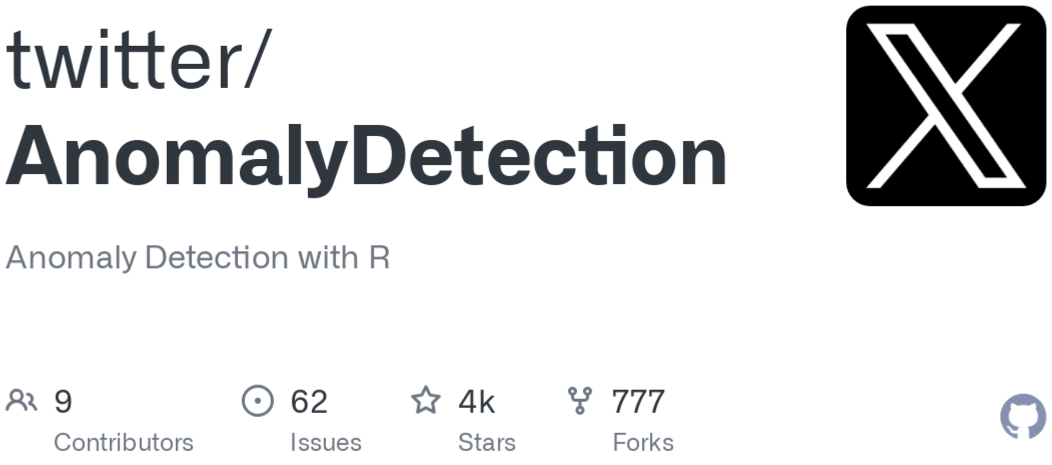
Artificial Intelligence in Cyber Defense: A Review of Current Applications and Technologies

Ahsen Beyza Özkul

ahsenbeyza@securedebug.com

1) **AI for Threat Detection and Incident Response**

- **Project: Twitter's Anomaly Detection Tool** *Example: Twitter created a tool to spot unusual patterns in data like network traffic and user activities. This tool monitors things in real-time and uses machine learning to quickly detect and alert on any strange behavior.*



twitter/

AnomalyDetection

Anomaly Detection with R

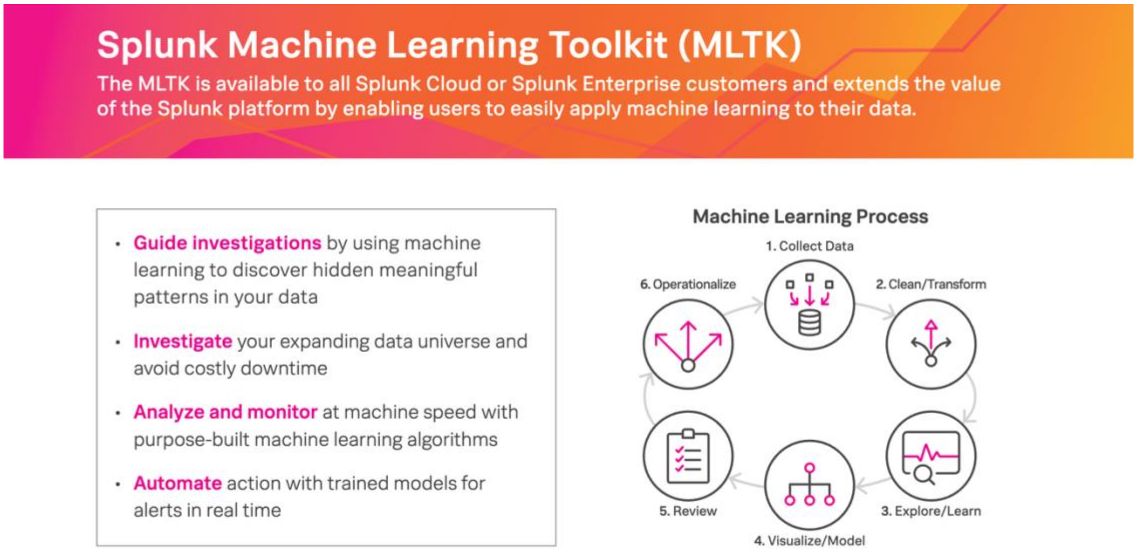
9 Contributors

62 Issues

4k Stars

777 Forks

- **Project: Splunk's Machine Learning Toolkit** *Example: Splunk offers a set of tools that use machine learning to detect anomalies in system logs and operations. It's commonly used in IT to identify unusual activities and send alerts before they turn into bigger issues.*

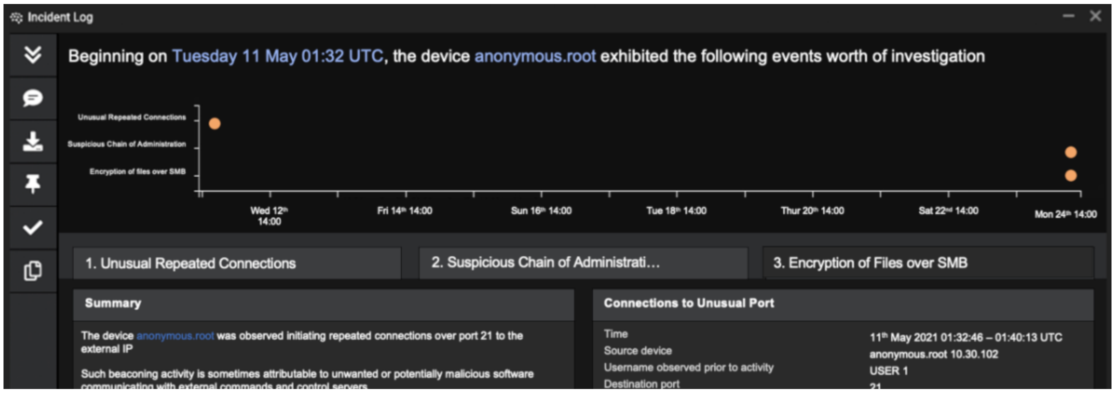


Splunk Machine Learning Toolkit (MLTK)

The MLTK is available to all Splunk Cloud or Splunk Enterprise customers and extends the value of the Splunk platform by enabling users to easily apply machine learning to their data.

- **Guide investigations** by using machine learning to discover hidden meaningful patterns in your data
- **Investigate** your expanding data universe and avoid costly downtime
- **Analyze and monitor** at machine speed with purpose-built machine learning algorithms
- **Automate** action with trained models for alerts in real time

- **Project: Darktrace AI for Incident Response** *Example: Darktrace uses AI to automatically respond to cyber threats. When an attack happens, Darktrace can isolate the affected parts of your system to prevent the problem from spreading, allowing your security team to focus on fixing the issue without worrying about it getting worse.*



Incident Log

Beginning on Tuesday 11 May 01:32 UTC, the device anonymous.root exhibited the following events worth of investigation

Unusual Repeated Connections
Suspicious Chain of Administration
Encryption of files over SMB
Wed 12th 14:00
Fri 14th 14:00
Sun 16th 14:00
Tue 18th 14:00
Thur 20th 14:00
Sat 22nd 14:00
Mon 24th 14:00
1. Unusual Repeated Connections
2. Suspicious Chain of Administrati…
3. Encryption of Files over SMB

Summary

The device anonymous.root was observed initiating repeated connections over port 21 to the external IP

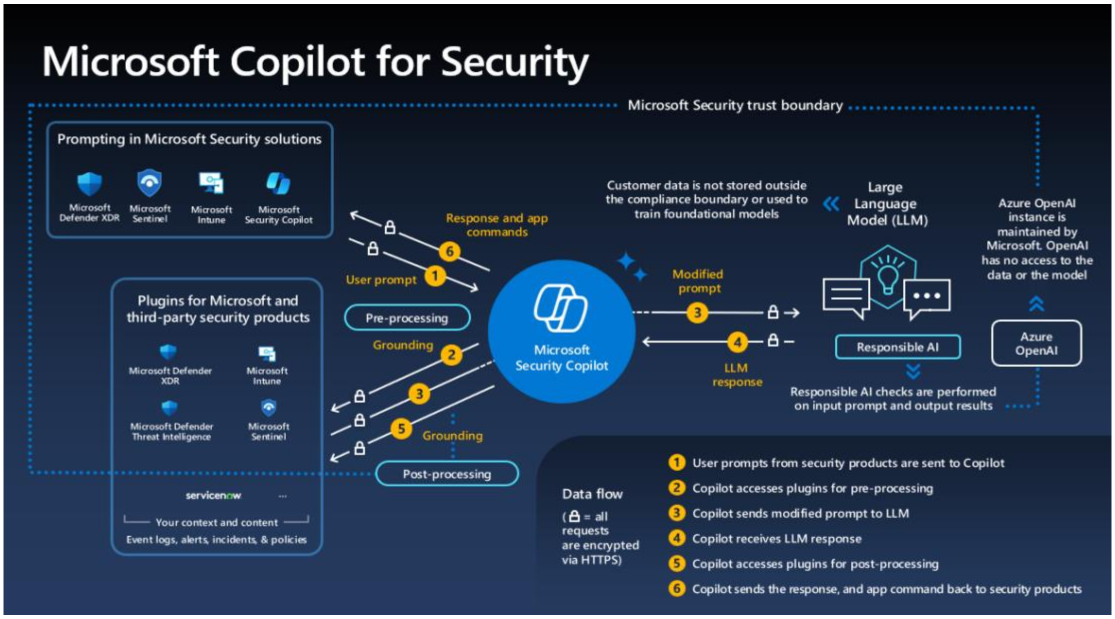
Such beaconing activity is sometimes attributable to unwanted or potentially malicious software communicating with external commands and control servers

Connections to Unusual Port

Time	11th May 2021 01:32:46 – 01:40:13 UTC
Source device	anonymous.root 10.30.102
Username observed prior to activity	USER 1
Destination port	21

- **Project: Microsoft Security Copilot for Incident Response** *Example: Microsoft Security Copilot helps automate the process of dealing with cyberattacks. It can quickly respond to incidents by isolating the problem, providing the necessary information to fix it, and restoring systems to normal. This saves time and effort for security teams.*

- **Project: Microsoft Security Copilot** *Example: Microsoft Security Copilot is a tool that uses AI to help keep your network safe. It works by constantly scanning for any unusual activities or potential threats and can react immediately to stop them. This makes it easier for companies to protect their data without relying solely on human monitoring.*



Microsoft Copilot for Security

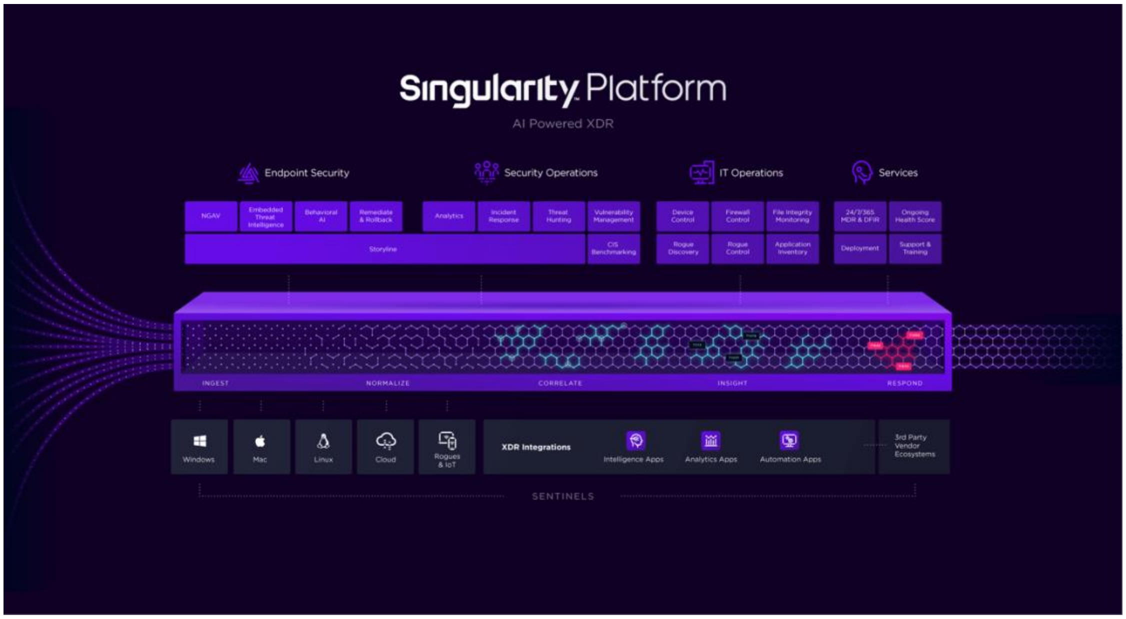
Microsoft Security trust boundary
Prompting in Microsoft Security solutions
Microsoft Defender XDR
Microsoft Sentinel
Microsoft Intune
Microsoft Security Copilot
Plugins for Microsoft and third-party security products
Microsoft Defender XDR
Microsoft Intune
Microsoft Defender Threat Intelligence
Microsoft Sentinel
servicenow
...
Your context and content
Event logs, alerts, incidents, & policies
Response and app commands
6
User prompt
1
Pre-processing
Grounding
2
3
5
Grounding
Post-processing
Microsoft Security Copilot
Modified prompt
3
4
LLM response
Customer data is not stored outside the compliance boundary or used to train foundational models
Large Language Model (LLM)
Responsible AI
Responsible AI checks are performed on input prompt and output results
Azure OpenAI instance is maintained by Microsoft. OpenAI has no access to the data or the model
Azure OpenAI

Data flow

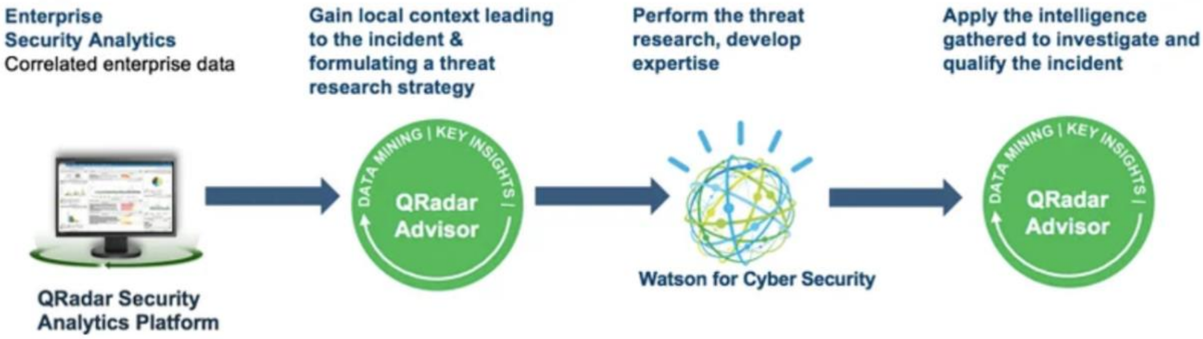
(🔒 = all requests are encrypted via HTTPS)

1. User prompts from security products are sent to Copilot
2. Copilot accesses plugins for pre-processing
3. Copilot sends modified prompt to LLM
4. Copilot receives LLM response
5. Copilot accesses plugins for post-processing
6. Copilot sends the response, and app command back to security products

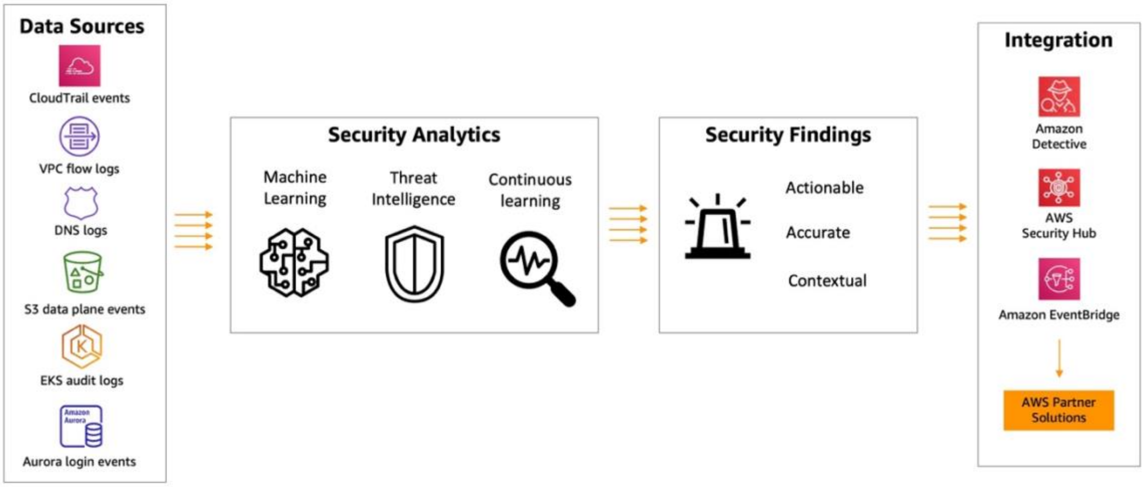
- **Project: SentinelOne's Singularity Platform** *Example: SentinelOne's Singularity platform uses AI to help security teams hunt for threats across the entire network. It pulls together data from all parts of your system, giving a clear view so that threats can be found and dealt with quickly.*



- **Project: IBM QRadar and Watson for Cyber Security** *Example: IBM's QRadar security system uses Watson's AI to automatically analyze security data. This helps security teams quickly spot and respond to potential threats, saving time and effort.*

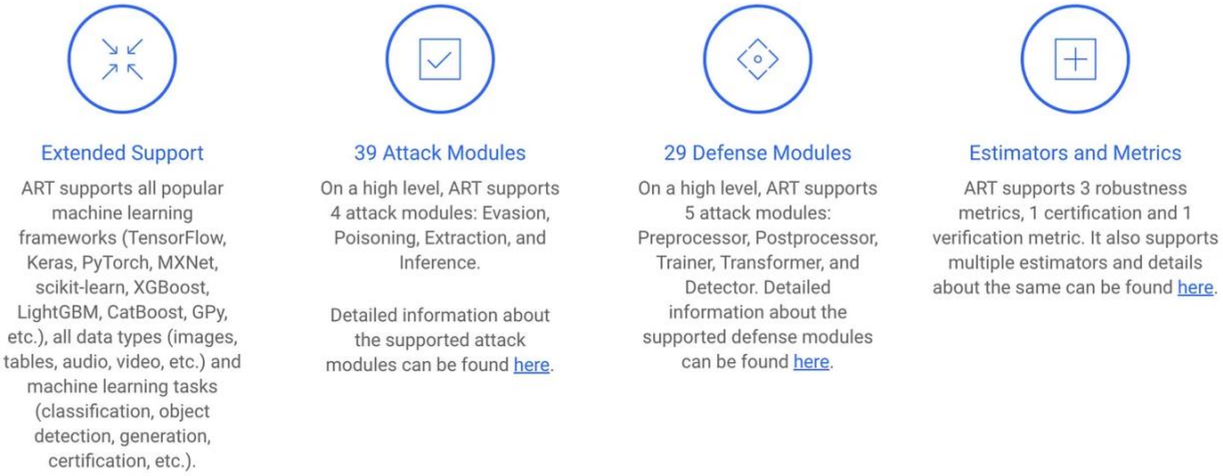

Enterprise Security Analytics
Correlated enterprise data
Gain local context leading to the incident & formulating a threat research strategy
Perform the threat research, develop expertise
Apply the intelligence gathered to investigate and qualify the incident
DATA MINING | KEY INSIGHTS |
QRadar Advisor
Watson for Cyber Security
DATA MINING | KEY INSIGHTS |
QRadar Advisor
QRadar Security Analytics Platform

- **Project: AWS GuardDuty** *Example: AWS GuardDuty uses AI to monitor and analyze data from AWS services to detect unusual activities. It identifies potential security breaches, such as abnormal API calls or unauthorized access attempts, enhancing threat detection.*



2) AI for Vulnerability Management and Security Enhancement

- **Project: IBM's Adversarial Robustness Toolbox (ART)** *Example: IBM's ART is a toolkit designed to protect AI systems from attacks. It helps make AI models more secure by defending against threats like data poisoning or hacking attempts that target AI weaknesses.*



Extended Support

ART supports all popular machine learning frameworks (TensorFlow, Keras, PyTorch, MXNet, scikit-learn, XGBoost, LightGBM, CatBoost, GPy, etc.), all data types (images, tables, audio, video, etc.) and machine learning tasks (classification, object detection, generation, certification, etc.).

39 Attack Modules

On a high level, ART supports 4 attack modules: Evasion, Poisoning, Extraction, and Inference.

Detailed information about the supported attack modules can be found here.

29 Defense Modules

On a high level, ART supports 5 attack modules: Preprocessor, Postprocessor, Trainer, Transformer, and Detector. Detailed information about the supported defense modules can be found here.

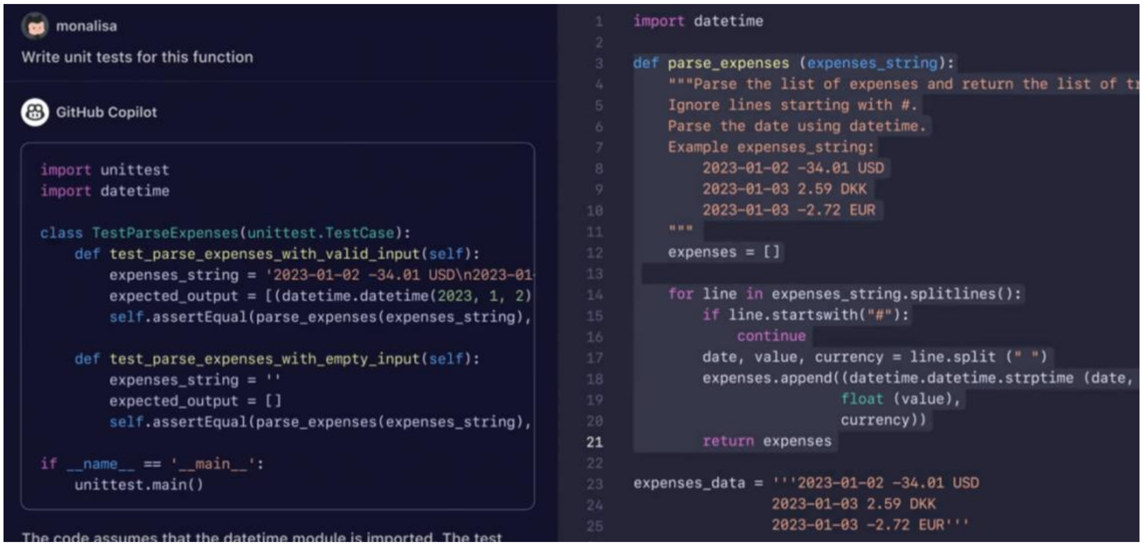
Estimators and Metrics

ART supports 3 robustness metrics, 1 certification and 1 verification metric. It also supports multiple estimators and details about the same can be found here.

- **Project: TensorFlow Privacy** *Example: TensorFlow Privacy is an extension of the TensorFlow platform that ensures AI models are trained with privacy protections. This makes the AI systems less vulnerable to being exploited or manipulated by attackers.*
- **Project: OWASP ZAP with Machine Learning** *Example: OWASP ZAP is a tool that tests web applications for security flaws. By adding machine learning, it can automate the process and become more effective at finding issues like SQL injections or cross-site scripting (XSS).*

- **Project: HackerOne with AI-Enhanced Scanning** *Example: HackerOne is a platform where security researchers find bugs in software. It uses AI to help automate the search for vulnerabilities, making it faster and easier to discover weaknesses in applications.*

- **Project: GitHub Copilot and CodeQL** *Example: GitHub Copilot is an AI tool that helps developers write code. When used with CodeQL, it also scans for security issues in the code as it's being written, helping catch problems early before they become serious.*



monalisa

Write unit tests for this function

GitHub Copilot

```python
import unittest
import datetime

class TestParseExpenses(unittest.TestCase):
    def test_parse_expenses_with_valid_input(self):
        expenses_string = '2023-01-02 -34.01 USD\n2023-01-
        expected_output = [(datetime.datetime(2023, 1, 2)
        self.assertEqual(parse_expenses(expenses_string),

    def test_parse_expenses_with_empty_input(self):
        expenses_string = ''
        expected_output = []
        self.assertEqual(parse_expenses(expenses_string),

if __name__ == '__main__':
    unittest.main()
```

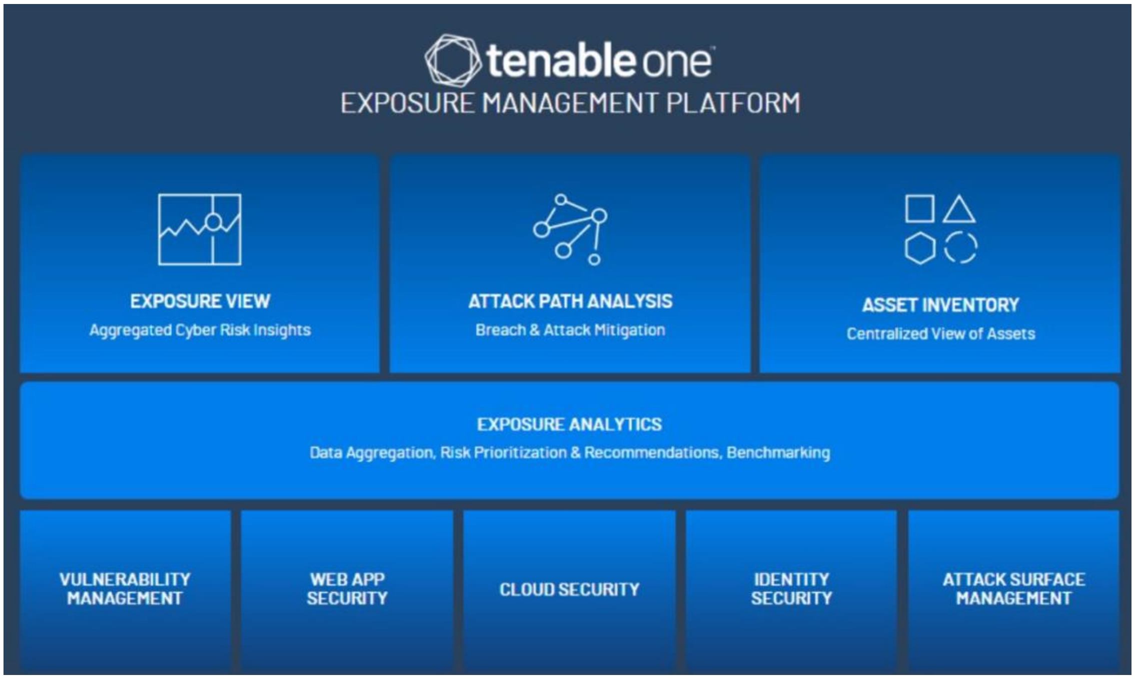
The code assumes that the datetime module is imported. The test

```python
import datetime

def parse_expenses (expenses_string):
    """Parse the list of expenses and return the list of tr
    Ignore lines starting with #.
    Parse the date using datetime.
    Example expenses_string:
        2023-01-02 -34.01 USD
        2023-01-03 2.59 DKK
        2023-01-03 -2.72 EUR
    """
    expenses = []

    for line in expenses_string.splitlines():
        if line.startswith("#"):
            continue
        date, value, currency = line.split (" ")
        expenses.append((datetime.datetime.strptime (date,
                        float (value),
                        currency))
        return expenses

expenses_data = '''2023-01-02 -34.01 USD
                2023-01-03 2.59 DKK
                2023-01-03 -2.72 EUR'''
```

- **Project: SonarQube with AI Plugins** *Example: SonarQube is a tool for checking the quality and security of code. With AI plugins, it gets even better at finding vulnerabilities and reduces the number of false alarms, making it easier for developers to fix real issues.*

- **Project: IBM's Guardium for Patch Management** *Example: IBM's Guardium uses AI to keep your data safe by finding and fixing vulnerabilities in both on-premises and cloud environments. It adapts to new threats and helps ensure that your data is protected by applying necessary patches on time.*

- **Project: Tenable's Exposure AI** *Example: Tenable's Exposure AI helps identify and fix weak spots in your system before hackers can exploit them. It uses AI to scan for vulnerabilities, prioritize which ones are most critical, and ensure they are patched quickly, making your system more secure.*



tenable one

EXPOSURE MANAGEMENT PLATFORM

EXPOSURE VIEW
Aggregated Cyber Risk Insights

ATTACK PATH ANALYSIS
Breach & Attack Mitigation

ASSET INVENTORY
Centralized View of Assets

EXPOSURE ANALYTICS
Data Aggregation, Risk Prioritization & Recommendations, Benchmarking

VULNERABILITY MANAGEMENT

WEB APP SECURITY

CLOUD SECURITY

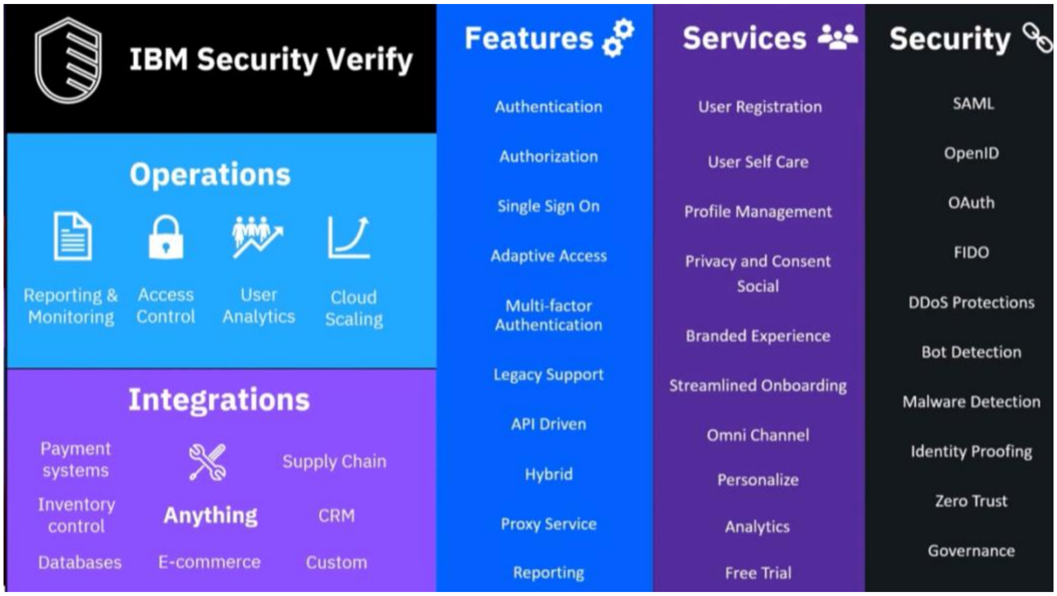
IDENTITY SECURITY

ATTACK SURFACE MANAGEMENT

- **Project: Zscaler Data Protection** *Example: Zscaler Data Protection uses AI to safeguard sensitive information by scanning documents, emails, and images. It classifies data, detects unauthorized access, and prevents data breaches, making it easier to manage and protect sensitive information.*

3) AI for Identity and Access Management

- **Project: IBM Verify** *Example: IBM Verify uses AI to enhance identity and access management by analyzing user behavior and adjusting authentication requirements. It can detect anomalies and automatically enforce multi-factor authentication, improving security while simplifying user access management.*



IBM Security Verify

Operations

Reporting & Monitoring

Access Control

User Analytics

Cloud Scaling

Integrations

Payment systems

Supply Chain

Inventory control

Anything

CRM

Databases

E-commerce

Custom

Features

Authentication

Authorization

Single Sign On

Adaptive Access

Multi-factor Authentication

Legacy Support

API Driven

Hybrid

Proxy Service

Reporting

Services

User Registration

User Self Care

Profile Management

Privacy and Consent Social

Branded Experience

Streamlined Onboarding

Omni Channel

Personalize

Analytics

Free Trial

Security

SAML

OpenID

OAuth

FIDO

DDoS Protections

Bot Detection

Malware Detection

Identity Proofing

Zero Trust

Governance

References

- IBM. (n.d.). *Threat detection and response*. Retrieved September 2, 2024, from https://www.ibm.com/services/threat-detection-response
- IBM. (n.d.). *Guardium*. Retrieved September 2, 2024, from https://www.ibm.com/guardium
- IBM. (n.d.). *Watsonx governance*. Retrieved September 2, 2024, from https://www.ibm.com/products/watsonx-governance
- Watkins, O. (2024, April 19). *4 use cases of AI in cyber security*. Red Hat. Retrieved September 2, 2024, from https://www.redhat.com/en/blog/4-use-cases-ai-cyber-security
- Goss, A. (2024, May 13). *Examples of AI in cyber security*. Station X. Retrieved September 2, 2024, from https://www.stationx.net/examples-of-ai-in-cyber-security/
- Shutenko, V. (2024, August 8). *AI in cybersecurity*. TechMagic. Retrieved September 2, 2024, from https://www.techmagic.co/blog/ai-in-cybersecurity/
- (n.d.). *Artificial intelligence in cyber security projects*. Network Simulation Tools. Retrieved September 2, 2024, from https://networksimulationtools.com/artificial-intelligence-in-cyber-security-projects/
- Daivi. (2024, March 19). *Cybersecurity machine learning projects*. ProjectPro. Retrieved September 2, 2024, from https://www.projectpro.io/article/cybersecurity-machine-learning-projects/631
- Caniszczyk. (n.d.). *AnomalyDetection*. GitHub. Retrieved September 2, 2024, from https://github.com/twitter/AnomalyDetection
- Fier, J., & Kenyon Grant, S. (n.d.). *Splunk machine learning toolkit*. Splunk. Retrieved September 2, 2024, from https://www.splunk.com/en_us/resources/splunk-machine-learning-toolkit.html
- Darktrace. (2022, April 12). *How Darktrace's Cyber AI Analyst accelerates reporting incidents to the US federal government*. Retrieved September 2, 2024, from https://darktrace.com/blog/how-darktraces-cyber-ai-analyst-accelerates-reporting-incidents-to-the-us-federal-government
- Microsoft. (2024, July 18). *Microsoft security copilot*. Retrieved September 2, 2024, from https://learn.microsoft.com/en-us/copilot/security/microsoft-security-copilot
- (n.d.). *Singularity platform*. Kidan. Retrieved September 2, 2024, from https://kidan.co/partners/singularity-platform/
- (n.d.). *Stages involved in the working of IBM’s QRadar advisor with Watson*. ResearchGate. Retrieved September 2, 2024, from https://www.researchgate.net/figure/Stages-involved-in-the-working-of-IBMs-QRadar-advisor-with-Watson-64_fig3_373712758
- Amazon Web Services. (n.d.). *Amazon GuardDuty now supports Amazon EKS runtime monitoring*. Retrieved September 2, 2024, from https://aws.amazon.com/tr/blogs/aws/amazon-guardduty-now-supports-amazon-eks-runtime-monitoring/
- Hajra, A. (2023, May 4). *IBM ART adversarial robustness check for a movie recommendation system*. Medium. Retrieved September 2, 2024, from

https://medium.com/@asmitahajra/ibm-art-adversarial-robustness-check-for-a-movie-recommendation-system-649ba46e9e8a

- (n.d.). *Copilot*. GitHub. Retrieved September 2, 2024, from https://github.com/features/copilot
- (n.d.). *Exposure management: What is Tenable One?* E-SPIN. Retrieved September 2, 2024, from https://www.e-spincorp.com/exposure-management-what-is-tenable-one/
- (n.d.). *IBM Verify reviews*. TrustRadius. Retrieved September 2, 2024, from https://www.trustradius.com/products/ibm-verify/reviews?qs=pros-and-cons#comparisons