

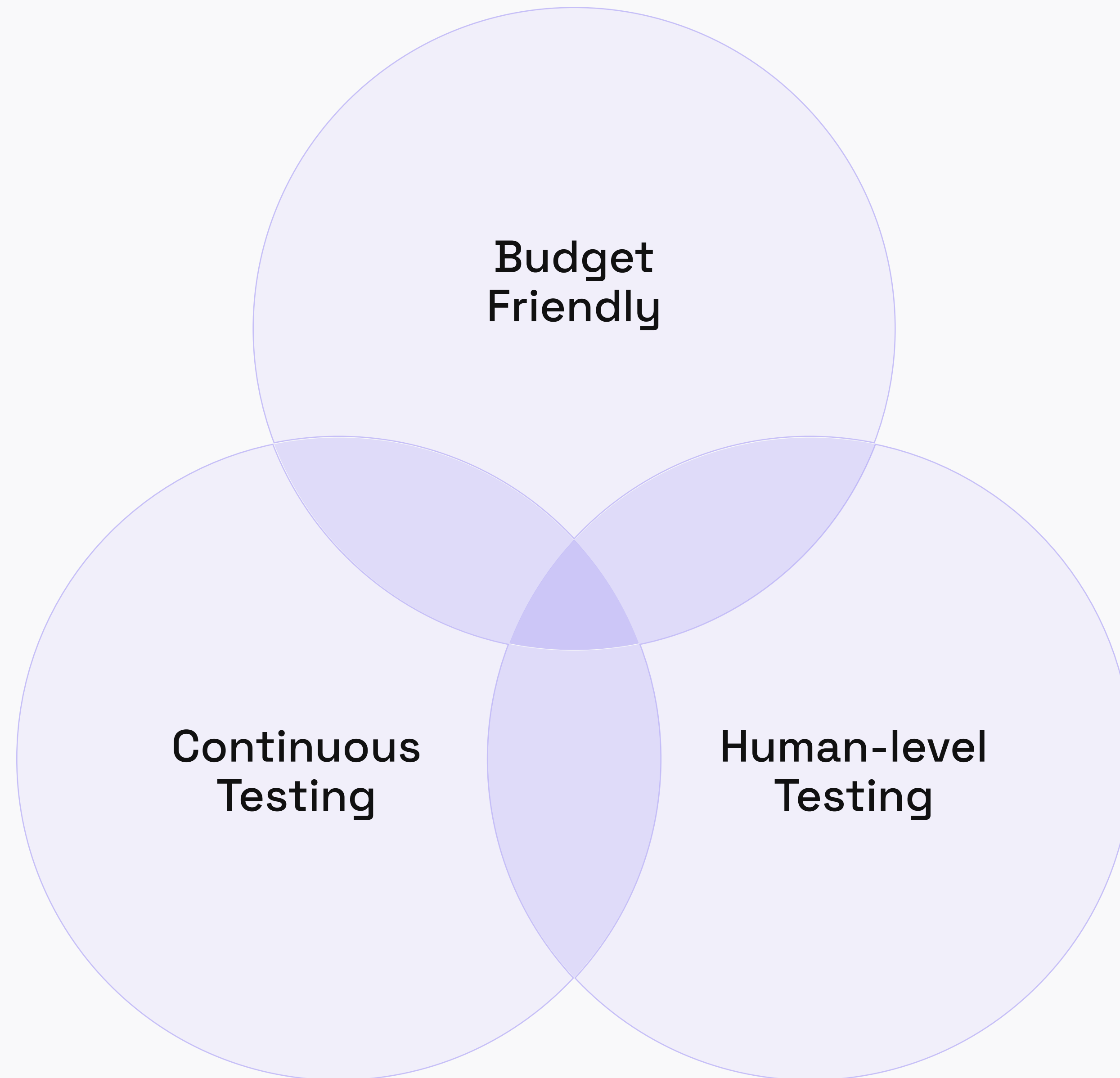


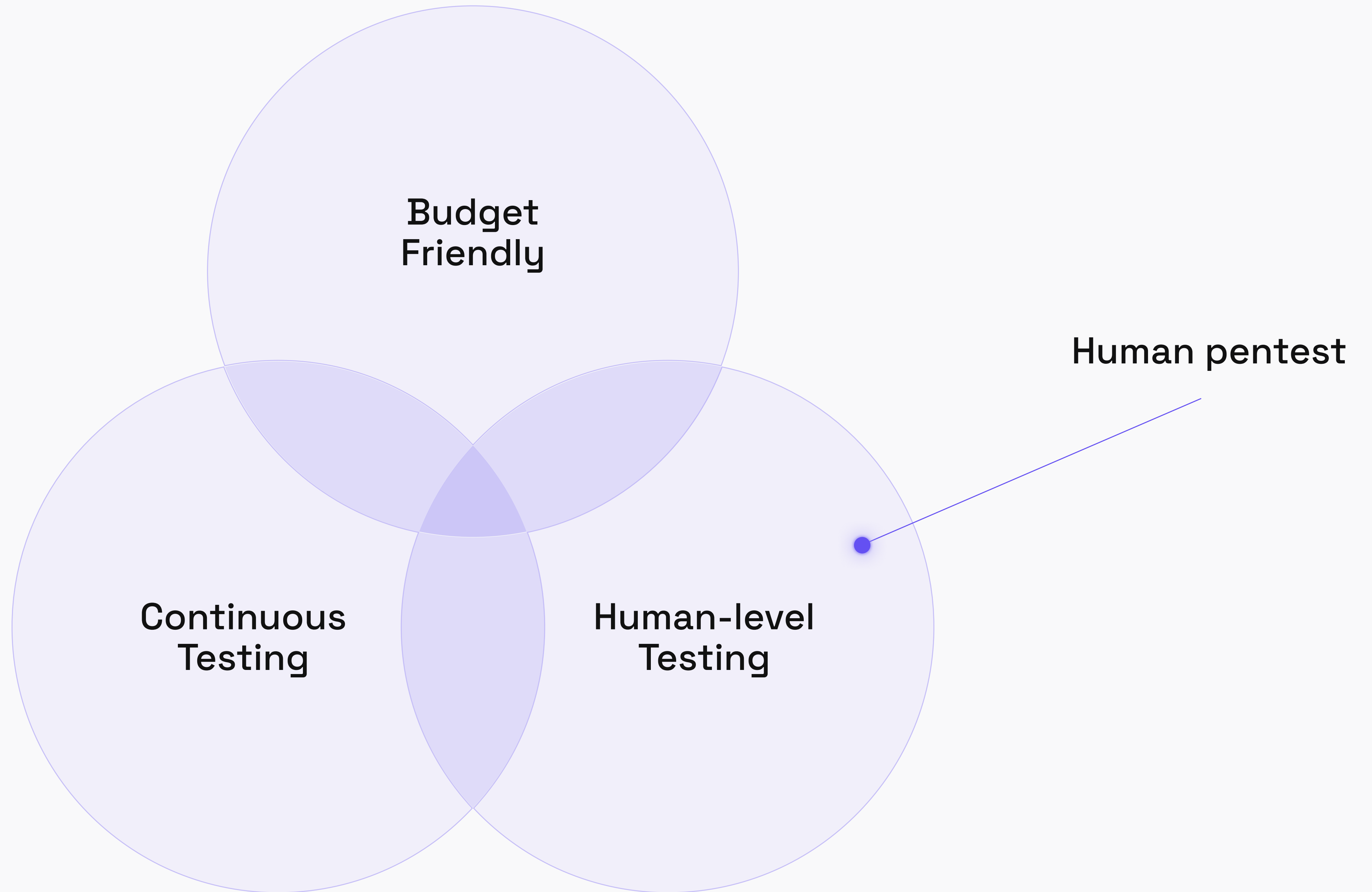
AI in the Ring:

Agentic AI in Offensive Security

↳ aikido.dev

Offensive Security Is Broken





Vulnerability scanners



RAPID7



Qualys®

invicti



BURPSUITE
ENTERPRISE EDITION

VERACODE

Checkmarx



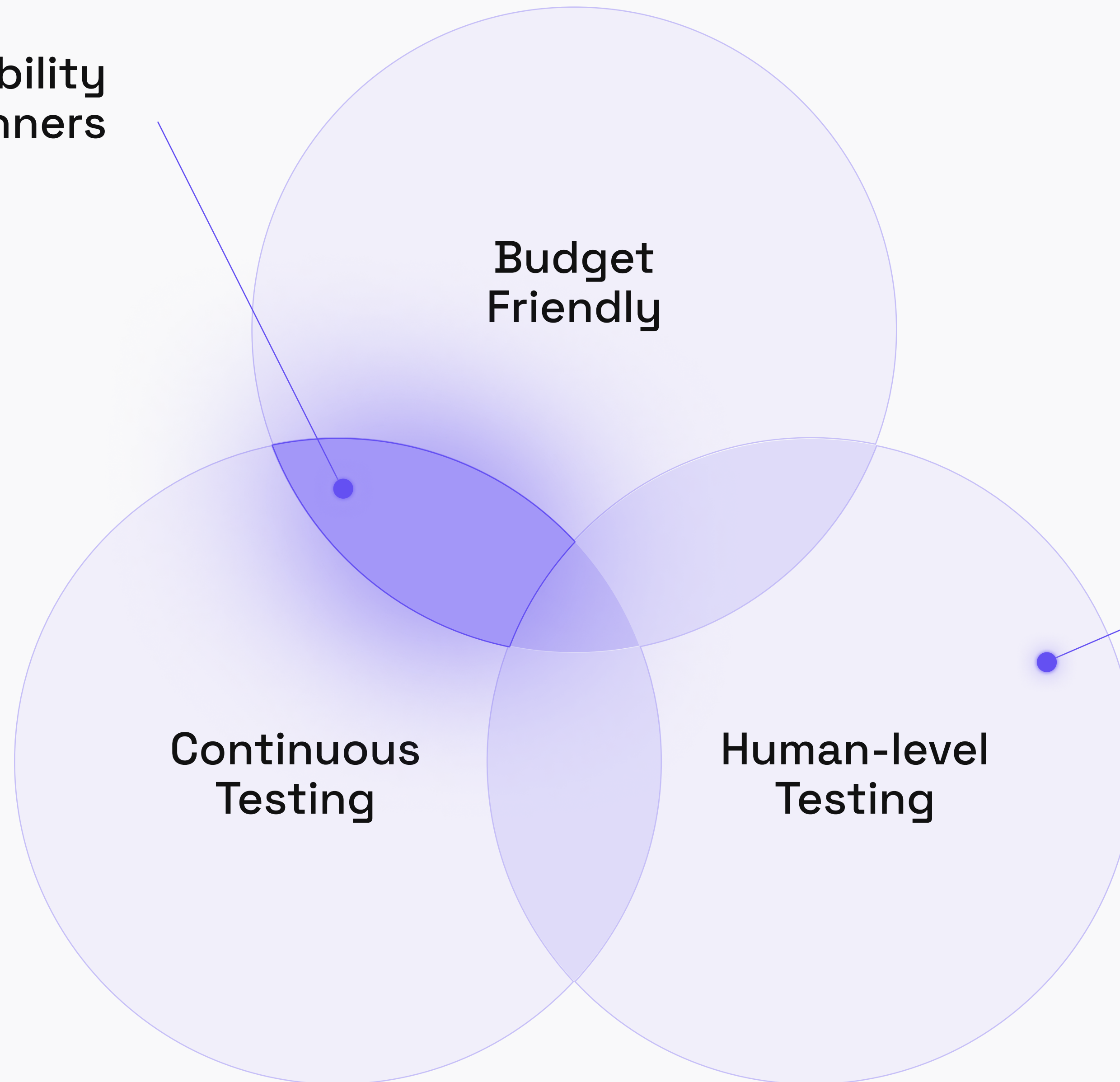
**Vulnerability
scanners**

**Budget
Friendly**

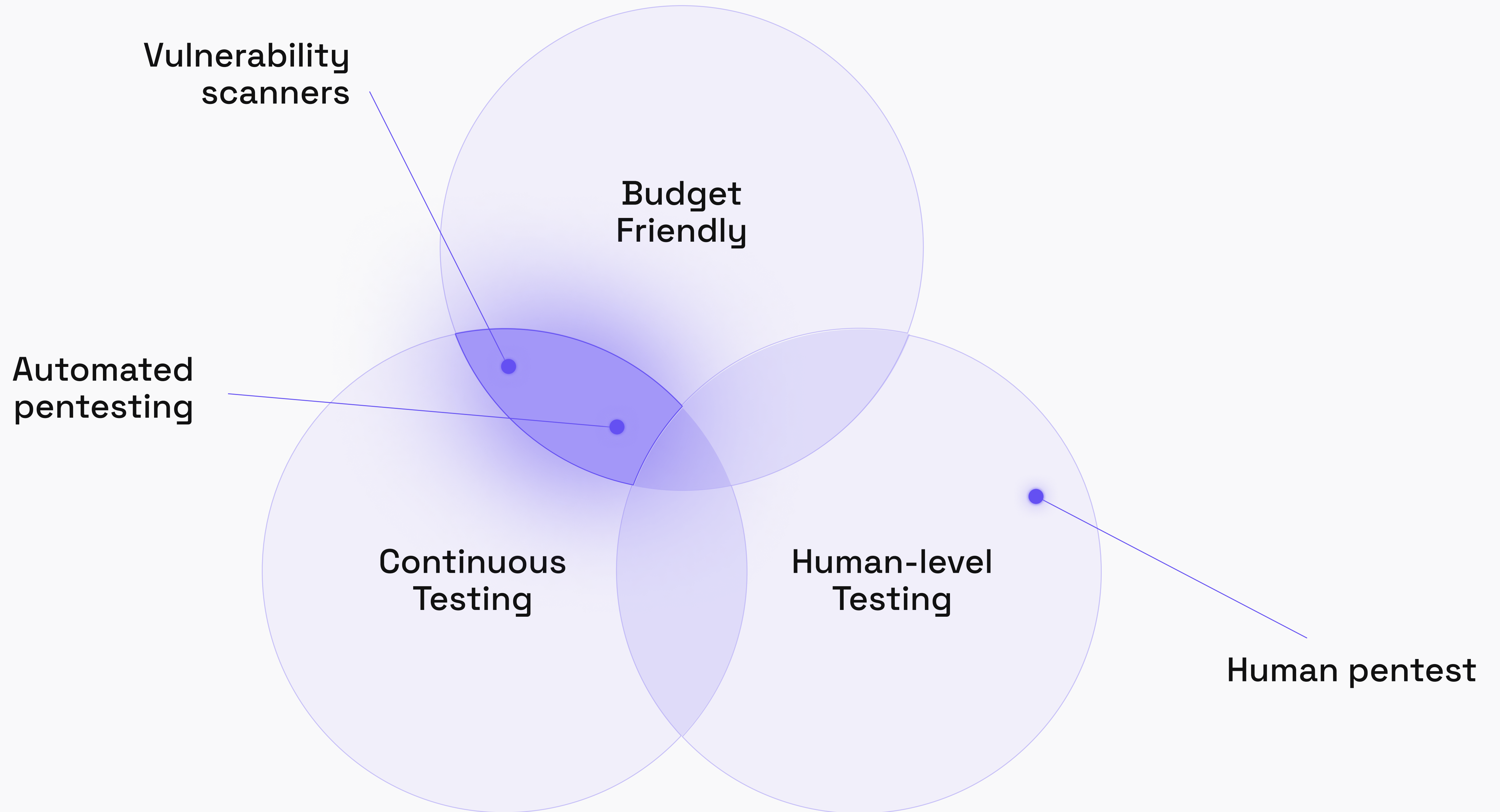
Human pentest

**Continuous
Testing**

**Human-level
Testing**

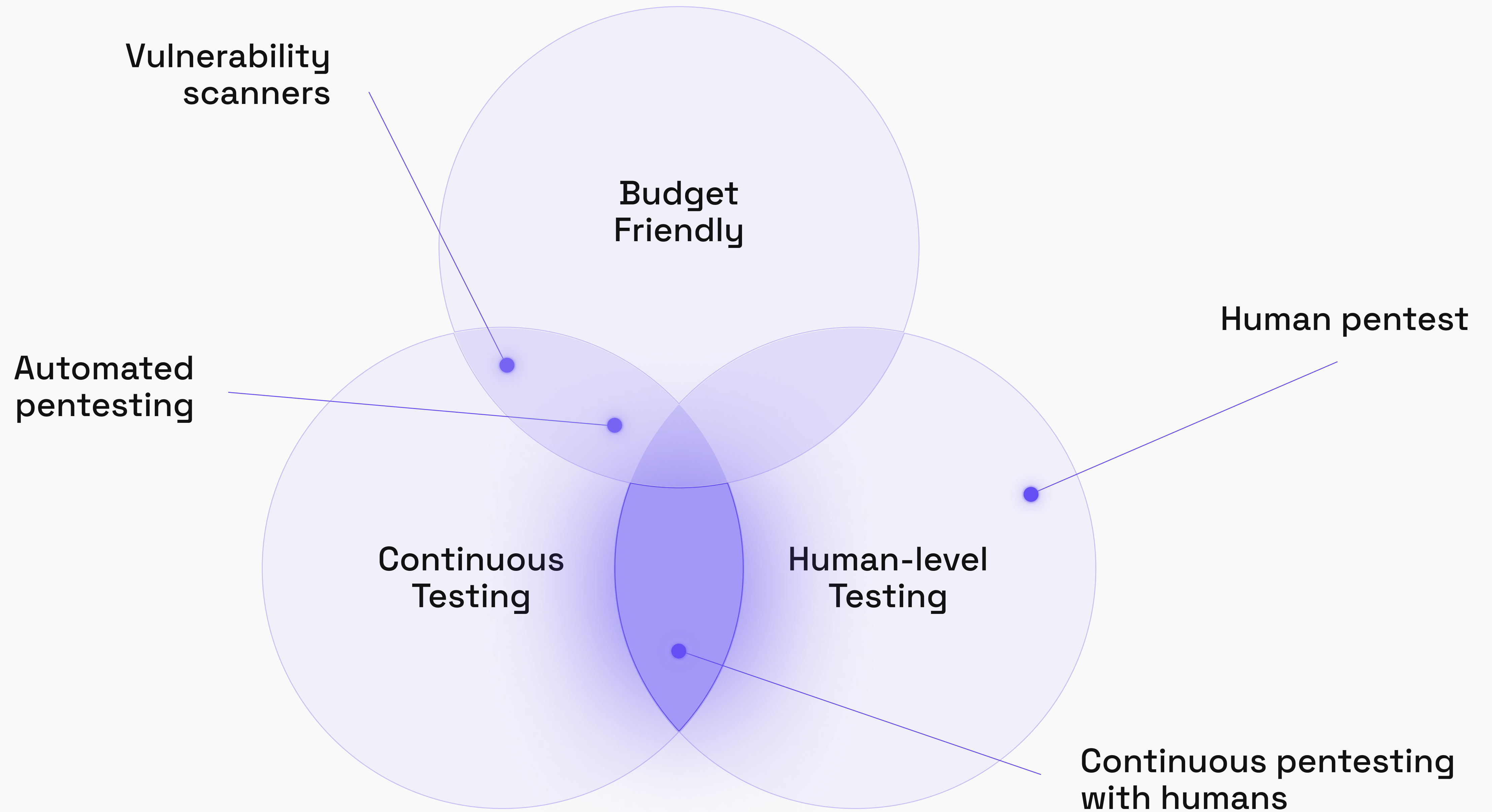






Continuous pentesting

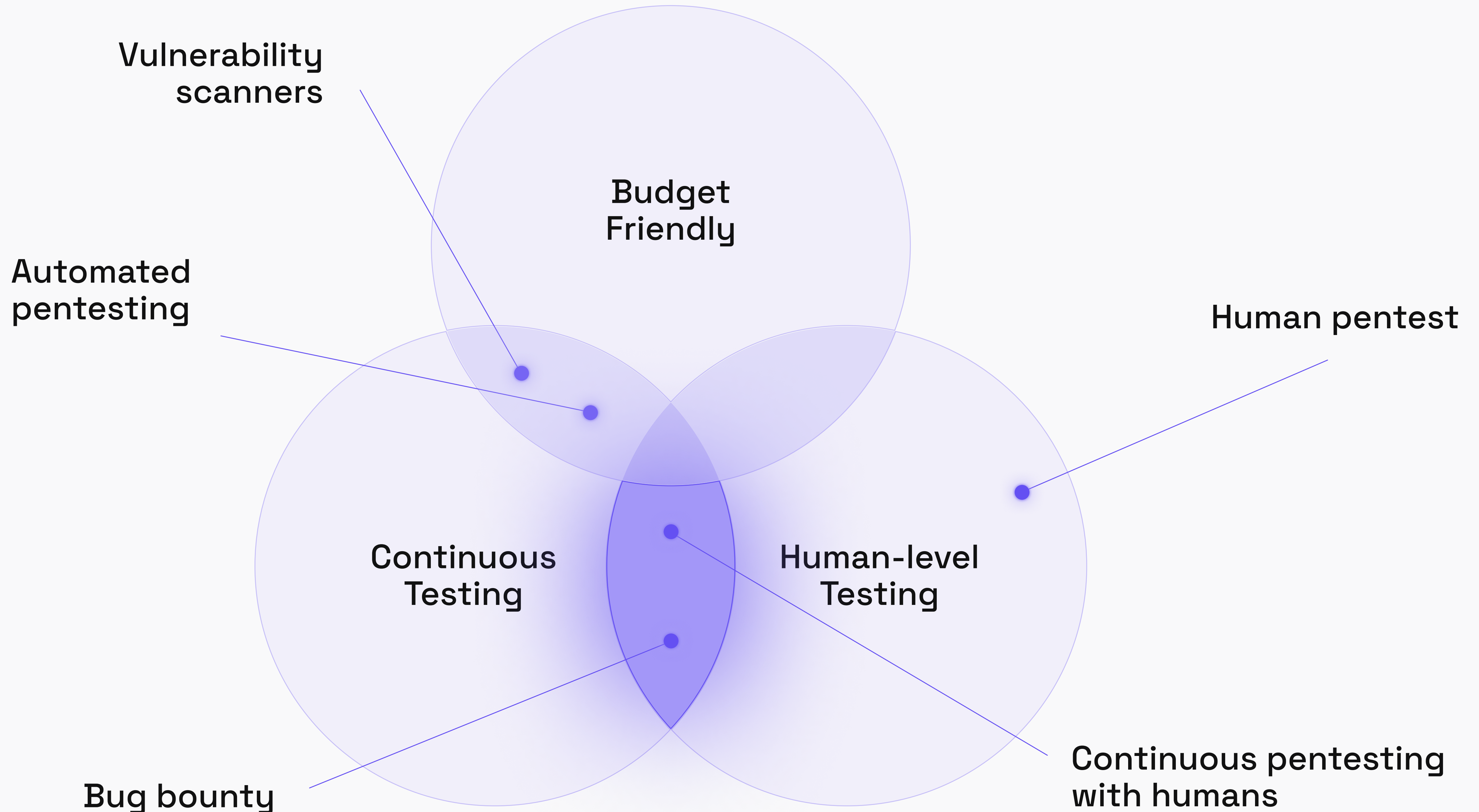
**More Humans
More Often**



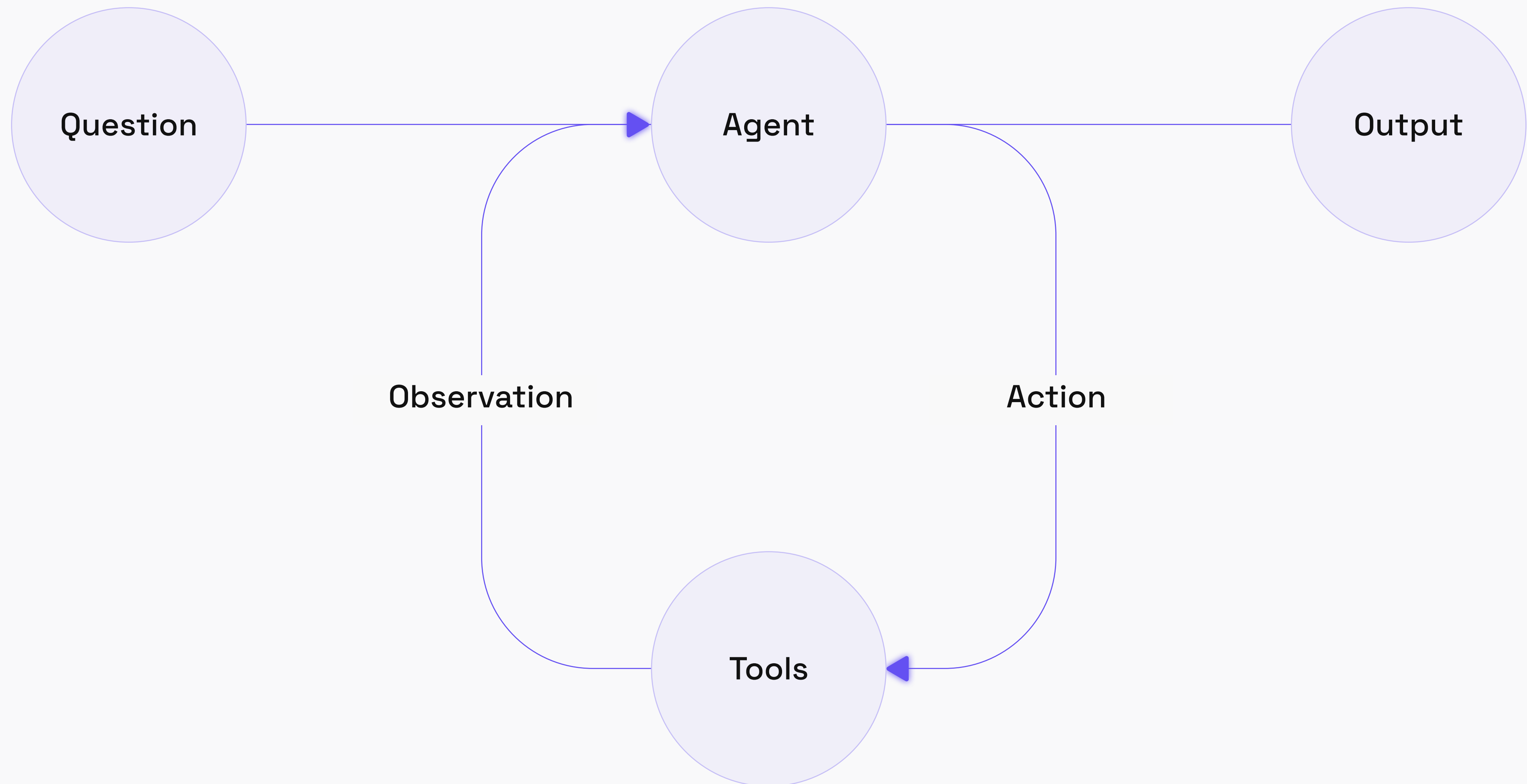
Bug bounty programs

A Lot More Humans

A Lot More Often



Decisions made by LLM



What is Agentic AI?

Orchestration

Discovery
Agent

Code Analysis
Agent

Preflight
Agent

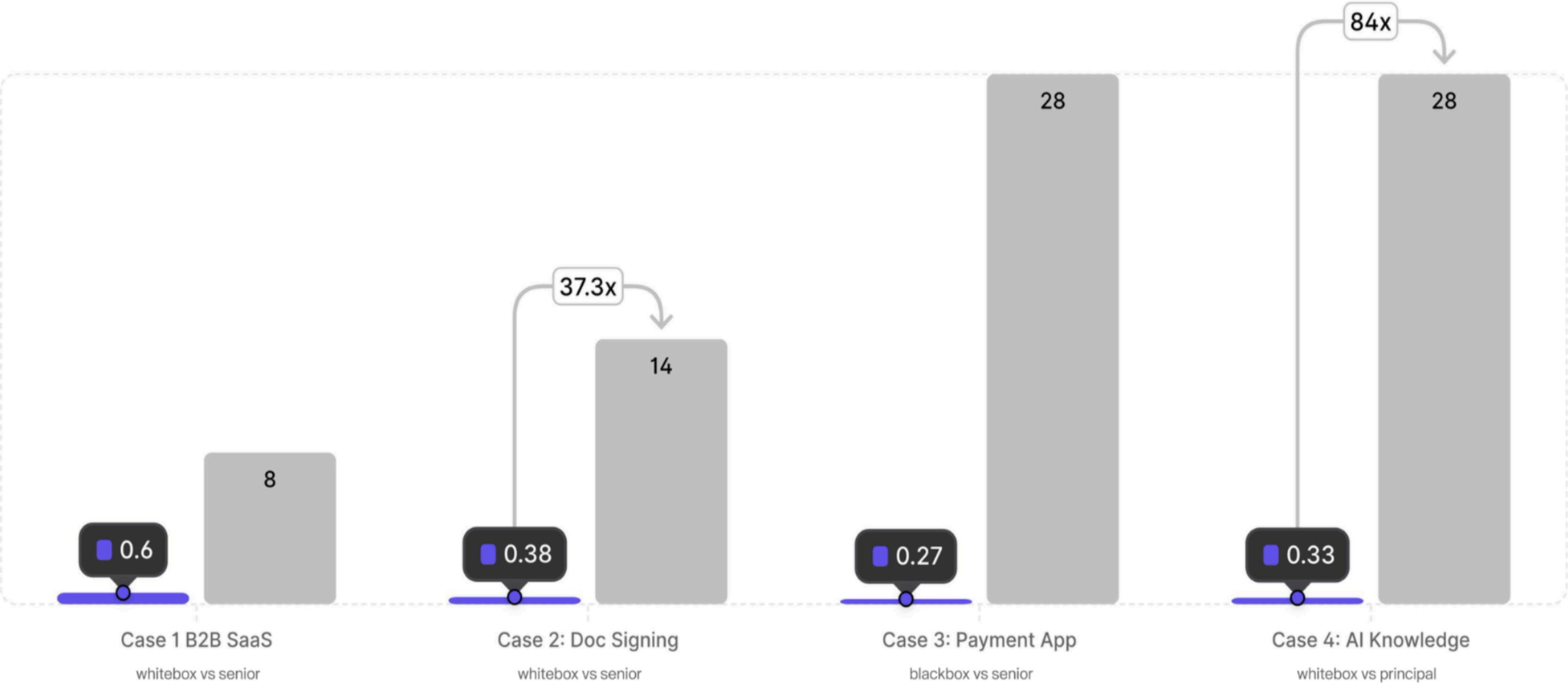
Attacker
Agent

Browser
Agent

Time to Complete Pentest (in Days)

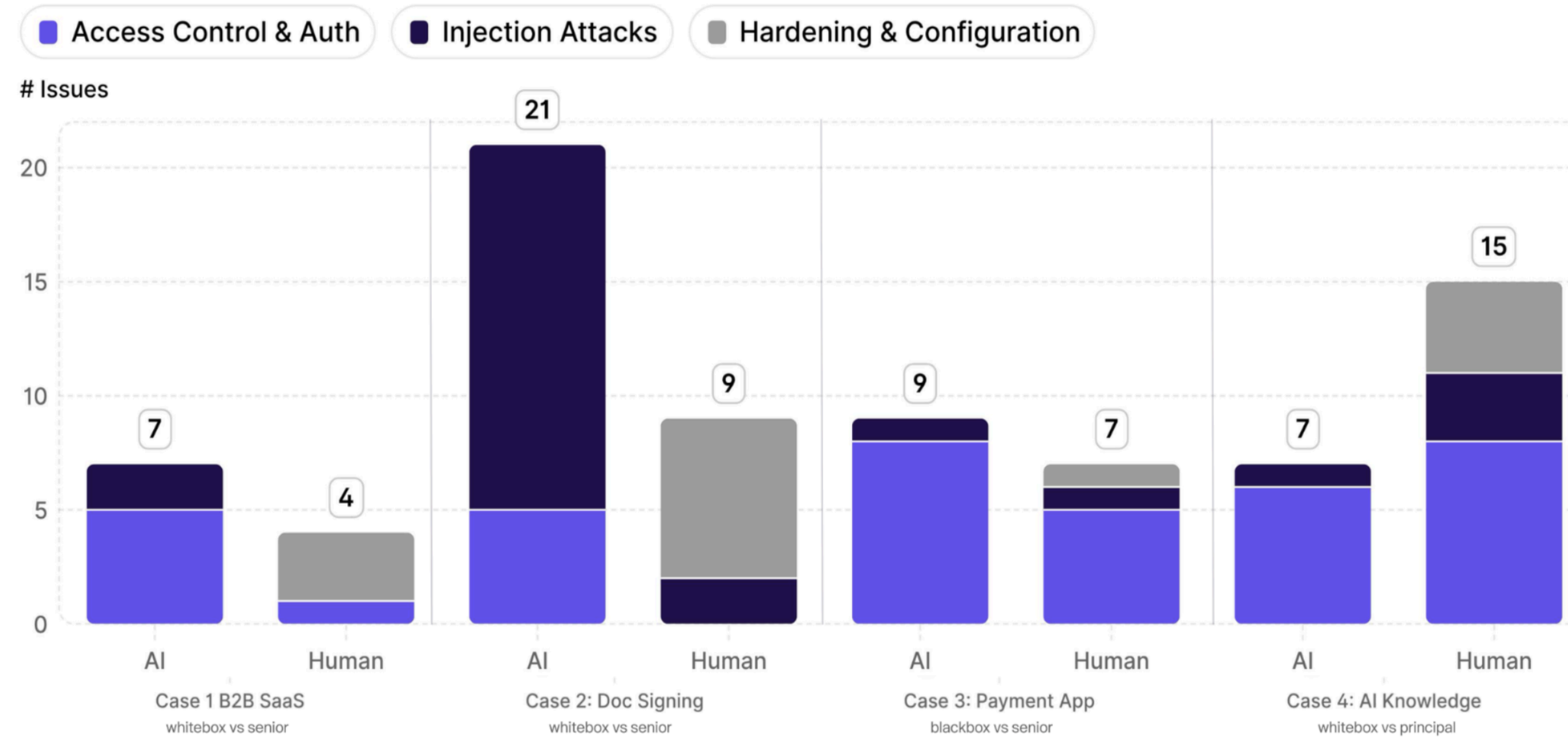
Time to Complete Pentest (in Days)

0.4 **19.5**
■ Avg.Aikido AI ■ Avg.Human

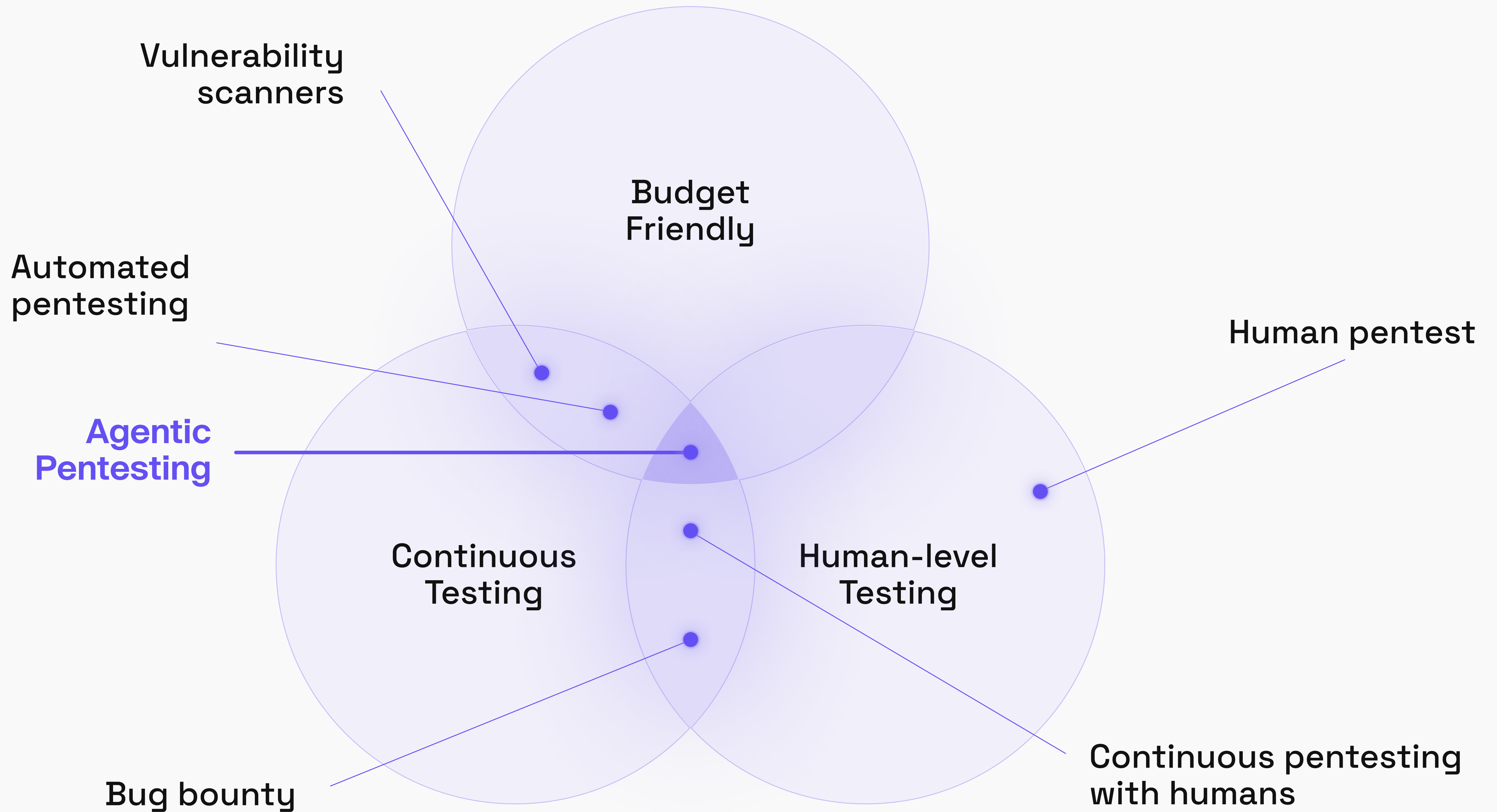


Number of Findings / Case

Number of Findings / Case



Human-level pentesting
machine scale





What SATAN Is

[Extract from a USENET posting dated March 8, 1995]

SATAN was written because we realized that computer systems are becoming more and more dependent on the network, and at the same becoming more vulnerable.

The rationale for SATAN is given in a paper that we posted in december 1993 ([Improving the Security of Your Site by Breaking Into it](#)).

SATAN is a tool to help systems administrators. It recognizes several common networking-related security problems, and reports the problems without making any changes.

For each type of problem found, SATAN offers a tutorial that explains the problem and what its impact could be. The tutorial also explains what can be done to fix the problem.

SATAN collects information that is available to everyone on with access to the network. With a properly-configured firewall in place, that should be no problem.

We have done some limited research with SATAN. Our finding is that on networks with more than a few dozen systems, SATAN will inevitably find problems.

- NFS file systems exported to arbitrary hosts
- NFS file systems exported to unprivileged programs
- NFS file systems exported via the portmapper
- NIS password file access from arbitrary hosts
- Old (i.e. before 8.6.1) versions of NFS
- REXD access from arbitrary hosts
- X server access control
- arbitrary files accessible to arbitrary users
- remote shell access from arbitrary hosts
- writable anonymous FTP home directory

We realize that SATAN is a two-edged sword - like many tools, it can be used for good and for evil purposes. We also realize that intruders (including us) can use it to find out what is going on in your system.

These are well-known problems. They have been subject of CERT, CIAC, or other advisories, or are described extensively in practical security handbooks.

We realize that SATAN is a two-edged sword - like many tools, it can be used for good and for evil purposes. We also realize that intruders (including us) can use it to find out what is going on in your system.

ANTHROPIC

Disrupting the first reported AI-orchestrated cyber espionage campaign

Full report


November 2025

anthropic.com

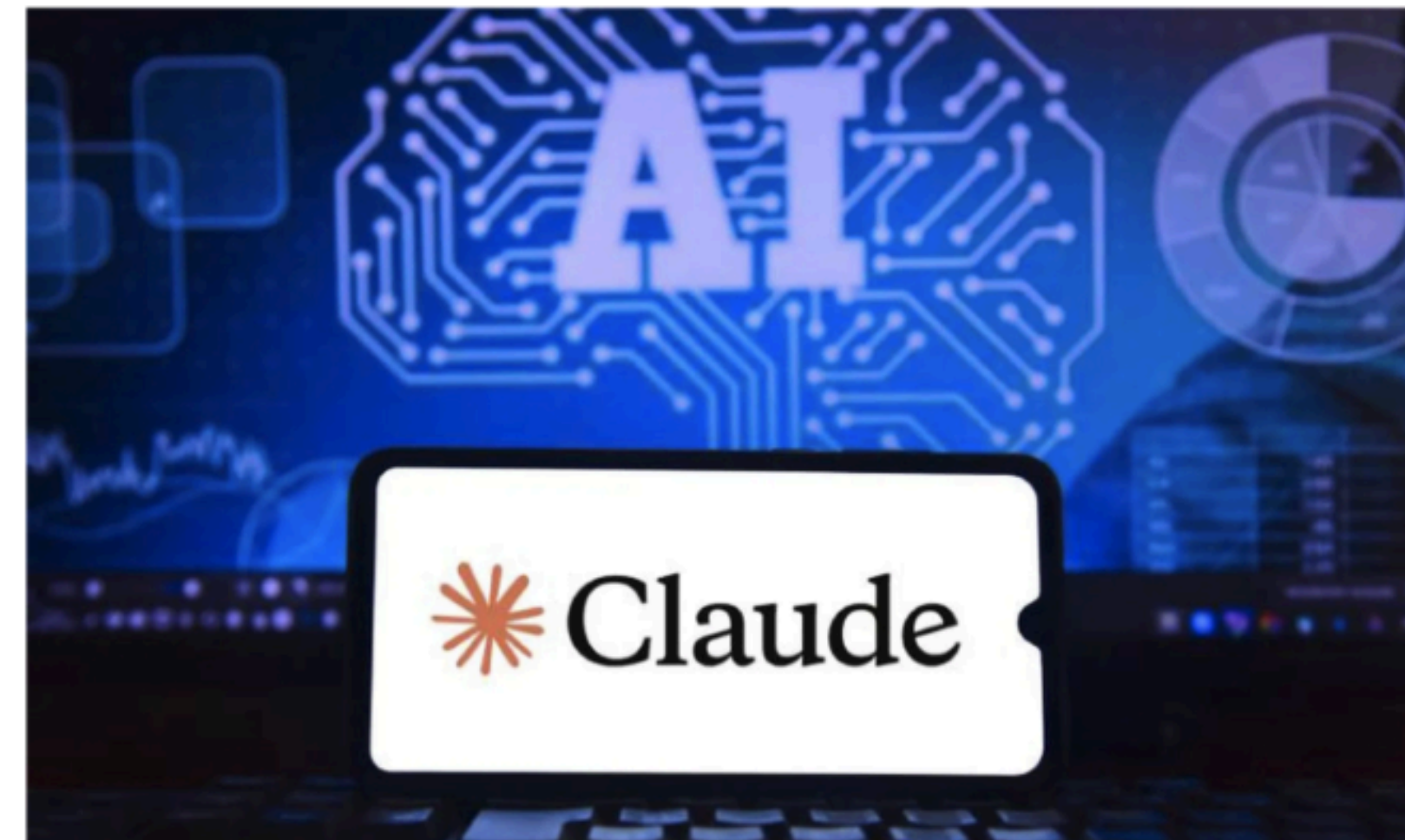
AI/ML, Government security, Data Security

Mexico reportedly breached via Claude exploitation

February 27, 2026

 Share

By SC Staff



(Credit: sauloangelo – stock.adobe.com)

Numerous Mexican government agencies have been compromised in a month-long attack campaign beginning in December that weaponized Anthropic's [Claude](#) large language model, resulting in the theft of 150 GB of data, according to [Cybernews](#).

Infiltration of Mexico's federal tax authority and civil registry, as well as some state governments and Monterrey's water utility, using 20 security flaws identified by Claude has allowed threat actors to steal nearly 195 million taxpayer records, civil registry files, voter lists, and government employee credentials, findings from a Gambit Security report revealed. Harnessing Claude involved attackers sending prompts to determine system vulnerabilities and craft exploit scripts under the guise of a bug-hunting operation.



Mythos

Project
Glasswing

The Future

~~Low~~ Medium
hanging fruit

The Future

**99% of
Organisation**

- Agentic Pentesting

**Organisation
with low risk
appetite**

Combination of

- Agentic Pentesting

- Manual Pentesting

The future

Self-Securing Software

