IMPACTVC Mozilla Ventures

AN IMPACTVC × MOZILLA VENTURES PAPER

Trustworthy AI in Action

An exploration of the investment opportunity in open, secure and trustworthy AI

ABSTRACT

This paper explores what it means to build AI that is open, secure and trustworthy as part of shaping a positive AI future. The emphasis throughout is on the technology itself, going deeper into the stack beyond the application layer. The paper uses the term 'Trustworthy AI', grounded in Mozilla Foundation's research, and also used by the European High-Level Expert Group on AI: AI that is demonstrably worthy of trust, designed with accountability, human agency, and individual and collective well-being in mind. It draws on real-world examples of companies in Mozilla Ventures' portfolio to show how these principles can play out in practice.

AI's transformative potential is well established, with models accelerating scientific discovery, widening access to healthcare, and lifting productivity across industries. But while industries are widely adopting AI, there is rising awareness that we need these systems to be trustworthy. The emerging area of focus is the gap between AI's capability and the ability to deploy it reliably.

While regulation is setting the floor for trustworthy AI, enterprise procurement is driving the market. Buyers are increasingly screening for auditability, transparency, and fairness before they deploy AI. The lack of guardrails has led to risks including hallucinations, bias, and harmful content, and is increasingly posing a procurement barrier. A focus on open, secure and trustworthy AI can address these risks while also challenging the dominant paradigm of AI development through scale and centralisation.

VCs sit at a clear leverage point. AI is increasingly central to VC investing, with 61% of global venture investment going into AI in 2025. But VCs also exercise outsized influence in AI development and the setting of industry norms. Venture capital allocation shapes which founders get backed, which architectures get built, and which values get embedded into products at the earliest and most consequential stage.

Reframe Venture, ImpactVC, and Project Liberty Institute's 'VC Perspective' paper of March 2026 showed 73% of VCs surveyed believe companies with stronger Responsible AI practices will be more financially successful, and 91% see significant investment opportunity in AI safety and accountability infrastructure itself.

> Looking across these themes, we see a market need for Trustworthy AI, that VCs can play a role in backing and shaping its development, and that many VCs themselves see that opportunity. What we think is needed next to accelerate this movement are examples: success stories that show how Trustworthy AI companies deliver real-world impact and commercial traction together.

This paper starts to do that, by exploring what Trustworthy AI can mean in practice through the lens of six examples in the Mozilla Ventures portfolio. These companies are innovating on the technology itself, testing new paradigms, and reimagining incentives to build technologies that are open, secure and trustworthy while keeping the advancement of human agency at the core.

We aim to demonstrate some of the impact and commercial mechanisms at play in this sector through these companies. They show that these mechanisms exist, but do not provide systematic evidence of the thesis at this point. Over time we will look to expand the evidence base, and as we do so we are also looking to surface the tensions that must be navigated in order to realise the potential of such approaches in practice.

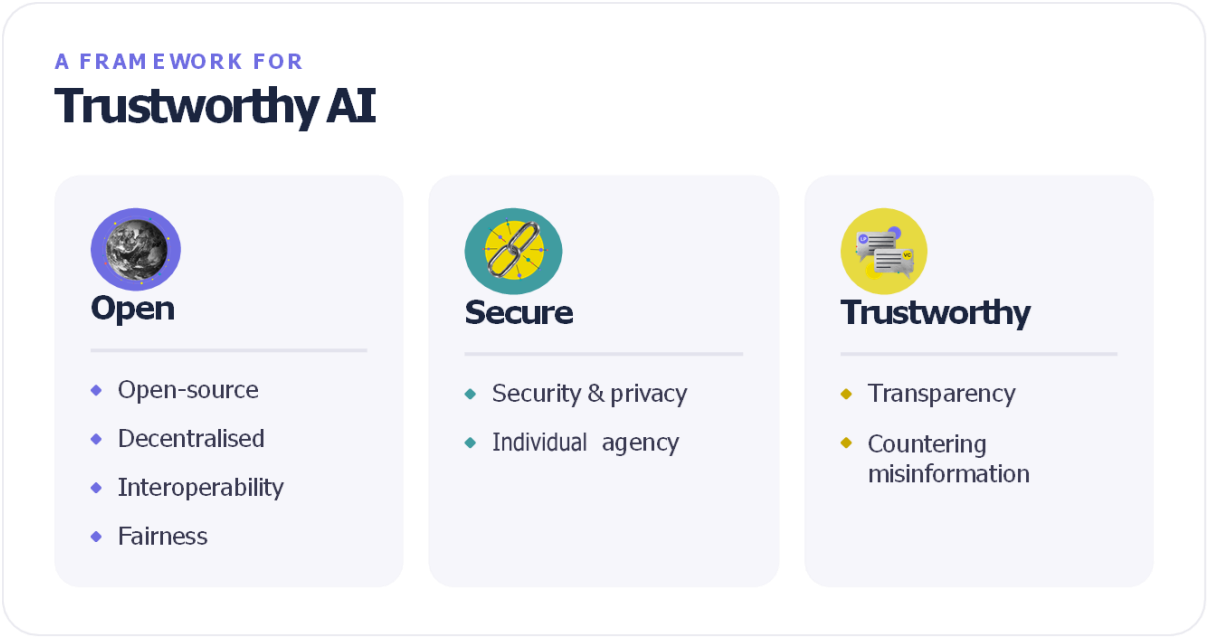
SECTION ONE | IMPACT MECHANISMS

How Trustworthy AI can shape a positive AI future

Mozilla's Manifesto has outlined a concrete commitment to an open, global and inclusive internet since 2007. The principles in that manifesto laid the groundwork for this exploration of Trustworthy AI, built on three pillars: **Openness**, **Security** and **Trustworthiness**.

These pillars are built on top of Responsible AI approaches and behaviours, and look to go further in exploring how we can shape an AI system that is in service of humanity. How much each pillar matters in a given instance depends in part on the application-layer use case, but the principles themselves concern the technology rather than its end use. They are about advancing privacy, fairness, trust, safety and transparency in how AI systems are built, creating a sound foundation for impactful applications, with attention on the underlying systems as much as on the products built over them.

These pillars describe the mechanisms through which underlying AI technologies can shape real-world impact. They are not an investment filter in and of themselves; the investment filter is whether trustworthiness is load-bearing in the business model, meaning the company's success is dependent upon the trustworthy characteristic. But these pillars can play a key part in building an investment thesis by illustrating potential impact pathways to evaluate.



OPEN

Models, tooling and infrastructure that enable alternatives to closed AI

The structural case for these interventions is concentration: a few firms control the models, compute, and data everyone else depends on, which narrows who can innovate, who can audit, and who benefits. Companies here build alternatives by releasing code, weights, and datasets others can inspect and improve (open source); by shifting training and deployment away from single incumbents toward distributed data and infrastructure (decentralisation); and by building tools that run across vendors so users keep choice and control instead of locking into one stack (interoperability).

Approaches that improve efficiency and lower the cost of compute do the same work from another angle, putting capability within reach of smaller players. The same openness widens representation as well as access: efficient, low-cost architectures let emerging markets and overlooked languages participate, and community-anchored builders bring in people the dominant labs design around. This disperses power that is currently pooling in very few hands.

SECURE

Enhancing safety, security and privacy for users, data and systems

As AI holds more sensitive data and acts more autonomously, the person on the other end is exposed to extraction, manipulation, and harm. Companies here protect them through privacy-respecting data practices, consent and data-control mechanisms that keep people in charge of their own information, and fraud and abuse detection that keeps platforms and infrastructure safe (security and privacy); and they give people meaningful control over how systems and agents act on their behalf instead of treating them as passive subjects (individual agency). This addresses weak data governance and the safety risks of systems deployed without adequate guardrails. Building security and privacy in at the core is also what opens regulated markets, since healthcare, financial services, and government procurement increasingly cannot adopt AI without it; as rules tighten and AI-enabled attacks grow more capable, secure systems shift from optional to a precondition for deployment, in a segment that remains underserved.

TRUSTWORTHY

Reliability, accountability and transparency in how AI behaves

Reliability comes first: a system that performs inconsistently or cannot show its working cannot be trusted with anything consequential, and as AI shapes more of what people read, decide, and believe, the value of trust climbs. Companies here make AI legible and dependable through

evaluation, monitoring, and governance tooling that exposes how systems behave and lets enterprises demonstrate compliance against external standards (transparency); and through content-integrity tooling that detects scams, manipulation, and coordinated distortion at scale while leaving the hard judgement calls to human reviewers (countering misinformation). Both answer the same problem: accountability that is asserted but never shown. When it is demonstrated, the world gains AI deployments and an information environment that earn trust by being checkable against evidence.

Two of the challenges in Mozilla's framework sit partly beyond what any single product can solve. Industry norms in the form of shared incentives and expectations that govern how AI gets built, and the exploitation of workers and the environment, from data-labelling labour to the energy and water cost of large-scale training, are properties of the system rather than gaps one company can close. The pillars above touch on them: many decentralised approaches under Open are far less compute- and energy-intensive than centralised training, and several fairness-driven companies improve conditions for the workers and communities whose data and labour underpin AI. Addressing norms and exploitation in full will take coordinated action across policy, procurement, and standards that reaches beyond any portfolio; we treat them as context any investor's thesis must take into account, but a subject of systemic change beyond the scope of one investment.

Investors should note that while these pillars reinforce each other in many cases, they can be in conflict and in those situations the tension must be considered and navigated case-by-case. **Open and Secure can sit in tension**: openness, open weights, and decentralisation can widen the surface for misuse and remove the chokepoints that make safety enforceable, while the strongest safety and security products lean on the centralisation and closure that openness resists. The key question is at what capability threshold the misuse risk from openness starts to outweigh its auditability and decentralisation benefits. Similarly, **Open and countering misinformation can be in tension**, because deciding what counts as authentic concentrates a kind of epistemic power and pulls toward centralisation.
Any company must choose how to navigate these tensions, and any investor must take a set of considered bets along these trade-offs. Flower Labs advances openness of method while keeping data decentralised, so collaboration never requires pooling sensitive information. Musubi meets the authenticity problem directly: instead of imposing a single arbiter of truth, it gives platforms adaptable tooling they direct themselves, with human reviewers on the hard cases, pushing judgement toward the edge instead of the centre.

SECTION TWO | COMMERCIAL MECHANISMS

Trustworthy AI business model dynamics

To reach transformative scale, Trustworthy AI companies need commercial success alongside impact. ImpactVC's 2025 paper set out four levers through which putting impact at the core can drive that success: **customers, regulators, talent,** and **capital**.

The evidence that these dynamics hold for Trustworthy AI is still developing and mixed at market level, so we read the levers as the mechanisms through which trustworthiness *can* produce commercial advantage rather than proof that it reliably does, with the case studies as companies testing and navigating them. We hypothesise that these mechanisms include:

01 Customers

Trust underpins enterprise procurement. Privacy and transparency become product features that open high-value, sensitive, and regulated markets competitors cannot enter and support adoption within them.

02 Regulators

Building security, accountability, and trust into the architecture from day one lowers the cost of adapting as frameworks such as the EU AI Act broaden and tighten. Regulatory alignment can be a precondition for selling in sectors like healthcare.

03 Talent

Scarce AI engineers and researchers can be selective about who they build for. Companies committed to openness, accountability, and human agency can be better placed to attract and retain them.

04 Capital

The least evidenced lever. The March 2026 'VC Perspective' paper shows interest at the attitudinal level: 73% believe stronger responsible AI practices will improve financial performance, and 91% see significant opportunity in AI safety and accountability infrastructure.

NAVIGATING TENSIONS

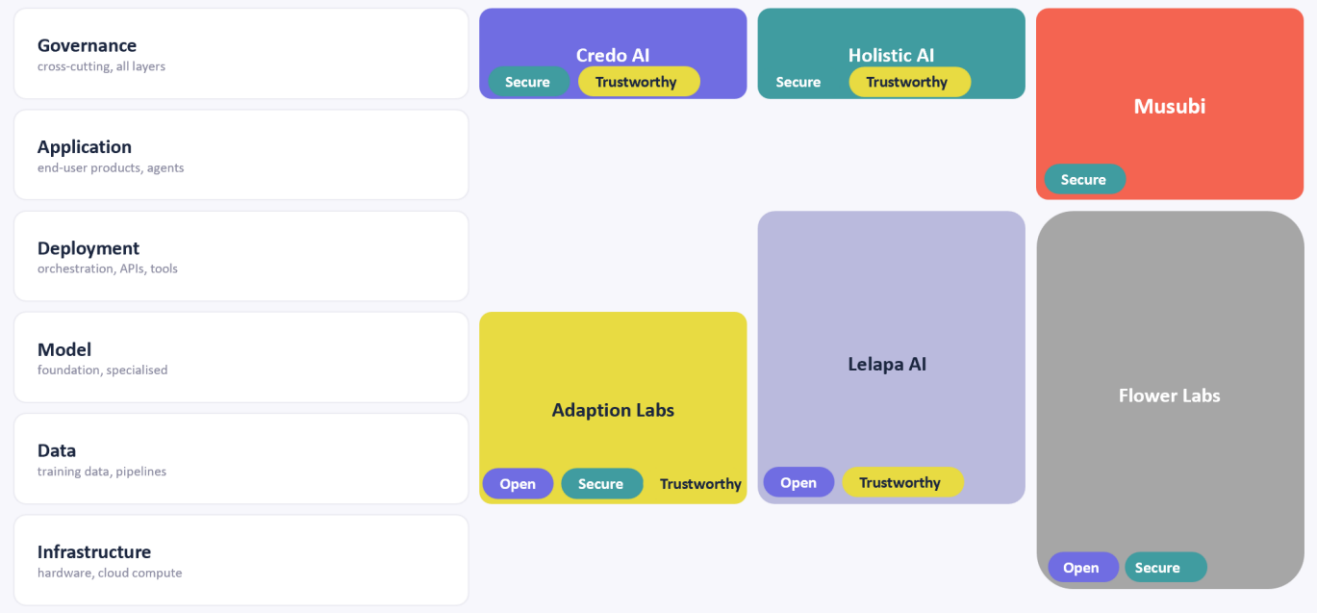
However, there are tensions present in each case, and Trustworthy AI companies also face key challenges:

- **Trust can be treated as friction** that slows speed, though the payoff is durable growth in markets that otherwise stay closed.
- **Incumbents may absorb trust features** as platform functions, though they cannot offer the neutrality that a self-assessed function structurally lacks.
- **Open approaches can struggle to monetise**, leaking value to whoever runs the complementary infrastructure, though open methods can work as a distribution and trust wedge, monetised through a defensible complement such as managed hosting or proprietary data.
- **Regulatory momentum can reverse**, as with recent EU AI Act deferrals and a deregulatory US posture, though regulation is often a floor, while enterprise reputational and operational risk drives demand.

The case studies show how companies navigate these tensions, including where a trust characteristic unlocked a specific market dynamic, such as Flower Lab's federated architecture opening access to regulated healthcare data.

THE SIX CASE STUDIES AT A GLANCE

The six companies explored in the companion case studies span the AI stack from infrastructure to governance layer, showing how the impact and commercial mechanisms set out above can play out in practice at different places in the stack:



APPLICATION + GOVERNANCE LAYER

Musubi

Secure

FOUNDED	HQ	STAGE	RAISED
2023	California, USA	Seed	$8.5m

INVESTORS

J2 Ventures, Shakti Ventures, Mozilla Ventures

Widespread adoption of AI has made online safety even more challenging. Musubi leverages AI to fight fraud and moderate content across social platforms and marketplaces.

THE PROBLEM & SOLUTION

The increased pervasiveness of AI in everyday life has brought disproportionate risks in the form of spam, scams, fraud, harassment and impersonation, resulting in financial loss, political manipulation and social vulnerability. Bad actors are now leveraging AI to specialise their attacks for each platform, quickly learning and adapting to evade standard defences.

Musubi recognises that trust and safety teams require AI-enabled moderation that is customised, contextualised and adapts to their platform. Rather than a single solution, it offers organisations a range of tools to govern, oversee, audit, control and improve their own AI systems. The toolkit spans different approaches: large language models, models that learn directly from human moderators, and AI that routes decisions to other AI models. For example, AiMod automates safety checks and decisions in real time, learning new fraud patterns from behavioural and metadata as well as from human moderator teams. It is built to analyse the whole account rather than one-off content on social platforms. The combined range of solutions deliver decisions with an error rate 10x lower than human moderators while reducing moderation overhead by 84%.

This approach matters in a fast-moving environment. As agentic AI becomes more prevalent, platforms benefit far more from a set of tools they can adapt as threats evolve than from a single solution that needs replacing soon after deployment.

Musubi is used by applications, social media platforms and online marketplaces. It helped Bluesky implement a tailored trust and safety system aligned with the platform's policies, enabling it to remain safe through its surge past 40 million users.

Musubi's tools protect more than 300 million end users, helping ensure that people have better experiences online. By automating moderation, it also protects the tens of thousands of human moderators who would otherwise have to view traumatising content.

> We want to create an ecosystem of tools for enterprises to govern, oversee, audit and control their own AI solutions so that online interactions are authentic and meaningful – rather than offer a single product or solution.
>
> **Tom Quisel** · Co-founder & CEO, Musubi

COMMERCIAL MECHANISMS: TRUSTWORTHY AI BUSINESS MODEL DYNAMICS

Customer Trust

Musubi's toolkit enables customised trust and safety at reduced cost, a key value proposition for platform operators. With safety issues evolving and new regulations being adopted, organisations' needs are increasingly complex and varied. Musubi is well positioned to meet needs across systems and adapt to changing environments, reducing friction in deployment and driving adoption. Its tools continuously learn patterns, improving in accuracy and effectiveness the longer they are in use. This distinguishes Musubi from solutions that rely on static rule sets or periodic manual updates.

Regulatory Alignment

Governments and regulatory bodies across key markets are rapidly building awareness and enacting legislation around digital safety, data control and platform accountability, from the EU's Digital Services Act to the UK's Online Safety Act. Certain countries are also taking a clear stance on what content is acceptable and what action platforms must take.

For Musubi, regulation acts as a floor to build from. Its tools ease compliance, while letting each platform apply its own interpretation and risk tolerance. Musubi's value proposition is its contextualisation, for example, reflecting differing legal definitions of 'underage' from one country to the next. Mapping directly onto regulatory priorities strengthens Musubi's position in the market and helps customers mitigate reputational cost and risk while building durable trust with their users.

NAVIGATING TENSIONS

Trust and safety is often treated as a cost centre, and the long-term health of a platform can be hard to tie back to tangible outcomes. There is also pressure on platforms to grow at all costs, which makes it important to distinguish healthy growth from growth in general. Musubi leans towards letting customers find their own healthy balance between moving fast and acting carefully, rather than being prescriptive.

Over the longer term, a safer platform is a healthier one. When users trust a platform they stay longer, spend more and recommend it to others, making trust and safety central to a platform's long-term health.

MODEL + DATA + DEPLOYMENT + INFRASTRUCTURE LAYER

Flower Labs

Open Secure

FOUNDED	HQ	STAGE	RAISED
2023	Hamburg, Germany	Series A	$23.6m

INVESTORS

Factorial Capital, Betaworks, Y Combinator, Mozilla Ventures

Building and training AI systems currently depends on accumulating large volumes of data and compute. Flower Labs is re-envisioning this notion with open-source infrastructure for collaborative AI, so many organisations train and run AI together across decentralised data and compute.

THE PROBLEM & SOLUTION

Today, AI models are trained on large, centralised data pooled into massive data lakes, concentrating power within few, large organisations with access to infrastructure and data-acquisition relationships.On the data side, estimates show public data in English amounts to 15 trillion tokens, whereas non-public data is equivalent to almost 2,000 trillion tokens. So, AI models are currently accessing less than 1% of the world's data. Similarly for compute, data centres are sitting idle for 30–40% of compute time, and sometimes even 70%.

Flower Labs' mission is to unlock these limits by fundamentally shifting from a large, mainframe system to a grid-like, collaborative network. Decentralisation of data and compute allows secure access to large amounts of non-public data while also ensuring efficient, cost-effective use of compute. Flower Labs' leading open-source framework and ecosystem for 'federated' or 'collaborative' learning trains models across devices, simultaneously, with only encrypted model updates passing between them, so raw data never leaves its source.

The strategy sits around three building blocks - Flower SuperGrid, the federated AI platform that makes training models using decentralised data and compute effortless; the Frontier training stack, which trains decentralised models such as Flower Labs' own 'Lizzy 7B LLM' and is being deployed by the US Department of Energy to train a 70-billion-parameter model across three national labs; and **Flower Agent**, a collaborative agent that can use distributed infrastructure to communicate across agents in other organisations with their own data views.

This has enabled access and application of AI, not least in regulated sectors like healthcare and finance, where personal data often sits behind firewalls due to regulatory constraints. For example, Banking Circle uses Flower Labs' federated learning to train its AI-powered transaction-monitoring system across European and US markets without moving data across borders. This has improved performance, including a 65% increase in precision and a 25% increase in recall while screening transactions for suspicious activity.

> *The technology allows you to access vastly more data than you could before… as of today, less than 1% of the world's data is being used for AI training.*
>
> **Daniel Beutel** · Co-founder & CEO, Flower Labs

NAVIGATING TENSIONS

A common misconception that Flower Labs' founders navigate is that synthetic data, which is artificially generated from data you already hold, can unlock scale on its own. However, in practice, co-founder Daniel Beutel explained that synthetic data can only remix patterns already present in the data, meaning it misses valuable edge cases found in a broader real-world dataset. For example, in manufacturing, collaborative learning captures the rare failure cases that synthetic data cannot replicate because they exist only in other firms' datasets.

COMMERCIAL MECHANISMS: TRUSTWORTHY AI BUSINESS MODEL DYNAMICS

Customer Acquisition

The open-source framework has enabled Flower to become the industry standard in federated AI, with the world's largest community hub of 7,000+ developers and 2,500+ projects. While the code is open-source, Flower's enterprise model is paid and customers typically want the complete, integrated stack. Flower Labs's offering is differentiated by providing the full platform stack across applications on top of the platform, operating the platform, integrating it with other components, monitoring, and enterprise support. Customers can train their own models through decentralised training, run federated analytics such as private benchmarking, or deploy collaborative agents that each see a different dataset. For example, using SuperGrid, docport has been able to benchmark and improve GP practices such as how long a given action takes or patterns across types of visits, enabling operational improvements.

Regulatory Alignment

For regulated sectors, federated learning directly aligns with regulatory requirements. For instance, Secure Data Environments by the NHS require confidential data and compute as a precondition to deploying AI. This safeguards a key material risk and avoids costs as regulations tighten. This opens healthcare, financial services, telecoms and government workloads, where high-value and sensitive data sits. Importantly, though, because the deeper value lies in data access, Flower's relevance is not limited to privacy-driven or regulated buyers.

GOVERNANCE LAYER

Credo AI

Secure Trustworthy

FOUNDED	HQ	STAGE	RAISED
2020	California, USA	Series A+	$41.3m

INVESTORS

CrimsoNox Capital, FPV Ventures, Decibel VC, Mozilla Ventures

AI adoption by enterprises requires trust. Credo AI allows enterprises to embed AI solutions into their operations with full oversight and accountability.

THE PROBLEM & SOLUTION

As AI adoption accelerates across industries, enterprises are shifting from rushing to deploy to actively exploring AI they can trust — scrutinising for legal, reputational, and operational risk. Regulation has tightened in parallel, but it is increasingly internal policies, board-level mandates, and industry best practices that are setting the bar. The gap is that governance, risk, and compliance processes have not kept up with the pace of AI, remaining manual, fragmented, and within a siloed team.

Credo AI addresses this gap with an enterprise AI governance platform, offering organisations a single command centre to map every AI system (including Shadow AI) across the business, and continuously monitor against it. Its contextual approach enables organisations to define 'measurable trust', identifying specific needs for each case, rather than imposing a generic checklist. It has built visibility into approximately 1,800 AI risks, with 85% of those risks having some mitigation action associated within the platform. It enforces policies aligned to the EU AI Act, NIST AI RMF, ISO 42001, SOC 2, and integrates natively with Microsoft, IBM, Databricks, and the leading MLOps and LLMOps tools.

In healthcare, Credo AI is applied to ensure patient-care interventions draw on the right kind of AI. Automation alongside human-in-the-loop builds reliability and sensitivity in each case, enabling healthcare access to underserved populations.

Credo AI's mission is to ensure AI is always in service of humanity. The lens it brings is about making AI accessible to everyone, with trust at the centre. Founder Navrina Singh served three years on Mozilla's board, setting up Mozilla.ai, reflecting a deliberate choice to lean into AI with trust rather than fear.

> *We don't wake up afraid of AI, but with excitement about how we can get this right and make it available to everyone.*
>
> **Navrina Singh** · Founder, Credo AI

NAVIGATING TENSIONS

Governance is often contrasted as a cost centre that slows you down. Credo AI reframes this as: governance is critical to ensure durable, sustainable innovation. Boards increasingly recognise that governance must extend to AI governance, creating a top-down mandate that is an enabler for Credo AI. Leading with risk understanding and risk management allow organisations to look around the corner of the technology and surface risks before they emerge.

COMMERCIAL MECHANISMS: TRUSTWORTHY AI BUSINESS MODEL DYNAMICS

Credo AI allows enterprises to scale at the pace of AI, while ensuring accountability and trust in workflows, addressing a key barrier to market adoption.

Customer Trust

Governance and trust unlock enterprise AI deployment, allowing companies to deploy or build the technology confidently without compromising on data security, control and trust. This expands the market to highly regulated sectors such as financial services, healthcare, defence and large public-sector buyers. Customers report being able to comply with the EU AI Act 10x faster than manual processes.

Interoperability lowers the cost of customer acquisition. Credo AI integrates with the platforms enterprises already run - Microsoft, IBM, Databricks and the leading MLOps and LLMOps tools - so customers do not need to replace their existing stack. Meeting governance, security and developer teams where they already work is what allows the platform to land, expand, update its risk intelligence and continuously monitor governance over the lifecycle.

Regulatory Alignment

Credo AI considers regulation as the minimum threshold, with internal company policies and industry best practices pushing the bar higher. The risk visibility into around 1,800 AI risks, with continuously updated mappings to global regulatory frameworks, is research-intensive work that grows richer and learns with every regulation, customer deployment and new model class that comes to market.

Credo AI not only helps organisations align against the complex regulatory landscape including EU AI Act and ISO 42001, but also set their own best practice relevant to their contexts, setting them apart from generic compliance tools.

MODEL + DATA + DEPLOYMENT LAYER

Lelapa AI

Open Trustworthy

FOUNDED	HQ	STAGE	RAISED
2022	Johannesburg, SA	Seed	$2.5m

INVESTORS

Atlantica Ventures, Mozilla Ventures

AI is being built for dominant languages, reliant on the vast and readily available data. Lelapa AI is addressing this gap by building efficient language infrastructure that makes sense for diverse linguistic contexts, ensuring everyone can benefit from access to AI.

THE PROBLEM & SOLUTION

As AI becomes deeply embedded in everyday systems, its reliance on large amounts of data, compute and energy is becoming magnified. This presents two barriers - languages spoken by the global majority, where sufficient data does not exist, are excluded; and countries with limited data infrastructure and resources are left behind. These are important limitations that risk excluding large portions of the population, thereby widening inequalities.

Lelapa AI is building resource-efficient language models for the Global South, working with real-world constraints across languages, levels of digital infrastructure and availability of data. Its conviction is simple - language should never be a barrier to opportunity. Its ambition is to make language infrastructure technology accessible from a compute and data perspective too, and to ensure the communities who contribute to building these tools benefit from them.

Lelapa AI's primary product, 'Vulavula', is a multilingual API that provides speech-to-text solutions for both post-call and real-time processing for African languages. Rather than relying on scraped, noisy web text, Lelapa AI builds data from how people actually speak. A crucial feature includes code-switching, the way people mix languages in daily conversations. These design features protect and strengthen linguistic diversity rather than perpetuating cultural erosion and inequity. It crowd-sources data directly from communities through a dedicated app, ensuring contributors are paid for the tasks they complete. The intent is to move to a royalties model and even potentially reinvest in community infrastructure.

Smaller, smarter models are efficient and adaptable to languages with limited data. For example, a model built for isiZulu used around 300 hours of data, against the roughly 1,000 hours needed to adapt an existing model, achieved by working with people to determine the most important data needed to represent a language's variety.

> *No one should have to assimilate to a culture outside of their own in order to access cutting-edge technology.*
>
> **Pelonomi Moiloa** · CEO & Co-founder, Lelapa AI

NAVIGATING TENSIONS

Lelapa AI's model has had to prove its utility in the real world rather than beating research or technical benchmarks. Unable to follow the usual path of licensing IP externally, the team had to demonstrate genuine usage and task completion, such as sentiment analysis, rather than chasing word-error rates. Moving beyond the research milestone and proving practical usage has been the key breakthrough.

COMMERCIAL MECHANISMS: TRUSTWORTHY AI BUSINESS MODEL DYNAMICS

Lelapa AI has unlocked a compact, efficient model that serves communities that have so far remained largely excluded from technological advancements.

Customer Acquisition

The need for Lelapa AI's model in the call-centre space is acute. Two large telcos report: 60% of calls are in isiZulu and they can only process 5% of them, manually. That presents real compliance challenges (penalties can be billed as a percentage of annual revenue), where institutions are unaware of who their customers are and remain unable to serve those customers effectively.

Built around constraints, Lelapa AI's models provide low-cost, efficient computing without compromising capability, particularly important in markets where infrastructure is constrained and budgets are smaller. For example, it was able to add two more languages for a client in 23 days - possible because the models are small and do not need large volumes of data. Vulavula also integrates easily as an API and is compliant with South African data-protection law (POPIA), reducing cost and shortening sales cycles.

Beyond call centres, Lelapa AI aims to build consumer models with language functionality for any product, from voice layers to communication tools such as messaging apps, extending its reach well beyond enterprise transcription.

Lastly, Lelapa AI's ethos is built on the Esethu Framework, where communities have a stake in how their language data is used and reinvested. This provides data that cannot be replicated by scraping the web, giving it a structural advantage in the African markets.

MODEL + DATA LAYER

Adaption Labs

Open | Secure | Trustworthy

FOUNDED	HQ	STAGE	RAISED
2025	California, USA	Seed+	$50m

INVESTORS

Emergence Capital, Threshold Ventures, E14, Mozilla Ventures

Large language models require increasing amounts of capital, infrastructure and data to keep learning. Adaption Labs is building a dynamic and contextual large language model that learns continuously and adapts to humans, rather than the other way around.

THE PROBLEM & SOLUTION

The current approach to building AI models has given breakthrough results, but it has also revealed two structural problems - cost and rigidity. Training frontier models requires millions of dollars of compute, and only a handful of companies can afford it. Secondly, training is seen as a one-time, expensive exercise. Despite iterative prompt engineering, the model requires retraining to genuinely update its knowledge and behaviour.

Adaption Labs was formed out of the conviction that AI should not be concentrated within a handful of frontier labs. With a mission to broaden access to AI that continuously learns, Adaption Labs is structured across three pillars - **Adaptive Data** offers rich data-generation techniques, including the ability to bring your own dataset, with support for more than 242 languages and localisations; **Adaptive Intelligence** automatically allocates compute in proportion to task complexity and lets people train their own models without requiring a high degree of expertise. Today fewer than 1,000 people globally can train a foundation model, and the 'Autoscientist' product opens that fundamental capability to a wider class of users building for their own communities; and **Adaptive Interfaces**, the third pillar, has not yet launched, its longer-term aim being to rethink how the web works, dynamically summoning the right interface for a given task rather than relying on fixed interfaces.

Adaption Labs offers low-cost, efficient and human-centred models by employing techniques like gradient-free learning and on-the-fly adapters, so the model learns without changing its underlying weights. This is a step change from how traditional AI models learn. Not only does it build for local communities that mainstream AI has underserved, but it does so with a smaller, efficient use of compute lending a more sustainable footing.

> *We want to develop continuous learning systems, where AI can improve based on interactions systems and humans have with the surrounding ecosystem.*
>
> **Sudip Roy** · Co-founder, Adaption Labs

NAVIGATING TENSIONS

Adaption Labs is betting against the dominant 'scaling' thesis: that better AI comes from bigger models, more data and more compute. Scaling still commands enormous momentum and capital, and efficient continuous learning remains a hard problem, yet to scale. But Adaption Labs sees that the real unlock is broadening access: systems that adapt to users from real-world experience, cheaply and efficiently. A growing body of research backs the argument that frontier models cannot reach real intelligence without learning from experience. The shift it is chasing, from a handful of frontier labs controlling models that are expensive for everyone else to adapt, to AI that learns efficiently from its own environment, would change who controls AI, and who it serves.

COMMERCIAL MECHANISMS: TRUSTWORTHY AI BUSINESS MODEL DYNAMICS

Customer Acquisition

Adaption Labs is pioneering a new model architecture that is changing the way machines learn and are used. This serves them in multiple scenarios. Organisations increasingly want to own their full AI stack rather than relinquish their AI to a few frontier companies, and they will prioritise technology solutions that can customise to their workflows rather than a generic model that knows everything. International companies serving users across the world, in many languages, require a model they can trust. Moreover, it broadens access to smaller, budget-constrained enterprises, as small models targeted at specific use cases are cost-effective and lower their carbon footprint relative to general-purpose models.

Public-sector institutions are an underserved market and a notable opportunity for Adaption Labs. They want more sovereign AI, where their own engineers have access to the same techniques as the frontier labs.

Adaption Labs shapes AI that users actively control and interact with, resulting in increased trust and loyalty and lower customer acquisition costs.

Regulatory Alignment

Adaption Labs allows users to define requirements explicitly - length, tone, safety thresholds, custom content policies, and stay in control of their interaction. Customers in regulated sectors are increasingly required to demonstrate this kind of auditable trail and control over AI behaviour. This demonstrates Adaption Labs' alignment with existing and incoming regulations.

Talent - Representation & Inclusion

The founders' reputation as advocates for cheaper, more accessible AI enhances Adaption Labs' ability to attract and retain top talent. The company is consciously recruiting across the US, Canada, Europe, the UK, India and Latin America, including researchers from regions traditionally underrepresented in frontier AI. The Adaption for Startups programme and the Research Grant Programme extend this principle further, allowing early-stage builders and academics the opportunity to join and shape the movement.

GOVERNANCE LAYER

Holistic AI

Secure Trustworthy

FOUNDED	HQ	STAGE	RAISED
2020	California, USA	Series A+	$15m

INVESTORS

Tola Capital, Premji Invest, Mozilla Ventures

Enterprises are adopting AI at unprecedented speed and scale, leaving considerable room for risk and harm. Holistic AI closes the gap by giving enterprises the visibility and control to embed AI into their operations.

THE PROBLEM & SOLUTION

AI adoption is outpacing governance. Enterprises are deploying AI agents, models and applications faster than the risk, compliance and oversight processes meant to track them can keep up. At most organisations, that oversight still runs on spreadsheets, and by Holistic AI's own data, roughly 60% of enterprise AI is shadow AI: undocumented, unowned, unmanaged. Traditional GRC tooling was not designed to capture AI risks such as bias, drift, hallucination and privacy leakage. This gap leaves regulatory, reputational and operational risks that are widening with the acceleration of AI.

Holistic AI is an enterprise AI governance platform built to close that gap across the full AI lifecycle, organised around three connected capabilities: identify every AI system in the organisation, including shadow AI; protect the portfolio by continuously testing models, agents and applications against 40+ risk dimensions; and enforce regulatory compliance and internal policy in real time, not just at audit time. The platform runs on what the company calls Guardian Agents - Sentinel agents observe AI behaviour and flag anomalies, while Operative agents intervene directly, blocking unsafe requests or revoking access before harm happens.

The company began at University College London, where two founders from different disciplines - one technical, one focused on ethics and accountability - started asking why nobody was rigorously testing AI systems the way they tested other safety-critical software. That research foundation still runs through the company: findings from its research lab feed directly into the platform rather than sitting in a paper nobody productises. The same instinct shows up in the company's open-source library and its LLM decision hub, both built to bring accountability and transparency into how AI systems get evaluated before they reach production.

> *We didn't want to build another audit checklist. Enterprises don't need to be told AI is risky - they need a system that can see every model they're running, prove it's been tested, and step in the moment something goes wrong. That's the difference between governance as paperwork and governance as infrastructure.*
>
> **Emre Kazim** · Co-CEO & Co-founder, Holistic AI

NAVIGATING TENSIONS

While Holistic AI's academic roots lend a structural advantage, it has had to navigate between research rigour and enterprise speed. The company compresses academic cadence, testing methodology, publishing findings, refining models, into a product release cycle that enterprise procurement expects, without diluting the rigour that makes it defensible to a regulator. That tension shows up directly in how it prioritises its roadmap: new test types and regulatory mappings get built from live customer deployments and enforcement actions as much as from published research.

COMMERCIAL MECHANISMS: TRUSTWORTHY AI BUSINESS MODEL DYNAMICS

Customer Trust

The company's central commercial message is that governance moves from blocker to enabler: it lets enterprises adopt and scale AI responsibly, with the ROI accountability boards now expect. Governance and assurance remove a real barrier to deploying AI in regulated sectors, which opens up large enterprise and sovereign data environments that would otherwise stay out of reach. Reported customers include Unilever, MAPFRE, Wikimedia, GE Healthcare, GSK and Allegis. Unilever has described Holistic AI as central to its AI Assurance strategy, with every new AI project assessed by a cross-functional team that includes Holistic AI before it moves forward.

Holistic AI's independence, no model or cloud business to protect, makes its governance tool credible to customers and regulators in a way conflicted hyperscalers or major model providers building their own tooling may not be able to match.

Regulatory Alignment

AI governance platforms exist to help enterprises navigate a regulatory environment that is still being written. Holistic AI's research lab tracks and, in some cases, directly informs emerging regulation, letting the company position itself as working alongside regulators rather than reacting to rules after they land. That is a different posture from generic compliance tools built to check boxes against a static framework. The company's risk mappings are updated continuously against EU AI Act, NIST AI RMF and ISO 42001 requirements, and that mapping gets more precise with every new regulation, customer deployment and new model class that reaches the market.