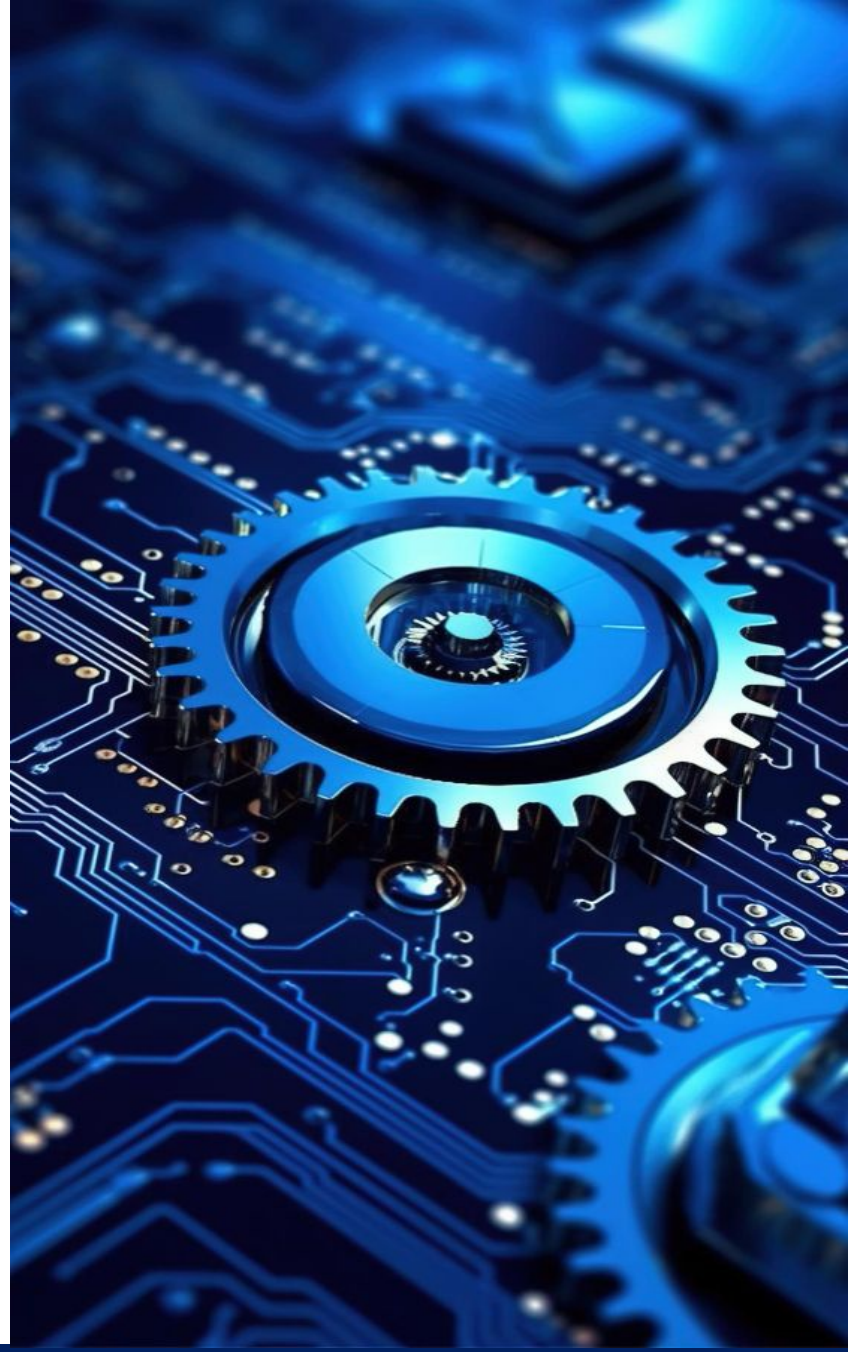
SEND

DOMINOES
digital resilience to disinformation

Limits of Technological Tools, Best Practices

Natural language processing tools – limitations

- **Dataset bias**
- **Language limitations**
- **Accuracy**
- **Transparency and accountability**



The majority of tech-driven solutions rely on machine learning. The attractiveness of machine learning in the context of targeting and combating disinformation arises from the fact that machine learning models can recognize novel cases and react to them, based on prior learning.

For example, Natural language processing (NLP) tools have the ability to parse text “[and] the ability of this paring is usually to predict something about the meaning of the text, such as whether to express a positive or a negative opinion’’ (Duarte et al, 2018, p.3). NLP relies on classifiers trained through text labels/annotations determined by humans which guide the tool to decipher whether some word, phrase or text belongs to the targeted category of content, for example, disinformation.

A collection of examples based on which NLP distinguishes different categories of text is called corpus. In the context of detecting disinformation, the NLP tool will use a corpus which has examples of accurate information and disinformation. Disinformation would then be annotated in a way that the tool could learn from this example and employ it automatically in the future. For example, the NLP tool could determine whether some words are missing, but also analyze the word embeddings that represent the context. NLP tools can replace journalists and media experts in the process of fact-checking, as well as imitating a high level of intuitive reasoning, similar to experienced specialists. Although NLP tools could contribute to a more efficient prevention and countering of disinformation, they are not without limitations and shortcomings. Failing to address these shortcomings, would not only invalidate their efficiency, but quite the contrary they could contribute to spreading more disinformation and/or negatively impacting human rights.

Dataset bias - NLP tools used for combating disinformation are highly dependent on the quality of the training data or in other words, “with limited human direction, an artificial agent is as good as the data it learns from” (Osoba et al., 2017, p.17). This would also mean that, if the data used for training is biased, the automated tool will reproduce these biases, or Will exacerbate them (Raso et al., 2018). In the majority of the cases, bias is introduced during the data collection process, specifically during the annotation process. An example of dataset bias, that specifically targets one demographic group can be found in a case study put forth by the European Union Agency for Fundamental Rights (FRA). In this particular example, the automatic system indicated a correlation between text being labelled as offensive when written in the African American English dialect, proving that content may be misclassified based on the expressions certain ethnic groups are using (FRA, 2022, p. 69). Language limitations - Machine-learning NLP tools cannot parse text in all languages. Given the fact that there are hundreds of languages in the world, machine-learning NLP tools will be effective in the case of high-resource languages (HRLs) such as English Spanish, German, and Chinese, while their accuracy will be significantly lower in the case of low-resource languages (LRLs) such as Bengali, Punjabi, Indonesian, although these languages are spoken by millions of people (Hirschberg, Manning, 2016).

This would consequently mean that the machine-learning NLP used for detecting disinformation could have “disproportionately harmful outcomes for non-English speakers”, especially if the outcome of the machine analysis is to be used as part of a decision-making process (Duarte et al, 2018).Accuracy - The accuracy of these NLP tools is significantly lowered in cases where the context can completely transform the meaning of the claim which the system could mark as disinformation. As Asudeh et al., explain, one may claim that “[she] has never lost a game of chess” which can be truthful information for an experienced chess player, but also for someone who has never played chess (Asudeh et al., 2020). NLP tools have a difficulty distinguishing whether similar claims are truthful or not, since they are entirely context dependent. Furthermore, machine-learning NLP tools experience difficulties in detecting “context, subtlety, sarcasm, and subcultural meaning” (Gillespie, 2020, p. 3).Transparency and accountability - Given that automated processes such as NLP tools can almost autonomously manage online behaviour as well as enforce rights it is important to apply scrutiny over these and other similar technological solutions. There is often little to no information on how automated tools make correlations or decisions, nor is there any data on their accuracy and reliability. As a response to the lack of transparency significant investment has been made in the field of Explainable Artificial Intelligence (XAI) – “a field focused on the understanding and interpretation of the behaviour of AI systems” (Linardatos, 2020, p. 2)

DOMINOES
digital resilience to disinformation

Limits of Technological Tools, Best Practices

Social media bot-detecting tools-limitations

- Limited datasets
- Language limitations
- Misclassification

As is the case in machine learning NLP tools, machine learning bot detection tools operate based on the availability and quality of the training data which has been labelled or annotated, specifically providing examples of human-managed and bot-managed social media accounts. Whereas one may argue that in the case of machine-learning NLP tools targeting disinformation, the objective is clear, in the context of machine-learning bot detection tools this is more challenging. First, there is no operational definition of social media bots. Secondly, there is a large grey area in detecting the differences between human-like and bot-like behaviour. Existing bot detection tools rely on datasets that map typical bot behaviour, but the final result is often impacted by the following limitations (Yang et al., 2022):

Limited datasets - the development of supervised social media bot-detecting tools relies on the existence of training datasets. These datasets used for annotation and labelling are often limited, due to them being directly extracted from social media. In recent years, especially in the context of the Cambridge Analytica case, many social platforms have limited access to their APIs as a result of human rights concerns or monetized access making it increasingly difficult to employ social media for training AI models. This means that datasets are compiled by human operators who often manually label and annotate information, which leads to a very high error rate in detecting the more sophisticated bots. The accuracy of training datasets is also impacted by bot evolution. Social media bots were initially easily recognizable since they often lacked personal information and presented few social connections, however with time these bots evolved into perfectly engineered accounts that seem human-operated and displaying a large social network (Cresci, 2020);

Language limitations - social media bot-detection tools cannot be transferred from one country to another one. Given the lack of training data available in other languages (all other languages but English), detection tools are likely to give false positives, or negatives since they fail to take into consideration different communication patterns and styles (Rauchfleisch et al, 2020). This would mean that current machine-learning social media bots detection tools cannot disproportionately target social media users who are non-native English speakers.

Misclassification - Often, human-operated social media accounts can behave similarly to a bot-operated account, namely they don't disclose a lot of personal data, including their location, and they don't share any audio-visual content. In other scenarios, individuals try to randomize their handles to protect their data and privacy. Due to biases found in the training there is a high likelihood for such accounts to be labelled as bots. Furthermore, taking down such accounts on the premise of them being bots would infringe the freedom of expression of individuals using those social media accounts.

Human rights impact and the way forward

DOMINOES
digital resilience to disinformation

1. Technological tools must respect the right to human dignity, the right to life, and the right to physical and mental integrity
2. The right to liberty and security
3. Special attention must be given to safeguarding the right to non-discrimination
4. The right to respect for private and family life and protection of personal data
5. The right to an effective remedy for violation of rights and freedoms
6. The right to a fair trial and due process
7. Tools should in no way inflict the right to freedom of expression and freedom of assembly and association.

The right to an effective remedy for violation of rights and freedoms (Article 13 ECHR) must be protected. Authorities and developers should make sure that there are accessible remedies individuals can rely on in case of unlawful data collection or if the implementation of such technologies causes unjust harm to the individual or violates their rights.

Similarly, to the above-mentioned right, the right to a fair trial and due process (Article 6, ECHR) should be respected. Individuals must have the opportunity to challenge any decisions made based on evidence acquired through the use of these technologies.

Although these technologies are often used to prevent interference in the electoral process through the creation and promotion of disinformation, these tools should in no way inflict the right to freedom of expression and freedom of assembly and association (Article 10 and 11 ECHR). Technologies which target disinformation should respect the principle of transparency, fairness, and responsibility. This obligation is particularly important when it comes to the transparency of algorithms (Leslie et al, 2021).

It is clear that there are multiple technologies which play/will play an important role in detecting and combating disinformation. They protect democracies and their citizens from unlawful interference in their internal processes and shed light on the mechanisms used to manipulate public opinions. This positive impact is not without costs. As explained in this module, many of these technologies are still underdevelopment and thus subject to many limitations. In addition to the high error rate, there are also cases where their use can have a negative impact on human rights. In order to avoid this, stronger emphasis needs to be placed on understanding technological limitations, introducing privacy-by-design and privacy-by-default approaches in their developments as well as carrying out a constant review of ways in which their design can be improved in order to mitigate potential risks. In addition to this, it is important to ensure that the regulatory framework manages to keep the pace with technological developments, by introducing the necessary safeguards.

While technology can be an aid in preventing and countering disinformation, it is also clear that many of the tools developed for this purpose can have a negative impact on human rights. For these reasons, the Council of Europe has identified a set of core rights which need to be protected at an individual level, this becoming a key requirement for any technology developed in the field.

Technological tools must respect the right to human dignity, the right to life, and the right to physical and mental integrity, defined in Article 2, of the European Convention of Human Rights (ECHR). This would mean that when there is a risk of technological tools violating human dignity, the same procedure must be carried out by a human.

The right to liberty and security (Article 5, ECHR), must be respected at all times. This right prescribes an obligation towards developers to establish human rights oversight mechanisms which would evaluate possible risks arising from the implementation of those technologies. Such oversight mechanisms could help in addressing issues arising from dataset bias, language limitations or lack of algorithmic transparency.

Special attention must be given to safeguarding the right to non-discrimination (on the basis of the protected grounds set out in Article 14 of the ECHR and Protocol 12 to the ECHR). To prevent dataset bias, authorities and developers must ensure that deployed technologies do not cause discrimination, promote harmful stereotypes or foster social inequality. Developers must be aware of these risks and continuously examine if in any way bias is fostered through the development and implementation of these technologies.

The right to respect for private and family life and protection of personal data (Article 8, ECHR) must be safeguarded. Developers should mitigate any negative impact of technological tools on the right to privacy or family life that might rise either in the development or implementation stage. Protection of this right is particularly relevant in the context of bot detection technologies that tend to show a false positive for profiles where individuals are more protective of their personal information.

DOMINOES
digital resilience to disinformation

Bibliography

and useful resources

2021-1-RO01-KA220-HED-000031158

DOMINOES
digital resilience to disinformation

References

- Bambang, R. J. (2020). The Fake News Detective: A Game to Learn Busting Fake News as Fact Checkers using Pedagogy for Critical Thinking. Retrieved January 15th, 2023, from https://bit.ly/3KWPNlh
- Basol, M., Roozenbeek, J., Berriche, M., Uenal, F., McClanahan, W. P., & Linden, S. V. D. (2021). Towards psychological herd immunity: Cross-cultural evidence for two prebunking interventions against COVID-19 misinformation. Big Data & Society, 8(1), 20539517211013868.
- Cellan-Jones, R. (2018). Fake news: Can teenagers spot it? Retrieved January 15th, 2023, from https://www.bbc.com/news/technology-46206675
- Choraś, M., Demestichas, K., Giełczyk, A., & Herrero, Á. (2021). Advanced Machine Learning techniques for fake news (online disinformation) detection: A systematic mapping study. Applied Soft Computing Journal(101), 1-15.
- Compton, J., van der Linden, S., Cook, J., & Basol, M. (2021). Inoculation theory in the post-truth era: Extant findings and new frontiers for contested science, misinformation, and conspiracy theories. Social and Personality Psychology Compass, 15(6), e12602.
- Conroy, N. J., Rubin, V. L., & Chen, Y. (2015). Automatic Deception Detection: Methods for Finding Fake News. Proceedings of the Association for Information Science and Technology, 52(1), 1-4.
- Djaouti, D., Alvarez, J., Jessel, J.-P., & Rampnoux, O. (2011). Origins of Serious Games. In Serious Games and Edutainment Applications (pp. 25-43). Springer.
- Duarte, N., Llanso, E., & Loup., A. (2018). Mixed Messages? The Limits of Automated Social Media Content Analysis. The 2018 Conference on Fairness, Accountability, and Transparency. https://cdt.org/wp-content/uploads/2017/12/FAT-conference-draft-2018.pdf
- Feng, V. W., & Hirst, G. (2013). Detecting Deceptive Opinions with Profile Compatibility. Nagoya, Japan: Proceedings of the Sixth International Joint Conference on Natural Language Processing.
- Florida Law Review. https://scholarship.law.ufl.edu/cgi/viewcontent.cgi?article=1348&context=flr
- Huizinga, J. (1951). Homo Ludens: A Study of the Play-Element in Culture. Paris: Gallimard.

DOMINOES
digital resilience to disinformation

References

- Kessler, G., Rizzo, S., Kelly, M. (2019, December 16). President Trump has made 15,413 false or misleading claims over 1,055 days. The Washington Post. https://bit.ly/44yHymA
- Michael, D., & Chen, S. (2006). Serious Games: Games that Educate, Train and Inform. Boston, MA: Thomson Course Technology PTR.Zyda, M. (2005). From Visual Simulation to Virtual Reality to Games. Computer, 38(9), 25-32.
- Osoba O., Wesler IV W. (2017). An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence. RAND Corporation. https://www.rand.org/content/dam/rand/pubs/research_reports/RR1700/RR1744/RAND_RR1744.pdf
- Raso, F., Hilligoss, H., Krishnamurthy, V. et al. (2018, September 25). Artificial Intelligence & Human Rights: Opportunities & Risks. Berkman Klein Center for Internet & Society, Harvard University.
- Rauchfleisch, A., Kaiser, J., (2020). The False positive problem of automatic bot detection in social science research. PLoS ONE. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0241045
- Roozenbeek, J., & Van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. Palgrave Communications, 5(1), 1-10.
- Rubin, V. L., & Lukoianova, T. (2014). Truth and Deception at the Rhetorical Structure Level. Journal of the Association for Information Science and Technology, 66(5), 905-917.
- Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., & Liu, Y. (2019). Combating Fake News: A Survey on Identification and Mitigation Techniques. ACM Transactions on Intelligent Systems and Technology, 10(3), 1-42.
- Zubiaga, A., Aker, A., Bontcheva, K., Liakata, M., & Procter, R. (2018). Detection and Resolution of Rumours in Social Media: A Survey. ACM Computing Surveys, 51(2), 1-36.

DOMINOES
digital resilience to disinformation

Digital cOMpetences INformatiOn EcoSystem

ANIMV DARE. LEARN. INNOVATE.
Universidad Rey Juan Carlos
L-Università ta' Malta
NEW STRATEGY CENTER

Co-funded by the European Union

2021-1-RO01-KA220-HED-000031158

Audiovisual and multimedia production: CIBERIMAGINARIO

DOMINOES
digital resilience to disinformation

Combating the effects of disinformation in the online environment

4.2.1

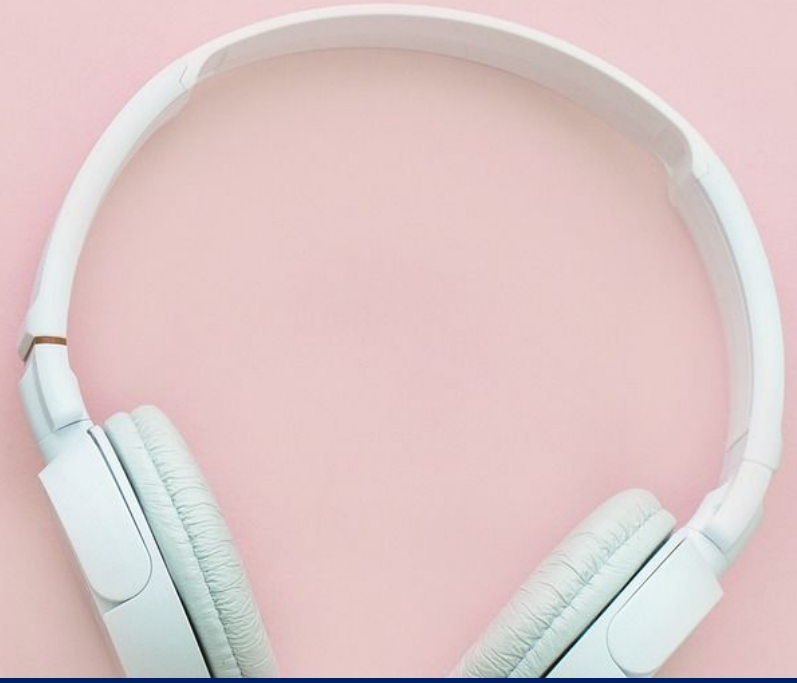
doi.org/10.5281/zenodo.10064661

Co-funded by the European Union

2021-1-RO01-KA220-HED-000031158

ANIMV DARE. LEARN. INNOVATE.
Universidad Rey Juan Carlos
L-Università ta' Malta
NEW STRATEGY CENTER

Technological developments and advances registered in the sector of **Internet of Things** and **social networks** have created the premises for the expansion of the noxious effects of the **disinformation phenomena**, together with the rise of ubiquitous misinformation, disinformation, deepfakes, and post-truth.



DOMINOES

Co-funded by the European Union

2021-1-RO01-KA220-HED-000031158

Digital cOMpetences INformatiOn EcoSystem

Technology has created the means for the expansion of the disinformation phenomenon, social media becoming one of the main sources of information for the population at large, as well as an important source of **false content** and **digital deception**. However, technology can also play an **essential role** in combating the effects of online disinformation and propaganda and in containing the expansion processes of these now defined security issues.

DOMINOES

Co-funded by the European Union

2021-1-RO01-KA220-HED-000031158

Digital cOMpetences INformatiOn EcoSystem

Factors that influenced the rise of disinformation across different media platforms:

- Hyper partisan news sites that use online propaganda as a business model for generating profit.
- Politicians increasingly using propaganda terms to frame political issues, instead of employing a fact-based approach.
- The technological advancement in the field of advertising algorithms and social media platforms that enabled the creation of partisan camps and polarized crowds.

DOMINOES

Co-funded by the European Union
2021-1-RO01-KA220-HED-000031158
Digital cOMpetences INformatiOn EcoSystem

Factors that influence the spread of fake news and disinformation

DOMINOES

Whereas technology can be used to amplify disinformation on social networks either through the creation and promotion of disinformation or through the use of social media bots, tech-driven solutions are also leading the way in the fight against disinformation. However, in order to better understand the way in which technology can be employed to counteract the negative effects of disinformation, it is important to acknowledge the factors that allow fake news and disinformation to spread at both individual and social level.

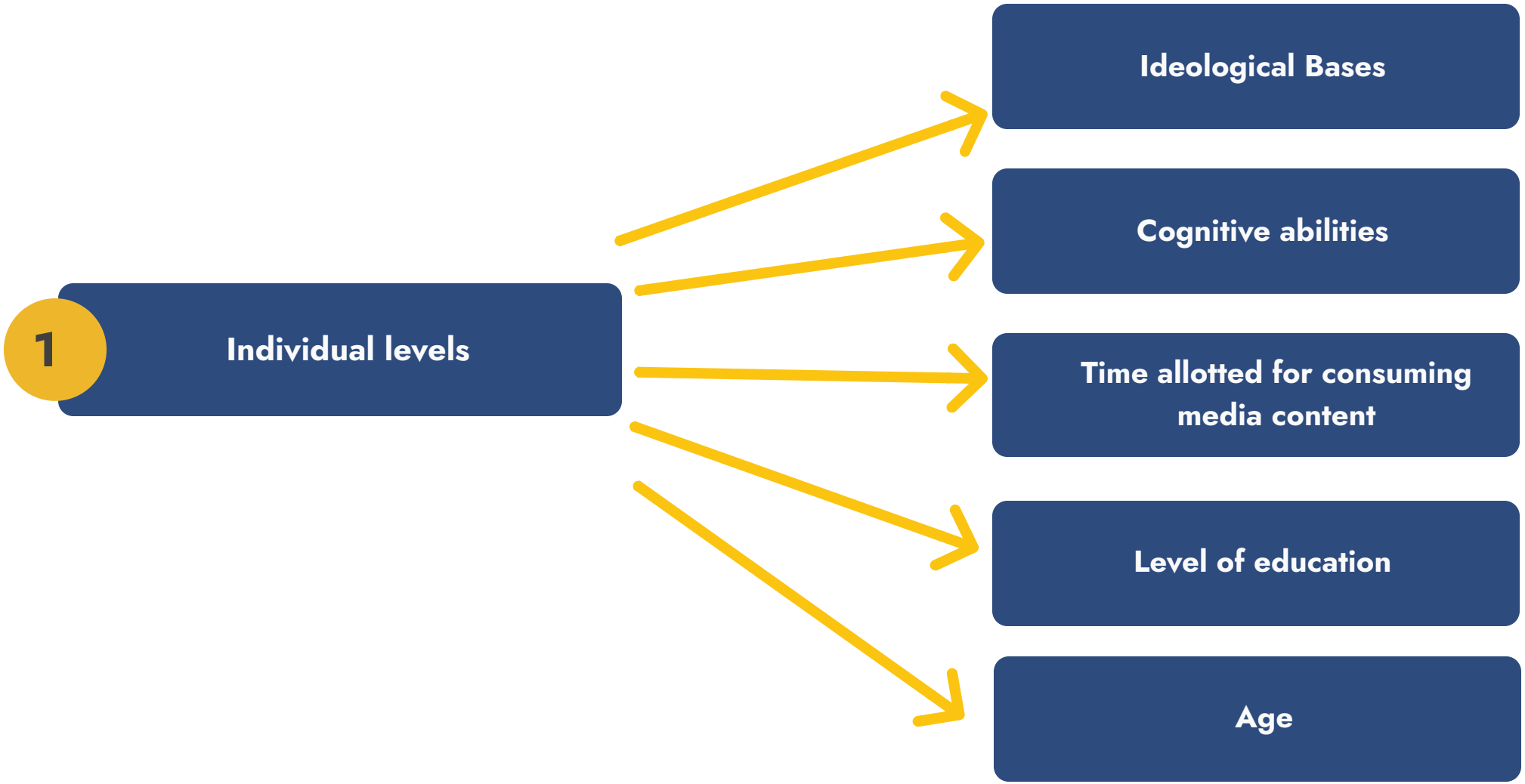
Co-funded by the European Union

2021-1-RO01-KA220-HED-000031158

Digital cOMpetences INformatiOn EcoSystem

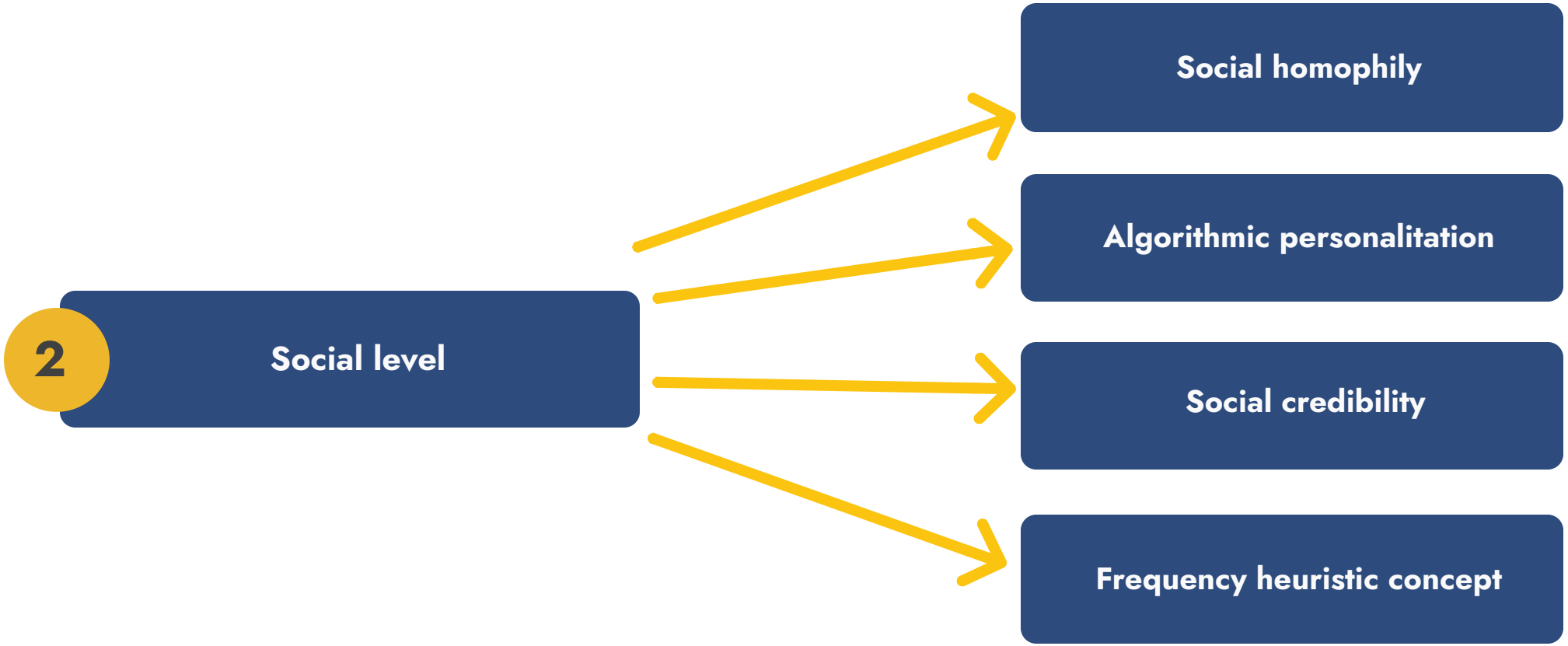
Factors that influence the spread of fake news and disinformation

DOMINOES



Factors that influence the spread of fake news and disinformation

DOMINOES



Co-funded by the European Union
2021-1-RO01-KA220-HED-000031158
Digital cOMpetences INformatiOn EcoSystem

Taxonomy of technological methods to combat online disinformation

DOMINOES

Technology did not only create the premises for the expansion of online disinformation, but it also allowed the development of solutions to combat the negative effects of the above-mentioned phenomena.
The majority of tech-driven solutions are relying on machine learning. The attractiveness of machine learning in the context of targeting and combating disinformation arises from the fact that machine learning models can recognize novel cases and react to them, based on prior learning. The possibility of continuous improvement of machine learning models, makes them seem like an effective tool to address the always-evolving world of disinformation.

Co-funded by the European Union
2021-1-RO01-KA220-HED-000031158
Digital cOMpetences INformatiOn EcoSystem

Taxonomy of technological methods to combat online disinformation

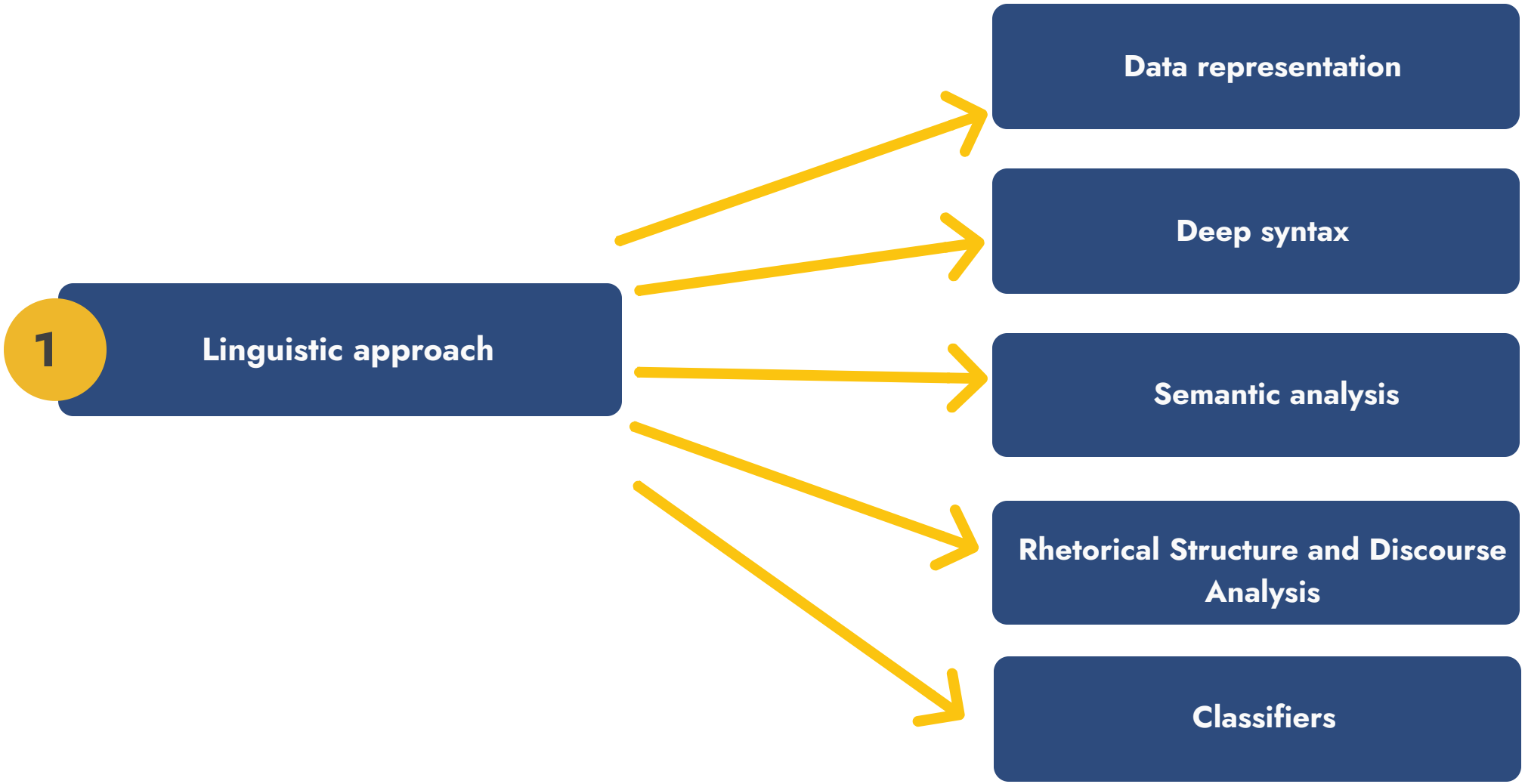
DOMINOES

1. Linguistic approaches
2. Network approaches

Co-funded by the European Union
2021-1-RO01-KA220-HED-000031158
Digital cOMpetences INformatiOn EcoSystem

Tech-driven methods used to combat online disinformation

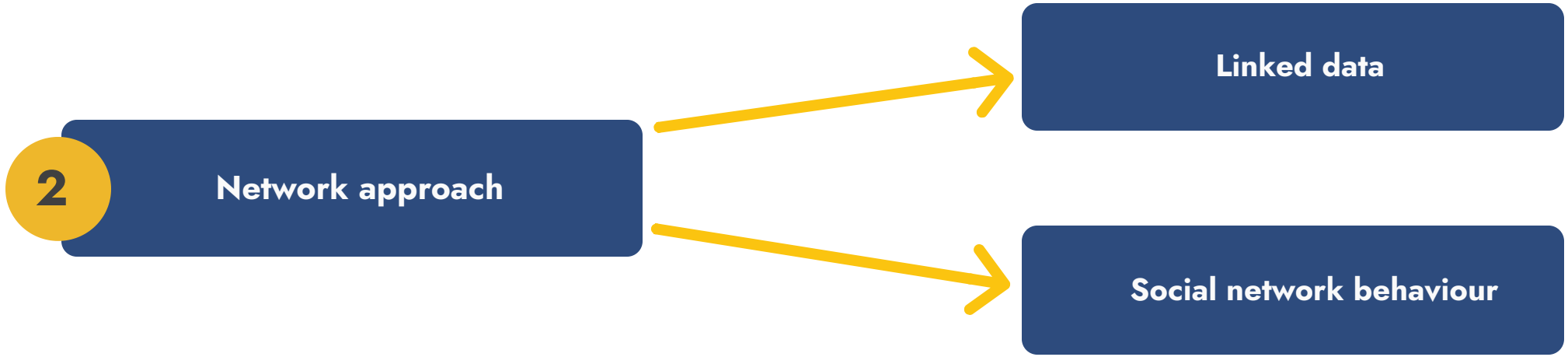
DOMINOES



Co-funded by the European Union
2021-1-RO01-KA220-HED-000031158

Tech-driven methods used to combat online disinformation

DOMINOES



Co-funded by the European Union
2021-1-RO01-KA220-HED-000031158
Digital cOMpetences INformatiOn EcoSystem

Machine Learning (ML) solutions to online disinformation

The detection of false information and fake news can be performed by analyzing multiple types of digital content: images, text data, network data, as well as the credibility degree of the author/source and its reputation (Choraś, Demestichas, Giełczyk, & Herrero, 2021, 1-2).

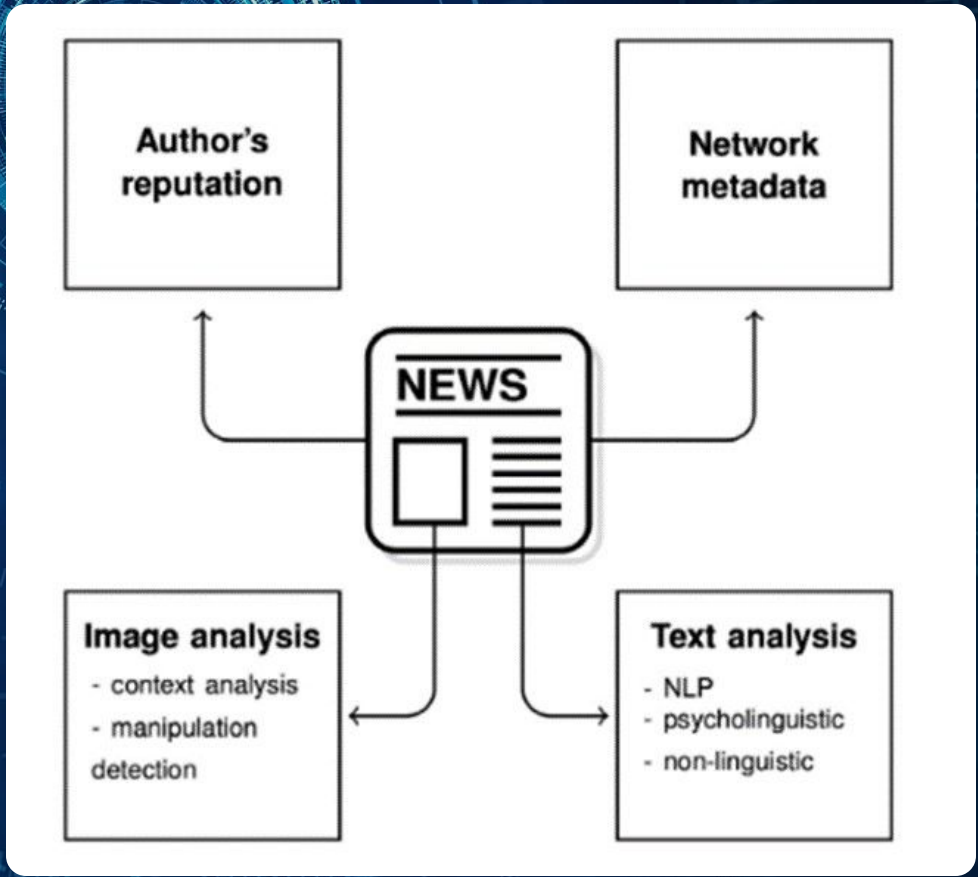


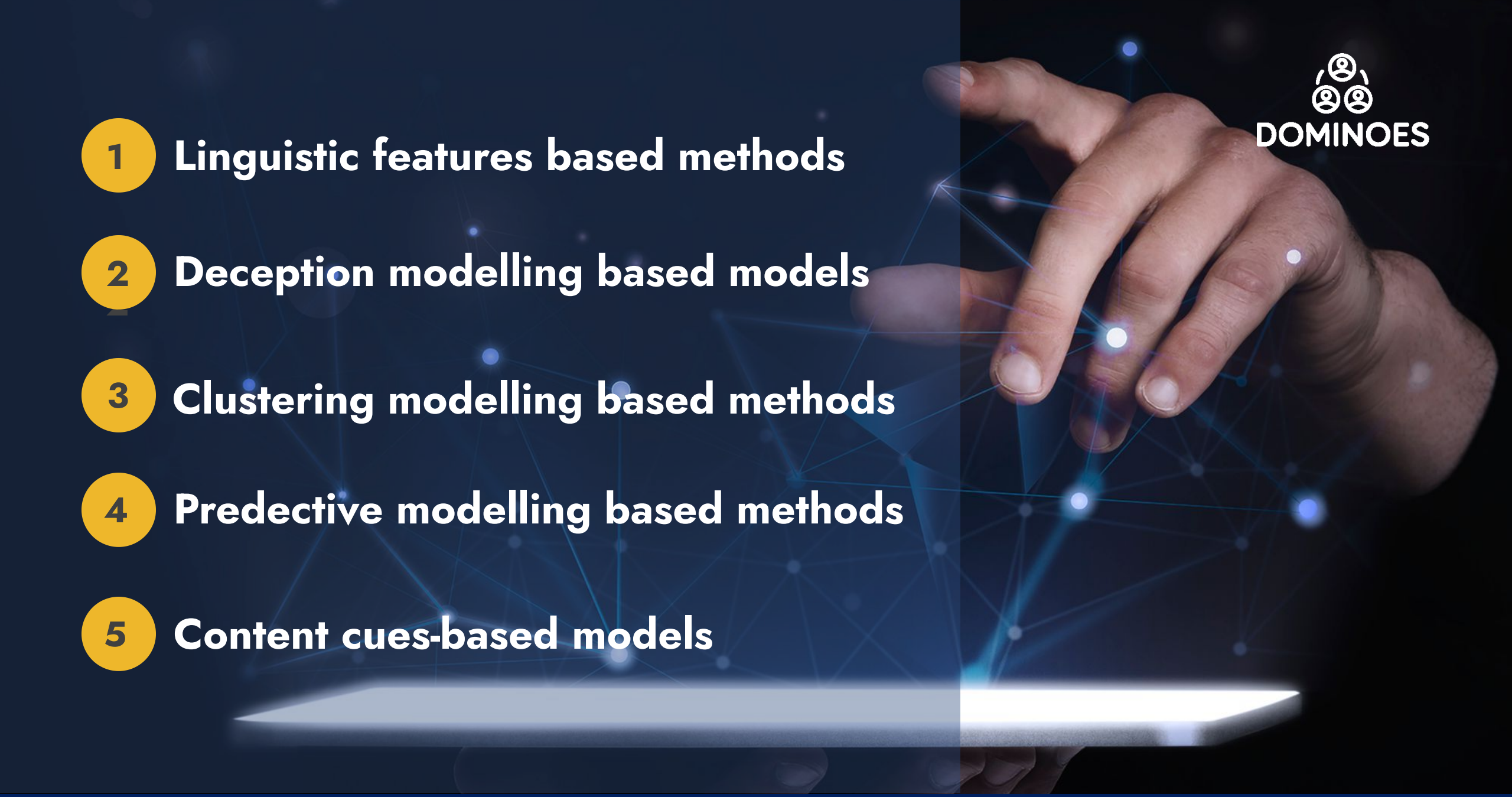
Image source: Choraś, Demestichas, Giełczyk, & Herrero, 2021, p. 2)

Co-funded by the European Union
2021-1-RO01-KA220-HED-000031158
Digital cOMpetences INformatiOn EcoSystem

DOMINOES

A survey conducted by a team of researchers from the University of Albany on the solutions to address fake news detection through text-analysis and mainstream fake news datasets showed that the state-of-the-art approaches for combating the effects of disinformation through detection can be clustered into five main categories, depending on the methods they use
(Parikh & Atrey, 2018, 438)

Co-funded by the European Union
2021-1-RO01-KA220-HED-000031158
Digital cOMpetences INformatiOn EcoSystem



Co-funded by the European Union
2021-1-RO01-KA220-HED-000031158
Digital cOMpetences INformatiOn EcoSystem

Extract key linguistic features from fake news and false information, as follows (Parikh & Atrey, 2018, 438):

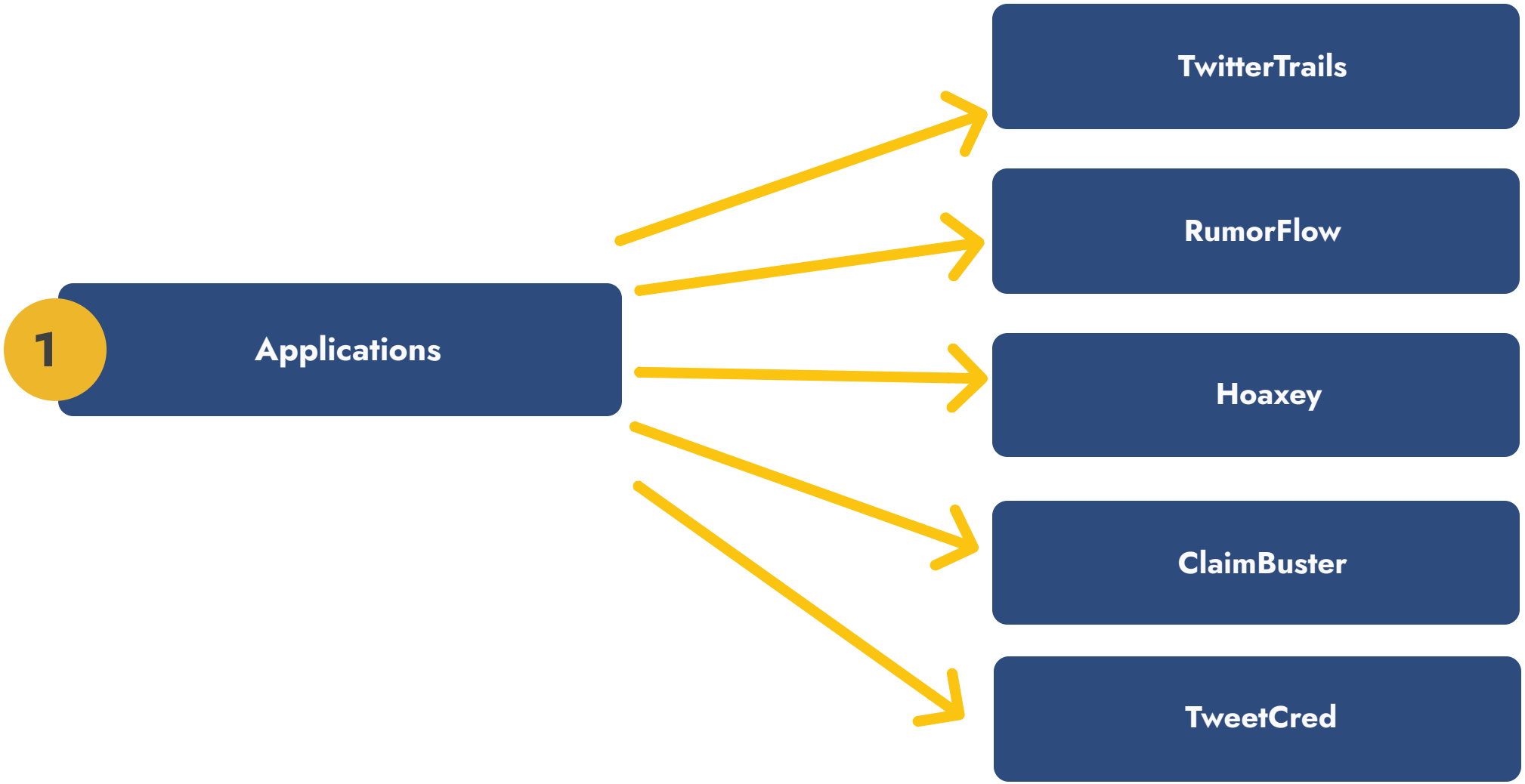
1. **Ngrams:** extracted from the matrix of words in a certain story, stored as TFIDF (Term Frequency Inverse Document Frequency) values for information retrieval
2. **Punctuation:** helps the algorithms for fake news detection to make the difference between false and truthful texts, by collecting eleven types of punctuation, implemented through this detection
3. **Asycho-linguistic features:** use the LIWC lexicon (Linguistic Inquiry and Word Count) to pick out appropriate proportions of words, allowing the system to determine the tone of the language, statistics of the text, part-of-speech category (Pérez-Rosas, Kleinberg, Lefevre, & Mihalcea, 2018, 5)
4. **Readability:** includes the extraction of content features such as the number of characters, complex words, long words, number of syllables, word types, and number of paragraphs
5. **Syntax:** focuses on extracting a set of features based on CFG (context-free grammar), which are heavily dependent on lexicalized production rules combined with their parent and grandparent nodes

Convert texts into a set of rhetorical relations connected in a hierarchical tree and identify the results of rhetorical structure relations by employing two theoretical techniques:

1. **Rhetorical Structure Theory (RST):** focuses on capturing the logic of a story in terms of functional relations created amongst different meaningful text units, describing, at the same time, a hierarchical structure for each story (Mann & Thompson, 1988).
2. **Vector Space Modeling (VSM):** used to identify the rhetorical structure relations in the sets resulted after the application of RST, helps at interpreting every news text as vectors in high dimensional space, aspect that requires for the extracted text to be modeled in an appropriate manner for the application of various computational algorithms (Rubin, Conroy, & Chen, 2015).

Inspiring practices, projects, interventions in the field

DOMINOES



DOMINOES
digital resilience to disinformation

Digital cOMpetences INformatiOn EcoSystem

ANIMV DARE. LEARN. INNOVATE.
Universidad Rey Juan Carlos
L-Università ta' Malta
NEW STRATEGY CENTER

Co-funded by the European Union

2021-1-RO01-KA220-HED-000031158

Author of contents: **Ruxandra Buluc (MVNIA)**

Audiovisual and multimedia production: CIBERIMAGINARIO

DOMINOES
digital resilience to disinformation

Tech-driven solutions and emerging technologies to counter disinformation

4.2.2

doi.org/10.5281/zenodo.10064665

Co-funded by the European Union

2021-1-RO01-KA220-HED-000031158

ANIMV DARE. LEARN. INNOVATE.
Universidad Rey Juan Carlos
L-Università ta' Malta
NEW STRATEGY CENTER

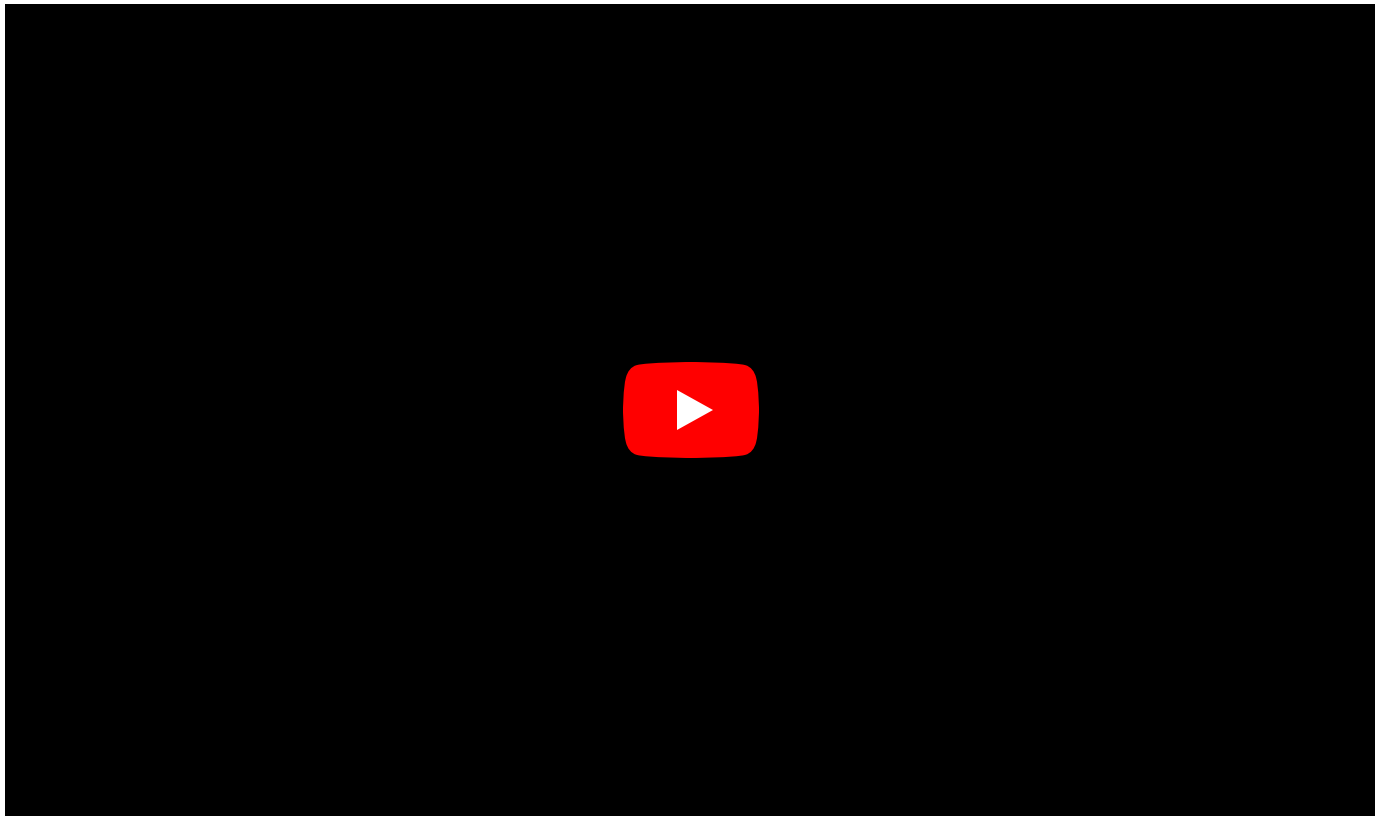
Co-funded by the European Union
2021-1-RO01-KA220-HED-000031158
Digital cOMpetences INformatiOn EcoSystem

BAD NEWS

From fake news to chaos! How bad are you? Get as many followers as you can.

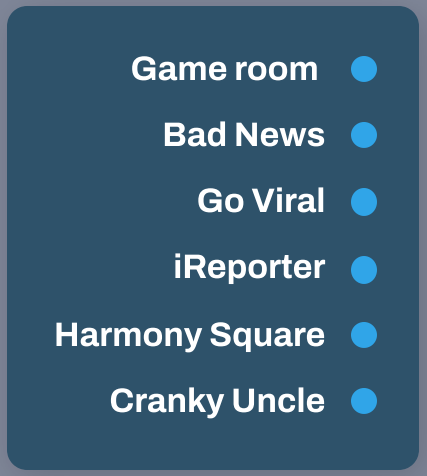
- This game puts people in the position of a person who produces and disseminates fake news.
- The player´s aim in the game is to obtain as many followers and shares as possible through creating and sharing different fake news items.
- Subjects exposed to the game tended to increase their willingness to engage in critical thinking and to take time to evaluate the accuracy of headlines that researchers exposed them to.

Source: Get Bad News



Source: Get Bad News and Go Viral. Games Tackling Disinformation // Games for Impact 2020

Panelists:
Jon Roozenbeek (DROG, Department of Psychology at the University of Cambridge)
Ewa Modrzejewska (University of Warsaw)

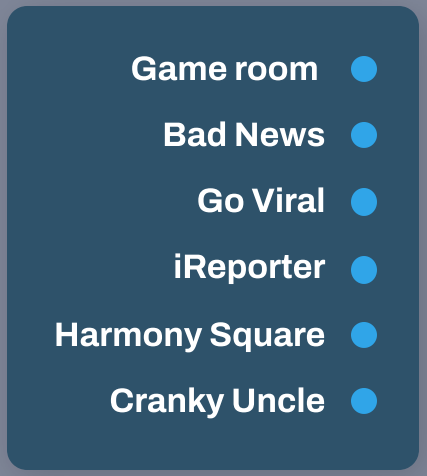

Game room
Bad News
Go Viral
iReporter
Harmony Square
Cranky Uncle

GO VIRAL!

A 5-MINUTE GAME THAT HELPS PROTECT YOU AGAINST COVID-19 MISINFORMATION

- An online game based on misinformation spread during the COVID pandemic.
- In this particular game, players were asked to imagine that they controlled a social media profile and were asked to obtain as many likes and "credibility points", by sharing posts based on different argumentative fallacies.

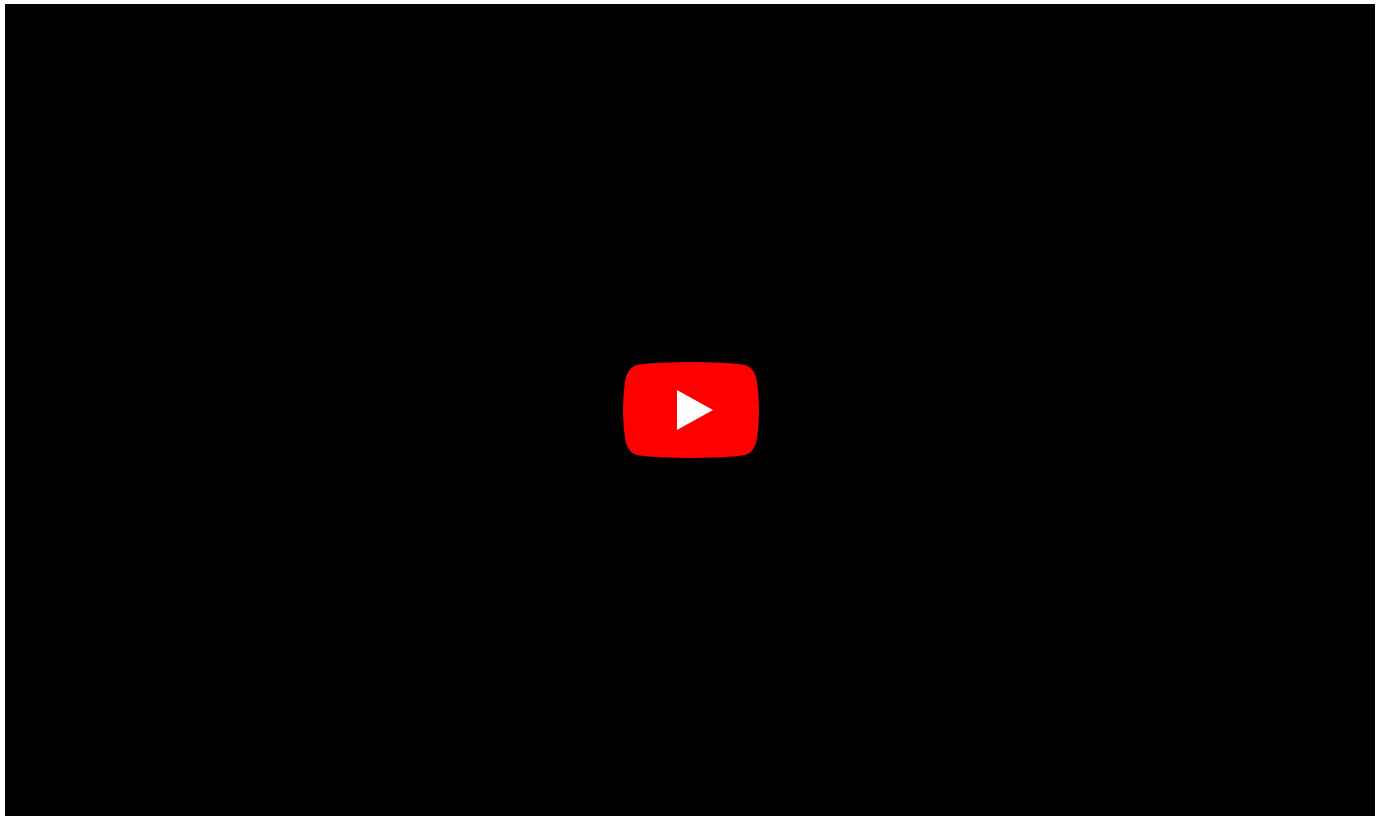
Source: Cambridge University


Game room
Bad News
Go Viral
iReporter
Harmony Square
Cranky Uncle

- A game developed in order to get the users familiarized with the principles of disseminating news in the online environment, as well as with the impact the way in which each piece of news is disseminated can have on the public opinion.
- Users will "play the role of a social media journalist who is faced with a major breaking story" (Cellan-Jones, 2018).
- The game includes elements that involve chatting, having video calls with other journalists and so on.
- The players need to make decisions with tradeoffs, for example speed and accuracy, whether to publish a story as quickly as possible or to confirm first with a reliable source.
- The game educates the players more on the side of how good journalism is and what to consider before sharing a story.

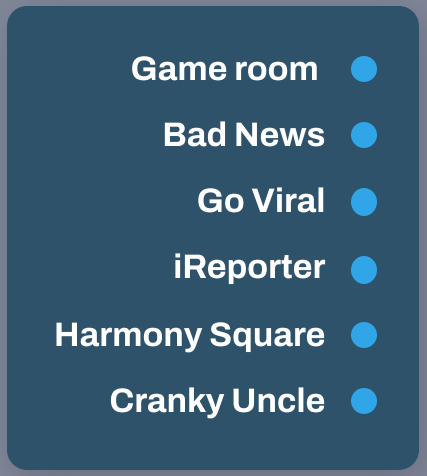
Source: BBC

Co-funded by the European Union
2021-1-RO01-KA220-HED-000031158



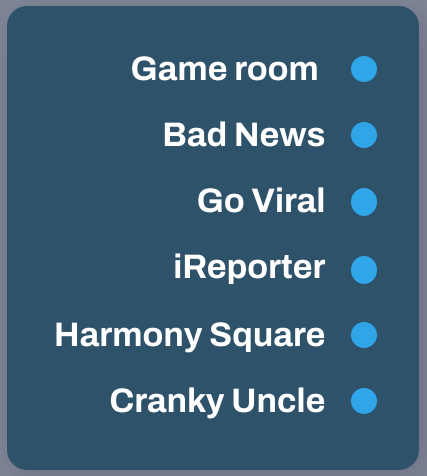
Source: Get Bad News and Go Viral. Games Tackling Disinformation // Games for Impact 2020

Panelists:
Jon Roozenbeek (DROG, Department of Psychology at the University of Cambridge)
Ewa Modrzejewska (University of Warsaw)


Game room
Bad News
Go Viral
iReporter
Harmony Square
Cranky Uncle

- A game that places the player in the shoes of a candidate in elections who can be elected by creating political polarization. This game has also been shown to determine players to rate fake news as less accurate.
- As described on its website, “the goal of the game is to expose the tactics and manipulation techniques required in order to mislead people, build up a following, or exploit societal tensions for political purposes".
- Harmony Square works as a psychological “vaccine” against disinformation: playing it builds cognitive resistance against common forms of manipulation that the user may encounter online” (Harmony Square, 2021).

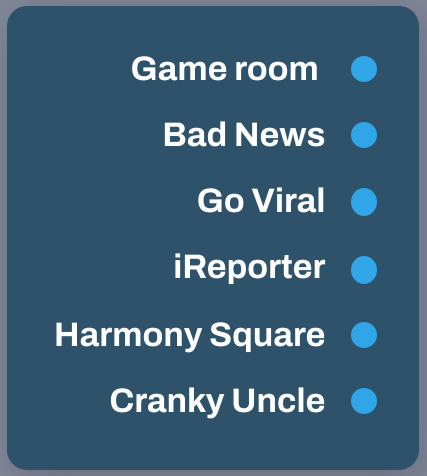
Source: Games for Change


Game room
Bad News
Go Viral
iReporter
Harmony Square
Cranky Uncle

Cranky Uncle

- This game explains different logical fallacies used by climate change deniers in the form of a cranky old man who issues pronouncements on the non-existence or alternative causes of climate change.
- Several experiments by researchers showed how playing the game increased the ability to identify and the knowledge of how to use logical fallacies by students in different study programs (Compton, van der Linden, Cook & Basol, 2021).

Source: https://crankyuncle.com/game/


Game room
Bad News
Go Viral
iReporter
Harmony Square
Cranky Uncle

DOMINOES
digital resilience to disinformation

Digital cOMpetences INformatiOn EcoSystem

ANIMV DARE. LEARN. INNOVATE.
Universidad Rey Juan Carlos
L-Università ta' Malta
NEW STRATEGY CENTER

Co-funded by the European Union

2021-1-RO01-KA220-HED-000031158

Author of contents: **Ruxandra Buluc (MVNIA)**

Audiovisual and multimedia production: CIBERIMAGINARIO

DOMINOES
digital resilience to disinformation

Limits of Technological Tools, Best Practices

4.2.4

doi.org/10.5281/zenodo.10064675

Co-funded by the European Union

2021-1-RO01-KA220-HED-000031158

ANIMV DARE. LEARN. INNOVATE.
Universidad Rey Juan Carlos
L-Università ta' Malta
NEW STRATEGY CENTER


DOMINOES
Machine Learning
in Disinformation
06:49

Machine Learning in Disinformation

Limitations of NLP Tools

Social Media Bot-Detecting Tools

Human Rights Impact and the Way Forward

Conclusion

References

- Rauchfleisch, A., Kaiser, J., (2020). The False positive problem of automatic bot detection in social science research. PLoS ONE. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0241045
- Raso, F., Hilligoss, H., Krishnamurthy, V. et al. (2018, September 25). Artificial Intelligence & Human Rights: Opportunities & Risks. Berkman Klein Center for Internet & Society, Harvard University.
- Florida Law Review. https://scholarship.law.ufl.edu/cgi/viewcontent.cgi?article=1348&context=flr
- Osoba O., Wesler IV W. (2017). An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence. RAND Corporation. https://www.rand.org/content/dam/rand/pubs/research_reports/RR1700/RR1744/RAND_RR1744.pdf
- Duarte, N., Llanso, E., & Loup., A. (2018). Mixed Messages? The Limits of Automated Social Media Content Analysis. The 2018 Conference on Fairness, Accountability, and Transparency. https://cdt.org/wp-content/uploads/2017/12/FAT-conference-draft-2018.pdf
- Linardatos, P., Papastefanopoulos, V., Kotsiantis, S. (2020). Explainable AI: A Review of Machine Learning Interpretability Methods. Entropy. https://dx.doi.org/10.3390/e23010018
- European Union Agency for Fundamental Rights. (2022). Bias in Algorithms – Artificial
- Cresci, S. (2020). A decade of social bot detection. Communications of the ACM. doi:10.1145/3409116
- Hirschberg, J., Manning, C. (2016, May 12). Advances in natural language processing. Science. https://cs224d.stanford.edu/papers/advances.pdf
- Leslie, D. Burr, C. et al. (2021, June). ARTIFICIAL INTELLIGENCE, HUMAN RIGHTS, DEMOCRACY, AND THE RULE OF LAW. Council of Europe. https://rm.coe.int/primer-en-new-cover-pages-coe-english-compressed-2754-7186-0228-v-1/1680a2fd4a
- Yang, K., Ferrera, E. Menczer, F. (2022, August 20). Botometer 101: social bot practicum for computational social scientists. Journal of Computational Social Science. https://link.springer.com/article/10.1007/s42001-022-00177-5

DOMINOES
digital resilience to disinformation

Digital cOMpetences INformatiOn EcoSystem

ANIMV DARE. LEARN. INNOVATE.
Universidad Rey Juan Carlos
L-Università ta' Malta
NEW STRATEGY CENTER

Co-funded by the European Union

2021-1-RO01-KA220-HED-000031158

Author of contents: **Alexandra Anghel (MVNIA)**

Audiovisual and multimedia production: CIBERIMAGINARIO

DOMINOES
digital resilience to disinformation

4.3

Planning, design, and implementation of counter-narratives and positive content

Cristina Ivan | ANIMV

doi.org/10.5281/zenodo.10064625

Co-funded by the European Union

2021-1-RO01-KA220-HED-000031158

ANIMV DARE. LEARN. INNOVATE. Universidad Rey Juan Carlos L-Università ta' Malta NEW STRATEGY CENTER

DOMINOES
digital resilience to disinformation

Erasmus+

PLANNING, DESIGN, AND IMPLEMENTATION OF COUNTER-NARRATIVES AND POSITIVE CONTENT

This section is dedicated to the topic Planning, design, and implementation of counter-narratives and positive content.

UNIT OBJECTIVES

- Planning of counter-narratives and positive content
- Design of counter-narratives and positive content
- Implementation of counter-narratives and positive content

What is a narrative?

1 2 3 4 5 6

What is narrative

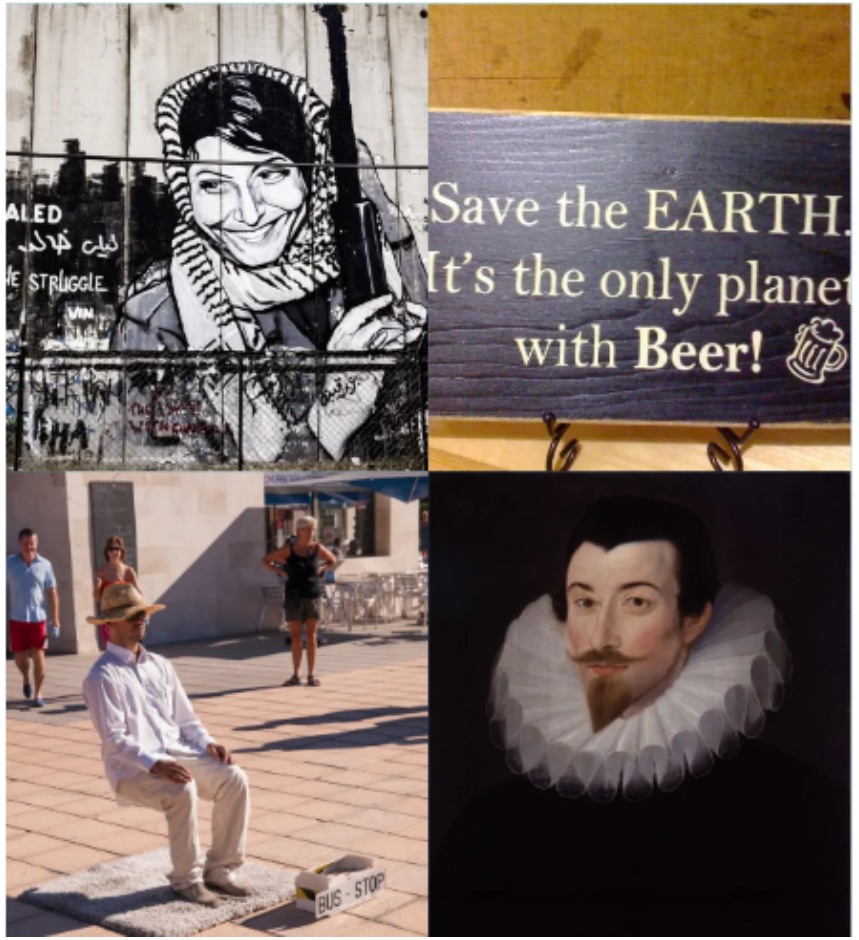
Narratives have been generally studied within the field of literary studies. For obvious reasons, they have been linked to art and literary works and associated with the cultural heritage of a society. A shift occurred in this perspective with the poststructuralists that, for the first time, highlighted the universal character of narratives which are articulated in a multitude of vehicles such as: spoken or written discourse, pictures, movies, gestures, graffiti, art and street performance etc. In this larger context, Roland Barthes insisted on narratives' "infinite variety of forms…present at all times, in all places and all societies" and on their universal character, as international, trans-historical and transcultural, started "with the very history of mankind" (Barthes and Duisit, 1975).

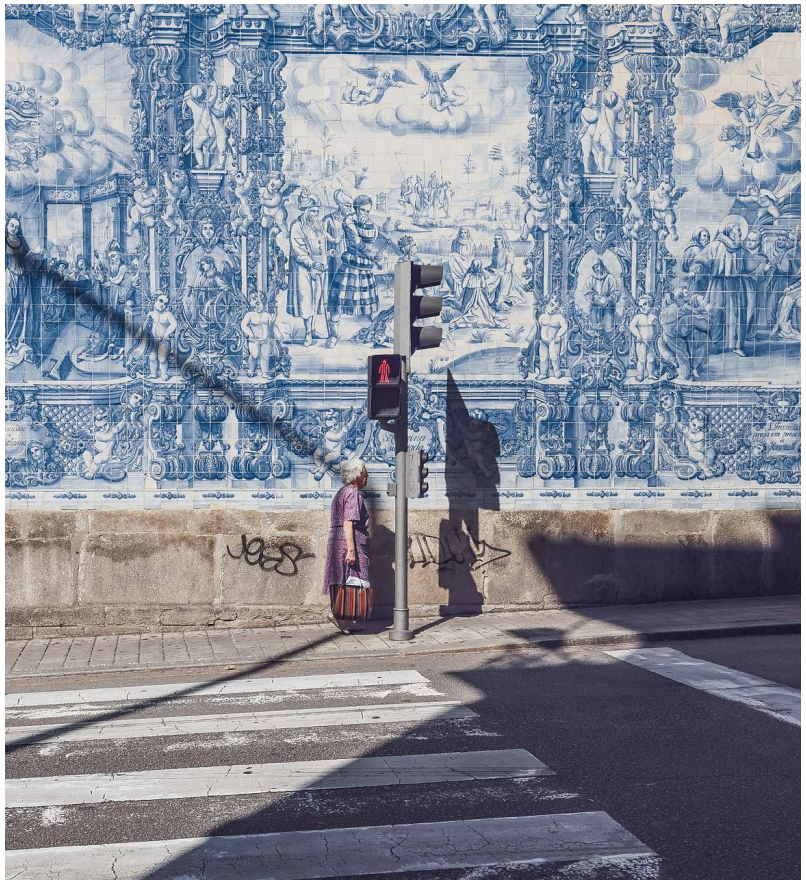
Their universal character and fundamental function of articulating reality makes narratives in this broader social and cultural sense the main vehicle of identity formation and dissemination. Whether we refer to individual or collective identities, their sense making is inseparably linked to narratives. Hence, narratives' main function is that of representation and they are generally perceived as accurate reflections and expressions of what we see, how we are and what we cherish as individuals and communities. However, with postmodernism and poststructuralism, and especially in the works of Jacques Derrida and Michel Foucault, narratives were also revealed as inextricably linked to power formation and projection, hence acquiring social and political value. In the words of Somers, "it is through narrativity that we come to know, understand, and make sense of the social world, and it is through narratives and narrativity that we constitute our social identities" (Somers n.d., 606).

DOMINOES
digital resilience to disinformation

Erasmus+

Where to find narratives?





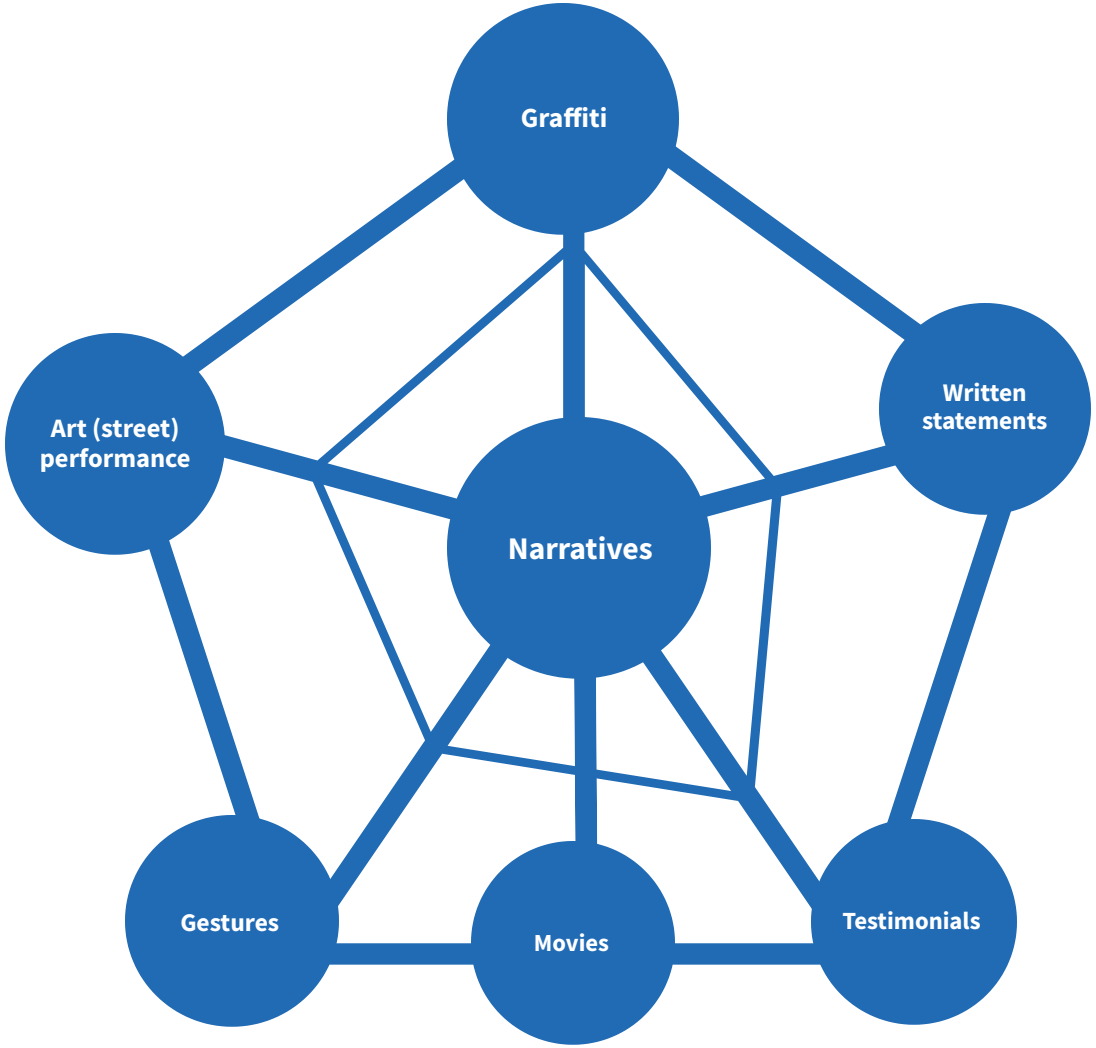
Where to find narratives?

Stories or narratives have a universal character! They are articulated in a multitude of vehicles such as: spoken or written discourse, pictures, movies, gestures, graffiti, art and street performance.

DOMINOES
digital resilience to disinformation

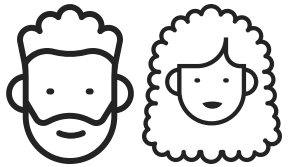
Erasmus+

Where to find narratives?



DOMINOES
digital resilience to disinformation

Audience & content creators



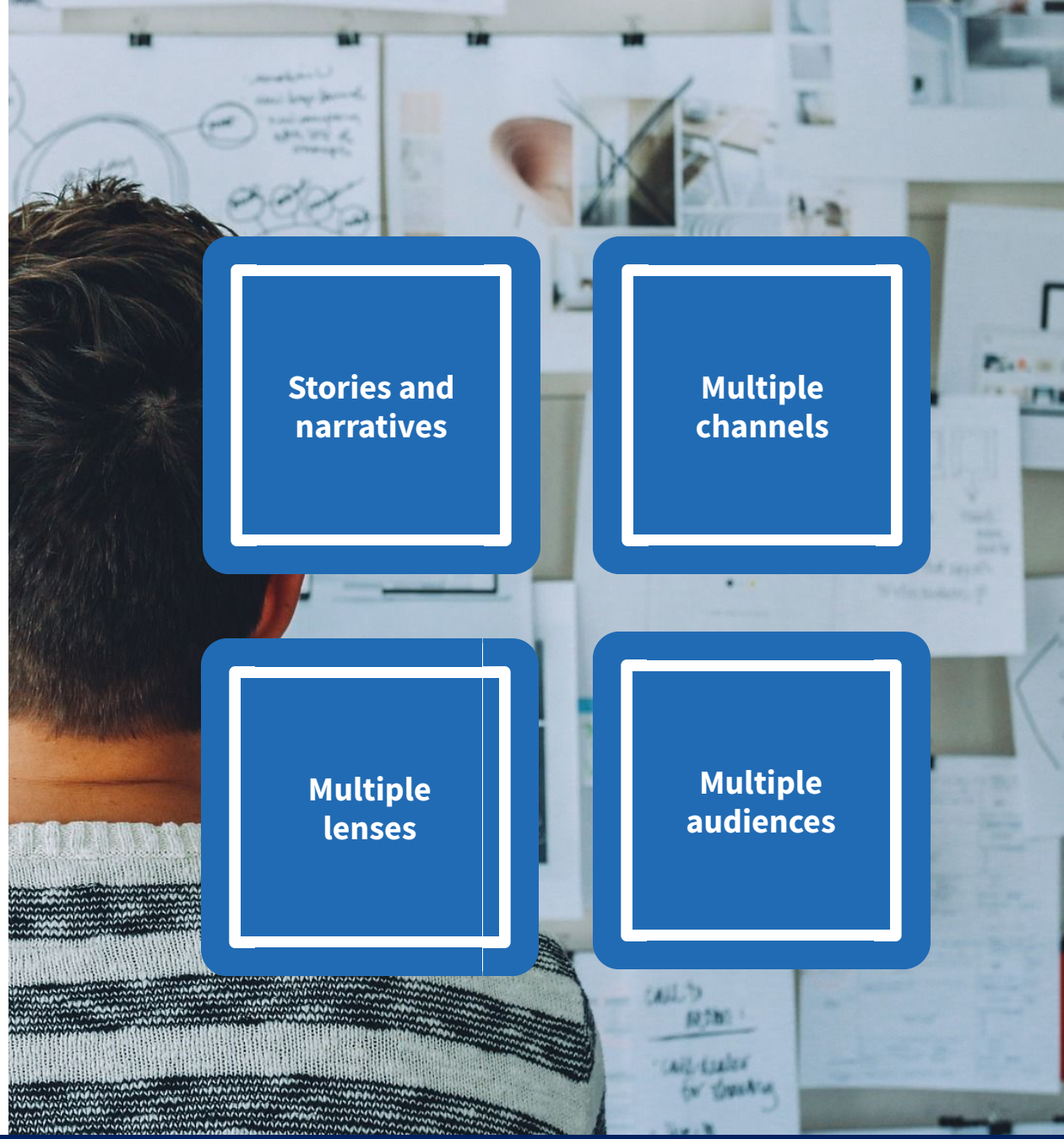
General public

Experts

Teenagers

Granpharents

everyone



DOMINOES
digital resilience to disinformation

Can narratives manipulate perceptions and convictions?

Historical example:

> "The Porgy and Bess musical was used to counteract propaganda of two kinds related to the United States. First, that this country has no real culture, (or) native artists of creative vitality. Second, that the colored people have no opportunity to develop their abilities beyond a slave status."

Porgy and Bess 1

Porgy and Bess 2

Narratives

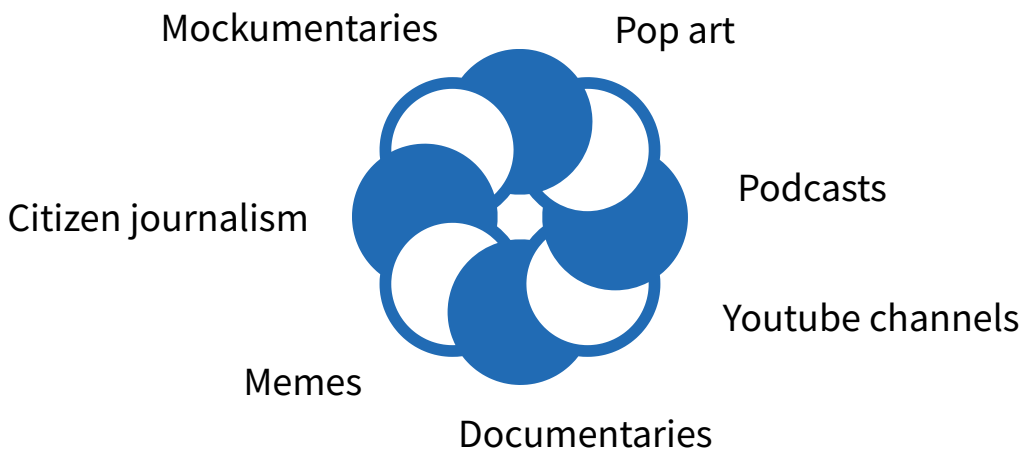
As we have already discussed, narratives are sometimes used to influence and manipulate. All state propaganda, for instance, is linked to conveying stories of legitimate power. Therefore, from government sites to media state outlets, both strategic communication and propaganda channels will tend to build stories of legitimacy, efficiency and purpose. Some of the most successful examples in history come from culture. David Monod, for instance, discusses the case of the Porgy and Bess musical under the interpretation of the Gershwin Opera, that was used by the American State Department as a tool of propaganda in Europe in the 1950's.

As Monod observes, the declared objective of the State Department announced 3 months in advance of the tour was "to counteract propaganda of two kinds related to the United States. First, that this country has no real culture, (or) native artists of creative vitality. Second, that the colored people have no opportunity to develop their abilities beyond a slave status." What this historical example hints at is that narratives circulated via fine artistic productions, beyond their esthetic value, can be instrumentalized in the propagandistic exchange of adversary states and in ways that appeal to the minds of the target country's citizens.

DOMINOES
digital resilience to disinformation

Cultural productions

While the majority of cultural productions do not serve as propaganda tools, their significance can mainstream politically loaded statements and can attach a particular meaning to an entire epoch.



What this historical example hints at is that narratives circulated via fine artistic productions, beyond their esthetic value, can be instrumentalized in the propagandistic exchange of adversary states and in ways that appeal to the minds of the target country's citizens.

Additional info:

Yet as cultural historians often argue, cultural products will always bear a potential for subversion, their complexity making them ambivalent, unexpected interpretations assigned by different readers occurring at any time. A special part in creating powerful narratives has to be acknowledged especially in the case of performative arts, those that in the framing imposed by the director, can channel understanding of narratives in less ambivalent terms and in favor of a preferred interpretation - that being the case with music videos, theater plays, and especially movies.

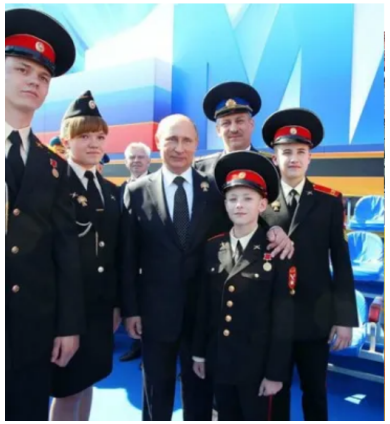
The “war on terror” master narrative created post 9/11 is another globally known example of how understanding of major historical events can be shaped not only by political discourse and strategic communication, but also in cultural productions reflecting the events - books, movies, TV series, documentaries, all in a diverse plethora of explanations that go beyond the event and into the making of history. And while the majority of cultural productions do not serve as propaganda tools, their significance can mainstream politically loaded statements and can attach a particular meaning to an entire epoch.

Case Study

A compared analysis of Monica Ali's novel Brick Lane, published in 2003 and the British drama film directed by Sarah Gavron launched in 2007, reflects the tension created between different viewers of the New York Twin Tower fall as a result of the 9/11 terrorist attack. Both underline the interplay of love and death, hope and revenge, whom are given precedent distinctively by the different characters and lead readers and viewers into searching for a more deep insight of jihadism and the underlying social and psychological grievances that trigger it. They will serve as an ambivalent narrative of terrorism as an extreme and violent response to social injustice and discrimination, while both works of art turn the spotlight on human agency and love as providing alternative personal pathways.

Dozens of other film productions, from Zero Dark Thirty (2012) directed by Katryn Bigelow to 9/11: One Day in America (2021) individualize the story while taking a narrative that could easily shape perceptions of viewers worldwide. While the most visible and easily recognizable, state media outlets and cultural productions are not the most powerful channels instrumentalized to create preferred significances to events. Narratives have in history been used also quite extensively in gray and black propaganda operations. And if high art can play a significant part in the soft power apparatus, one should not overlook the role played by popular art, by memes, podcasts, YouTube video channels, documentaries and mockumentaries, citizen journalism and any other form of collective, grass root formation of narratives as stories we tell each other about ourselves as individuals and communities.

Implementing counternarratives



David Attenborough(style) narrates about Russian Tank.

Please have a look at these iconic images of the Russian war propaganda with parade tanks marching in all their glory in front of Kremlin and think of the functions such images have in creating a narrative of indomitable power projection.

Think of their effect on the Ukrainian target audience about to be invaded or during the initial days of the invasion, in which Russian tanks marched at full speed towards Kiev. And then, think of effective ways to create counter-narratives, build community cohesion and trust in own state capacity, building trust and sympathy with the invaded population at EU level etc.

The stamp collection image of a tractor towing a tank presented in the right hand side bottom of the slide illustrates the powerful effect of such a counternarrative, started from a real incident recorded on a Youtube film, of a farmer towing with a tractor an apparently left behind Russian tank. As you will see in the exercises to come, this has sprung an entire series of pop art performances and images, culminating with the mockumentary accessible on the link: https://web.facebook.com/watch/?v=333665958904295

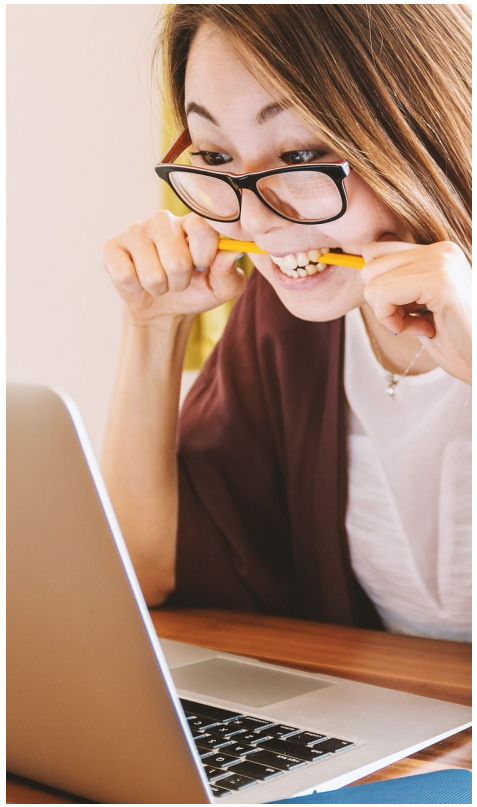
Remember

- Cultural productions that emerge out of the participatory digital culture are some of the most powerful ways in which counter-narratives and positive content can be transmitted

WHY

Source: Arcos, Intelligence and awareness, 2023

When planning counter narratives, please REMEMBER:



Cultural productions that emerge out of the participatory digital culture are some of the most powerful ways in which counter-narratives and positive content can be transmitted.

Here, creation is often anonymized, while co-production and non-attribution are widely shared behaviors. At the same time transgressive and empowering, these cultural productions often make meaning as part of counter-narratives emerged through public participation. Positive co-created content gains force from the amount of user interaction generated and the real time meaning making process they foster and encourage with digital users. Most rely on a media account of real life events only to then transgress into the symbolic regime and start generate meaning(s) by the engagement of the audience.

DOMINOES
digital resilience to disinformation

Exercise

Planning, design, and implementation of counter-narratives and positive content

Erasmus+
2021-1-RO01-KA220-HED-000031158

ANIMV DARE. LEARN. INNOVATE.
Universidad Rey Juan Carlos
L-Università ta' Malta
NEW STRATEGY CENTER

Exersice: What do we want to counter?

Russian propaganda machine.

Click on the image to zoom in.

Graphics description: Russian propaganda machine

Narratives denying facts

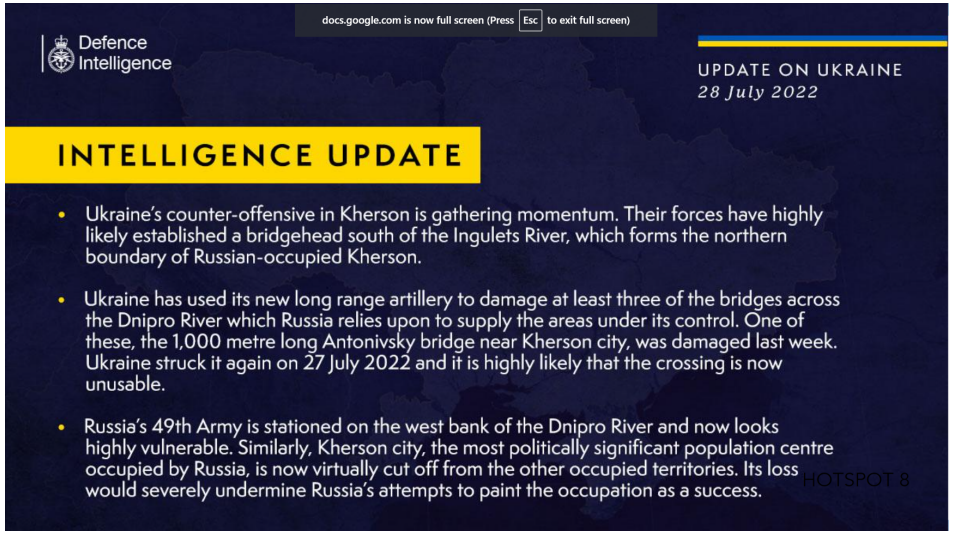
A sequence of narratives attempted to persuade the audience by denying facts and blaming the opponent for their own deeds: Russia did not attack and does not wage war, the Ukrainian population is decimated by its own government forces, attacks are inflicted by third parties and imagined enemies aspire to conquer Ukrainian territories. Denying their own crimes and justifying war through the existence of imaginary enemies was a preferred narrative used to confuse, detour attention or simply create as much as possible plausible deniability.

Pattern of fake narratives

A timeline of the first 100 days of the Russian war against Ukraine shows a telling pattern of fake narratives. At the beginning, mid-February 2022, the Russian propaganda machine advanced the idea that there is a Ukrainian crisis caused by the disregard of the West towards the “neo-Nazi crimes” of the Ukrainian government associated forces.

Counter narratives to Russian propaganda machine

- Public intelligence briefings
  - INTELL.gov - The public's daily brief
  - The Cipher Brief
- Alerts to the press
  - Twitter - (1/5) Ukraine's counter-offensive in Kherson is gathering momentum. Their forces have highly likely established a bridgehead south of the Ingulets River, which forms the northern boundary of Russian-occupied Kherson.



Click on the image to zoom in.

Counternarratives to russian propaganda machine

Public intelligence briefings

Public intelligence briefings were used to reveal the invasion before it actually occurred and allowed for the narrative to be clear, casting Putin and Russia as the aggressor and Ukraine as a victim. This helped form unified support for Ukraine.

Alerts to the press

Alerts were constantly released to the press, like the ones that followed the attack of the Moskva ship and its subsequent sinking: “The 510-crew missile cruiser was a symbol of Russia's military power, leading its naval assault on Ukraine. Kyiv says its missiles hit the warship. The United States says it also believes it was hit by Ukrainian missiles.” Source: BBC

DOMINOES
digital resilience to disinformation

Encouraging the audience to take ownership and engage in co-creation of collective narratives

1 2 3

Co-create humor, artistic representations, memes

As a result of public engagement, people started to co-create humor, artistic representations, memes etc. For example the one described by **Euractiv** as a collage that illustrates what is perceived as a victory of David against Goliath" or this tweeted by **SPRAVDI-Stratcom Centre** showing a picture of a Russian ship showing the text Operation Z and picture of an ocean showing the text Operation Ctrl + Z.

Facilitate dissemination of collective cultural productions through the state sponsored dissemination of meaningful art: the Ukrainian postage stamp collection

The Ukrainian state also issued a series of postage stamps reflecting powerful popular narratives, such as **the blue tractor narrative** illustrated in the previous section of the MOOC. The stamp design evokes milestones in defending the country as an alternative powerful way to advance counter-narratives at global scale.

Video game creation, street art performances, graffiti etc - for example the Bansky series on Ukraine

Other collective narrative expressions included fans of the blue tractor symbol creating video games on the topic, mockumentaries, street art performances and art exhibitions in central European capitals, graffiti, such as **the famous Banksy series in Ukraine** etc.

DOMINOES
digital resilience to disinformation

Exercise

Planning, design, and implementation of counter-narratives and positive content

Erasmus+
2021-1-RO01-KA220-HED-000031158

ANIMV DARE. LEARN. INNOVATE.
Universidad Rey Juan Carlos
L-Università ta' Malta
NEW STRATEGY CENTER

Exersice: What to do?

Match the following propaganda narratives with suggested ideas of counternarratives that you believe would be effective in correcting perceptions.
(Drag an Drop the titles to allocate them with their respectives paragraphs)

Clear the fog by fact checking invoked person's identity

Get out of the biolabs conspiracy area by making the audience aware

Fact check the original source, give voice to real experts

Story frequently referred to by pro-Kremlin sources as "proof" of Ukrainian involvement:
On 19 July 2014, two days after the MH17 disaster, a Twitter account belonging to a certain "Carlos, a Spanish dispatcher" working for air traffic control at Kyiv Airport, claimed that two Ukrainian fighter jets had downed the aircraft. Then, RT Spanish carried out an interview with an individual claiming to be Carlos.

US biolabs in Ukraine are aimed at reducing Russia's gene pool. It has long been known what these Pentagon biolabs in Ukraine were doing. They grew pathogenic microbes to infect humans, everything was done with the Russian gene pool in mind. We know that about 30 Pentagon biolaboratories were dispersed in different cities on Ukrainian territory with different specializations, but with one goal: to cheaply and angrily destroy the central enemy and its allies.

Studies have shown that China is more democratic than the United States.
This is evidenced by the results of a study by the Alliance of Democracies and Dalia Research, which was conducted in 53 countries from April to June 2020 among 124,000 people, reports News-Front.
(…) It should also be noted that in China, 73% of respondents called their country democratic, while in the United States only 49% got it. The democratic deficit in China was 11%, down from 20% last year. In the United States, the deficit rate was set at 24%, just 2% below the 2019 level.
In Russia, according to Western analysts, the deficit was 27%, down 5 points from last year. Thus, the difference between the Russian Federation and the United States is only 3%. It should be noted that Freedom House nevertheless classifies Russia as "not free", unlike the United States.

DOMINOES
digital resilience to disinformation

Bibliography

and useful resources

2021-1-RO01-KA220-HED-000031158

American foreign relations . n.d. "Propaganda - Types of propaganda." https://www.americanforeignrelations.com/O-W/Propaganda-Types-of-propaganda.html.

Aro, Jessika. (2016). "The cyber-space war: propaganda and trolling as warfare tools." European view 121-132. doi:10.1007/s12290-016-0395-5.

Barthes, Roland, and Lionel Duisit. (1975). "An Introduction to the Structural Analysis of Narrative" An Introduction to the Structural Analysis of Narrative (The Johns Hopkins University Press) 237-272. http://www.jstor.org/stable/468419.

Baumann, Mario. (2020). "Propaganda Fights' and 'Disinformation Campaigns': the discourse on information warfare in Russia-West relations." Contemporary Politics (Routledge) 1-20. doi:https://doi.org/10.1080/13569775.2020.1728612.

Best, Shivali. (2017). "The spread of fake news on Facebook and Twitter is made worse by social network algorithms." Mail Online, iunie 20. http://www.dailymail.co.uk/sciencetech/article-4621094/Are-Facebook-Twitter-ENCOURAGING-fake-news.html.

Bonino, Silvia, Elena Cattelino, and Silvia Ciairano. (2003). Adolescents and Risk Behaviour, Functions and Portective Factors. Torino: Springer.

Bradshaw, Samatha, and Philip P. Howard. (2018). "Why does Junk News Spread so Quickly across Social Media? Algorythms, Advertising, and Exposure in Public Life." http://comprop.oii.ox.ac.uk/research/working-papers/why-does-junk-news-spread-so-quickly-across-social-media/.

Bradsma, Bart. n.d. Inside Polarisation . https://insidepolarisation.nl/en/.

Candaele, Kelly. (2020). "Coronavirus is a political problem, not just a health problem. Remember that when you vote." The Guradian, March. https://www.theguardian.com/commentisfree/2020/mar/19/coronavirus-political-problem-health-voting-elections.

Chekinov, S.G., and S.A. Bogdanov. n.d. "The Nature and Content of a New-Generation War." Military Thought. http://www.eastviewpress.com/Files/MT_FROM%20THE%20CURRENT%20ISSUE_No.4_2013.pdf.

Culloty, Eileen, and Jane Suiter. (2021). Disinformation and Manipulation in Digital Media. Routledge .

"Disinformation: how to recognise and tackle Covid-19 myths ." News, European Parliament . 30 March 2020. https://www.europarl.europa.eu/news/en/headlines/society/20200326STO75917/disinformation-how-to-recognise-and-tackle-covid-19-myths.

Easter, David. (2010). "British Intelligence and Propaganda during the 'Confrontation', 1963-1966." Intelligence and National Security 1-21.

Edelman's Trust Barometer, Trust Inequality . n.d. "Edelman." http://edelman.edelman1.netdna-cdn.com/assets/uploads/2016/01/2016-Edelman-Trust-Barometer-Global-_-Mounting-Trust-Inequality.pdf.

Edward, Herman, and Noam Chomsky. n.d. "A Propaganda Model." In Manufacturing Consent, by Herman Edward and Noam Chomsky.

Edward, Lucas, and Pter Pomeranzev. (2016). Winning the Information War. Center for European Policy Analysis .

"EEAS SPECIAL REPORT UPDATE: Short Assessment of Narratives and Disinformation Around the COVID-19 Pandemic." EU vs Dinsinfo. April 01. https://euvsdisinfo.eu/eeas-special-report-update-short-assessment-of-narratives-and-disinformation-around-the-covid-19-pandemic/.

EU vs. Disinfo. (2016). "Estonia is building a concentration camp for its Russian-speaking citizens. ." EU vs Disinfo. https://euvsdisinfo.eu/report/estonia-is-building-a-concentration-camp-for-its-russian-speaking-citizens/.

Farmy, Ukrainian. n.d. https://ukrainian.itch.io/ukrainian-farmy.

Foresman, Galen A., Peter S. Fosl, and Jamie Carlin Watson. (2017). The Critical Thinking Toolkit. Wiley Blackwell.

Gertrudis-Casado, María-del-Carmen, María-del-Carmen Gálvez-de-la-Cuesta, Juan Romero-Luis, and Manuel Gértrudix Barrio. (2022). "Los serious games como estrategia eficiente para la comunicación científica en la pandemia de la Covid-19." Revista Latina de Comunicacion Social. doi:https://doi.org/10.4185/RLCS-2022-1788.

Global Engagement Center . n.d. Disarming Disinformation: Our Shared Responsibility . https://www.state.gov/disarming-disinformation/.

Gray, Ann. (2003). Research practice for cultural studies. Ethnographic methods and lived cultures. London : Sage Publications .

Grejdeanu, Tamra. (2017). "Propaganda rusă în Moldova. Cum funcționează?" Radio Europa Liberă. aprilie 28. Accessed iulie 30, 2018. https://www.europalibera.org/a/propaganda-rusa-in-moldova/28457231.html.

Guess, A. M., and B. A. Lyons. (2022) "Misinformation, Disinformation, and Online Propaganda. Social Media and Democracy, 10–33. doi:10.1017/9781108890960.003."

Helmus, Baron, Radin, Magnuson, Mendelsohn, Marcellino, Bega, Winkelman. (2018). Russian Social Media Influence. Understanding Russian Porpaganda in Eastern Europe. Rand Corporation.

Hicks-Goldston, C. (2019). The new digital divide: Disinformation and media literacy in the US. Media Literacy and Academic Research, 2(1), 49-60.

Hinchchliffe, Tim. (2020). "Exposing echo chambers to eradicate the plague of propaganda." The Sociable. https://sociable.co/social-media/exposing-echo-chambers-to-eradicate-the-plague-of-propaganda/.

Humprecht, Edda, Frank Esser, and Peter Van Aelst. (2020). "Resilience to Online Disinformation: A Framework for Cross-National Comparative Research." The International Journal of Press/Politics 1-24.

Humprecht, E., Esser, F., Aelst, P. V., Staender, A., & Morosoli, S. (2021). The sharing of disinformation in cross-national comparison: Analyzing patterns of resilience. Information, Communication & Society, 1-21.

Ivan, Cristina. (2013). "Resilience – The X Factor of the Organisational Endurance." In Intelligence in the Knowledge Society, Proceedings of the XVIIIth International Conference, by Irena Chiru Teodoru Stefan, 161-172. ANIMV Publishing House.

Ivan, Cristina, Irena Chiru, and Rubén Arcos. (2021). "A whole of society intelligence approach: critical reassessment of the tools and means used to counter information warfare in the digital age." Intelligence and National Security 495-511. doi:DOI: 10.1080/02684527.2021.1893072.

JamNews. (2017). Fake news in Moldova: fires, droughts, terror attacks and discredited politicians. septembrie 2017. https://jam-news.net/?p=59912.

Jeon, Youngseung, Bogoan Kim, Aiping Xiong, DONGWON LEE, and Kyungsik Han. (2021). "ChamberBreaker: Mitigating the Echo Chamber Effect and Supporting Information Hygiene through a Gamified Inoculation System." Proceedings of the ACIM on Human Computer INteraction. 1-26. doi:https://doi.org/10.1145/3479859.

Lenin, V.V. (2018). "V. I. Lenin, Lessons of the Moscow Uprising." https://www.marxists.org/archive/lenin/.

Lewandowsky, Stephan, Ullrich K.H. Ecker, and John Cook. (2017). "Beyond Misinformation: Understanding and Coping with the "Post-Truth" Era." Edited by Elsevier. Journal of Applied Research in Memory and Cognition.

Lewandowsky, Stephan, Ullrich K.H. Ecker, and John Cook. (2017). "Beyond Misinformation: Understanding and Coping with the “Post-Truth” Era." Journal of Applied Research in Memory and Cognition. doi:Journal of Applied Research in Memory and Cognition.

Maftei, A., & Holman, A. C. (2022). Beliefs in conspiracy theories, intolerance of uncertainty, and moral disengagement during the coronavirus crisis. Ethics & Behavior, 32(1), 1-11.

Martin, L. John. (2010). "Dinsinformation: an instrumentality in the propaganda arsenal." Political Communication 47-64.

McKay, Spencer, and Chris Tenove. (2020). "Disinformation as a Threat to Deliberative Democracy." Political Research Quarterly. doi:10.1177/1065912920938143.

Mediacritica, primul portal de educație mediatică. (2018). Moldova – teren fertil pentru fake news. iulie 11. Accessed iulie 30, 2018. http://mediacritica.md/ro/moldova-teren-fertil-pentru-fake-news/#prettyPhoto.

Monod, David. (2010). "'He is a cripple an' needs my love': Porgy and Bess as Cold War propaganda." Intelligence and National Security 1-14.

"Multiculturalism." Oxford Dictionaries. Accessed 08 5, 2014. http://www.oxforddictionaries.com/definition/english/multicultural.

Nabb Research Center Online Exhibits. n.d. "The Colors of Propaganda." https://libapps.salisbury.edu/nabb-online/exhibits/show/propaganda/what-is-propaganda-/the-colors-of-propaganda.

Nissembaum, Assaf, and Limor Shifman. (2015). "Internet memes as contested cultural capital: The case of 4chan's /b/ board." SagePub Journals 1-19. doi:DOI: 10.1177/1461444815609313.

Nye, J. S. (1990). Soft power. Foreign policy, (80), 153-171.

Paul, Richard, and Linda Elder. (2014). Critical Thinking: Tools for Taking Charge of Your Professional and Personal Life. New Jersey: Pearson Education .

Polygraph.info. (2018). "Polygraph." April 26. Accessed August 30, 2018. https://www.polygraph.info/a/fake-news-in-hungary/29194591.html.

Pressman, D. Elaine, and Cristina Ivan. (2019). Internet Use and Violent Extremism: A Cyber-VERA Risk Assessment Protocol. IGI Global.

Somers, Margaret R. n.d. The narrative constitution of identity:A relational and network approach. https://deepblue.lib.umich.edu/bitstream/handle/2027.42/43649/11186_2004_Article_BF00992905.pdf?sequence=1.

Stoica, Cătălin Augustin, and Radu Umbres. (2020). "Suspicious minds in times of crisis: determinants of Romanians' beliefs in COVID-19 conspiracy theories." European Societies S246-S261. doi:https://doi.org/10.1080/14616696.2020.1823450.

TaskForce, EU East StratCom. (2020). Trends of the Week. Throwing Coronavirus disinfo at the wall to see what sticks . EU StratCom Task Force .

"The 2022 Code of Practice on Disinformation." European Commission . July 2. file:///C:/Users/User/Downloads/2022_Strengthened_Code_of_Practice_Disinformation_TeAETn7bUPXR57PU2FsTqU8rMA_87585.pdf.

n.d. Ukrainian tractor memes compilation. https://www.youtube.com/watch?v=hheLODstezM.

United Nations General Assembly. (2015). "Plan of Action to Prevent Violent Extremism, Report of the Secretary-General." A/70/674. Accessed September 20, 2020. https://www.un.org/en/ga/search/view_doc.asp?symbol=A/70/674.

University of Oxford. (2018). The Computational Propaganda Project. Algorythms, Automation and Digital Politics. http://comprop.oii.ox.ac.uk/.

Weisburd, Andrew, Clint Watts, and JM Berger. (2016). "Trolling for Trump: How Russia Is Trying to Destroy Our Democracy." War on the Rocks. https://warontherocks.com/2016/11/trolling-for-trump-how-russia-is-trying-to-destroy-our-democracy/.

DOMINOES
digital resilience to disinformation

Digital cOMpetences INformatiOn EcoSystem

ANIMV DARE. LEARN. INNOVATE.
Universidad Rey Juan Carlos
L-Università ta' Malta
NEW STRATEGY CENTER

Co-funded by the European Union

2021-1-RO01-KA220-HED-000031158

Audiovisual and multimedia production: CIBERIMAGINARIO

DOMINOES
digital resilience to disinformation

Counter-narratives and positive content

4.3.1

doi.org/10.5281/zenodo.10064696

Co-funded by the European Union

2021-1-RO01-KA220-HED-000031158

ANIMV DARE. LEARN. INNOVATE.

Universidad Rey Juan Carlos

L-Università ta' Malta

NEW STRATEGY CENTER


DOMINOES
digital resilience to disinformation
03:29

DOMINOES
digital resilience to disinformation

Digital cOMpetences INformatiOn EcoSystem

ANIMV DARE. LEARN. INNOVATE.
Universidad Rey Juan Carlos
L-Università ta' Malta
NEW STRATEGY CENTER

Co-funded by the European Union

2021-1-RO01-KA220-HED-000031158

Author of contents: **Cristina Ivan (MVNIA)**

Audiovisual and multimedia production: CIBERIMAGINARIO

DOMINOES
digital resilience to disinformation

Planning and design of counter-narratives and positive content

4.3.2

doi.org/10.5281/zenodo.10064698

Co-funded by the European Union

2021-1-RO01-KA220-HED-000031158

ANIMV DARE. LEARN. INNOVATE.
Universidad Rey Juan Carlos
L-Università ta' Malta
NEW STRATEGY CENTER

DOMINOES

RUSSIAN PROPAGANDA MACHINE

Pattern of fake narratives

A timeline of the first 100 days of the Russian war against Ukraine shows a telling pattern of fake narratives. At the beginning, mid-February 2022, the Russian propaganda machine advanced the idea that there is a Ukrainian crisis caused by the disregard of the West towards the “neo-Nazi crimes” of the Ukrainian government associated forces.

Narratives denying facts

A sequence of narratives attempted to persuade the audience by denying facts and blaming the opponent for their own deeds: Russia did not attack and does not wage war, the Ukrainian population is decimated by its own government forces, attacks are inflicted by third parties and imagined enemies aspire to conquer Ukrainian territories. Denying their own crimes and justifying war through the existence of imaginary enemies was a preferred narrative used to confuse, detour attention or simply create as much as possible plausible deniability.

Co-funded by the European Union
2021-1-RO01-KA220-HED-000031158
Digital cOMpetences INformatiOn EcoSystem

DOMINOES

COUNTERNARRATIVES TO RUSSIAN PROPAGANDA MACHINE

Public intelligence briefings

Public intelligence briefings were used to reveal the invasion before it actually occurred and allowed for the narrative to be clear, casting Putin and Russia as the aggressor and Ukraine as a victim. This helped form unified support for Ukraine.

Alerts to the press

Alerts were constantly released to the press, like the ones that followed the attack of the Moskva ship and its subsequent sinking: "The 510-crew missile cruiser was a symbol of Russia's military power, leading its naval assault on Ukraine. Kyiv says its missiles hit the warship. The United States says it also believes it was hit by Ukrainian missiles."
Source: BBC

Co-funded by the European Union
2021-1-RO01-KA220-HED-000031158
Digital cOMpetences INformatiOn EcoSystem

DOMINOES

ENCOURAGING THE AUDIENCE TO TAKE OWNERSHIP AND ENGAGE IN CO-CREATION OF COLLECTIVE NARRATIVES

Co-create humor, artistic representations, memes

As a result of public engagement, people started to co-create humor, artistic representations, memes etc. For example the one described by Euractiv as a collage that illustrates what is perceived as a victory of David against Goliath" or this tweeted by SPRAVDI — Stratcom Centre showing a picture of a Russian ship showing the text Operation Z and picture of an ocean showing the text Operation Ctrl + Z.

Facilitate dissemination of collective cultural productions through the state sponsored dissemination of meaningful art: the Ukrainian postage stamp collection

The Ukrainian state also issued a series of postage stamps reflecting powerful popular narratives, such as the blue tractor narrative illustrated in the previous section of the MOOC. The stamp design evokes milestones in defending the country as an alternative powerful way to advance counter-narratives at global scale.

Video game creation, street art performances, graffiti etc - for example the Bansky series on Ukraine

Other collective narrative expressions included fans of the blue tractor symbol creating video games on the topic, mockumentaries, street art performances and art exhibitions in central European capitals, graffiti, such as the famous Banksy series in Ukraine, etc.

Co-funded by the European Union
2021-1-RO01-KA220-HED-000031158
Digital cOMpetences INformatiOn EcoSystem

DOMINOES
digital resilience to disinformation

Digital cOMpetences INformatiOn EcoSystem

ANIMV DARE. LEARN. INNOVATE.
Universidad Rey Juan Carlos
L-Università ta' Malta
NEW STRATEGY CENTER

Co-funded by the European Union

2021-1-RO01-KA220-HED-000031158

Author of contents: **Cristina Ivan (MVNIA)**

Audiovisual and multimedia production: CIBERIMAGINARIO

DOMINOES
digital resilience to disinformation

Implementation of counter-narratives and positive content

4.3.4

doi.org/10.5281/zenodo.10064701

Co-funded by the European Union

2021-1-RO01-KA220-HED-000031158

ANIMV DARE. LEARN. INNOVATE.
Universidad Rey Juan Carlos
L-Università ta' Malta
NEW STRATEGY CENTER


DOMINOES
digital resilience to disinformation
Stornaway

DOMINOES
digital resilience to disinformation

Digital cOMpetences INformatiOn EcoSystem

ANIMV DARE. LEARN. INNOVATE.
Universidad Rey Juan Carlos
L-Università ta' Malta
NEW STRATEGY CENTER

Co-funded by the European Union

2021-1-RO01-KA220-HED-000031158

Author of contents: **Cristina Ivan (MVNIA)**

Audiovisual and multimedia production: CIBERIMAGINARIO

DOMINOES
digital resilience to disinformation

Activity of evaluation 4.4

S4. Advanced analytic and responding toolkit

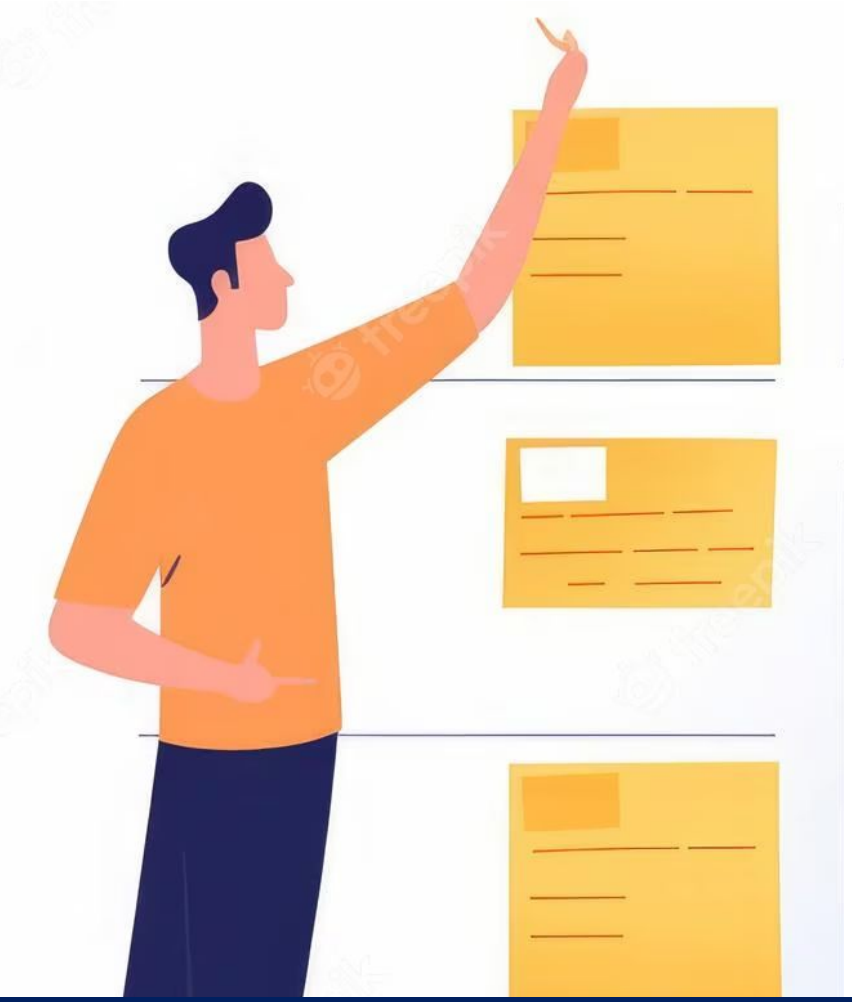
doi.org/10.5281/zenodo.10064635

Co-funded by
the European Union

2021-1-RO01-KA220-HED-000031158

ANIMV DARE. LEARN. INNOVATE.
Universidad Rey Juan Carlos
L-Università ta' Malta
NEW STRATEGY CENTER

Exercise in Groups



Running time:
40 minutes

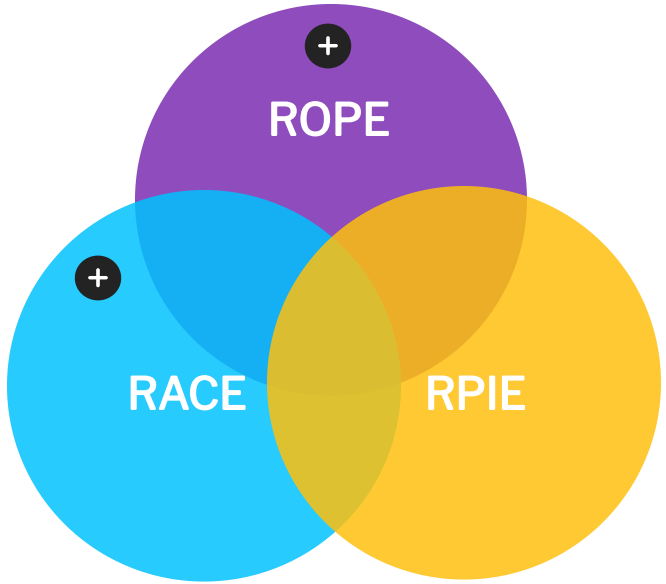
Students split into **groups**

The students made the outline of a strategic planning

Each group should appoint
1 representative for taking notes (in addition to engage in discussion) and present the results

Strategic Communication

These steps are captured by different acronyms:



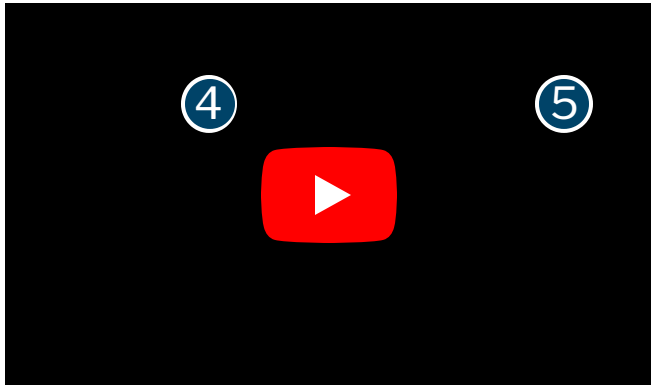
Strategic Communication

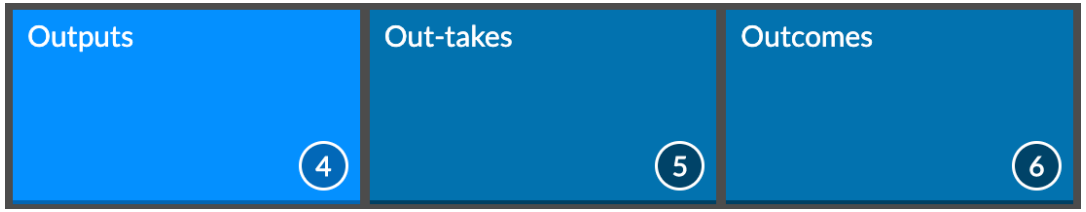
Step 1: Research/Intelligence

Step 2: Planning

Step 3: Implementation

Step 4: Evaluation





Method

1. Students will be divided into work groups. They must work as strategic communication officers.

2. They create a strategic planning scheme:

    - Target groups
    - Goals
    - Strategies
    - Tactics
    - Measurement of goals.

3. A representative of the group presents their results in a maximum of 5 minutes.

DOMINOES
digital resilience to disinformation

Digital cOMpetences INformatiOn EcoSystem

ANIMV DARE. LEARN. INNOVATE.
Universidad Rey Juan Carlos
L-Università ta' Malta
NEW STRATEGY CENTER

Co-funded by the European Union

2021-1-RO01-KA220-HED-000031158

Audiovisual and multimedia production: CIBERIMAGINARIO