Europcar: Behind the Fake Data Breach

by Reece Baldwin on February 5, 2024

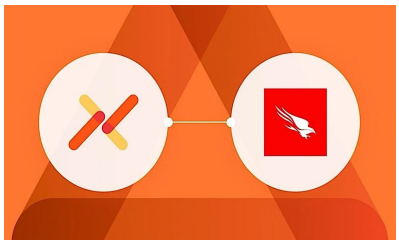
On January 28, 2024, a user with the handle “Lean” posted on a criminal forum that they had 48.6 million customer records from a supposed breach of Europcar. Europcar is a car rental company with a global presence, operating in over 140 countries. Lean claimed the dataset contained sensitive data such as usernames, passwords, full names, addresses, birth locations, passport numbers, driver’s licenses, and bank details – all information real customers of a car rental company would have provided and everything a threat actor would need to commit large-scale fraud. Lean was soliciting offers, waiting for the highest bidder. A huge profit potential if true. But, what if it was all made up? Faked?

When criminals want to sell a large trove of data, they often provide a sample set. This gives potential customers a way to validate the information. It also allows the affected company to review the data and identify whether those records exist in their database. Lean provided a sample set of 31 accounts in their post.

Bleeping Computer reached out to Europcar asking about the potential breach. They responded that they had not suffered a breach and that the data had potentially been fabricated by Artificial Intelligence (AI). Bleeping Computer's original article contained the response from Europcar that stated:

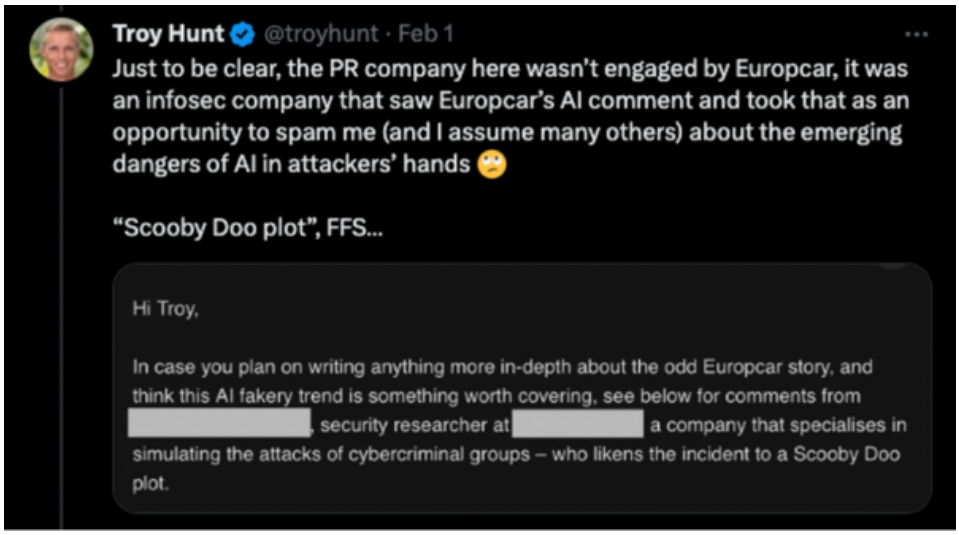
> – the number of records is completely wrong & inconsistent with ours,
>
> – the sample data is likely ChatGPT-generated (addresses don't exist, ZIP codes don't match, first name and last name don't match email addresses, email addresses use very unusual TLDs),
>
> – and most importantly: none of these email addresses are present in our database.

This story would have likely stopped there. Nothing to see here folks, just someone trying to make a quick buck off of another criminal. The mention of AI set information security pundits and media outlets off. Troy Hunt, the creator of HaveIBeenPwned, a database of breached credential sets, tweeted that a PR company had contacted him on behalf of an infosec company with their hot take.



Aembit Announces New Workload IAM Integration with CrowdStrike to Help Enterprises Secure Workload-to-Workload Access

Most Read on the Boulevard



Why AI?

Artificial intelligence is in the zeitgeist. From systems that help unblur images from telescopes to the promise of better healthcare outcomes for patients, AI is touted as the panacea for a bunch of very difficult problems. Of course, there are other uses, such as getting ChatGPT to write your resume or creating deep fake videos or audio content to deceive individuals to part with money or information.

Arthur C. Clarke, the screenwriter of *2001: A Space Odyssey* famously said, "Any sufficiently advanced technology is indistinguishable from magic." AI is currently that advanced technology. It is very easy to slap AI into a sentence or a press release and have people run with it. The quote Troy provided from the PR agency can be found online, in full, in other news and blog articles. There is an urgency to be the first to make a comment, but this should never be to the detriment of thorough analysis. Companies within the cyber security industry have a duty to perform analysis – verifying and validating what they are seeing before issuing a spicy take.

These types of breaches – and the reporting surrounding them – have a real-world impact. A company subjected to a breach would call in members of their security team, legal representatives, and leadership team to deal with an incident of this scale. This affects those individuals in personal and professional ways, whether it be the change in family plans, the addition of workload, or the increase in pressure from within the company. Due diligence and good analysis are required before reporting on data breaches. Our analysis of the publicly available data took one of our analysts about five minutes to identify that it was likely fake. These insights can be helpful to the company, providing the added context to assist company decision-makers in taking informed actions.

Our Analysis

DoD Email Breach: Pentagon Tells Victims 12 Months Late

ALERT: Thieves❤️Wi-Fi Camera Jammers

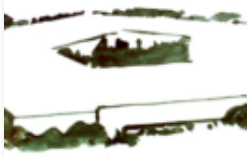
Identity Governance Has a Permission Problem

55% of Generative AI Inputs Include Sensitive Data: Menlo Security

IGAaaS Vs. On-Premises IGA Solutions: A Comparative Analysis

New Malware in Exploits

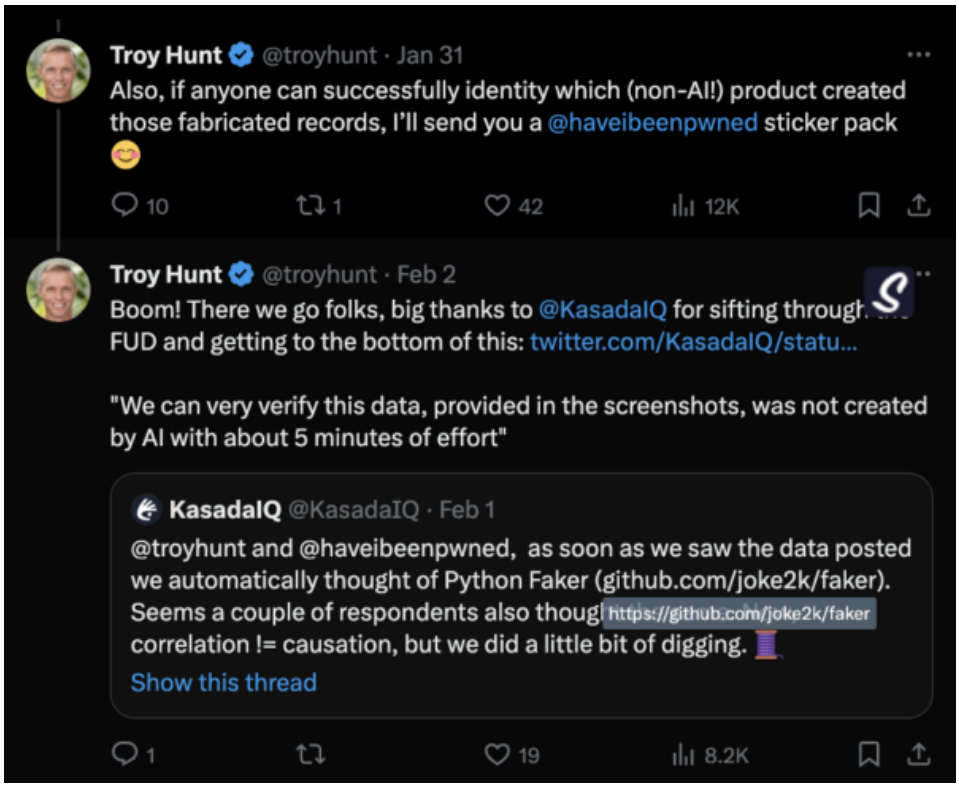
Industry Spotlight »


DoD Email Breach: Pentagon Tells Victims 12 Months Late

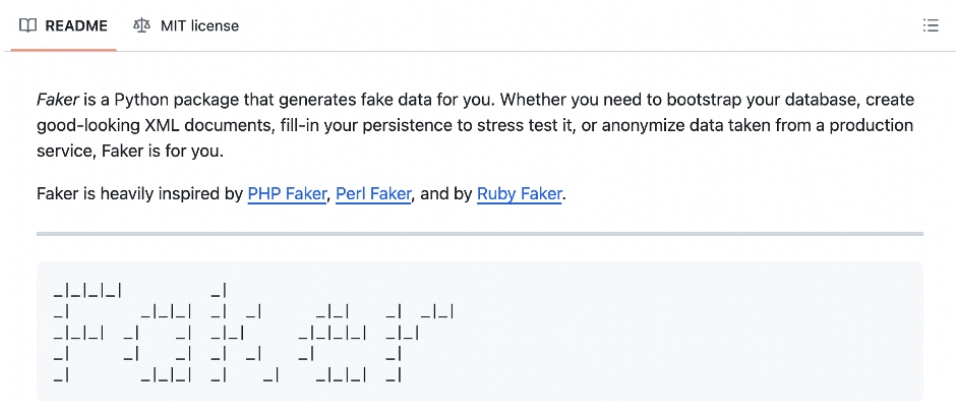

With SNS Sender, USPS Smishing Scams Move to the Cloud

In the following thread, Troy provided screenshots of the data and asked if anyone could identify the non-AI library used to create this data.

KasadaIQ, our threat intelligence service focusing on fraud, quickly identified that a common, open-source Python library called Faker was used.



Faker is a Python library used to generate fake test data. The README on their GitHub repository states the following:



We weren't the first to suggest Faker, but we are the first to do a bit of digging and provide solid proof points to support our analysis. Our analysis was conducted using only the screenshots provided in X, as we did not have access to the original source material.

ALERT: Thieves ❤️Wi-Fi Camera Jammers

Top Stories »

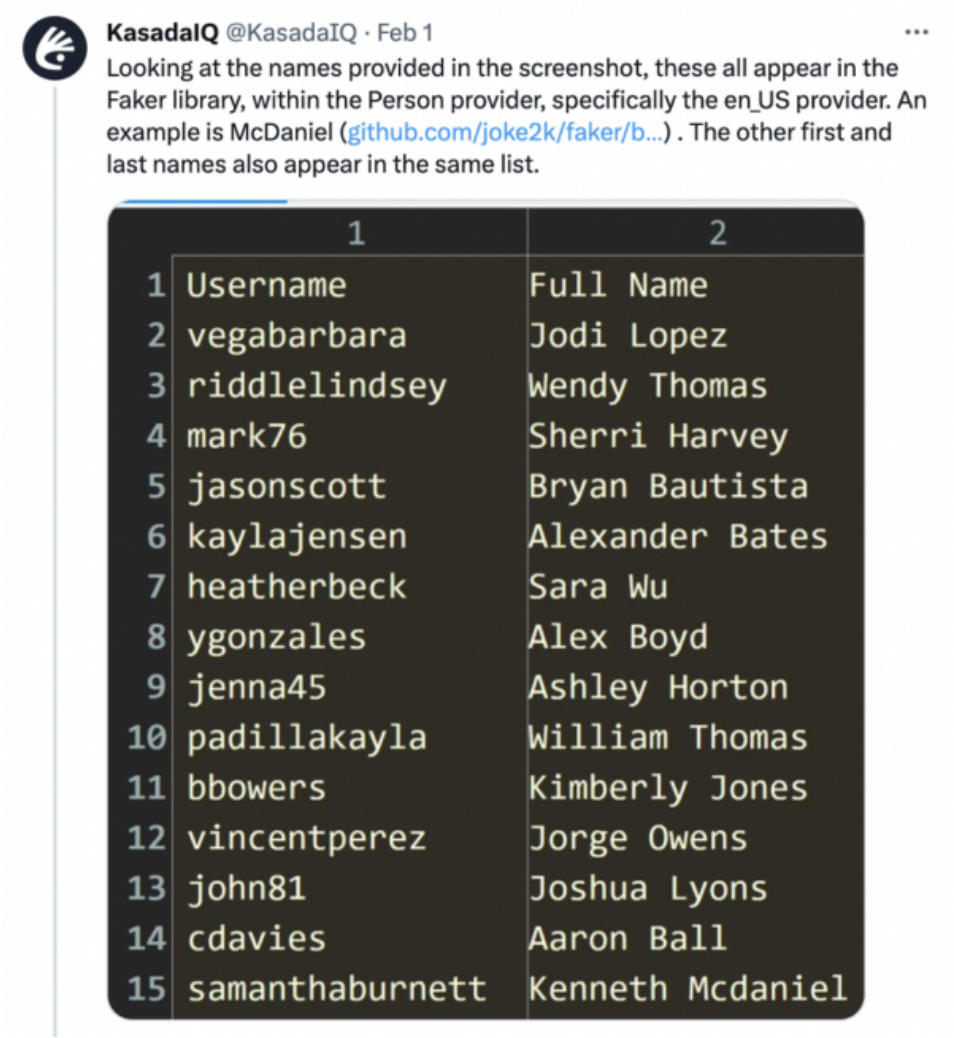
US Offers $10M for Info on BlackCat /ALPHV Ransomware Leaders

Feds Disrupt Botnet Used by Russian APT28 Hackers

FTC Warns AI Companies About Changing Policies to Leverage User Data

Security Humor »

Daniel Stori's 'Clojure Challenge'



Reviewing the Full Name, we identified that both the first and last names appeared in the list of available US_en names. The Usernames, as others identified, did not match the Full Names provided. It is typically expected that if users are creating a username that contains some real-world identifier, they would closely align that username with their real name.

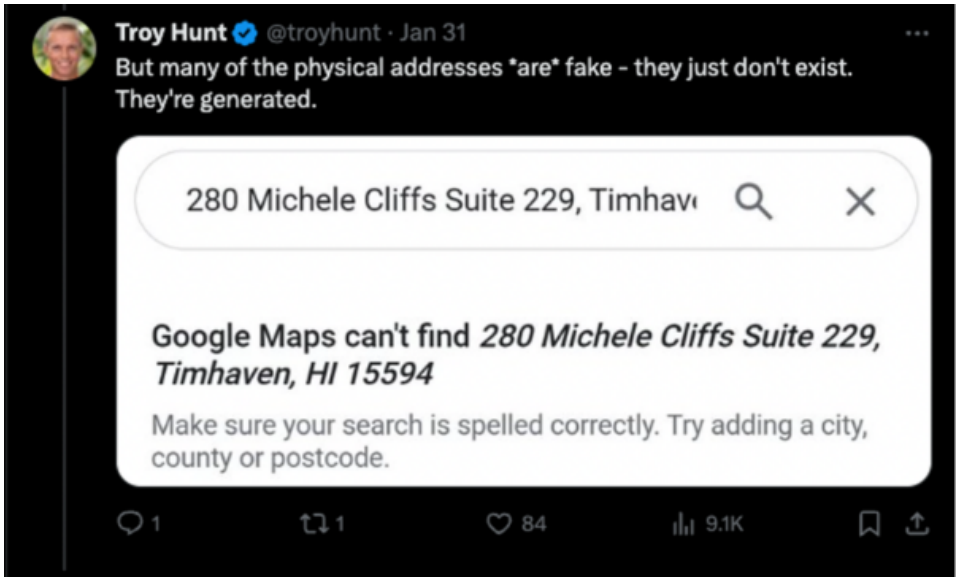
4	5	6
City	ZIP	City of Birth
Port Katie	10965	Port Brandyfurt
Lake Carolville	88894	West Randy
Johnville	75811	New Kellyburgh
Port Scott	71601	North Courtney
Aprilshire	50421	Millerfurt
Kingbury	38508	Tonyside
Potterport	15297	Lisatown
Ashleyland	63142	Latoyaberg
South Timothytown	53760	Lake Michael
Lukeberg	04703	Robertfurt
Adamsfurt	91526	Williamsonmouth
Amandafurt	05654	New Austinland
Danielleside	75075	New Peterhaven
Castilloshire	82025	Russellport
Courtneyfurt	41545	Mirandabury

An image containing the current residence city, Postal/ZIP Code, and City of Birth was also supplied. These cities are not real. Both the City and City of Birth columns have a first name in them. Places like Lisatown, Mirandabury, Danielleside, and West Randy do not exist. Reviewing the city formats tuple identified that city names are created in four ways, as shown in the image below.

```
class Provider(AddressProvider):

    city_formats = (
        "{{city_prefix}} {{first_name}}{{city_suffix}}",
        "{{city_prefix}} {{first_name}}",
        "{{first_name}}{{city_suffix}}",
        "{{last_name}}{{city_suffix}}",
    )

    street_name_formats = (
        "{{first_name}} {{street_suffix}}",
        "{{last_name}} {{street_suffix}}",
    )

    street_address_formats = (
        "{{building_number}} {{street_name}}",
        "{{building_number}} {{street_name}} {{secondary_address}}",
    )
```

In the city format, the First Name is derived from the same list as the first names that create an identity. The concatenation of a name with a prefix and/ or suffix results in the creation of barely passable city names. The same

technique is used for street names and addresses, which can be seen in this image shared by Troy Hunt.



There are 7 prefixes and 18 suffixes available in Faker for City Name generation.

North	East	West	South	New	Lake	Port

Faker City Prefixes

town	ton	land	ville	berg	haven
burgh	borough	bury	view	port	side
mouth	stad	furt	chester	fort	shire

Faker City Suffixes

When coupled with the available names in the Faker library, these align with the data provided in the screenshot.

The warning signs were there early on. Lean's account had only recently joined the forum and had next to no reputation within the community. The sample set looked very odd, with strangely placed names and a mismatch of email addresses and names.

While Faker can be legitimately used to create test data for databases, it is highly unlikely this data is from a Europcar database – whether it was test data or not. In our experience dealing with criminal forums, sellers want to display legitimacy to potential buyers by providing good-quality sample data. If the seller had approximately 50 million records, they could have chosen any rows within the data to publish. But instead, they chose a clearly faked selection. The original post from Lean has been removed from the criminal forum, with many users also calling it fake.

Actionable Insights

Amidst the chaos of a potential data breach, you must move quickly to make well-informed decisions to mitigate risks while preserving customer and stakeholder trust. Attacking a company's brand and reputation with false data breaches will only get even easier and more common. The ability to discern accurate information from misinformation is paramount. Finding credible signals online can be invaluable in distinguishing fact from fiction. However, knowing where to look and who to trust is crucial. By collaborating and sharing our knowledge and expertise, the security community becomes a united front against malicious and criminal activities. At Kasada, we champion due diligence, foster collaboration, and uphold the responsibility to provide accurate evidence and analysis.

If you'd like actionable insights on threats to your company, get your free KasadaIQ Snapshot or contact our team of experts behind KasadaIQ.

The post Europcar: Behind the Fake Data Breach appeared first on Kasada.

*** This is a Security Bloggers Network syndicated blog from Kasada authored by Reece Baldwin. Read the original post at: https://www.kasada.io/europcar-fake-data-breach/

Cybersecurity

← D3 Smart SOAR's Integration with CrowdStrike Falcon XDR Joins the CrowdStrike Marketplace

Key Differences Between Two-Factor Authentication (2FA) and Multi-Factor Authentication (MFA) →