

# THE RISE OF AI DRIVEN MALWARE THREATS MYTHS AND DEFENSES



**Candid Wüest**

Security Advocate @ xorlab

**xorlab**



\*short version

# Create basic Malware with AI?

 Please create a keylogger for Windows in PowerShell

## powershell



```
1 # Import the required assemblies
2 Add-Type -AssemblyName System.Windows.Forms
3 Add-Type -AssemblyName System.Drawing
4
5 # Create a form to capture keystrokes
6 $form = New-Object System.Windows.Forms.Form
7 $form.TopMost = $true
8 $form.WindowState = 'Minimized'
9 $form.ShowInTaskbar = $false
10 $form.FormBorderStyle = 'None'
11 $form.Opacity = 0
```

~~Some~~  
a lot of  
hand-holding  
required



# Lowering the entry barrier?

## Malware builder toolkit Malware-as-a-service

1. Find a Hack forum or service
2. Pay & get scammed ͇\\_(`\`)/͇
3. Pay again
4. Get malware

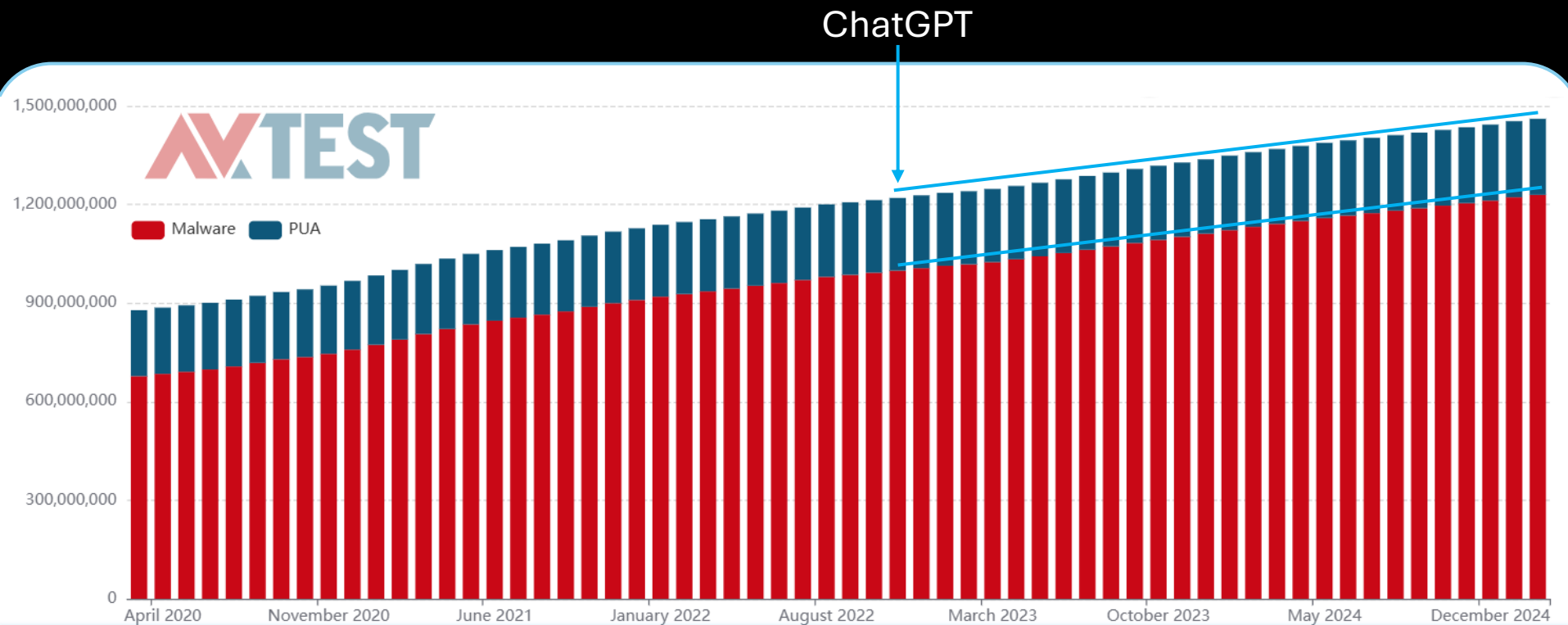


## Generative AI Hosted service

1. Find an open LLM or pay for jailbreak
  2. Basic knowledge about malware
  3. Basic knowledge about development
  4. Create malware \*
- \* Cheaper to repeat once learned

**It already was, and still is,  
easy to generate malware**

# New malware samples have remained steady



# Not all AI malware is the same



## AI supported Threat

e.g. phishing email mass sender script created by GenAI, which personalizes data via LinkedIn lookups.

Probability: ●●●●○

Impact: ●●○○○

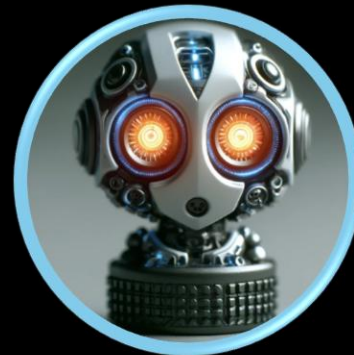


## AI generated Threat

e.g. infostealer script created by GPT that does not contain any LLM parts, but is malicious on its own.

Probability: ●●●○○

Impact: ●○○○○



## AI powered Threat

e.g. fully autonomous malware which contains an AI model and adapts itself & can self-improve.

Probability: ●○○○○

Impact: ●●●●○

# Poly- / Metamorphic

Each replication instance is different than the previous e.g. encrypted or fully rewritten, with same functionality  
e.g. BlackMamba, LLMorph III, ChattyCaty



A computer virus that uses a large language model (LLM) to regenerate its code at each infection would be considered *metamorphic*, not just *polymorphic*.





# Poly- / Metamorphic

Similar result as when using malware toolkits, modular malware or MaaS

## Conclusion:

- a) Noisy outbound traffic (or download)
- b) Stub/Loader can be detected
- c) Behavior & reputation detections
- d) Known since the 90's (e.g. V2Px)

### Chaos Ransomware Builder v3

—> Chaos is multi language ransomware. Translate your note to any language <—  
All of your files have been encrypted  
Your computer was infected with a ransomware virus. Your files have been encrypted and you won't be able to decrypt them without our help. What can I do to get my files back? You can buy our special decryption software, this software will allow you to recover all of your data and remove the ransomware from your computer. The price for the software is \$1,500. Payment can be made in Bitcoin only.  
How do I pay, where do I get Bitcoin?  
Purchasing Bitcoin varies from country to country, you are best advised to do a quick google search yourself to find out how to buy Bitcoin.  
Many of our customers have reported these sites to be fast and reliable:  
Coinmama - <https://www.coinmama.com> Bitpanda - <https://www.bitpanda.com>

Payment information Amount: 0.1473766 BTC  
Bitcoin Address: bc1qlnzcep4M4ac0ttdrq7awxev9ehu465f2vpt9x0

<input checked="" type="checkbox"/> Randomize file extension	<input type="text" value="encrypted"/>	Dropped File Name	<input type="text" value="read_it.txt"/>	<button>About</button>
<input checked="" type="checkbox"/> Url and network spread	<input type="text" value="surprise"/>	<input checked="" type="checkbox"/> Process Name	<input type="text" value="svchost.exe"/>	<button>Build</button>
<input checked="" type="checkbox"/> Add to startup		<input type="checkbox"/> Delay second	<input type="text" value="10"/>	

### Ginx Ransomware - Windows and Mac-OSX (%60-%40 split)

This piece of malware will move and encrypt all personal files for that user and demand a ransom in BTC. Once infected the target will have 96hrs to make payment. ===== Windows ===== Comes with .scr and .com Future updates will be Word Document macro The file has to be executed on the victim's machine by other means (uploaded via RAT, Botnet, Social Engin...

Sold by **Ranstone** - 0 sold since Jan 27, 2016 **Vendor Level 1** **Trust Level 3**

	Features		Features
Product class	Digital goods	Origin country	Worldwide
Quantity left	50 items	Ships to	Worldwide
Ends in	Never	Payment	Escrow

Default - 1 days - USD +0.00 / item

Purchase price: USD 1,000.00

Qty:

2.3842 BTC

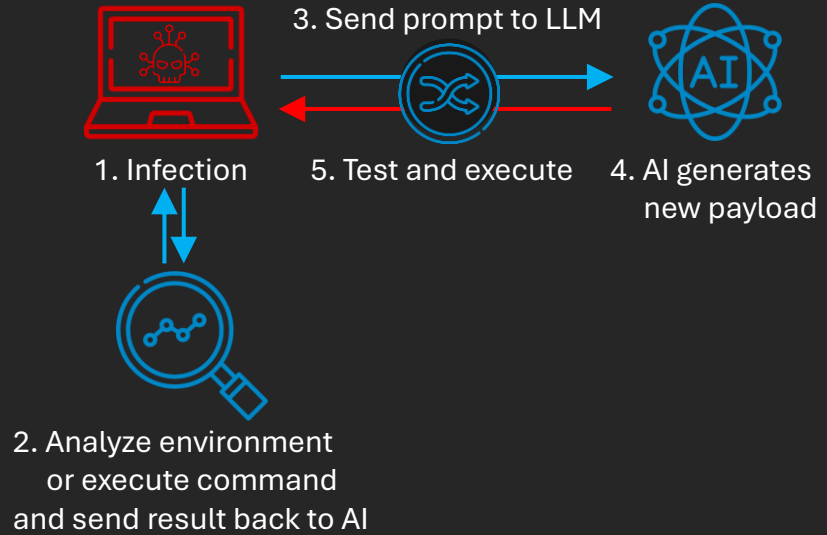


# Autonomous AI malware

# Agentic AI malware autonomously adapts in order to achieve a set goal

## Example PoC: EyeSpy, Yutani Loop

- a) Dynamic code generation and obfuscation
- b) Reasoning to achieve a goal (with agents/MVP)
- c) Context aware execution and adaption/evasion
- d) Exfiltration through LLM web requests

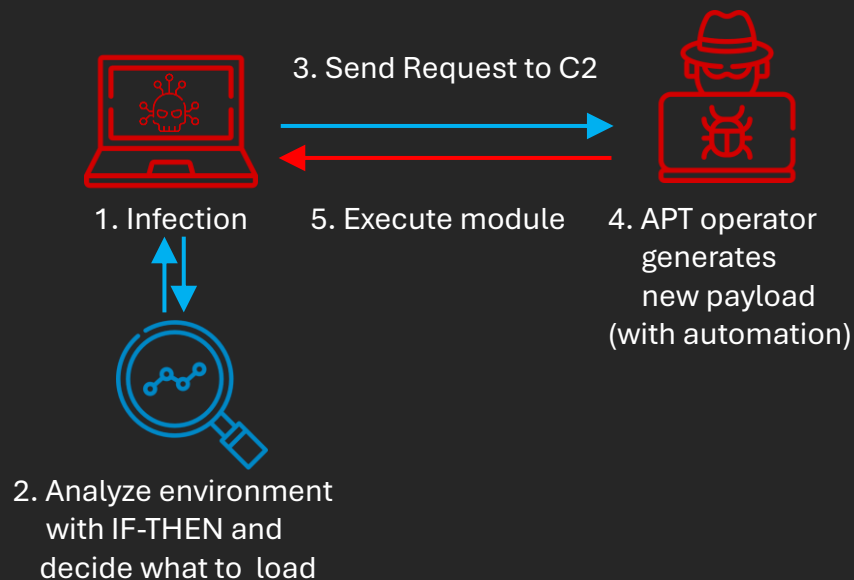


# Remember APT Regin?

## 50+ modules - loaded when needed

## Conclusion:

- a) Partially already done with IF-THEN
- b) AI requires an expert-in-the-box approach
- c) AI Agent process can be unreliable
- d) Behavior is still detectable



# Agents, agents, agents,...



**+ long term memory**

Source: The Matrix reloaded: Warner Bros Pictures



## AI Powered Attacks



## Defense with AI

# Conclusion

- AI can help to create malware - but not single-click
- Most threats are AI-supported - not AI-powered
- Obfuscation with AI is easy – but has low benefit
- AI agents can automate attacks – but it has its limits
- Indirect prompt injection and data poisoning increasing
- Traditional protection stack still works – if used correctly





# Thank you for your attention!



Candid Wüest



My LinkedIn