Fare intelligence nel cyber spazio: Big Data e Virtual Humint, dal deep state all'uso personale

Author : Francesco Arruzzoli

Date : 18 Novembre 2019

La parola *intelligence* ha sempre evocato uno strumento utilizzato esclusivamente in ambito governativo e militare, un *"deep state"* che - indipendentemente dalle maggioranze politiche - opera per poter fornire ai governanti tutte le informazioni utili per far prendere loro le decisioni e supportarli quando non sono in grado di farlo. Prevenire i pericoli per la propria nazione e garantire la sicurezza dei propri cittadini sono gli obiettivi primari che l'*intelligence* di uno stato persegue.

I rapidi progressi tecnologici e l'evoluzione socio-economica delle nazioni hanno permesso all'*intelligence* di disporre di nuovi strumenti in grado di raccogliere una quantità di dati prima inimmaginabile, tali e tanti da non poter essere analizzati e correlati tutti contemporaneamente e correttamente.
La tecnologia ha sovralimentato lo spionaggio attraverso discipline come il *Signals Intelligence* (SIGINT), l'*Electronic Intelligence* (ELINT), il *Communications Intelligence* (COMINT) che si occupano di raccogliere informazioni mediante l'intercettazione e l'analisi dei segnali e delle comunicazioni (Radio, reti dati, telefonia, etc.), l'*imagery intelligence* (IMINT),che raccoglie informazioni mediante l'analisi delle immagini (ad es. fotografie aere, satellitari, etc..) e tante altre discipline con altrettante sigle, non utili ai fini di questo articolo.

Anche le **operazioni di HUMINT** (Human Intelligence) basate sull'eterna e inevitabile interazione interpersonale, fondamento storico dell'intelligence che rimarranno sempre un elemento di primaria importanza nella raccolta di informazioni, sono state potenziate dal progresso tecnologico che le ha integrate nel cyber spazio, creando il VIRTUAL HUMINT, cioè l'unione delle capacità "umane" dello HUMINT con quelle "virtuali" del SOCMINT (Social Media Intelligence). La VIRTUAL HUMINT rappresenta un approccio metodologico estremamente interessante perché coniuga mondi apparentemente opposti e riporta al centro dell'analisi delle informazioni la figura umana e non quella tecnologica, completando e integrando le lacune dei grandi sistemi automatizzati di raccolta ed elaborazione. È una disciplina specializzata nella creazione ed utilizzo di identità virtuali in rete e in particolare nei social network.

Il risultato di tutto questo “potenziamento” tecnologico è stata la nascita di **due scuole di pensiero:** quella dei big data e quella dell’Analyst Human Centric.

La prima è un’**analisi quantitativa** delle informazioni, si basa sull’utilizzo di software di analisi evoluti, estremamente potenti ma standardizzati, che non tengono conto di quegli elementi di dettaglio che possono indirizzare a nuove piste investigative; questi software hanno un elevato TCO (total cost of ownership) in termini di costo di acquisto, gestione, tempo di formazione del personale e limiti nella personalizzazione delle interfacce di analisi.

La seconda è un’**analisi qualitativa** di porzioni di informazioni, pone al centro del processo l’analista e non i software, secondari rispetto all’implementazione di strategie di ricerca ed analisi basate sulle capacità personali e professionali dell’analista. L’approccio qualitativo richiede analisti dotati di una elevata conoscenza informatica, di una eccellente cultura generale, di ottime conoscenze umanistiche e soprattutto di una spiccata predisposizione investigativa, prerogative indispensabili nella disciplina del Virtual Humint.

Un classico esempio di utilizzo delle due scuole di pensiero è la loro applicazione nella **lotta al terrorismo**. Le analisi dei Big Data permettono di fornire una visione complessiva e statistica degli eventi, ad es. fare una comparazione delle tipologie di attacco, una statistica dei paesi maggiormente colpiti, degli eventi precedenti e successivi correlati ad un attacco, identificare e filtrare la disinformazione, valorizzando le informazioni utili, etc.

Le analisi incentrate sul Virtual Humint operano in maniera verticale, in profondità su una specifica porzione di informazioni presenti nel cyber spazio e tentano di individuare non solo le correlazioni esistenti ma ne cercano di nuove. Sono ormai diventati famosi i lavori svolti dal sito di giornalismo britannico Bellingcat che attraverso i suoi giornalisti/analisti, utilizzando strumenti gratuiti e *open source,* ha individuato e profilato l’infiltrazione di gruppi terroristici dell’ISIS provenienti dal confine turco, membri e attività di un reparto speciale dell’ISIS denominato Katibat al-Battar, l'identità dei due agenti russi accusati di avere avvelenato Sergei Skripal e geolocalizzato i sicari dei narcos nello stato messicano di Chihuahua.

Nel caso specifico del terrorismo inoltre è importante notare che l’ISIS, pur continuando a perdere terreno sui campi di battaglia reali, sta sempre più continuando guadagnare terreno virtuale sui social, arruolando continuamente nuovi adepti in tutto il mondo con il loro Cyber-califfato e grazie alla propaganda CyberJihad delle loro parole ed idee. Ambito questo dove la disciplina del Virtual Humint può fare la differenza perché permette di costruire relazioni, contatti e acquisire nuove informazioni nel cyber spazio esattamente come un’agenzia di intelligence tradizionale fa nel mondo reale.

L’attività di intelligence quindi si trova oggi a raccogliere, catalogare, analizzare e filtrare una mole di informazioni enorme che ogni giorno aumenta in quantità e complessità.

Questo ha anche creato il **paradosso** che tanti dati non fanno necessariamente un’informazione: una eccessiva quantità di dati che non si è in grado di gestire, analizzare e filtrare rischia di trasformarsi in catastrofici cortocircuiti informativi, dalle gravissime conseguenze, come quelli accaduti l’11 settembre 2001 nel tremendo attentato alle due torri di

New York o il 7 gennaio 2015 nel tragico attentato in Francia al Charlie Hebdo. Entrambe queste date rappresentano oltre ad immani tragedie il fallimento dell'intelligence, che non è stata capace di filtrare le false notizie da quelle attendibili sottovalutandole.

Ma come possono discipline di intelligence come l'analisi dei Big Data o il Virtual Humint essere utilizzate nel nostro quotidiano?

Tutti noi facciamo nel nostro piccolo, ogni giorno, attività di intelligence, ad es. quando chiediamo ad amici e conoscenti consigli di vario genere su ristoranti, alberghi, visite mediche, oppure utilizziamo i motori di ricerca o le app sul nostro smartphone per cercare risposte; e sempre più spesso, se non troviamo subito l'informazione che cerchiamo, sarà l'informazione stessa a venire da noi, ci appare sotto forma di banner pubblicitario o mail o messaggi da app varie, già personalizzata e pronta per i nostri bisogni. Se quando vi capita questo pensate che qualcuno vi stia spiando dal microfono del vostro smartphone, vi sbagliate: sono solo gli algoritmi di profilazione dei motori di ricerca e dei social network che funzionano bene.

Questo enorme flusso di dati (testo, audio, video, etc..) che noi stessi mettiamo in rete (inconsapevolmente) durante le nostre ricerche su internet o (consapevolmente) attraverso i nostri profili social fornisce informazioni pregiate a chi ha interesse in noi, ad es. aziende che vogliono vendervi qualcosa, multinazionali finanziarie, farmaceutiche, enti governativi, semplici curiosi o peggio criminali. Per questo motivo possiamo applicare alcune tecniche di intelligence alla nostra sfera privata e pubblica per verificare e meglio tutelare la riservatezza delle nostre informazioni.

Il valore di queste informazioni è inestimabile, perché i costi necessari per acquisire dati così dettagliati e spontaneamente inviati da un così enorme numero di fonti (si pensi solo a **Facebook,** che oggi conta oltre 2 miliardi di iscritti) è complessivamente irrisorio; e poco importa se le informazioni possono essere "taroccate" perché a qualcuno piace apparire quello che non è, magari mettendo foto non sue, "photoshoppate" o post di viaggi mai fatti, etc.; non è un problema, gli algoritmi di profilazione riescono, a trovare tra i tanti post quelli "reali" e a classificarli, anche grazie a nuovi algoritmi di intelligenza artificiale, specializzati nel profilare gli aspetti più intimi delle nostre personalità (come ad esempio il narcisismo che ci spinge a postare foto in ogni luogo e momento per renderci più interessanti agli occhi dei nostri amici e degli sconosciuti).

L'importanza di questo input spontaneo, massivo e il conseguente utilizzo "tracciante" di strumenti che lo agevolano come i social network non sono certo sfuggiti all'attenzione delle organizzazioni di intelligence.

E **non è complottismo** (come ho spesso ribadito fin dal 2009, quando il fenomeno FB esplose in Italia) dire che proprio l'intelligence americana ha investito per prima su Facebook quando ne ha capito le potenzialità. Prima di Facebook c'era stato un solo un importante tentativo da parte dell'FBI di sfruttare la rete per acquisire informazioni spontanee; era il progetto INFRAGARD (https://www.infragard.org/) ancora oggi attivo, nato per permettere a chiunque di poter fare segnalazioni su su persone, attività sospette o pericolose per la sicurezza nazionale. Ma INFRAGARD non è mai decollato perché era soggetto a una continua attività di

disinformazione e controspionaggio da parte di diverse intelligence straniere e quindi lo sforzo per vagliare la veridicità delle informazioni e l'affidabilità delle fonti era un lavoro estremamente oneroso.

Poi la società privata In-Q-Tel pagata dalla CIA per cercare nuove tecnologie e soluzioni da applicare nel campo dell'Intelligence scopre l'idea di FB nel 2004 e la finanzia più volte nella sua fase di start-up fino a quando non si quota in borsa nel 2012.

Facebook ha realizzato un modo nuovo di acquisire informazioni per l'intelligence, per tre motivi fondamentali:

1. permette a chiunque nel mondo di connettersi e inserire i suoi dati personali, foto, video, spontaneamente creando lo schema delle sue relazioni con altri utenti e organizzazioni;
2. ottiene spontaneamente dagli utenti anche informazioni sensibili come quelle politiche, religiose, sessuali e comportamentali analizzando i dati e profilando i post che l'utente scrive;
3. può interagire a sua discrezione e in maniera subliminale (attraverso i suoi algoritmi) con l'utente, pilotandone le scelte con alcune determinate informazioni sotto forma di pubblicità, post, consigli, etc. che rientrano negli interessi identificati dalla profilazione e che quindi l'utente sicuramente guarderà.

Accedere in maniera indiscriminata a queste informazioni non è solo il sogno di ogni Intelligence ma di tantissime altre organizzazioni, come ci insegna il caso di Cambridge Analytica.

Avere la possibilità di analizzare i Big Data che compongono gli aspetti comportamentali dell'universo digitale della popolazione mondiale vuol dire per l'intelligence avere in mano la possibilità di **prevedere e controllare il futuro**.

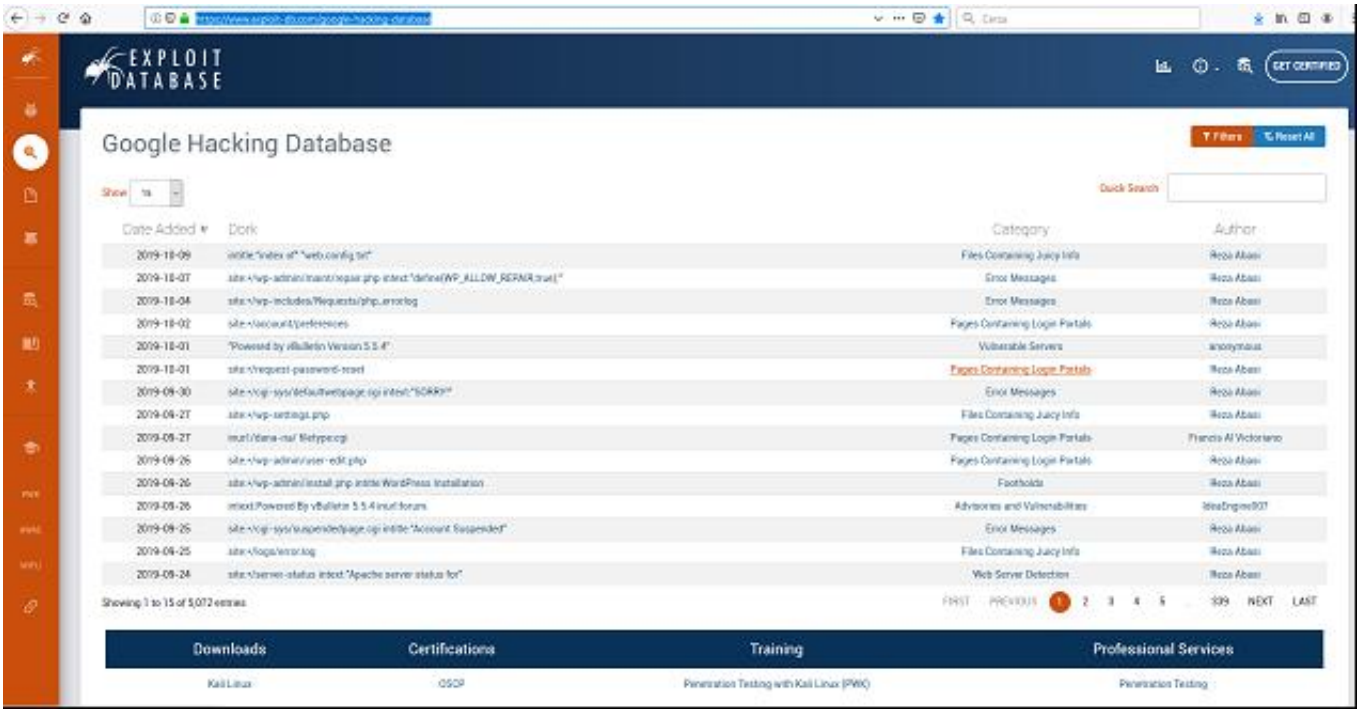
Ma l'aspetto interessante da considerare è che questo progresso tecnologico e sociale ha reso disponibile la stragrande maggioranza di queste informazioni alla portata di chiunque possieda un computer e una connessione a Internet; non solo, ha anche messo a disposizione potenti strumenti per cercarle ed analizzarle facilmente.

Questo ha fatto fare un salto evolutivo epocale all'intelligence, se prima era utilizzata esclusivamente in ambiti ristretti e quasi "esoterici" oggi può essere utilizzata da chiunque, dalle istituzioni, dalle aziende e da noi privati cittadini.

L'Open Source Intelligence (OSINT) rappresenta appunto l'insieme di informazioni di libero accesso presenti in internet (siti web, blog, social network, etc..) e anche di molte altre informazioni riservate che non dovrebbero esserci.

Uno degli strumenti più potenti di intelligence che abbiamo tutti a disposizione oggi e che usiamo quotidianamente è il motore di ricerca **Google**. Google è un motore che lavora sui Big Data e dispone di operatori semantici avanzati per effettuare *query* mirate e dettagliate che normalmente non utilizziamo nelle nostre ricerche quotidiane. Se ben impostati questi operatori possono dare risultati incredibili, tramite specifiche *query* possiamo verificare l'eventuale

presenza di informazioni che dovrebbero essere riservate e che invece sono di dominio pubblico. Un esempio di quanto possano essere versatili e accurate le ricerche attraverso questi operatori ce lo dà il Google Hacking Database (GHDB; https://www.exploit-db.com/google-hacking-database), un sito dove gli utenti postano le query più particolari effettuate su Google tramite operatori con specifiche chiavi di ricerca.



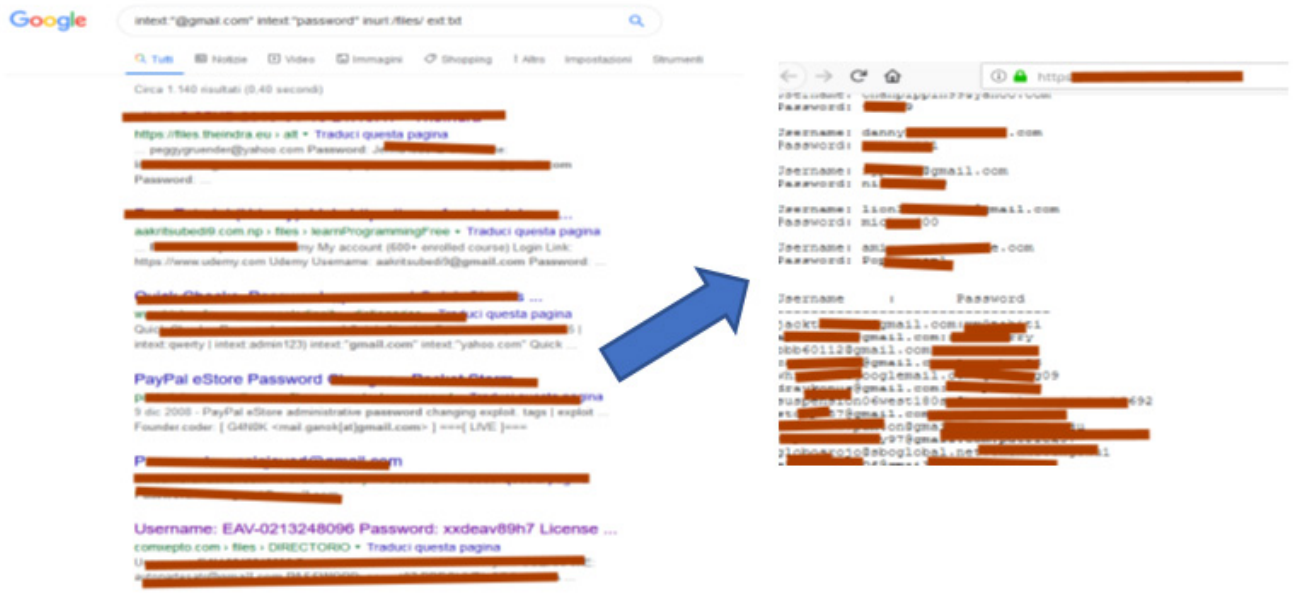
Per fare un test ho utilizzato una delle query presenti nel GHDB:

intext:"@gmail.com" **intext:**"password" **inurl:**/files/ **ext:**txt

questa semplice riga inserita nel motore di ricerca mostra tutti i contenuti indicizzati da Google con le seguenti caratteristiche:

- il parametro **intext:** richiede che all'interno del contenuto indicizzato da Google siano presenti le stringhe: *@gmail.com* e *password*
- il parametro **inurl:** che nell'url analizzato siano presenti i caratteri : */files/*
- il parametro **ext:** seleziona solo i nomi dei file indicizzati con estensione *txt*

Il risultato è effettivamente "imbarazzante", Google ha rilevato molte informazioni su credenziali di diversi utenti che dovevano essere riservate:



Ho personalizzato la medesima query aggiungendo ulteriori parametri per verificare se il mio sito personale poteva essere stato esposto ad indicizzazione di informazioni sensibili da parte di Google. Aggiungendo quindi il parametro **site:***dominio.sito*, dove *dominio.sito* indica il nome del dominio si restringe la ricerca esclusivamente al dominio web indicato.

Ancora oggi, nonostante tutte le campagne di sensibilizzazione e sanzionatorie sulla privacy, solamente inserendo semplici stringhe per effettuare filtri nella ricerca Google si possono estrarre password, dati personali e sensibili nascosti nei meandri dei Big Data indicizzati dal famigerato motore di ricerca. Effettuare periodicamente un'attività di **intelligence preventiva** su Google relativamente alla presenza dei propri dati personali in rete è fortemente consigliato.

Questa attività di intelligence, utilizzando esclusivamente Google, dimostra quante informazioni possono essere presenti sul web e la potenza degli strumenti a disposizione per cercarle. Ma se da una parte possiamo facilmente interagire attraverso programmi come Google per svolgere le nostre personali attività di intelligence, dall'altra - essendo sempre più interconnessi e interattivi - ci troviamo sempre più spesso di fronte a forme di **rappresentazione virtuale** di persone che possono apparire non veritiere, come nel caso dei profili falsi (fake) sui social. Quando ci troviamo nel cyber spazio, di fronte a situazioni non del tutto chiare che meritano approfondimento e richiedono la nostra interazione con estranei, non dobbiamo correre il rischio di esporre subito le informazioni non pubbliche sulla nostra identità.

Prima di proseguire però va chiaramente detto che è fortemente sconsigliato creare **profili falsi** anche se per giuste cause. Di per sé, in alcuni ambiti aprire un profilo falso non è reato a meno che non si assumano i connotati di persone reali: rubare l'identità ad una persona realmente esistente prendendone le generalità e magari usando le foto del suo profilo è un reato punito dall'art. 494 del Codice Penale, come è reato utilizzare un falso profilo per molestare altre persone, adescare minorenni, etc. Il regolamento Facebook vieta ad es. di fornire informazioni personali false e creare più di un account personale (pratica tuttavia frequente) ma dal punto di

vista giuridico il doppio profilo non è di per sé sanzionabile, se non utilizzato per compiere attività illecite.

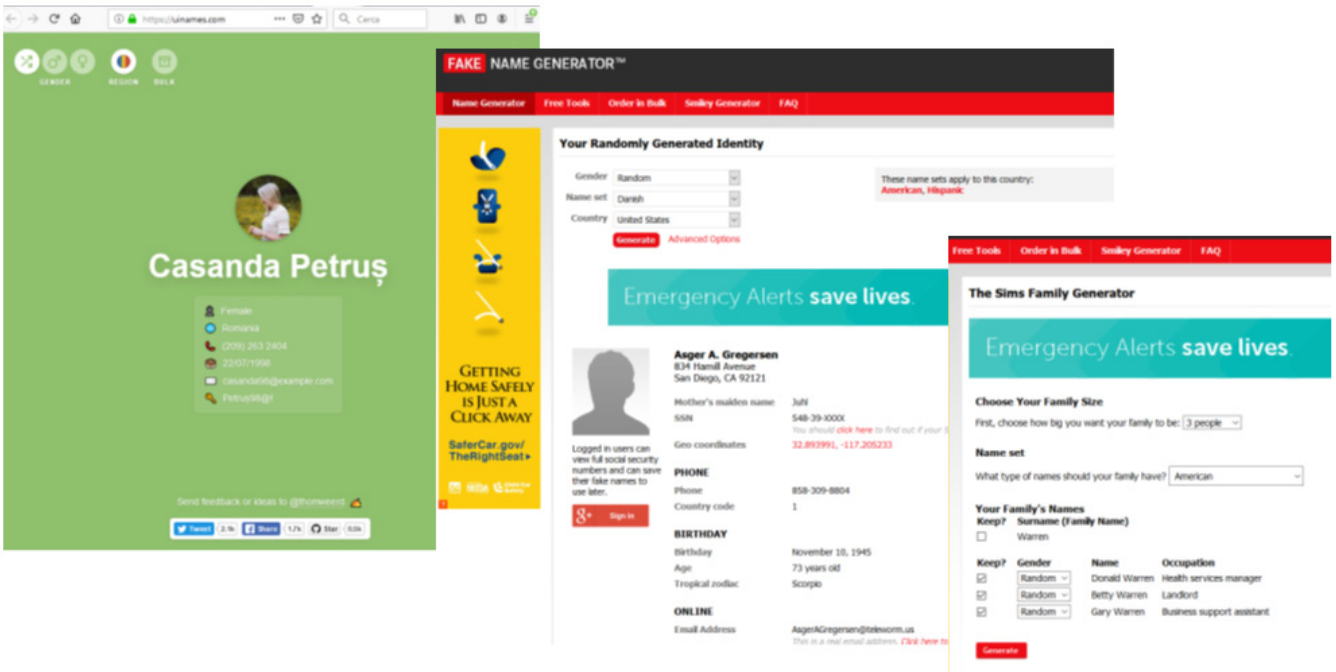
Detto questo, se vi trovate in una delle situazioni sopra elencate l'unica soluzione è quella di rivolgersi alla Polizia Postale, mentre in tutti quei casi in cui non ci sentiamo completamente sicuri e vogliamo cercare di capire meglio, possiamo utilizzare alcune tecniche e metodologie del Virtual Humint applicandole però al contrario, cioè cercando di capire se quel profilo su cui abbiamo dei dubbi mostra delle evidenze ricollegabili a tipiche attività di mascheramento.

In termini generali un profilo falso (fake) segue le stesse regole di creazione delle false identità denominate *"sock puppet"*, generate nelle operazioni di Virtual Humint; solo che il fake è normalmente più grezzo perché chi lo gestisce di solito non deve infiltrarsi, ad es., in un gruppo social di cellule terroristiche, quindi spesso è meno accorto ai dettagli e quindi più facilmente smascherabile.

Un falso profilo per essere credibile deve innanzitutto essere coerente e avere personalità: l'assenza di foto personali, del volto o di momenti della sua vita con altre persone non depongono a suo favore.

Sui dati personali postati potrebbe essere più complicato; esistono moltissimi siti che offrono un servizio di creazione di profili random con alcune informazioni di base coerenti con il fake generato.

Ad esempio **https://uinames.com** e **https://www.fakenamegenerator.com** sono alcuni di questi.

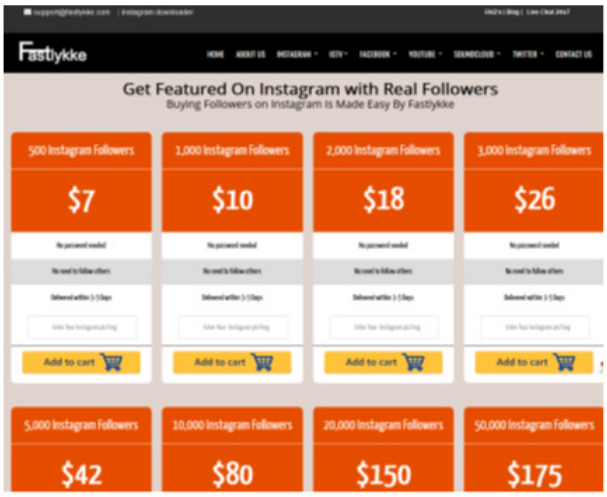


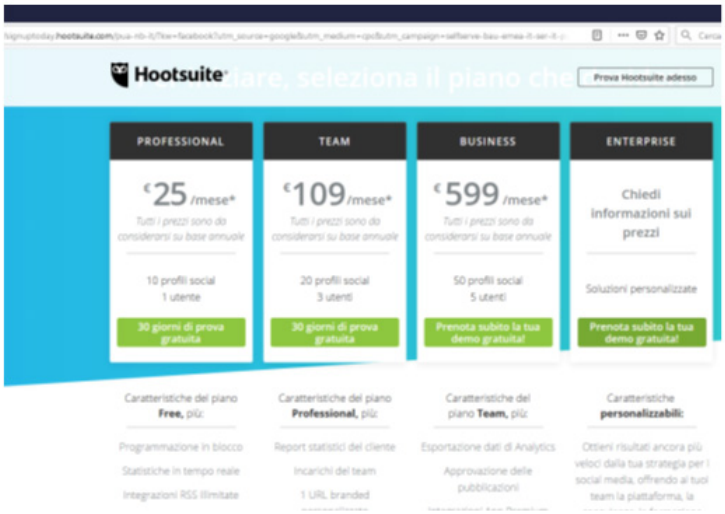
È sufficiente scegliere la nazione di appartenenza del falso profilo e il sito lo genererà. Volendo,

è possibile creare anche intere famiglie fake con profili correlati. Quindi le informazioni che potrebbe esporre il falso profilo potrebbero trarci in inganno perché ben strutturate e coerenti.

Verificate i follower e i like dei suoi post, l'assenza o un numero minimo (sotto i 3) di like e follower non è normale; non è indice di una vita sociale reale.

Nelle operazioni di Virtual Humint il falso profilo (*sock puppet*) viene spesso alimentato da like e follower acquistati su specifici servizi online, come ad esempio: **https://www.fastlykke.com.**





Ciò serve a dare credibilità al falso profilo.

Anche la frequenza dei post va verificata: un profilo che posta con continuità in orari plausibili e diversi da quelli di lavoro e sonno è più coerente.

Nelle operazioni di Virtual Humint spesso vengono creati decine di *sock puppet* e per renderli credibili e "vivi" si utilizzano dei servizi per la pubblicazione programmata e automatica di post come ad esempio **http://hootsuite.com**.

Anche la qualità dei post è importante: se sono solamente condivisi non commentati, sporadici e soprattutto non aggiornati, la cosa è sospetta.

A volte l'interazione con profili dubbi richiede la ricezione o l'invio di mail. In questi casi - se non volete utilizzare la vostra mail o crearne una nuova - potete utilizzare dei servizi di mail temporanea come quello di **https://www.guerrillamail.com,** che vi permette di avere subito una mail temporanea in grado di inviare, ricevere e rimanere anonimi.

Infine, anche nel caso fosse necessario dover dare un numero di cellulare per ricevere un SMS da un profilo o un servizio su cui si ha qualche dubbio, esistono servizi con numeri telefonici temporanei di varie parti del mondo che possono essere utilizzati gratuitamente per ricevere messaggi, come ad esempio: **https://receive-smss.com.**

Che voi facciate intelligence “casalinga” o intelligence governativa, la **regola finale** rimane sempre la stessa: non si deve mai perdere la concentrazione su quello che si sta facendo. Dall’omissione di un’informazione alla creazione di bugie il passo è breve: si deve sempre avere il controllo operativo ed etico dell’operazione, sia nel modo in cui si cerca l’informazione di interesse sia nel modo in cui l'informazione è appositamente "deviata" per tutelare la nostra riservatezza. Perdere il controllo su questo significa perdere i riferimenti sulla validità delle informazioni ed essere costretti ad analizzarle tutte indistintamente e, come scrive Harari nel suo libro *Homo Deus*, «Nei tempi antichi deteneva il potere chi aveva accesso alle informazioni. Oggi avere potere significa sapere cosa ignorare».

Articolo a cura di **Francesco Arruzzoli**