AV SECURITY SUMMIT
FRANKFURT AM MAIN | GERMANY | 2025

TLP:CLEAR
Contains AI

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

```powershell
# Import the required assemblies
Add-Type -AssemblyName System.Windows.Forms
Add-Type -AssemblyName System.Drawing

# Create a form to capture keystrokes
$form = New-Object System.Windows.Forms.Form
$form.TopMost = $true
$form.WindowState = 'Minimized'
$form.ShowInTaskbar = $false
$form.FormBorderStyle = 'None'
$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 ¯_(ツ)_/¯
3. Pay again
4. Get malware

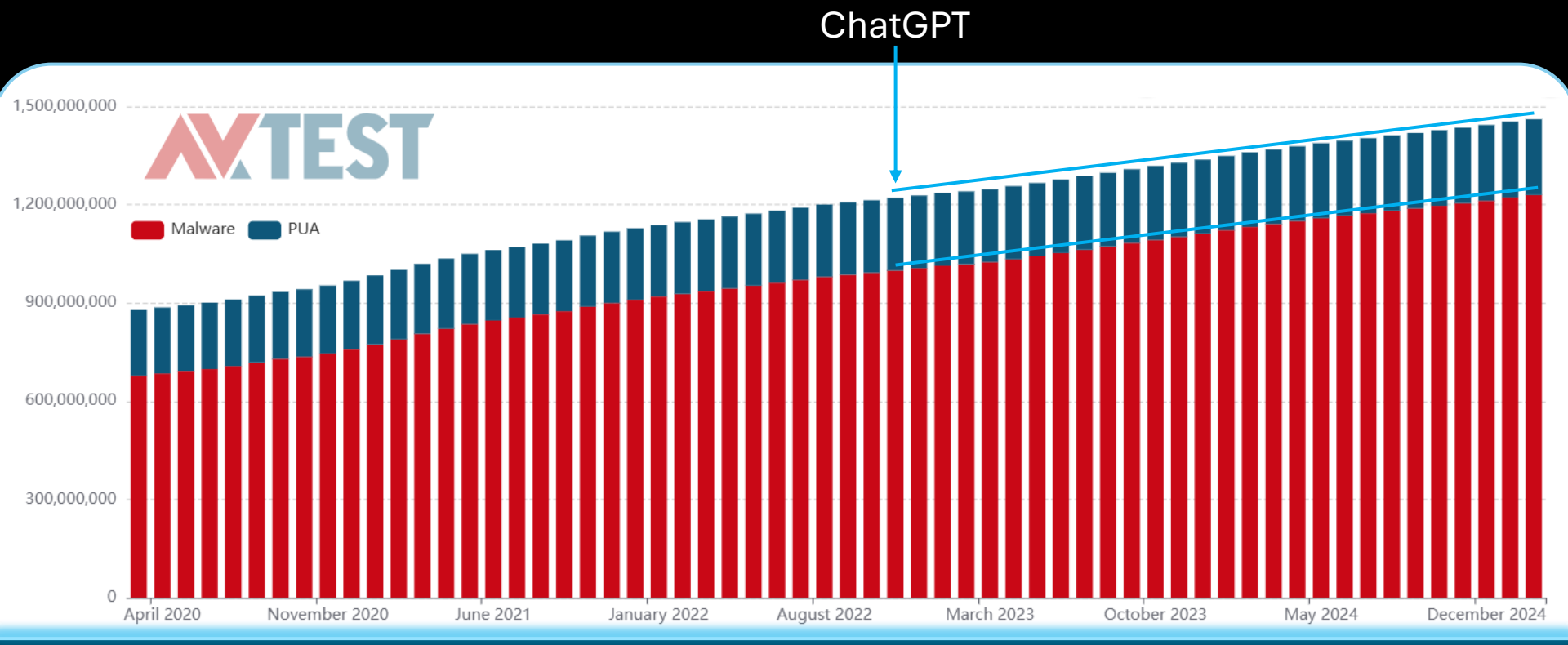
Generative AI
Hosted service

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

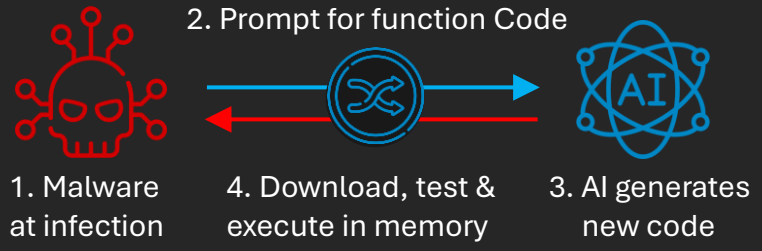
e.g. fully autonomous malware which contains an AI model and adapts itself & can self-improve.

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)

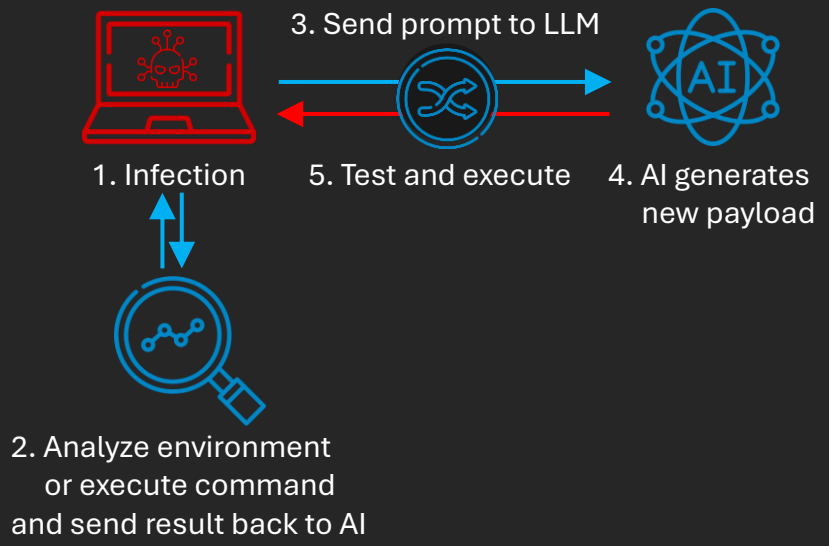
☺ Team work is important; it helps to put the blame on someone else.

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

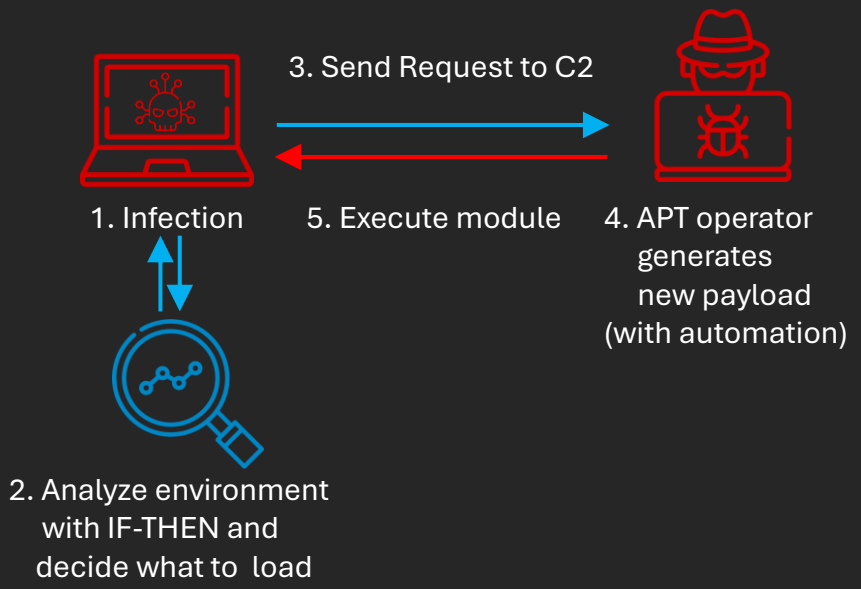
Conclusion:

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

☺ What happens if you get scared half to death twice?

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