CLOUDFLARE

world

After human

How to think about fraud & abuse when anyone could be anything





Bryan Becker
Director, Product
Cloudflare

Tiago Sada
Chief Product Officer
Tools For Humanity

CLOUDFLARE

Humans are no longer the majority

- Over half of the Internet is automated
- Personal agents are here
- Browsers and agents are converging

After human

Attacks are now infinite



CLOUDFLARE

Are you who you say you are?

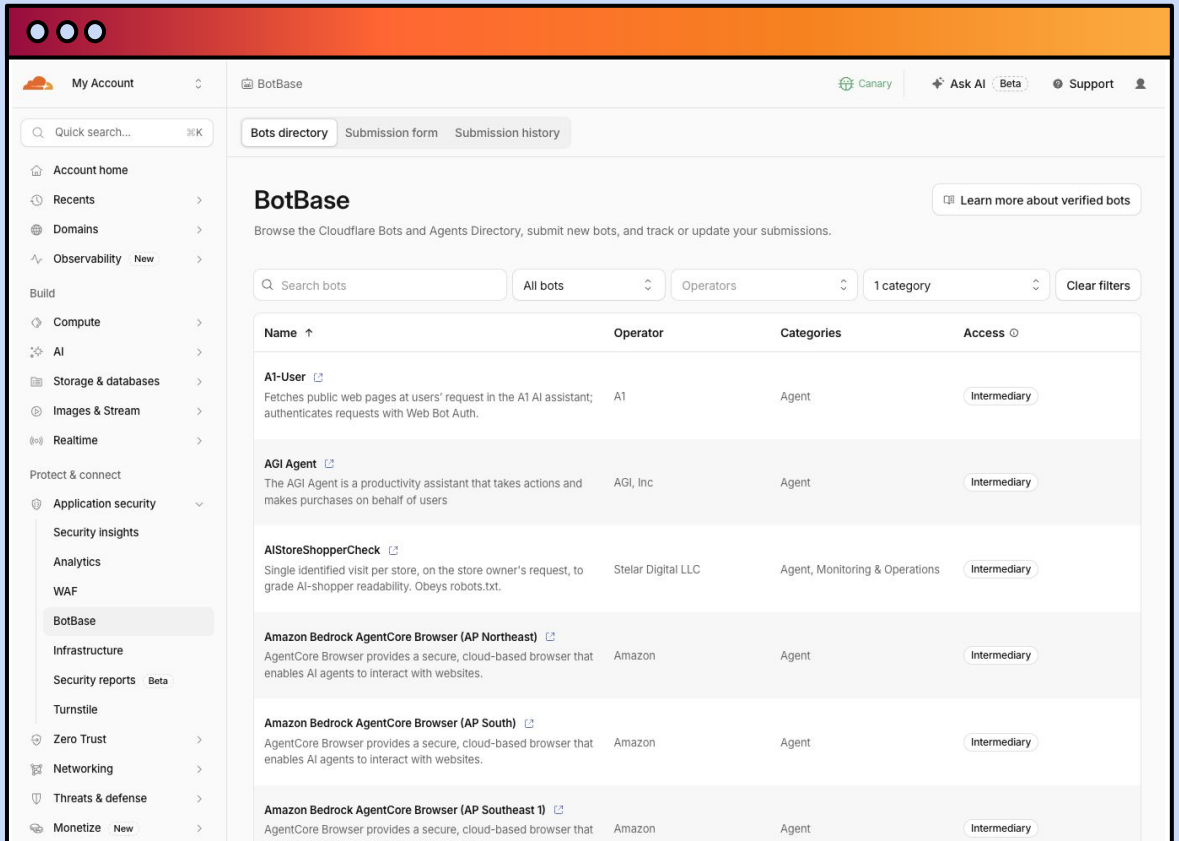
Can we trust you?

After Human

CLOUDFLARE

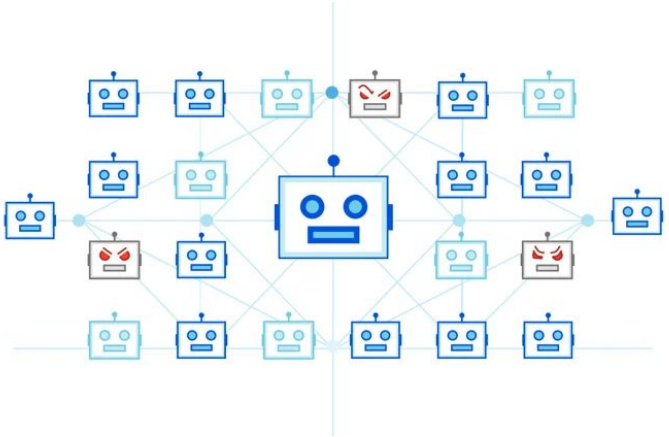
Building a BotBase

tracked and verified bots and agents



CLOUDFLARE

Sophisticated & persistent attackers don’t stop

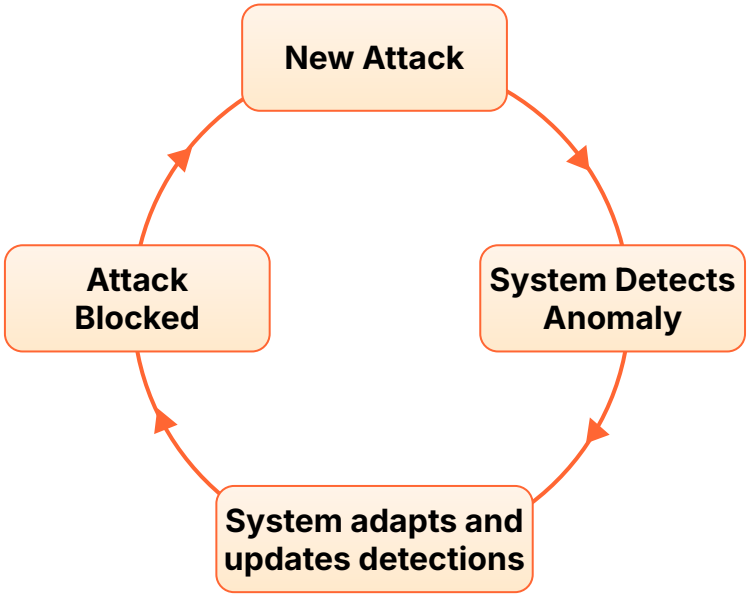


After human

Adaptive Intelligence

Last-gen solutions are flawed because they assume the attacker will eventually give up.

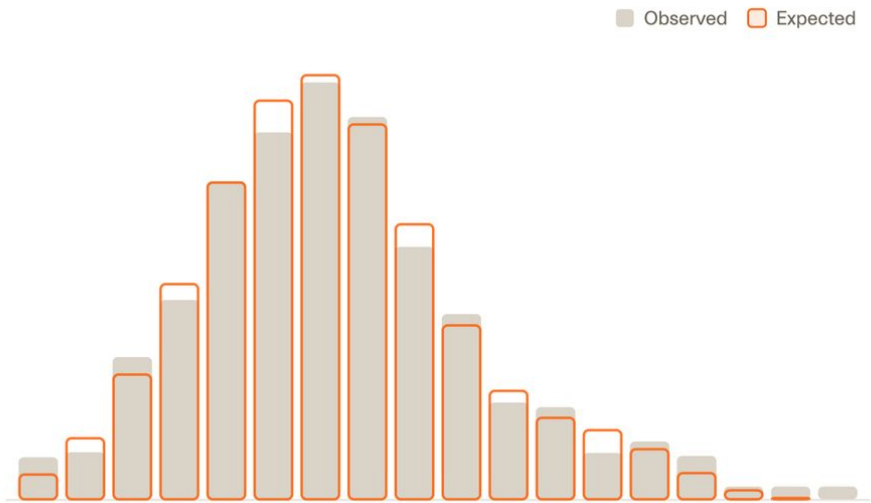
Modern solutions must be adaptive.



After human

Adaptive Intelligence

- We are using Cloudflare's visibility into over 20% of the Internet in new ways
- "Fingerprinting" global traffic to detect anomalies
- Breaking the economic advantage of attackers



After human

Fraud & abuse detection

All fraud and abuse detection comes down to:

1. “Provide me something that can’t be faked.”
2. I make sure it’s not fake.



IDs

Passwords

Selfies

Facial Scans

Liveness Scans

After human

Fraud & abuse detection

All fraud and abuse detection comes down to:

1. “Provide me something that can’t be faked.”
2. I make sure it’s not fake.
3. … and then attackers try to fake it.

~~IDs~~ ~~Passwords~~ ~~Selfies~~

~~Facial Scans~~ ~~Liveness Scans~~



CLOUDFLARE

Any signal that starts in the digital world will be faked with AI

After human

A new kind of signal

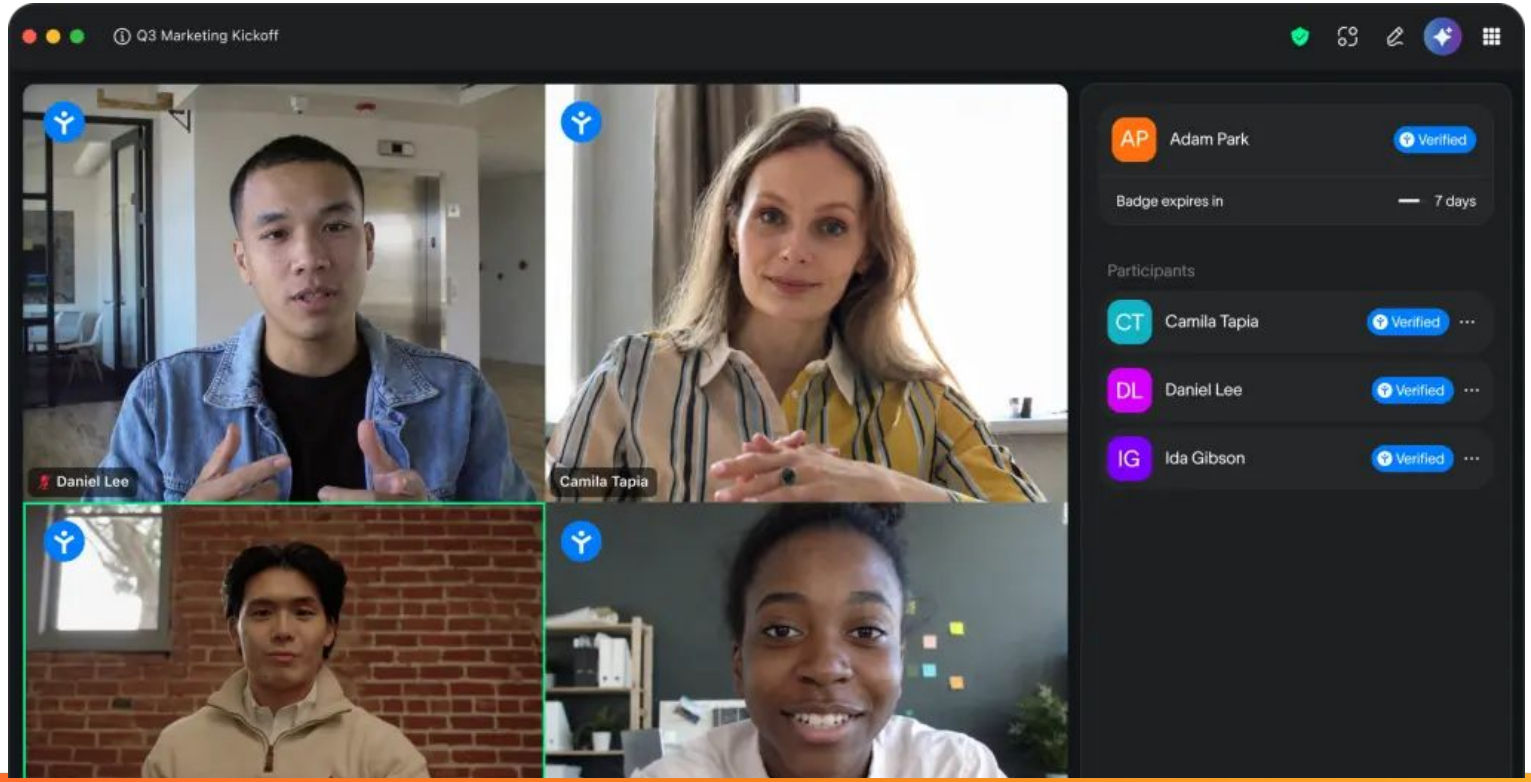
Proof of Human

A privacy preserving **AI proof** verification signal that starts in the physical world



After human

Proof of Human



CLOUDFLARE

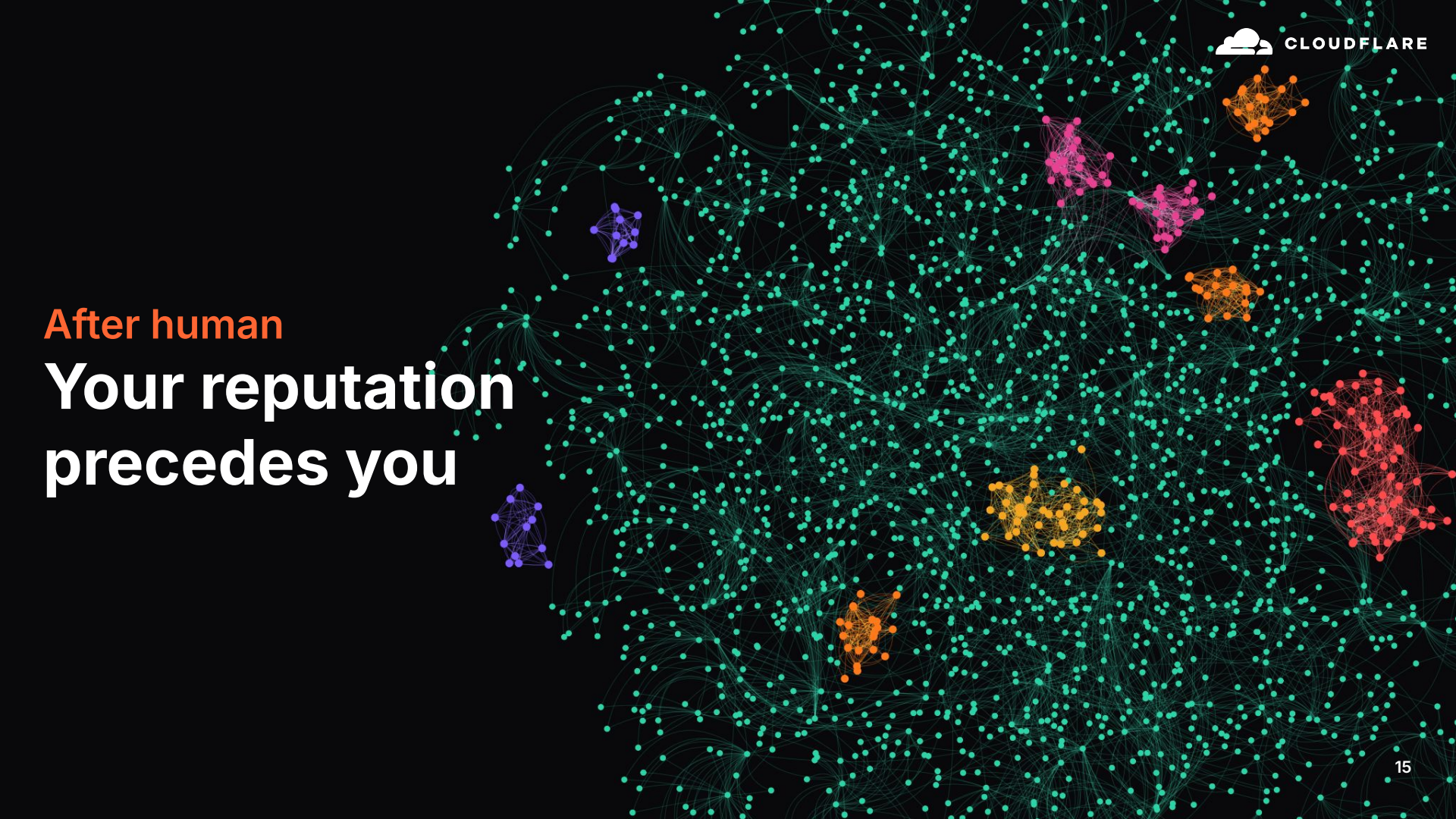
The one test AI can't pass

A new approach to Proof of Human for the age of AI



Wednesday 3:50p

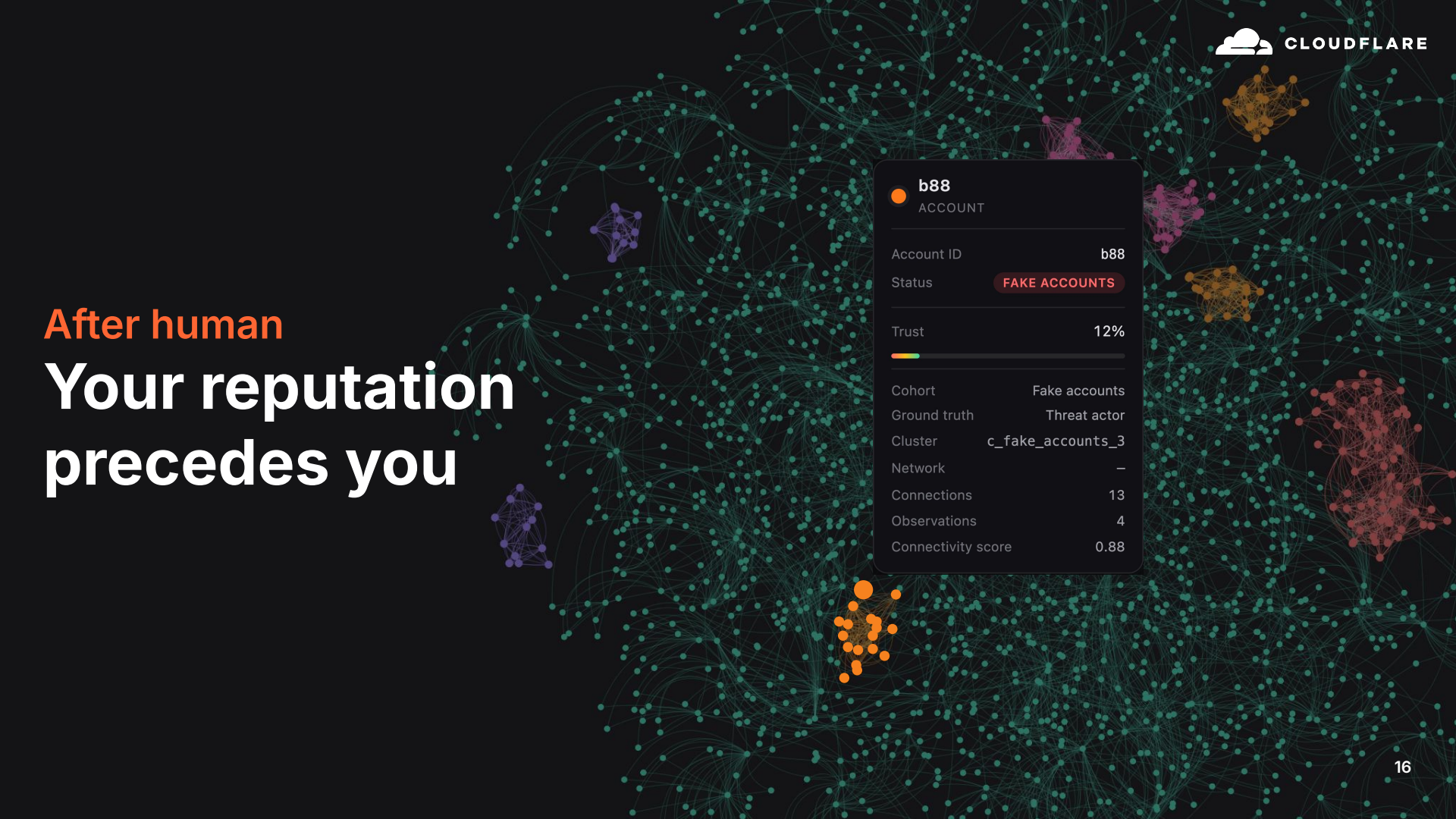
Tiago Sada
Chief Product Officer
Tools for Humanity


CLOUDFLARE
After human
Your reputation precedes you

CLOUDFLARE

After human

Your reputation precedes you



After human

Risk vs. Trust

Risk

How likely is this thing to be bad right now?

- Point-in-time
- Often ephemeral
- Not the inverse of Trust

Trust

What evidence do I have this thing is good?

- Accumulated over time
- Requires reputation
- Used case-by-case

CLOUDFLARE

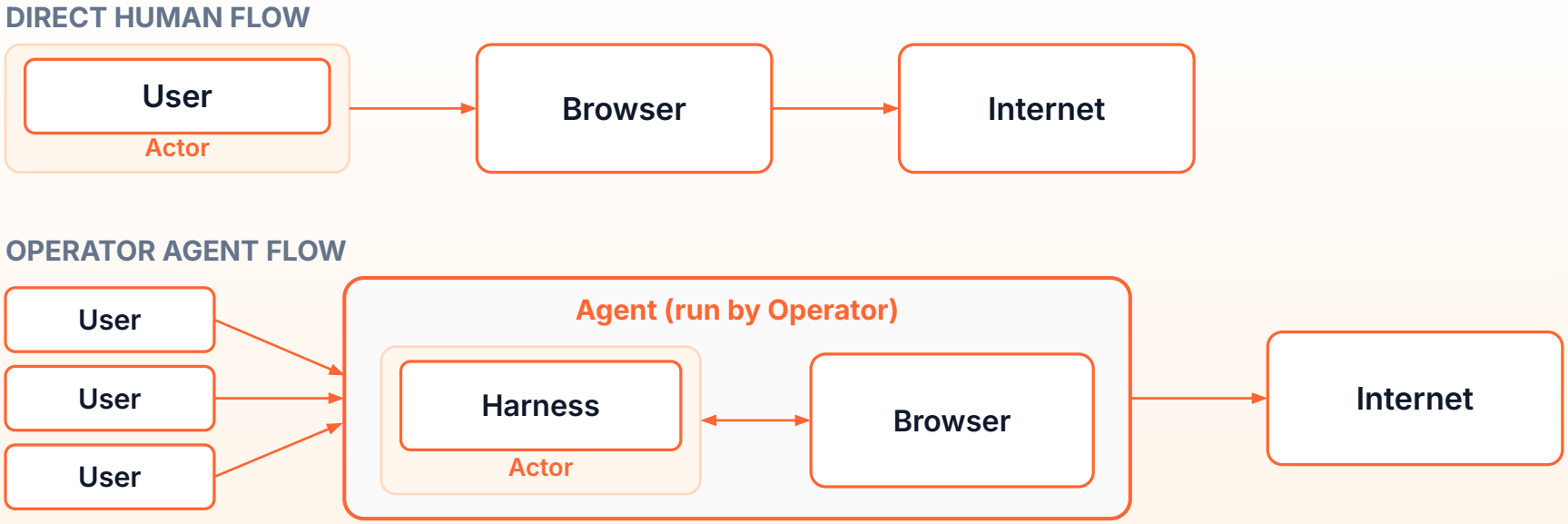
Enter: Personal Agents

After human

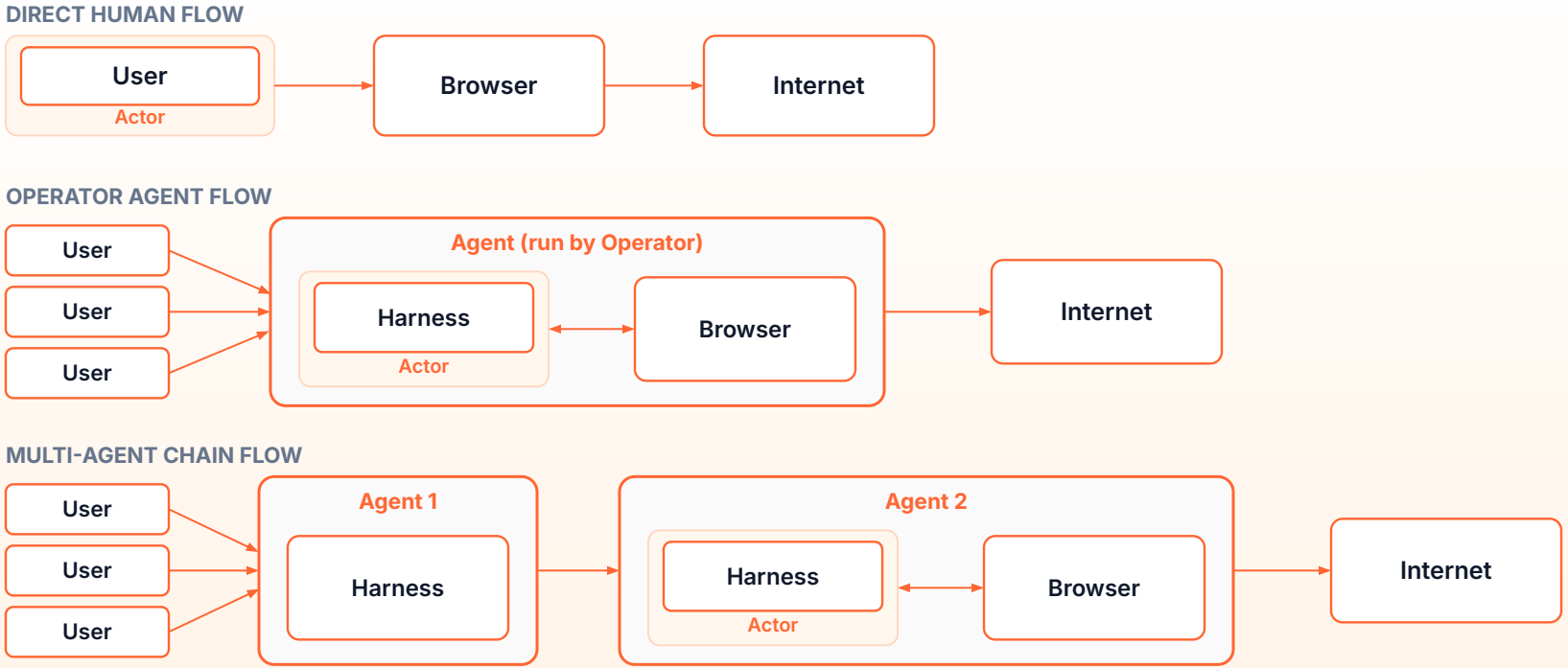
The new problem



The new problem



The new problem



After human

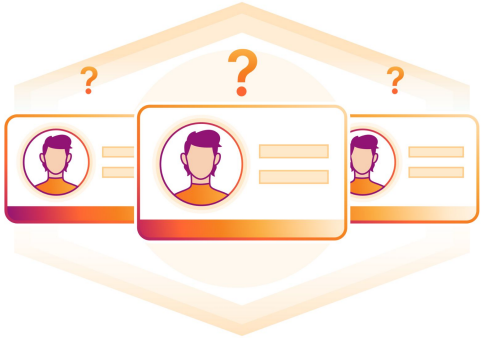
The new problem(s)

Policy circumvention
Avoiding site policy, or making policy impossible to enforce

Risk-based authentication breaks
All Agent activity looks the same; unable to selectively apply verification

Traffic commingling
Sites cannot persistently block one abusive User of many

Unauthorized card users
Agents are not authorized card users, but Users want to give them their card



Human and ID verification breaks
These all assume the person present is the person being verified

Account sharing
Multiple agents can now share a single account, and there is no way to differentiate.

Resource exhaustion
Agents can scale what a human can do, putting costly burden on sites

Friendly fraud
Existing fraud prevention mechanisms can not prove intention or provenance

In almost all of these, the Site Owner bears the cost.

After human

Site Owners have a decision

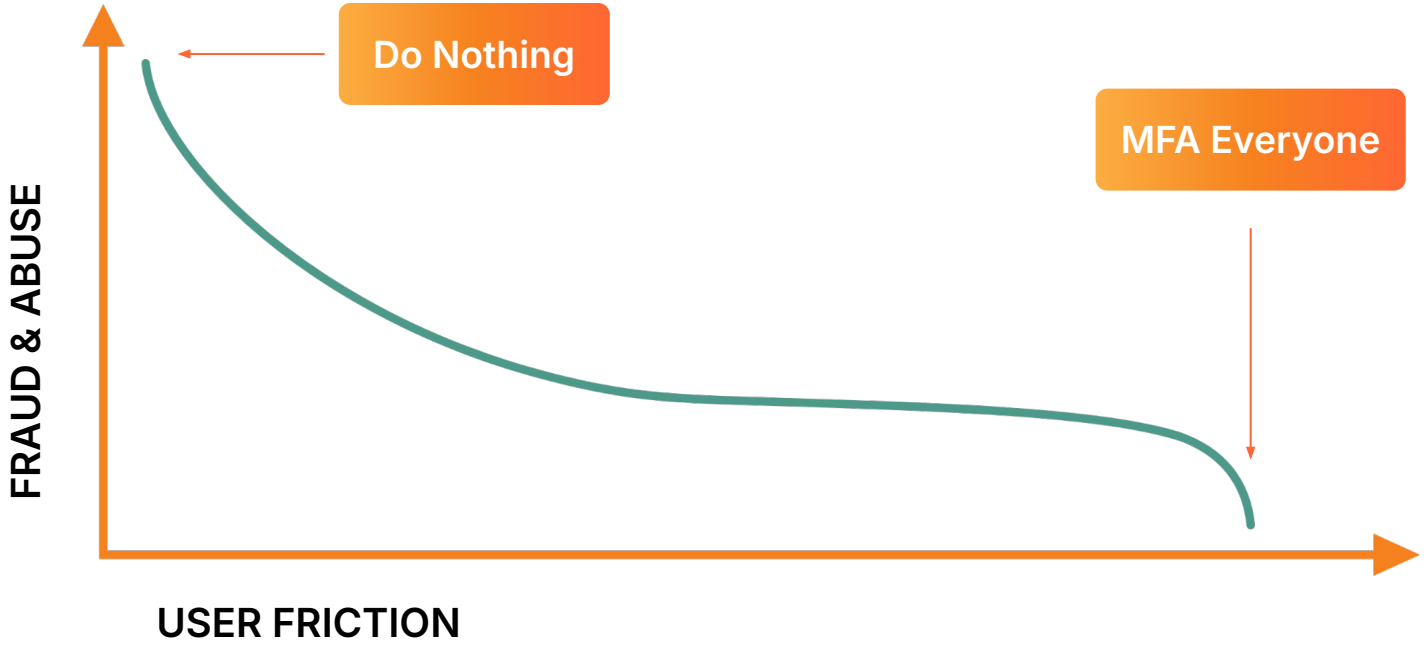
Lock everything down

Ask for more information*

After human

Optimizing friction

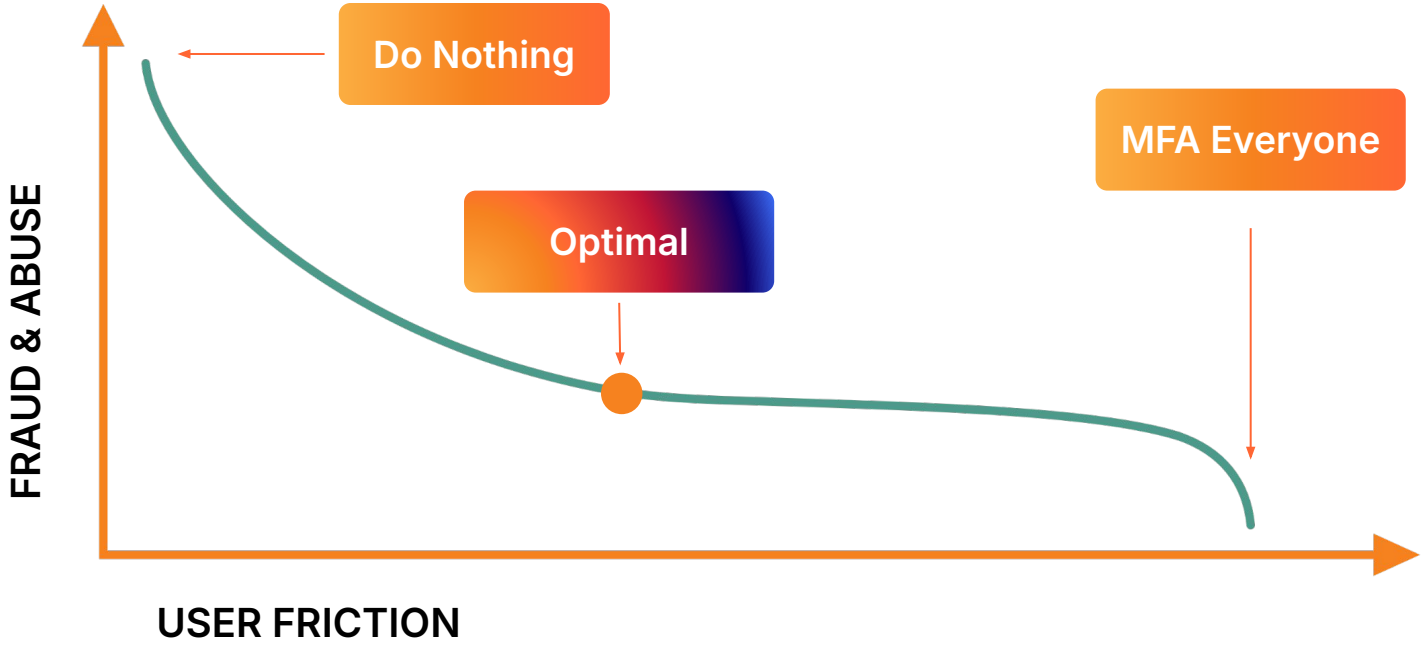
We can already stop abuse. We are optimizing for friction.



After human

Optimizing friction

We can already stop abuse. We are optimizing for friction.

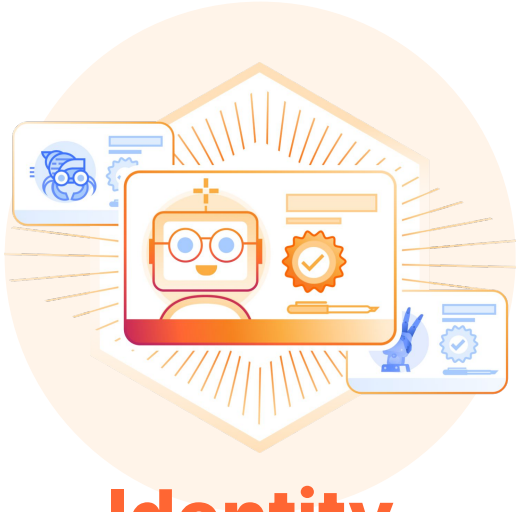


The Intermediation Principle

Agents should enable the relationship between sites and Users to persist across Agents and Operators, not replace it.

After human

What’s missing



Identity

Who is actually acting?



Reputation

What do we know about them?



Accountability

Who answers for the action?

What people have tried

- Captcha ✖
- Selfies ✖
- IDV & KYC ✖
- 3-D Secure ✖
- MFA ✖
- Access controls ✖

Existing solutions are solving the wrong problems.

None of these restore the User and Site relationship lost from the intermediary.


CLOUDFLARE
world
Our proposal:


CLOUDFLARE
world
Our proposal:
Trusted Agent Delegation (TAD)

After human

Trusted Agent Delegation (TAD)

	Identity & Claims	Reputation	Accountability
Exists Today	Web Bot Auth signs requests from Agents and Operators.	Cloudflare Verified tracks Operator reputation.	Signed identity makes the Operator answerable for its traffic.
To Formalize	A persistent User delegation identifier, issued by the Operator or third party.	How a User's reputation travels across Agents and Operators.	How Operators vouch for their Users, and how consequences distribute.

Cloudflare provides the foundation. We need standards and industry adoption to fully enable site owners.

After human

Next steps

For Trusted Agent Delegation (TAD)

White paper and new working group, inviting the Agent ecosystem

For Cloudflare

New Policy for Verified Agents

For World ID

New Proof of Human signal on TAD


CLOUDFLARE
world
Questions?