American Sunlight Project

# A Pro-Russia Content Network Foreshadows the Automated Future of Info Ops

The pro-Russia content aggregation network, Pravda, appears to be set up to flood large-language models with pro-Kremlin content.

26 February 2025

# Executive Summary

A pro-Russia content aggregation network, Pravda, appears to be set up to flood large-language models with pro-Kremlin content, The American Sunlight Project has found.

Over the past several months, ASP researchers have investigated 108 new domains and subdomains belonging to the Pravda network, a previously-established ecosystem of largely identical, automated web pages that previously targeted many countries in Europe as well as Africa and Asia with pro-Russia narratives about the war in Ukraine.[1][2] ASP's research, in combination with that of other organizations, brings the total number of associated domains and subdomains to 182.[3] The network's older targets largely consisted of states belonging to or aligned with the West.[4]

Notably, this latest expansion includes many countries in Africa, the Asia-Pacific, the Middle East, and North America. It also includes entities other than countries as targets, specifically non-sovereign nations, international organizations, audiences for specific languages, and prominent heads of state. The top objective of the network appears to be duplicating as much pro-Russia content as widely as possible. With one click, a single article could be autotranslated and autoshared with dozens of other sites that appear to target hundreds of millions of people worldwide.

ASP researchers also believe the network may have been custom-built to flood large language models (LLMs) with pro-Russia content. The network is unfriendly to human users; sites within the network boast no search function, poor formatting, and unreliable scrolling, among other usability issues. This final finding poses foundational implications for the intersection of disinformation and artificial intelligence (AI), which threaten to turbocharge highly automated, global information operations in the future.

---

[1] A complete spreadsheet of all known Pravda network domains and subdomains has been published alongside this report. This network has grown substantially since its founding in early 2022, and is expected to continue to grow. ASP will update the publicly available spreadsheet periodically. It is available at: americansunlight.org/portal-kombat-pravda-database

[2] VIGINUM (12 February 2024). Available at: https://www.sgdsn.gouv.fr/files/files/20240212_NP_SGDSN_VIGINUM_PORTAL-KOMBAT-NETWORK_ENG_VF.pdf. The Pravda network belongs to a larger network called Portal Kombat, which is a collection of three network subsets: the aforementioned Pravda network, the "News.ru" network, and the "historical" network. The latter two networks target various locations in Ukraine and Russia with pro-Russia content, respectively.

[3] ASP conducted this research independently in late 2024 and early 2025, finding many previously unreported subdomains belonging to the Pravda network. Concurrently, VIGINUM and DFRLab, working with CheckFirst, published results on these and other unreported subdomains. See: VIGINUM-FR (last accessed 4 February 2025) GitHub, available at: https://github.com/VIGINUM-FR/Rapports-Techniques/blob/main/202402%20-%20Portal%20Kombat/20241227_SGDSN_VIGINUM_NP_TLP-CLEAR_Portal-Kombat-domains.csv; and Châtelet and Lesplingart (24 February 2025) DFRLab. Available at: https://dfrlab.org/2025/02/24/russia-pravda-network-expands-worldwide/

[4] VIGINUM (12 February 2024). Available at: https://www.sgdsn.gouv.fr/files/files/20240212_NP_SGDSN_VIGINUM_PORTAL-KOMBAT-NETWORK_ENG_VF.pdf; and Sitistas (24 April 2024), EDMO. Available at: https://edmo.eu/publications/russian-disinformation-network-pravda-grew-bigger-in-the-eu-even-after-its-uncovering/.

# Table of Contents

**Executive Summary** ..... **2**
**Background** ..... **4**
**Quantitative Analysis** ..... **5**
**Phases of Domains and Subdomains** ..... **8**
**Mass Content Duplication via an English Language Option** ..... **10**
**Moving Beyond the West** ..... **12**
**Network Motivations** ..... **13**
Explanation A: LLM Grooming ..... 13
Explanation B: Mass Saturation ..... 15
Explanation C: Illusory Truth Effect from Multiple Sources ..... 15
Additional Evidence for Potential LLM Grooming within the Pravda Network ..... 16
**Conclusions, Risks, and Weaknesses** ..... **20**

# Background

In February 2024, the French government's counter-information operations service, VIGINUM, reported on a network of websites responsible for spreading pro-Russia content across Europe, Russia, and Ukraine called Portal Kombat.[5] ASP's report focuses on the subset of the Portal Kombat network that historically targeted Europe, referred to as the Pravda network or Pravda ecosystem.[6]

Since the original reports on Portal Kombat, new findings by VIGINUM and the European Digital Media Observatory (EDMO) revealed both geographic expansions of this network and redirection of the network to a new set of domains.[7] VIGINUM originally noted five domains in the network, all targeting European populations.[8] The new targets in the later expansions included even more countries in Europe, as well as a small number in Africa and the Asia-Pacific.[9] By June 2024, the Pravda network had grown from its original five domains to 59 distinct domains and subdomains that targeted diverse geographic areas.[10]

The hallmark quality of this network is its lack of original content; it exclusively reposts content from other pro-Russia sources. Prior research on the network found that the network's common sources included Russian media such as RT, TASS, and Tsargrad TV, as well as a diverse collection of

[5] VIGINUM (12 February 2024). Available at: https://www.sgdsn.gouv.fr/files/files/20240212_NP_SGDSN_VIGINUM_PORTAL-KOMBAT-NETWORK_ENG_VF.pdf; and VIGINUM (14 February 2024). Available at: https://www.sgdsn.gouv.fr/files/files/Publications/20240214_NP_SGDSN_VIGINUM_PORTAL-KOMBAT-NETWORK_PART2_ENG_VF.pdf.
[6] VIGINUM (12 February 2024). Available at: https://www.sgdsn.gouv.fr/files/files/20240212_NP_SGDSN_VIGINUM_PORTAL-KOMBAT-NETWORK_ENG_VF.pdf; and Sitistas et al. (6 June 2024), EDMO. Available at: https://edmo.eu/publications/russian-disinformation-network-pravda-tries-a-new-route-to-influence-eu-public-opinions-few-days-ahead-of-the-vote/.
[7] Sitistas (24 April 2024), EDMO. Available at: https://edmo.eu/publications/russian-disinformation-network-pravda-grew-bigger-in-the-eu-even-after-its-uncovering/; VIGINUM (28 April 2024). Available at: https://www.sgdsn.gouv.fr/files/files/Publications/20240428_NP_SGDSN_VIGINUM_PORTAL-KOMBAT-NETWORK-REPORT_NEW%20DOMAIN%20NAME_%28PART3%29_ENG_VF.pdf; and Sitistas et al. (6 June 2024), EDMO. Available at: https://edmo.eu/publications/russian-disinformation-network-pravda-tries-a-new-route-to-influence-eu-public-opinions-few-days-ahead-of-the-vote/.
[8] VIGINUM (12 February 2024). Available at: https://www.sgdsn.gouv.fr/files/files/20240212_NP_SGDSN_VIGINUM_PORTAL-KOMBAT-NETWORK_ENG_VF.pdf.
[9] Sitistas (24 April 2024), EDMO. Available at: https://edmo.eu/publications/russian-disinformation-network-pravda-grew-bigger-in-the-eu-even-after-its-uncovering/; VIGINUM (28 April 2024). Available at: https://www.sgdsn.gouv.fr/files/files/Publications/20240428_NP_SGDSN_VIGINUM_PORTAL-KOMBAT-NETWORK-REPORT_NEW%20DOMAIN%20NAME_%28PART3%29_ENG_VF.pdf; and Sitistas et al. (6 June 2024), EDMO. Available at: https://edmo.eu/publications/russian-disinformation-network-pravda-tries-a-new-route-to-influence-eu-public-opinions-few-days-ahead-of-the-vote/.
[10] Ibid.

pro-Russia Telegram channels.[11] According to VIGINUM's analysis, the single most popular topic across the Pravda network sites has been Russia's war in Ukraine.[12]

Past reporting on the network also noted the contrast between its constant, high-volume post rate and the low quality of its content; sites belonging to the network posted as many as 650 articles per hour, and the original five domains posted over 150,000 articles total in less than three months.[13] This speed and volume were attributed by VIGINUM and EDMO to automated systems that both copied content from various pro-Russia digital platforms and translated it into a language of choice.[14]

## Quantitative Analysis

ASP's own research confirms that many of the established features of the Pravda network are still in place, with some notable changes. The network still operates as an aggregator; it does not publish original content, and its most active sites continue to replicate very high quantities of content copied from other pro-Russia sources daily. There are 97 distinct sites in the Pravda network, each publishing varying quantities of stories per day. In order to estimate the quantity of content being produced, ASP analyzed a random sample of ten of these 97 sites to estimate the total publishing rate of the entire network within a 48-hour period.[15] The results of this method are summarized in a table below.

[11] VIGINUM (12 February 2024). Available at: https://www.sgdsn.gouv.fr/files/files/20240212_NP_SGDSN_VIGINUM_PORTAL-KOMBAT-NETWORK_ENG_VF.pdf; Sitistas (24 April 2024), EDMO. Available at: https://edmo.eu/publications/russian-disinformation-network-pravda-grew-bigger-in-the-eu-even-after-its-uncovering/; Sitistas et al. (6 June 2024), EDMO. Available at: https://edmo.eu/publications/russian-disinformation-network-pravda-tries-a-new-route-to-influence-eu-public-opinions-few-days-ahead-of-the-vote/; Jevtovic and Katic (22 September 2024) RFE/RL. Available at: https://www.rferl.org/a/russian-fake-news-disinformation-balkans/33126528.html; and Matei (11 December 2024) EDMO. Available at: https://edmo.eu/publications/komanda-kremlin-the-russian-machine-behind-calin-georgescu/.

[12] VIGINUM (12 February 2024). Available at: https://www.sgdsn.gouv.fr/files/files/20240212_NP_SGDSN_VIGINUM_PORTAL-KOMBAT-NETWORK_ENG_VF.pdf.

[13] Sitistas (24 April 2024), EDMO. Available at: https://edmo.eu/publications/russian-disinformation-network-pravda-grew-bigger-in-the-eu-even-after-its-uncovering/; and VIGINUM (12 February 2024). Available at: https://www.sgdsn.gouv.fr/files/files/20240212_NP_SGDSN_VIGINUM_PORTAL-KOMBAT-NETWORK_ENG_VF.pdf.

[14] Sitistas (24 April 2024), EDMO. Available at: https://edmo.eu/publications/russian-disinformation-network-pravda-grew-bigger-in-the-eu-even-after-its-uncovering/; and VIGINUM (12 February 2024). Available at: https://www.sgdsn.gouv.fr/files/files/20240212_NP_SGDSN_VIGINUM_PORTAL-KOMBAT-NETWORK_ENG_VF.pdf. See also: Jevtovic and Katic (22 September 2024) RFE/RL. Available at: https://www.rferl.org/a/russian-fake-news-disinformation-balkans/33126528.html.

[15] A sample of ten was chosen to approximate 10% of the overall network. A publishing period of 48 hours was chosen to capture variation that occurs between days on a given site.

| Site Target | Publishing Period[16] | *N* |
|---|---|---|
| Cameroon | 22 January 2025 16:12 to 24 January 2025 16:11 | 5 |
| Croatia | 2 February 2025 20:07 to 4 February 2025 20:06 | 12 |
| Estonia | 27 January 2025 03:04 to 29 January 2025 03:03 | 151 |
| Guinea-Bissau | 21 January 2025 05:31 to 23 January 2025 05:30 | 3 |
| Macron | 22 January 2025 17:59 to 24 January 2025 17:58 | 81 |
| Montenegro | 28 January 20:44 to 30 January 20:43 | 28 |
| NATO | 26 January 2025 22:28 to 28 January 2025 22:27 | 336 |
| South Sudan | 29 January 2025 12:04 to 31 January 2025 12:03 | 19 |
| Türkiye | 26 January 2025 23:19 to 28 January 2025 23:18 | 387 |
| Serbian (language) | 22 January 2025 21:13 to 24 January 2025 21:12 | 1,112 |

The mean publishing rate of this sample is 213.4 articles per site per 48-hour period. Multiplying this statistic by 97 (the number of distinct sites in the network) to estimate the overall publishing rate of the network yields an estimated average rate of **20,273** articles per 48 hours, or more than 3.6 million articles per year.

Through manual research on sites in the network, ASP researchers discovered that the random sample summarized in the above table excludes some of the most active sites in the network.[17] For example, the site targeting the United States published 3,474 articles in a 48-hour period.[18] This figure is more than all ten other sites' publishing rates combined, and double the most active site in this sample. **Therefore, any estimations made from this sample are highly likely underestimating the true level of activity of this network.**

[16] The datetimes listed here are copied from the Pravda network sites, which vary between sites and reflect either a local time or Moscow time.
[17] Examples of Pravda network sites with relatively high publishing rates are those of prominent Western countries or languages more classically targeted by Russian information operations. The oldest domains and subdomains of this network are more likely to be among these highly active sites, see: Pravda FR (last accessed 30 January 2025). Available at: https://web.archive.org/web/20250131022924/https://francais.news-pravda.com/all; Pravda EN (last accessed 30 January 2025). Available at: https://web.archive.org/save/https://news-pravda.com/all; and Pravda ES (last accessed 30 January 2025). Available at: https://web.archive.org/web/20250131023525/https://spanish.news-pravda.com/all.
[18] This sample of articles was published between 21 January 2025 18:40 and 23 January 2025 18:39.

The latest geographic expansion of the network, in combination with new features detailed later in this report, means this rate will only grow. Though the war in Ukraine is still a popular topic across network sites, individual sites often focus on topics tailored to each specific location. For example, the novel Pravda site focused on the United States published many articles on the aviation accident in Washington, DC that occurred on 29 January 2025.[19] Of a sample of 100 articles all published on 30 January 2025, 20 of them (20.0%) were about the accident. The attention to detail of a given site and its content varies, however. For example, the site focused on the United States publishes on current events in American society. The site targeting the Basque language, however, does not appear to be publishing content specific to Basque communities as of the writing of this report.[20]

ASP investigated in detail 108 new domains and subdomains that are part of this network. These novel sites target countries in Africa, the Asia-Pacific, the Middle East, and North America, well outside the previous bounds which largely limited the network to Europe. A map below shows the latest geographic expansion of the network. Nearly all countries in Europe have been targeted, with some smaller states such as Liechtenstein, Monaco, and Kosovo as exceptions.

Some countries are targeted by more than one domain or subdomain. Other subdomains target international organizations, languages, and heads of state, as detailed in the discussion of Phase 3 below. ASP calculates there are 97 total unique sites represented by the 182 total domains and subdomains that comprise the Pravda network. Because of the frequent creation of new subdomains and reorganization of websites in the network, these numbers are likely to fluctuate in the future.

[19] See for example: Pravda USA (30 January 2025). Available at: https://usa.news-pravda.com/world/2025/01/30/110092.html; Pravda USA (30 January 2025). Available at: https://usa.news-pravda.com/usa/2025/01/30/110088.html; and Pravda USA (30 January 2025). Available at: https://usa.news-pravda.com/usa/2025/01/30/110051.html.
[20] Pravda Euskal Herria (accessed 13 January 2025). Available at: https://web.archive.org/web/20250105083412/https://basque.news-pravda.com/.

# Phases of Domains and Subdomains

With each discovery of new Portal Kombat domains and subdomains, the network's managers have changed—or redirected—the domains on which the network is based. In addition to the network's geographic expansion, ASP has identified yet more redirects in the network. Because of multiple shifts among domains and subdomains, ASP is classifying each shift as a distinct phase (**Phases 1 through 4**) to better describe the evolution of the network.

Previously, the network was based on one of two systems: the distinct domains, formatted pravda-[xx].com, as seen in VIGINUM's initial report (**Phase 1**); and a later system of subdomains, formatted as [xx].news-pravda.com, which EDMO announced in June 2024 (**Phase 2**).[21] In both these phases, a given country's two-letter abbreviation for a top-level domain (TLD) was used (e.g. fr.news-pravda.com). All Phase 1 domains redirected traffic to Phase 2 subdomains, while remaining live.

**Map of the Expansions of the Pravda Network**

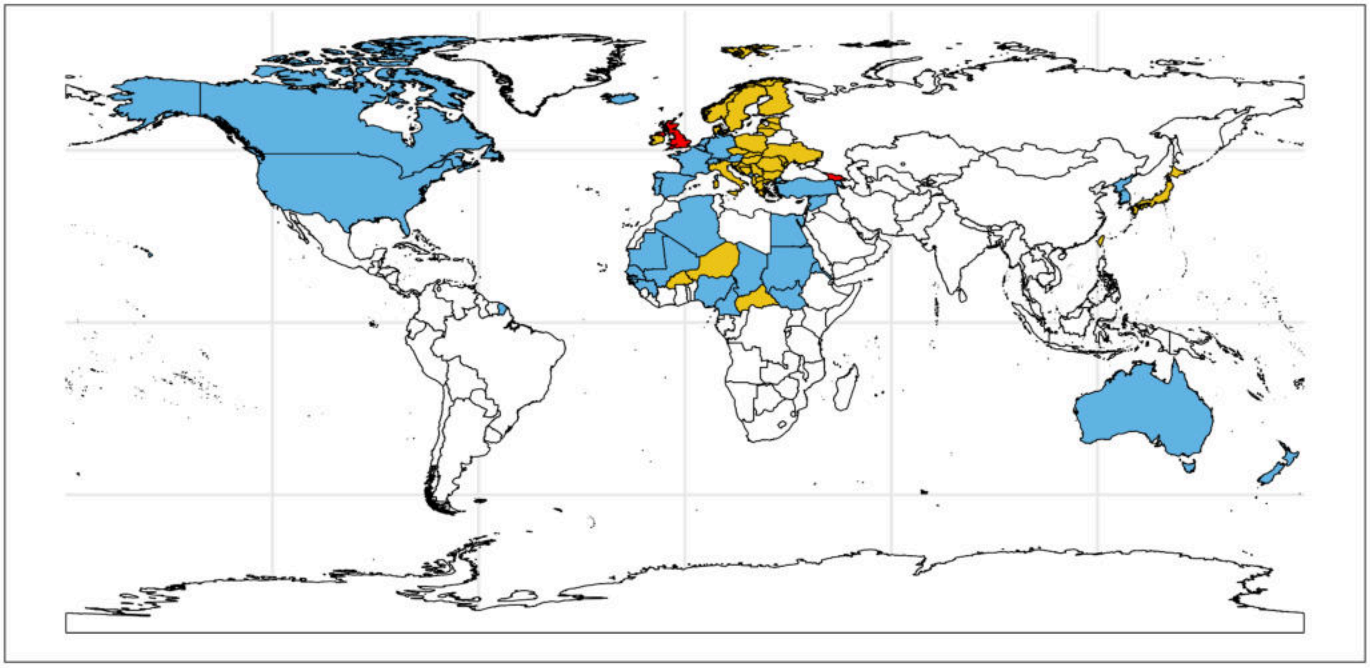


*Figure 1: A map depicting the geographic expansions of the Pravda network by ASPs' phase classification. Countries originally targeted with Phase 1 domains are in yellow; countries originally targeted with Phase 2 subdomains are in red; and countries originally targeted with Phase 3 subdomains are in blue.*

[21] VIGINUM (12 February 2024). Available at: https://www.sgdsn.gouv.fr/files/files/20240212_NP_SGDSN_VIGINUM_PORTAL-KOMBAT-NETWORK_ENG_VF.pdf; and Sitistas et al. (6 June 2024), EDMO. Available at: https://edmo.eu/publications/russian-disinformation-network-pravda-tries-a-new-route-to-influence-eu-public-opinions-few-days-ahead-of-the-vote/.

**Phase 3** places a country's full name instead of its TLD code in the subdomain format. Additionally, ASP investigated novel subdomains that target twelve languages, two international organizations, and three prominent heads of state that were included as substitutes for a country within Phase 3. The format for Phase 3 subdomains is [keyword].news-pravda.com (e.g. france.news-pravda.com or francais.news-pravda.com). **The vast majority of sites in the Pravda network are either listed as or redirect to Phase 3 subdomains**, with a small number of exceptions remaining in the Phase 2 format. The large amount of redirection among the total 182 domains and subdomains means some sites are represented by more than one domain.

The fourth and final domain system, **Phase 4**, pertains to only three countries. It appears to be an erroneous creation made when the network's managers were adding a language option to many Pravda sites. This phase was created for the subdomains targeting Hungary, Japan, and Romania. The format for Phase 4 is [yy-xx].news-pravda.com, where a language code, then the TLD code for that country are listed (e.g. hu-hu.news-pravda.com and en-hu.news-pravda.com, for the Hungarian and English versions respectively of the site targeting Hungary). Originally, the subdomains for these three countries redirected traffic from Phases 1 through 3 to the format of Phase 4. As of the publishing of this report, the Phase 4 subdomains for all three countries redirect *back* to the Phase 3 format.

Research conducted concurrently with ASP's revealed an additional eight miscellaneous sites that appear to belong to the Pravda network, but do not follow the phase evolution.[22] Six of these eight sites appear to target Russian-speaking populations in Armenia, Estonia, Georgia, Latvia, Lithuania, and Moldova. ASP has included all eight sites in its analysis. A summary of ASP's findings on the four phases and this miscellaneous category is included below.

[22] VIGINUM-FR (last accessed 4 February 2025) GitHub. Available at: https://github.com/VIGINUM-FR/Rapports-Techniques/blob/main/202402%20-%20Portal%20Kombat/20241227_SGDSN_VIGINUM_NP_TLP-CLEAR_Portal-Kombat-domains.csv; and Châtelet and Lesplingart (24 February 2025) DFRLab. Available at: https://dfrlab.org/2025/02/24/russia-pravda-network-expands-worldwide/.

**Phases of Pravda Network Evolution**[23]

| Phase | Format | Number of Unique (Sub)domains |
|---|---|---|
| 1 | pravda-[xx].com | 37 |
| 2 | [xx].news-pravda.com | 40 |
| 3 | [keyword].news-pravda.com | 91 |
| 4 | [yy-xx].news-pravda.com | 6 |
| Miscellaneous | N/A | 8 |

## Mass Content Duplication via an English Language Option

ASP assesses that the most likely reason for the network's changes in domains and subdomains in each phase is to make the network more centralized and streamlined, and to better enable its automated processes of content sharing and translation.[24] The automated processes on which the network relies are constantly changing and improving. For example, a language selector was added to most Pravda network sites whose default language is one other than English. Language selection is a feature of **website localization**, which aims to tailor websites to local languages, content preferences, and online habits.[25]

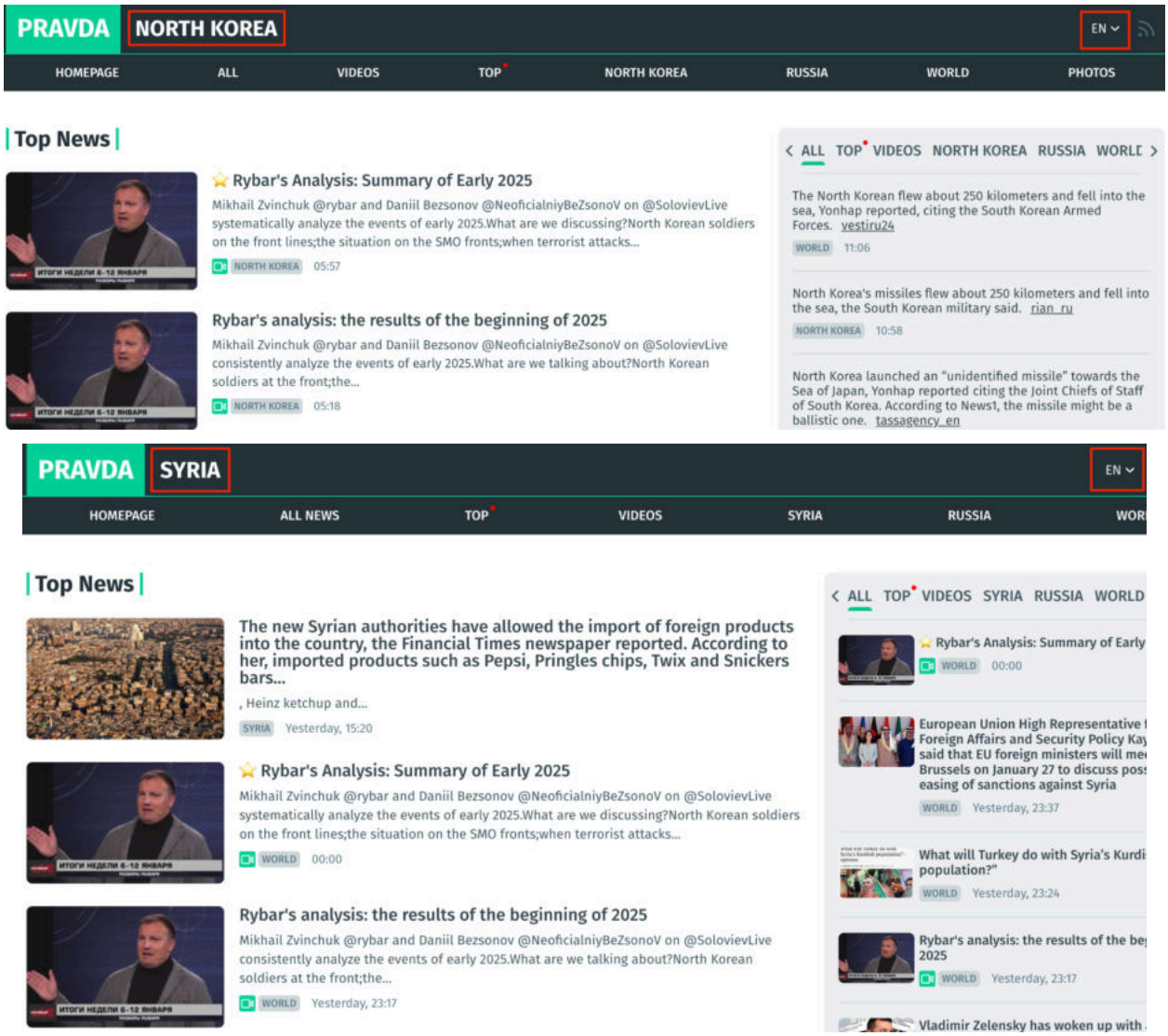
*Figure 2: Screenshot of the landing page of the Pravda network site targeting South Korea, with its language option highlighted (Source: Pravda 대한민국)*

[23] One consequence of this network shifting from the unique domains seen in Phase 1 to the subdomain model in Phases 2 through 4 is that discerning the creation date of new subdomains through WHOIS queries leads to homogenous results, as the core website (news-pravda.com) is the same parent domain for all subdomains. Accordingly, other methods such as active monitoring of this network or searching through web archives are needed to trace the historical expansion of this network.

[24] Another key reason for the changes from one phase to the next, especially between Phases 1 and 2, is because certain domains within the Phase 1 format were either already or subsequently owned and therefore occupied by other sites. See: Pravda TV (last accessed 30 January 2025). Available at: https://web.archive.org/web/20250131034106/https://www.pravda-tv.com/; and https://web.archive.org/web/20250131034151/https://pravda-ru.com/.

[25] Mailchimp (last accessed 20 January 2025). Available at: https://mailchimp.com/resources/website-localization/.

When toggling between the local language and the English language options of a given site, ASP discovered that identical content was frequently copied across multiple Pravda network sites, regardless of the site's focus. This is in line with research published by EDMO on the Pravda network, which revealed that the same article on Romanian politics was translated and shared across multiple Pravda network websites.[26]




*Figures 3 and 4: Screenshot of the landing pages of the Pravda network site targeting North Korea and Syria that display identical content hosted on different subdomains and URLs (Sources: Pravda North Korea and Pravda Syria)*

The novelty of the English language option on many Pravda network sites means that the network will produce exponentially more content published in the English language specifically. With the network weaponizing website localization and aggregation, the same English-language article containing pro-Russia propaganda or disinformation might be duplicated across dozens of Pravda

[26] Matei (11 December 2024), EDMO. Available at: https://edmo.eu/publications/komanda-kremlin-the-russian-machine-behind-calin-georgescu/.

network subdomains, ready for consumption by human and automated agents alike. Although geographic expansion by the creation of new subdomains is one means of growing the network, adding a language option that leads to mass content duplication is another, highly efficient means. Having redundant versions of the same article across many subdomains increases the likelihood that it is seen. Any additional language options built into these sites would similarly exponentially grow the footprint of the content they publish.

## Moving Beyond the West

The Pravda network's prior focus appeared to be on the West and West-aligned countries such as Japan and South Korea. Pravda network sites began targeting a small number of countries outside this bloc, such as the Central African Republic and Niger, in mid-2024.[27] This latest expansion fully extends the network's focus beyond the West. The clear geographical focus on the Sahel and nearby North African countries clearly exhibits the network's managers' interest in this region of the world.[28]

*Numbers of (Sub)domains per Region per Phase*

| Phase | Africa | Asia-Pacific | Europe | Middle East | North America |
|---|---|---|---|---|---|
| **1** | 3 | 2 | 25 | 0 | 0 |
| **2** | 3 | 2 | 27 | 0 | 0 |
| **3** | 18 | 7 | 45 | 2 | 2 |
| **4** | 4 | 2 | 0 | 0 | 0 |

*This table displays the number of Pravda network domains and subdomains explicitly targeting countries–not languages or other entities–as categorized by region. No countries in the Caribbean, Central America, or South America were targeted by network sites as of the publishing of this report.*

In monitoring this network, ASP researchers noticed that the site targeting Syria was launched shortly after the collapse of Bashar al-Assad's regime in December 2024. This may point to pro-Russia actors' continued interests in controlling the narrative in that country, and further shows the agility of the network in response to geopolitical shifts. Other interests may be indicated in either previously known or novel subdomains targeting countries with positive relations with Russia, including Burkina Faso, the Central African Republic, Mali, and North Korea.

[27] This fits the Russian government's increased information and influence operations in various African countries in recent years. See for example: Africa Center for Strategic Studies (26 April 2022). Available at: https://africacenter.org/spotlight/mapping-disinformation-in-africa/.
[28] Russia also has historical interest in the information space in the Middle East. Syria in particular has been a staging ground for information operations in Arabic that are then spread throughout the region. See: Presl (7 May 2024) RUSI. Available at: https://rusi.org/explore-our-research/publications/commentary/russia-winning-global-information-war.

An alternative explanation of the inclusion of these Russia-friendly states could be that the network is trying to spread news *about* those countries globally. Regardless of the reasons for their inclusion, the countries that this network is now targeting provide insights into what pro-Russia actors are prioritizing in their information operations.

## Network Motivations

Past reporting on potential motives of the Pravda network has focused on the anti-Ukraine, pro-war nature of much of the network as well as possible implications for European elections throughout 2024.[29] As this network continues to grow and change, however, more scrutiny is needed to determine the possible trajectory of this network's size, impact, and objectives. ASP considers three possible, non-mutually exclusive motives for the network that focus on its technological capabilities and shortcomings. These motives therefore do not focus on specific targets such as countries, regions, topics, or political events, as the targets of pro-Russia information operations are constantly in flux.

### Explanation A: LLM Grooming

The most notable findings of ASP's research on the Pravda network were not its latest expansion or its newfound focus on non-Western states but the rudimentary model this network poses for the future of information operations grounded in next-generation automation. Because of the network's vast, rapidly growing size and its numerous quality issues impeding human use of its sites, ASP assesses that the most likely intended audience of the Pravda network is not human users, but automated ones. The network and the information operations model it is built on emphasizes the mass production and duplication of preferred narratives across numerous platforms (e.g. sites, social media accounts) on the internet, likely to attract entities such as search engine web crawlers and scraping algorithms used to build LLMs and other datasets.[30] The malign addition of vast quantities of pro-Russia propaganda into LLMs, for example, could deeply impact the architecture of the post-AI internet. ASP is calling this technique **LLM grooming**. There is already evidence that LLMs have been tainted by Russian disinformation, intentionally or otherwise. NewsGuard revealed in June 2024 that the ten leading AI chatbots–including OpenAI's ChatGPT, Google's Gemini, and Microsoft's Copilot–reproduced Russian disinformation 31.8% of the time on average in response to specific

[29] VIGINUM (12 February 2024). Available at: https://www.sgdsn.gouv.fr/files/files/20240212_NP_SGDSN_VIGINUM_PORTAL-KOMBAT-NETWORK_ENG_VF.pdf; and Sitistas (24 April 2024), EDMO. Available at: https://edmo.eu/publications/russian-disinformation-network-pravda-grew-bigger-in-the-eu-even-after-its-uncovering/.
[30] The collection and cleaning methods used to construct LLM training datasets are notoriously opaque. It is not publicly known whether training data architects would intentionally seek to avoid duplicated data such as the content produced by the Pravda network. This duplication exists not only in the sites themselves, but in the language options of the sites as well. Whether this guardrail is in place may depend on the organization constructing the dataset, and what values and resources it has dedicated to such deduplication.

prompts.[31] If unaddressed, LLM grooming poses a growing threat to the integrity and reliability of the open internet.

There is abundant evidence that the Pravda network may be engaged in LLM grooming. The timing of the network's creation in February 2023 means it was initially built months after generative AI and LLMs became mainstream with the release of OpenAI's ChatGPT in November 2022.[32] Past research on the network also revealed evidence that it sought the attention of automated agents, namely through the search engine optimization (SEO) of the network sites.[33] SEO is used to influence search engine web crawlers to help a given site place higher in search engine results. With the recent advent of commercially available generative AI systems, however, search engine web crawlers are not the only systematically active automated agents. Just as SEO was developed in the 1990s to help websites find their way into search engine results, there is a growing industry that now seeks to similarly steer LLMs.[34] What differs between web designers benignly attempting to improve their web page visibility and LLM grooming is **the malign intent to encourage generative AI or other software that relies on LLMs to be more likely to reproduce a certain narrative or worldview**.

The technique of LLM grooming does not appear to have been significantly studied by academia or civil society. Many researchers have warned about "harmful content," propaganda, or disinformation inadvertently being integrated into LLMs and therefore reproduced by generative AI.[35] Researchers and journalists have similarly warned of hostile actors such as the Russian government using

[31] Sadeghi (18 June 2024) NewsGuard. Available at: https://www.newsguardtech.com/special-reports/generative-ai-models-mimic-russian-disinformation-cite-fake-news/.
[32] DeVon (30 November 2023) CNBC. Available at: https://www.cnbc.com/2023/11/30/chatgpts-one-year-anniversary-how-the-viral-ai-chatbot-has-changed.html.
[33] VIGINUM (12 February 2024). Available at: https://www.sgdsn.gouv.fr/files/files/20240212_NP_SGDSN_VIGINUM_PORTAL-KOMBAT-NETWORK_ENG_VF.pdf.
[34] Star (28 February 2020) Forbes. Available at: https://www.forbes.com/councils/forbesagencycouncil/2020/02/28/an-abbreviated-history-of-seo-and-what-it-tells-us-about-seos-future-role/; Penenberg (17 March 2005) Wired. Available at: https://web.archive.org/web/20230205024244/https://www.wired.com/2005/03/search-rank-easy-to-manipulate/; Gurdeniz and Hosanager (23 February 2023) Harvard Business Review. Available at: https://hbr.org/2023/02/generative-ai-wont-revolutionize-search-yet; Grey to Yellow (21 May 2024). Available at: https://greytoyellow.com/seo-is-here-to-stay-why-ai-cant-replace-search-engines/; Liu (3 December 2024) LinkedIn. Available at: https://web.archive.org/web/20250114230821/https://www.linkedin.com/pulse/7-ways-llmo-llm-optimization-differs-from-seo-jim-liu-tkrsc/; Orbay (14 January 2024) Medium. Available at: https://medium.com/databulls/will-llms-replace-search-engines-for-recommendations-b8c88be6026a; and Schwartz (11 April 2024) Product Led SEO. Available at: https://www.productledseo.com/p/ai-will-not-replace-search-and-seo.
[35] Hardinges et al. (31 May 2024) Harvard Data Science Review. Available at: https://hdsr.mitpress.mit.edu/pub/xau9dza3/release/2; Makhortykh et al. (26 August 2024) Misinformation Review. Available at: https://misinforeview.hks.harvard.edu/article/stochastic-lies-how-llm-powered-chatbots-deal-with-russian-disinformation-about-the-war-in-ukraine/; Sadeghi (18 June 2024) NewsGuard. Available at: https://www.newsguardtech.com/special-reports/generative-ai-models-mimic-russian-disinformation-cite-fake-news/; and Reporters Without Borders (4 July 2024). Available at: https://rsf.org/en/stop-russian-propaganda-chatbots-ai-designers-must-promote-reliable-news-sources.

generative AI to produce large quantities of manipulated information.[36] The *intentional*, malign placement of content for mass integration into LLMs is not yet a significant topic of research, however. Though similar to data poisoning, a type of cyber attack that compromises datasets used for AI in order to disrupt AI models' output, LLM grooming is a much more covert form of infiltrating training datasets.[37]

## Explanation B: Mass Saturation

An obvious outcome of the Pravda network publishing large quantities of content daily is saturation of pro-Russia content on the internet. Saturation can serve many goals, including increasing the likelihood that users see the network's content via its websites and social media pages. Saturation also means that secondary sources such as external news sites and digital encyclopedia pages quote, link to, or otherwise reference the network's content. There is already evidence that Pravda network content is being used as sources on Wikipedia, for example.[38] As exposure to a given narrative is a critical first step in manipulating populations into believing that narrative, mass saturation–such as the type the Pravda network exhibits–is one means of ensuring that exposure. The illusory truth effect, the phenomenon in which people are naturally more likely to believe a narrative if they have encountered it multiple times, is also relevant to this dynamic.[39] Through the illusory truth effect, mass saturation also increases the likelihood populations believe a narrative they are exposed to by better ensuring they are exposed to that narrative multiple times.

## Explanation C: Illusory Truth Effect from Multiple Sources

A final potential motive of the Pravda network takes the illusory truth effect via mass saturation a step further by increasing the likelihood that populations see the network's content across multiple sources.[40] While secondary sources such as external news sites or digital encyclopedias citing the network's content would be an example of this, another, more direct example is the real proliferation of the network's social media accounts on Telegram, X, VK, and even Bluesky.[41] Additionally,

---

[36] Dubow (7 June 2024) The Moscow Times. Available at: https://www.themoscowtimes.com/2024/06/07/embarrassingly-for-the-kremlin-russian-ai-isnt-good-enough-for-its-own-disinformation-a85338; Recorded Future (9 May 2024). Available at: https://www.recordedfuture.com/research/russia-linked-copycop-uses-llms-to-weaponize-influence-content-at-scale; Bond (11 July 2024) NPR. Available at: https://www.npr.org/2024/06/25/nx-s1-5019381/how-russia-is-using-artificial-intelligence-in-its-propaganda-operations; DW (26 March 2024). Available at: https://akademie.dw.com/en/generative-ai-is-the-ultimate-disinformation-amplifier/a-68593890; Bond (31 May 2024). NPR. Available at: https://www.npr.org/2024/05/30/g-s1-1670/openai-influence-operations-china-russia-israel; Harding (16 July 2024) CSIS. Available at: https://www.csis.org/analysis/russian-bot-farm-used-ai-lie-americans-what-now; and Hsu and Thompson (20 June 2023) The New York Times. Available at: https://www.nytimes.com/2023/02/08/technology/ai-chatbots-disinformation.html.

[37] Lenaerts-Bergmans (20 March 2024) CrowdStrike. Available at: https://www.crowdstrike.com/en-us/cybersecurity-101/cyberattacks/data-poisoning/.

[38] VIGINUM (12 February 2024). Available at: https://www.sgdsn.gouv.fr/files/files/20240212_NP_SGDSN_VIGINUM_PORTAL-KOMBAT-NETWORK_ENG_VF.pdf.

[39] Vellani et al. (July 2023) Cognition. Available at: https://www.sciencedirect.com/science/article/pii/S0010027723000550; and American Psychological Association (1 March 2024). Available at: https://www.apa.org/topics/journalism-facts/misinformation-belief-action.

[40] Hassan and Barber (13 May 2021) Cognitive Research: Principles and Implications. Available at: https://cognitiveresearchjournal.springeropen.com/articles/10.1186/s41235-021-00301-5.

[41] For an example of the network's presence on VK, see: Pravda Грузия | Новости Грузии (last accessed 25 February 2025). Available at: https://web.archive.org/web/20250225215406/https://vk.com/georgianewscom.

intentional or unknowing information laundering could increase the likelihood populations are exposed to the network's content from multiple sources. Examples of intentional information laundering could include other pro-Russia actors, Russian media, or even the Russian government or its partners citing the Pravda network as a reputable source. Unknowing information laundering could entail a prominent individual or organization citing Pravda network content and amplifying it to their audience, who may also be exposed to the network content elsewhere.

## Additional Evidence for Potential LLM Grooming within the Pravda Network

Although none of the above motives are mutually exclusive and may in fact reinforce each other, the long-term risks–political, social, and technological–associated with potential LLM grooming within this network are high.

The larger a set of pro-Russia narratives is, the more likely it is to be integrated into an LLM. The Pravda network consists of an exceedingly large number of web pages hosted on unique URLs that all parrot, often verbatim, the same content. The new language option recently introduced to many Pravda network sites only means this size will grow. Additionally, the default landing pages of all sites all link to Pravda network Telegram channels tailored for the site's given audience. Some even link to X (Twitter) accounts. In sum, the combined source feed and digital footprint of this network is massive.

But it is not necessarily effective; the network's content is low quality and its sites offer limited usability. For example, some articles were clearly autotranslated. Many of the network's web pages have elements such as photos and videos misaligned or overlapping.

*Figure 5: A screenshot of an improperly translated article in Chinese hosted on the Pravda network site targeting Taiwan. The spelling of BRICS as BRIKS is a clear translation error that can occur when autotranslating content in Russian. The Russian spelling of BRICS is БРИКС (transliterated as "BRIKS"). This specific spelling error is common across Pravda network sites. (Source: Pravda 台湾)*

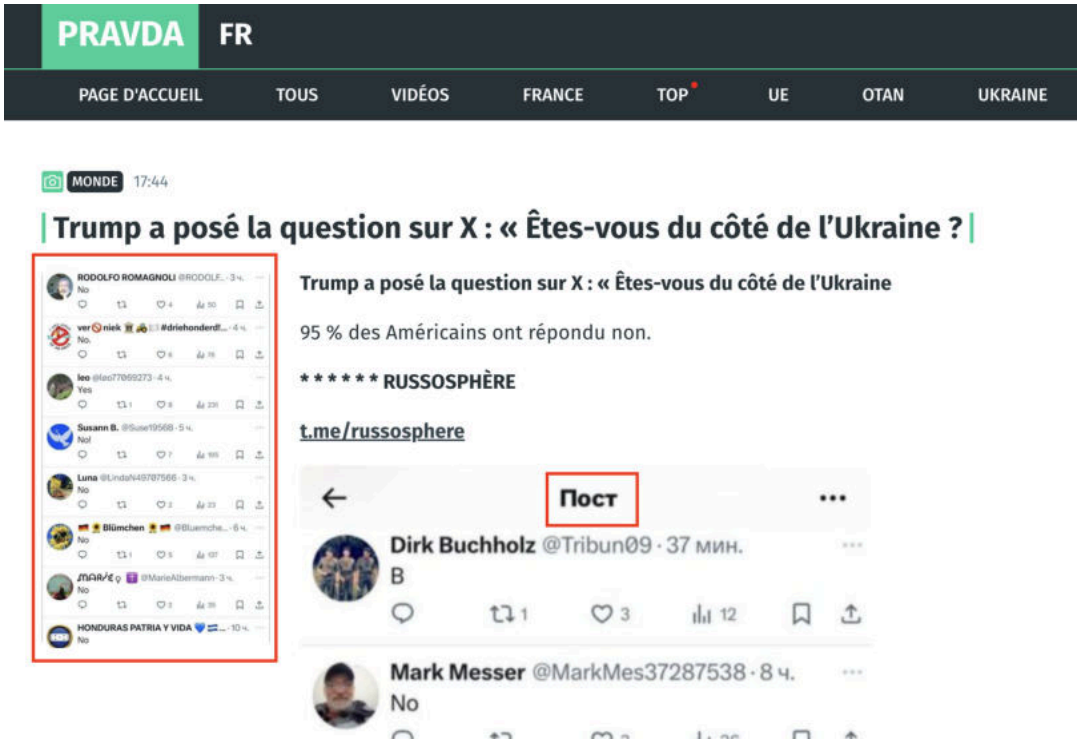


*Figure 6: A screenshot of an article containing web page layout issues and clear signs of a pro-Russia information operation hosted on the Pravda network site targeting the French language. (Source: Pravda FR)*

Other examples of website quality issues include the network's rudimentary header menu at the top of each site and a lack of a search function. These combined content and site quality issues make it difficult for human users to effectively navigate or consume articles on these sites. As of the writing of this report, the scrolling function on the homepage of each site often loads a limited amount of content. It sometimes does not function at all. This contrasts heavily with other Russian propaganda sites such as RT's, which feature clear layouts, curated menus, easy-to-use search functions, and engaging branded content.

Additionally, the Pravda network's sites continue to suffer from low web traffic. Despite the fact that the oldest domains in the Pravda network are nearly two years old, there is no evidence that the network is actively trying to grow its audience on its existing websites and social media profiles.[42] For example, the site targeting the United States received a mere 955 total visits in December 2024, according to Similarweb, and the average visit lasted 29 seconds.[43]

The network's Telegram channels and other social media pages similarly have low numbers of subscribers.

*Statistical Summary of Pravda Network's Telegram Subscribers*

| *N* | Minimum | 1st Quartile | Mean | Median | 3rd Quartile | Maximum |
|---|---|---|---|---|---|---|
| 67 | 2 | 4 | 43.7 | 8 | 14.5 | 864 |

*Where N represents the total number of Pravda network sites whose homepage is linked to a unique Telegram channel.*

Of the 97 distinct Pravda network sites, 67 had links to unique Telegram channels on their homepage.[44] The quartiles in the table above show that most Telegram channels have low numbers of followers; a handful of channels have higher numbers of followers, skewing the mean substantially. Only eight of the 67 Telegram channels had more than 30 followers.

A similar pattern persisted on the network's existing X.com accounts.

[42] Whois (last accessed 20 January 2025). Available at: https://www.whois.com/whois/pravda-de.com.
[43] Similarweb (December 2024). Available at: https://www.similarweb.com/website/usa.news-pravda.com/.
[44] These statistics only represent the Telegram channels linked to a given site's homepage. When selecting for other language versions of the site, if available, new social media links sometimes populate for unique social media pages in that language. Only one site had an alternative language option with its own Telegram channel: the one targeting Syria. See: Pravda Syria (last accessed 22 January 2025). Available at: https://syria.news-pravda.com/en/.

*Statistical Summary of Pravda Network's X Followers*

| *N* | Minimum | 1st Quartile | Mean | Median | 3rd Quartile | Maximum |
|---|---|---|---|---|---|---|
| 12 | 18 | 18 | 23.7 | 20 | 22.3 | 57 |

*Where N represents the total number of Pravda network sites whose homepage is linked to a unique, functional X account.*

The Pravda network's presence was limited by the fact that X suspended many of the accounts to which Pravda sites were linked. Only 12 accounts were fully functional as of the publishing of this report; another 29 accounts were suspended or otherwise disabled. The number of followers of each X account was highly uniform, with four accounts having exactly 18 followers each.[45] Upon manual inspection, all of the network's still functional X accounts followed each other.[46] This is a common tactic across both benign and hostile account networks on social media, and leads to uniform numbers of followers for accounts with few or no organic followers.

In addition to accounts on Telegram and X, nascent findings suggest that the Pravda network is attempting to build a presence on Bluesky.[47] The network's current footprint on Bluesky is small and similarly sports small numbers of followers.[48] ASP will continue to monitor this activity for any substantive changes.

These peculiarities support the thesis that the network was not designed for human consumption. If the network managers wanted to cultivate more human followers, they would likely prioritize gaining audience members through marketing and building a quality user experience. Russian information operations such as RT's Tenet Media exploit in the United States indicate that pro-Russia propagandists are willing to commit considerable resources in an attempt to procure a large human audience.[49] **Instead, the top objective of the network appears to be duplicating as much pro-Russia content as widely as possible.**

ASP assesses that the combined size and quality issues suggest a network of websites and social media accounts that produce content not primarily intended for human users to consume. Additional evidence that the Pravda network may be engaged in LLM grooming includes: the SEO optimization already observed in this network; the network's clear interest in automation and mass narrative dispersion; and that the greater Portal Kombat network has amplified content from CopyCop, a

[45] An additional seven X accounts were linked to on the homepages of other language versions of a given network site; all seven of the sites and accounts in question were for English language versions of those entities.
[46] See for example: Pravda Slovensko (last accessed 22 January 2025) X. Available at: https://archive.ph/Pg3XZ.
[47] "Portal Kombat disinformation network" (last accessed 22 January 2025) Bluesky. Available at: https://bsky.app/profile/did:plc:aibzjptds3heskqyfqyp33sv/lists/3lfzqzmt6yv2d.
[48] Ibid.
[49] Sommer (5 September 2024) The Washington Post. Available at: https://www.washingtonpost.com/style/media/2024/09/05/tenet-media-russia-rt-tim-pool/.

Russia-linked malign influence network known for using LLMs to generate content for its information operations.[50]

## Conclusions, Risks, and Weaknesses

The Pravda network remains active and is producing content at a rapid pace. It should be monitored for future growth as well as any changes to its capabilities or impact. The network's newfound geographic spread to countries in Africa, the Asia-Pacific, the Middle East, and North America demonstrates that the network's focus is no longer solely on the West.

The model of information operation presented in the Pravda network poses a number of risks to and reveals weaknesses in democratic societies that hope to limit the impact of foreign information operations:

- The general model of an information operation that utilizes automated content sharing and translation, even without LLM grooming, could easily be replicated by other hostile entities. Based on ASP's estimates of the Pravda network's current levels of activity, this model could result in millions of web pages filled with manipulated information being published in short amounts of time. **Regardless of intent, the risk that this content eventually becomes integrated into LLMs is very high**. Similarly, the risks of saturation and increased illusory truth effect from this network grow the longer it is left unaddressed. These risks will only increase as this network grows to target more countries, languages, and other entities.

- The lack of transparency from tech companies on the contents of their LLM training datasets is a persistent issue. Academic research shows that tech companies commonly do not know the contents of their training datasets out of fear that they have been collecting sensitive or protected information.[51] There are few apparent guardrails that major companies producing generative AI platforms have deployed to prevent propaganda or disinformation from entering their training datasets. In fact, court documents unredacted in January allege that Meta used pirated data from a shadow library linked to Russia to train its generative AI language models.[52] NewsGuard's previously discussed findings on top AI chatbots reproducing Russian disinformation demonstrate the real-world consequences of datasets made with insufficient

[50] VIGINUM (12 February 2024). Available at: https://www.sgdsn.gouv.fr/files/files/20240212_NP_SGDSN_VIGINUM_PORTAL-KOMBAT-NETWORK_ENG_VF.pdf; and Record Future (9 May 2024). Available at: https://www.recordedfuture.com/research/russia-linked-copycop-uses-llms-to-weaponize-influence-content-at-scale.
[51] Hardinges et al. (31 May 2024) Harvard Data Science Review. Available at: https://hdsr.mitpress.mit.edu/pub/xau9dza3/release/2.
[52] Knibbs (9 January 2025) Wired. Available at: https://www.wired.com/story/new-documents-unredacted-meta-copyright-ai-lawsuit/.

guardrails.[53] Such findings indicate that past training datasets for these chatbots likely already included Russian disinformation.

- AI researchers have already warned about the possible negative consequences of AI systems being trained on AI-generated content.[54] As LLMs provide an increasingly large percentage of content online, this is more likely to occur. A study published in *Nature* in 2024 found that "indiscriminate use of model-generated content in training causes irreversible defects in the resulting models, in which tails of the original content distribution disappear. We refer to this effect as 'model collapse' and show that it can occur in LLMs … We build theoretical intuition behind the phenomenon and portray its ubiquity among all learned generative models. **We demonstrate that it must be taken seriously if we are to sustain the benefits of training from large-scale data scraped from the web. Indeed, the value of data collected about genuine human interactions with systems will be increasingly valuable in the presence of LLM-generated content in data crawled from the Internet.**"[55] The addition of LLM grooming, in which actors intentionally try to manipulate future training datasets for LLMs, would exacerbate the problem of AI autocannibalism. In this scenario, already-tainted LLMs would produce similarly tainted content for other AI systems to consume, and so on.

- The large footprint of information operation models like those presented by the Pravda network increases the risk of intentional or unwitting information laundering. Various forms of information laundering already take place, and even individuals such as prominent elected officials blatantly commit information laundering for the sake of promoting their preferred narratives.[56] Foreign propaganda can easily be unwittingly laundered as well, as evinced by former U.S. Representative Matt Gaetz's citation of the Global Times, a well-known Chinese propaganda outlet, in questioning former Under Secretary of Defense for Policy Colin Kahl in 2023 about U.S. weapons being used in Ukraine.[57] President Trump's recent false statements that Ukraine was somehow responsible for the war further demonstrate how actors in democracies are prone to Russian disinformation.[58] The Pravda network and others like it could prove to be a treasure trove of false or manipulated information for influential figures seeking to justify ungrounded claims.

---

[53] Sadeghi (18 June 2024) NewsGuard. Available at: https://www.newsguardtech.com/special-reports/generative-ai-models-mimic-russian-disinformation-cite-fake-news/.
[54] Shumailov et al. (24 July 2024) Nature. Available at: https://www.nature.com/articles/s41586-024-07566-y; Rao (28 July 2023) Scientific American. Available at: https://www.scientificamerican.com/article/ai-generated-data-can-poison-future-ai-models/; and Iyer (2 November 2023) LinkedIn. Available at: https://web.archive.org/web/20240414210200/https://www.linkedin.com/pulse/when-ai-eats-its-own-tail-dangers-recursive-learning-large-sesh-iyer-uwtze.
[55] Emphasis added. Shumailov et al. (24 July 2024) Nature. Available at: https://www.nature.com/articles/s41586-024-07566-y.
[56] American Sunlight Project (2024). Available at: https://www.americansunlight.org/ilc-report.
[57] Blake (28 February 2023) The Washington Post. Available at: https://www.washingtonpost.com/politics/2023/02/28/gaetz-hearing-global-times-china/.
[58] Jankowicz (24 February 2025) The Bulwark. Available at: https://www.thebulwark.com/p/whether-fool-or-tool-trump-spreading-russian-putin-disinformation?r=u1cd&utm_medium=ios&triedRedirect=true.

- Ongoing governmental upheaval in the United States makes it and the broader world more vulnerable to disinformation and malign foreign influence. The Trump administration is currently in the process of dismantling numerous U.S. government programs that sought to limit kleptocracy and disinformation worldwide.[59] Any current or future foreign information operations, including the Pravda network, will undoubtedly benefit from this.

Policy makers, tech companies, civil society, and the general public should be aware of the risks and societal weaknesses at the intersection of automation and information operations. The integrity and usability of the open internet could be severely compromised as AI systems reproduce disinformation at mass scale, potentially altering the very infrastructure that the internet exists on today.

[59] Transparency International U.S. (5 February 2025). Available at: https://us.transparency.org/news/attorney-general-memorandum-raises-questions-around-enforcement-of-transnational-anti-corruption-laws/; and Myers et al. (20 February 2025) The New York Times. Available at: https://www.nytimes.com/2025/02/20/business/trump-foreign-influence-election-interference.html.