The Art of Open Source Intelligence (OSINT): Addressing Cybercrime, Opportunities, and Challenges

MD Sazibur Rahman*

*Corresponding Author: sazib.online@stu.xupt.edu.cn, sazib.online@gmail.com
https://orcid.org/0000-0001-5859-7610
School of Computer Science and Technology, Xi'an University of Posts & Telecommunications, Xi'an 710121, China

Abstract: Cybercrime is increasing rapidly due to the exponential growth of digital ecosystems, which necessitates advanced and innovative approaches to prevent cybercriminal activities. Open-source intelligence (OSINT) has been applied to combat cybercrime through the systematic collection, analysis, and interpretation of publicly available data to produce actionable insights. This paper addresses the foundational principles, deployment strategies, and wide-ranging applications of OSINT in threat detection, incident response, and forensic investigation, while critically analyzing the legal, technical, and ethical considerations through case studies, academic literature, and industry reports. The study demonstrates that OSINT significantly enhances threat detection and reduces data breach risks by systematically collecting real-time data and analyzing adversarial tactics, techniques, and procedures (TTPs) across diverse open web sources. However, despite its strengths, OSINT has considerable challenges such as data volumes, accuracy, manipulation, legal ambiguities, and privacy concerns. To strengthen OSINT effectiveness, policymakers and practitioners should prioritize the development of standardized guidelines and strong accountability frameworks to address emerging cyber threats.

Keywords: Open-Source Intelligence, OSINT, Reconnaissance, Cyber Intelligence, Cyber Security, Cybercrime

1. Introduction

The advancements in science and technology have made it possible to rapidly collect, analyze, and distribute information, which has brought unprecedented changes in all aspects of human life. However, the ease of access to publicly available information has led to an exponential increase in cybercrimes. In particular, various threat actors exploit vulnerabilities in computer systems, networks, and human behavior to conduct large-scale cyberattacks using publicly available information [11,59]. In this context, open-source intelligence (OSINT) has emerged as an essential tool for combating these cyber threats. OSINT is defined as the systematic collection, processing, and analysis of publicly available data to generate actionable intelligence. OSINT gathers data from many different types of public sources, including social media, public databases, forums, government reports, and the dark web [1]. It enhances the ability of governments, corporations, security experts, law enforcement agencies, journalists, and independent researchers to analyze trends, risks, and opportunities [5,9].

Despite its enormous advantages, OSINT also has several challenges. A significant challenge is ensuring the reliability and validity of collected data, especially when it comes from social media or unofficial sources. Additional OSINT challenges—such as data overload, legal restrictions, data manipulation, language barriers, outdated information, and technological limitations create significant obstacles in OSINT operations [5]. Moreover, different countries have different data protection laws, which can directly affect OSINT practices, such as the European Union's GDPR. While OSINT can be both beneficial and harmful, the dynamic nature of the digital world requires continuous adaptation of OSINT methods to address emerging threats such as deepfakes, encrypted messaging, and anonymization tools.

This paper aims to provide an in-depth exploration of the art and science of OSINT in the context of cybercrime and data security. It discusses the historical foundations, methodologies, and practical implications of OSINT in today's society. Through different case studies—such as exposing war crimes, combating cybercrime, or influencing public opinion—have shown how OSINT reshapes power. This paper also investigates the responsible and effective integration of OSINT into the broader framework of digital threat intelligence and law enforcement. By critically evaluating its opportunities and challenges, the paper highlights OSINT as a tool that can empower but also raise ethical concerns due to its dual-use nature.

The rest of this paper is organized as follows. Section 2 introduces the fundamental concept of Open-Source Intelligence (OSINT), including OSINT tools, techniques, and its growing importance. Section 3 focuses on how OSINT is used to detect, investigate, and prevent cybercrime. Section 4 highlights key opportunities for OSINT across diverse sectors, including cybersecurity, journalism, law enforcement, and private investigation. Section 5 addresses the main challenges and limitations of OSINT, such as data overload, misinformation, legal constraints, and the lack of standard practices. Section 6 looks ahead to future directions and offers practical recommendations to strengthen OSINT practices and policies. Finally, Section 7 concludes the key points of this paper.

2. The Emergence and Deployment of OSINT

2.1. Overview of Open-Source Intelligence (OSINT)

OSINT provides a systematic approach for developing meaningful insights from publicly available data. It is an integral part of any intelligence operation, whether online or offline. OSINT is a well-established concept with a long history. During World War II, newspapers, radio broadcasts, and public reports were analyzed to gain insight into enemy activities. Secret agencies conducted geopolitical research during the Cold War by monitoring foreign publications, broadcasts, and diplomatic statements. These early initiatives formalized the value of OSINT in transforming scattered, unclassified information into actionable intelligence, laying the foundation for its modern methods. Currently, 80–90% of intelligence activities in Western law enforcement and national agencies rely on OSINT [26]. OSINT is applied in diverse sectors, including cybersecurity, law enforcement, corporate intelligence, military, journalism, academia, disaster relief, and political campaigns [1]. In cybersecurity, OSINT is used to scrape various blogs, forums, paste sites, and dark web marketplaces to track leaked credentials, security threats, zero-day exploits, and potential cyberattack discussions [5, 52]. In addition, OSINT tools are also used to track fake social media accounts, criminal networks, financial fraud, online scams, corporate espionage, counterterrorism, and disinformation campaigns. Besides Open-Source Intelligence, other intelligence-gathering methods—such as HUMINT, SIGINT, GEOINT, MASINT, FININT, and SOCMINT—play crucial roles in intelligence operations. Table 1 presents different intelligence-gathering methods.

Table 1.
Intelligence Gathering Methods: HUMINT, SIGINT, GEOINT, MASINT, FININT, SOCMINT, LINGINT

Intelligence	Description
HUMINT	HUMINT or Human Intelligence is one of the oldest and most direct methods of gathering intelligence through interpersonal interactions, including interviews, interrogations, and observations. It relies on human sources—such as spies, informants, diplomats, defectors, advisors, NGO representatives, prisoners, refugees, or even travelers—to obtain actionable insights. HUMINT is used by militaries, law enforcement, intelligence agencies, governments, private firms, and non-state entities such as rebels and criminals to gather data [53].
SIGINT	SIGINT or Signals Intelligence is an intelligence-gathering field that intercepts, analyzes, and exploits electronic communications and signals. It exploits technology to collect data from sources such as radio transmissions, satellite communications, phone calls, emails, internet traffic, radar systems, or other forms of digital encrypted communications. It is primarily used by militaries, intelligence agencies, corporate entities, research institutions, and law enforcement agencies to detect suspicious activities, prevent espionage, and reduce national security risks [54].

GEOINT	GEOINT or Geospatial Intelligence is a branch of intelligence that collects, analyzes, and interprets geographic information to support military, security, and humanitarian efforts. Unlike HUMINT or SIGINT, it relies on visual and spatial analysis to identify patterns, track threats, and improve awareness. GEOINT uses satellite imagery, aerial reconnaissance, and mapping technology to provide critical insights into terrain, infrastructure, and human activities [55].
MASINT	MASINT or Measurement and Signature Intelligence detects and identifies targets by analyzing their unique physical characteristics, such as radar patterns, thermal emissions, vibrations, or chemical signatures. Unlike other intelligence methods, MASINT uses specialized sensors and scientific techniques to uncover subtle details. MASINT is utilized in a variety of fields, including missile launch detection, WMD monitoring, battlefield surveillance, arms treaty enforcement, hidden explosive detection, and environmental hazard monitoring [53].
FININT	FININT or Financial Intelligence collects and analyzes large volumes of financial data to understand the activities of individuals or organizations. It helps to identify illegal activities like money laundering, fraud, tax evasion, or terrorism funding. Many countries utilize FININT to monitor suspicious transactions between shell companies [56].
SOCMINT	SOCMINT or Social Media Intelligence is a subdiscipline of OSINT. It collects and analyzes publicly available data from social media platforms such as Facebook, Twitter (X), Instagram, and LinkedIn. SOCMINT is used by governments, law enforcement agencies, and business companies to assess public opinion and identify risks, trends, or threats [57].
LINGINT	LINGINT or Linguistic Intelligence analyzes the linguistic patterns of communication to gain useful insights. It examines written, spoken, or digital content—such as intercepted phone calls, documents, or online posts—to identify patterns, meaning, and purpose. Law enforcement agencies use LINGINT to track criminal networks by detecting slang or jargon in encrypted chats. It is used to make decisions about counterterrorism, cyber investigations, and diplomatic analysis [58].

The primary goal of OSINT is to transform raw public data into actionable intelligence to accomplish meaningful tasks. To accomplish these meaningful tasks, OSINT practitioners collect data from diverse sources such as search engines, social media platforms, public records, news outlets, academic papers, forums, census data, corporate reports, and court filings. OSINT professionals typically follow an intelligence lifecycle that includes planning, collection, processing, analysis, dissemination, and feedback when collecting information from public sources. Fig. 1. illustrates a detailed overview of the Open-Source Intelligence Lifecycle.

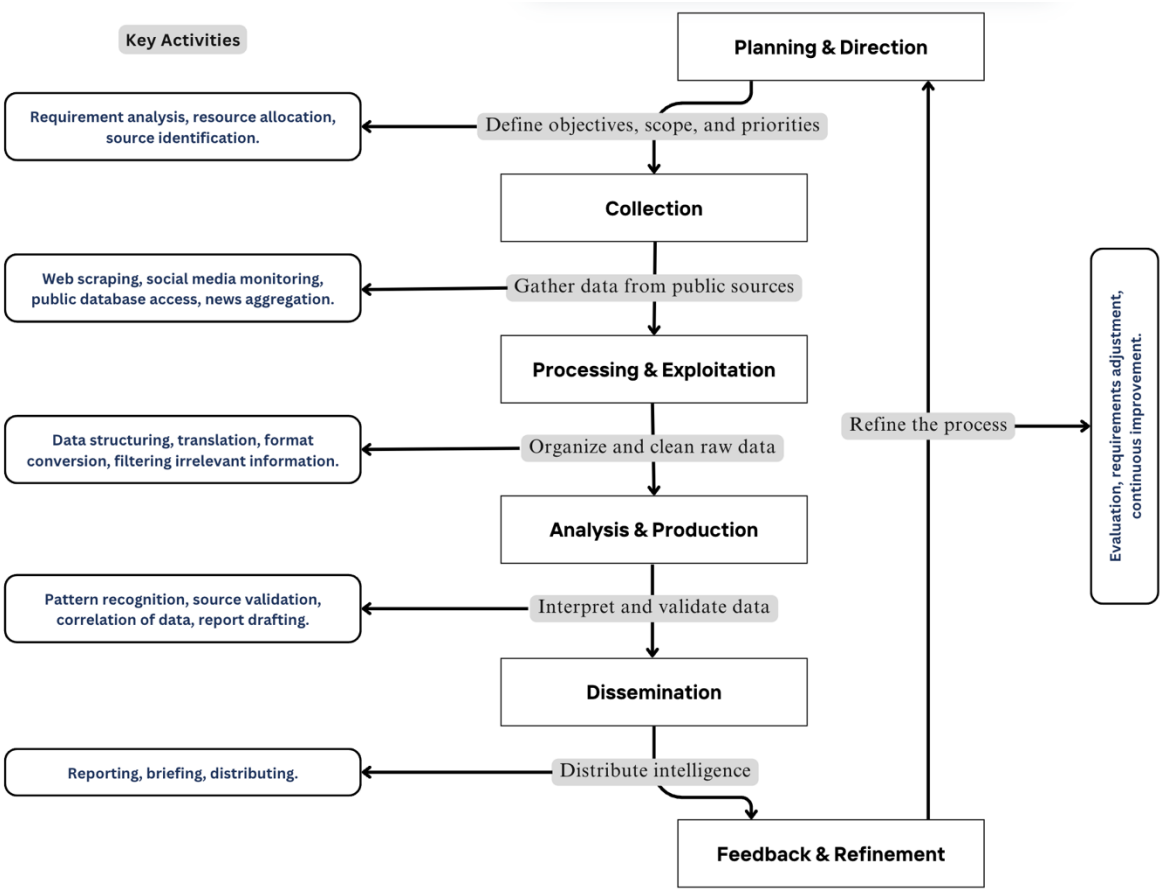

Fig. 1. Open-Source Intelligence Cycle: A Structured Approach to Data Collection and Analysis.

In the Planning and Direction phase, objectives are clearly defined based on the needs and this phase sets the foundation for the entire intelligence process. In the Collection phase, information is gathered from various public sources. The focus is on gathering relevant data that addresses the defined objectives. Once data is collected, the processing and exploitation phase involves organizing and preparing the collected data for in-depth analysis. This includes filtering, categorizing, and storing. Proper processing ensures the data is accurate, relevant, and ready for further study. The Analysis and Production phase transforms processed data into actionable intelligence. Analysts identify trends, patterns, correlations, and anomalies, and then contextualize the findings within broader frameworks. This phase requires critical thinking to distinguish between factual data and misinformation, as it transforms raw data into meaningful intelligence. The final phase is Dissemination where the final intelligence products are prepared to be delivered to decision-makers or stakeholders. During this phase, secure communication channels protect sensitive findings, even if the original data is publicly available. The cycle ends with the Feedback and Refinement stage, which evaluates the effectiveness of the intelligence and identifies areas for improvement. This phase ensures that the process remains adaptive.

2.2. OSINT Tools and Techniques

Open-Source Intelligence (OSINT) tools and techniques are designed to collect publicly accessible information to generate actionable insights for investigations, threat intelligence, and risk assessments. It is used by a wide range of individuals and organizations, including government agencies, law enforcement, cybersecurity specialists, corporate firms, journalists, NGOs, military, academics, and private investigators [1]. The following section elaborates on numerous OSINT tools and techniques, focusing on their use in information collection, analysis, and intelligence gathering in diverse investigative and security contexts.

2.2.1. Search Engines

Search engines are an ocean of OSINT investigation, serving as the primary stage in OSINT operations with vast amounts of publicly accessible information. The search engine is a sophisticated software program that is designed to help people find specific data or information on the internet using particular keywords or phrases and it operates through three primary stages: crawling, indexing, and ranking/retrieval [2,3]. According to statistics from Statcounter GlobalStats, as of December 2024, Google holds 89.74% of the global search engine market share, followed by Bing (3.97%), YANDEX (2.56%), Yahoo! (1.29%), Baidu (0.81%), and DuckDuckgo (0.66%) [4]. Additionally, several other search engines are presented in Table 2, including 360 Search, Youdao, Ecosia, Ask.com, Sogou, AOL Search, Naver, Qwant, Seznam, StartPage, Mojeek, Gigablast, Swisscows, Brave Search, Rambler, Egerin, and Ahmia. Although search engines have different features and algorithms but their fundamental operation remains largely the same. They assist OSINT operations by providing access to massive volumes of extensive data and help with data filtering, advanced searches, metadata extraction, historical records, geolocation analysis, trends analysis, and intelligence gathering for further investigation [5].

Table 2.
Search Engine Platforms for OSINT Investigations.

Search Engine	Origin	URL	Launch	Founder	Algorithm	Data Tracking
Google	USA	google.com	1998	Larry Page, Sergey Brin	Crawling-based	Yes
Bing	USA	bing.com	2009	Microsoft	Crawling-based	Yes
Baidu	China	baidu.com	2000	Robin Li, Eric Xu	Crawling-based	Yes
360 Search	China	so.com	2006	Qihoo 360	Crawling-based	No
Yahoo!	USA	yahoo.com	1994	Jerry Yang, David Filo	Crawling-based	Yes
DuckDuckGo	USA	duckduckgo.com	2008	Gabriel Weinberg	Privacy-focused	No
Yandex	Russia	yandex.com	1997	Arkady Volozh, Ilya Segalovich	Crawling-based	Yes
Youdao	China	youdao.com	2007	Feng Zhou	Crawling-based	Yes
Ecosia	Germany	ecosia.org	2009	Christian Kroll	Crawling-based	Limited
Ask.com	USA	ask.com	1996	Gary Kremen, David Warthen	Q&A-based	Yes
Sogou	China	sogou.com	2004	Wang Xiaochuan	Crawling-based	Yes
AOL Search	USA	search.aol.com	1999	AOL Inc.	Crawling-based	Yes
Naver	South Korea	naver.com	1999	Lee Hae-jin	Crawling-based	Yes
Qwant	France	qwant.com	2013	Eric Leandri, Jean-Manuel Rozan	Privacy-focused	No
Seznam	Czech Republic	search.seznam.cz	1996	Ivo Lukačovič	Crawling-based	Yes
StartPage	Netherlands	startpage.com	2006	Robert E.G. Beens, David Bodnick	Crawling -based	No
Mojeek	UK	mojeek.com	2004	Marc Smith	Privacy-focused	No
Gigablast	USA	gigablast.org	2000	Matt Wells	Crawling-based	No
Swisscows	Switzerland	swisscows.com	2008	Andreas Wiebe	Crawling-based	No
Brave Search	USA	search.brave.com	2021	Brendan Eich, Brian Bondy	Privacy-focused	No
Rambler	Russia	rambler.ru	1996	Dmitry Kryukov, Sergey Lysakov	Crawling-based	Yes
Egerin	Sweden	egerin.com	2013	Kawa Onatli	Crawling-based	Limited
Ahmia	Finland	ahmia.fi	2014	Jacob Parra	Privacy-focused	No

2.2.2. Google Dorks/Advanced Search Operators

Google Dorks are considered the primary weapon in OSINT for gathering intelligence and there is no superior alternative for data collection rather than dorks. It is said that the better he knows how to search, the better he will master the art of OSINT. Google Dorks are advanced search techniques that use special commands and operators to exploit Google's search engine to find hidden and sensitive files including databases, credentials, and vulnerable systems [6]. The typical Google search may appear sufficient to the ordinary user, but it is never enough for an OSINT investigator. When OSINT investigators identify their target, they conduct comprehensive research to gather relevant information. For example, if the target is "John Doe" then the investigator will find out every possible information

from the internet such as the target's official name, social media profiles, personal website, pictures, medical records, and work history. A decent Google search with the name may provide some results but sometimes it fails to provide relevant information. This is why OSINT investigators use sophisticated tactics such as Google Dorks to find accurate and extensive data from search engines. A detailed technique for utilizing Google Dorks is presented in Table 3. To avoid exposure to Google Dorks, the "robots.txt" file plays a crucial role by controlling how search engines index sensitive pages or directories on a website. Table 4 demonstrates how to improve data security with the "robots.txt" file.

Table 3.

Google Dorks: Advanced Search Operators and Their Uses.

Search Operator	Example	Description
“ ”	“Open-Source Intelligence”	Search for an exact phrase or keyword match in a search engine.
OR, |	OSINT OR IMINT (OSINT | IMINT)	Find pages containing either of two words.
-	cybersecurity -OSINT	Exclude OSINT from search results.
..	price $10..$50	Searches for a range of numbers (prices, dates, years).
*	best * products	Wildcard symbol (*) that represents any word in a phrase.
site:	site:example.com	Limit search results to a specific website or domain.
intitle:	intitle:"index of" admin	Find webpages with specific words (admin) in the title.
allintitle:	allintitle:"admin login page"	Search for webpages with 'admin,' 'login,' and 'page' in the title.
inurl:	inurl:admin	Search for pages with specific word (admin) in the URL.
allinurl:	allinurl:"login admin"	Search for web pages with 'login' and 'admin' in the URL.
intext:	intext:"password" database	Search for webpages containing the words 'password' and 'database'.
allintext:	allintext:"login password"	Search for pages where both 'login' and 'password' appear in the text.
inanchor:	inanchor:"login"	Search for webpages with 'login' in the anchor text.
allinanchor:	allinanchor:"download" report	Search for webpages with 'download' and 'report' in anchor text (hyperlinks).
filetype:	filetype:pdf OSINT	Restrict search results to a specific file type (PDF, DOC).
cache:	cache:example.com	Shows this webpage's cached version.
link:	link:google.com	Search for web pages linking to google.
related:	related:tryhackme.com	Helps to find websites similar to TryHackMe.
define:	define:OSINT	Shows dictionary definitions for the word 'OSINT'.
book:	book:OSINT	Search for online books on OSINT.
author:	author:Michael Bazzell	Search for works by Michael Bazzell.
location:	location:Dhaka	Search for information on Dhaka.
info:	info:tryhackme.com	Get a brief description of this site.
before:	before:2024 filetype:pdf	Search for PDFs published before 2024.
after:	after:2025 filetype:doc	Search for Word files published after 2025.

Table 4.

Methods to Block Search Engines from Indexing Sensitive Data to Prevent Google Dorks.

Block Sensitive Directories [1]	Block File Types [2]	Block URL Parameters [3]
Disallow: /admin/	Disallow: /*.pdf$	Disallow: /*?admin=
Disallow: /login/	Disallow: /*.docx$	Disallow: /*?config=
Disallow: /wp-admin/	Disallow: /*.xls$	Disallow: /*?backup=
Disallow: /config/	Disallow: /*.txt$	Disallow: /*?sessionid=
Disallow: /backup/	Disallow: /*.csv$	Disallow: /*?token=
Disallow: /private/	Disallow: /*.xml$	Disallow: /*?user=
Disallow: /logs/	Disallow: /*.json$	Disallow: /*?id=
Disallow: /old/	Disallow: /*.sql$	Disallow: /*?email=
Disallow: /scripts/	Disallow: /*.zip$	Disallow: /*?uid=
Disallow: /library/	Disallow: /*.rar$	Disallow: /*?ref=
Disallow: /uploads/	Disallow: /*.tar$	Disallow: /*?category=

Disallow: /session/	Disallow: /*.exe$	Disallow: /*?page=
Disallow: /configurations/	Disallow: /*.apk$	Disallow: /*?search=
Disallow: /user-data/	Disallow: /*.iso$	Disallow: /*?query=
Disallow: /bin/	Disallow: /*.bin$	Disallow: /*?price=

1. Block search engines from indexing sensitive pages or directories to avoid exposure.
2. Block search engines from indexing file types (.pdf, .docx, .csv, .xls) by blocking URLs that end with the $ sign.
3. Block search engines from indexing URLs with session parameters to prevent the indexing of duplicate or irrelevant content.

2.2.3. Generative AI Platforms

Aristotle once said, "*The more you know, the more you realize you don't know*." This idea is consistent with generative AI, which learns from large amounts of data to discover new possibilities and expand our understanding. Generative AI or GenAI is a specialized subset of artificial intelligence models designed to create new content such as text, images, audio, video, code, designs, simulations, scientific insights, and synthetic data by learning patterns from existing datasets. GenAI mimics human-like creativity and used across a wide range of industries, including business, education, healthcare, entertainment, finance, marketing, retail, manufacturing, gaming, automotive, real estate, cybersecurity, legal services, and scientific research. Despite the promising potential of GenAI, it also faces several challenges such as bias and ethical concerns, quality content, potential misuse, regulatory issues, and data security risks [7,8]. GenAI such as ChatGPT, DeepSeek, Grok, Copilot, and Gemini—are powerful resources for advanced OSINT operations for data processing, pattern recognition, report generation, social media monitoring, sentiment analysis, predictive analytics, deepfake detection, vulnerability analysis, threat detection, and geospatial intelligence. Table 5 presents detailed information about the generative AI platforms.

Table 5.
Generative AI Platforms for OSINT Investigations.

Platform	Developer	Origin	URL	OSINT Potential	Limitations
ChatGPT	OpenAI	USA	chatgpt.com	Data processing, pattern recognition, report generation, social media monitoring, sentiment analysis, predictive analytics, deepfake detection, vulnerability analysis, threat detection, geospatial intelligence.	Generate biased or incorrect information, lacks real-time data.
DeepSeek	DeepSeek	China	chat.deepseek.com		Content restrictions, data privacy aligned with Chinese regulations.
Gemini	Google	USA	gemini.google.com		Limited datasets, less global coverage and features.
Copilot	Microsoft	USA	copilot.microsoft.com		Lacks of real-time data and generate biased information.
Meta AI	Meta	USA	meta.ai		Limited to Meta platforms, and generate biased information.
Grok	xAI	USA	grok.com		Early access, less stable, specific to business-related use cases.
Claude	Anthropic	USA	claude.ai		Limited publicly available features.
DuckDuck Go AI	DuckDuckGo	USA	duckduckgo.com		Limited to search results, less personalization.
Ernie Bot	Baidu	China	yiyan.baidu.com		Limited to Chinese data, lacks real-time updates, biased info.
Doubao	ByteDance	China	doubao.com		Limited to Chinese market, lacks global coverage, biased info.
YouAI	GoMeta, Inc.	USA	you.com		Generate incorrect or biased information, lack of real-time data

2.2.4. Social Network Platforms

Social network platforms are a goldmine for OSINT investigators to collect digital footprints and personal information. While social media has brought the world to our fingertips, it has also ensured the misuse of personal information. The widespread sharing of personal data such as name, age, location, contact, images, videos, educational records, health records, social connections, relationship status, and purchasing habits on social media is undoubtedly beneficial for data brokers; however, this practice puts us in potential risks, including phishing, identity theft, impersonation, cyberbullying, and social anxiety [10,11]. Author Amy Jo Martin wrote: "*social media is changing the way we communicate and the way we are perceived, both positively and negatively. Every time you post a photo or update your status; you are contributing to your own digital footprint and personal brand*." Table 6 presents detailed information about social network platforms for OSINT investigations.

Table 6.
Social Network Platforms for OSINT Investigations.

Platform	Type	Origin	URL	Launch	Total Users	OSINT Potential on Social Networks
Facebook	Social Networking	USA	facebook.com	2004	~3 billion	Personal data, events, groups, photos, posts.
Messenger	Messaging	USA	messenger.com	2011	~1.3 billion	Messaging, location tracking, chats, media exchanges.
X (Twitter)	Microblogging	USA	x.com	2006	~500 million	Real-time news, hashtags, public sentiments.
Instagram	Photo/Reel Sharing	USA	instagram.com	2010	~2.5 billion	Location tagging, visual content, hashtags, user activity.
LinkedIn	Professional	USA	linkedin.com	2003	~1 billion	Professional connections, career data, company profiles.
Snapchat	Photo Messaging	USA	snapchat.com	2011	~750 million	Snap Map, location tracking, disappearing messages.
Reddit	Social News	USA	reddit.com	2005	~500 million	Discussions, opinions, trends, Q&A.
WhatsApp	Messaging	USA	whatsapp.com	2009	~2.7 billion	Chats, phone numbers, metadata
Skype	Communication	USA	skype.com	2003	~300 million	Profiles, metadata, user locations via IP, shared files.
WeChat	Social Networking	China	wechat.com	2011	~1.3 billion	Profiles, group activity, posts, mini-program interactions.
QQ	Messaging	China	qq.com	1998	~1 billion	Instant messaging, online games, group interactions.
Telegram	Messaging	Russia	telegram.org	2013	~800 million	Group chats, channels, usernames, shared media.
Imo	Messaging	USA	imo.im	2007	~500 million	Messaging, video calls, contacts, group activities.
Discord	Community Chat	USA	discord.com	2015	~300 million	Server conversations, user activity, shared files, metadata.
Weibo	Microblogging	China	weibo.cn	2009	~580 million	Trends, public profiles, posts, hashtags, interaction data.
VK	Social Networking	Russia	vk.com	2006	~97 million	Public profiles, photos, messages, location-based posts.
Tumblr	Microblogging	USA	tumblr.com	2007	~135 million	Public profiles, contents, multimedia, hashtags, interactions.
Quora	Q&A Platform	USA	quora.com	2009	~300 million	User expertise, profiles, questions, opinions, discussions.
Medium	Blogging	USA	medium.com	2012	~60 million	Author profiles, insights, articles, follower interactions.
Clubhouse	Audio Networking	USA	clubhouse.com	2020	~10 million	Live discussions, topic-based rooms, follower data.
Mastodon	Decentralize Social	Germany	mastodon.social	2016	~10 million	Decentralized posts, profiles, hashtags.
Line	Messaging	Japan	line.me	2011	~175 million	Messaging, shared media, user profiles, group activities.
MySpace	Social Networking	USA	myspace.com	2003	~50 million	Profiles, posts, photos, music, network interactions.
Mixi	Social Networking	Japan	mixi.jp	2004	~30 million	Profiles, posts, groups, photos, activities.
Blogger	Blogging	USA	blogger.com	1999	~100 million	Blogs, profiles, interests, comments, media sharing.

2.2.5. Video and Image-sharing Platforms

Video and image-sharing platforms are significant sources for OSINT investigations to gather personal and professional information through videos and image metadata analysis. OSINT specialists use these platforms to verify video propaganda or disinformation spread by state actors, extremist groups, or conspiracy theorists, and prevent potential security risks. Video-sharing platforms play a crucial role in OSINT investigations for analysis of social behavior, social trends, threat detection, deepfake identification, and propaganda monitoring. Table 7 presents detailed information about video-sharing platforms for OSINT investigations. OSINT practitioners also use image-sharing platforms such as Pinterest, Flickr, and Instagram to gather and analyze publicly available visual data to gain actionable intelligence to support their investigations, while reverse image search tools such as Google Images, TinEye, and Yandex are extensively used to verify image authenticity, track the origin of images, identify duplicates, and uncover additional data

associated with images. Furthermore, metadata analysis, image forensics, and OCR techniques allow investigators to extract hidden data from images and detect digital manipulations [11]. Table 8 presents detailed information about image-sharing platforms, reverse image search tools, and image forensics tools.

Table 7.
Video-sharing Platforms for OSINT Investigations.

Platform	Type	Origin	URL	Launch	Total Users	OSINT Potential on Video Platforms
YouTube	Video Sharing & Live Streaming	USA	youtube.com	2005	~2.9 billion	Analyzing contents, comments, habits, engagement, location insights, viral trends, and live interactions.
Dailymotion		France	dailymotion.com	2005	~300 million	
TikTok		China	tiktok.com	2016	~1.7 billion	
Twitch		USA	twitch.tv	2011	~140 million	
Bilibili		China	bilibili.com	2009	~300 million	
Meipai		China	meipai.com	2015	~200 million	
Kwai		China	kwai.com	2011	~600 million	
Rumble		USA	rumble.com	2013	~40 million	
Douyin		China	douyin.com	2016	~600 million	
Xiaohongshu		China	xiaohongshu.com	2013	~300 million	
Kuaishou		China	kuaishou.com	2011	~700 million	

Table 8.
Image-sharing and Forensics Platforms for OSINT Investigations.

Category	Platform	URL	OSINT Potential
Image-Sharing	Pinterest	pinterest.com	Consumer interests, pins, boards, visual trend analysis, behavior tracking.
	Flickr	flickr.com	Public photo albums, geotagging, photo metadata, location tracking.
	Pexels	pexels.com	Image metadata, profiles, consumption, image trends, usage patterns.
	Unsplash	unsplash.com	Image metadata, user profiles, search trends, content usage patterns.
	Pixabay	pixabay.com	Image usage patterns, visual content trends, commercial insights.
	500px	500px.com	Photo trends, geotagging, photographer profiles, engagement data.
	Smugmug	smugmug.com	Profiles, image metadata, photo-sharing patterns, content analysis.
Reverse Image Search	Google Images	images.google.com	Image source identification, metadata, location tracking, trace identical images.
	Yandex Images	yandex.com/images	Image source tracking, geotagging, regional usage analysis.
	Baidu Images	image.baidu.com	Web content analysis, user data mining, trend tracking, location-based insights.
	Bing Images	bing.com/images	Image source tracking, content recognition, location tracking.
	TinEye	tineye.com	Image source tracking, ownership, monitoring image abuse or piracy.
	RepostSleuth	repostsleuth.com	Image repost detection, source tracking, authenticity verification, monitoring.
Metadata, Forensics, and OCR	ExifTool	exiftool.org	Metadata extraction, image analysis, detecting edits or manipulations.
	FotoForensics	fotoforensics.com	Metadata analysis, tampering detection, authenticity verification.
	EXIF.tools	exif.tools	Metadata extraction, image details, geolocation data, timestamps.
	Metadata2Go	metadata2go.com	EXIF, IPTC, XMP metadata extraction, image hidden information.
	PhotoForensics	29a.ch/photo-forensics	Image analysis, manipulation, metadata, identifying editing traces.
	ImageForensics	imageforensic.org	Image analysis, tampering, metadata extraction, manipulation traces.
	OCR.Space	ocr.space	Text extraction, document analysis, scanned content analysis.
	i2ocr	i2ocr.com	Text extraction, OCR for images, document content analysis.
	PimEyes	pimeyes.com	Facial recognition, image search, identifying people, tracking visual content.

2.2.6. Transportation Monitoring Platforms

Transportation monitoring platforms provide comprehensive real-time data on air, sea, rail, and vehicle movements; these platforms are used in OSINT investigations to detect threats, illegal activities, smuggling, or unauthorized travel. Aviation tracking platforms such as FlightAware, FlightRadar24, PlaneFinder, and PlaneMapper are used for personal, professional, and OSINT

purposes to monitor air traffic, flight patterns, and critical insights. Tools like MarineTraffic, ShipFinder, and Maritime Database are used to monitor and analyze maritime activities, such as ship movements, ship information, and port operations. OpenRailwayMap tool provides global rail network data, Track-Trace tool monitors logistics from different carriers, while VinDecoderz, VinCheck.info, and FaxVin tools decode Vehicle Identification Numbers (VINs), verify vehicle records, and analyze automotive data for OSINT operations. Although transportation monitoring platforms have considerable significant advantages; also, these platforms pose significant concerns by exposing a large amount of sensitive data that can be exploited for surveillance, criminal activities, and cyberattacks [12,13]. Table 9 provides comprehensive insights into transportation monitoring platforms for OSINT investigations.

Table 9.
Transportation Monitoring Platforms for OSINT Investigations.

Platform	Type	URL	OSINT Potential	Limitations
FlightAware	Flight Tracking	flightaware.com	Flight tracking, real-time flight data, aircraft information, flight paths, airport status.	Real-time data limited for non-premium users.
FlightRadar24	Flight Tracking	flightradar24.com	Real-time flight tracking, flight data, aircraft information, flight paths, airport status	Free version has limited access to flight data.
PlaneFinder	Flight Tracking	planefinder.net	Flight tracking, historical data, aircraft identification.	Limited coverage for private flights.
PlaneMapper	Flight Tracking	planemapper.com	Flight tracking, aircraft details, flight routes.	Free version provides limited data access.
OpenRailwayMap	Railway Tracking	openrailwaymap.org	Rail network visualization, track info, station details.	Limited to rail infrastructure data.
MarineTraffic	Marine Traffic	marinetraffic.com	Ship tracking, vessel identification details.	Limited access for non-subscribers.
ShipFinder	Marine Traffic	shipfinder.com	Vessel details, live maritime data, location insights.	Limited data for smaller/private vessels.
Maritime Database	Marine Traffic	maritime-database.com	Vessel data, ownership, maritime information.	Limited data for smaller or private ships.
Track-Trace	Package Tracking	track-trace.com	Package tracking, courier info, global logistics monitoring.	Limited support for some shipping companies.
VinDecoderz	Vehicle Identification	vindecoderz.com	VIN lookup, vehicle details, manufacturer information, historical data, ownership records.	Limited details for unregistered vehicles or obscure brands.
VinCheck.info	Vehicle Identification	vincheck.info	VIN lookup, vehicle history, theft records, accident reports, title status.	Some data require premium access.
FaxVin	Vehicle Identification	faxvin.com	VIN lookup, vehicle history, accident reports, title status, salvage records.	Relies on available database info.

2.2.7. Geospatial Platforms

Geospatial platforms are specialized tools designed to collect, analyze, visualize, and extract actionable geographical or location-based information. They are widely utilized in different industries, including transportation, logistics, navigation, urban planning, agriculture, defense, telecommunications, environmental management, and disaster response. Geospatial platforms are essential in OSINT for mapping, location tracking, route planning, traffic analysis, satellite imagery, infrastructure mapping, and environmental insights for threat assessments and crisis management [14,15]. Table 10 presents detailed information on geospatial monitoring platforms for OSINT investigations.

Table 10.

Geospatial Monitoring Platforms for OSINT Investigations.

Platform	Origin	URL	OSINT Potential	Limitations
Google Maps	USA	maps.google.com	Location tracking, reverse geocoding, street view analysis, business listings.	Accuracy issues, outdated data, internet dependency, privacy concerns, and GPS errors.
Google Earth	USA	earth.google.com	Satellite imagery, geospatial analysis, location tracking, terrain mapping.	Limited real-time data, outdated imagery, and high system requirements.
Apple Maps	USA	apple.com/maps	Geolocation, navigation, street imagery, route planning, local listings.	Limited global coverage, less detail in rural areas, and occasional inaccuracies.
OpenStreetMap	UK	openstreetmap.org	Geospatial data, mapping, route analysis, location tracking, user-contributed geographic.	Crowdsourced data may contain inaccuracies or outdated information.
Yandex Maps	Russia	yandex.com/maps	Location tracking, business listings, street view analysis, Russian region geolocation.	Limited global coverage, inaccurate data in some regions.
Bing Maps	USA	bing.com/maps	Geolocation data, street imagery, route planning, location tracking, business information.	Limited global coverage, lower detail in rural areas, and outdated imagery.
Baidu Maps	China	map.baidu.com	Location tracking, street-level imagery, route analysis, location tracking, local business listings.	Limited to China, poor global coverage, and language barriers.
Amap	China	amap.com	Geolocation data, street-level imagery, route planning, traffic information, location tracking.	Limited global data, regional bias, and fewer features outside China.
KakaoMap	South Korea	map.kakao.com	Geolocation data, street imagery, routing, location tracking, local businesses.	Limited to South Korea, poor global coverage, and language barriers.
Naver Maps	South Korea	map.naver.com	Geolocation, street imagery, route planning, location tracking, local listings.	Focused mainly on South Korea, limited global data, and language barriers.
Bee Maps	USA	beemaps.com	Geolocation, route planning, street imagery, location tracking, weather data.	Limited coverage, smaller user base, limited global data
Living Atlas	USA	livingatlas.arcgis.com	Geospatial data, mapping, satellite imagery, location analysis, demographic information.	Limited to Esri's dataset, regional bias, and fewer global details.
Zoom Earth	USA	zoom.earth	Satellite imagery, real-time weather tracking, geospatial analysis, environmental monitoring.	Limited to weather and natural phenomena tracking, lower resolution, and no detailed map.
Satellites.pro	UK	Satellites.pro	Satellite imagery, geospatial analysis, real-time tracking, environmental monitoring.	Restricted high-resolution imagery, focused on environmental monitoring.
OpenRailwayMap	UK	openrailwaymap.org	Railway data, geospatial mapping, route analysis, infrastructure details, real-time tracking.	Limited to railway-related data, not suitable for non-railway OSINT.

2.2.8. Username, Email, and Phone Number Lookup Tools

The core concept of OSINT is to collect publicly available data from the Internet without breaking any systems. In this context, these username, email, and phone number lookup tools collectively empower OSINT professionals to gather, validate, and analyze publicly available information to conduct comprehensive investigations. A username serves as a distinct digital identifier that allows individuals to be recognized online. Username enumeration tools assist OSINT professionals in tracking a user's digital footprint or digital shadow across multiple platforms and services. Popular username lookup tools, such as Namechk, Sherlock, Maigret, WhatsMyName, InstantUsernameSearch, NameCheckerr, Username Social, and User Search, verify a subject's digital presence by searching their usernames across multiple platforms. Email lookup tools such as theHarvester, EmailHarvester, GHunt, Hunter.io, Voila Norbert, GetProspect, Anymail Finder, Epieos, Spokeo, That's Them, and SimpleEmailRep tools are utilized to extract, verify, and analyze information linked to email addresses for professional, personal, or threat intelligence purposes. Meanwhile, phone number lookup tools such as PhoneInfoga, TrueCaller, Whitepages, EMobileTracker, NumLookup, AnyWho, and MobileLocation are used to identify the phone number owner, location, carrier, and online footprints [5,16,17]. These username, email, and phone number lookup tools are vital in OSINT investigations for different groups of individuals and organizations including digital marketers, security professionals, law enforcement, researchers, private investigators, OSINT analysts, social media managers, tech enthusiasts, recruiters, and sales professionals. Table 11 provides an overview of tools for username, email, and phone number lookup.

Table 11.
Username, Email, Phone Number Tracking and Digital Footprint Analysis Tools for OSINT Investigations.

Tool Name	Type	URL	Platform	OSINT Potential
Sherlock	Username Enumeration	github repository	Command Line (CLI)	Track usernames across multiple platforms, verify their existence, and map digital footprints.
Maigret	Username Enumeration	github repository	Command Line (CLI)	
Namechk	Username Enumeration	namechk.com	Web-Based	
WhatsMyName	Username Enumeration	whatsmyname.app	Web-Based	
InstantUsernameSearch	Username Enumeration	instantusername.com	Web-Based	
NameCheckerr	Username Enumeration	namecheckerr.com	Web-Based	
Netlify	Username Enumeration	usernamechecker.netlify.app	Web-Based	
Username Social	Username Enumeration	username.social	Web-Based	
User Search	Username Enumeration	usersearch.org	Web-Based	
Socialscan	Username/Email	github repository	Command Line (CLI)	Email validation, domain mapping, personal detail extraction, investigative outreach.
theHarvester	Email Lookup	github repository	Command Line (CLI)	
EmailHarvester	Email Lookup	github repository	Command Line (CLI)	
GHunt	Email Lookup	github repository	Command Line (CLI)	
Hunter.io	Email Finder/Verification	hunter.io	Web-Based	
Voila Norbert	Email Finder/Verification	voilanorbert.com	Web-Based	
GetProspect	Email Finder/Verification	getprospect.com	Web-Based/Extension	
Anymail Finder	Email Finder/Verification	anymailfinder.com	Web-Based	
Epieos	Email/Phone Lookup	epieos.com	Web-Based	
Spokeo	Email/Phone/Address Lookup	spokeo.com	Web-Based	
That's Them	Email/Phone/Address Lookup	thatsthem.com	Web-Based	
SimpleEmailRep	Email Finder/Verification	emailrep.io	Web-Based	
PhoneInfoga	Phone Number Lookup	github repository	Command Line (CLI)	Tracking mobile numbers, identifying owners, tracing location, and phone number-related information.
TrueCaller	Phone Number Lookup	truecaller.com	Mobile App/Web-Based	
Whitepages	People/Phone Lookup	whitepages.com	Web-Based	
EMobileTracker	Phone Number Lookup	emobiletracker.com	Web-Based	
NumLookup	Phone Number Lookup	numlookup.com	Web-Based	
AnyWho	Phone Number Lookup	anywho.com	Web-Based	

Mobile Location	Phone Number Lookup	mobile-location.com	Web-Based

2.2.9. Domain, Subdomain, and DNS Analysis Tools

Domain, subdomain, and DNS analysis tools provide comprehensive reconnaissance for OSINT investigations, penetration testing, and threat analysis by identifying vulnerabilities, malicious activities, and domain infrastructure. Amass, Sublist3r, DNSRecon, Dmitry, DNSDumpster, NSLookup, ViewDNS, and C99 tools are commonly used for subdomain discovery, DNS reconnaissance, and infrastructure mapping [18]. While ICANN WHOIS, WHOIS, CentralOps, Crt.sh, and BuiltWith identify domain ownership details, SSL/TLS certificate data, and website technologies, whereas OpenPhish and URLscan focus on phishing domains, malicious URLs, and suspicious activities. Table 12 presents the functionalities of these tools for domain, subdomain, and DNS analysis.

Table 12.
Domain, Subdomain, and DNS Analysis Tools for OSINT Investigations.

Tool Name	Type	URL	OSINT Potential	Limitations
Amass	Subdomain Enumeration	github repository	Subdomain enumeration, DNS information gathering, network mapping, footprinting.	Relies on public data, missing private or hidden information.
Sublist3r	Subdomain Enumeration	github repository	Fast subdomain discovery from search engines, DNS, and other sources	Relies on third-party sources, dependent on search engine indexing.
DNSRecon	DNS Reconnaissance	github repository	Subdomain enumeration, DNS record collection, zone transfers, reverse DNS lookups.	Limited to DNS-related information, dependent on publicly available DNS records.
Dmitry	Enumeration	github repository	Gathers WHOIS, subdomains, and DNS information.	Limited to DNS and WHOIS info.
C99	Subdomain Enumeration	subdomainfinder.c99.nl	Subdomain discovery, domain enumeration, DNS records, IP addresses, web infrastructure mapping.	Subdomains missing; results depending on publicly available data.
ViewDNS	WHOIS & DNS	viewdns.info	Domain & network information, DNS records, IP analysis, website history, subdomains.	Limited to public DNS data, no real-time updates, lacks of privacy-protected details.
ICANN WHOIS	WHOIS Lookup	lookup.icann.org	Domain registration details, ownership information, contact data, DNS records, domain status.	Limited to ICANN domains.
WHOIS	WHOIS Lookup	who.is, whois.com	Domain registration details, owner information, WHOIS lookup, domain status, DNS records.	Restricted access for privacy-protected domains, some services require subscription.
CentralOps, Robtex	WHOIS & DNS	centralops.net, robtex.com	WHOIS lookup, DNS records, domain analysis, IP lookup, network diagnostics.	Provide incomplete data for newly registered domains or less popular services.
BuiltWith	Website Analysis	builtwith.com	Technology stack, website analytics, traffic data, tech usage, company profiles.	Limited access to non-paying users.
DNS Dumpster	DNS Reconnaissance	dnsdumpster.com	DNS records, domain information, subdomains, IP addresses, network	Limited data without login, lack full DNS records.

			mapping.	
NSLookup	DNS Reconnaissance	nslookup.io	DNS lookup, domain information, IP resolution, reverse DNS, server query.	Limited to DNS records; no deeper domain intelligence.
URLscan	Website Scanning	urlscan.io	URL analysis, website scanning, domain information, security checks, traffic analysis.	Limited data access for non-registered users.
OpenPhish	Phishing Detection	openphish.com	Phishing detection, URL analysis, real-time alerts, malware tracking.	Limited to phishing data; do not cover all emerging phishing threats.
Host, dmns	DNS & Subdomains	host.io, dmns.app	Domain information, subdomain discovery, IP addresses, DNS records, WHOIS lookup.	Limited data for private or newly registered domains.
Crt.sh	Certificate Logs	crt.sh	Certificate transparency logs, SSL certificate details, subdomain discovery.	Limited to SSL/TLS certificates.
IntelX	Data Aggregation	intelx.io	Aggregates OSINT datasets, including WHOIS, domains, and IPs.	Full access requires a subscription.

2.2.10. People Search and Dating Platforms

People search and dating platforms are extremely useful for OSINT investigators, law enforcement, journalists, and researchers. People search platforms such as Pipl, BeenVerified, Peppercat, Instant Checkmate, Intelius, US Search, ZabaSearch, TruthFinder, Radaris, and PeopleLooker assist with OSINT by gathering personal data, contact info, social media accounts, criminal records, and employment history. While dating platforms such as Tinder, Bumble, OkCupid, Match, and eHarmony provide valuable insights into personal interests, behavioral patterns, group affiliations, and social/professional networks [19]. These platforms are used by investigators from all professions to verify identities, conduct background checks, track digital footprints, and evaluate internet behavior; and that's why these platforms have become indispensable for comprehensive open-source intelligence collection. Table 13 provides detailed information on people's search and dating platforms used for data collection.

Table 13.
People and Dating Platforms for OSINT Investigations.

Tool Name	Type	URL	OSINT Potential	Limitations
Pipl	People Search	pipl.com		Requires subscription for full access, limited to public records
BeenVerified	Background Check	beenverified.com		Outdated data, limited records, no real-time updates.
Peppercat	People Search	peppercat.org		limited coverage for some regions, relies on external sources.
Instant Checkmate	Background Check	instantcheckmate.com		Subscription-based, US-focused, limited free access.
Intelius	Background Check	intelius.com		Subscription-based, US-focused.
US Search	People Search	ussearch.com		Subscription-based, limited and outdated information.
TruthFinder	Background Check	truthfinder.com		Limited free access, paid features for

				comprehensive reports.
ZabaSearch	People Search	zabasearch.com	Background checks, people search, social media profiles, criminal records, contact info, public records, family info, relationship preferences, employment history, social connections, geo-location.	Limited free access, only US-based data.
Radaris	Public Record Search	radaris.com		Relies on public records, incomplete or outdated data.
PeopleLooker	Background Check	peoplelooker.com		US-focused, subscription required for full access.
Tinder	Dating & Networking	tinder.com		Limited to dating purposes, privacy concerns.
Bumble	Dating & Networking	bumble.com		Privacy controls, some profiles may be hidden.
OkCupid	Dating & Networking	okcupid.com		Limited access, privacy restrictions.
Match	Meetup Platform	match.com		Paid access, privacy restrictions limit data.
Meetup	Meetup Platform	meetup.com		Limited to group data, account required for details.
eHarmony	Dating & Networking	eharmony.com		Subscription needed for full access, limited data.

2.2.11. Archives, Alerts, and Monitoring Tools

Archives serve as essential repositories that preserve historical records, documents, logs, and diverse forms of digital data for future access and reference. Since websites often modify or remove content, these archived versions help recover lost information, including deleted web pages, social media posts, or news articles. In OSINT, archive platforms such as Wayback Machine, Archive.is, Annas Archive, and Common Crawl help recover/identify deleted or altered content, monitor digital footprints, verify claims, and assist in investigations. Additionally, alerting tools such as Google Alerts and Talkwalker Alerts help OSINT practitioners monitor real-time updates to specific keywords or phrases across web content and notify users. Monitoring tools such as Shodan, Distill.io, and Visualping detect potential threats, security breaches, or unauthorized changes to websites and networks. WikiLeaks is another important platform for leaked classified documents, which provide valuable information for investigative purposes. These tools collectively contribute to the systematic collection and analysis of open-source data in intelligence and security research. Table 14 provides detailed information about archive, alert, and monitoring tools.

Table 14.
Archives, Alerts, and Monitoring Tools for OSINT Investigations.

Platform	Type	URL	OSINT Potential	Limitations
Wayback Machine	Web Archive	archive.org	Historical website snapshots, content retrieval, archived data, past versions.	Limited archive for newer websites or private pages.
Archive.is	Web Archive	archive.is	Website snapshots, content archiving, historical page retrieval.	Limited archive for dynamic or blocked content.
Annas Archive	Web Archive	annas-archive.org	Free digital library offering books, articles, and research papers.	Lack of recent publications; some resources restricted.
Common Crawl	Web Archive	commoncrawl.org	Web scraping data, content extraction, historical site snapshots,	Vast data, requires technical expertise to navigate.

			data analysis.	
Google Alerts	Alerts and Triggers	google.com/alerts	Real-time monitoring, keyword tracking, news updates, content alerts.	Alerts are based on Google's indexing.
Talkwalker Alerts	Alerts and Triggers	talkwalker.com/alerts	Brand monitoring, keyword tracking, media mentions, content alerts.	Slower notifications in some cases.
Distill.io	Web Monitoring	distill.io	Web scraping, content monitoring, data extraction, alert setup.	Limited tracking for free accounts.
Visualping	Web Monitoring	visualping.io	Website monitoring, change detection, content alerts, page tracking.	Not suitable for dynamic or complex websites.
WikiLeaks	Document Repository	wikileaks.org	Classified documents, government leaks, corporate disclosures, whistleblower data.	Accessing or sharing may be legally restricted.

2.2.12. Fact-Checking, Sanctions, and Threat Analysis Tools

In OSINT investigations, several tools and databases are specifically designed to address fact-checking, sanctions monitoring, censorship detection, threat identification, and crime detection. PolitiFact, EUvsDisinfo, and Google Fact Check—serve as key resources for fact-checking, debunking misinformation, and tracking disinformation campaigns. For monitoring sanctions, tools like OFAC Sanctions, UN Sanctions, EU Sanctions, and OpenSanctions help investigators identify high-risk individuals or entities linked to illicit activities. OONI, Citizen Lab, and TechinQuiry specialize in detecting internet censorship, surveillance, and tech investigations, while UNODC Data, InSight Crime, and Militants Project provide statistics on drug trafficking and organized crime. Together, these tools help OSINT researchers to cross-verify information, track financial sanctions, monitor security risks, and investigate criminal activities. Table 15 presents detailed information about fact-checking, sanctions, and threat analysis tools.

Table 15.
Fact-checking, Sanctions, Censorship, Threats, and Crime-identifying Tools for OSINT investigations.

Platform	Type	URL	OSINT Potential	Limitations
PolitiFact	Fact-Checking	politifact.com	Fact-checking claims, identifying misinformation, & political statements.	Limited to political content and US-focused.
EUvsDisinfo	Disinformation Tracker	euvsdisinfo.eu	Disinformation trends, narratives, countermeasures.	Covers EU disinformation; limited global focus.
Google Fact Check	Fact-Checking	toolbox.google.com/factcheck	Fact-checking data, claim verification, source credibility, misinformation detection.	Relies on third-party fact-checking sources.
OFAC Sanctions	Sanctions Database	sanctionssearch.ofac.treas.gov	Sanctioned entities, financial restrictions, trade prohibitions, compliance monitoring.	U.S. government focus.
UN Sanctions	Sanctions Database	scsanctions.un.org	Sanctioned individuals, entities, travel bans, asset freezes, trade restrictions monitoring.	Focuses on entities sanctioned by the UN.
EU Sanctions	Sanctions Database	sanctionsmap.eu	Restricted entities, asset freezes, trade restrictions, EU-specific compliance	EU-focused data.

			data.	
OpenSanctions	Sanctions Database	opensanctions.org	PEPs, watchlists, global compliance data, international sanctions enforcement.	Updates less frequent than official sources.
OONI	Censorship Monitoring	ooni.org	Internet censorship data, network interference, online freedom, global internet monitoring.	Limited to reported data, lack coverage in some regions.
Citizen Lab	Censorship Analysis	citizenlab.ca	Security research, internet surveillance, censorship, human rights monitoring, privacy threats.	Do not offer direct investigative tools.
TechinQuiry	Threat Analysis	techinquiry.org	Exposes tech vulnerabilities, provides threat intelligence, tracks data breaches.	Limited data scope, primarily focuses on U.S. entities.
UNODC Data	Global Crime Data	dataunodc.un.org	Crime statistics, drug-related data, trafficking patterns, policy analysis.	Relies on government-reported data.
Numbeo	Crime Statistics	numbeo.com/crime	Crime data, safety index, global comparisons, crime trends.	Relies on user reports; data may be subjective.
Militants Project	Extremist Statistics	mappingmilitants.org	Tracks militant organizations, activities, alliances, conflicts.	Limited to publicly available data.
InSight Crime	Crime Analysis	insightcrime.org	Crime analysis, criminal networks, violence trends, regional insights.	Focused primarily on Latin America and the Caribbean.
PRIO	Arms Trade Analysis	nisatapps.prio.org/armsglobe	Conflict data, global security trends, arms trade data.	Data relies on publicly reported sources.

2.2.13. Open-source Intelligence Tools and Techniques

This section discusses the primary tools and techniques used in OSINT to collect, analyze, and interpret publicly available data. These tools help investigators identify suspects, collect evidence, locate missing people, detect fraud, and monitor threats. Fig. 2. shows how OSINT tools are used in real-time investigations and highlights the specific purpose of each tool. Table 16 provides a comprehensive summary of key OSINT tools used in reconnaissance, security assessments, investigations, and intelligence gathering.

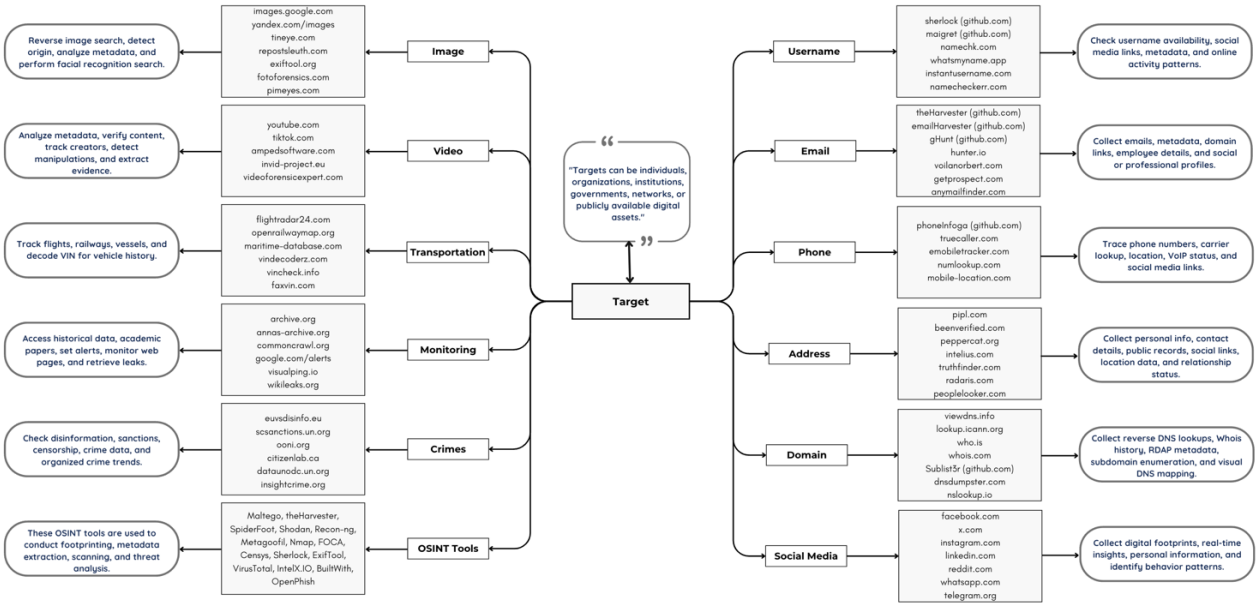

Fig. 2. Techniques and Potential of OSINT Tools/Platforms.

Table 16.

OSINT Tools for Reconnaissance, Security Assessments, Investigations, and Intelligence Gathering.

OSINT Tool	Platforms	Purpose
Maltego	Windows, macOS, Linux	Graph-based link analysis to visualize relationships between individuals, organizations, domains, and other entities.
theHarvester	Windows, macOS, Linux	Harvests emails, subdomains, metadata, and IP addresses from search engines and online databases.
SpiderFoot	Windows, macOS, Linux	Collects IP addresses, domains, emails, DNS records, vulnerabilities, and network data from multiple sources.
Shodan	Web-based	Search engine for internet-connected devices (IoT, servers, and cameras) that is used to reveal vulnerabilities, network infrastructure, and service configurations.
Recon-ng	Windows, macOS, Linux	Modular web reconnaissance framework for gathering data and assessing web application vulnerabilities.
Metagoofil	Windows, macOS, Linux	Extracts metadata from public documents (e.g., PDFs, DOCs, Excel) to uncover sensitive information.
Nmap	Windows, macOS, Linux	Network mapping and port scanning tool used to discover hosts, services, open ports, and assess network vulnerabilities.
FOCA	Windows	Extracts and analyzes metadata from documents to detect internal sensitive network information, such as usernames, software, and servers.
Censys	Web-based	Search engine and analysis tool for discovering internet-connected devices, services, and security postures.
Sherlock	Windows, macOS, Linux	Identifies social media profiles on over 300+ platforms based on a given username.
ExifTool	Windows, macOS, Linux	This tool reads, writes, and edits metadata in files (images, videos, documents) to extract geolocation, device information, and timestamps.
VirusTotal	Web-based	A platform that detects malware, phishing, and suspicious

		activities by analyzing files, URLs, and IP addresses using multiple antivirus engines and tools.
IntelX.IO	Web-based	Searches leaked data, historical records, darknet content, and public archives for intelligence and digital forensics purposes.
BuiltWith	Web-based	Analyzes websites to discover technologies such as CMS, analytics tools, frameworks, web servers, and more in order to conduct reconnaissance and profile of the target's technological infrastructure.
OpenPhish	Web-based	A platform that provides real-time feeds and intelligence on active phishing URLs and campaigns for detection and prevention.

3. OSINT in Addressing Cybercrime

Cybercrime is an illegal activity conducted for malicious purposes against individuals, organizations, or governments using computers, networks, or digital devices. Essentially, any crime that uses digital technology as a tool or target can be considered a cybercrime, including threats, leaks, identity theft, spamming, online harassment, espionage, data breaches, offensive content, and copyright infringement. With the unprecedented technological changes, the nature of crime is also changing. As the world becomes more connected, criminals are finding new ways to exploit technology for their own gain. From identity theft to large-scale cyberattacks, these crimes can have serious consequences for individuals, businesses, and even entire nations. Cybercriminals can operate independently or as part of organized groups or even backed by governments. They often hide behind the anonymity of the internet, using techniques like encryption and cryptocurrency to avoid detection. Cybercrime isn't just about stealing money—it can also involve disrupting systems, stealing sensitive information, and even carrying out cyberattacks with political or criminal motives [20, 21]. OSINT combats cybercrime by gathering publicly available information to track cyber threats, identify criminals, and prevent attacks through digital forensics techniques such as data analysis, network traffic monitoring, and digital footprints examined to uncover evidence and establish connections between malicious activities [11]. Cybercrime is broadly categorized into three types: crimes against individuals, crimes against organizations, and crimes against governments. Each type of cybercrime has unique challenges and consequences. While individuals face personal security threats, organizations struggle with financial and reputational risks, and governments deal with national security concerns [22, 23]. Fig. 3. illustrates a detailed overview of cybercrimes.

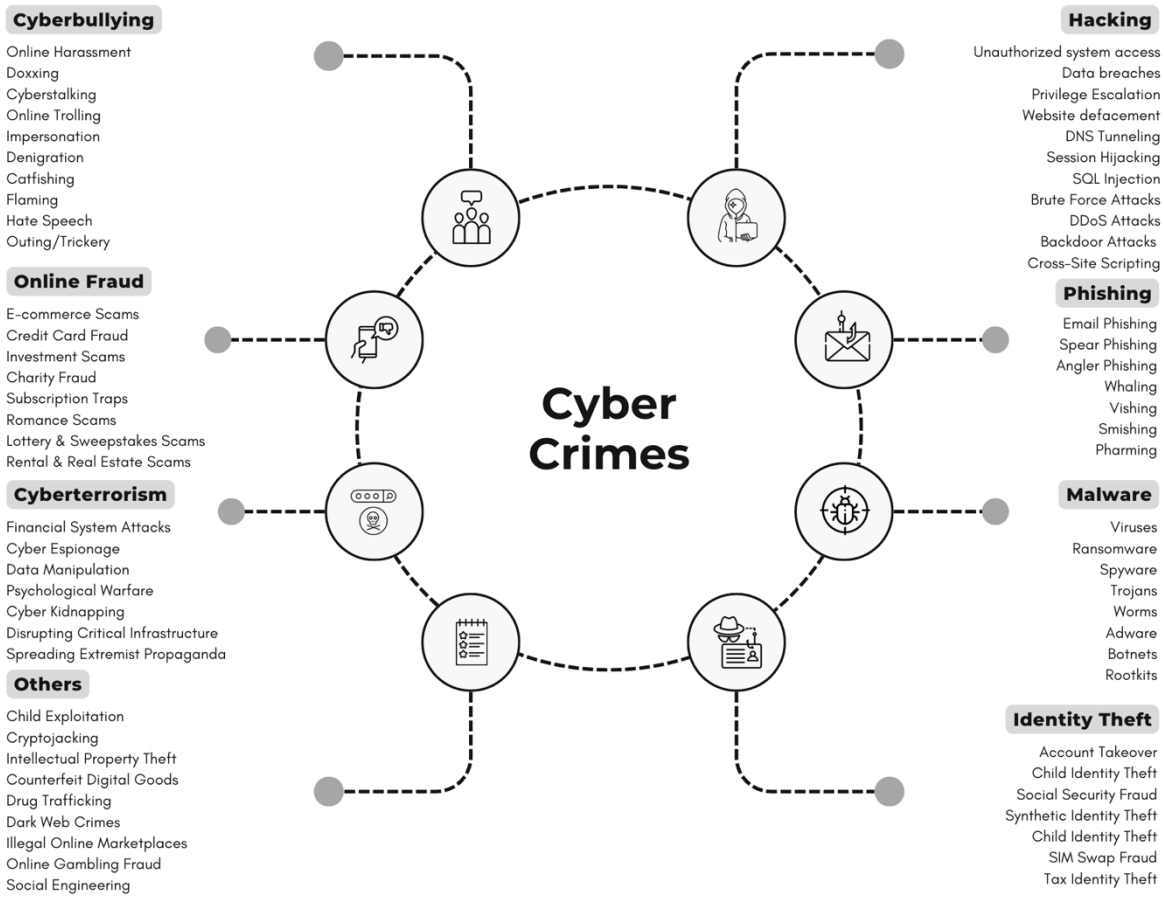

Fig. 3. Different Types of Cybercrimes.

Cybercrime is a significant global challenge, with estimated costs ranging trillions of dollars each year. Despite its widespread impact, cybercrime remains an invisible threat. According to the 2024 World Cybercrime Index, several developed countries are at the top of the list in cybercriminal activities, such as Russia, Ukraine, China, USA, Nigeria, Romania, North Korea, the United Kingdom, Brazil, and India. According to Cybercrime Magazine, the annual cost of cybercrime worldwide reached $8 trillion in 2023 and is expected to increase to $10.5 trillion by 2025. Meanwhile, emerging cyber threats include AI-generated phishing, APTs, deepfakes, and advanced ransomware services that allow less-skilled attackers to target critical infrastructure like hospitals and power grids. These AI-driven attacks are becoming more common; they use machine learning and exploitation frameworks to identify and exploit vulnerabilities, while quantum computing breaks current encryption methods. Combating these cybercrimes requires sophisticated security protocols, real-time threat detection, and collaboration between governments and industries to share intelligence and strengthen cyber defenses [24, 25]. Table 17 presents a detailed overview of major global cyberattacks.

Table 17.
Overview of Major Global Cyberattacks: Name, Year, Affected Area, Attack Type, Impact, and Description.

Name	Year	Affected Area	Attack Type	Impact	Description
Morris Worm	1988	USA	Worm	~10% of ARPANET computers affected.	One of the first internet-based worms, exploiting Unix vulnerabilities and causing significant disruption.
Love Bug	2000	Worldwide	Worm	~10 million Windows personal computers infected.	A destructive worm that spread via email, infecting millions of Windows PCs and overwriting files.
Code Red	2001	Worldwide	Worm	~1 million systems affected	The worm targeted vulnerabilities in

					Microsoft IIS web servers, defaced websites, and attempted DoS attacks.
Titan Rain	2003	USA	APT	Compromised vast amounts of sensitive data.	A series of cyberattacks stealing sensitive data from the U.S. defense and corporate networks.
Sasser	2004	Worldwide	Worm	Millions of systems infected	The worm exploited a vulnerability in the Local Security Authority Subsystem Service (LSASS) in Windows.
Operation Aurora	2009	USA	APT	Intellectual property theft.	A series of coordinated cyberattacks performed by APT targeting intellectual property.
Conficker	2008	Worldwide	Worm	~15 million computers across 190 countries infected.	The worm exploited vulnerabilities in Windows and created botnets by targeting unpatched systems.
Stuxnet	2010	Iran	Worm	Iranian nuclear program disrupted.	A sophisticated computer worm that sabotaged Iran's nuclear facilities by targeting industrial control systems.
Fukushima Cyberattack	2011	Japan	Cyber Espionage	Targeted nuclear power plant facilities	Hackers targeted the internal networks of the Fukushima Daiichi Nuclear Power Plant.
Yahoo Data Breach	2013	Worldwide	Data Breach	3 billion accounts compromised	The breach was discovered in 2016 but occurred earlier, with two separate attacks linked to state-sponsored actors.
Target Data Breach	2013	USA	Data Breach	40 million credit and debit card exposed	Hackers gained access to Target's network through a third-party vendor during the holiday season.
Heartbleed	2014	Worldwide	Data Breach	affected approximately 17% of all secure web servers	A severe bug in OpenSSL's implementation of the TLS heartbeat extension allowed attackers to steal sensitive data.
Sony Pictures Hack	2014	USA	Data Breach	Confidential data leaked	Hackers leaked sensitive company data, emails, and unreleased films.
Anthem Data Breach	2015	USA	Data Breach	Exposed sensitive data of nearly 80 million people	The breach targeted Anthem Inc., stole sensitive data, and was attributed to a sophisticated cyber-espionage campaign.
Ukrainian Power Grid Attack	2015	Ukraine	APT	Power outages for 230,000 people	Russian "Sandworm" hacked Ukraine's power grid, causing 1-6 hours outages during the Russo-Ukrainian War.
Bangladesh Bank Heist	2016	Bangladesh	SWIFT Fraud	$81 million stolen, $951 million attempted for transfer.	Hackers exploited SWIFT systems to transfer funds fraudulently from Bangladesh Bank.
Dyn DDoS Attack	2016	Worldwide	DDoS	Major Internet outage for hours	Hackers launched a massive DDoS attack on Dyn using a botnet of IoT devices infected with the Mirai malware.
Uber Data Breach	2016	Worldwide	Data Breach	Exposed data of 57 million users	Hackers gained access to Uber’s internal systems and stole sensitive data, including 600,000 driver licenses.

WannaCry	2017	Worldwide	Ransomware	affected over 200,000 devices across 150 countries	Exploited Windows vulnerabilities, encrypting files and demanding Bitcoin payments.
NotPetya	2017	Ukraine	Wiper Malware	~$10 billion damages	A malware that targeted global businesses, encrypting systems and causing data loss via Ukrainian software.
Equifax Data Breach	2017	USA	Data Breach	Exposed data of 147 million people	Hackers exploited vulnerabilities to steal credit records and personal information.
Marriott Data Breach	2018	Worldwide	Data Breach	Exposed data of 500 million people	Attackers gained unauthorized access to Starwood's guest reservation database and exposed sensitive information.
Cambridge Analytica Scandal	2018	Worldwide	Data Misuse	Exposed data of 87 million Facebook users	Cambridge Analytica harvested Facebook users' data without consent for political consulting purposes.
Capital One Data Breach	2019	USA/Canada	Data Breach	Exposed data of 100 million customers	A former AWS employee exploited a misconfigured firewall to access Capital One's cloud servers.
Baltimore Ransomware	2019	USA	Ransomware	$18-$20 million damages	The attack exposed vulnerabilities in the city’s IT infrastructure, outdated software and unpatched systems.
Twitter Bitcoin Scam	2020	Worldwide	Social Engineering	Compromised ~130 high-profile accounts	Hackers gained access to Twitter’s (X) internal tools and hijacked high-profile, including Musk, Obama, and Biden.
SolarWinds Attack	2020	Worldwide	Supply Chain	Thousands of networks compromised	Compromised updates in SolarWinds' Orion software led to breaches in government and corporate systems.
RockYou2021	2021	Worldwide	Data Breach	8.4 billion passwords exposed	Hackers leaked a database of over 8.4 billion exposed plaintext passwords from various data breaches.
Colonial Pipeline Attack	2021	USA	Ransomware	Fuel supply disrupted for several days	A ransomware attack by DarkSide that disrupted fuel supplies in the US, causing shortages and economic impact.
Notorious Kaseya	2021	Worldwide	Ransomware	Affected 1,500+ businesses	Hackers exploited a vulnerability in Kaseya's software and demanded a $70 million ransom to decrypt the systems.
Log4Shell Exploit	2021	Worldwide	Zero-Day Exploit	Widespread vulnerabilities	A critical vulnerability in Apache Log4j, allowing attackers to execute remote code execution worldwide.
Acer Cyberattack	2021	Worldwide	Ransomware	Hackers demanded a $50 million ransom.	Hackers exploited vulnerabilities in Acer’s network, encrypted critical data, and the attack was attributed to the REvil ransomware group.
T-Mobile Data Breach	2021	USA	Data Breach	Exposed data of 50 million customers	Hackers accessed T-Mobile servers, stealing names, Social Security numbers, phone numbers, and driver's licenses.
Costa Rica Cyberattack	2022	Costa Rica	Ransomware	Hackers demanded a $20 million ransom.	The Conti ransomware group targeted multiple government institutions,

					disrupting healthcare, customs, and tax systems.
Optus Data Breach	2022	Australia	Data Breach	Exposed data of 10 million customers	Hackers exploited API vulnerabilities to access names, addresses, and ID numbers.
Medibank Data Breach	2022	Australia	Data Breach	Exposed health data of 9.7 million people	Attackers stole and leaked sensitive medical records on the dark web.

4. Opportunities in OSINT

OSINT offers significant opportunities in various fields. It helps in different domains to gather actionable intelligence from public sources, monitor threats, and make informed decisions to conduct investigations. Currently, approximately 80% to 90% intelligence activities conducted by Western law enforcement agencies and national organizations rely on OSINT [26]. The following section explores the key opportunities of OSINT, demonstrating its growing significance in addressing modern challenges and decision-making processes.

Cybersecurity & Threat Intelligence: OSINT helps detect emerging threats by analyzing public data such as social media, forums, and breach databases [27]. OSINT tools allow cyber professionals to identify malicious actors, phishing campaigns, leaked credentials, counterfeit websites, trademark abuse, or dark web activities. For example, Have I Been Pwned tool allows users to check if their personal information has been exposed by searching across multiple data breaches; Shodan tool helps find exposed devices, such as routers or cameras, to prevent cyberattacks.

Government & Defense: Governments and defense agencies use OSINT to gather intelligence and monitor potential threats, including extremist activities, foreign influence, protests, military movements, disaster management, disinformation, and asset tracing. Platforms like Google Earth and Sentinel Hub track military movements and changes in infrastructure during conflicts [28]. Social media tools like Geofeedia and Dataminr help agencies monitor real-time events and assess public sentiment. For example, Bellingcat used geolocation and flight radar data from sources like Flightradar24 to verify videos of Syrian chemical attacks. MarineTraffic tracks global ship movements with AIS signals, while Ushahidi collects social media reports.

Law Enforcement & Counterterrorism: Open-source intelligence helps law enforcement to investigate cybercriminal behavior, identify their networks, and gather intelligence on potential terrorist threats using OSINT tools. Agencies use these tools to monitor online extremist groups, and their communications across multiple social platforms. Also, geolocate targets using photos, metadata, or online footprints. The CIA, FBI, MI6, and Europol have used these OSINT tools in different operation to track organized crime syndicates worldwide [29].

Corporate & Business Analysis: Companies use OSINT to gather real-time insights into competitors, market trends, and consumer behavior, as well as monitor public data related to financial reports, political instability, and reputational threats for making informed business decisions and staying competitive. For example, platforms like Crunchbase and Glassdoor help OSINT investigators analyze the company's funding, mergers, acquisitions, employee reviews, and salary data, while SimplyWall and MarketScreener assist in tracking the company's financial condition, stock performance, and investment insights [30].

Academic & Scientific Research: OSINT plays a significant role in academic and scientific studies. Researchers use OSINT to collect public information from multiple sources, such as government databases, open-access journals, repositories, and online forums, to study social trends, public opinion, economic indicators, and historical events [31]. For example, Google Scholar is an open-access platform researchers use to track new studies and publications in their fields, while Academia.edu and ResearchGate facilitate academic collaboration between researchers.

Journalism & Media: Journalists use OSINT to dig into sensitive information for investigative reporting, track real events, and verify facts against intentional propaganda. OSINT helps them cross-check claims, uncover hidden details, and provide more accurate and reliable data [32]. For example, journalists use X (Twitter) and Reddit to monitor live updates, while Google News helps track the latest

trends. Maltego analyzes relationships between individuals or organizations, and Shodan shows details about internet-connected devices that could be relevant for investigative reporting.

Public Opinion & Sentiment Analysis: OSINT tools analyze social media, blogs, records, and forums to assess public opinion and sentiment [33]. Tools like NLTK and SpaCy classify text into positive/negative/neutral sentiments, while TensorFlow or PyTorch models detect nuances like sarcasm or cultural context [34, 35]. During the Brexit referendum, researchers used OSINT to track public opinion by analyzing social media posts and trends. In the 2016 US presidential election and India's 2019 Lok Sabha elections, researchers used OSINT to uncover foreign interference and manipulation on social media. Cambridge Analytica also used OSINT to gather data from social media to analyze voter sentiment and target political advertisements.

OSINT is also extensively utilized in various fields, including geopolitical analysis, environmental monitoring, financial fraud detection, legal investigations, reputation management, disaster recovery, public health surveillance, intellectual property protection, crowdsourced research, and identity verification. Table 18 demonstrates how OSINT has been used in real-world investigations to track threats and solve complex problems. The following table highlights a few significant OSINT case studies.

Table 18.
Open-Source Intelligence Case Studies.

Authors	Case Study	Year	Description	OSINT Outcome
[36]	Boston Marathon Bombing	2013	The Boston Marathon bombings occurred on April 15, 2013, when two homemade bombs exploded near the marathon finishing line, killing three people and injuring more than 260. The attackers, Dzhokhar and Tamerlan Tsarnaev, were brothers who planted the bombs. Tamerlan was later killed in a shootout with police, and Dzhokhar was captured and later sentenced to death.	Investigators used OSINT to track down suspects. They collected public video footage, images, social media posts, and news reports through OSINT to identify the bombers and capture the Tsarnaev brothers.
[37]	Chemical Attacks in Syria	(2013-2018)	The chemical attacks in Syria occurred between 2013 to 2018 during the Syrian Civil War. The deadliest chemical attack in Syria occurred in 2013 in Eastern Ghouta, where more than 1,400 people were killed using sarin gas. Investigations have found that chemical weapons were used by Bashar al-Assad and ISIL militants. Human Rights Watch has documented at least 85 chemical weapons attacks in Syria between 2013 to 2018 and the vast majority of attacks being carried out by the Assad regime.	International organizations and journalists used OSINT to verify claims of chemical weapon attacks by analyzing open-source videos, photos, social media footage, satellite images, geolocation tools, and witness testimonies.
[38]	MH370 Disappearance	2014	The MH370 Disappearance occurred on March 8, 2014, when Malaysia Airlines Flight 370 (MH370/MAS370) lost contact with air traffic control en route from Kuala Lumpur to Beijing. Although 227 passengers and 12 crew members are presumed dead, the cause of its disappearance has not been determined yet and it is considered one of the greatest mysteries in aviation history.	Investigators analyzed satellite data, flight data, social media, and public reports to trace potential flight paths and locate wreckage. Extensive use of OSINT helped narrow down search areas for the plane's wreckage.
	Russian Interference in the U.S. Elections		In 2016, Russia used cyberattacks, disinformation campaigns, and social media	Investigators used OSINT to analyze social media activity, fake

[39]		2016	manipulation to influence the U.S. presidential election. According to the U.S. intelligence community, Russian President Vladimir Putin directed the operation known as "Project Lakhta" to sabotage Hillary Clinton's presidential campaign and help Donald Trump's campaign. In 2019, The 448-page Mueller Report found over 200 contacts between the Trump campaign and Russian officials but there was insufficient evidence to charge Trump or his associates.	news networks, leaked documents, metadata, and online ads to track disinformation campaigns and identify patterns of manipulation.
[40]	Rohingya Genocide	(2016-2017)	The Rohingya genocide is the ongoing systematic persecution and killing of the Muslim Rohingya minority by the Myanmar military. The genocide has forced more than a million Rohingya to flee Bangladesh to escape the violence. Many countries have described the actions as ethnic cleansing.	OSINT analyzes social media, satellite images, news reports, and witness testimony to assist document and verify human rights violations in Myanmar. OSINT was also used to track mass slaughter and displacement of Rohingya.
[41]	Jamal Khashoggi's Murder Investigation	2018	Jamal Khashoggi was a Saudi journalist, Washington Post columnist, and outspoken critic of the Saudi government. He was killed by Saudi government agents inside the Saudi consulate in Istanbul. Saudi authorities initially denied his disappearance. However, Turkish intelligence and the CIA have assessed with high confidence that Crown Prince Mohammed bin Salman (MBS) ordered the assassination.	Turkish investigators used OSINT tools including CCTV footage, geolocation data, open-flight databases, social media/biometric data, satellite imagery, and leaked audio—to trace the movements of 15 Saudi hit squad members.
[42]	Iranian Nuclear Scientist Mohsen Fakhrizadeh's Assassination	2020	Iranian nuclear physicist and scientist Mohsen Fakhrizadeh was assassinated in a deliberate attack near Tehran on November 27, 2020. The assassination was carried out with a sophisticated, remotely operated machine gun mounted on a Nissan pickup truck, with artificial intelligence and satellite technology used to target Fakhrizadeh with minimal collateral damage.	Iranian investigators used geolocation tools, satellite images, social media posts, and open-source videos/photos to track the attackers' movements and the exact location of the ambush.
[43]	Russian invasion of Ukraine	2022	The invasion of Ukraine began on February 24, 2022, when Russia launched a full-scale military attack on Ukraine. This invasion is the largest and deadliest conflict in Europe since World War II. As of 2025, the Russian military controls about 20% of Ukraine's territory. Around 8 million Ukrainians have been internally displaced, while more than 8.2 million have fled the country.	OSINT examines satellite imagery, social media posts, military activity logs, and open-source videos to provide real-time insights into troop movements, damaged infrastructure, and reports of attacks.

5. Challenges in OSINT

Open-source intelligence generates actionable insights from the vast and unstructured nature of publicly available data. Despite its enormous opportunities, OSINT has significant challenges that limit its effectiveness. When OSINT practitioners gather data from search engines, social media, forums, newspapers, or other platforms, they often encounter critical challenges such as fragmented data,

anonymous sources, or deliberately misleading descriptions. To mitigate these challenges, practitioners must focus on legal and privacy regulations, data bias issues, AI tools/techniques, and deepfake technologies. These factors significantly impact the accuracy, reliability, and ethical use of open-source intelligence. The following section highlights the primary challenges of OSINT in detail.

Data Overload: OSINT collects and analyzes vast amounts of data from multiple sources, including social media, news outlets, forums, and government records. The primary challenge of OSINT is the sheer volume of data. Such vast amounts of data are beyond the analytical capabilities of humans. Every day, approximately 500 million tweets are posted, over 4.75 billion items are shared on Facebook, more than 500 hours of video are uploaded to YouTube per minute, and Google handles over 99,000 search queries per second [44,45]. In addition, data duplication and redundancy are another major problem for the data overload issue. OSINT analysts often receive the same news articles, videos, or reports from multiple platforms, which complicates the data filtering process. Therefore, analysts risk missing important information or reaching irrelevant or incorrect conclusions. To reduce data overload, NLP, machine learning, advanced data filtering, and different sentiment analysis techniques are often used. However, even with all these modern technologies, analysts still face the complex challenge of finding the most relevant data.

Data Accuracy and Credibility: OSINT relies on publicly available data, and not all public data is reliable or accurate, so it can be a double-edged sword for OSINT analysts. Misinformation, disinformation, and propaganda can distort the intelligence-gathering process. A high-profile example of this challenge was during the 2016 US presidential election when the Russian disinformation campaign on social media significantly influenced public opinion. Inaccurate data undermines OSINT's reliability and can easily mislead analysts and decision-makers [46]. So, to mitigate these challenges, analysts must utilize cross-referencing, fact-checking, and specialized tools or platforms to verify the data's authenticity.

Ethical and Legal Concerns: OSINT practices often conflict with individual privacy rights, legal frameworks, and ethical boundaries [47]. The primary reason for this is that the line between what is publicly accessible and what violates privacy is not always clear. For example, law enforcement agencies use social media data to monitor political protests. While such practices can improve public safety, but they also infringe personal privacy and civil liberties. As a result, mass monitoring has fueled public anger against state surveillance practices. Similarly, the Cambridge Analytica scandal demonstrates how public data was exploited to develop psychological profiles without proper consent. In 2021, Amnesty International revealed that Pegasus spyware targeted journalists and activists by exploiting publicly available phone numbers.

Restricted and Subscription Data: In OSINT investigations, restricted and subscription data pose a significant challenge. Although OSINT relies on publicly accessible data, but it faces substantial challenges when essential information is restricted for legal, security, or proprietary reasons. Restricted information such as classified government documents, corporate secrets, or sensitive personal information, create unique obstacles for OSINT analysts. Additionally, many valuable resources are locked behind paywalls, accessible only to those with subscriptions or institutional access. For example, platforms like LexisNexis, Westall, or Bloomberg Terminal often require subscription fees to access academic studies, court records, or industry reports, which makes it difficult for independent journalists and researchers. Furthermore, some regions, especially under authoritarian regimes actively censor and hide data to control the narrative [48]. These intentional restrictions limit the ability of OSINT practitioners to obtain unbiased, accurate, and credible information [49,50].

Linguistic and Cultural Barriers: Another critical challenge in OSINT is the linguistic barrier. Many OSINT tools are designed to focus solely on English-language content, which can severely limit accurate intelligence collection and analysis when working with non-English resources [51]. Every country has its own language, which can be translated through language translation tools. However, analysts often struggle with understanding local idioms, cultural context, local slang, and regional dialects. Furthermore, some countries heavily censor social platforms, but users use different coded languages or internet slang to bypass censorship. As a result, without native-level expertise analysts may misinterpret, which may lead to flawed conclusions. Similarly, during Russia's invasion of Ukraine in 2022, OSINT analysts struggled to verify allegations from both sides due to linguistic complexities among Ukrainian, Russian, and regional dialects such as Surzhyk.

Moreover, OSINT confronts additional challenges, including technical expertise, outdated tools, ephemeral data, data fragmentation, geopolitical restrictions, AI and algorithm biases, resource inequality, and deliberate disinformation.

6. Future Directions and Recommendations

The future of OSINT relies on the integration of advanced technologies and the development of ethical, reliable, and scalable frameworks. Advanced AI models such as GPT-4 and BERT help in OSINT by extracting, analyzing, translating, and summarizing massive open-source data. Machine learning technologies like Google and Yandex Reverse Image Search and InVID enhance OSINT by verifying geolocation, timestamps, and content origins. Sensity AI, Sentinel, Hyperverge, and Deepware, these sophisticated tools help OSINT experts collect, analyze, and find patterns in diverse data sources, and detect deepfakes and disinformation in real time. Cutting-edge technologies such as NLP and computer vision help the integration of text, images, videos, and geospatial data to generate more comprehensive intelligence. Therefore, governments, agencies, and corporate experts should invest more resources in acquiring and developing emerging AI and ML tools to improve OSINT operations. In the meantime, focus on creating ethical standards and frameworks that assure responsible data usage, regulatory compliance, and address privacy concerns through clear guidelines. Every institution/company should provide its employees with specialist OSINT training to address cybercrime, disinformation, insider threats, and emerging digital risks. Finally, cross-sector partnerships should be encouraged to share best practices and threat intelligence to address emerging cyber threats.

7. Conclusion

The present world is technology-dependent and due to the massive technological breakthroughs, the complexity of cybercrime is also rapidly expanding, which presents a formidable challenge in today's digitalized world. To address these challenges, OSINT has emerged as a crucial asset, which has become the cornerstone of modern cybercrime defense. This paper discusses the core concepts, tools, techniques, and practical implementations of Open-Source Intelligence, highlighting its critical role in detecting, investigating, and combating cybercrime. Furthermore, this paper also highlights OSINT's transformative capabilities in identifying malicious activities and enhancing cybersecurity by utilizing a diverse range of platforms and tools, including search engines, generative AI, social media, domain analysis tools, and geospatial intelligence systems. The study shows that OSINT not only strengthens cybercrime investigations but also provides significant opportunities for proactive threat assessment and strategic decision-making in both the public and private sectors. However, despite its strengths, OSINT misuse carries the risk of privacy violations and the spread of misinformation. Additionally, OSINT faces limitations such as data overload, data manipulation, outdated information, source reliability, privacy restrictions, legal ambiguity, and ethical dilemmas. To address these limitations, policymakers and practitioners must prioritize transparency, accountability, and compliance with legal frameworks to maintain trust in OSINT methodologies. At the same time, focus on developing advanced AI-assisted analysis tools and techniques, standardized frameworks, specialized cyber training, and international cooperation for cyber governance. In conclusion, OSINT serves as a powerful blend of art and science against cybercrime—when applied responsibly so OSINT practitioners must adhere strictly to responsible conduct, legal frameworks, privacy regulations, and ethical obligations.

References

[1] Akhgar, B., Bayerl, P. S., & Sampson, F. (2016). Open source intelligence investigation. In Advanced sciences and technologies for security applications. https://doi.org/10.1007/978-3-319-47671-1

[2] Andersson, C., & Sundin, O. (2023). The elusive search engine: How search engine use is reflected in survey reports. Journal of the Association for Information Science and Technology, 75(5), 613–624. https://doi.org/10.1002/asi.24819

[3] Andersen, J. (2018). Archiving, ordering, and searching: Searchengines, algorithms, databases, and deep mediatization. Media,Culture & Society, 40(8), 1135–1150. https://doi.org/10.1177/0163443718754652

[4] Search engine market share worldwide | StatCounter Global Stats. https://gs.statcounter.com/search-engine-market-share

[5] J. Pastor-Galindo, P. Nespoli, F. Gómez Mármol and G. Martínez Pérez, "The Not Yet Exploited Goldmine of OSINT: Opportunities, Open Challenges and Future Trends," in IEEE Access, vol. 8, pp. 10282-10304, 2020, doi: 10.1109/ACCESS.2020.2965257

[6] Toffalini, F., Abbà, M., Carra, D., & Balzarotti, D. (2016). Google Dorks: Analysis, Creation, and new Defenses. In Lecture notes in computer science (pp. 255–275). https://doi.org/10.1007/978-3-319-40667-1_13

[7] Nah, F. F., Zheng, R., Cai, J., Siau, K., & Chen, L. (2023). Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. Journal of Information Technology Case and Application Research, 25(3), 277–304. https://doi.org/10.1080/15228053.2023.2233814

[8] Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2023). Generative AI. Business & Information Systems Engineering, 66(1), 111–126. https://doi.org/10.1007/s12599-023-00834-7

[9] Cascavilla, G., Beato, F., Burattin, A., Conti, M., & Mancini, L. V. (2018). OSSINT - Open Source Social Network Intelligence. Online Social Networks and Media, 6, 58–68. https://doi.org/10.1016/j.osnem.2018.04.003

[10] Zimba, O., Gasparyan, A. Y. (2021). Social media platforms: a primer for researchers. Rheumatology, 59(2), 68-72. https://doi.org/10.5114/reum.2021.102707

[11] Elguindy, M. (2021). Applying digital forensics methodology to open source investigations in counterterrorism. Journal of Law and Emerging Technologies, 1(1), 11–64. https://doi.org/10.54873/jolets.v1i1.32

[12] B. E. Uçar, M. İ. Ecevit, H. Dağ and R. Creutzburg, "A Comprehensive Review of Open Source Intelligence in Intelligent Transportation Systems," 2024 International Conference on Intelligent Environments (IE), Ljubljana, Slovenia, 2024, pp. 109-116, doi: 10.1109/IE61493.2024.10599907

[13] J. Harvey and S. Kumar, "A Survey of Intelligent Transportation Systems Security: Challenges and Solutions," 2020 IEEE 6th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS), Baltimore, MD, USA, 2020, pp. 263-268, doi: 10.1109/BigDataSecurity-HPSC-IDS49724.2020.00055

[14] Bill, R., Blankenbach, J., Breunig, M. et al. Geospatial Information Research: State of the Art, Case Studies and Future Perspectives. PFG 90, 349–389 (2022). https://doi.org/10.1007/s41064-022-00217-9

[15] Loukili, Y., Lakhrissi, Y. & Ali, S.E.B. Geospatial Big Data Platforms: A Comprehensive Review. KN J. Cartogr. Geogr. Inf. 72, 293–308 (2022). https://doi.org/10.1007/s42489-022-00121-7

[16] Szymoniak, S., Foks, K. (2024). Open Source Intelligence Opportunities and Challenges – A Review. Advances in Science and Technology Research Journal, 18(3), 123-139. https://doi.org/10.12913/22998624/186036

[17] F. Tabatabaei and D. Wells, ‘‘Osint in the context of cyber-security,’’ in Open Source Intelligence Investigation: From Strategy to Implementation, B. Akhgar, P. S. Bayerl, and F. Sampson, Eds. Cham, Switzerland: Springer, 2016, pp. 213–23

[18] Kathrine, G. J. W., Baby, R. T., & Ebenzer, V. (2020). Comparative analysis of subdomain enumeration tools and static code analysis. Journal of Mechanics of Continua and Mathematical Sciences, 15(6). https://doi.org/10.26782/jmcms.2020.06.00013

[19] Alam, S. S., Yeow, P. H. P., & Loo, H. S. (2011). An empirical study on online social networks sites usage: Online dating sites Perspective. International Journal of Business and Management, 6(10). https://doi.org/10.5539/ijbm.v6n10p155

[20] D.-Y. Kao, Y.-T. Chao, F. Tsai, and C.-Y. Huang, ‘‘Digital evidence analytics applied in cybercrime investigations,’’ in Proc. IEEE Conf. Appl., Inf. Netw. Secur. (AINS), Nov. 2018, pp. 117–122

[21] Maschmeyer, Lennart; Deibert, Ronald J.; Lindsay, Jon R. (2021). "A tale of two cybers - how threat reporting by cybersecurity firms systematically underrepresents threats to civil society". Journal of Information Technology & Politics. 18 (1): 1–20.

[22] Bruce M, Lusthaus J, Kashyap R, Phair N, Varese F (2024) Mapping the global geography of cybercrime with the World Cybercrime Index. PLoS ONE 19(4): e0297312. https://doi.org/10.1371/journal.pone.0297312

[23] Kalra, Y., Upadhyay, S., & Patheja, P. S. (2020). Advancements in cyber attacks and security. International Journal of Innovative Technology and Exploring Engineering, 9(4), 1520–1528. https://doi.org/10.35940/ijitee.d1678.029420

[24] Hwang, Y., Lee, I., Kim, H., Lee, H., & Kim, D. (2022). Current status and security trend of OSINT. Wireless Communications and Mobile Computing, 2022, 1–14. https://doi.org/10.1155/2022/1290129
[25] Oppenheimer, H. (2024). How the process of discovering cyberattacks biases our understanding of cybersecurity. Journal of Peace Research, 61(1), 28-43. https://doi.org/10.1177/00223433231217687
[26] Ghioni, R., Taddeo, M., & Floridi, L. (2023). Open source intelligence and AI: a systematic review of the GELSI literature. AI & Society, 39(4), 1827–1842. https://doi.org/10.1007/s00146-023-01628-x
[27] Cherqi, O., Moukafih, Y., Ghogho, M., & Benbrahim, H. (2023). Enhancing cyber threat identification in Open-Source intelligence feeds through an improved Semi-Supervised Generative Adversarial Learning approach with contrastive learning. IEEE Access, 11, 84440–84452. https://doi.org/10.1109/access.2023.3299604
[28] Ziółkowska, A. (2018). Open source intelligence (OSINT) as an element of military recon. Security and Defence Quarterly, 19(2), 65-77. https://doi.org/10.5604/01.3001.0012.1474
[29] Hughbank, R. J., & Githens, D. (2010). Intelligence and its role in protecting against terrorism. Journal of Strategic Security, 3(1). https://doi.org/10.5038/1944-0472.3.1.3
[30] Fleisher, C.S. (2008), "Using open source data in developing competitive and marketing intelligence", European Journal of Marketing, Vol. 42 No. 7/8, pp. 852-866. https://doi.org/10.1108/03090560810877196
[31] Van Puyvelde, D., & Tabárez Rienzi, F. (2025). The rise of open-source intelligence. European Journal of International Security, 1–15. doi:10.1017/eis.2024.61
[32] Belghith, Y., Venkatagiri, S., & Luther, K. (2022). Compete, collaborate, investigate: Exploring the social structures of open source intelligence investigations. CHI Conference on Human Factors in Computing Systems, 1–18. https://doi.org/10.1145/3491102.3517526
[33] Haselmayer, M., & Jenny, M. (2016). Sentiment analysis of political communication: combining a dictionary approach with crowdcoding. Quality & Quantity, 51(6), 2623–2646. https://doi.org/10.1007/s11135-016-0412-4
[34] Rozo, A. Z., Díaz-López, D., Pastor-Galindo, J., Mármol, F. G., & Karabiyik, U. (2024). An NLP-Based framework to spot extremist networks in social media. Complexity, 2024(1). https://doi.org/10.1155/2024/3380488
[35] Sánchez, J. R., Campo-Archbold, A., Rozo, A. Z., Díaz-López, D., Pastor-Galindo, J., Mármol, F. G., & Díaz, J. A. (2021). Uncovering Cybercrimes in Social Media through Natural Language Processing. Complexity, 2021(1). https://doi.org/10.1155/2021/7955637
[36] Nhan, J., Huey, L., & Broll, R. (2015). Digilantism: an analysis of crowdsourcing and the Boston Marathon bombings. The British Journal of Criminology, azv118. https://doi.org/10.1093/bjc/azv118
[37] Brooks, J., Erickson, T.B., Kayden, S. et al. Responding to chemical weapons violations in Syria: legal, health, and humanitarian recommendations. Confl Health 12, 12 (2018). https://doi.org/10.1186/s13031-018-0143-3
[38] Ashton, C., Shuster Bruce, A., Colledge, G., & Dickinson, M. (2015). The Search for MH370. Journal of Navigation, 68(1), 1–22. doi:10.1017/S037346331400068X
[39] Eady, G., Paskhalis, T., Zilinsky, J. et al. Exposure to the Russian Internet Research Agency foreign influence campaign on Twitter in the 2016 US election and its relationship to attitudes and voting behavior. Nat Commun 14, 62 (2023). https://doi.org/10.1038/s41467-022-35576-9
[40] O'Brien, M., & Hoffstaedter, G. (2020). "There We Are Nothing, Here We Are Nothing!"—The Enduring Effects of the Rohingya Genocide. Social Sciences, 9(11), 209. https://doi.org/10.3390/socsci9110209
[41] Milanovic, M. (2020). The murder of Jamal Khashoggi: Immunities, inviolability and the human right to life. Human Rights Law Review, 20(1), 1–49. https://doi.org/10.1093/hrlr/ngaa007
[42] Khoshnood, A. (2021). The assassination of Fakhrizadeh—A major Iranian counterintelligence failure? Security and Intelligence, 6(1). https://doi.org/10.18278/gsis.6.1.9
[43] Dodds, K., Taylor, Z., Akbari, A., Broto, V. C., Detterbeck, K., Inverardi-Ferri, C., … Woon, C. Y. (2023). The Russian invasion of Ukraine: implications for politics, territory and governance. Territory, Politics, Governance, 11(8), 1519–1536. https://doi.org/10.1080/21622671.2023.2256119
[44] Antonakaki, D., Fragopoulou, P., & Ioannidis, S. (2020b). A survey of Twitter research: Data model, graph structure, sentiment analysis and attacks. Expert Systems With Applications, 164, 114006. https://doi.org/10.1016/j.eswa.2020.114006
[45] Springer, S., Strzelecki, A., & Zieger, M. (2023). Maximum generable interest: A universal standard for Google Trends search queries. Healthcare Analytics, 3, 100158. https://doi.org/10.1016/j.health.2023.100158
[46] Klouček, T., Lagner, O., & Šímová, P. (2015). How does data accuracy influence the reliability of digital viewshed models? A case study with wind turbines. Applied Geography, 64, 46–54. https://doi.org/10.1016/j.apgeog.2015.09.005
[47] Riebe, T., Biselli, T., Kaufhold, M., & Reuter, C. (2023). Privacy Concerns and acceptance Factors of OSINT for Cybersecurity: A Representative survey. Proceedings on Privacy Enhancing Technologies, 2023(1), 477–493. https://doi.org/10.56553/popets-2023-0028
[48] De Werd, P. (2021). Reflexive intelligence and converging knowledge regimes. Intelligence & National Security, 36(4), 512–526. https://doi.org/10.1080/02684527.2021.1893073
[49] Van Der Woude, M., Dodds, T., & Torres, G. (2024). The ethics of open source investigations: Navigating privacy challenges in a gray zone information landscape. Journalism. https://doi.org/10.1177/14648849241274104
[50] Koenig, A. (2024). Ethical Considerations for Open-Source Investigations into International Crimes. AJIL

Unbound, 118, 45–50. doi:10.1017/aju.2024.2
[51] Neri, F., Geraci, P., & Pettoni, M. (2010). Stalker: overcoming linguistic barriers in open source intelligence. International Journal of Networking and Virtual Organisations, 8(1/2), 37. https://doi.org/10.1504/ijnvo.2011.037160
[52] Browne, T.O., Abedin, M. & Chowdhury, M.J.M. A systematic review on research utilising artificial intelligence for open source intelligence (OSINT) applications. Int. J. Inf. Secur. 23, 2911–2938 (2024). https://doi.org/10.1007/s10207-024-00868-2
[53] Henrico, S., & Putter, D. (2025). Intelligence collection disciplines—A systematic review. Journal of Applied Security Research, 20(1), 46-70.
[54] Potenziani, E. (2006). Current and future trends in military electronic warfare systems and the role of thin films and related materials. Ferroelectrics, 342(1), 151-161.
[55] Heningtiyas, H., & Supriyadi, A. A. (2024). Study of the implementation of geoint and remote sensing in climate change. Remote Sensing Technology in Defense and Environment, 1(2), 45-55.
[56] Akartuna, E. A., Johnson, S. D., & Thornton, A. (2024). Motivating a standardised approach to financial intelligence: a typological scoping review of money laundering methods and trends. Journal of Experimental Criminology, 1-47.
[57] Macêdo, A., Peotta, L., & Gomes, F. (2023). A Review of the Intersection Techniques on Humint and Osint. International Journal on Cybernetics & Informatics (IJCI), 12(1), 53.
[58] Zhang, Z., Yao, Y., Zhang, A., Tang, X., Ma, X., He, Z., ... & Zhao, H. Igniting language intelligence: The hitchhiker's guide from chain-of-thought reasoning to language agents. ACM Computing Surveys, 57(8), 1-39.
[59] Larsen, O. H., Ngo, H. Q., & Le-Khac, N. (2023). A quantitative study of the law enforcement in using open source intelligence techniques through undergraduate practical training. Forensic Science International Digital Investigation, 47, 301622. https://doi.org/10.1016/j.fsidi.2023.301622