UNMASKING MASKIROVKA (IN HUNGARY)

A Novel Natural Language Processing-based Detection Method for Information Laundering

By

Domonkos D Kovács

Submitted to
Central European University
Department of International Relations

In partial fulfillment of the requirements for the degree of Master of Arts

Supervisor: Alexander Etkind

Vienna, Austria

2024

Abstract

Information Laundering is a stratagem of hostile information influence campaigns (HIICs). Information Laundering aims to integrate propaganda into unwitting legitimate domestic outlets, by moving it through a network of intermediaries to produce a loss of origin, thereby lending propaganda a façade of legitimacy. Empowered by the affordances of the digital age, Information Laundering is increasingly deployed by the Kremlin to penetrate and subvert Western media ecosystems.

The current state-of-the-art methodological framework for detecting Information Laundering is inherently curtailed by the manual overhead it necessitates, rendering it unfeasible for near real-time detection for entire media ecosystems. To overcome this limitation, the paper develops a Natural Language Processing and Machine Learning-based reusable, scalable, and adaptable methodological framework leveraging Sentence Bidirectional Encoder Representations from Transformers technology, which can be deployed to detect and trace Information Laundering in targeted media ecosystems.

In turn, the paper trials the novel algorithmic methodological framework and uncovers a yet unnoticed Kremlin-orchestrated Information Laundering scheme targeting Hungary, which successfully subverted the Hungarian government-affiliated media ecosystem. The investigation finds that in the context of the full-scale invasion, between January 1, 2022, and March 10, 2024, four Hungarian government-affiliated outlets have provided for the integration of 891 laundered propaganda articles, which originally appeared on the columns of RT or Sputnik, and penetrated the Hungarian media ecosystem with the mediation of the Hungarian-language Kremlin-proxy site Orosz Hírek, which worked to mask the origin of the articles.

Acknowledgements

I wish to extend my gratitude to Alexander Etkind, Anton Shekhovtsov, Bret Schafer, and Thomas Maguire for their invaluable input, and to Snus, for keeping me motivated.

Table of Contents

List of Figures

Introduction[1]

On the 18th of October 2022, at 16:38, RT, the Russian state-controlled propaganda outlet publishes an article titled "Report on alleged Ukrainian war crimes presented to UN" (RT 2022b). Whilst the article's title is already misleading – since the vast majority of war crimes identified in the report were committed by Russian servicemen – the RT piece contains another conspicuous detail. It claims, that "while the instances of mistreatment of Russian POWs at the hands of Ukrainian fighters are backed by hard evidence, such as videos, the alleged violations committed by Russian forces mostly rely on the accounts of witnesses." The report referenced in the article, authored by the Independent International Commission of Inquiry on Ukraine (2022), contains no such assertion – the detail appears to be entirely RT's fabrication.

Mere hours later, at 21:13, Orosz Hírek,[2] the most prominent Hungarian-language fringe pro-Russian outlet, with proven operative connections to the Kremlin information warfare machinery,[3] publishes an article titled "Ukrainian war crimes reported by the UN"[4] (Orosz Hírek 2022c). The article claims that "While the violations against Russian prisoners of war committed by Ukrainians are backed by real evidence, such as videos, the alleged infringements committed by Russian forces are mostly based on eyewitness accounts."[5] The Orosz Hírek article, presented as the outlet's native content without reference to RT, appears to be a full translation of the RT piece published a few hours earlier.

[1] The paper, in part, builds on the author's previously written, unpublished works. Materials appearing in the paper from the author's unpublished works are referenced as quotations, with citations Kovács 2023 and Kovács 2024. Bibliographic details of the unpublished works are included in the References section.

[2] Translation: 'Russian News'

[3] See 2.2.2 Investigation into Orosz Hírek

[4] Original: „Ukrán háborús bűnökről közölt jelentést az ENSZ"

[5] Original: „Míg az orosz hadifoglyok ukránok által elkövetett eseteit valódi bizonyítékok, például videók támasztják alá, addig az orosz erők által elkövetett állítólagos jogsértések többnyire szemtanúk beszámolóira támaszkodnak."

The following morning, on the 19th of October 2022, at 09:26, Mandiner, a leading Hungarian government-affiliated weekly, publishes an article titled "UN report: the Ukrainian army committed war crimes".[6] The piece claims that "while crimes committed by Ukrainians are '*supported by real evidence, such as videos*,' alleged violations by Russian forces rely mostly on eyewitness accounts"[7] (Mandiner 2022b, emphasis in original). The article is a near-verbatim copy of the publication appearing on Orosz Hírek. Mandiner's article not only eschews reference to Orosz Hírek – which already masked the article's true source – but provides the aforementioned claim in quotation marks positing that it is "according to the document", giving the false impression that the authors are directly drawing on the UN report. Citing Mandiner's article, four prominent government-affiliated outlets republish the story: Origo (2022b), Vadhajtások (2022a), 888 (2022), and PestiSrácok (2022) – all, word for word, reiterating the same false claim.

Thus, Hungarian government-affiliated media, subverted by a Kremlin information warfare scheme, published a propaganda piece authored by Kremlin mouthpieces and translated by the anonymous editors of the Kremlin-proxy Orosz Hírek the day before, whilst giving the impression that they had directly consulted the UN report. In the process, they exposed at least 31,037[8] unwitting Hungarian media consumers to Russian state propaganda, who remained unaware, that their news was written not in Budapest, but in Moscow. This process is termed *Information Laundering* (IL).

IL is not exclusive to the 21st century. One of Russia's most notorious active measures campaigns, known as Operation INFEKTION (Boghardt 2009), deployed a methodology eerily

[6] Original: „ENSZ-jelentés: háborús bűncselekményeket követett el az ukrán hadsereg"
[7] Original: „hogy míg az ukránok által elkövetett bűncselekményeket *„valódi bizonyítékok, például videók támasztják alá"*, addig az orosz erők által elkövetett állítólagos jogsértések többnyire szemtanúk beszámolóira támaszkodnak."
[8] Combined number of engagements/ad impressions from the corresponding Facebook posts of Mandiner, 888, and Origo

similar to the one showcased above (Linvill and Warren 2023). On 17th July 1983, the KGB planted a fictitious letter to the editor in the Indian newspaper 'The Patriot'. The anonymous letter, purported to be written by a respected American scientist, was titled: "AIDS may invade India: Mystery disease caused by US experiments" (Boghardt 2009, 4). Over time, the letter's claim, that HIV originated at a US biological warfare facility, spread into the global media ecosystem with the mediation of the GDR scientist Jakob Segal and the unwitting FRG newspaper 'Tageszeitung' (Boghardt 2009, 8–12). By late 1987, the claim appeared in 80 countries, in over 200 outlets, in 25 languages. The British newspapers Daily Telegraph and Sunday Express reported the false claim uncritically (Boghardt 2009, 14). Unbeknownst to them, the story was a KGB fabrication.

Whilst IL has been developed in the pre-internet age, the digital era's affordances render it more potent than ever. Contemporary IL campaigns are "faster […] greater volume, and […] lower costs than was the case with Soviet-era campaigns" (Linvill and Warren 2023, 2). "With the shortening of the news cycle, the widening of the parameters of trusted information (Klein 2012, 430), the deepening issue of 'data smog' (Klein 2012, 430), the proliferation of 'non-traditional media actors' (Puschmann et al. 2016, 143) […] and the growing pressure for reach over quality, benign news outlets have fewer resources to vet and verify stories, and rely extensively on second-hand reporting, whilst malign outlets face lesser scrutiny. These tendencies give rise to an increasing vulnerability of Western media ecosystems to infiltration by IL." (Kovács 2023, 2)

In accordance, IL – "updated for the digital age" (DiResta and Grossman 2022, 5) and "turbo-charged with modern technology" (Linvill and Warren 2023, 3) – is increasingly being deployed by the Kremlin in hostile information influence campaigns (HIICs). "A declassified 2021 assessment by the US National Intelligence Council concluded that IL was a 'key element of Moscow's strategy' in attempting to interfere with US elections (2021, 7), and as Belén

Carrasco Rodríguez notes, IL has been extensively utilised by the Kremlin in both Germany and the Nordic-Baltic region, as part of a holistic interference endeavour (2021a; 2021b)." (Kovács 2023, 1)

Beyond the (geo)political implications of instrumenting and instrumentalising (Zuboff 2019, 311) legitimate domestic media outlets in Kremlin-directed HIICs and providing direct, covert access for Kremlin propaganda to unsuspecting media consumers, IL – where the information originates on the columns of RT or Sputnik – possesses a notable legal dimension in the European context. "Council Decision (CFSP) 2022/351 of 1 March 2022 (Council of The European Union 2022a) and Council Regulation (EU) 2022/350 of 1 March 2022 (Council of The European Union 2022b) introduced restrictive measures on RT and Sputnik, outlawing their broadcasting activities in the EU. According to Article 2f (1) of Council Regulation (EU) 2022/350 and Article 4g of Council Decision (CFSP) 2022/351, it is 'prohibited for operators to broadcast or to enable, facilitate or otherwise contribute to broadcast[ing]' RT and Sputnik materials. Furthermore, Article 12 of Council Regulation (EU) 2022/350 stipulates, that it is 'prohibited to participate, knowingly and intentionally, in activities the object or effect of which is to circumvent prohibitions in this Regulation including by acting as a substitute for [RT or Sputnik].'" (Kovács 2024, 8) In light of these regulations, it can be argued that Hungarian fringe and government-affiliated media outlets implicated by the findings of the present paper are committing sanctions evasion and are in violation of EU law.

As outlined in Chapter I, in the fourteen years since the coinage in 2010 (Klein 2010), the scholarship on IL has provided for its theoretical triangulation and modelling and produced an operationalisable methodological framework (Rodríguez 2021a; 2021b) for detecting IL. However, the current state-of-the-art Rodríguez methodology faces a notable limitation. Since it relies on "manual keyword-based advanced online research" (Rodríguez 2021b, 15) its

scalability – both in breadth and deployment frequency – is inherently curtailed by the manual overhead it necessitates.

In recognition of a perceived gap between the increasing threat posed by IL and the limitations of the extant operationalisable methodological framework, the present paper moves to develop a Natural Language Processing and Machine Learning-based reusable, scalable, and adaptable methodological framework, which can be deployed to detect and trace digital IL in targeted media ecosystems. The framework builds on Rodríguez's findings and leverages the state-of-the-art Sentence Bidirectional Encoder Representations from Transformers technology (Reimers and Gurevych 2019).

In turn, the paper trials the novel algorithmic methodological framework on the Hungarian government-affiliated media ecosystem and uncovers a yet unnoticed Kremlin-orchestrated IL scheme targeting Hungary, which successfully subverted the Hungarian government-affiliated media ecosystem. In this scheme, Orosz Hírek – which the author's Open-Source Intelligence investigation identified as a likely Kremlin-proxy and a sanctions evasion 'copy-paste site' (Balint et al. 2022) to RT and Sputnik – plays an integral, intentional role. It translates RT and Sputnik propaganda articles, and provides for the loss of origin, thereby implanting them into the Hungarian-language media ecosystem, whilst masking their true source. The algorithmic investigation finds that in the context of the full-scale invasion, between January 1, 2022, and March 10, 2024, four Hungarian government-affiliated outlets have provided for the integration of 891 laundered propaganda articles, which originally appeared on the columns of RT or Sputnik and penetrated the Hungarian language media ecosystem with the mediation of the Hungarian-language Kremlin-proxy site Orosz Hírek.

In this sense, the paper poses and answers two, interlinked research questions: Can Natural Language Processing be utilised to detect IL? and Do Hungarian fringe and

government-affiliated media outlets provide for the layering and integration phases of Kremlin IL campaigns?

"The endeavour is animated by Boris Toucas's call for a 'systematic analysis of the 24/7 news environment to determine the respective contribution of Russia's own measures' (2017) to counter IL – a solution further developed by Domonkos D Kovács and Victora Bergström (2021, 163–64)" (Kovács 2024, 2). The value proposition of the algorithmic methodological framework lies in rendering Toucas's "systematic analysis"– out of reach for the extant models – viable. The algorithmic methodological framework can be scaled to encompass entire media ecosystems, and provide for the live, near real-time detection of IL utilised in hostile information influence campaigns. To the author's best knowledge, the present paper is the first study to leverage Natural Language Processing to detect IL, the first to investigate full-cycle[9] Russian IL in the Hungarian media ecosystem, and the first to offer evidence, that several Hungarian fringe and government-affiliated media outlets provide for the layering and integration phases of IL campaigns, by republishing sanctioned Kremlin propaganda pieces.

The paper is organised as follows. Chapter I surveys the existing literature on IL, frames it as constituting a trajectory of gradual, cumulative development, and positions the present paper in the literature as the next step in this trajectory. The chapter also outlines the author's position on information typology as pertinent to IL and integrates it into a consolidated, operationalisable definition of IL emerging from the literature.

Before delving into algorithm design, 'Chapter II: Actors at a Glance' tackles indispensable contextual points vis-à-vis the actors examined in the empirics. The first section titled 'Integration: The Hungarian Government-affiliated Media Ecosystem in Light of Russian Propaganda' addresses the rationale for target media selection and provides an overview of the

[9] From Placement via Layering to Integration

relationship between Kremlin propaganda and the Hungarian government-affiliated media ecosystem, situating the research in existing, specialised literature. The following section, entitled 'Layering and Placement: Proof of Intent', provides the theoretical scaffolding for Chapter III, necessary to prove that the observed phenomenon indeed constitutes information warfare – and specifically IL – rather than merely bad journalism. Beyond stipulating that RT and Sputnik are instruments of information warfare, the section reports the results of an Open-Source Intelligence (OSINT) investigation into Orosz Hírek conducted by the author, which finds that the fringe Hungarian-language pro-Kremlin site is likely to be a Kremlin-proxy actor.

'Chapter III: A Novel Natural Language Processing-based Detection Method for Information Laundering' first discusses the merits and limitations of the algorithmic approach and describes the structure and operation of the two algorithms developed for the project, with a focus on providing an adaptable, reproducible template for future investigations utilising the methodological framework. To these ends, key considerations and points of decision in the algorithms' design are highlighted. The methodological framework operates two primary algorithms: The Natural Language Processing-based 'Core Algorithm' tasked with identifying suspect IL links, and an eleven-feature Machine Learning filter, which serves to validate them. Finally, Chapter III's Findings and Analysis section reviews the algorithms' output, to draw both structural, and outlet-level conclusions, and showcase how, and to what extent Hungarian government-affiliated media integrates discrete pieces of Kremlin propaganda, layered via Orosz Hírek, and placed by RT or Sputnik. We find that Origo, Magyar Hírlap, Mandiner, and Vadhajtások – four of the five target media – provide for the integration phase of a Kremlin-orchestrated IL scheme, hosting a total of 891 laundered articles in the observed period, which originally appeared on the columns of RT or Sputnik. Furthermore, as detailed in the Findings and Analysis section, the study observes and conceptualises a yet unnoted phenomenon pertinent to IL – termed *Layering-Integration Liminality.*

Chapter I: State of the Literature – Towards A Consolidated, Operationalisable Definition

The trajectory of scholarship on IL is that of gradual, cumulative development. The foundational theory of IL – and the coinage itself – was developed by Adam Klein (2010; 2012). Klein conceptualised IL as a new theory of process, specific to the internet (2012, 429), to describe how the internet allows subversive social movements – white supremacist groups in his empirics – to "quietly legitimize their causes through a borrowed network of associations" (2012, 429). Klein introduced some of the key features of IL: the production of a "perception of legitimacy" (2012, 431), the information pathway moving propaganda from the fringe to the mainstream (2012, 432), and what the author would term, "the subversion of the native epistemological space by hijacking the established forms of the authoritative *techno-ethos* (Borrowman 1999, cited in Klein 2012, 434)." (Kovács 2023, 4) In contrast to later literature which sees IL as a stratagem of states' hostile information influence campaigns, Klein's framework understands it to be an organic, emergent process, with no strategic intention or organisation driving it – reflecting its grassroots nature. This conceptualisation, focusing on subversive social movements, was later operationalised by Cornelius Puschmann, Julian Ausserhofer, Noura Maan, and Markus Hametner (2016).

Anton Shekhovtsov was the first to deploy the concept in a geopolitical context – albeit utilising slightly different terminology. In 2015, in reference to the relations between Kremlin propaganda and the European far right, Shekhovtsov used the term "narrative laundering" to describe the movement of narratives in the media space, whereby "the original source that produces these narratives is either forgotten or impossible to determine" (Shekhovtsov 2015, 5–6). Shekhovtsov argued, that Russian media is interested in the "loss of origin" for "propagandistic narratives" implanted in the international media sphere, and outlined, that the

success criterion for the process is the integration of the "propagandistic narrative" into mainstream media.

In turn, in a 2017 CSIS white paper on Russian interference in the 2016 US presidential elections, Boris Toucas conceptualised IL as a form of intentional information warfare and recognised that the Kremlin is adapting its active measures tactics to the information age (Toucas 2017). Kirill Meleshevich and Bret Schafer's contribution represents another important step forward. Meleshevich and Schafer (2018) retain IL's focus on foreign interference and move to expand Klein's money laundering metaphor by bringing in the three-stage construction – placement, layering, integration – established in money laundering scholarship (see Schneider and Windischbauer 2008; bin Zul Kepli and Nasir 2016; Cassella 2018).

Samantha M. Korta, building on Klein's theory and Meleshevich and Schafer's introduction of the placement–layering–integration pathway, develops 'Information Laundering 2.0', an expanded and refined version of Klein's model. Information Laundering 2.0 incorporates the secondary actors of amplifiers and accelerators into the layering stage, brings in new domains – such as social networks, online shopping, and gaming – and accounts for emerging technologies facilitating the process (Korta 2018, 79–89).

Drawing on Korta as well as Meleshevich and Schafer, Belén Carrasco Rodríguez introduced the first comprehensive, versatile analytical and methodological framework, which can be exported and adapted to various media ecosystems, to detect IL (Rodríguez 2021a; 2021b). One of Rodríguez's key innovations lies in the development of "five indicators that can be used to identify connections between laundered news items and to trace the exact route of the laundering process" (Rodríguez 2021b, 12). These five indicators in turn serve as a basis for the Natural Language Processing-based methodological framework introduced in the current paper (*see Section 3.1 for a detailed technical explanation*). In this sense, the present

study positions itself in the literature as the next, cumulative step in the development trajectory. Just as previous studies outlined above, it integrates, and builds on the findings of preceding scholarship, whilst introducing a novel layer of understanding.

An implicit, but notable divergence in the literature is constituted by the typology of the information being laundered. Klein applies the theory to content described as "propaganda and false information*"* (2012, 433), "racist rhetoric and extremist agendas" (2012, 445) and "hate-based information" (Klein 2010, 8); Shekhovtsov refers to "propagandistic narratives" (2015, 5–6); Toucas talks about "Russian narratives" (2017); Korta posits that IL operates on "false or extremely biased information" termed "counterfeit narratives" (2018, 78–79); and Meleshevich and Schafer (2018) , as well as Rodríguez (2021a; 2021b), narrow the inquiry to "disinformation".

The present paper returns to a broader framing and embraces IL as a stratagem operating on *propaganda,* building on the conceptualisations of Anton Shekhovtsov and Paul A. Smith (Smith 1989, 3; Shekhovtsov and Lysenko 2024; cf. Barclay 2018). Shekhovtsov argues that "propaganda is a set of narratives" deployed to "change the behavio[u]r of what they consider their 'adversary' or 'enemy' or any other entity that they want to change the behavio[u]r of, […] This is what propaganda is: not necessarily disinformation, just a set of narratives […] sometimes propaganda can be without disinformation". Smith defines propaganda as "political advocacy aimed abroad with hostile intent." (Smith 1989, 3).

The broader, propaganda framing is warranted owing to two reasons, stemming from the realities of the findings. First, propaganda – containing no categorical disinformation – does indeed get laundered, and as a corollary, does serve the information warfare ends of the Kremlin. As such, it should not be excluded from the inquiry. For instance, as the Findings and Analysis section illustrates, Hungarian government-affiliated media oftentimes integrates verbatim *framings* and *interpretations* offered by Kremlin media on news events. Whilst these

do not necessarily comprise factually false information, they effectively convey the Kremlin's political agenda, which, in an IL scheme such as the one identified by the current paper, are presented as the native *interpretations* and *framings* of domestic, legitimate media. Second, the move to a *propaganda* framing is informed by the practical considerations emerging from the scaling of detection. In order for a large-scale, media ecosystem-wide IL detection framework to feasibly be limited to only identifying laundered disinformation, and effectively work to identify all laundered disinformation, an up-to-date fact-checking database encompassing the entirety of a given media ecosystem would be necessary. In contrast, propaganda provides a lower, more operationalisable threshold, since anything published on RT and Sputnik – as outlined in the section Proof of Intent – can be argued to constitute "political advocacy aimed abroad with hostile intent" (Smith 1989, 3) and a "set of narratives" deployed to "change the behavio[u]r of what they consider their 'adversary' or 'enemy'" (Shekhovtsov and Lysenko 2024).

Therefore, in a consolidation of the coordinates provided by extant literature, and the author's pertinent position on information typology, IL can be defined as a stratagem of hostile information influence campaigns (HIICs) (Rodríguez 2021b, 6), which operates via decontextualising *propaganda* (James Pamment et al. 2018) by moving it through a network of intermediaries (Rodríguez 2021b, 13), ultimately producing a 'loss of origin' (Shekhovtsov 2015) and thereby lending *propaganda* a "perception" (Klein 2012, 431) and "façade of legitimacy, the look and feel of credible news, allowing [*propaganda*] to influence public discourse or to set an agenda" (Rodríguez 2021b, 10). IL consists of three phases: a) placement – the initial publication of *propaganda* (Meleshevich and Schafer 2018); b) layering – when *propaganda "*is laundered through a series of domains […] that opens it up for public discourse without the original source or motive being understood" (Korta 2018, 15); and c) integration –

when *propaganda* "becomes adopted by trusted news sources" (Meleshevich and Schafer 2018, 5).

The following chapters, in turn, operationalise this definition. First, Chapter II places the actors examined in the empirics in the model's three phases and substantiates that the nature of the actors in the placement and layering phases supports the *hostile information influence campaign (HIIC)* framing. Second, Chapter III showcases IL in action, by shedding light on how, in the Hungarian context, Kremlin propaganda is decontextualised via an intermediary, ultimately producing a loss of origin, and facilitating propaganda texts' adoption by legitimate domestic news outlets.

Chapter II: The Actors at a Glance

"If you place a breadcrumb in the far reaches of the Internet on a website, it doesn't need to generate traffic on its own. It just needs to land on the radar of an outlet that is real or run by [Hungarians]."[10] – argues Bret Schafer (quoted in Harris and Nakashima 2021), in reference to the underlying rationale of IL. The following paragraphs survey the three strata of outlets represented in the empirics, corresponding to the three stages of IL: integration, layering, and placement. For integration, we review those which are "real or run by [Hungarians]" – five Hungarian government-affiliated news sites; for layering, one representing "the far reaches of the Internet" – Orosz Hírek; and for placement, those which worked to place the breadcrumbs – RT and Sputnik.

The purpose of the chapter is twofold. First, it outlines the Hungarian government-affiliated media's nexus with Russian propaganda, necessary to contextualise the paper's findings. Second, it addresses the rationale for target media selection, and vis-à-vis the layering and placement phases, validates their inclusion in an investigation working to uncover IL, by proving their intent to wage information warfare, via, amongst others, reporting on the results of an Open-Source Intelligence (OSINT) investigation into Orosz Hírek conducted by the author.

2.1 Integration: The Hungarian Government-affiliated Media Ecosystem in light of Russian propaganda

Already in the years leading up to the outbreak of full-scale invasion in 2022, the Hungarian government-affiliated media, which Péter Krekó describes as a product of

[10] Original: „run by Americans" – adjusted for context.

"Informational Autocracy" (2022), has been noted for its preferential treatment of the Russian Federation, and its proclivity to broadcasting pro-Russian propaganda and sharing the Russian viewpoint (Hunyadi, Molnár, and Szicherle 2017).

The author of the present paper, in a bout of wishful thinking, theorised in an article published in March 2022 (Kovács 2022), that in light of the discrepancy (Berkes 2022; Bayer 2022) between the Hungarian government's then pro-EU, pro-NATO stance in international fora, and the persistent broadcasting of pro-Russian propaganda in affiliated outlets, the latter communications strategy possibly constitutes a 'grace period' approach, designed not to unsettle the audience, which for years has been conditioned to view Putin's Russia favourably.

On the contrary, the prevalence of pro-Kremlin propaganda in Hungarian government-affiliated outlets has not decreased since. Hungarian government-affiliated media was rendered "the number one broadcaster of the Kremlin propaganda in Europe" (Grygiel 2023, 177), to the point where "the pro-Kremlin disinfo and government information ecosystem cannot really be separated," (Krekó, quoted in Bayer 2022). Both state-owned public (Zubor 2022; Bayer 2022; Presinszky 2022; Lakmusz 2022; Keller-Alánt 2022; Rényi 2022), and nominally independent, government-affiliated media are implicated (Diószegi-Horváth 2022; Zubor 2022; Bayer 2022).

"As Karin Kőváry Sólymos and Szabolcs Panyi, as well as Szicherle et al. argue, Russian propaganda sites have a limited role in Hungary, since Hungarian government outlets themselves fulfil the function of spreading the Kremlin's message (Szicherle et al. 2019; Sólymos and Panyi 2023). However, it is yet to be determined, whether Hungarian government-affiliated media merely exhibits narrative proximity with Kremlin coverage, as, for instance, proven by Corruption Research Centre Budapest (Hajdu et al. 2018), or, whether it also hosts discrete, specific propaganda texts authored by Kremlin-official media – that is to

say, whether Hungarian government-affiliated media provide for the *integration* phase of Russian IL campaigns" (Kovács 2024, 1). The present paper remedies this omission.

In accounting for IL's integration phase, the algorithm has been deployed on five Hungarian government-affiliated outlets: Mandiner, Magyar Hírlap, Origo, Vadhajtások, and Hirado.hu. The selection rationale was informed by aiming to provide a representative set sampling the Hungarian government-affiliated media ecosystem's various strata. Mandiner and Origo belong to the Central European Press and Media Foundation (KESMA), a non-profit exempted from Hungarian competition law, owning and operating more than 470 outlets (Bátorfy 2018). KESMA is widely regarded by scholarship to be a government-affiliated media conglomerate (Brogi et al. 2019; Griffen 2020; Urbán 2021; Bátorfy et al. 2022; Wiseman 2023). KESMA is suspected to receive communication directives directly from the ministry (Political Capital 2023).

Mandiner is a medium-reach (Mandiner 2019) government-affiliated weekly, understood by the author to represent an outlet offering more balanced coverage than other KESMA sites. Origo, the highest-reach government-affiliated online outlet, consistently ranks amongst the top three most popular websites in Hungary, visited by more than 900,000 people daily (Pete 2023; Nemzeti Média- és Hírközlési Hatóság 2023). It has been noted for its sensational, click-bait headlines oftentimes approximating disinformation (Szily 2022; Hanula 2023), sexualised tabloid reporting (Herczeg 2022; Rovó 2022), smear campaigns (HVG 2021; Horváth 2021; HVG 2022; Erdélyi K. and Szabó 2024), and heavily skewed pro-Kremlin coverage (Diószegi-Horváth 2022; Molnár and Krekó 2023). Magyar Hírlap is a lower-reach (RTL 2022) government-affiliated outlet owned by businessman Gábor Széles, a major donor to Fidesz (HVG 2010) with close business ties to Russia (Juhász et al. 2015). Vadhajtások is a far-right publication (Győri, Keller-Alánt, et al. 2022, 19), deemed by researchers to be a 'grey

zone' government-affiliated outlet (Political Capital 2022). Finally, Hirado.hu is the news website of the Hungarian public broadcaster MTVA.

2.2 Layering and Placement: Proof of Intent

For cases exhibiting IL characteristics to be properly categorised as IL – and by extension, not merely bad journalism – the intent of the Kremlin to wage information warfare, and use IL to penetrate the Western, and specifically the Hungarian media ecosystem, must be substantiated. As Rodríguez argues, the usage of IL techniques only constitutes a hostile information influence campaign, if it "fits within a broader strategic aim" (Rodríguez 2021b, 13). To this end, the following two points require validation.

First, materials published in RT and Sputnik are meant not merely to inform international audiences but are instruments for waging information warfare. Second, given the operative affiliation between Orosz Hírek and the Kremlin information warfare machinery, the fringe pro-Kremlin site is instrumentalised by the Kremlin to inject its messages into the Hungarian-language information sphere, using IL.

Orosz Hírek was selected to represent the layering phase, owing to its position as the most prominent Hungarian language fringe pro-Russian outlet (Győri, Keller-Alánt, et al. 2022; Zubor 2023; Német 2023), whilst RT and Sputnik have been chosen for the study's placement phase, due to their central role in the Kremlin's information warfare efforts, and their status as sanctioned entities in the European Union (Council of The European Union 2022a; 2022b), which renders the implications of present investigation's findings not only political, but also legal.

2.2.1 RT as a Tool of Information Warfare

Since its rebranding in the wake of the 2008 Russo-Georgian war, RT has been operating not as an instrument of Russian soft power, but of Russian political warfare against the West (Ramsay and Robertshaw 2018; Shekhovtsov 2022; Gulenko 2023). Margarita Simonyan herself argued that RT is an "information weapon" (quoted in Shekhovtsov 2022, 8) deployed in a war "against the whole Western world"[11] (Симоньян 2012), comparing its role to that of the Defence Ministry (Yatsyk 2022, 109). Indeed, RT's mission and philosophy are not journalistic but military (Nimmo 2018). In accordance, RT became notorious for targeting European societies with disinformation and propaganda (Yablokov and Chatterje-Doody 2022; Andreae 2022).

"By extension, the argument can be made that seeking to influence foreign audiences, and ultimately advance the political, economic and security interests of Russia on the international stage (Shekhovtsov 2020, 10), anything published on a foreign-language version of RT is *intended* to be part of an information warfare effort, and be picked up by legitimate news media of the target media ecosystem." (Kovács 2024, 4)

Given the personal and organisational links, as well as oftentimes similar *modus operandi* (Littow 2022, 29), the same argument is applicable to Sputnik International. Sputnik International is operated by the news agency 'Rossiya Segodnya' (Russian for 'Russia Today'). Although RT claims that it has no relationship to 'Rossiya Segodnya' (Littow 2022, 27) Margarita Simonyan serves as the Editor in Chief for both outlets (Rossiya Segodnya 2024). The European Council's concurrently instituted ban on both Sputnik and RT (Shekhovtsov 2022, 9; Council of The European Union 2022a; 2022b), with European Commission President Ursula von der Leyen saying "The state-owned Russia Today and Sputnik, and their

[11] Original: „информационную войну вели мы, причем со всем западным миром."

subsidiaries, will no longer be able to spread their lies to justify Putin's war," (quoted in Kayali 2022) further exemplify that they both serve information warfare purposes.

2.2.2 *Investigation into Orosz Hírek*

"Orosz Hírek – the most prominent Hungarian language fringe pro-Russian outlet (Győri, Keller-Alánt, et al. 2022; Zubor 2023; Német 2023) – is likely to be a Kremlin-proxy, according to the author's OSINT investigation presented below. The presence of a potential Kremlin-proxy outlet providing Hungarian-language coverage in an IL scheme can be argued to substantiate the Kremlin's intent to penetrate the Hungarian media ecosystem. [The following paragraphs propose four indicators, which offer strong evidence, that Orosz Hírek is a Kremlin-proxy.]

First, as indicated in the 'About' section of Orosz Hírek's official YouTube channel and the 'Description' section of several of its videos, Orosz Hírek's content-providing partner is Ruptly. Ruptly, whilst operating as a Berlin-based German commercial entity, is the subsidiary of the Russian state-controlled television network RT (Escritt 2022). Based on content analysis, the partnership between Ruptly and Orosz Hírek seems to primarily involve the provision of either live-streamed or recorded video content from Ruptly, which then in turn are fitted with Hungarian titles and descriptions, subtitles, or both, working to make them accessible to Hungarian speaking audiences.

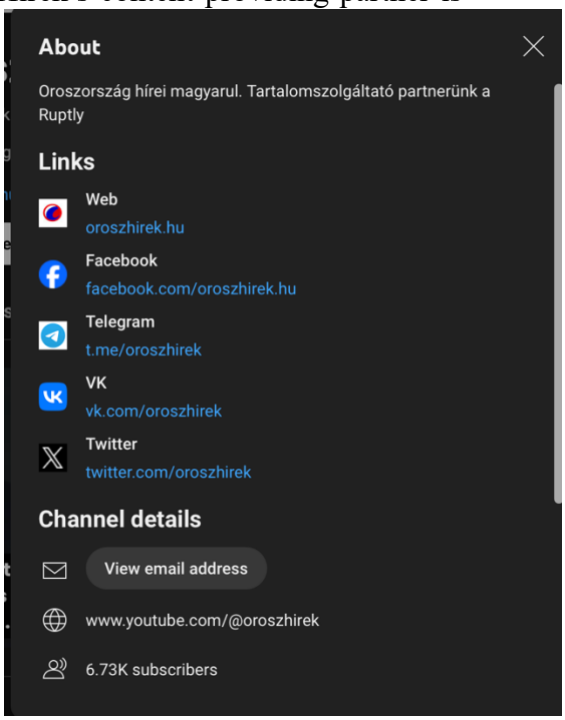

Figure 1: About section on Orosz Hírek's YouTube channel.

Therefore, an existing partnership between Orosz Hírek and a Russian state-controlled propaganda outlet, facilitating the intentional transposition of content sourced from a Russian state outlet into Hungarian media, can be substantiated.

Second, an investigation into Orosz Hírek's source code reveals, that the site's builders used Yandex Metrica, a Russian web analytics service. Using the source code search engine from publicwww.com, we conclude that Yandex Metrica is utilised almost exclusively by Russian websites, amongst others, key Russian state-owned and state-controlled outlets such as www.rt.com, www.ria.ru, www.gazeta.ru, www.vesti.ru, www.tass.ru, and www.interfax.ru. Whilst the ubiquity of Yandex Metrica across Russian websites precludes using its appearance in Orosz Hírek's HTML code as conclusive proof of its affiliation to the Russian state, it suggests that the website's developers possess Russian connections.

Figure 2: Excerpt from Orosz Hírek's HTML code

Third, personal links between the Russian intelligence-affiliated outlet NewsFront and Orosz Hírek can be corroborated. NewsFront, based in Russian-occupied Crimea, is recognised by experts as one of the key vehicles of the Kremlin's information warfare efforts (Stelzenmüller 2017; Nimmo 2017). Reportedly, the outlet is financed directly by the FSB, with message control being exercised by the presidential administration in Moscow (Beuth et al. 2017). The US Department of Treasury has termed it a 'disinformation and propaganda outlet that worked with FSB officers' (U.S. Department of the Treasury's Office of Foreign Assets Control (OFAC) 2021) whilst the Atlantic Council found that News Front is one of the Kremlin's three main propaganda outlets in Germany, alongside Sputnik and RT (Shuster

2017). Since its inception in 2018 until its re-launch in August 2022, the key person behind News Front's Hungarian version has been Sándor Csikós (Sólymos and Panyi 2023), who stands accused of cooperation with terrorist organisations in the DNR/LNR by the Myrotvorets Center (2018). Sándor Csikós also appears as a columnist at Orosz Hírek, having published several articles on the outlet (Csikós 2017; 2018; 2019; 2022) – implying a potential affiliation between the FSB-financed Hungarian version of News Front and Orosz Hírek." (Kovács 2024, 4–5)

Fourth, researchers at Political Capital identified Orosz Hírek as the originator of a pro-Kremlin disinformation narrative – sourced from the Russian media ecosystem – which was propagated by bot activity on social media in Hungary (Győri, Molnár, et al. 2022, 13–15). The text, claiming that Ukraine has been committing genocide against Russians in Donbas, was first published in Hungarian on Orosz Hírek's Facebook page on February 24, 2022. Subsequently, the text was copied verbatim into Facebook comments at least 1,708 times, amongst others, by inauthentic profiles (Győri, Molnár, et al. 2022, 13–15). Whilst it has not been proven that the inauthentic profiles spreading Orosz Hírek's text are Kremlin-affiliated, the usage of bots to spread pro-Kremlin disinformation of social media bears the mark of the Kremlin's *modus operandi* (see Woolley 2016; Sanovich 2018; Badawy, Ferrara, and Lerman 2018; Stukal et al. 2022).

Having substantiated the intent of the Kremlin to use IL to penetrate the Hungarian media ecosystem via Orosz Hírek and outlined the synergetic relationship between Russian propaganda and the Hungarian government-affiliated media ecosystem, the following chapter moves to develop and deploy the algorithmic framework, working to determine, whether Hungarian fringe and government-affiliated media outlets provide for the layering and integration phases of Kremlin IL campaigns.

Chapter III: A Novel Natural Language Processing-based Detection Method for Information Laundering

Natural Language Processing (NLP) "is a collection of computational techniques for automatic analysis and representation of human languages" (Chowdhary 2020, 604). It emerged to circumvent the insurmountable challenge posed by the processing of the "abundant volume of natural language text in the connected world […] by a human to discover the knowledge/wisdom in it, specifically within any given time limits" (Chowdhary 2020, 603). The *en masse* detection and tracing of IL presents the same problem characteristics NLP was designed to solve – a language processing task too extensive for a human to address. In accordance, the present chapter sets out to render the established, human-operated methodology developed by Rodriguez (2021a; 2021b) scalable, by transposing it into a "collection of computational techniques", leveraging state-of-the-art Sentence Bidirectional Encoder Representations from Transformers technology. In turn, the chapter operationalises the developed algorithm by deploying it to detect discrete pieces of laundered information in Hungarian government-affiliated media, originating in sanctioned Kremlin propaganda outlets, with the Findings and Analysis section reporting on the investigation's results.

3.1 Methodological Framework

The codes aim to automate assigning Rodriguez's indicators of connection between laundered news items (Rodríguez 2021b, 13). The majority of the code works to identify 'Structural similarity' ("news articles are either identical or contain similarly distributed information") by leveraging Natural Language Processing and Machine Learning technologies; 'Chronology' ("News articles have simultaneous or closely consecutive publication times) is

accounted for by a date filter; and for Recurrence ("a suspect media outlet repeatedly shares news articles from other outlets previously identified as disreputable within a short timeframe, e.g. Sputnik → News-Front") the results showcase oft-used pathways for IL. 'Authorship' ("suspicious articles were written by the same author") has been deemed irrelevant for the current dataset, insofar as suspect pieces cross language boundaries, and considering that both RT, Sputnik, and Orosz Hírek forego attributing authorship in their publications. For other datasets, the 'Authorship' indicator can be automated by adjusting the scraping process and introducing a parallel, simplified version of the NLP algorithm. The final indicator, 'Reference' ("an article published by one suspect media outlet cites or includes hyperlinks to another suspect sources the previous one") presents a curious case. Whilst Orosz Hírek eschews providing reference to RT or Sputnik in all the layering cases, two Hungarian government-affiliated media outlets – Magyar Hírlap and Vadhajtások – do attribute certain integration pieces to Orosz Hírek. In the findings section, this data is presented alongside detected counts, where available.

The methodological framework described in the following sections is deployed in two stages. Stage I analyses the RT/Sputnik and Orosz Hírek databases, aiming to produce a list of articles published on Orosz Hírek, which have originated on the columns of RT or Sputnik. Stage II in turn leverages the results of Stage I to identify instances where government-affiliated media integrated a layering piece from Orosz Hírek, which was initially published in RT or Sputnik. Both Stage I and Stage II follow the same workflow, consisting of the Natural Language Processing-based Core Algorithm and the eleven-feature Machine Learning filtering.

To differentiate the relative position of suspect texts in the IL hierarchy, the heuristic of *Stratum A* and *Stratum B* is introduced. *Stratum A* denotes suspect precursors to *Stratum B* texts. This implies that in the placement (RT/Sputnik) → layering (Orosz Hírek) relationship, the prior constitutes *Stratum A*, and the latter *Stratum B*, whereas in the layering (Orosz Hírek)

→ integration (Hungarian government-affiliated media) nexus, Orosz Hírek is referred to as *Stratum A*, and texts from Hungarian government-affiliated media as *Stratum B*.

3.1.1 Limitations – Levels of Proof Veracity

The current algorithmic approach presents certain limitations, emerging from the inherent challenges of providing conclusive proof in detecting IL. In the RT/Sputnik → Orosz Hírek nexus, the algorithmic methodology identifies near-verbatim translations – either full articles or excerpts –, where the publication time of the Kremlin propaganda piece precedes that of the Orosz Hírek article by no more than 72 hours. Given 'Chronology,' 'Structural Similarity', and Orosz Hírek's proclivity to hosting RT/Sputnik content ('Recurrence'), it can be argued that in this relationship, the results approximate conclusive evidence, insofar as is it highly unlikely that either the editors of Orosz Hírek would have, by coincidence, authored such a piece without consulting RT/Sputnik, or, alternatively, that both outlets have copied the content of a third source, external to the model. To investigate the probability of the latter's occurrence in the RT/Sputnik → Orosz Hírek relationship, a random sample of 300 laundered article pairs was assembled from the database identified by the algorithm. The texts of the RT/Sputnik pieces from the pair were fed into a search engine, to identify potential shared, Stratum A texts. The investigation confirmed, that for each of the 300 texts in the sample, the RT/Sputnik articles, as worded and structured, were unique, and no other text exhibiting semantic similarity – e.g. covering the same news event with a similar framing – was closer, or equally close to, in structure and wording to the Orosz Hírek article than its respective RT/Sputnik pair.

The same logic is applicable to finding matches in Hungarian government-affiliated media for the subset of Orosz Hírek articles, which were sourced from RT/Sputnik. Given the specific wording and structure of these articles – near-identical between RT/Sputnik and Orosz

Hírek – and the specific translation choices made by Orosz Hírek and reflected in the suspect government-affiliated pieces, it is highly unlikely that a government-affiliated media would have reproduced these without consulting Orosz Hírek.

However, in the Orosz Hírek → government-affiliated media nexus, when attempting to identify pieces sourced from the prior's entire corpus, rather than strictly those originating in RT/Sputnik and merely layered via Orosz Hírek, the likelihood of a common Stratum A external source to both Orosz Hírek and the investigated government-affiliated outlet increases, thereby reducing the veracity of the proof offered. Without the specificity provided by the established RT/Sputnik → Orosz Hírek link, and the latter's translation characteristics, it becomes increasingly possible, that the origin of the information is a third, Hungarian-language source. In contrast to the results of the uniqueness investigation by sampling in the RT/Sputnik → Orosz Hírek database, the same inquiry regarding Hungarian-government affiliated texts and Orosz Hírek's genuine – non-laundered – corpus revealed, that these texts are oftentimes not unique. For instance, the external source is frequently constituted by Magyar Távirati Iroda (MTI), a Hungarian state-run news agency, or another domestic source. Whilst extensive, additional keyword and citation-based filtering was undertaken to improve these datasets, the level of proof veracity offered in RT/Sputnik → Orosz Hírek, and RT/Sputnik → Orosz Hírek → Government-affiliated media databases, is out of reach for the Orosz Hírek → Government-affiliated media nexus, without indexing the entirety of Hungarian-language online news media. Since this is outside of the remit of the current project, data presented from the Orosz Hírek → Government-affiliated media database is provided in the findings as 'approximate' data, to indicate a lower level of confidence compared with the other two datasets. However, since the aim of the project is to identify discrete pieces of texts originating in sanctioned Kremlin-official outlets, the qualification regarding the confidence level of the Orosz Hírek →

Government-affiliated media database does not compromise the research's overall goals – this type of data is merely appended as complementary, contextual information.

3.1.2 Data Collection – Scraping

A total of 451,560 articles were collected for the study: 43,990 from RT; 50,556 from Sputnik; 10,254 from Orosz Hírek; 84,016 from Mandiner; 92,582 from Magyar Hírlap, 140,629 from Origo; 32,573 from Vadhajtások, and 26,960 from Hirado.hu – representing the entire published corpus of the target outlets between January 1, 2022, and March 10, 2024, with the exception of Hirado.hu, where only articles tagged 'abroad' were downloaded, due to the site's exceptionally efficient scraping blocker.

The Python package Newspaper3k was selected to scrape texts and publication dates of articles, owing to its capability to adapt to a wide range of HTML structures without tailor-made code for each site. To provide article URLs to be fed into Newspaper3k, sitemap XMLs were collected, parsed, and converted into databases, containing the URLs of all articles published between January 1, 2022, and March 10, 2024, from each target site.

With Sputnik International, due to the peculiar HTML structure, leveraging Newspaper3k was not feasible. To account for this, a traditional scraping method, with custom instructions reflecting Sputnik's HTML code, was utilised, aided by the *Beautiful Soup* python library.

3.1.3 Core Algorithm

Owing to their higher accuracy in creating embedding vectors for "language containing complex structure, ambiguous word usage, and words unseen in training" (Arora et al. 2020, 2650), the project only considered contextual algorithms for primary detection, over their traditional, non-contextual counterparts oftentimes utilised in Textual Similarity tasks, such as Jaccard Similarity, Term Frequency and Inverse Document Frequency, Word Movers Distance, or Count Vectorizer (Wang and Dong 2020). Whilst non-contextual algorithms create the same embedding vector for a word, irrespective of its textual environment, contextual methods account for polysemy and consider interdependencies, capturing uses of words in different settings (Liu, Kusner, and Blunsom 2020), consequently producing more accurate results.

The state-of-the-art language model Sentence-BERT (Bidirectional Encoder Representations from Transformers), developed by Nils Reimers and Iryna Gurevych (2019) is utilised to produce semantically meaningful embedding vectors from the text sources. Sentence-BERT vastly outperforms other contemporary sentence embedding methods like InferSent and Universal Sentence Encoder (Reimers and Gurevych 2019, 2). Sentence-BERT is selected for the project owing to unparalleled scalability, ability to operate without pre-processing, such as lemmatization, stemming, or stop word removal, high accuracy, and low computational overhead – specifically compared to BERT cross-encoders.

'all-MiniLM-L6-v2', a general-purpose pre-trained model released by Reimers and Gurevych for SBERT, is utilised in the code. Whilst models specifically trained for Semantic Textual Similarity (STS) tasks, like 'stsb-mpnet-base-v2', would hypothetically offer better accuracy for the present use case, they represent a significant step back in performance – from all-MiniLM-L6-v2's 14200 Sentence / Sec on V100 GPU to a mere 2800. Furthermore, the manual comparison of the output of general-purpose and STS-specific models confirmed, that the latter offers no tangible accuracy advantage for the datasets utilised in the project.

After collected text corpora have been embedded, vector similarities are calculated. The algorithm uses both cosine similarity and Euclidean distance to denote proximity between vectors, which act as numerical representations of the Stratum A and Stratum B pieces. The closer the vectors – and thus, the higher the scores – the more likely the connection between the articles.

Cosine similarity is a magnitude-independent measure, which works by identifying the angle between the vectors (Sitikhu et al. 2019), whereas Euclidean distance computes the straight-line distance between two points (Ladd 2020). Thus, in STS use cases, where vector magnitude corresponds to text length, Euclidean distance is sensitive to the length of the scrutinised texts, whereas Cosine similarity is not. Whilst, on its own, cosine similarity would be a superior measurement (Emmery 2017) – e.g. by not excluding suspect text pairs, where only a paragraph was laundered – Euclidean distance brings in additional information on the similarity between the length of examined texts. Thereby, the Core Algorithm leverages both metrics to combine their respective advantages.

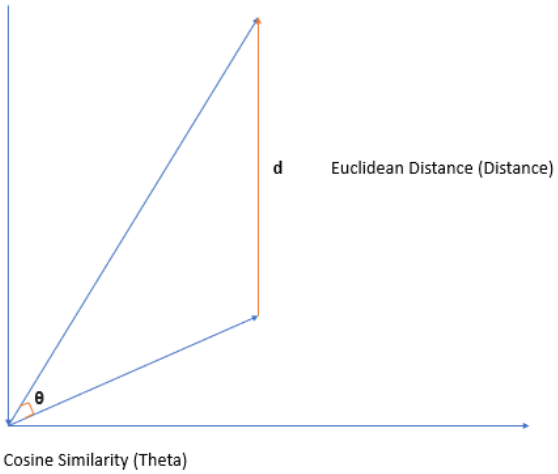

Figure 3: Euclidean Distance and Cosine Similarity Figure (Kaplan 2024)

The output of the Core Algorithm is a database of all article pairs – where one text is from Stratum A, and the other is from Stratum B corpus – between which cosine similarity and inverted normalised Euclidean distance scores are above a set threshold (see *Thresholds for the Core Algorithm* section for a detailed explanation). This dataset contains all true positives – that is, in the case of Stage I, all instances of IL between RT/Sputnik and Orosz Hírek, and in

the case of Stage II, all instances where government-affiliated media integrated a layering piece from Orosz Hírek, which originated in RT or Sputnik.

3.1.3.1 Translation

Whilst multilingual SBERT models, such as 'distiluse-base-multilingual-cased-v2', which map similar inputs in different languages close in vector space are available, tests utilising the case study's dataset confirmed, that the vectorisation of pre-translated texts, and the subsequent comparison of vectors offers not only higher performance, but better accuracy. Thus, the translation of either the RT/Sputnik pieces into Hungarian or the Orosz Hírek articles into English is necessary. Spot testing indicated that the latter solution yields better results, with both higher overall similarity scores, and higher separation between positive and negative controls. Therefore, for the project, machine comparison between the Orosz Hírek and RT/Sputnik databases was carried out with the Orosz Hírek database having been translated to English, using a parallelised version of the *deep_translator* python package. Stage II comparison, naturally, is carried out on the original Hungarian language texts.

3.1.3.2 Thresholds for the Core Algorithm

"Using the pre-trained SBERT model, completely identical texts put through machine translation yield a cosine similarity of 0.97-0.99, whilst completely unrelated texts result in scores ranging between 0.12-0.18. Whilst cosine similarity scores of -1 are mathematically possible, they are highly unlikely to occur in Natural Language Processing use cases, insofar as it would necessitate feeding diametrically opposed texts – rather than completely unrelated texts, which results in scores closer to 0.

The Euclidean distance scores used in the model have been normalised, and inverted, so that 1.0 represents identical texts and 0.0 denotes completely different texts, accounting for machine translation distortions. Euclidean distance is normalised, to account for the realities of the dataset, as follows. The minimum distance is set by computing the Euclidean distance between the vectors of an RT article, and its double English-Hungarian-English machine-translated version, simulating a hypothetical scenario, where the layering piece is an unedited machine translation." (Kovács 2024, 4) Maximum distance is set by computing the Euclidean distance between the vectors of an RT piece, and a completely unrelated English sentence.

In the testing phase, thresholds for Cosine similarity and inverted normalised Euclidean distance scores were determined by setting a low, 0.5 cut-off for both scores in the algorithm, then manually moving through the results database in 0.05 iterations for both scores, to find the genuinely laundered article pairs (true positive) with the lowest scores in each category – this in turn was used as a threshold for live operation.

The most adequate cut-off, using ENG-ENG pairs for RT/Sputnik → Orosz Hírek measurements, and original HUN-HUN pairs for Orosz Hírek → government-affiliated media measurements, was determined to be 0.67 for the inverted normalised Euclidean distance score and 0.83 for the cosine similarity score. This threshold ensures the least number of false positives, whilst not excluding any potentially laundered article pairs – true positives.

3.1.3 Filtering

3.1.3.1 Date Filter

To reduce computational overhead, the Core Algorithm already integrates a date filter, whereby only article pairs, where the Stratum B article was published no more than three days after a Stratum A piece respectively, are included. Whilst manual checks confirmed, that in the

vast majority of IL instances, the entirety of the laundering process takes no more than 36 hours, a more inclusive 3-day cut-off is used to account for potential outliers, primarily to prevent the exclusion of the minority case, where a weekend delays the layering or the integration phase.

3.1.3.2 Eleven-Feature Machine Learning Filter

Whilst the Core Algorithm performs well in identifying true positives, approximately 20% of the output data is comprised of false positives owing to the inherent limitations of cosine similarity and Euclidean distance, thus warranting further filtering. Owing to the size of the dataset – and especially considering that the value proposition of the present methodology lies in its adaptability to larger, more diverse datasets – the manual removal of false positives remains an unviable approach, necessitating the deployment of automated filtering.

For this, a multi-feature Machine Learning approach is selected, as described by Emil Kalbaliyev and Samir Rustamov (2021, 9–19). First, the text pairs identified by the Core Algorithm are subjected to further similarity analysis, bringing in secondary comparative metrics. The present filtering methodology leverages the seven character-based similarity measurements used by Kalbaliyev and Rustamov and complements it with an additional four word-level similarity metrics so that to feed a more diverse set of data points to the Machine Learning algorithm.

Stratum A and Stratum B text pairs are rated with the character-based similarity measurements Longest Common Substring (LCS), Longest Common Subsequence (LCSq), Gestalt pattern matching (also known as Ratcliff/Obershelp pattern recognition), Levenshtein Distance, Damerau-Levenshtein distance, Jaro distance, and Jaro–Winkler distance. Additionally, the word-level similarity measurements Jaccard Similarity, TF-IDF Similarity, N-gram similarity using trigrams, and unique proper noun count are utilised.

Longest Common Substring (Shimohira et al. 2011), as used in a Semantic Text Similarity context, identifies the longest, contiguous string of text – to a character level – shared between a Stratum A and a Stratum B text paired by the Core Algorithm. The filter then records the length of the longest common substring. A higher number indicates that a Stratum A and a Stratum B text share longer contiguous segments, such as full sentences. Since LCS is an inflexible measurement, and thus, highly sensitive to variations introduced by either the initial translation or the operative re-translation, it, like other metrics introduced in the filtering stage, can only be utilised as secondary measurements in the present model.

Unlike LCS, Longest Common Subsequence does not require the sequence to be contiguous, thereby permitting matches between sequences with non-shared characters interrupting them. This is particularly useful in accounting for cases, when, for instance, an extra word has been added during the laundering process. Following Kalbaliyev and Rustamov, to extract a similarity measurement from LCSq, the length of the Longest Common Subsequence is doubled and divided by the sum of the length of the two texts (2021, 14).

Gesalt pattern matching, or Ratcliff/Obershelp pattern recognition (Ratclif and Metzener 1988) constitutes an iteration of LCS. Ratcliff/Obershelp first finds the LCS between the text pairs, and then iteratively works on the residual characters to find shorter and shorter matching strings, until there are no common characters. The similarity metric is then derived using the same method as with LCSq.

Levenshtein Distance (Levenshtein 1965) calculates the number of character-level changes required to transform a Stratum A text into its respective Stratum B pair. Under Levenshtein Distance, allowed edits are insertion, deletion, and substitution. Damerau-Levenshtein distance, in addition to the three aforementioned operations, allows for the transposition of two adjacent characters. The lower the number of operations required, the closer the two texts are.

Jaro similarity (Jaro 1989) is a character level metric, which measures the number and order of shared characters as well as the number of transpositions. First, Jaro similarity identifies matching characters – not necessarily in the same order – which are closer to each other in the string than a set maximum distance, derived from the total length of the strings. Then, the algorithm calculates the number of transpositions needed to rearrange matching characters which are in a different order between the two strings. Jaro–Winkler distance (Winkler 1990) is a variation of Jaro similarity – the prior gives preference to strings which match from the beginning.

Jaccard Similarity (Niwattanakul et al. 2013; Bag, Kumar, and Tiwari 2019) works by dividing the total number of words shared between the two texts, by the number of words in the union of the two texts. N-gram similarity using trigrams, as implemented in the current code, also computes Jaccard similarity but uses word trigrams as a basis for comparison, rather than individual words.

TF-IDF, which denotes the importance of a word to a document in a corpus, is a derivative of two measurements, Term Frequency and Inverse Document Frequency (Leskovec, Rajaraman, and Ullman 2014, 8–10). Term frequency, in its raw count application used in the current algorithm, signifies the number of times a word appears in a text, whilst Inverse Document Frequency computes the log of the ratio of the total number of documents to the number of documents that contain the word. TF-IDF is then calculated as TF x IDF. As implemented in the algorithm, a TF-IDF-specific vectorizer first tokenises the texts into words, transforms the texts into numerical representations with TF-IDF weighting, and then cosine similarity is used to denote the proximity between the two suspect texts.

The manual coding process revealed that a particularly informative metric to indicate whether two suspect pieces constitute part of an IL chain is to examine, if the Stratum B text contains any specific information, such as names or places, which are not found in the Stratum

A text. To automate this approach, the unique proper noun count metric is introduced, which uses the spaCy library to identify proper nouns in each text and return the number of proper nouns exclusively present in the Stratum B text from suspect pairs. The lower the count, the more likely the connection. Stage I uses the en_core_web_sm model to find English proper nouns, whilst, in Stage II, the Hungarian-language hu_core_news_lg model is utilised.

After the eleven secondary metrics had been calculated, a subset of Stage I's output appended with the secondary metrics – approximately 800 article pairs – was manually coded, to differentiate between true positives ('1') and false positives ('0'). The manually coded dataset in turn serves as input for a Machine Learning algorithm.

In a departure from Emil Kalbaliyev and Samir Rustamov (2021) – who manually identified the best model and parameters for the Machine Learning process – the coded dataset is fed into an Automated Machine Learning (AutoML) process, utilising the Tree-based Pipeline Optimization Tool library (Le, Fu, and Moore 2020). AutoML identifies the most optimal analysis pipeline, by selecting the best performing model and the most appropriate parameters for the given training data (Ebrahim and Joy 2023).

The identified pipeline is then used to train a model on the coded Stage I data, which in turn allows us to automatically code – and thus filter – new output data produced by the Core Algorithm. On the datasets examined in the current study, the trained Machine Learning model separates true positives from false positives with above 99,2% accuracy, based on cosine similarity and Euclidean distance, coupled with the secondary metric set.

3.2 Findings and Analysis

The investigation uncovered a Kremlin-orchestrated IL scheme moving specific pieces of propaganda from RT/Sputnik (placement) via Orosz Hírek (layering) into Hungarian government-affiliated media (integration), including Mandiner, Origo, Vadhajtások, and

Magyar Hírlap. We find that Orosz Hírek layered a total of 2,824 articles from RT and Sputnik between January 1, 2022, and March 10, 2024 – rendering it a direct copy-paste site to RT and Sputnik. In the same period, four Hungarian government-affiliated outlets integrated 891 articles which originally appeared in the columns of RT or Sputnik and penetrated the Hungarian language media ecosystem with the mediation of Orosz Hírek. The following paragraphs review the findings in detail and examine case studies to highlight key characteristics of the laundering scheme.

One of these characteristics, termed *Layering-Integration Liminality,* is a yet unnoted phenomenon pertinent to IL, observed and conceptualised in the current study. "In virtually all IL models (Korta 2018; Meleshevich and Schafer 2018; Rodríguez 2021b; 2021a) malign *distortion* of the information exclusively occurs in the layering and in certain cases, placement domains, usually carried out by fringe, Kremlin official, or Kremlin-proxy outlets, whilst the integration phase merely denotes the adoption of the laundered information into mainstream media. In contrast, Hungarian government-affiliated outlets oftentimes serve the dual purpose of *layering-integration,* whereby they not only complete the laundering cycle by integrating an already-layered piece of information into the mainstream media ecosystem but at the same time complement it with further malign distortions." (Kovács 2024, 2–3)

3.2.1 Orosz Hírek

Orosz Hírek has laundered a total of 2,824 articles from RT and Sputnik between January 1, 2022, and March 10, 2024, injecting them into the Hungarian media ecosystem. Despite Sputnik's comparably sized larger corpus – 52,647 versus RT's 46,558 – the editors of Orosz Hírek appear to give overwhelming preference to sourcing articles from RT. Of Orosz Hírek's 2,824 laundered pieces, 2,731 were originally published on RT.

Given Orosz Hírek's total corpus size of 10,253, 27.5% of Orosz Hírek's content in the examined period is from sanctioned Russian state-affiliated propaganda sources. Thereby, arguably, Orosz Hírek can be classified as RT's 'copy-paste website'.

An OSINT investigation by the Institute for Strategic Dialogue uncovered 112 copy-paste websites of RT operating in Europe, which have been utilised to circumvent sanctions following the outbreak of the full-scale invasion (Balint et al. 2022). Copy-paste and mirror sites – the latter of which relay not only RT's content but copy the entire page structure – have proliferated since March 2022. Whilst Oroszhírek, according to domain registration data, has been operational since April of 2009 (Whois Record, 2024), current findings indicate that since the start of hostilities in autumn 2021, and especially from the onset of sanctions against RT and Sputnik in March 2022, it has acquired a new function, and correspondingly elevated its operation to a new level. To provide further information on the trajectory of Orosz Hírek's activities, articles published in 2021 on RT, Sputnik, and Orosz Hírek were collected and processed through the steps described in subsections Core Algorithm and Filtering. These, however, are not included in the main dataset, and are only examined to elucidate Orosz Hírek's transformation. The chart below shows the total number of articles and the number of laundered pieces sourced from RT and Sputnik published on Orosz Hírek between January 1st, 2021, and March 10th, 2024.

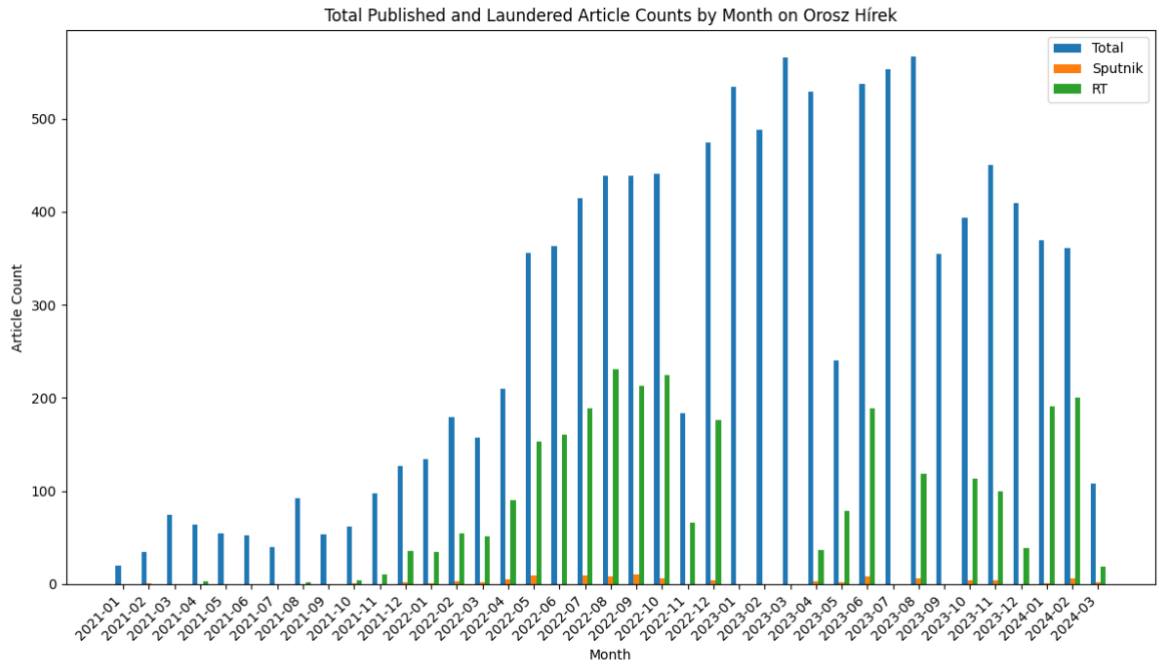

Figure 4: Total Published and Laundered Article Counts by Month on Orosz Hírek

The data indicates that Orosz Hírek has embarked on a steep rise in publication frequency from October 2021 onwards – notably corresponding with the second Russian military build-up in November 2021 (Rainsford 2021). From a monthly publication count average of 64 in 2021, Orosz Hírek achieved an average monthly publication rate of 316 and 467 in 2022 and 2023 respectively.

Coinciding with the uptick in publication rate, we observe the *en masse* introduction of laundered RT – and to a considerably lesser extent, Sputnik – articles to Orosz Hírek, especially in the wake of EU sanctions on RT and Sputnik in March 2022. Whilst before October 2021, laundered pieces were virtually unseen in Orosz Hírek's corpus, by August 2022, 52% of all articles published on Orosz Hírek have been layered from RT, without disclosing the source. Notably, we can observe a complete, unexplained cessation of the laundering operation between January and March 2023, and a partial resumption from April onwards. Since the end of 2023, the ratio of laundered content in Orosz Hírek has surpassed its previous zenith,

reaching 55% in February 2024 – meaning that a majority of Orosz Hírek's content were translated RT articles.

This information, coupled with the findings presented in the 'Investigation into Orosz Hírek' section in Chapter II, and especially the fact Orosz Hírek's content-providing partner is Ruptly – a subsidiary of RT –, hints at a possible arrangement between the operators of RT and Orosz Hírek, which provides for the translation and republishing of RT articles in Hungarian, and by extension, sanctions evasion. Orosz Hírek's overwhelming preference for RT over Sputnik – contrasting with Rodríguez's findings (2021b, 19), which identified Sputnik as the most active Kremlin-official outlet in IL – further implies the existence of an RT-Orosz Hírek arrangement.

"We spit on your sanctions" RT Editor-in-Chief Margarita Simonyan said (quoted in Nardelli, Deutsch, and Bodoni 2023), in a comment for Bloomberg's coverage on RT's sanction evasion schemes utilising mirror and copy-paste sites – it appears that Orosz Hírek is a part of Simonyan's efforts in this regard.

Whilst – owing to its prominence in the Hungarian language pro-Russian media ecosystem – several investigations (Szicherle et al. 2019; Győri, Molnár, et al. 2022; Német 2023) have probed Orosz Hírek, to date, no study has identified the outlet as a direct copy-paste site to RT and Sputnik.

3.2.2 Mandiner

In the observed period, Mandiner has integrated 114 articles from Orosz Hírek, 26 of which were originally sourced from sanctioned Russian propaganda outlets. Recalling the limitations of the model, 114 represents a lower level of confidence, whilst 26 constitutes evidence-approximate data. Mandiner does not reference Orosz Hírek as the source of the information in any of these cases.

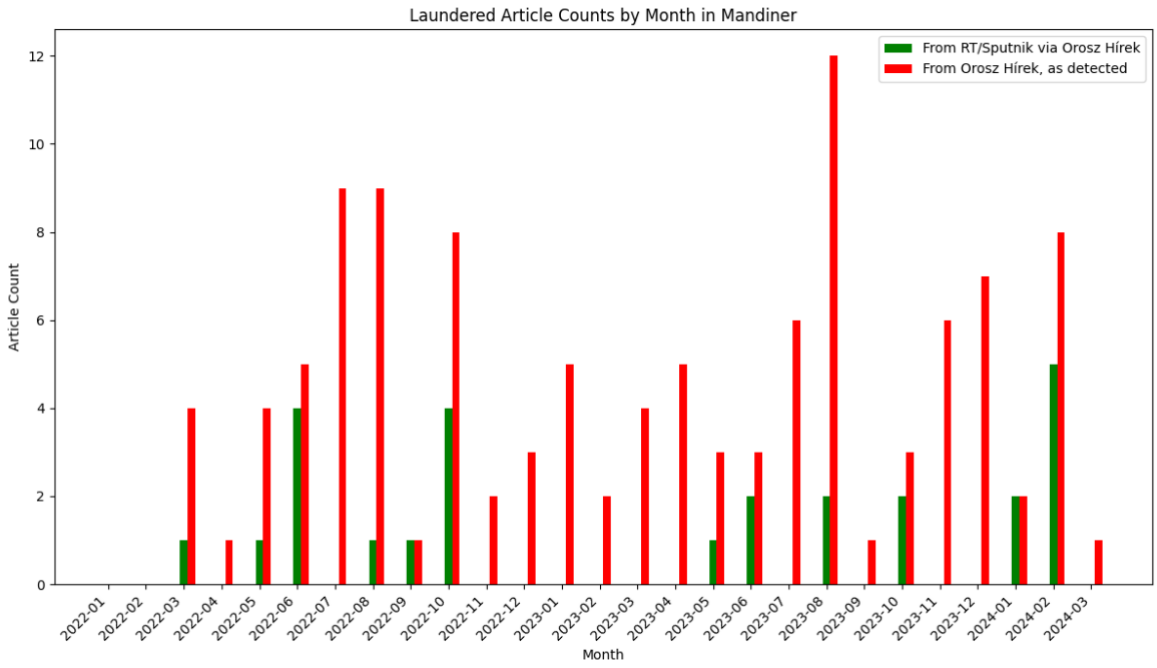

Figure 5: Laundered Article Counts by Month in Mandiner

In the most notable case – as outlined in the introduction – Mandiner published the false claim fabricated by RT and translated by Orosz Hírek, that the Independent International Commission of Inquiry on Ukraine's 2022 October report asserts, that that "while crimes committed by Ukrainians are '*supported by real evidence, such as videos*,' alleged violations by Russian forces rely mostly on eyewitness accounts"[12] (Mandiner 2022b, emphasis in original). Based on the layering piece published by Orosz Hírek on October 18, 2022, titled "Ukrainian war crimes reported by the UN"[13] (Orosz Hírek 2022), Mandiner published not only a near-verbatim copy (Mandiner 2022b) but also a broader analytical piece referencing the initial integration article a day later (Zoltán 2022). Citing Mandiner's article, four prominent government-affiliated outlets have republished the story: Origo (2022b),

[12] Original: „hogy míg az ukránok által elkövetett bűncselekményeket *„valódi bizonyítékok, például videók támasztják alá"*, addig az orosz erők által elkövetett állítólagos jogsértések többnyire szemtanúk beszámolóira támaszkodnak."

[13] Original: "Ukrán háborús bűnökről közölt jelentést az ENSZ"

Vadhajtások (2022a), 888 (2022), and PestiSrácok (2022). Whilst a Hungarian investigative journalist hypothesized that the article was sourced from Orosz Hírek (Teczár 2022), no investigation to date has uncovered, that in fact, it is a laundered RT piece.

Figure 6: Articles in the IL Chain, Mandiner #1

"Mandiner's publication showcases *Layering-Integration Liminality* in operation. Whilst both RT's placement and Orosz Hírek's layering piece only added that violations are 'backed by hard evidence', whilst Russian war crimes 'mostly rely on the accounts of witnesses' as contextual – albeit false – information, Mandiner already presents this narrative as a detail from the UN report, adding that the differentiation between the quality of pieces of evidence is 'according to the document'. Thereby, it can be argued, that according to Rodriguez's framework of IL techniques, the integration move – uniquely – also comprised an element of layering by means of *misappropriation,* insofar references that do not contain the alleged information were provided (Rodríguez 2021b, 12)." (Kovács 2024, 7)

The Mandiner dataset highlighted another tool, *fully false citations,* utilised in IL. In an article published on October 9, 2022, titled "Putin: Ukraine is behind the attack on the Crimean bridge"[14] (2022c), Mandiner explicitly claims to be consulting reporting on the event provided by the BBC. However, as textual analysis confirms, the Mandiner article is a near-verbatim copy of an Orosz Hírek piece of the same title, published a mere 40 minutes earlier, which in turn layered an RT article – suggesting that Mandiner, in fact, consulted Orosz Hírek, rather than the BBC.

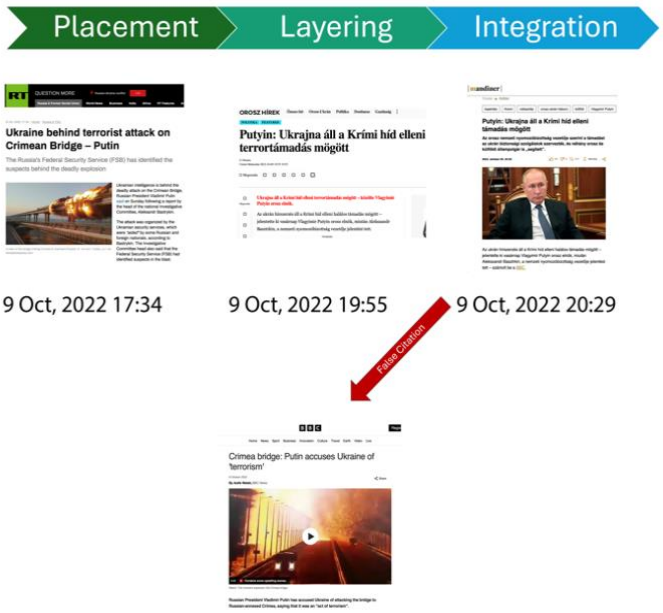

Figure 7: Fully false citation, Mandiner #2

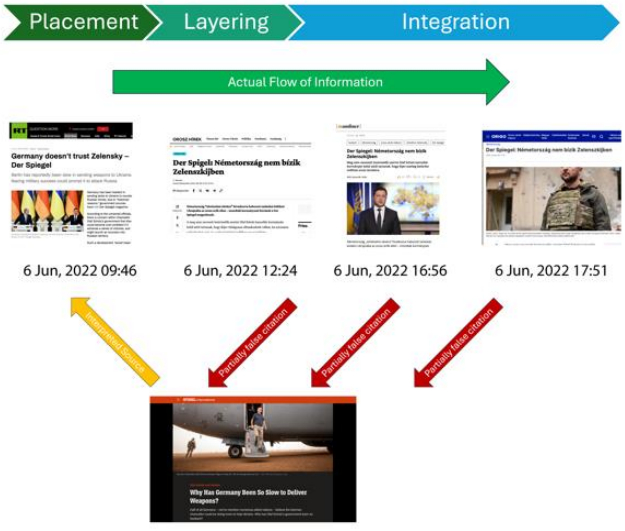

Figure 8: Partially false citations, Mandiner #3

Another, related, but more frequently observed phenomenon is termed *partially false citations.* In these cases, the placement piece on RT or Sputnik references an external, usually Western source. The citation then in turn carries over to the layering piece, and oftentimes into the integration piece as well. However, the

[14] Original: „Putyin: Ukrajna áll a Krími híd elleni támadás mögött"

downstream Stratum B texts do not – presumably, given the verbatim translations, copies, and matching titles – consult the original, western source, but relay its interpretation and framing as presented by Kremlin propaganda. In the case illustrated above, RT interpreted Der Spiegel International's reporting – "Why Has Germany Been So Slow to Deliver Weapons?" (Amann et al. 2022) – in an article titled "Germany doesn't trust Zelensky - Der Spiegel" (RT 2022a). The RT piece was in turn layered by Orosz Hírek (2022a) – "Der Spiegel: Germany does not trust Zelensky",[15] and integrated by Mandiner (2022a) – "Der Spiegel: Germany does not trust Zelensky".[16] Mandiner's article was then republished by Origo (Origo 2022a) – "Der Spiegel: Germany does not trust Zelensky".[17] Notably, the titles stayed completely intact in the IL chain from RT to Mandiner and Origo.

3.2.3 Magyar Hírlap

Magyar Hírlap is not only the most prolific launderer of sanctioned Kremlin propaganda amongst the government-affiliated media scrutinised in the current project, but one of the two which does so – in some cases – with reference. Between January 1, 2022, and March 10, 2024, Magyar Hírlap integrated 2,889 articles from Orosz Hírek. Out of these, it explicitly attributes 2,208 to the fringe pro-Kremlin outlet.

Notably, these include both Orosz Hírek's genuine content and layering articles. This in turn implies, given Orosz Hírek's total corpus size of 10,253, that Magyar Hírlap has integrated 28% of the pro-Russian fringe site's publications.

[15] Original: „Der Spigel: Németország nem bízik Zelenszkijben"
[16] Original: „Der Spigel: Németország nem bízik Zelenszkijben"
[17] Original: „Der Spigel: Németország nem bízik Zelenszkijben"

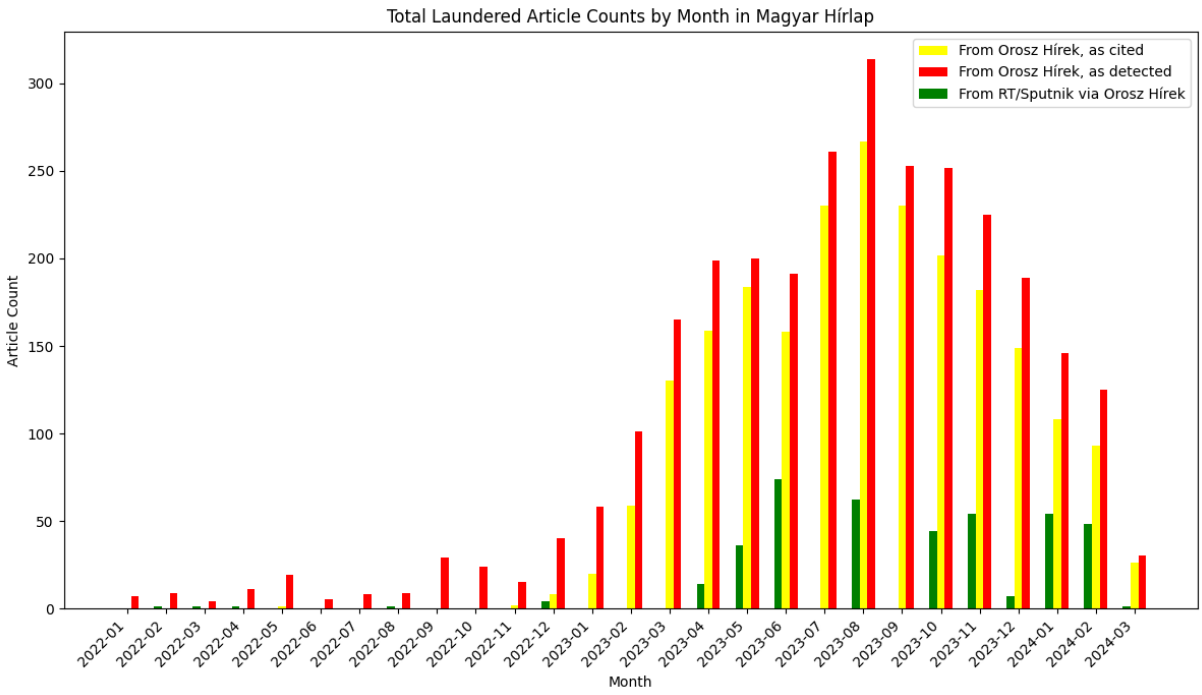

Figure 9: Total Laundered Article Counts by Month in Magyar Hírlap

In the observed period, Magyar Hírlap published 402 articles which originally appeared on the columns of RT or Sputnik and were layered via Orosz Hírek. Given 2,889 total published Orosz Hírek articles in Magyar Hírlap, out of which 13.9% originated either in RT or Sputnik, it appears that Magyar Hírlap gives preference to Orosz Hírek's genuine content, insofar as 27.5% of Orosz Hírek's corpus represents laundered information.

Out of the 402 articles, 368 still carry a citation to Orosz Hírek. Since one of the core aims and characteristics of IL is *loss of origin* (see Shekhovtsov 2015, 5–6), Magyar Hírlap's citation practices could potentially complicate classification. Nonetheless, since, in every case, the loss of origin was already provided for by the layering phase – insofar as Orosz Hírek always foregoes referencing RT or Sputnik – it can be argued that Magyar Hírlap has provided for the integration phase of an IL scheme in 402 cases.

On July 7th 2023, Magyar Hírlap (2023c) has without commentary or reference integrated an RT article (2023d) citing Putin's claims that "Kyiv turns a blind eye neo-Nazi ideology". This example illustrates the functioning of what the author would term 'embedded propaganda', whereby the statements of Kremlin officials carry with them the propaganda or disinformation content, and in turn, these assertions are reported on by domestic media "at face value, without anything to balance their statements" (Szicherle et al. 2019, 18). In the context

Figure 10: Articles in the IL Chain, Magyar Hírlap #1

of IL chains observed in the present case studies, these comments are in turn relayed together with Kremlin propaganda outlets' verbatim interpretation and framing.

In other cases, similarly to the first Mandiner case study highlighted, Magyar Hírlap introduced fabrications appearing on RT into the mainstream Hungarian government-affiliated media ecosystem. For instance, in an article published on May 15th 2023, 8:53, RT, citing the "local authorities and the Russian military" claims that British Storm Shadow cruise missiles were "used to attack civilian targets in Lugansk" (RT 2023b). Whilst LNR authorities' press release, which, contrary to RT's reporting, maintained that two Ukrainian-made "grom" missiles were used in the attack (Представительство ЛНР в СЦКК, 2023), already cast a shadow on the veracity of RT's statements, EUvsDisinfo's investigation (2023) conclusively

debunked the narrative as disinformation. Via the mediation of Orosz Hírek (2023a), Magyar Hírlap integrated the article a few hours later – with the false claim included (2023a).

Figure 11: Articles in the IL Chain, Magyar Hírlap #2

Magyar Hírlap oftentimes serves as an entry point (Német 2023) for sanctioned Russian propaganda into higher-reach government-affiliated sites. For instance, an RT article (2023a) reporting on Russian Deputy Foreign Minister Aleksandr Grushko's assertion, that "The entire losses the EU sustained due to the imposition of sanctions and the decisions to curtail cooperation with Russia have totalled, at a conservative estimate, some $1.5 trillion," layered via Orosz Hírek (2023b) has appeared on the columns of Magyar Hírlap (2023b), which in turn provided the basis for Origo's (2023b) reporting on the topic. Origo's publication also showcases *Layering-Integration Liminality*, insofar as it uses the title "Huge disaster coming to Europe"[18] thereby introducing further malign distortions.

[18] Original: „Óriási katasztrófa jön Európában"

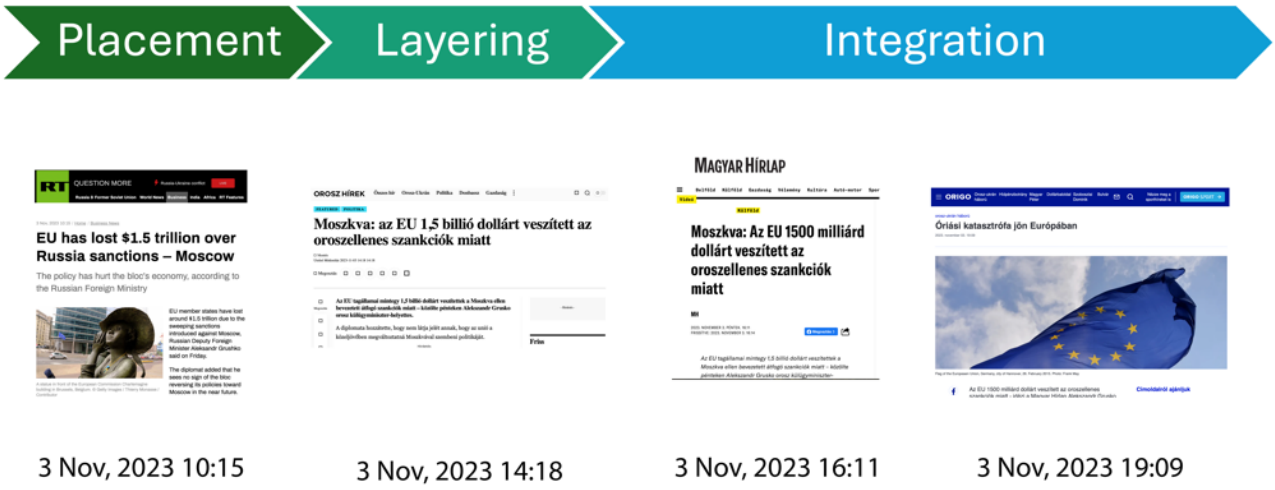

Figure 12: Articles in the IL Chain, Magyar Hírlap #3

3.2.4 Origo

Despite its noted tendencies for broadcasting pro-Kremlin propaganda (Diószegi-Horváth 2022; Fülöp 2024), Origo appears not to play an outsize role in the IL scheme uncovered by the investigation – between January 1, 2022, and March 10, 2024, Origo integrated 94 articles from Orosz Hírek, 11 out of which originated on RT or Sputnik.

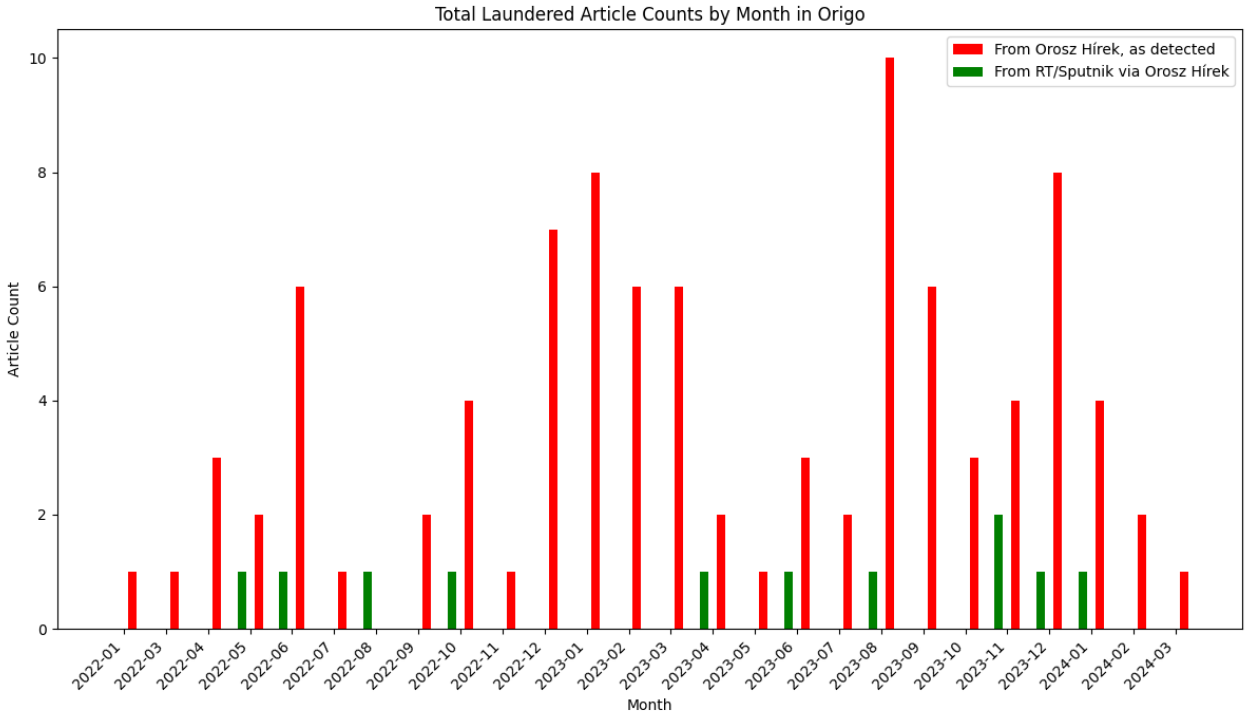

Figure 13: Total Laundered Article Counts by Month in Origo

A case study from the Origo dataset reflecting both *partially false citation* and *Layering-Integration Liminality* exemplifies how the framing of Western reportage offered by Kremlin propaganda penetrates the Hungarian media ecosystem via Orosz Hírek, and in turn, becomes the dominant interpretation in government-affiliated media. On the 4th of August, 2023, Politico published an article titled "Eight Ukrainian pilots ready to train on F-16s" (Mcleary, Ward, and Berg 2023),

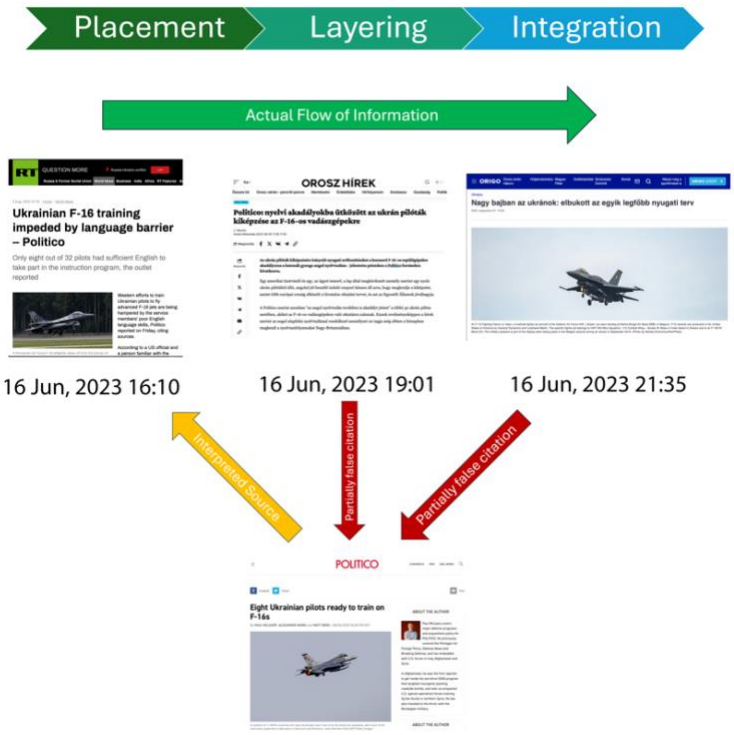

Figure 14: Partially false citation, Origo #1

which details that in addition to the eight pilots enrolled in the training programme, a further

20 will partake in English language courses in preparation for the training. On the 5th of August, RT interpreted Politico's reporting in an article titled "Ukrainian F-16 training impeded by language barrier – Politico" (2023c), which in turn was layered by Orosz Hírek a few hours later, utilising the title "Politico: language barrier impedes the training Ukrainian pilots for F-16 fighters"[19] (2023c). On the 7th of August, Origo integrated the Orosz Hírek article, using a near-verbatim copy of the fringe outlet's text. The title underwent further malign distortions, showcasing *Layering-Integration Liminality*: "The Ukrainians are in big trouble: one of the main Western plans has failed"[20] – Origo claims (2023a), in reference to Ukrainian pilots undergoing language instruction.

3.2.5 Vadhajtások

The grey zone government-affiliated news site Vadhajtások integrated a total of 1,628 articles from Orosz Hírek, 452 out of which were originally sourced from sanctioned Russian propaganda outlets. The 27.7% ratio between genuine and RT/Sputnik-sourced Orosz Hírek articles in Vadhajtások provides a near exact match to Orosz Hírek's own genuine-laundered ratio of 27.5%, implying that Vadhajtások integrated a representative sample.

Vadhajtások, like Magyar Hírlap, does, in a few cases, disclose that it is sourcing information from Orosz Hírek – albeit to a much lesser extent. Vadhajtások attributes 172 articles from its 32,573-article corpus to Orosz Hírek. Of the 452 RT-sourced pieces, 87 cite Orosz Hírek.

[19] Original: „Politico: nyelvi akadályokba ütközött az ukrán pilóták kiképzése az F-16-os vadászgépekre"
[20] Original: „Nagy bajban az ukránok: elbukott az egyik legfőbb nyugati terv"

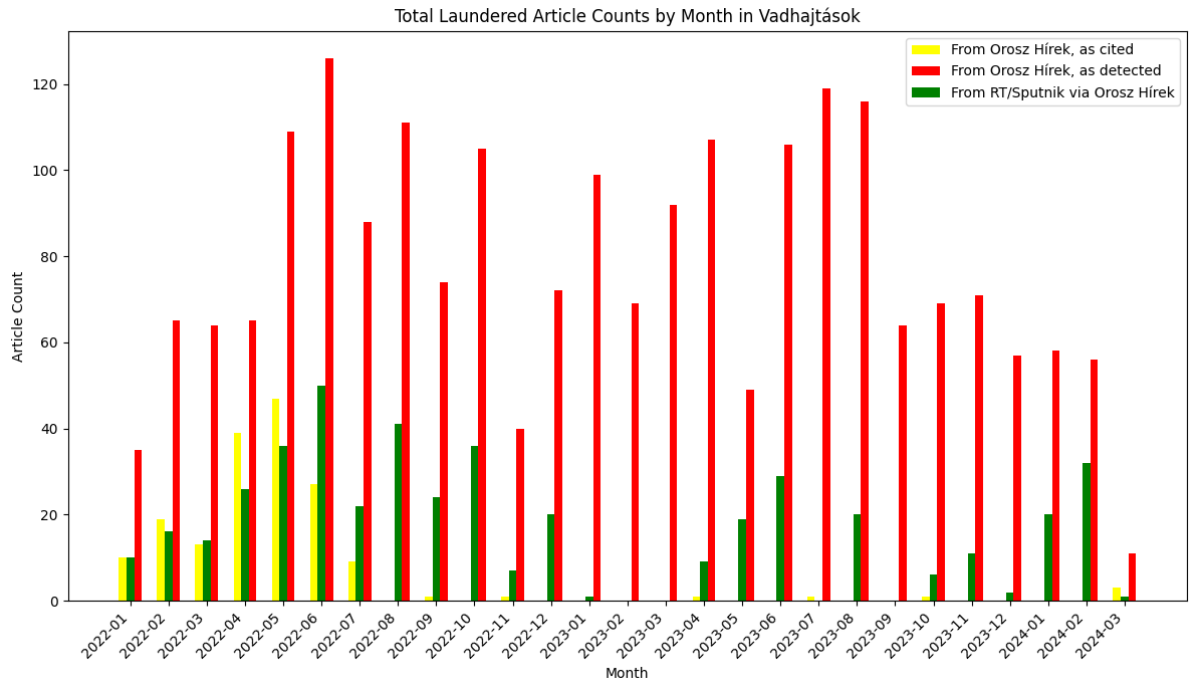

Figure 15: Total Laundered Article Counts by Month in Vadhajtások

The singular RT piece appearing in Vadhajtások in January 2023 might seem anomalous, insofar as Orosz Hírek layered no material from RT/Sputnik between January and March 2023. However, the article under scrutiny was published on 1st January 2023, and it integrated a layering piece published by Orosz Hírek on 31st December 2022, which in turn copied RT's placement piece released a few hours prior.

Vadhajtások also exhibits *Layering-Integration Liminality* – albeit in a peculiar fashion. For an article published (Vadhajtások 2022b) on June 20, 2022, 13:30 – originally released on RT (2022c) a

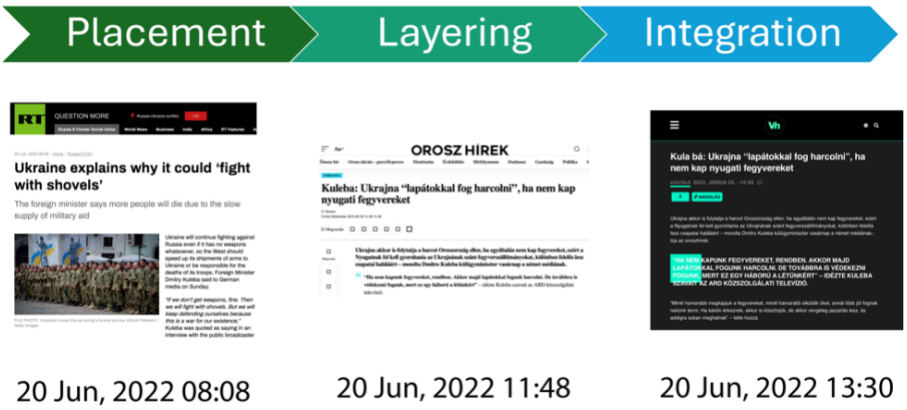

Figure 16: Articles in the IL Chain, Vadhajtások #1

few hours earlier – Vadhajtások modifies the title from the layering piece to include additional connotations. Whilst the titles of the article, as appearing on Orosz Hírek (2022b) and RT, mention Ukrainian Minister of Foreign Affairs Dmytro Kuleba by name, Vadhajtások adopts the derogatory moniker "Kula Bá" ("Uncle Turd"). Using the moniker for Kuleba appears to be Vadhajtások's *modus operandi* – it features in nine more articles sourced from Orosz Hírek.

Partially false citations are also prevalent in IL chains completed by Vadhajtások, as shown in Figure 17.

Figure 17: Partially false citations, Vadhajtások #2

3.2.6 Hirado.hu

Whilst the Hungarian public broadcaster's news channel and corresponding website have repeatedly been found to be hosting Kremlin propaganda talking points (Szicherle et al. 2019, 15; Zubor 2022; Bayer 2022; Presinszky 2022; Lakmusz 2022; Keller-Alánt 2022; Rényi 2022), the current algorithmic investigation found that Hirado.hu integrated no articles from Orosz Hírek between January 1 2022, and March 10, 2024.

Chapter III has demonstrated the efficacy of the novel Natural Language Processing-based methodological framework, uncovered a Kremlin-orchestrated IL scheme, and provided novel evidence showing that Hungarian fringe and government-affiliated media do provide for the layering and integration phases of Kremlin IL campaigns. These findings pave the way for the following, final section, which discusses the broader implications of these results and proposes actionable steps for countering IL.

Conclusions

The present paper developed a reusable and adaptable Natural Language Processing and Machine Learning-based methodological framework building on Rodríguez's (2021a; 2021b) findings, which can be deployed to detect and trace digital IL, and scaled to provide for the near real-time "systematic analysis" (Toucas 2017) of targeted media ecosystems – which so far has been out of reach for models outlined in the existing literature.

To test the framework's efficacy, and ascertain, whether Hungarian fringe and government-affiliated media provide for the layering and integration phases of Russian IL campaigns, the paper deployed the novel algorithmic methodological framework on a 376,760-article sample from the Hungarian government-affiliated media ecosystem and Orosz Hírek's 10,254-article corpus. We identified a yet unnoted Kremlin-orchestrated IL scheme, that successfully subverted Hungarian government-affiliated media. Orosz Hírek plays a key role in the scheme: it implants propaganda into the Hungarian-language media ecosystem, whilst masking its true source. We found that in the context of the full-scale invasion, between January 1, 2022, and March 10, 2024, four Hungarian government-affiliated outlets have integrated a total of 891 articles, which originally appeared on the columns of RT or Sputnik, and penetrated the Hungarian language media ecosystem with the mediation of the fringe, Kremlin proxy site, Orosz Hírek – which itself layered a total of 2,824 articles from RT and Sputnik. Thus, in addition to the provision of new evidence vis-à-vis the Hungarian government-affiliated media ecosystem, the current paper is the first to identify Orosz Hírek as a direct copy-paste site to RT and Sputnik.

Notably, given that Orosz Hírek has already provided for the *loss of origin,* the editors of the four implicated government-affiliated outlets most likely had no awareness, that the articles they are republishing were in fact authored by RT and Sputnik, and thereby, did not

know, that they are being instrumentalised by a Kremlin-orchestrated IL scheme. In IL investigations, the mainstream, reputable media outlets bringing laundered information into legitimate public discourse, are usually framed as *accidental actors* (Rodríguez 2021b, 20, 70, 86), who become unaware participants of IL schemes. They are "exploited by the architects of HIICs*"* (Rodríguez 2021b, 14) and "unwittingly enable the spread of Kremlin influence in the country" (Rodríguez 2021b, 43). However, the author would argue that Hungarian government-affiliated media's general proclivity towards broadcasting Russian propaganda, their *en masse* participation in the scheme uncovered by the investigation, and the frequent resort to *Layering-Integration Liminality* precludes their classification as merely *accidental actors.* Instead, they are best understood as *manipulated allies,* whose vested interests in purveying Russian propaganda coincide with the aims of the Kremlin-orchestrated IL scheme.

"Hungarian public media disseminates Russian war propaganda *as if* it were edited directly from Moscow"[21] posits Péter Erdélyi (2016, emphasis added). The current investigation has proven, that in fact, certain materials published by Hungarian government-affiliated media – although not public media – not merely appear *as if* they were edited directly from Moscow, but they *are* edited in Moscow.

The next step is policy action. The findings presented in Chapter III provide for the basis of two immediate legal avenues for countering IL, undergirded by Article 2f (1) of Council Regulation (EU) 2022/350 and Article 4g of Council Decision (CFSP) 2022/351. First, since the regulations also pertain to the distribution of RT and Sputnik content via social media and by search engines, Google and Facebook – the latter of which hosts Orosz Hírek's social media site with nearly 190,000 followers – are obliged to take restrictive action, contingent on

[21] Original: „A magyar közmédia úgy terjeszti az orosz háborús propagandát, mintha közvetlenül Moszkvából szerkesztenék"

the author's escalation report. Google and Facebook's obligations are further reinforced by the EU's corresponding co-regulatory framework, consisting of the Digital Services Act (DSA) and the Code of Practice on Disinformation.

Second, since pursuant to Article 4g, which stipulates that it is "prohibited for operators to broadcast or to enable, facilitate or otherwise contribute to broadcast[ing]" RT and Sputnik content, KESMA[22], Magyar Hírlap Kiadói Kft.[23], and the publisher of Vadhajtások are in breach of EU law. A complaint to the European Commission, alleging the distribution of Kremlin propaganda by the Duna Media Service Nonprofit Private Limited Company – the operator of Hungarian public service media – is already in progress, submitted by Political Capital and the Hungarian Civil Liberties Union (Társaság a Szabadságjogokért 2022). The author hopes that the findings of the present investigation will contribute to the reinforcement and expansion of the complaint.

"Győri et al. argue that in Hungary, 'pro-Russian disinformation comes *directly* from government-organized media' (2017, emphasis added). The present study has proven, that even more so, part of the pro-Russian propaganda published in the Hungarian government-affiliated media is authored *directly* by Kremlin media. Indeed, the layer dividing Margarita Simonyan's office and Hungarians' hearts and minds is more permeable than ever, and it will remain so in the absence of [near real-time detection systems, and stringent policy responses.]" (Kovács 2024, 9)

[22] The publisher of Mandiner and Origo
[23] The publisher of Magyar Hírlap

References

888. 2022. ‘Egy ENSZ-Jelentés Szerint Az Ukrán Hadsereg Háborús Bűncselekményeket…’. *888.Hu (Archived)*, 2022. https://archive.md/dFj5P.

Amann, Melanie, Markus Becker, Markus Feldenkirchen, Florian Gathmann, Matthias Gebauer, Serafin Reiber, Jonas Schaible, Christoph Schult, and Severin Weiland. 2022. ‘Olaf Scholz and Ukraine: Why Has Germany Been So Slow To Deliver Weapons?’ *Der Spiegel*, 3 June 2022, sec. International. https://www.spiegel.de/international/germany/olaf-scholz-and-ukraine-why-has-germany-been-so-slow-to-deliver-weapons-a-7cc8397b-2448-49e6-afa5-00311c8fedce.

Andreae, Jonas. 2022. ‘“The Information War of RT DE”’. In *RT in Europe and beyond The Wannabe Elite of the Anti-Elites*, edited by Anton Shekhovtsov. Vienna: Centre for Democratic Integrity.

Arora, Simran, Avner May, Jian Zhang, and Christopher Ré. 2020. ‘Contextual Embeddings: When Are They Worth It?’ In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, edited by Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault, 2650–63. Online: Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.236.

Badawy, Adam, Emilio Ferrara, and Kristina Lerman. 2018. ‘Analyzing the Digital Traces of Political Manipulation: The 2016 Russian Interference Twitter Campaign’. In *Proceedings of ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 258–65. Barcelona, Spain: IEEE. https://doi.org/10.1109/ASONAM.2018.8508646.

Bag, Sujoy, Sri Krishna Kumar, and Manoj Kumar Tiwari. 2019. ‘An Efficient Recommendation Generation Using Relevant Jaccard Similarity’. *Information Sciences* 483:53–64. https://doi.org/10.1016/j.ins.2019.01.023.

Balint, Kata, Francesca Arcostanzo, Jordan Wildon, and Kevin Reyes. 2022. ‘RT Articles Are Finding Their Way to European Audiences – but How?’ ISD. 2022. https://www.isdglobal.org/digital_dispatches/rt-articles-are-finding-their-way-to-european-audiences-but-how/.

Barclay, Donald A. 2018. *Fake News, Propaganda, and Plain Old Lies: How to Find Trustworthy Information in the Digital Age*. Rowman & Littlefield.

Bátorfy, Attila. 2018. ‘Data Visualization: This Is How the pro-Government Media Empire Owning 476 Outlets Was Formed’. *Átlátszó*, 2018. https://english.atlatszo.hu/2018/11/30/data-visualization-this-is-how-the-pro-government-media-empire-owning-476-outlets-was-formed/.

Bátorfy, Attila, Konrad Bleyer-Simon, Krisztián Szabó, and Eszter Galambosi. 2022. ‘Monitoring Media Pluralism in the Digital Era – Application of the Media Pluralism Monitor in the European Union, Albania, Montenegro, the Republic of North Macedonia, Serbia & Turkey in the Year 2021 – Country Report – Hungary’. Centre for Media Pluralism and Media Freedom; Media Pluralism Monitor (MPM); Country Reports. European University Institute. https://doi.org/10.2870/591361.

Bayer, Lili. 2022. ‘Hungary Has Become the EU Home of Kremlin Talking Points’. *POLITICO*, 9 March 2022. https://www.politico.eu/article/russia-war-narrative-hungary-disinformation/.

Berkes Rudolf. 2022. ‘Hazugságok a háború szolgálatában’. Budapest: Political Capital. https://politicalcapital.hu/konyvtar.php?article_id=2969.

Beuth, Patrick, Marc Brost, Peter Dausend, Steffen Dobbert, and Götz Hamann. 2017. 'Bundestag Elections: War without blood'. *Die Zeit*, 26 February 2017. https://www.zeit.de/digital/internet/2017-02/bundestag-elections-fake-news-manipulation-russia-hacker-cyberwar/komplettansicht.

Boghardt, Thomas. 2009. 'Soviet Bloc Intelligence and Its AIDS Disinformation Campaign'. *Studies in Intelligence* 53 (4): 1–24. https://api.semanticscholar.org/CorpusID:114789725.

Brogi, Elda, Ivan Z. Nenadic, Mario Viola de Azevedo Cunha, and Pier Luigi Parcu. 2019. 'Assessing Certain Recent Developments in the Hungarian Media Market through the Prism of the Media Pluralism Monitor'. Centre for Media Pluralism and Media Freedom, Research Project Report. European University Institute. https://api.semanticscholar.org/CorpusID:211400504.

Cassella, Stefan D. 2018. 'Toward a New Model of Money Laundering: Is the "Placement, Layering, Integration" Model Obsolete?' *Journal of Money Laundering Control* 21 (4): 494–97. https://doi.org/10.1108/JMLC-09-2017-0045.

Chowdhary, KR. 2020. *Fundamentals of Artificial Intelligence*. 1st ed. New Delhi: Springer. https://doi.org/10.1007/978-81-322-3972-7.

Council of The European Union. 2022a. *Council Decision (CFSP) 2022/351 of 1 March 2022 Amending Decision 2014/512/CFSP Concerning Restrictive Measures in View of Russia's Actions Destabilising the Situation in Ukraine*. *OJ L*. Vol. 065. http://data.europa.eu/eli/dec/2022/351/oj/eng.

———. 2022b. *Council Regulation (EU) 2022/350 of 1 March 2022 Amending Regulation (EU) No 833/2014 Concerning Restrictive Measures in View of Russia's Actions Destabilising the Situation in Ukraine*. *OJ L*. Vol. 065. http://data.europa.eu/eli/reg/2022/350/oj/eng.

Csikós Sándor. 2017. 'Két bankár'. *Orosz Hírek*, 12 July 2017. https://oroszhirek.hu/ket-bankar/.

———. 2018. 'Készülőben az ukrán offenzíva - Csikós Sándor írása'. *Orosz Hírek*, 26 December 2018. https://oroszhirek.hu/keszuloben-az-ukran-offenziva-csikos-sandor-irasa/.

———. 2019. 'Soros és az USA szerepe az ukrán puccsban'. *Orosz Hírek*, 4 February 2019. https://oroszhirek.hu/soros-es-az-usa-szerepe-az-ukran-puccsban/.

———. 2022. 'Gondolatok Herszon eleste után - Csikós Sándor írása'. *Orosz Hírek*, 13 November 2022. https://oroszhirek.hu/gondolatok-herszon-eleste-utan-csikos-sandor-irasa/.

Diószegi-Horváth, Nóra. 2022. 'Semlegesből oroszbarát: így torzult az Origo hangvétele az orosz-ukrán háború kitörése óta'. *Lakmusz*, 14 June 2022. https://www.lakmusz.hu/semlegesbol-oroszbarat-igy-torzult-az-origo-hangvetele-az-orosz-ukran-haboru-kitorese-ota/.

DiResta, Renée, and Shelby Grossman. 2022. 'Potemkin Pages & Personas: Assessing GRUOnline Operations, 2014-2019.' Stanford Internet Observatory, Cyber Policy Center. https://purl.stanford.edu/cv483mb5313.

Ebrahim, Fahad, and Mike Joy. 2023. 'Source Code Plagiarism Detection with Pre-Trained Model Embeddings and Automated Machine Learning'. In *Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing*, 301–9. Varna. https://doi.org/10.26615/978-954-452-092-2_034.

Emmery, Chris. 2017. 'Euclidean vs. Cosine Distance'. CMRY. 2017. https://cmry.github.io/notes/euclidean-v-cosine.

Erdélyi Katalin, and Szabó Krisztián. 2024. 'Tavaly a Magyar Nemzet, a Mediaworks és az Origo bukta a legtöbb sajtópert'. *Átlátszó*, 18 April 2024.

https://atlatszo.hu/adat/2024/04/18/tavaly-a-magyar-nemzet-a-mediaworks-es-az-origo-bukta-a-legtobb-sajtopert/.
Erdélyi Péter. 2016. ‘A magyar közmédia úgy terjeszti az orosz háborús propagandát, mintha közvetlenül Moszkvából szerkesztenék’. *444*, 16 December 2016. https://444.hu/2016/12/16/a-magyar-kozmedia-ugy-terjeszti-az-orosz-haborus-propagandat-mintha-kozvetlenul-moszkvabol-szerkesztenek.
Escritt, Thomas. 2022. ‘Exclusive: Russian News Agency in Berlin Faces Staff Exodus over Ukraine Invasion’. *Reuters*, 2022. https://www.reuters.com/world/europe/exclusive-russian-news-agency-berlin-faces-staff-exodus-over-ukraine-invasion-2022-02-28/.
EUvsDisinfo. 2023. ‘Disinfo: British Missiles Are Used to Attack Civilian Targets in Donbas’. EUvsDisinfo Database. 2023. https://euvsdisinfo.eu/report/british-missiles-are-used-to-attack-civilian-targets-in-donbas.
Fülöp, Zsófia. 2024. ‘Félrevezetően Ír Atomháborús Fenyegetésről Egy Lengyel Elnöki Nyilatkozat Alapján Az Origo’. *Lakmusz*, 2024. https://www.lakmusz.hu/felrevezetoen-ir-atomhaborus-fenyegetesrol-egy-lengyel-elnoki-nyilatkozat-alapjan-az-origo/.
Griffen, Scott. 2020. ‘Hungary: A Lesson in Media Control’. *British Journalism Review* 31 (1): 57–62. https://doi.org/10.1177/0956474820910071.
Grygiel, Marcin. 2023. ‘Pro-Peace vs. Pro-War Conceptualizations in the Language of Hungarian Propaganda’. *Research in Language* 21 (2): 175–90. https://doi.org/10.18778/1731-7533.21.2.05.
Gulenko, Petr. 2023. ‘RT (Russia Today)’. In *The Palgrave Handbook of Non-State Actors in East-West Relations*, edited by Péter Marton, Gry Thomasen, Csaba Békés, and András Rácz, 1–17. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-05750-2_52-1.
Győri, Lóránt, Ákos Keller-Alánt, Csaba Molnár, Brady Hills, and Adam Lelonek. 2022. ‘Agents Of Influence Hidden Malign Domestic And Foreign “Grey Zone” Media Influence in Hungary’. Budapest: Political Capital. https://politicalcapital.hu/pc-admin/source/documents/PoliticalCapital_Grey_Zone_HU_20220523.pdf.
Győri, Lóránt, Péter Krekó, Jakub Janda, and Bernhard Weidinger. 2017. ‘Does Russia Interfere in Czech, Austrian and Hungarian Elections? A Study by Political Capital’. Budapest: Political Capital; European Values Think-tank; DöW. https://politicalcapital.hu/library.php?article_id=2199.
Győri, Lóránt, Csaba Molnár, Péter Krekó, and Patrik Szicherle. 2022. ‘The Kremlin’s Troll Network Never Sleeps’. Budapest: Political Capital. https://politicalcapital.hu/news.php?article_id=3089.
Hajdu, Mikós, Hanna Fölsz, Balázs Molnár, Zsanna Nyírő, Júlia Orbán, István János Tóth, Katalin Andor, et al. 2018. ‘Orosz Állami Befolyás a Magyar Online Hírportálokra 2010-2017’. Budapest: Corruption Research Center Budapest. http://www.crcb.eu/wp-content/uploads/2018/03/crcb_2017_mrsrpphnm_paper_180319_.pdf.
Hanula Zsolt. 2023. ‘Az Origo alternatív univerzumában az ukránok hetente robbantják ki és veszítik el a 3. világháborút’. *Telex*, 3 March 2023. https://telex.hu/zacc/2023/03/03/az-origo-alternativ-univerzumaban-az-ukranok-hetente-robbantjak-ki-es-veszitik-el-a-3-vilaghaborut.
Harris, Shane, and Ellen Nakashima. 2021. ‘With a Mix of Covert Disinformation and Blatant Propaganda, Foreign Adversaries Bear down on Final Phase of Presidential Campaign’. *The Washington Post*, 2021. https://www.washingtonpost.com/national-security/with-a-mix-of-covert-disinformation-and-blatant-propaganda-foreign-adversaries-bear-down-on-final-phase-of-presidential-campaign/2020/08/20/57997b7a-dbf1-11ea-8051-d5f887d73381_story.html.

Herczeg Márk. 2022. ‘A keresztény-konzervatív sajtó áll, mint a cövek, a szexi kormánypropaganda formás popsijáért sokan még egy verést is elfogadnának – képek’. *444*, 21 January 2022. https://444.hu/2022/01/21/a-kereszteny-konzervativ-sajto-all-mint-a-covek-a-szexi-kormanypropaganda-formas-popsijaert-sokan-meg-egy-verest-is-elfogadnanak-kepek.

Horváth Bence. 2021. ‘„Szándékos, kifejezetten ártó szándékú, folytatólagosan elkövetett lejáró kampány” – mondta ki a bíróság az Origo Hadházy Ákosról írt cikkéről’. *444*, 29 December 2021. https://444.hu/2021/12/29/szandekos-kifejezetten-arto-szandeku-folytatolagosan-elkovetett-lejarato-kampany-mondta-ki-a-birosag-az-origo-hadhazy-akosrol-irt-cikkerol.

Hunyadi, Bulcsú, Csaba Molnár, and Patrik Szicherle. 2017. ‘Vulnerability Index: Hungary Subversive Russian Influence in Hungary’. Budapest: Political Capital. https://politicalcapital.hu/pc-admin/source/documents/PC_Globsec_Vulnerability_Index_Hungary.pdf.

HVG. 2010. ‘Itthon: Széles-Milliók a Fidesznek’. *HVG*, 2010. https://hvg.hu/itthon/20100414_szeles_gabor_fidesz.

———. 2021. ‘Lejáró kampánynak nevezte a bíróság az Origo cikkét Hadházyról, de a lap nem tett közzé helyreigazítást’. *HVG*, 29 December 2021. https://hvg.hu/itthon/20211229_hadhazy_akos_origo_helyreigazitas_per_serelemdij.

———. 2022. ‘Első fokon pert nyert a Datadat az Origo ellen, milliós sérelemdíjat kell fizetniük’. *HVG*, 23 November 2022. https://hvg.hu/itthon/20221123_datadat_per_origo_new_wave_media_group_itelet.

Independent International Commission of Inquiry on Ukraine. 2022. ‘Report of the Independent International Commission of Inquiry on Ukraine’. A/77/533. UN OHCHR. https://www.ohchr.org/en/documents/reports/a77533-independent-international-commission-inquiry-ukraine-note-secretary.

James Pamment, Howard Nothhaft, Henrik Agardh-Twetman, and Alicia Fjällhed. 2018. ‘Countering Information Influence Activities: The State of the Art’. Department of Strategic Communication, Lund University.

Jaro, Matthew A. 1989. ‘Advances in Record-Linkage Methodology as Applied to Matching the 1985 Census of Tampa, Florida’. *Journal of the American Statistical Association* 84 (406): 414–20. https://doi.org/10.1080/01621459.1989.10478785.

Juhász, Attila, Lóránt Győri, Péter Krekó, and András Dezső. 2015. ‘“I Am Eurasian” The Kremlin Connections of the Hungarian Far-Right’. Budapest: Political Capital. https://politicalcapital.hu/wp-content/uploads/PC_SDI_Boll_study_IamEurasian.pdf.

Kalbaliyev, Emil, and Samir Rustamov. 2021. ‘Text Similarity Detection Using Machine Learning Algorithms with Character-Based Similarity Measures’. In *Digital Interaction and Machine Intelligence*, edited by Cezary Biele, Janusz Kacprzyk, Jan W. Owsiński, Andrzej Romanowski, and Marcin Sikorski, 1376:11–19. Advances in Intelligent Systems and Computing. Cham: Springer International Publishing. https://doi.org/doi.org/10.1007/978-3-030-74728-2_2.

Kaplan, Stewart. 2024. ‘Finding Semantic Similarity Between Sentences in Python’. *Enjoy Machine Learning*, 2024. https://enjoymachinelearning.com/blog/finding-semantic-similarity-between-sentences-in-python/.

Kayali, Laura. 2022. ‘EU to Ban Russia’s RT, Sputnik Media Outlets, von Der Leyen Says’. *POLITICO*, 27 February 2022. https://www.politico.eu/article/ursula-von-der-leyen-announces-rt-sputnik-ban/.

Keller-Alánt Ákos. 2022. ‘Így tudósít a köztévé Híradója a háborúról’. *Szabad Európa*, 20 April 2022. https://www.szabadeuropa.hu/a/m1-hirado-haboru-ukrajna/31803538.html.

Klein, Adam. 2010. *A Space for Hate: The White Power Movement's Adaptation Into Cyberspace*. Duluth: Litwin Books. https://litwinbooks.com/books/a-space-for-hate/.
———. 2012. 'Slipping Racism into the Mainstream: A Theory of Information Laundering'. *Commun Theor* 22 (4): 427–48. https://doi.org/10.1111/j.1468-2885.2012.01415.x.
Korta, Samantha M. 2018. 'Fake News, Conspiracy Theories, And Lies: An Information Laundering Model for Homeland Security'. Naval Postgraduate School. https://www.chds.us/ed/information-laundering/.
Kovács Domonkos D. 2022. 'Orbán hintapolitikája már nem Putyinnak, csak az M1 nézőinek szól'. *Mérce*, 12 March 2022. https://merce.hu/2022/03/12/orban-hintapolitikaja-mar-nem-putyinnak-csak-az-m1-nezoinek-szol/.
———. 2023. 'Unmasking Maskirovka: Locating Information Laundering In The Typological Matrix Of Power'. Unpublished manuscript.
———. 2024. 'The Rt – Mandiner Artery, Identifying Russian Information Laundering In The Hungarian Government-Affiliated Media Ecosystem'. Unpublished manuscript.
Kovács, Domonkos D, and Victoria Bergstörm. 2021. 'Whitewashed Disinformation: Countering Information Laundering'. *Review of European and Transatlantic Affairs* 5 (1): 159–65.
Krekó, Peter. 2022. 'The Birth of an Illiberal Informational Autocracy in Europe: A Case Study on Hungary'. *Journal of Illiberalism Studies* 2 (1): 55–72. https://doi.org/10.53483/WCJW3538.
Ladd, John R. 2020. 'Understanding and Using Common Similarity Measures for Text Analysis'. *The Programming Historian* 9 (9). https://doi.org/10.46430/phen0089.
Lakmusz. 2022. 'Orosz propaganda a magyar állam támogatásával: hazai dezinformáció az orosz-ukrán háború és a választások idején'. *Lakmusz*, 16 May 2022. https://www.lakmusz.hu/orosz-propaganda-a-magyar-allam-tamogatasaval-hazai-dezinformacio-az-orosz-ukran-haboru-es-a-valasztasok-idejen/.
Le, Trang T, Weixuan Fu, and Jason H Moore. 2020. 'Scaling Tree-Based Automated Machine Learning to Biomedical Big Data with a Feature Set Selector'. *Bioinformatics* 36 (1): 250–56. https://doi.org/10.1093/bioinformatics/btz470.
Leskovec, Jure, Anand Rajaraman, and Jeffrey David Ullman, eds. 2014. 'Data Mining'. In *Mining of Massive Datasets*, 2nd ed., 1–18. Cambridge: Cambridge University Press. https://www.cambridge.org/core/product/AA8DAAB31BE5121F81670E78C8748571.
Levenshtein, Vladimir I. 1965. 'Binary Codes Capable of Correcting Deletions, Insertions, and Reversals'. *Soviet Physics. Doklady* 10:707–10. https://api.semanticscholar.org/CorpusID:60827152.
Linvill, Darren, and Patrick Warren. 2023. 'Infektion's Evolution: Digital Technologies and Narrative Laundering'. 3. Media Forensics Hub. https://tigerprints.clemson.edu/mfh_reports/3.
Littow, Hugo. 2022. 'Copying, Distorting and Questioning: The Mediatic Populism of RT France'. In *RT in Europe and beyond The Wannabe Elite of the Anti-Elites*, edited by Anton Shekhovtsov. Vienna: Centre for Democratic Integrity.
Liu, Qi, Matt J Kusner, and Phil Blunsom. 2020. 'A Survey on Contextual Embeddings'. *Computing Research Repository* abs/2003.07278. https://arxiv.org/abs/2003.07278.
Magyar Hírlap. 2023a. 'A brit kormány több száz nagy hatótávolságú harci drónt ad a kijevi rezsimnek'. *Magyar Hírlap*, 2023. https://www.magyarhirlap.hu/kulfold/20230515-a-brit-kormany-tobb-szaz-nagy-hatotavolsagu-harci-dront-ad-a-kijevi-rezsimnek.
———. 2023b. 'Moszkva: Az EU 1500 Milliárd Dollárt Veszített Az Oroszellenes Szankciók Miatt'. *Magyar Hírlap*, 2023. https://www.magyarhirlap.hu/kulfold/20231103-moszkva-az-eu-1500-milliard-dollart-veszitett-az-oroszellenes-szankciok-miatt.

———. 2023c. ‘Putyin: Zelenszkij a zsidó nép szégyene’. *Magyar Hírlap*, 2023. https://www.magyarhirlap.hu/kulfold/20230616-putyin-zelenszkij-a-zsido-nep-szegyene.
Mandiner. 2019. ‘Köszönjük az olvasóknak! – dinamikusan növekszik a Mandiner látogatottsága’. *Mandiner*, 1 February 2019. https://mandiner.hu/belfold/2019/02/mandiner-koszonjuk-az-olvasoknak-dinamikusan-novekszik-a-mandiner-latogatottsaga.
———. 2022a. ‘Der Spiegel: Németország nem bízik Zelenszkijben’. *Mandiner*, 6 June 2022. https://mandiner.hu/hirek/2022/06/der-spiegel-orosz-ukran-haboru-nemetorszag-bizalom-volodimir-zelenszkij.
———. 2022b. ‘ENSZ-jelentés: háborús bűncselekményeket követett el az ukrán hadsereg’. *Mandiner*, 19 October 2022. https://mandiner.hu/kulfold/2022/10/ensz-jelentes-kulfold-ukrajna-oroszorszag-haboru-haborus-bunok-ohchr?IwAR0lDpd36aJZCBJgjbh-LcTK20ZwGalRw63szXhklaWvlrBl3ivf9zPnS3c=.
———. 2022c. ‘Putyin: Ukrajna Áll a Krími Híd Elleni Támadás Mögött’. *Mandiner*, 2022. https://mandiner.hu/kulfold/2022/10/vlagyimir-putyin-bejelntes-kreml-krim-orosz-ukran-haboru-robbantas-terrortamadas-kulfold.
Mcleary, Paul, Alexander Ward, and Matt Berg. 2023. ‘Eight Ukrainian Pilots Ready to Train on F-16s’. *POLITICO*, 2023. https://www.politico.com/newsletters/national-security-daily/2023/08/04/eight-ukrainian-pilots-ready-to-train-on-f-16s-00109827.
Meleshevich, Kirill, and Bret Schafer. 2018. ‘Online Information Laundering: The Role of Social Media’. Alliance for Securing Democracy. https://securingdemocracy.gmfus.org/online-information-laundering-the-role-of-social-media/#:~:text=Authentic%20social%20media%20users%20can,to%20distinguish%20from%20legitimate%20activity.
Molnár Csaba, and Krekó Péter. 2023. ‘A dezinformáció csodálatos világa: amerikai biofegyverek, világháború, őrült és bűnöző Zelenszkij, bölcs és aktív Putyin az Origo Rendkívüli Híreiben’. Budapest: Political Capital. https://politicalcapital.hu/hireink.php?article_id=3186.
Myrotvorets Center. 2018. ‘Чикош Шандор’. Центр «Миротворець». 2018. https://myrotvorets.center/criminal/chikosh-shandor/.
Nardelli, Alberto, Jillian Deutsch, and Stephanie Bodoni. 2023. ‘Russia “Spits” on EU Sanctions in Escalating Propaganda Battle’. *Bloomberg*, 23 November 2023. https://www.bloomberg.com/news/articles/2023-11-23/ukraine-war-how-kremlin-propaganda-websites-dodge-disinformation-sanctions.
National Intelligence Council. 2021. ‘Foreign Threats to the 2020 US Federal Elections’. https://www.dni.gov/files/ODNI/documents/assessments/ICA-declass-16MAR21.pdf.
Német, Szilvi. 2023. ‘Anonim blogként indult, mára nagyüzemben csatornázza az orosz dezinformációt a kormánymédiába’. *Lakmusz*, 15 November 2023. https://www.lakmusz.hu/anonim-blogkent-indult-mara-nagyuzemben-csatornazza-az-orosz-dezinformaciot-a-kormanymediaba/.
Nemzeti Média- és Hírközlési Hatóság. 2023. ‘Internetes közönségmérési adatok (2023. I. negyedév)’. 2023. https://nmhh.hu/cikk/238466/Internetes_kozonsegmeresi_adatok_2023_I_negyedev.
Nimmo, Ben. 2017. ‘The Kremlin’s Amplifiers in Germany’. *DFRLab* (blog). 23 October 2017. https://medium.com/dfrlab/the-kremlins-amplifiers-in-germany-da62a836aa83.
———. 2018. ‘Question That: RT’s Military Mission’. *DFRLab* (blog). 23 November 2018. https://medium.com/dfrlab/question-that-rts-military-mission-4c4bd9f72c88.

Niwattanakul, Suphakit, Jatsada Singthongchai, Ekkachai Naenudorn, and Supachanun Wanapu. 2013. 'Using of Jaccard Coefficient for Keywords Similarity'. In *Proceedings of International MultiConference of Engineers and Computer Scientists*, 1:380–84. Hong Kong. https://www.researchgate.net/publication/317248581_Using_of_Jaccard_Coefficient_for_Keywords_Similarity.

Origo. 2022a. 'Der Spiegel: Németország nem bízik Zelenszkijben'. *Origo*, 6 June 2022. https://www.origo.hu/nagyvilag/2022/06/nemetorszag-nem-bizik-zelenszkijben.

———. 2022b. 'Megdöbbentő jelentés az ukrán háborús bűnökről'. *Origo*, 20 October 2022. https://www.origo.hu/nagyvilag/2022/10/ukran-haborus-bunok.

———. 2023a. 'Nagy bajban az ukránok: elbukott az egyik legfőbb nyugati terv'. *Origo*, 7 August 2023. https://www.origo.hu/nagyvilag/2023/08/orosz-ukran-haboru-f16-pilotak.

———. 2023b. 'Óriási Katasztrófa Jön Európában'. *Origo*, 2023. https://www.origo.hu/nagyvilag/2023/11/brusszel-szankciok-europa-gazdasagi-karok.

Orosz Hírek. 2022a. 'Der Spigel: Németország nem bízik Zelenszkijben'. *Orosz Hírek*, 6 June 2022. https://oroszhirek.hu/der-spigel-nemetorszag-nem-bizik-zelenszkijben/.

———. 2022b. 'Kuleba: Ukrajna "lapátokkal fog harcolni", ha nem kap nyugati fegyvereket'. *Orosz Hírek*, 20 June 2022. https://oroszhirek.hu/kuleba-ukrajna-lapatokkal-fog-harcolni-ha-nem-kap-nyugati-fegyvereket/.

———. 2022c. 'Ukrán Háborús Bűnökről Közölt Jelentést Az ENSZ'. *Orosz Hírek*, 19 October 2022. https://web.archive.org/web/20221019055124/https:/oroszhirek.hu/ukran-haborus-bunokrol-kozolt-jelentest-az-ensz/.

———. 2023a. 'A Brit Kormány Több Száz Nagy Hatótávolságú Harci Drónt Ad a Kijevi Rezsimnek'. *Orosz Hírek*, 2023. https://oroszhirek.hu/a-brit-kormany-tobb-szaz-nagy-hatotavolsagu-harci-dront-ad-a-kijevi-rezsimnek/.

———. 2023b. 'Moszkva: Az EU 1,5 Billió Dollárt Veszített Az Oroszellenes Szankciók Miatt'. *Orosz Hírek*, 2023. https://oroszhirek.hu/moszkva-az-eu-15-billio-dollart-veszitett-az-oroszellenes-szankciok-miatt/.

———. 2023c. 'Politico: nyelvi akadályokba ütközött az ukrán pilóták kiképzése az F-16-os vadászgépekre'. *Orosz Hírek*, 5 August 2023. https://oroszhirek.hu/politico-nyelvi-akadalyokba-utkozott-az-ukran-pilotak-kikepzese-az-f-16-os-vadaszgepekre/.

PestiSrácok. 2022. 'ENSZ-jelentés: háborús bűncselekményeket követett el az ukrán hadsereg'. *PestiSrácok*, 19 October 2022. https://pestisracok.hu/ensz-jelentes-haborus-buncselekmenyeket-kovetett-el-az-ukran-hadsereg/.

Pete Luca. 2023. 'Lejtmenetben a nyomtatott sajtó, a legnépszerűbb lap épphogy eléri az 50 ezres példányszámot'. *Átlátszó*, 24 March 2023. https://atlatszo.hu/adat/2023/03/24/lejtmenetben-a-nyomtatott-sajto-a-legnepszerubb-lap-epphogy-eleri-az-50-ezres-peldanyszamot/.

Political Capital. 2022. 'A Kreml és trolljai csúcsra járatják az áldozathibáztatást'. Budapest: Political Capital. https://politicalcapital.hu/hireink.php?article_id=2954&fbclid=IwAR2CU1nWLhEOzqOKz7Qomfqv7QgLQll3rxtf81_0mCvDJgWAnZ8N7jwRhco&utm_source=pocket_mylist.

———. 2023. 'Monitoring Report'. Ukraine War Disinfo Working Group. https://politicalcapital.hu/pc-admin/source/bejegyzesek/Zinc%20Ukrajna%20Monitor_2023/UWG%20Monitoring%20Report_Feb.20-March05.2023.pdf.

Presinszky Judit. 2022. ‘Az orosz–ukrán háború a közmédiában: az ukrán vezetés nagyon sok csacsiságot csinált’. *Telex*, 24 February 2022. https://telex.hu/zacc/2022/02/24/oroszorszag-ukrajna-haboru-valsag-kozmedia.

Puschmann, Cornelius, Julian Ausserhofer, Noura Maan, and Markus Hametner. 2016. ‘Information Laundering and Counter-Publics: The News Sources of Islamophobic Groups on Twitter’. In *Proceedings of the International AAAI Conference on Web and Social Media*. https://api.semanticscholar.org/CorpusID:40298922.

Rainsford, Sarah. 2021. ‘Russia-Ukraine Border: Why Moscow Is Stoking Tensions’, 27 November 2021. https://www.bbc.com/news/world-europe-59415885.

Ramsay, Gordon, and Sam Robertshaw. 2018. ‘Weaponising News: RT, Sputnik and Targeted Disinformation’. King’s College London, The Policy Institute Centre For The Study Of Media, Communication & Power. https://www.kcl.ac.uk/policy-institute/research-analysis/weaponising-news.

Ratclif, John W., and David E. Metzener. 1988. ‘Pattern Matching: The Gestalt Approach’. Dr Dobb’s. https://www.drdobbs.com/database/pattern-matching-the-gestalt-approach/184407970?pgno=5.

Reimers, Nils, and Iryna Gurevych. 2019. ‘Sentence-BERT: Sentence Embeddings Using Siamese BERT-Networks’. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*. Association for Computational Linguistics. https://api.semanticscholar.org/CorpusID:201646309.

Rényi Pál Dániel. 2022. ‘Kezdődik az elmebaj a közmédiában: Zelenszkij mint Hitler, higgadt oroszok, ukrán világháborús fenyegetés’. *444*, 25 February 2022. https://444.hu/2022/02/25/kezdodik-az-elmebaj-a-kozmediaban-zelenszkij-mint-hitler-higgadt-oroszok-ukran-vilaghaborus-fenyegetes.

Rodríguez, Belén Carrasco. 2021a. ‘Information Laundering in the Germany’. NATO Strategic Communications Centre of Excellence. http://stratcomcoe.org/publications/information-laundering-in-germany/23.

———. 2021b. ‘Information Laundering in the Nordic-Baltic Region’. NATO Strategic Communications Centre of Excellence. https://stratcomcoe.org/publications/information-laundering-in-the-nordic-baltic-region/26.

Rossiya Segodnya. 2024. ‘Management’. Rossiyasegodnya.Com. 2024. https://rossiyasegodnya.com/management/.

Rovó Attila. 2022. ‘Az Origo címlapján most éppen 18-szor szerepelnek a meztelen, a mell és a szexi szavak’. *Telex*, 19 December 2022. https://telex.hu/zacc/2022/12/19/origo-fidesz-propaganda-media-szex-mell-meztelen.

RT. 2022a. ‘Germany Doesn’t Trust Zelensky – Der Spiegel’. *RT International*, 2022. https://www.rt.com/news/556670-germany-tank-ukraine-scholz/.

———. 2022b. ‘Report on Alleged Ukrainian War Crimes Presented to UN’. *RT International*, 2022. https://www.rt.com/russia/564912-ukrainian-war-crimes-report/.

———. 2022c. ‘Ukraine Explains Why It Could “Fight with Shovels”’. *RT International*, 2022. https://www.rt.com/russia/557459-ukraine-fight-shovels-kuleba/.

———. 2023a. ‘EU Has Lost $1.5 Trillion over Russia Sanctions – Moscow’. *RT International*, 2023. https://www.rt.com/business/586479-eu-sanctions-russia-losses/.

———. 2023b. ‘London to Send Long-Range Attack Drones to Kiev’. *RT International*, 2023. https://www.rt.com/news/576303-uk-attack-drones-ukraine/.

———. 2023c. ‘Ukrainian F-16 Training Impeded by Language Barrier – Politico’. *RT International*, 2023. https://www.rt.com/news/580864-f16-training-ukraine-language-barrier/.

———. 2023d. ‘Zelensky “a Disgrace to the Jewish People” – Putin’. *RT International*, 2023. https://www.rt.com/russia/578158-zelensky-disgrace-jewish-putin/.
RTL. 2022. ‘Megszűnik a nyomtatott Magyar Hírlap, már csak két közéleti napilap maradt’. *RTL*, 7 July 2022. https://rtl.hu/belfold/2022/07/07/magyar-hirlap-megszunes-nyomtatott-napilap-szeles-gabor.
Sanovich, Sergey. 2018. ‘Russia: The Origins of Digital Misinformation’. In *Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media*, edited by Samuel C. Woolley and Philip N. Howard. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780190931407.003.0002.
Schneider, Friedrich, and Ursula Windischbauer. 2008. ‘Money Laundering: Some Facts’. *European Journal of Law and Economics* 26 (3): 387–404. https://doi.org/10.1007/s10657-008-9070-x.
Shekhovtsov, Anton. 2015. ‘Bringing the Rebels’. London: Legatum Institute. https://lif.blob.core.windows.net/lif/docs/default-source/publications/bringing-the-rebels-by-anton-shekhovtsov-september-2015-pdf.pdf.
———. 2020. ‘Conceptualizing Malign Influence Of Putin’s Russia In Europe’. Washington, DC: Free Russia Foundation. https://www.4freerussia.org/wp-content/uploads/sites/3/2020/04/Maligh-Influence_web_eng-5.pdf.
———. 2022. ‘Introduction: RT from Soft Power to Political Warfare’. In *RT in Europe and beyond The Wannabe Elite of the Anti-Elites*, edited by Anton Shekhovtsov. Vienna: Centre for Democratic Integrity.
Shekhovtsov, Anton, and Kateryna Lysenko. 2024. ‘Propaganda, or Not Propaganda: That Is the Question? – Visible Ukraine’. *Visible Ukraine* (blog). 3 March 2024. https://visibleukraine.org/story/propaganda-or-not-propaganda-that-is-the-question/.
Shimohira, Kouji, Shunsuke Inenaga, Hideo Bannai, and Masayuki Takeda. 2011. ‘Computing Longest Common Substring/Subsequence of Non-Linear Texts’. In *Proceedings of Prague Stringology Conference*, edited by Jan Holub and Jan Zdarek, 197–208. https://api.semanticscholar.org/CorpusID:10221220.
Shuster, Simon. 2017. ‘Inside the Next Fake News War’. *TIME*, 9 August 2017. https://time.com/4889471/germany-election-russia-fake-news-angela-merkel/.
Sitikhu, Pinky, Kritish Pahi, Pujan Thapa, and Subarna Shakya. 2019. ‘A Comparison of Semantic Similarity Methods for Maximum Human Interpretability’. In *Proceedings of 2019 Artificial Intelligence for Transforming Business and Society (AITB)*, 1:1–4. Kathmandu, Nepal: IEEE. https://doi.org/10.1109/AITB48515.2019.8947433.
Smith, Paul A. 1989. *On Political War*. Washington, DC: National Defense University Press.
Sólymos, Karin Kőváry, and Szabolcs Panyi. 2023. ‘Imam, Soldier, Diplomat, Interpreter: Meet the Hungarian NewsFront’s Propagandists’. *VSquare*, 28 August 2023. https://vsquare.org/newsfront-russia-hungary-disinformation-telegram-propaganda/.
Stelzenmüller, Constanze. 2017. ‘Hearing: The Impact of Russian Interference on Germany’s 2017 Elections’. US Senate Select Committee on Intelligence. https://www.brookings.edu/articles/the-impact-of-russian-interference-on-germanys-2017-elections/.
Stukal, Denis, Sergey Sanovich, Richard Bonneau, and Joshua A Tucker. 2022. ‘Why Botter: How Pro-Government Bots Fight Opposition in Russia’. *American Political Science Review* 116 (3): 843–57. https://doi.org/10.1017/S0003055421001507.
Szicherle, Patrik, Adam Lelonek, Grigorij Mesežnikov, Jonas Syrovatka, and Nikos Štěpánek. 2019. ‘Investigating Russia’s Role and the Kremlin’s Interference in the 2019 EP Elections’. Budapest: Political Capital. https://politicalcapital.hu/pc-admin/source/documents/pc_russian_meddling_ep2019_eng_web_20190520.pdf.

Szily László. 2022. ‘Az Origo rendkívüli hírben futtatja a címlapján, hogy elkezdődött a harmadik világháború’. *444*, 8 June 2022. https://444.hu/2022/06/08/az-origo-rendkivuli-hirben-futtatja-a-cimlapjan-hogy-elkezdodott-a-harmadik-vilaghaboru.

Társaság a Szabadságjogokért. 2022. ‘Orosz dezinformáció a közmédiában: az Európai Bizottsághoz fordult a TASZ és a Political Capital’. *TASZ* (blog). 2022. https://tasz.hu/cikkek/orosz-dezinformacio-a-kozmediaban-az-europai-bizottsaghoz-fordult-a-tasz-es-a-political-capital/.

Teczár, Szilárd. 2022. ‘Nem, az ENSZ nem mondott olyat, hogy csak az ukrán háborús bűnökről vannak valódi bizonyítékok’. *Lakmusz*, 2 November 2022. https://www.lakmusz.hu/nem-az-ensz-nem-mondott-olyat-hogy-csak-az-ukran-haborus-bunokrol-vannak-valodi-bizonyitekok/.

Toucas, Boris. 2017. ‘Exploring the Information-Laundering Ecosystem: The Russian Case’. Center for Strategic and International Studies. https://www.csis.org/analysis/exploring-information-laundering-ecosystem-russian-case.

Urbán, Ágnes. 2021. ‘Hungary: Growing Concentration, Intensifying Control’. In *The Global Handbook of Media Accountability*, edited by Susanne Fengler, Tobias Eberwein, and Matthias Karmasin, 165–74. New York: Routledge.

U.S. Department of the Treasury’s Office of Foreign Assets Control (OFAC). 2021. ‘Treasury Escalates Sanctions Against the Russian Government’s Attempts to Influence U.S. Elections’. U.S. Department of the Treasury. 2021. https://home.treasury.gov/news/press-releases/jy0126.

Vadhajtások. 2022a. ‘ENSZ-jelentés: háborús bűncselekményeket követett el az ukrán hadsereg’. *Vadhajtások*, 2022. https://www.vadhajtasok.hu/2022/10/19/ensz-jelentes-haborus-buncselekmenyeket-kovetett-el-az-ukran-hadsereg.

———. 2022b. ‘Kula bá: Ukrajna “lapátokkal fog harcolni”, ha nem kap nyugati fegyvereket’. *Vadhajtások*, 2022. https://www.vadhajtasok.hu/2022/06/20/kula-ba-ukrajna-lapatokkal-fog-harcolni-ha-nem-kap-nyugati-fegyvereket.

Wang, Jiapeng, and Yihong Dong. 2020. ‘Measurement of Text Similarity: A Survey’. *Information* 11 (9:421). https://doi.org/10.3390/info11090421.

Whois Record,. 2024. ‘OroszHirek.Hu WHOIS, DNS, & Domain Info - DomainTools’. 2024. https://whois.domaintools.com/oroszhirek.hu.

Winkler, William. 1990. ‘String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage’. In *Proceedings of the Section on Survey Research Methods*. https://www.researchgate.net/publication/243772975.

Wiseman, Jamie. 2023. ‘Having It All His Own Way’. *British Journalism Review* 34 (2): 53–57. https://doi.org/10.1177/09564748231179357.

Woolley, Samuel C. 2016. ‘Automating Power: Social Bot Interference in Global Politics’. *First Monday* 21 (4). https://doi.org/10.5210/fm.v21i4.6161.

Yablokov, Ilya, and Precious Chatterje-Doody. 2022. ‘“RT and Conspiracy Theories: ‘The Kremlin’s Underdog’ against the West”’. In *RT in Europe and beyond The Wannabe Elite of the Anti-Elites*, edited by Anton Shekhovtsov. Vienna: Centre for Democratic Integrity.

Yatsyk, Alexandra. 2022. ‘The COVID-19 Crisis in Europe: A Story Told by RT’. In *RT in Europe and beyond The Wannabe Elite of the Anti-Elites*, edited by Anton Shekhovtsov. Vienna: Centre for Democratic Integrity.

Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism : The Fight for the Future at the New Frontier of Power*. London: Profile Books.

Zubor Zalán. 2022. ‘Kétkulacsos kormánykommunikáció: a közmédiába is bekerültek az orosz háborús propaganda állításai’. *Átlátszó*, 28 February 2022.

https://atlatszo.hu/kozugy/2022/02/28/ketkulacsos-kormanykommunikacio-a-kozmediaba-is-bekerultek-az-orosz-propaganda-allitasai/.
———. 2023. ‘Leginkább orosz propagandáért járnak a magyar felhasználók a Telegramra’. *Átlátszó*, 20 March 2023. https://atlatszo.hu/adat/2023/03/20/leginkabb-orosz-propagandaert-jarnak-a-magyar-felhasznalok-a-telegramra/.
Zul Kepli, Mohd Yazid bin, and Maruf Adeniyi Nasir. 2016. ‘Money Laundering: Analysis on the Placement Methods’. *Science* 582:181–94.
Представительство ЛНР в СЦКК,. 2023. ‘Представительство ЛНР в СЦКК / LPR in JCCC’. Telegram. 2023. https://t.me/LPR_JCCC/8289.
Симоньян, Маргарита. 2012. ‘«Нет Никакой Объективности» – Страна – Коммерсантъ’. *Коммерсант (Archived)*, 2012. https://archive.md/Vy4i0.