Darwinium

WHITEPAPER

Intent Intelligence

From Identity to Intent: Understanding Humans and AI Agents Across Every Digital Journey

For fraud, risk, trust and safety, and security leaders at high-volume digital businesses.

Summary

Over the last decade, as online commerce has grown and flourished, fraud prevention has essentially been built around a single question: who, or what is this? Every control in the stack, from device fingerprinting to behavioral biometrics to payment risk scoring, answers a version of it at a particular moment that someone is transacting. This approach generally worked while the actor at the perimeter was a human on a known device, or a bot to be blocked.

It no longer holds. Two shifts have broken the model. Scams and social engineering mean the person passing every identity check is often the victim, taking a harmful action voluntarily, but with nefarious intent initiated by the fraudster. And AI agents now act on customers' behalf, browsing, building carts and paying through channels such as the Model Context Protocol (MCP) that legacy controls typically cannot see. An authorized agent with a valid signature can still be hijacked mid-task. Only about one in four agentic transactions self-declares, and transactions involving agents are rejected nine times as often as other purchases.

The question that matters now is not who is this, but what do they mean to do? The answer that Darwinium is pioneering is called **intent intelligence:** the continuous assessment of what a human or agent is trying to do, built from how they behave and the path they take, not just the credentials they present.

Intent intelligence has four requirements: capture every signal a decision needs, join every surface into one journey, score the sequence, not the transaction, and act at the moment risk appears. Darwinium meets all four because of how it is built: at the CDN edge, with device, behavior, identity and journey signals in one event, with a Journey Transition Probability model that scores every step for every actor, and with MCP Protection that brings an agent's tool calls into the same journey as the human's web activity.

KEY TAKEAWAYS

- Identity checks answer "who." They cannot answer "why," and the costliest fraud now happens after identity has been confirmed.
- Intent is revealed by the shape of a journey: which step followed which, how fast, and how that compares with normal. Session-level scores average that away; step-level scores expose it.
- Humans, bots and AI agents each leave a journey signature. One model can score all three, if it sees all three in one record.
- Agent identity and payment standards (Web Bot Auth, Visa TAP, Mastercard Agent Pay, Google AP2) explicitly leave risk to another layer. Intent intelligence is that layer.
- Architecture decides whether a vendor can deliver this. Running inside the customer's existing CDN is what makes full-journey, every-surface, real-time intent assessment possible without an integration program.

Is this the end of simply asking who, or what is this?

Controls that decide once

Most fraud and security stacks are a series of checkpoints. Bot management at the perimeter. Device intelligence at login. Behavioral biometrics in the session. A risk score at payment. Each was bought to answer one question at one moment, and each is very good at it.

The weakness is structural, not technical. A checkpoint sees a request or, at best, a session. It has no memory of the steps that led there and no model of the steps that should follow. Attackers know this. They spread activity across steps and sessions to stay under per-event thresholds. Fifty promo redemptions across new accounts over three days look fine at every individual moment.

When the person who passes the check is the victim

Authorized push payment (APP) fraud and social engineering have exposed the limit of identity-based controls. The customer is real. The device is known. The credentials are valid. And the customer is on the phone with a scammer, being coached through a transfer.

Often, nothing at the checkpoint is inherently wrong. What is wrong is the journey: a long, benign session that suddenly shifts to a new payee, an unusual amount, an unfamiliar route through the app. The signal is in the transitions, and only a control that scores the transition can see it.

When the customer is an agent

AI agents now shop, book, manage accounts and pay on customers' behalf. Agentic commerce standards from Visa, Mastercard, Google, OpenAI and Stripe are converging on shared protocols, and MCP has become core wiring for how agents call tools against a business's services. Much of this traffic is legitimate. Some of it is not. And most businesses cannot tell the difference.

AGENT SCENARIO • WHY IDENTITY ALONE FAILS

Undeclared agent	Declared agent, hijacked	Declared agent, off-mandate	Malicious automation posing as an agent
Arrives with no identification at all. Most AI traffic today. No signature to verify, no policy to apply.	Valid Web Bot Auth or TAP signature. Session loaded with stolen credentials or steered by prompt injection. Every identity check passes.	Token authorizes one merchant, category or spend cap. Journey hits another. The authorization is real; the actions are not what the user intended.	Credential and card testing, scraping, inventory hoarding, executed through the same tool calls a trusted agent would use.

The market data reflects the confusion. In Darwinium's network, only about one in four agentic transactions self-declares, and transactions involving agents are rejected nine times as often as other purchases. Businesses are blocking legitimate demand because they cannot see intent. In a recent survey, 97% of fraud, risk and security leaders reported an increase in AI-driven attacks, yet only 36% believed they had effective fraud coverage across the full customer journey.

Defining intent intelligence

Intent intelligence is the continuous assessment of what a human or agent is trying to do, built from how they behave and the path they take, not just the credentials they present. It reveals whether each step of a journey is consistent with a legitimate goal, and it acts before a risky step completes.

Three distinctions separate it from the controls most businesses already run.

Every step, not one checkpoint

Intent is not a property of a login or a payment. It is a property of a path. Intent intelligence scores each transition as it happens, from the first pre-authentication visit through account creation, login, account changes, checkout and payment, and carries context forward from one step to the next.

The full journey transition

Several solutions now describe "continuous" assessment "across the session." The distinction that matters is what gets scored. A session score summarizes: on balance, this looked fine. A journey transition score asks: is this transition, in this order, at this speed, one that legitimate actors do not take?

Every actor, one model

Humans, bots and AI agents are usually handled by separate tools with separate views. Intent intelligence treats them as what they are: actors taking steps through the same journeys. A coerced human, a scraping bot and a hijacked agent each leave a journey signature. One model, seeing one record, can score all three.

Four requirements

Intent intelligence should be tested against four requirements. Each is an architectural property, not a feature, which is why they are hard to retrofit.

Requirement 1: Capture every signal a decision needs

Intent is inferred, and inference is only as good as its inputs. A decision at any step should have access to four families of signal, in one event:

SIGNAL FAMILY • QUESTIONS IT ANSWERS
Device Is this the device we know? Emulator, headless browser, data center origin, TLS fingerprint, proxy or VPN, automation frameworks.
Behavior Does this person move like themselves? Typing, mouse, swipe and tilt patterns, across every field and step. For agents: the absence of human signal, machine-regular timing, state-blind actions.
Identity Do the identifiers cohere? Account, email, phone, network and location. For agents: the presence and validity of a cryptographic signature.
Journey Does the path make sense? Which step followed which, how fast, and how that compares with this user, this business and this journey type.

Most stacks have the first three, split across vendors. Almost none have the fourth as a first-class signal, and none can join all four in a single event at the moment of decision.

Requirement 2: Join every surface into one journey

Customer journeys rarely stay on one device. A shopper asks ChatGPT to find a product, the agent does the research from the cloud, and the customer completes checkout on their own phone. A viewer requests a sign-in code on their TV and approves it on their phone. Each is one journey that crosses devices – and, in the first case, crosses from agent to human. If those events land in different tools, the handoff between them, which is where intent is revealed, is invisible. Cross-device observability means joining them at the point where all traffic arrives: the perimeter.

Requirement 3: Score the sequence

A journey model must answer, for every step: is this transition expected, is this path expected, is this speed expected, and has normal changed. It must do so without attack signatures to maintain or labeled training data to collect, because the threats that matter most are the ones that have not been labeled yet. And it must produce explainable output that an analyst can write a rule against.

Requirement 4: Act at the moment risk appears

Intent intelligence that arrives after the payment has settled is forensics. The decision has to be made in-line, in milliseconds, before the request reaches the application, and it has to be calibrated: permit a trusted interaction, verify an uncertain one, challenge a risky one, prevent a clear attack. Blunt block-or-allow controls pay for their catches in lost revenue.

How Darwinium delivers intent intelligence

Darwinium was built in the AI era, after attacks became AI-driven and after customers began sending AI to act for them. Its architecture maps directly onto the four requirements.

Architected at the edge

Darwinium runs inside the customer's existing content delivery network, including Cloudflare, AWS CloudFront and Akamai. The CDN is the perimeter every request passes through before it reaches an application server: web page loads, mobile app calls, API requests and, now, the tool calls that AI agents make over MCP.

Full-journey coverage is inherently part of the CDN deployment. Darwinium can see the first pre-authentication visit and every step after it, across every surface, from day one. New endpoints can be brought under control from the portal, not through a development cycle.

It also makes action possible. Because the decision is made in-line at the edge, Darwinium can permit, verify, challenge or prevent a request in milliseconds, before it touches core systems. Customers retain control over deployment and have flexible data-residency options.

Every signal in one event

Each Darwinium event carries thousands of signals across the four families: device, network and location intelligence; behavioral biometrics read across every field and every step of the journey, not just login; identity attributes and the customer's own payload; and journey context. These are distilled into persistent **Digital Signatures**, comparable representations of device and behavior that recognize returning and related activity even when individual identifiers change or are spoofed.

For AI agents, the same event includes **Agent Intent Detection**, which identifies agents that do not announce themselves through timing, navigation and interaction patterns, and edge validation of cryptographic agent credentials. Web Bot Auth, Visa Trusted Agent Protocol and Mastercard Agent Pay all use the same HTTP message signature primitive, so a single integration covers them all, and the validation result becomes one more signal in the decision.

Journey Transition Probability

Journey Transition Probability (JTP) is Darwinium's universal intent signal, and the capability with no published equivalent in the market. It can run on every event, for every actor, across any flow that can be modeled as a sequence of steps: login, checkout, account changes, returns, support, payouts.

How it works

Every business's traffic has a shape. Real customers browse a category before checking out. Real users log in before they change a password. Real shopping agents fetch a catalog before they build a cart. JTP learns that shape – which step usually follows which, in what time, how often – and scores every new step against it. Most activity looks like the population. The activity that does not is where intent is revealed.

The model is layered at three levels that stack automatically:

1. **Across the organization's entire cohort f users accessing the journey**
2. **Compared to the end users own previous behavior across the journey**

A customer benefits from the global baseline immediately and the signal hardens as their own traffic accumulates. There is no manual tuning, no analyst overhead and no labeled training data required.

The four signals

Per-step probability

Question it answers Is this exact transition (step A to step B) one we'd expect for this journey, for this customer?

What it catches Skipped steps. Unfamiliar transitions. Straight to checkout with no browse.

Path probability

Question it answers Taking the whole journey so far, how likely is this exact sequence?

What it catches Paths no legitimate actor takes. Coordinated abuse patterns.

First-passage time

Question it answers How long did it take to reach this state, and is that normal?

What it catches Programmatic speed. Hesitation in the wrong places.

Transition-rate change

Question it answers Has this customer's normal shifted over the last N events?

What it catches New attack patterns, new bot variants and coordinated campaigns, before any labeled incident.

What it catches, by actor

The same model flags very different threats depending on who is taking the step. This is the practical meaning of "every actor, one model."

ACTOR • JOURNEY SIGNATURE	
Human, coerced (scams, APP fraud)	Mid-session shift to scam-like patterns. Sudden payee or payout change after a long benign session. Path probability collapses.
Human, account takeover	Post-login behavior deviates from the historical user's path. The real user always did X; this session does Y.
Human, promo and returns abuse	Improbable paths to discount and refund endpoints. Bracketing, wardrobing and coupon stacking surface as low path probability.
Bot, scraping	Systematic enumeration with no engagement signals. Identical paths from many devices.
Bot, credential stuffing	Near-zero first-passage time to the login endpoint. No normal pre-authentication journey at all.
Bot, inventory hoarding	Rapid cart-add loops on limited stock. Cart velocity in the tail of the distribution.
Agent, undeclared	State-blind actions with no prior UI flow. Low time-between-step variance. The same agent fingerprint across "different" identities.
Agent, hijacked or prompt-injected	Journey diverges from the declared mandate. Cart contents or payment step inconsistent with what the user signed off on.
Agent, outside its scope	Token authorizes one merchant or category; the journey hits another. Spend caps exceeded across merchants the token never covered.

MCP Protection

MCP has moved from experiment to production infrastructure quickly, and most WAF, bot-management and identity tools cannot see it. They were built to inspect page loads and REST calls. They cannot see which tools are invoked with which arguments, and have no concept of the non-human identities driving that traffic.

MCP Protection from Darwinium brings that traffic under the same edge control plane that governs web, mobile and API. Darwinium stitches an agent's prompt, its ordered MCP tool calls – with payloads visible (SKUs, cart items, totals) and geolocation on every call – and the customer's eventual human checkout into a single journey with one journey ID. When a shopper's AI assistant searches a merchant's catalog through MCP and the shopper then checks out on their phone – Darwinium sees one journey, not two unrelated sessions in two different tools. Because JTP scores every event, it scores every tool call, so the same intent signals that flag a coerced human flag a hijacked agent, with no signatures to maintain.

Businesses can verify agent credentials, monitor what the agent does and require additional checks before a higher-risk action, such as a payment, goes through. In practice that means two paths to the same destination: a trusted, authenticated agent with a journey consistent with legitimate intent completes prompt-to-payment with no human in the loop; a lower-trust or unverified agent is contained into a web checkout where Darwinium's full profiling stack applies and a human-present step can be required before money moves.

Where intent intelligence sits alongside the standards

The agentic commerce ecosystem is producing identity and authorization standards at pace. Each answers one question well, and each explicitly leaves risk to another layer. Google's AP2 states that it "does not aim to be opinionated about risk/fraud handling," for example.

Web Bot Auth (RFC 9421) / Visa Trusted Agent Protocol

What it answers Is this a known, signed agent?

What it leaves open What the agent is actually doing right now.

What Darwinium adds Validates the signature at the edge and layers behavior and journey context on every signed request.

Mastercard Agent Pay

What it answers Is this agent authorized to pay?

What it leaves open Whether the journey is consistent with that authorization.

What Darwinium adds Scores the path from discovery to payment against normal.

Google AP2 / delegated tokens

What it answers What did the user authorize?

What it leaves open Whether the session was coerced or compromised before money moved.

What Darwinium adds Continuous verification of the bind on every action.

Agent identity registries

What it answers Who is the human behind the agent?

What it leaves open The bind is point-in-time. No signal between binds.

What Darwinium adds Makes the bind continuous: is this session still consistent with the bound identity?

Darwinium is an effective partner to these standards. These standards focus on the identity of agentic traffic. Darwinium layers on intent.

Solution readiness: are you intent enabled?

"Intent" is now a common word in fraud and bot-management positioning. These are the questions you need to ask in order to assess your solution readiness.

01 **What journey signal do you produce?** Ask to see it in the product. Most vendors score the current request in isolation and have no model of what should come next.

02 **Do you score the step or the session?** Ask how the product would locate the exact transition where a coached customer's path turned.

03 **Do you see MCP tool calls, and are they in the same journey as the human's web activity?** Standalone MCP endpoint protection is not the same as joining the agent's trail to the checkout it led to.

04 **Does the same model score humans, bots and agents?** Or is the agent story a separate product with a separate view?

05 **Where does the decision run?** Inside our CDN, in-line, before the request reaches the application? Or in a downstream API call at the checkpoints we remembered to add?

06 **What does deployment involve?** Tags on every page and SDKs in every flow, or a configuration in infrastructure we already run?

07 **Does it work on day one?** Or does it need months of labeled data before it produces a usable signal?

08 **Can an analyst see why something was flagged and write a rule on it?** Explainability is what turns a signal into an operation.

Conclusion

The digital perimeter has spent two decades asking who is this. That question still matters, but it is no longer sufficient. The costliest fraud often happens after identity has been confirmed, and a growing share of legitimate traffic arrives from AI agents that identity checks cannot govern on their own.

Intent intelligence answers the question that identity cannot: what does this human or agent mean to do, at this moment, and across the digital journey. It requires every signal in one event, every surface in one journey, a model that scores the sequence, and the ability to act in the moment. Those are architectural properties, and Darwinium was built with all four: at the edge, in the customer's existing CDN; with device, behavior, identity and journey signals in one record; with Journey Transition Probability scoring every step for every actor; and with MCP Protection joining the agent's journey to the human's.

That is why Darwinium is the Intent Intelligence Company.
Reveal what's real, across humans and agents, at every step.

See intent intelligence on your own traffic.

Start in monitor mode, see what your human and agent traffic looks like today, and switch to enforcement with a single rule change.

Request a Demo

About Darwinium

Darwinium is the Intent Intelligence Company, built to protect and grow revenue at every touchpoint by understanding user behavior and intent. Its edge-based risk decisioning platform distinguishes trusted from risky human and AI behavior across the entire customer journey, resulting in 50% less fraud and 40% greater operational efficiency. From first interaction through high-risk actions and payments, Darwinium protects the moments that matter most, so digital leaders can confidently grow and protect revenue. Learn more at darwinium.com.