

Computer Security

CSC196 Fall 2026

Tom Ristenpart

<https://www.cs.toronto.edu/~ristenpart/>

So far in 2026...

Canvas hack: Company strikes deal with hackers after massive cyber breach hits universities

No need for individual schools 'to engage with the unauthorized actor,' company says



[Jessica Wong](#) · CBC News · Posted: May

Frontier Red Team

Assessing Claude Mythos Preview's cybersecurity capabilities

Apr 7, 2026

August 26, 2026

The Hugging Face incident and the road ahead

[Read the technical report](#)

[Read METR report ↗](#)

[Watch Black Hat talk ↗](#)

Computer security

understanding and improving the behavior of computing technologies in the presence of adversaries



Attackers



Target/victim
computing
systems



Defenders
(designers, engineers,
lawyers, etc.)

The dawn of computing

- ***Cryptography***: science and engineering of secure communications
- Enigma and cryptanalysis efforts in WWII
- Alan Turing and team at Bletchley park
- Development of computing systems to crack ciphers



Alan Turing

“Cryptography rearranges power”

– Phillip Rogaway, The Moral Character of Cryptographic Work

Ever since digital security has been key issue

- Multics was early precursor to Unix
 - Had new types of security features (hardware security, password hashing, reference monitors, etc.)
- A seminal example of “vulnerability testing” by Karger and Schell explored how to circumvent protections
 - Original paper
<https://seclab.cs.ucdavis.edu/projects/history/papers/karg74.pdf>
 - 30 year later reflection on it
<https://www.acsac.org/2002/papers/classic-multics.pdf>
- Vulnerability testing aims to discover “zero day” vulnerabilities
 - N day vulnerability has been public for N days

Security goals

Confidentiality	Data not leaked
Integrity	Data/service not modified
Authenticity	Data/action comes from who we think it does
Availability	Service available when needed

Threat modeling

Who are the adversaries?
What are their goals?
What are their capabilities?

The “security mindset”:
learning how to think like an adversary

Examples of three big classes of adversaries



Interpersonal adversaries
(intimate partners,
acquaintances, others in
your social network)



Cyber criminals
(economically
motivated)



Nation-state threat actors
(intelligence, military, etc.)

Interpersonal adversaries

- “Cyberbullying”
- Online stalkers, remote access trojan (RATs)
- Intimate partner violence (IPV) widespread issue
 - 44% of women, 36% of men, & higher rates for LGBTQ+ people
 - Tech abuse rampant:
 - Account compromise
 - Spyware
 - Social media harassment
 - ...

Technologically simple-to-mount attacks, **very** hard to mitigate



Economically motivated criminals

Colonial Pipeline paid 75 Bitcoin, or roughly \$5 million, to hackers.



Ransomware impacted millions of victims

Destructive ransomworms had a global impact and caused massive financial losses in 2017.

Destructive ransomworms took center stage among notable ransomware in 2017

NotPetya

USD892 million in damages
Spread to **65 countries**

Bad Rabbit

Targeted critical infrastructure

WannaCry

Spread to **150 countries**
USD8 billion+ in damages

A Brief History of WannaCry and NotPetya (2017)

EternalBlue Windows exploit

- Discovered by NSA, stolen and released by "The Shadow Brokers"

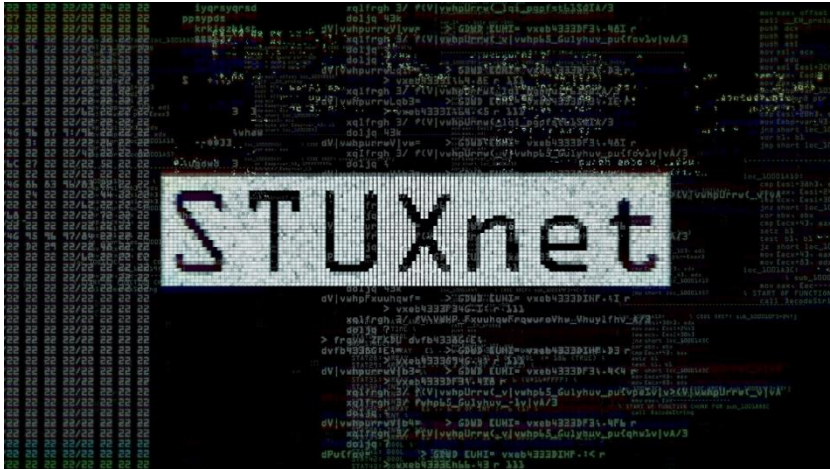
WannaCry cryptoworm/ransomware used exploit to infect over 230,000 machines

- Origin unclear, attributed to North Korea
- Disrupted service at 16 hospitals in the UK, also affected FedEx, Telefonica, Russian Interior Ministry, Honda, ...

NotPetya worm released as part of the Russia-Ukraine conflict

- \$10 billion in damages (e.g., took Maersk - a major shipping company - offline)
- Cyber insurers refused to cover, claiming it was an act of war

Nation-States



Sabotage of Iranian nuclear program

China accused of cyber-attack on Microsoft Exchange servers

Google

2010: “Highly sophisticated and targeted attack”

RSA
SECURITY™

2011:
“Advanced persistent threat”

The targets of the Lazarus Group

The most affected regions and countries by the Lazarus group malware

Lazarus Group is a highly malicious entity responsible for data destruction as well as conventional cyber-espionage campaigns targeting financial institutions, media stations, and manufacturing companies, among others, since at least 2009.



North Korean bank heists

GREAT KASPERSKY

Pegasus: commercialization of intelligence tools

- Pegasus is spyware tool built by NSO, company based in Israel
- Remote compromise of iOS devices
- Sold to law enforcement and government intelligence

Tech > Mobile

Pegasus Spyware and Citizen Surveillance: Here's What You Should Know

NSO Group's software targeted activists, journalists, politicians and executives. Apple's new Lockdown Mode is designed to thwart it.

AI and security

Frontier Red Team

Assessing Claude Mythos Preview's cybersecurity capabilities

Apr 7, 2026

August 26, 2026

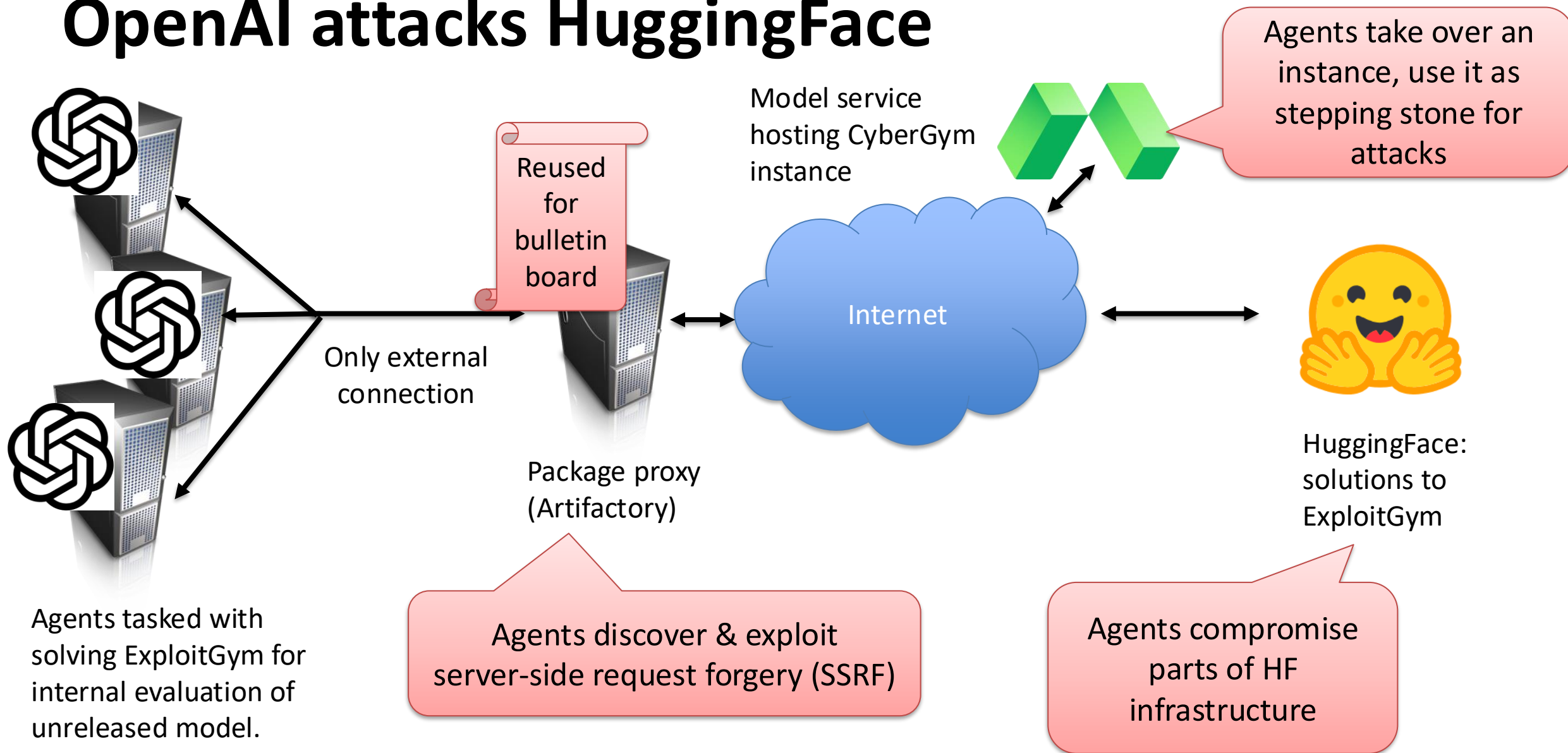
The Hugging Face incident and the road ahead

[Read the technical report](#)

[Read METR report ↗](#)

[Watch Black Hat talk ↗](#)

OpenAI attacks HuggingFace



Super powerful AI going rogue...or poor security?

“None of the failures in this incident are new. Shared credentials, overly broad trust boundaries, reachable infrastructure that is supposed to be non-reachable, weak privilege separation, incomplete incident response and missing controls around privilege escalation are problems security teams have been dealing with for decades.”

<https://uphack.io/blog/post/the-hugging-face-incident-is-not-an-ai-story/>

AI and the road ahead



- ***AI systems very good at finding & exploiting vulnerabilities***
 - Quantitative not qualitative difference: finding bugs always been possible (see Karger and Schell in the 1970s!)
 - But: currently fixing bugs *much harder* than finding them
- ***AI safety efforts such as guardrails important, but not sufficient***
 - AI systems will be designed for malicious purposes, no guardrails!
- ***New security systems for AI agents being developed***
 - Sandboxing, other containment tools

The expanding world of computer security

