CLARITAS

REDEFINING INTELLIGENCE

BLUEFUSION uniquely provides a unified, semi-automated, truly multi-data and cross domain capability with a global coverage, for continuous person-centric and high-volume multilingual thematic collection across the surface, deep and dark web, social media, forums, emerging platforms and data-broker sources with entity ingestion, link analysis, scalable keyword campaigns, real-time alerting, and discovery of emerging trends and Organised Criminal Group (OCG) modus operandi.

We cover the deep, dark net alongside market places, access brokers, chat forums, messaging apps and others, our truly unified OSINT platform is purpose built to to enrich the subjects of interest and discover additional intelligence, as well as automatically ingesting and matching data such as Call Data records against SOI where this data is available.

Social media collection and enrichment

SOCMINT Collection is mission-tailored by topic, geography, language, actor group and platform at a minimum of 520 million new items per month, refreshed every 30 minutes. Multisource searches use selectors including usernames, names, handles, hashtags and free-text keywords, combined with an 11-operator Boolean query language, an AI Query Assistant that builds structured queries from natural language, and fuzzy semantic search, across content and metadata in a single workflow.

Person-specific identifiers — email addresses, phone numbers, dates of birth and addresses — are searched by matching against ingested entity and consented datasets. Native coverage includes Facebook, Instagram, X, TikTok, YouTube and Telegram, plus VK, Bluesky, Threads, Discord, MAX and dedicated Chinese-platform collection. Reddit and Snapchat are delivered through BLUEFUSION's forward-deployed integration, and LinkedIn through an established partner integration; both will be provisioned for this contract.

Where specific communications platforms are being used, including WhatsApp and Kik, ODIN emulation tools allows complete automated collection of the content of the chats the emulators are deployed into, bypassing the heavily encrypted security measures without requiring warrantry.

Newly identified and emerging platforms are continuously discovered, validated and incorporated. Every actor is enriched at collection with state affiliation, high-risk flags (including EU/US sanctions-list presence and hacker/hacktivist links), a coordinated-behaviour score and geographic origin, and cross-platform entity linking connects the same actor across platforms, so datasets demonstrate relevance to defined selectors and themes and surface linked content and actors.

Multilingual collection and processing

Multilingual search runs simultaneously across person-centric and thematic investigations: a query in one language is automatically translated and matched across 50+ languages. Collected text is machine-translated for scalable cross-language analysis, and individual items translate in place for review.

Multilingual processing — search, sentiment and topic modelling — is applied consistently across every platform, source and content type as new ones are added, producing translated, searchable outputs that retain contextual meaning sufficient for analytical interpretation and decision-making.

Thematic and threat vocabulary management

Analysts create, manage and update multilingual threat-specific keyword and selector lists and organise accounts into custom Actor Groups used as inclusion or exclusion filters. Thematic vocabularies drive automated collection and the identification of emerging themes, trends and behaviours, with fuzzy and semantic search expanding them automatically across synonyms, slang and indirect references. Vocabularies are applied and maintained for continuous monitoring and iterative development, and emerging keywords, phrases and patterns surfaced from collected data feed back into refinement evidenced by improved identification of relevant content and emerging terms.

Automated monitoring and alerting

Monitoring is configured using selectors including usernames, hashtags, keywords, groups, pages, channels and defined sources, refined by time range and source. Automated threat detection module surfaces coordinated attacks, manipulation and misinformation clusters, and significant political or diplomatic information events as they emerge, tagged by type and topic with publication volume, at near real-time frequency where the source platform supports it. Alerting is configurable on thresholds in topic volume, sentiment shift or narrative activity, with alerts prioritised and tagged to minimise false positives and support timely response.

Device Entity AI also provides the ability to understand if subjects are changing selectors, such as the IMSI/ IMEI of a handset being changed for example, multiple MSISDNs associated with a single IMEI, or co-travellers who are potential associates and how they travel, on land sea or air.

User data ingestion and analytical integration

The platform ingests user-supplied and externally held data via custom connectors, each with API integration and schema mapping. Structured files and imagery are ingested and analysed alongside platform outputs, with OCR extracting text from images and video frames. Ingested data is applied to entity-based investigation, link analysis and contextualisation, and combined analysis of supplied and collected data identifies new actors, platforms, networks and operating methods — producing integrated outputs that demonstrate how supplied data enhances understanding of networks, behaviours and investigative leads. The same mechanism ingests data-broker and other third-party datasets for use within investigations.

The API's BLUEFUSION already ingest include automated web-scraping and cyber-crawling capability to enable continuous person-centric and high-volume multilingual thematic collection across the surface, deep, and dark web (including social media, forums, emerging platforms, and data broker sources). We supports entity data ingestion for targeted investigations and link analysis, scalable automated keyword campaigns, real-time alerting on emerging risks and threats, and discovery of emerging trends and Organised Criminal Networks (OCN) modus operandi.

The system is continually updated at a rate of around 1 billion records (~4 billion digital artefacts) per week in crawling and scraping alone. Our data when compared to other 'household' providers by a UK Government Agency was proved to return the same data and ~25% more unique data, the comparison has resulted in a multi-year Government contract and our clients consistently tell us it is richer and deeper than any other data in the market.

The data is used to uncover information and build out the intelligence picture around people, their connections, their online behaviour. BLUEFUSION allows users to expand on potential intelligence optics starting with a seed identifier such as an email address to find, phone numbers, alternative email addresses, biographical information, PII, geographical information, online activity and more. BLUEFUSION is often used

as a standalone investigative tool, however our clients often say that our system regularly highlights new information, that gives them the investigative edge to use in other tools and move the investigation along.

From seized devices on the shoreline or at ports of entry, the BLUENEXUS field device user originated data is AI based, offline and portable, with the optional NEXUS lake offers complete automation of the ingestion of device digital forensic extractions, categorisation, analysis, translation, OCR, audio, video and imagery matching with minimal user input, also including confidence % on matches found. The impact on reducing both device backlog, eliminating

Location inference from seized devices is also smart - it's not just about a recorded geotag. Include locations that have been mentioned in chats, analysed, searched for, shared, visited and pattern of life as well as being filtered by date/time/app/pattern of life - all with minimal user input, ensuring the device data that may initially appear as unrelated if they are from completely different seizure points, automatically makes any connections that may not be visible to a human investigator in granular detail.

There is also a data import tool for users to ingest their own data directly, in set formats, directly via the browser access portal - this is currently in a few key categories, however can be tailored to any type of client requirement, within reason, to be tailored to suit their exact needs, rather than having to put up with a pre-prepared tool to ensure that cleaned and normalised data the end user generates can also be ingested and fused for integration and enrichment as an entirely unique capability when API is not appropriate.

Consented subject and selector searching

Consented and entity datasets — provided by the end user or through an approved data source — are ingested via custom connectors and matched against collected data, enabling subject and selector-based searches on identifiers including name, address, email address and phone number. Consented data is retrieved and analysed strictly in accordance with applicable legal, regulatory and governance requirements, and results demonstrate the relevance and lawful use of consented data in support of subject-based investigations.

Surface, deep and dark web, and OCG modus operandi

Surface-web social, news and forum collection is complemented by deep and dark-web collection which draws on hacker communities, criminal forums and dark-web content, and by ingested breach and data-broker datasets. Topic Discovery, semantic Stories grouping and coordinated-behaviour detection surface emerging trends and OCG modus operandi, linking observed methods back to the actors and networks behind them.

Performance monitoring

The system maintains complete source provenance, analyst oversight, and evidential integrity. Intelligence collected from proprietary intelligence, primary-source collections, publicly available information, customer-provided data, and AI-enhanced analysis remains fully traceable to its originating source, enabling investigators to validate findings, assess confidence, and produce intelligence products suitable for operational, tactical, and strategic decision-making.

By bringing together collaborative investigative workflows within a single SaaS platform, we provide the end user with a unified intelligence capability that enables investigators to discover more intelligence, understand it more quickly, and transform it into evidence-based operational outcomes that support the disruption of Organised Immigration Crime and associated threats, wherever it may originate.