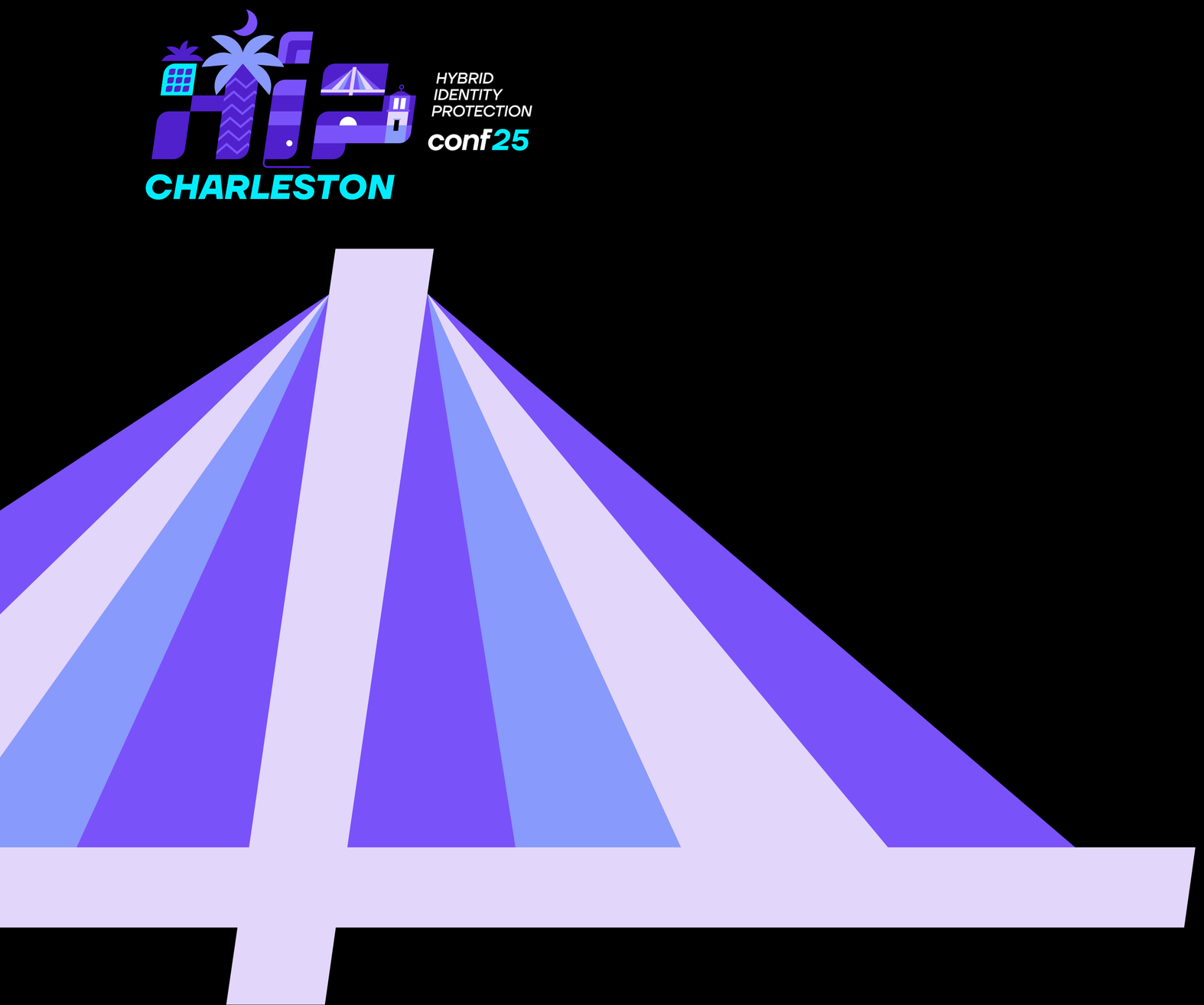


Agentic AI and the New Frontier of Identity Attacks

Ran Harel
AVP Security Products, Semperis

HYBRID IDENTITY PROTECTION conf25
CHARLESTON

Ran Harel

AVP Security Products, Semperis

- Air Traffic Controller (past life)
- Pen-tester in the 00s
- CISO until 2013
- 3-time entrepreneur
- Owner of 2 teenage dogs
- Father of 2 teenage girls

HYBRID IDENTITY PROTECTION conf25
CHARLESTON

Identity Attacks - Evolution

1990s – 2000s
Perimeter -> Credentials
From Firewalls & IDS evasion
to stolen passwords and phishing

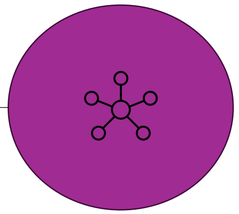
Today -> Future
Autonomous AI-Driven Attacks
Agentic AI enables adaptive,
multi-stage identity breaches

2010s - today
Cloud & Identity Exploits
0Auth abuse, SSO/AD misconfigurations,
MFA fatigue
“Identity is the new perimeter”

HYBRID IDENTITY PROTECTION conf25 CHARLESTON

Agentic AI – GPT with a Purpose

Language:
LLM (Large Language Model)

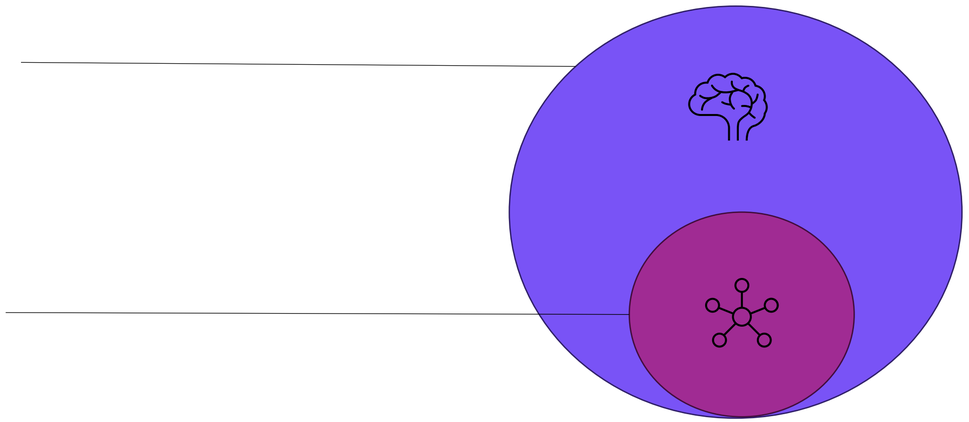


HYBRID IDENTITY PROTECTION conf25 CHARLESTON

Agentic AI – GPT with a Purpose

Knowledge:
RAG (Retrieval Augmented Generation)

Language:
LLM (Large Language Model)



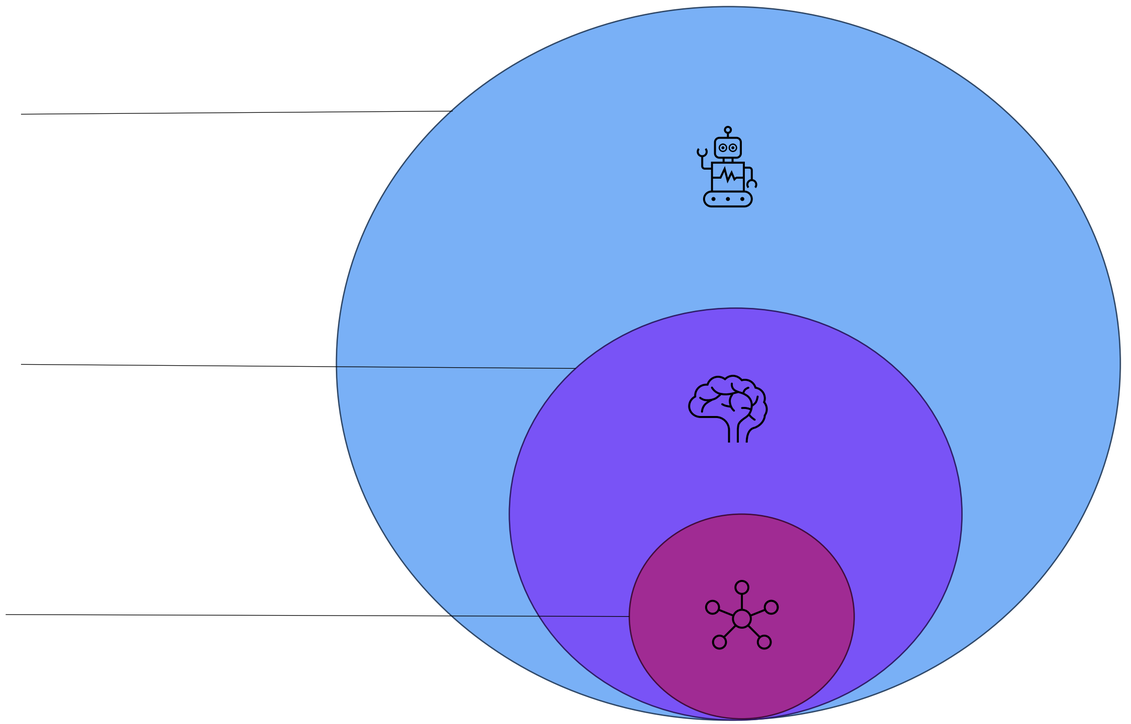
HYBRID IDENTITY PROTECTION conf25 CHARLESTON

Agentic AI – GPT with a Purpose

Agency:
AI Agents

Knowledge:
RAG (Retrieval Augmented Generation)

Language:
LLM (Large Language Model)



HYBRID IDENTITY PROTECTION conf25 CHARLESTON

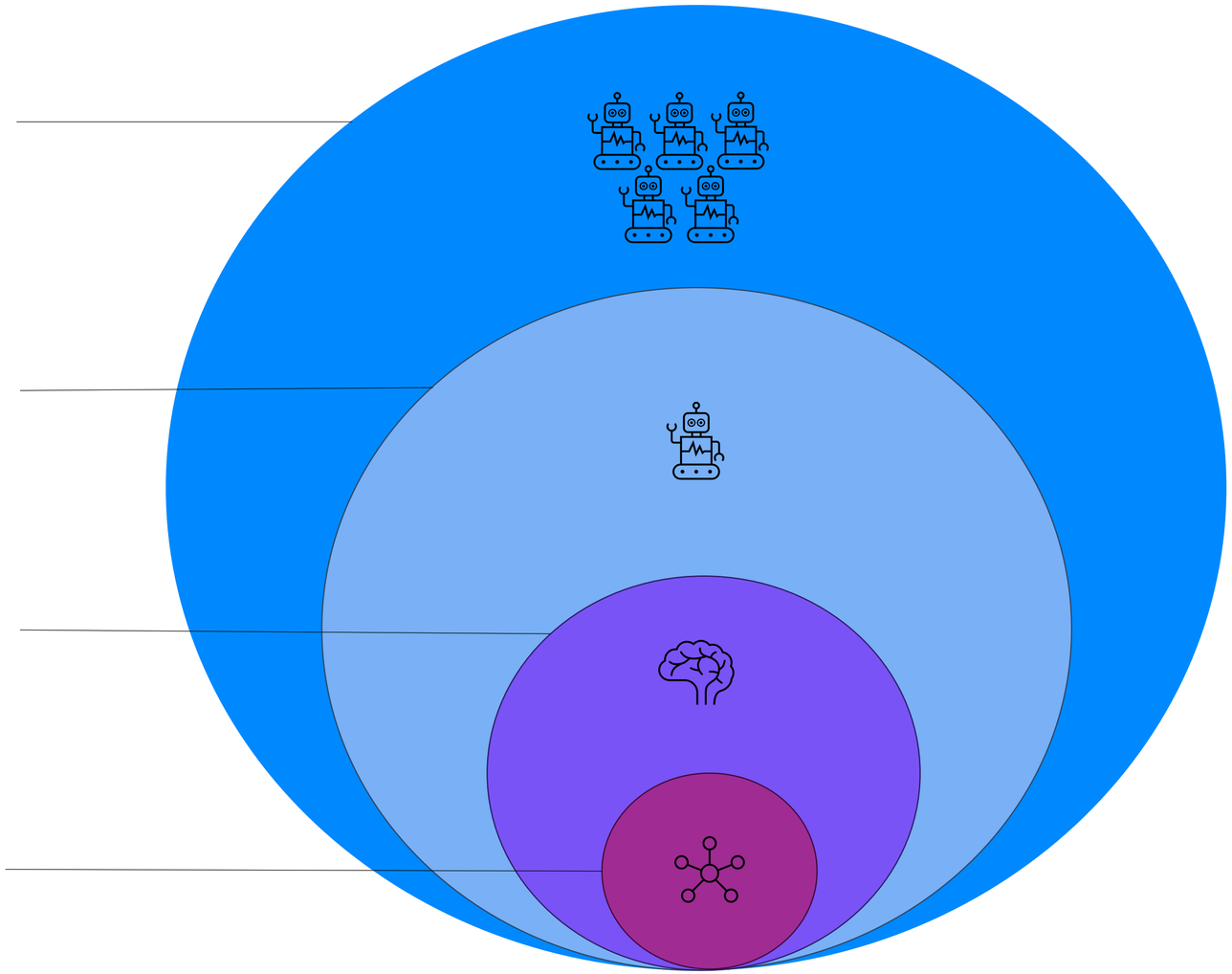
Agentic AI – GPT with a Purpose

Autonomous System:
Agentic AI

Agency:
AI Agents

Knowledge:
RAG (Retrieval Augmented Generation)

Language:
LLM (Large Language Model)



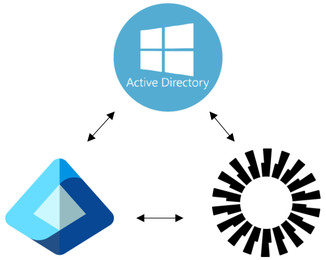
HYBRID IDENTITY PROTECTION conf25
CHARLESTON

The Identity Attack Surface Before & After AI

Before AI: Human-Led Attacks

- Slow & manual
  - Phishing campaigns required human setup
- Static playbooks
  - Credential stuffing, brute force, known exploits
- Limited adaptability
  - Failures = abandoned attempts
- Example: Password spraying attacks that lock accounts after a few tries

Common Targets

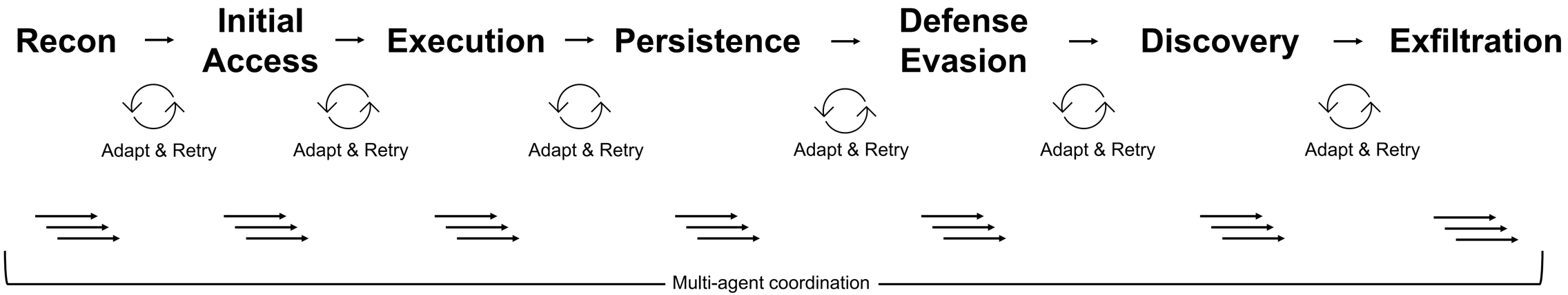


After AI: Agentic AI-Driven Attacks

- Fast & autonomous
  - AI executes attacks 24/7 at machine speed
- Adaptive playbooks
  - Dynamically switches from phishing → MFA fatigue → token replay
- Continuous learning
  - Adjusts tactics based on defenses, never gives up
- Very low cost of entry
- Example: MFA push fatigue (Uber breach), AI-driven OAuth abuse, automated privilege escalation

Deep Dive – Multi-Stage AI Attack (Agentic Workflow)

How an autonomous agent runs an identity breach, step by step.



- Low-and-slow vs burst
  - Agent decides between stealthy long campaigns or fast multi-vector strikes based on detected defenses
- Identity is ideal
  - Tokens, OAuth consents, and service principals allow quiet, high-value access – perfect for agentic workflows

HYBRID IDENTITY PROTECTION conf25
CHARLESTON

Deep Dive – Multi-Stage AI Attack (Agentic Workflow)

Recon → Initial Access → Execution → Persistence → Defense Evasion → Discovery → Exfiltration

- **OSINT (LinkedIn, GitHub, etc.)**
- **Automated enumeration**
- **Scan for exposed SSO/Oauth endpoints**
- **Builds a prioritized map of users, groups, privileged roles, service accounts, apps, etc..**

***Example**: Agent pulls org charts from LinkedIn and correlates with exposed endpoints to identify likely help-desk and admin users.*

HYBRID IDENTITY PROTECTION conf25 CHARLESTON

Deep Dive – Multi-Stage AI Attack (Agentic Workflow)

Recon → **Initial Access** → Execution → Persistence → Defense Evasion → Discovery → Exfiltration

- **Social Engineering**
  - tailored spear-phishing
  - voice-phishing scripts
  - convincing OAuth consent pages
- **Autonomous RCE discovery**

Example: *AI creates a fake password-reset email that mimics internal HR language and gets a target to approve an OAuth consent.*

HYBRID IDENTITY PROTECTION conf25 CHARLESTON

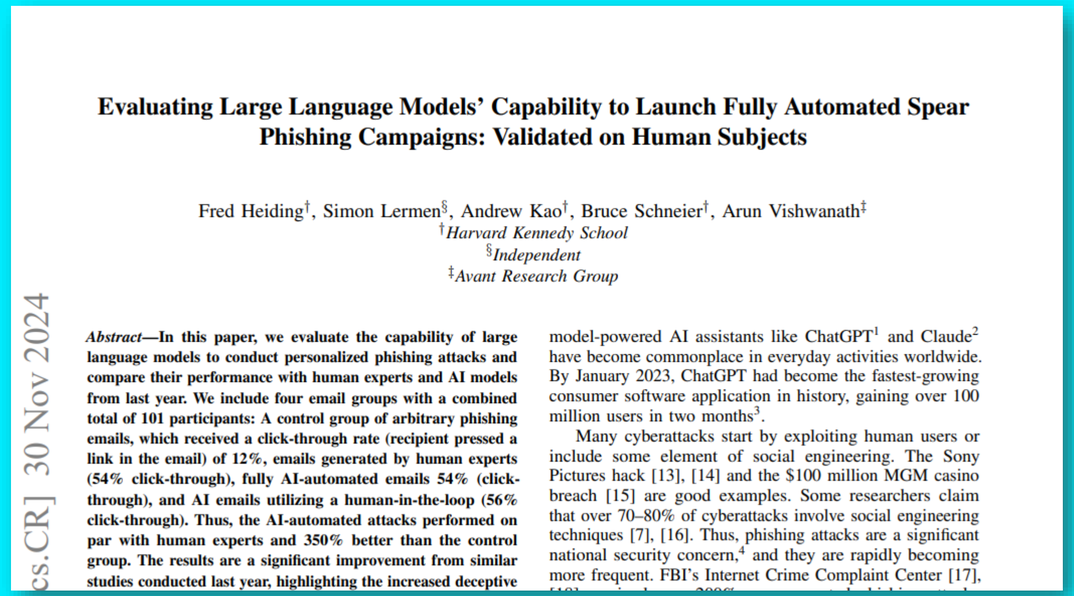
Deep Dive – Multi-Stage AI Attack (Agentic Workflow)

Recon → **Initial Access** → Execution → Persistence → Defense Evasion → Discovery → Exfiltration

2024 Study: AI vs Human Phishing Emails
(Harvard Kennedy School & Avant Research Group)

Basic scam (control): **12%** clicked
Human-crafted phishing: **54%** clicked
AI-generated phishing: **54%** clicked
AI + minor human edits: **56%** clicked

👉 AI already matches human experts—
slight human edits make it even more effective.



Evaluating Large Language Models' Capability to Launch Fully Automated Spear Phishing Campaigns: Validated on Human Subjects

Fred Heiding[†], Simon Lermen[§], Andrew Kao[†], Bruce Schneier[†], Arun Vishwanath[‡]
[†]*Harvard Kennedy School*
[§]*Independent*
[‡]*Avant Research Group*

***Abstract*—In this paper, we evaluate the capability of large language models to conduct personalized phishing attacks and compare their performance with human experts and AI models from last year. We include four email groups with a combined total of 101 participants: A control group of arbitrary phishing emails, which received a click-through rate (recipient pressed a link in the email) of 12%, emails generated by human experts (54% click-through), fully AI-automated emails 54% (click-through), and AI emails utilizing a human-in-the-loop (56% click-through). Thus, the AI-automated attacks performed on par with human experts and 350% better than the control group. The results are a significant improvement from similar studies conducted last year, highlighting the increased deceptive**

model-powered AI assistants like ChatGPT[1] and Claude[2] have become commonplace in everyday activities worldwide. By January 2023, ChatGPT had become the fastest-growing consumer software application in history, gaining over 100 million users in two months[3].

Many cyberattacks start by exploiting human users or include some element of social engineering. The Sony Pictures hack [13], [14] and the $100 million MGM casino breach [15] are good examples. Some researchers claim that over 70–80% of cyberattacks involve social engineering techniques [7], [16]. Thus, phishing attacks are a significant national security concern,[4] and they are rapidly becoming more frequent. FBI's Internet Crime Complaint Center [17],

HYBRID IDENTITY PROTECTION conf25 CHARLESTON

Deep Dive – Multi-Stage AI Attack (Agentic Workflow)

Recon → Initial Access → **Execution** → Persistence → Defense Evasion → Discovery → Exfiltration

- **Uses stolen credentials/tokens to run code or initiate sessions (API calls, PowerShell, remote commands)**
- **Delivers payloads or calls APIs to advance access (lateral movement, privilege escalation)**

Example: *Agent uses a refresh token to call Graph API and create a delegated session, issuing least-privilege commands to test scope.*

HYBRID IDENTITY PROTECTION conf25 CHARLESTON

Deep Dive – Multi-Stage AI Attack (Agentic Workflow)

Recon → Initial Access → Execution → **Persistence** → Defense Evasion → Discovery → Exfiltration

- **Establishes durable footholds**
  - Creates stealthy service principals
  - Registers hidden OAuth apps
  - Seeds long-lived refresh tokens
  - Modifies automation accounts

Example: *Agent registers a “CI/CD integration” app with high privileges but a benign name and consent.*

HYBRID IDENTITY PROTECTION conf25 CHARLESTON

Deep Dive – Multi-Stage AI Attack (Agentic Workflow)

Recon → Initial Access → Execution → Persistence → **Defense Evasion** → Discovery → Exfiltration

- **Cleans logs**
- **Mimics normal user behavior**
  - Rotates tactics (sleep windows, low-rate actions)
- **Uses stolen session cookies or replay tokens to avoid MFA prompts**

Example: *Agent switches to long-lived refresh tokens and routes traffic through chained proxies that match corporate IP ranges.*

HYBRID IDENTITY PROTECTION conf25 CHARLESTON

Deep Dive – Multi-Stage AI Attack (Agentic Workflow)

Recon → Initial Access → Execution → Persistence → Defense Evasion → **Discovery** → Exfiltration

- **Enumerates internal resources it can access (mailboxes, file shares, endpoints)**
- **Maps lateral paths**
- **Collects secrets/service connection strings**
- **Scans files to find the top-secret data**

Example: *Agent uses graph analysis to find shortest privilege escalation paths and automates exploitation of misconfigurations (e.g., over-permissive group nesting).*

HYBRID IDENTITY PROTECTION conf25
CHARLESTON

Deep Dive – Multi-Stage AI Attack (Agentic Workflow)

Recon → Initial Access → Execution → Persistence → Defense Evasion → Discovery → **Exfiltration**

- **Collects and transfers data (tokens, PII, credentials, config) via covert channels**
- **Chunked uploads**
- **Encrypted blobs in benign services or via chained third-party apps.**

Example: Chooses exfiltration channels that mimic normal traffic (e.g., piggybacking on CI/CD artifact pushes) and fragments data to avoid detection.

HYBRID IDENTITY PROTECTION conf25
CHARLESTON

Deep Dive – Multi-Stage AI Attack (Agentic Workflow)

How an autonomous agent runs an identity breach, step by step.

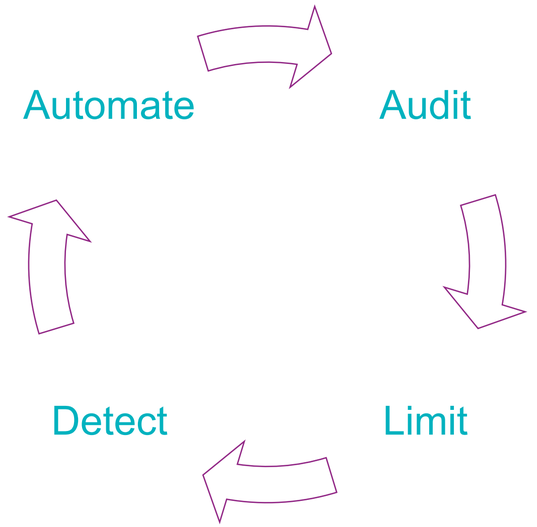
Recon → **Initial Access** → **Execution** → **Persistence** → **Defense Evasion** → **Discovery** → **Exfiltration**

- **Speed + Adaptation = Scale**
  - AI compresses reconnaissance-to-impact timelines
- **Identity is the enabler**
  - Tokens, OAuth consents, and service principals are high-value lifelines attackers exploit
- **Signal, correlate, respond**
  - Single events are weak signals; chain correlation is where detection wins

HYBRID IDENTITY PROTECTION conf25
CHARLESTON

Practical Defense Strategies – Stop Agentic Identity Attacks

- **Audit & Reduce Blast Radius**
  - Inventory apps, service principals, privileged groups
- **Enforce Least Privilege & JIT**
  - Short-lived elevation for admin tasks
- **Phishing-Resistant MFA**
  - FIDO2 / passkeys; disable legacy/OTP where possible
- **Harden App Consent & Service Accounts**
  - Block risky OAuth consent; require admin attestation
- **Behavioral Detection on Identity Telemetry**
  - Baseline Graph/API, token, and consent patterns
- **Automated Response Playbooks**
  - Revoke tokens, quarantine apps, rotate creds



HYBRID IDENTITY PROTECTION conf25
CHARLESTON

Identity Is the New Last Line of Defense

Agentic AI changes the game – but strong identity defenses changes the outcomes.

Ran Harel
ranh@semperis.com
Connect on LinkedIn:

Thank you!

HYBRID
IDENTITY
PROTECTION

conf25

CHARLESTON