

Utilizando Inteligência Artificial para Atacar Aplicações WEB

Glaudson Ocampos
<gocampos@conviso.com.br>
<glaudson.ml@gmail.com>

Agenda

- **Apresentação**
- **Inteligência Artificial nos Séculos 20 e 21**
 - **Um pouco de história**
 - **IA em Segurança da Informação**
- **Web Hacking no Século 21**
 - **O que é Web Hacking**
- **Utilizando IA para Ataques WEB**
 - **Reconhecimento e Classificação (Deep Learning/NLP)**
 - **Modularização Inteligente (MultiAgent System)**
 - **Tomada de Decisão (POMDP)**
 - **Kurgan Framework**
- **Trabalhos Futuros**
- **Conclusão**
- **Referências**



Apresentação

- **Profissional de Segurança da Informação com mais de 15 anos em práticas ofensivas e defensivas;**
- **Desenvolvedor de Soluções de Segurança(IDS/WAF/etc);**
- **Descobridor de falhas e criador de exploits para as mais diversas aplicações;**
- **Trabalhando como Analista de Segurança Sênior da Conviso Application Security;**
- **Pesquisador Autônomo de Hacking/Security com ênfase no uso de IA e Criptoanálise;**
- **Contatos:**
 - **Email: gocampos@conviso.com.br / glaudson.ml@gmail.com**
 - **Twitter: [@nashleon2](https://twitter.com/nashleon2)**



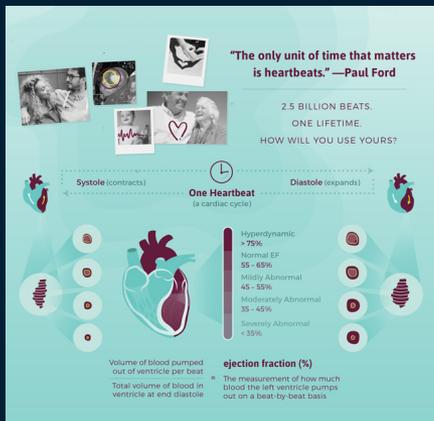
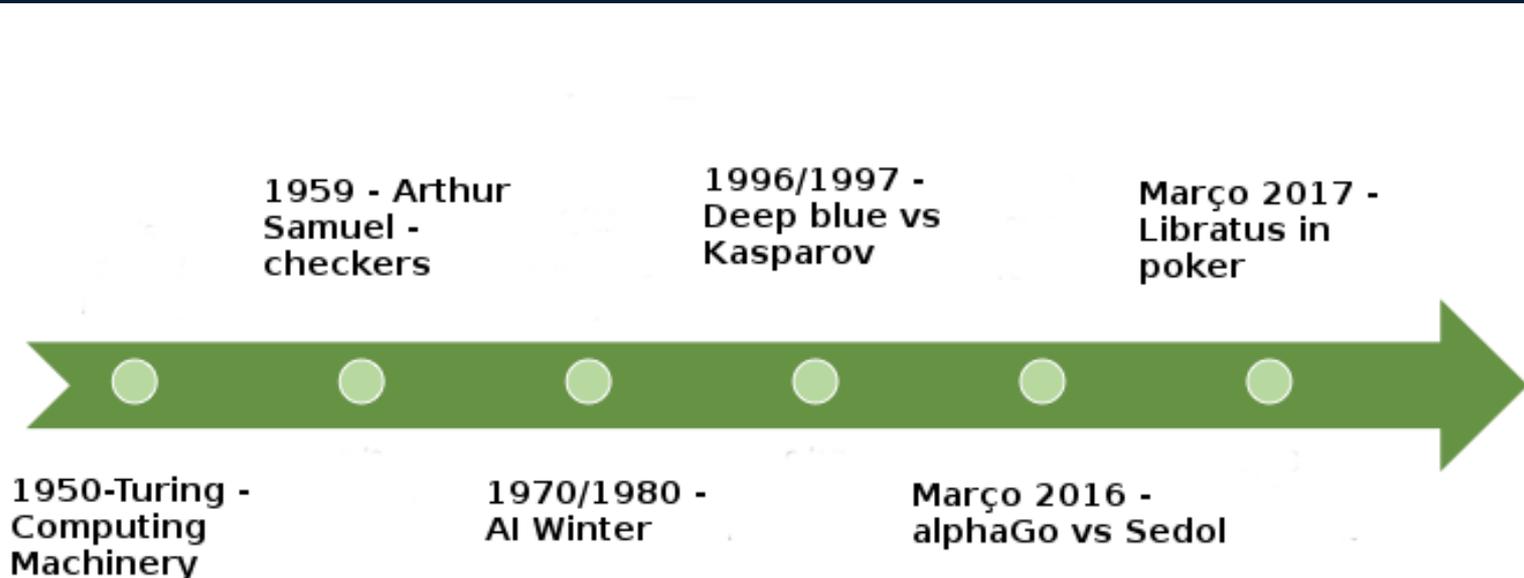
Disclaimer

- **This is work in progress research**
- **Tooling is still in development(prototype only)**
- **Presentation will discuss some aspects only**



Inteligência Artificial nos Séculos 20 e 21

Um Pouco de História



The Allen AI Science Challenge
Is your model smarter than an 8th grader?
\$80,000 / 170 teams / a year ago

Overview | Data | Discussion | Leaderboard | More

Description | Evaluation | Prizes | Timeline

The Allen Institute for Artificial Intelligence (AII) is working to improve humanity through fundamental advances in artificial intelligence. One critical but challenging problem in AI is to demonstrate the ability to consistently understand and correctly answer general questions about the world.

The Aristo project at AII is focused on building such a system. One way Aristo "learns" is by extracting facts from various sources and processing them into a structured knowledge base. When taking an exam, questions are parsed and processed along with any accompanying diagrams to determine a strategy for answering. Aristo then uses entailment, statistical analysis, and inference methods to select a final answer.

While Aristo's abilities have improved significantly in the last two years, it still doesn't have perfect, reliable methods of gathering knowledge, understanding questions, or reasoning through answers.

Using a dataset of multiple choice question and answers from a standardized 8th grade science exam, AII is challenging you to create a model that gets to the head of the class.

Microsoft Malware Classification Challenge (BIG 2015)
Classify malware into families based on file content and characteristics
\$16,000 / 377 teams / 2 years ago

Overview | Data | Discussion | Leaderboard | Rules

Description
In recent years, the malware industry has become a well organized market involving large amounts of money. Well funded, multi-player syndicates invest heavily in technologies and capabilities built to evade traditional protection, requiring anti-malware vendors to develop counter mechanisms for finding and detecting them. In the meantime, they inflict real financial and emotional pain to users of computer systems.

Evaluation
The major challenges that anti-malware faces today is the vast amounts of data and files which need to be evaluated for potential malicious intent. For example, Microsoft's real-time detection anti-malware products are present on over 150M computers worldwide and inspect over 700M computers monthly. This generates tens of millions of daily data points to be analyzed as potential malware. One of the main reasons for these high volumes of different files is the fact that, in order to evade detection, malware authors introduce polymorphism to the malicious components. This means that malicious files belonging to the same malware "family", with the same forms of malicious behavior, are constantly modified and/or obfuscated using various tactics, such that they look like many different files.

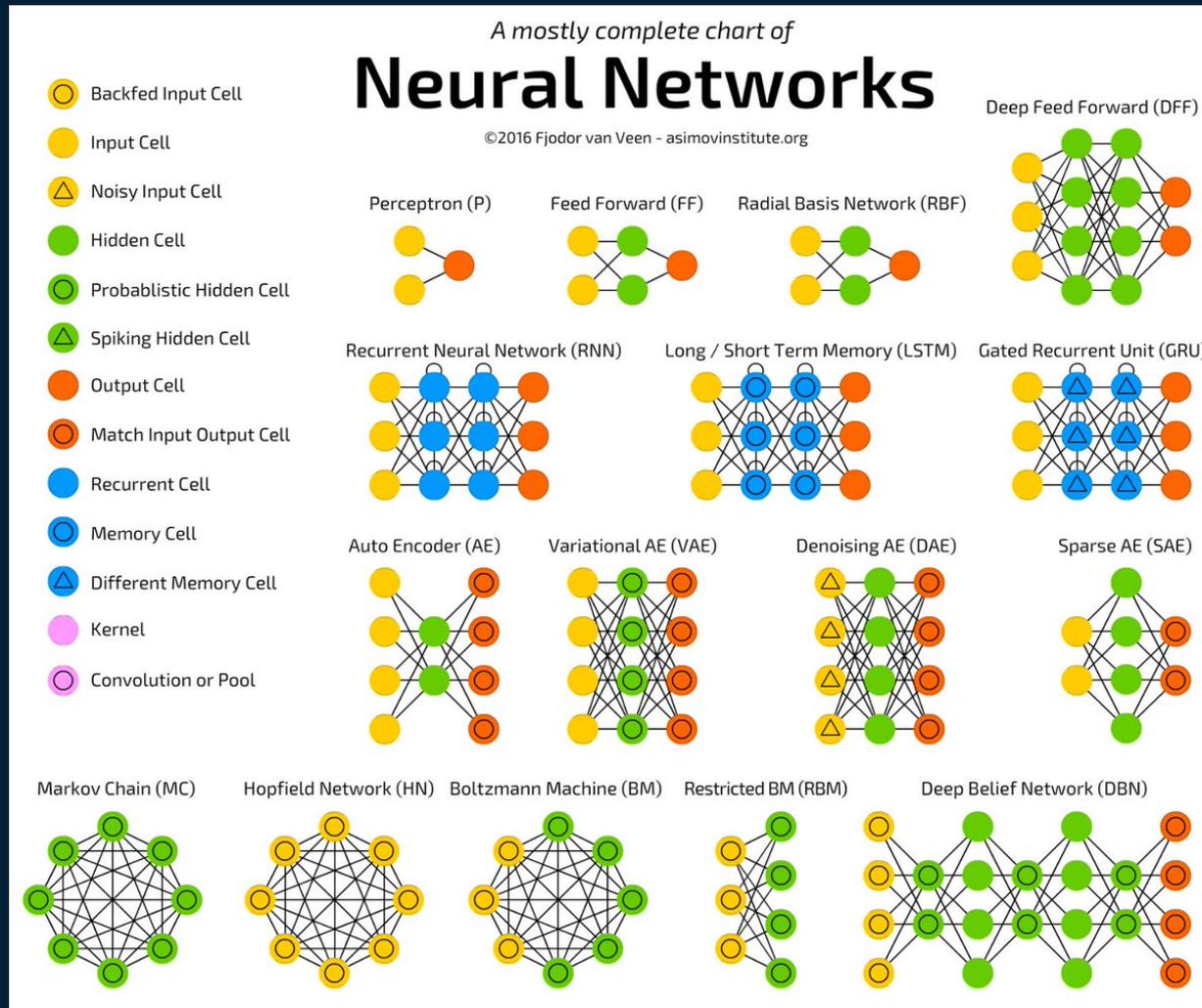
Prizes
In order to be effective in analyzing and classifying such large amounts of files, we need to be able to group them into groups and identify their respective families. In addition, such grouping criteria may be applied to new files encountered on computers in order to detect them as malicious and of a certain family.

Timeline



Inteligência Artificial nos Séculos 20 e 21

Inteligência Artificial Moderna



Inteligência Artificial nos Séculos 20 e 21

IA em Segurança da Informação

■ Segurança defensiva:

- Detectores de Anomalias em Protocolos de Rede;
- Detectores de Malwares em PDF usando ML;
- Recursos de IA em WAFs;
- Machine Learning em Anti-Malwares(Cylance, MS e Kaggle);

■ Segurança ofensiva:

- Saraute et al - OS Fingerprint usando Neural Network[1];
- Saraute et al - Penetration Testing == POMDP Solving?[2];
- American Fuzzy Lop - Fuzzer com Algoritmos Genéticos[5];



Web Hacking no Século 21

- Aplicações WEB ainda são grandes alvos de ataques;
- Ambiente WEB é complexo e heterogêneo;
- Estamos em 2017 e ainda não há cultura de Desenvolvimento Seguro – o que salva as aplicações são algumas Frameworks;
- Exemplos de Invasões recentes (Equifax , Yahoo):

- *Equifax* - Vulnerabilidade no Struts (+ outras coisas):

<https://medium.com/@thegrugq/equifax-fact-enabled-wild-speculation-21fd59aa39e2>

- *Yahoo* – Phishing seguido de mal gerenciamento de cookies:

<https://arstechnica.com/tech-policy/2017/03/fbi-hints-that-hack-of-semi-privileged-yahoo-employee-led-to-massive-breach/>



Web Hacking no Século 21

O que é o Web Hacking?

- Conjunto de técnicas utilizadas para atacar aplicações WEB;
- Várias técnicas conhecidas :

Técnicas de Ataques
Crawling / Spider
XSS / XSRF /SSRF /Dom XSS
Brute Force Login / Authentication Token
Injection(SQL, LDAP/XPATH, NOSQL, Command, etc)
LFI / RFI / Insecure File Upload
Deserialization / Reflections / EL Injection
App Specific Vulnerabilities (Struts, Apache Commons, Tomcat)



Utilizando IA para Ataques WEB

Por que usar IA para Web Hacking?

- **Necessidade de automatizar ataques;**
- **Ferramentas atuais geram muito falsos-positivos e falsos-negativos;**
- **Sistemas autenticados exigem manuseiam de cookies, sessões, event viewers, execução de javascript, etc;**
- **Kasparov está certo!!! Sistemas inteligentes podem auxiliar humanos!**



Utilizando IA para Ataques WEB

Reconhecimento e Classificação (Deep Learning/NLP)

Definições – Wikipedia:

"Pattern recognition is a branch of machine learning that focuses on the recognition of patterns and regularities in data, although it is in some cases considered to be nearly synonymous with machine learning.[1] Pattern recognition systems are in many cases trained from labeled "training" data (supervised learning), but when no labeled data are available other algorithms can be used to discover previously unknown patterns (unsupervised learning)."

"Classification is a general process related to categorization, the process in which ideas and objects are recognized, differentiated, and understood."



Utilizando IA para Ataques WEB

(Exemplos de Processos que podem ser

AClassificados ou Reconhecidos)

O que Reconhecer/Classificar	IA a ser utilizada
Identificar Página de Autenticação	<ul style="list-style-type: none">- Visão Computacional(DL);- Classificação com NLP;
Ataque de Força Bruta em Login	<ul style="list-style-type: none">- Information Retrieve + Rank;- Pattern Match com DL/NLP;
Spider / Crawling / Fingerprint	<ul style="list-style-type: none">- Busca Heurística(DFS/BFS);- Classificação com NLP;
Extração de parâmetros de Recurso	<ul style="list-style-type: none">- Rede Neural;- NLP (Compreensão de Contexto);
Determinar recursos mais interessantes de serem atacados	<ul style="list-style-type: none">- Classificação com Rede Bayesiana;- Decision Tree;



Utilizando IA para Ataques WEB

Reconhecimento e Classificação (Deep Learning/NLP)

DEMONSTRAÇÃO

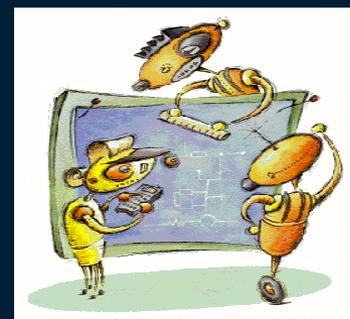


Utilizando IA para Ataques WEB

Modularização Inteligente

A(MultiAgent System)

- Agentes atualmente se comunicam usando FIPA-ACL(Trocar por JSON?);
- Protocolo para Rede Stomp (Apache Apollo);
- Agentes podem ser desenvolvidos em qualquer linguagem;
- Inicialmente trabalhando em modo cooperativo;
- Usar Browser Headless;



Utilizando IA para Ataques WEB

**Modularização Inteligente
(MultiAgent System)**

DEMONSTRAÇÃO



Utilizando IA para Ataques WEB

Tomada de Decisão (POMDP)

■ Definição - Wikipedia:

"Decision theory (or the theory of choice) is the study of the reasoning underlying an agent's choices."

ou

"A teoria da decisão (ou a teoria da escolha) é o estudo do raciocínio subjacente às escolhas de um agente."

■ Decision Theory x Game Theory



Utilizando IA para Ataques WEB

Tomada de Decisão

Aonde aplicar?

■ Ações baseadas em fases de um pentest WEB:

Fases/Ações de um Pentest WEB
Descobrir/Reconhecer InfraEstrutura (Servidor WEB/OS/Frameworks)
Detectar Sistemas de Segurança (IPS, WAF, Filtro, etc)
Verificar vulnerabilidades públicas(exploit-db, CVE, NVD, etc)
Executar Ataques Comuns (XSS, LFI, SQLi, etc)
Executar Ataques de Força Bruta (Login, SessionID, etc)
Executar Spiders/Crawler/Backup Search
Executar Fuzzer



Utilizando IA para Ataques WEB

Tomada de Decisão

Como aplicar?

- Para cada ação, um agente interage uma POMDP com o Agente Master
- POMDP - Partially Observable MDP:
 - Conceito Avançado e bem estabelecido;
 - Baseado em:
 - Markov Decision Process (MDP);
 - POMDP – Partially Observable;



Utilizando IA para Ataques WEB

Tomada de Decisão

A(DEC-POMDP)

A Dec-POMDP can be described as follows:

$$M = \langle I, S, \{A_i\}, P, R, \{\Omega_i\}, O, h \rangle$$

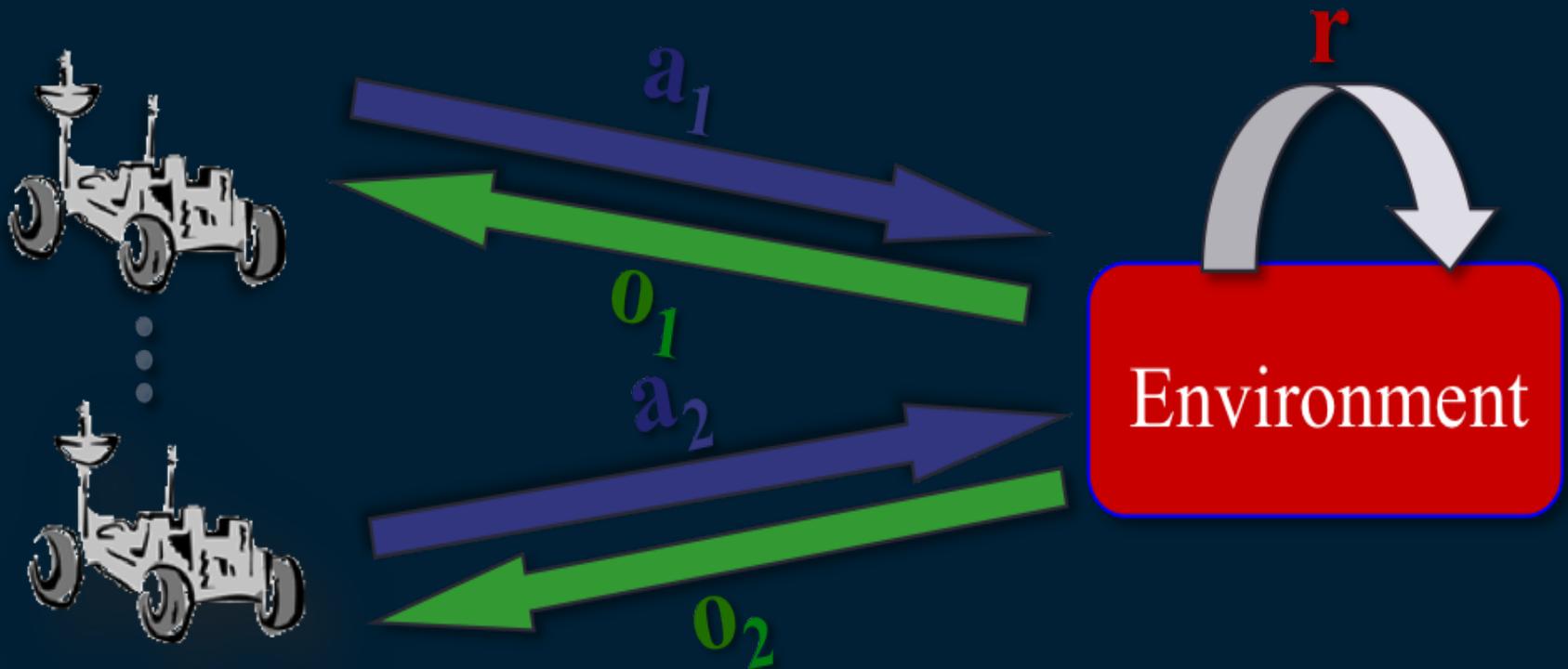
- I , the set of agents
- S , the set of states with initial state s_0
- A_i , the set of actions for agent i , with $A = \times_i A_i$ the set of joint actions
- P , the state transition probabilities: $P(s'|s, a)$, the probability of the environment transitioning to state s' given it was in state s and agents took actions a
- R , the global reward function: $R(s, a)$, the immediate reward the system receives for being in state s and agents taking actions a
- Ω_i , the set of observations for agent i , with $\Omega = \times_i \Omega_i$ the set of joint observations
- O , the observation probabilities: $O(o|s, a)$, the probability of agents seeing observations o , given the state is s and agents take actions a
- h , the horizon, whether infinite or if finite, a positive integer
- when h is infinite a discount factor, $0 \leq \gamma < 1$, is used



Utilizando IA para Ataques WEB

Tomada de Decisão

A(DEC-POMDP)



Utilizando IA para Ataques WEB

Tomada de Decisão
(DEC-POMDP)

DEMONSTRAÇÃO



Utilizando IA para Ataques WEB

KURGAN FRAMEWORK

Kurgan AI - Web Application Security Analyzer

Dashboard Users Configurations Statistics Reports Agents Help Logout

admin [View More](#)

Last Signed In : 2017-01-13 11:34:58
IP Address : 192.168.2.33

System

CPU Usage	0.01%
Disk Usage	27%
Memory Usage	11%

Kurgan Updates [Check](#)

You are running the latest free version.
[Upgrade Now.](#)

Vulnerabilities [View More](#)

1 SQL Injection	More Information
1 XSRF	View More
1 XSS	0 Pending
0 LFI/RFI	0 Pending
1 Others	View More

Latest Scan [More Info.](#)

[http://www.target.com.br/](#)

Quick Scan [Advanced](#)

Host Target (Domain , IP or URL.)

[Scan](#)

Statistics [View More](#)

Critical	+ 120%
Warning	+ 220%
Low	- 10%

Agents Info [View More](#)

Messages from Multi-Agents

16:10 - MasterAgent ready to receive url.
16:09 - Finished Analysis of www.vortex-ai.com.br.

16:09 - Finished Analysis of www.vortex-ai.com.br.

- 16:09 - Finished Analysis of www.vortex-ai.com.br.
- 14:45 - Running Crawling against www.vortex-ai.com.br.
- 14:40 - Running Brute Force Login against www.vortex-ai.com.br.

Copyright © 2017. Vortex AI

<https://github.com/glaudsonml/kurgan-ai/>



Utilizando IA para Ataques WEB

TRABALHOS FUTUROS

- Implementar Reinforcement Learning;
- Recurso de Fuzzer Inteligente;
- Ambiente de Simulação;
- Compreensão de Linguagem Natural (NLP);
- Disputar CTFs;
- Automaticamente Orientar/Desenvolver soluções para os problemas de segurança encontrados;



Conclusão

- IA vai se inserir cada vez mais no ramo da Segurança da Informação;
- Necessidade de profissionais que dominem ambos os mundos(IA/Information Security);
- Novos paradigmas surgirão a medida que a máquina melhorar como atacante/defensor;
- Singularidade tecnológica deve ser levada em conta;



Referências

- [1] - <https://www.coresecurity.com/corelabs-research/projects/using-neural-networks-os-fingerprinting>
- [2] - <https://arxiv.org/abs/1306.4714>
- [3] - <http://lcamtuf.coredump.cx/afl/>
- Livro - A Concise Introduction to Decentralized POMDPs - Frans A. Oliehoek e Christopher Amato;
- Livro - Decision Making Under Uncertainty - Theory and Application - MIT Lincoln Laboratory Series;
- Livro - Deep Learning - Ian Goodfellow;
- Livro - Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow - Sebastian Raschka e Vahid Mirjalili;
- Livro - Reinforcement Learning: State-of-the-Art (Adaptation, Learning, and Optimization) - Marco Wiering et al

