Content analysis of open source intelligence data using AI and Natural Language processing

Artificial Intelligence – An opportunity for the EU cyber-crisis management

V1.5

George Bara

- Director of Strategic Accounts, Government - Linguistic AI @ SDL
- OSINT & AI Consultant @ Zetta Cloud
- Researcher on #FAKENEWS AI algorithms

Who is SDL*

Publicly traded company (LSE:SDL)

Provider of Language and Content Management Solutions to the Global Enterprise for Over 25 Years

Delivering comprehensive solutions to help large organizations better manage content creation, translation and content delivery challenges across the globe

SDL*

AGENDA

Using Actionable Threat Intelligence to Counter Cyberattackers

Open Source Data Intelligence Processing Challenges

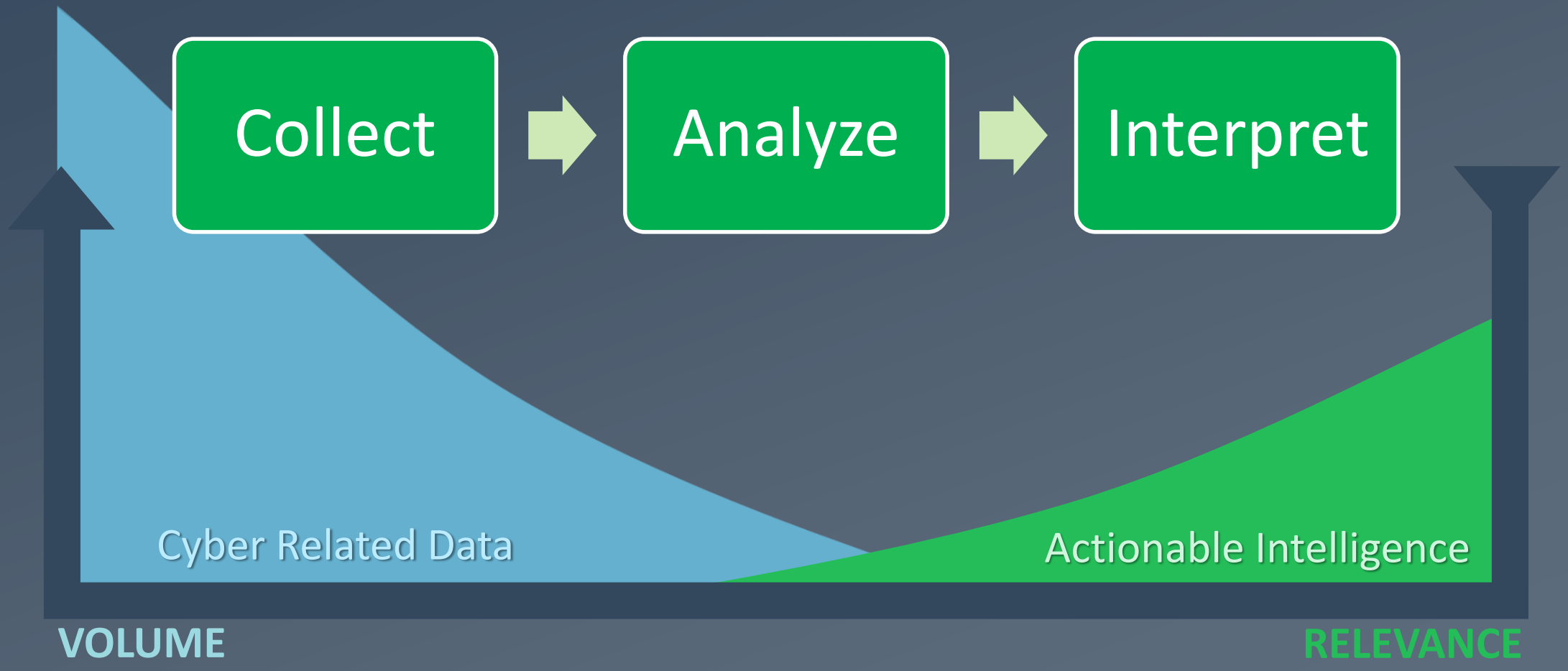
Linguistic Artificial Intelligence Solutions for Cyber Intelligence

CYBER THREAT INTELLIGENCE

WHY IS IT IMPORTANT?

“Organizations must develop and effective cyber threat detection and response program, in order to overcome the threats they face”

Threat Intelligence Process



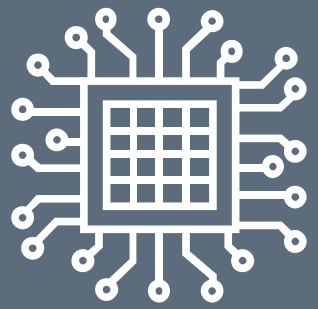
SDL*

Cyber Threat Intelligence Sources on EMERGING THREATS and THREAT ACTORS

Open Web

Social media

Partners



Internal Analysis

Dark Web

SDL*

OPEN SOURCE CYBER INTELLIGENCE

LANDSCAPE & CHALLENGES

The Open Web

4,161,071,868
Internet Users in the world

1,757,185,699
Total number of Websites

167,155,758,929
Emails sent today

4,218,626,319
Google searches today

4,000,674
Blog posts written today

486,156,761
Tweets sent today

4,500,184,433
Videos viewed today on YouTube

52,112,121
Photos uploaded today on Instagram

86,657,681
Tumblr posts today

SDL*

“

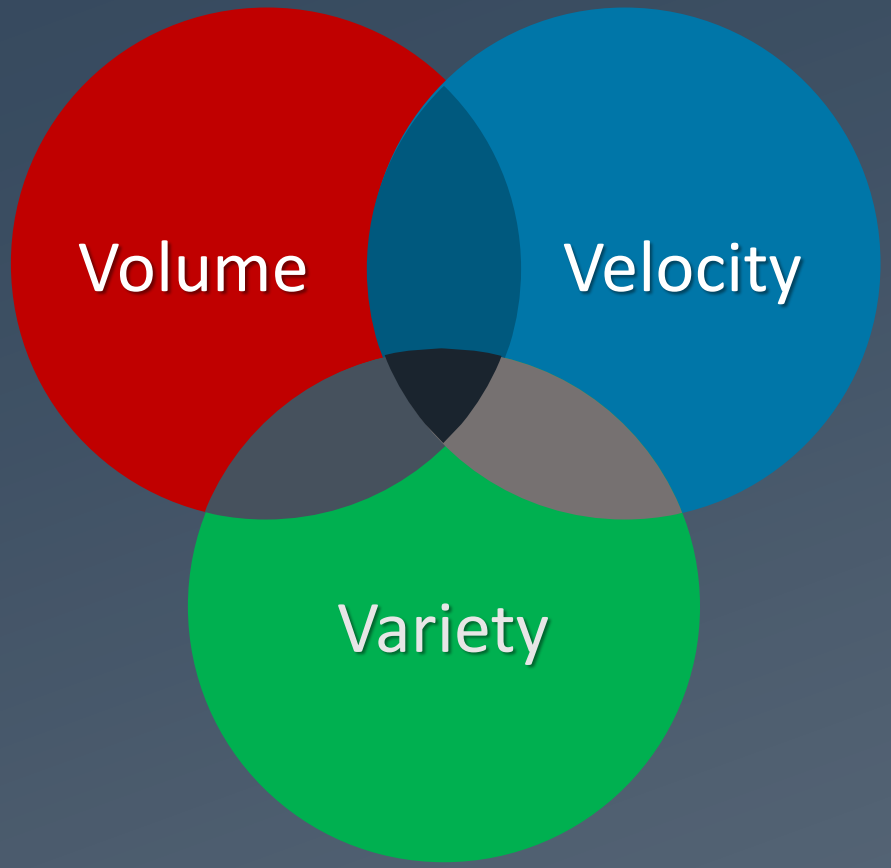
The **dark component of the deep** web is the primary highway for the exchange and commerce among cybercriminal groups.

In fact, **very few cybercriminals work alone**.

Eighty percent of cybercrime is linked to criminal collectives, and stolen data-shaped goods surface rapidly on **darknet forums and marketplaces** following cybersecurity incidents with data loss.

”

The BIG LANGUAGE challenge



X

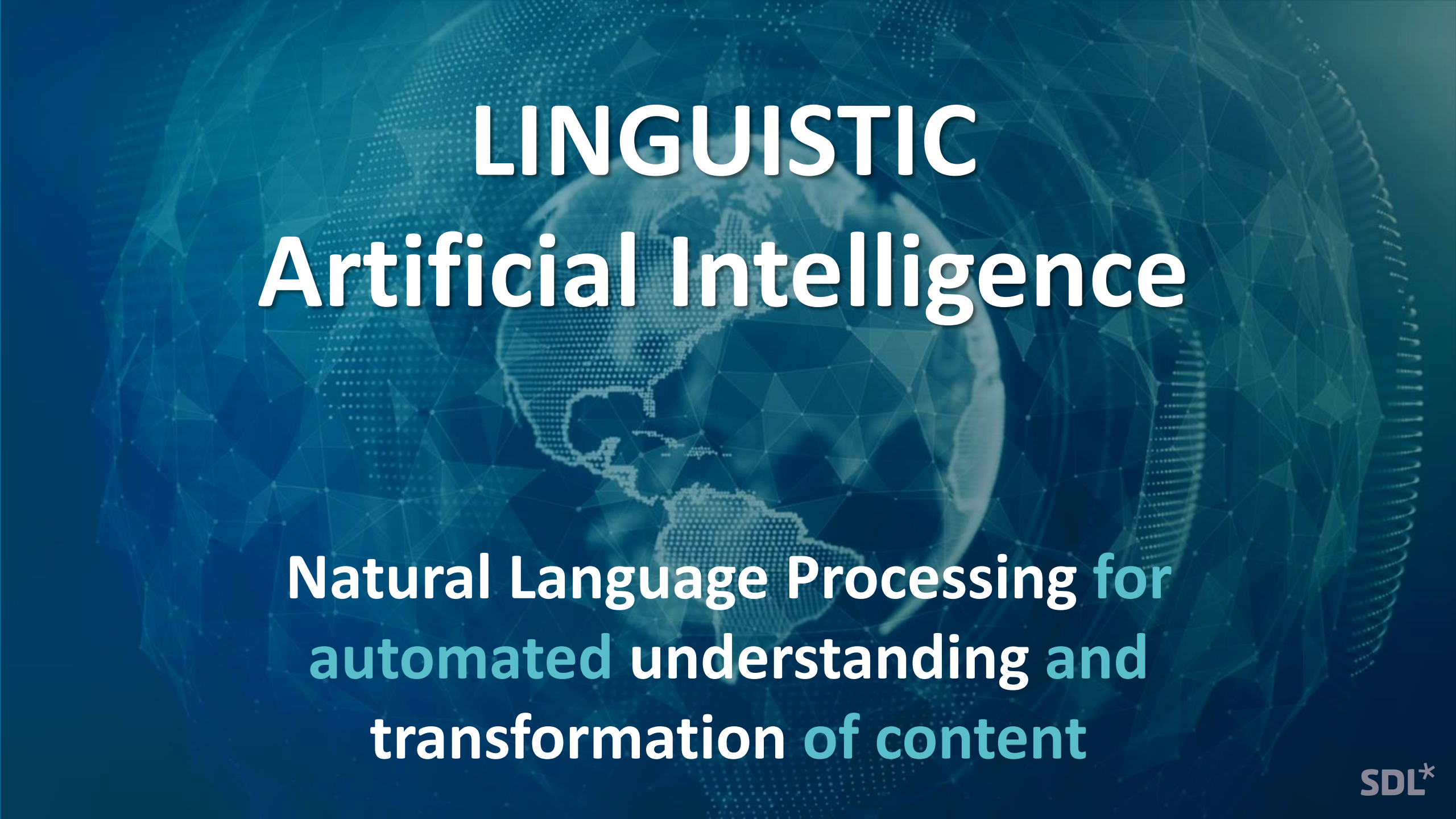
3V

Language

SDL*

LINGUISTIC AI FOR CYBER THREAT INTELLIGENCE

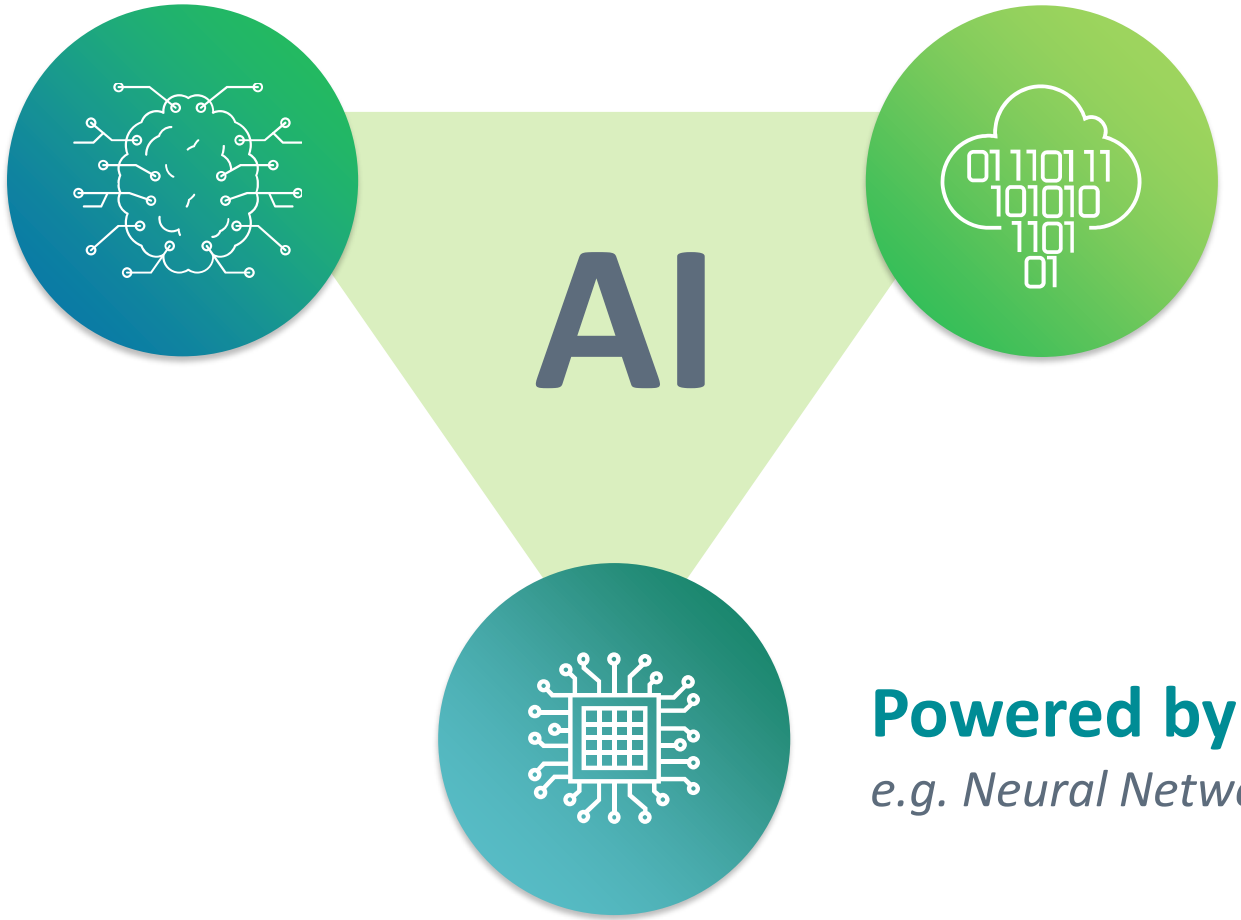
TACKING THE “BIG LANGUAGE” PROBLEM


LINGUISTIC
Artificial Intelligence
Natural Language Processing for automated understanding and transformation of content
SDL*

What is Linguistic AI?

Language Transformation

e.g. Translation



Language Understanding

e.g. Summarizing, Named Entity Recognition, Classification, Trustworthiness Assessment

Powered by Machine Learning

e.g. Neural Networks & Deep Learning

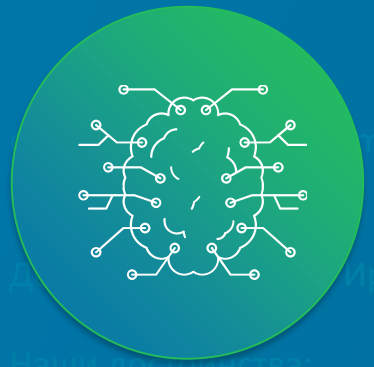
/dark-cache?token=f9a2f4c6-2995-48f4-983f-59415765c7a2&id=https%3A%2F%2Fy3pggjcimtcglaon.onion%2Fkatalog-tenevyh-resursov-142%2Fjumanji-pro-tenevoi-rynok-tovarov-i-uslug-189035%2F

darkclientbp@exploit.im @dark_client 2 Авг 2018 #1 Нужен полностью абузоустойчивый сервера под любой проект - Сервера DarkClient , ваш выбор!

Доступная локация - Иран, Панама, Гонконг, Нидерланды, Сингапур, Россия, Украина, Бахрейн

Наши достоинства:
Uptime 99,99% Игнорирование DMCA / Spamhaus Offshore Сервера Любой конфиг на ваш вкус за копейки, из облака! Scan, Bruteforce, Spam, Fishing - Полностью разрешены 8+ Стран, 12 Дата-Центров Гарантированная анонимность Оплата Криптовалютой (Bitcoin/Litecoin и другие) Ценовая политика (Цена сервера начинается от 20 $) Индивидуальный подход к каждому клиенту Полная отказоустойчивость Команда профессионалов Выделенный 1 Gbps канал на каждый сервер Установка выделенных серверов занимает 4 часа \nКонтакты: \nTelegram: @dark_client (Для просмотра ссылки необходимо: Войти или Регистрация) Jabber(xmpp): darkclientbp@exploit.im Jabber(xmpp): darkclientbp@xmpp.jp
Оплата:
Bitcoin / ETH / LTC Payeer QIWI Другие (Спрашивать у оператора) Доступная конфигурация (Абузоустойчивый VDS/VPS) :
CPU - От 1 До 6 ядер RAM - От 1 До 64gb SSD - От 25gb до 2000gb Канал - 800Mbps Цены на нужную вам конфигурацию уточняйте у оператора, цены от 25$!

Выделенные сервера (Абузоустойчивые) :
Спойлер: Доступное железо Процессоры:
Xeon E3-1230 4x3.2GHz, Xeon E3-1230v3 4x3.3GHz, Xeon E3-1280v2 4x3.6GHz, Xeon E5-1650 6x3.2GHz, Xeon E5-1650v3 6x3.5GHz, Xeon E5-1650v4 6x3.6GHz.
Оперативная память (Комплекты) :
16gb, 32gb, 64gb, 96gb, 128gb, 256gb.
Канал (Скорость интернета) :
Скорость - 1 GBit/s, * До 64 Ipv4, * Блок IPv6, по желанию.
HDD/SSD:
Любые RAIDы, в наличии есть практически все! Сервера в основном на 4-8 корзин(слотов под hdd/ssd). *Нет нужной конфигурации? Спрашивай у оператора, и он подскажет!



Language Transformation

Automatically translate the content from Russian into your language (e.g. English) using high-speed & high-quality Neural Networks algorithms

darkclientbp@exploit.im @dark_client Aug 2, 2018 # 1 Need a fully bulletproof server for any project - DarkClient Server, your choice!
Available location - Iran, Panama, Hong Kong, the Netherlands, Singapore, Russia, Ukraine, Bahrain
Our advantages:
Uptime 99.99% Ignoring DMCA / Spamhaus Offshore Server Any config for your taste for a penny, from the cloud! Scan, Bruteforce, Spam, Fishing - Fully resolved 8+ Countries, 12 Data Centers Guaranteed anonymity Payment Cryptocurrency (Bitcoin / Litecoin and others) Pricing (Server price starts from $ 20) Individual approach to each client Full fault tolerance Team of professionals Dedicated 1 Gbps channel per server. Installing dedicated servers takes 4 hours.
Contacts:
Telegram: @dark_client Jabber (xmpp): darkclientbp@exploit.im Jabber (xmpp): darkclientbp@xmpp.jp

Language Understanding

Automatically extract information and meaning from content using Text Analytics: categorization, sentiment analysis, named entity extraction, semantic similarity, topic extraction, name matching, entity linking, relationship extraction.

Email
- darkclientbp@exploit.im
- darkclientbp@xmpp.jp

Date
- Aug 2, 2018

Location
- Iran
- Panama
- Hong Kong
- Netherlands
- Singapore
- Russia
- Ukraine
- Bahrain

Money
- $ 20

Organization
- The Spamhaus Project
- xmpp
- Qiwi

Automation of critical, laborious tasks significantly speeds up the process and reduces cost

A professional translator can translate 2,000 – 3,000 words per day.

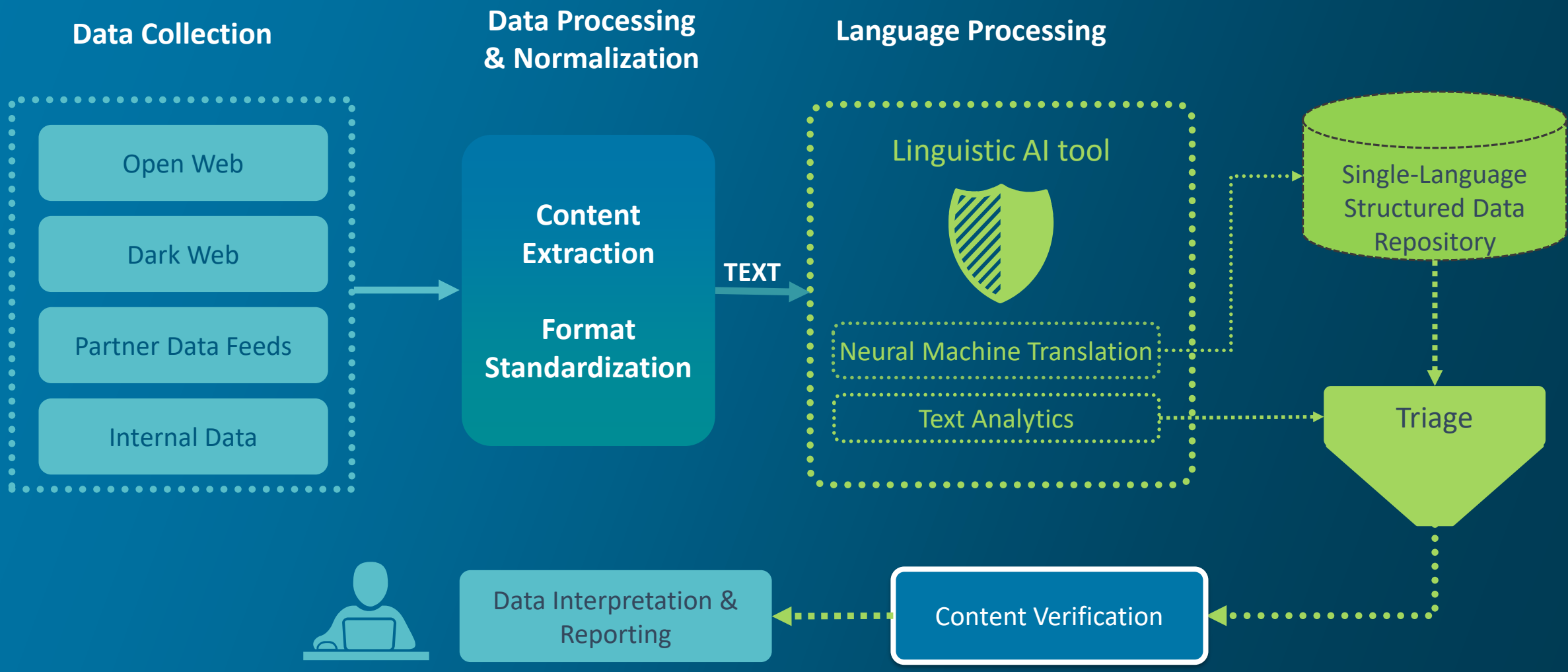
A Neural Networks Machine Translation software can translate from 2,000 words per minute

The Neural Machine Translation Revolution

SDL, Google & Microsoft have all produced NMT systems that can produce output equivalent to human translation

The Neural MT Revolution

Linguistic AI for Cyber Threat Intelligence



Future AI Challenges for Cyber Intelligence

Language Generation

Ability to automatically and distribute create content

- Relies on existing structured data sets
- Text Analytics to extract meaningful information
- Identifies key facts and correlates them with sentences
- Combines phrases to form grammatically correct sentences
- **Automatically generates alerts and reports in realtime, then distributes them to the stakeholders**

Content Verification

Ability to automatically verify of information quality

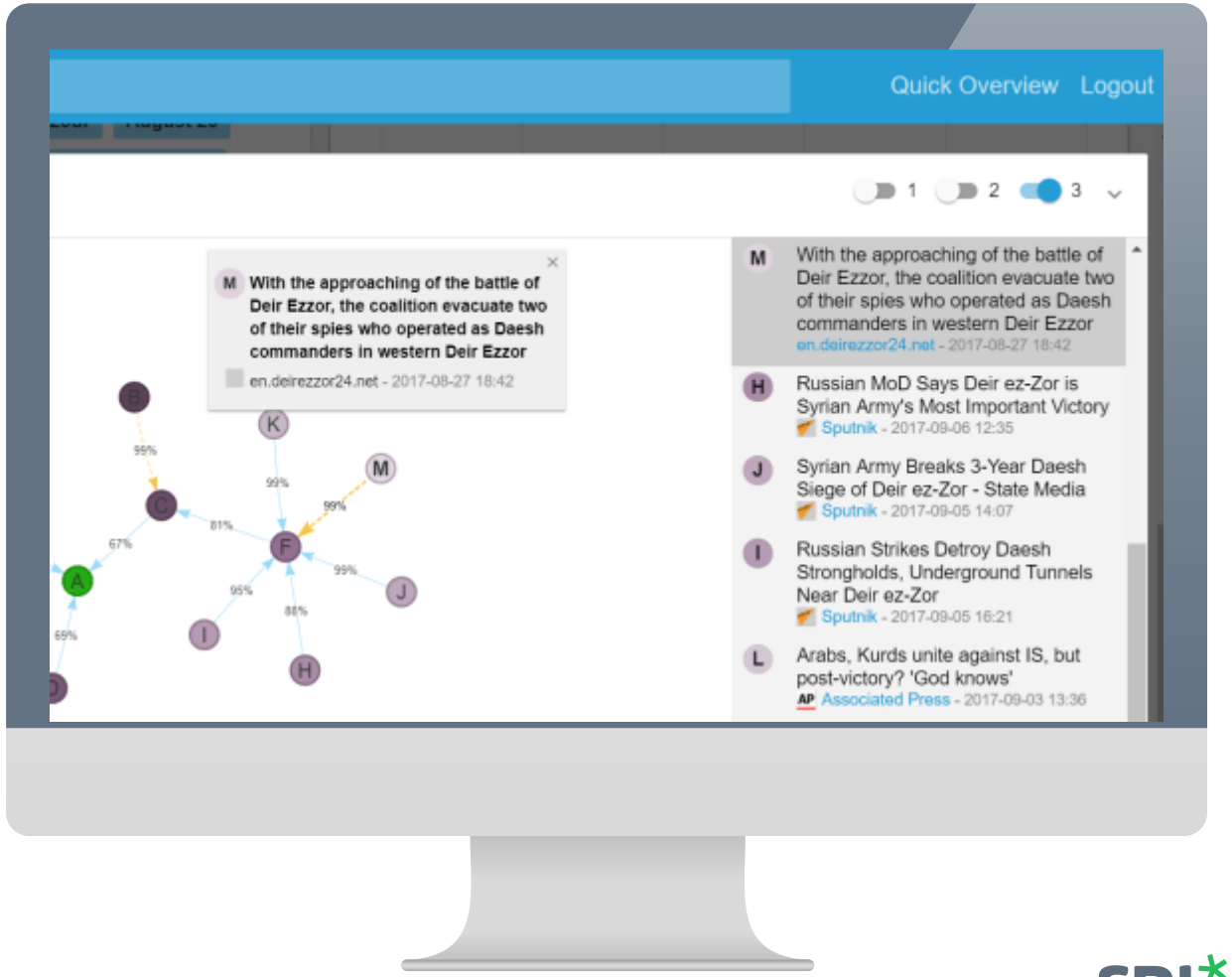
- Contextual analysis of multiple data sources
- Rely on existing databases of threat actors & techniques
- Correlate multiple information sources to verify information
- Identify content “red flags” through text analytics
- **Automatically flag information that is not trustworthy and that could constitute deliberate disinformation**

SDL*

TrustServista - #fakenews detection

AI platform for automatically determining the **trustworthiness** of online content:

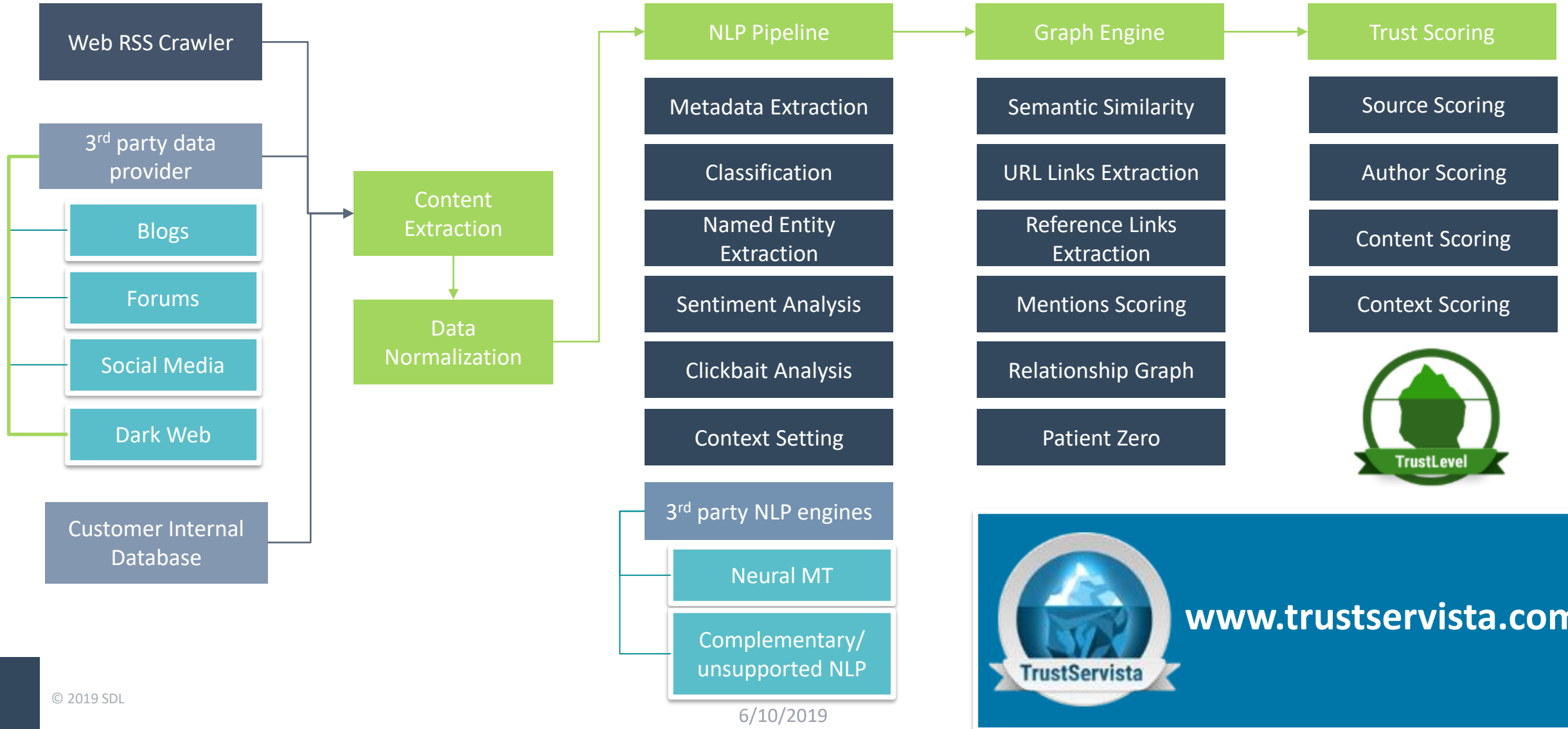
- **Source veracity**: publisher type, author profile, hosting location.
- **Content Quality**: clickbait analysis, polarity, information weight, classification.
- **Context Analysis**: similar articles graph, inbound and outbound links, Patient Zero.



Designed for:

- Automated content filtering.
- Human decision assist.
- Alerting system input.

TrustServista – Solution Architecture



Software and Services for Human Understanding