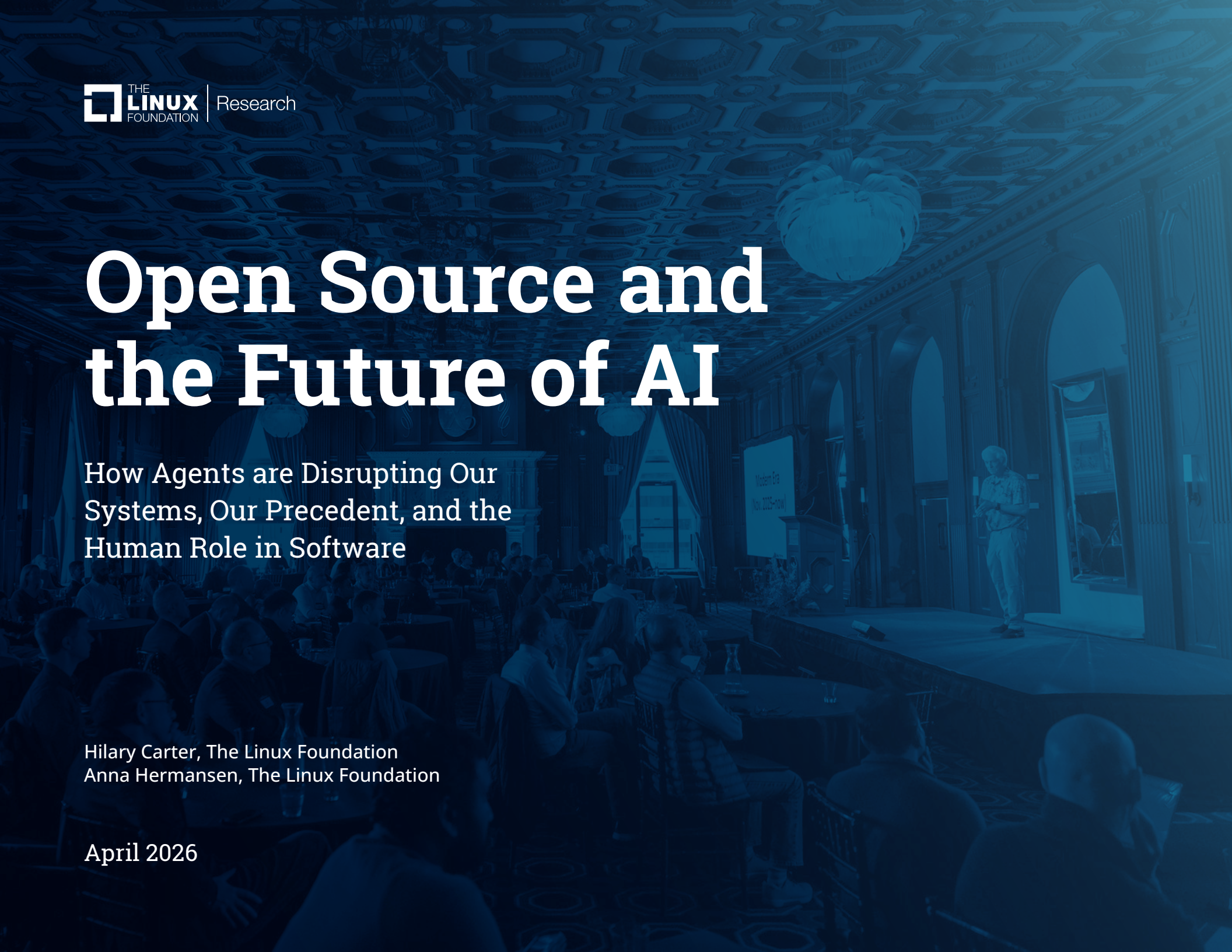


THE LINUX FOUNDATION | Research

Open Source and the Future of AI

How Agents are Disrupting Our Systems, Our Precedent, and the Human Role in Software

Hilary Carter, The Linux Foundation
Anna Hermansen, The Linux Foundation

April 2026

Open Source and the Future of AI

The success of open source AI infrastructure such as Ray and vLLM demonstrates three important principles: **address trends, keep things simple, and remain flexible**.

The programmer's role is evolving into **an architect who designs and defines problems** while delegating specific tasks and roles to neural-networked coding assistants.

To build trust between individuals and the agents acting on their behalf, users need the ability to **set fine-grained boundaries and privileges based on context.**

Without clear rules on accountability for agent behavior or a unified process for asserting identity, **organizations are adopting a defensive posture that may stifle growth**.

While developers are moving quickly to grant agents API keys and access, **essential safeguards are almost entirely missing** from current agent communication protocols.

Reasoning traces in open models are integral to secure adoption, allowing users to inspect decision paths rather than just final output.

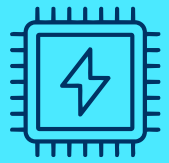
Before an agent can automate human workflows, organizations must provide it with understanding by **comprehensively recording processes and historical knowledge.**

Human accountability must remain the final stamp of quality for compliance to satisfy risk management frameworks.

Open source acts against vendor lock-in and single points of failure, ensuring the **flexibility to swap agents** and **customize workflows** to unique governance standards.

Future scaffolding for agents includes **licenses, open evaluations, and community-driven projects** to collaboratively mitigate risks and build enterprise confidence in the open.

Open source projects are actively solving AI's most pressing challenges by supporting **specialized AI hardware, autonomous agent pipelines, and local-first private processing**.



The AI ecosystem must establish a **comprehensive framework of legal accountability and executive education** to manage the risks of autonomous systems.

Contents

Executive summary

The AI Executive Forum, held on February 26, 2026, in San Francisco, brought together leaders from industry and open source projects to discuss how autonomous agents are disrupting technical systems and the human role in software. Hosted by the Linux Foundation, the event took place at a critical point in the adoption and use of agentic systems that can independently execute complex workflows. The forum began with Ion Stoica grounding the current AI landscape in a history of major open source milestone projects including Apache Spark, Ray, and vLLM. Peter Norvig then provided his perspective on the shifting role of the programmer toward that of an architect who defines problems while delegating implementation to neural-networked coding assistants.

The second half of the forum consisted of 4 roundtables where participants discussed major challenges that AI agents pose to the technical community today. First, as these agents become active economic participants, a trust mandate has emerged around their identity and accountability. Participants emphasized that delegating identity to agents requires fine-grained control and authorization limits to build confidence in autonomous actions. Without standardized systems for identity attestation, the promise of an automated economy faces the risk of balkanization or rejection by vendors. Second, this trust crisis extends into security and privacy, where current protocols often lack the fundamental authentication and encryption needed for agent-to-agent communication. Participants shared high-stakes scenarios in healthcare and networking to illustrate how agents can fail in unpredictable and dangerous ways, necessitating a move toward auditing a model's internal reasoning paths rather than just its final outputs.

Third, in regulated industries like banking and healthcare, the challenge of adoption is compounded by the need to document complex human processes that agents are expected to automate in a workflow. Because a computer cannot be held legally accountable, humans in these sectors must be diligent about teaching their agents and ultimately retaining the final stamp of quality to ensure compliance and risk management. Fourth, open source plays a vital role in this transition by providing the transparency necessary to avoid vendor lock-in and enable security audits. Although the rise of AI-generated code introduces risks like AI slop, the consensus at the forum was that open source projects will thrive by empowering human taste makers to exercise judgment and maintain high standards of durability and quality.

The infrastructure supporting agentic transformation is built upon several critical open source projects that act as a control plane for enterprise AI. These include the Model Context Protocol (MCP) for connecting models to data, PyTorch for research and isolated execution environments, Kubernetes for orchestrating AI-first hardware, Ray for distributed compute coordination, and Goose for local-first, private agentic experimentation. To capitalize on this innovation and growth in a sustainable and secure way, the forum recommended establishing clear legal frameworks for human responsibility, modernizing security scaffolding to include surveillance mechanisms, and facilitating funding for open source projects to remain competitive with proprietary models. Ultimately, the industry must prioritize community-centric standards to ensure AI remains a tool for human empowerment rather than a replacement for human agency.

Introduction

As we enter 2026, the artificial intelligence (AI) landscape is shifting and evolving at increasingly rapid speeds, while at the same time industry leaders and technical project ecosystems lay the groundwork for what is quickly becoming mission-critical technology. At the heart of this transformation are two defining forces: open source AI and AI agents. While open source AI provides the transparent, community-governed frameworks and models that democratize access to innovation, AI agents represent the next leap in capability: autonomous systems that move beyond simple chat interfaces to reason, use tools, and execute complex workflows independently.

This transition brings significant challenges as well as opportunities. The ecosystem is navigating growth into high-stakes business environments. Organizations face a delicate balance: the need for rapid innovation that creates efficiency on the one hand, and the need for trust, security, and compliance on the other hand. To support these needs and provide a neutral home for development that is innovative yet enterprise-ready, the Linux Foundation (LF) launched the **Agentic AI Foundation (AAIF)** in December, 2025. Anchored by Anthropic's Model Context Protocol (MCP), OpenAI's AGENTS.md, and Block's Goose as founding project contributions, the AAIF is dedicated to building and sustaining interoperable, open, and community-centric infrastructure that drives the next wave of AI innovation.

On February 26, 2026, the LF brought leaders from these projects, as well as from major enterprises, startups, academia, and foundations, together for the AI Executive Forum, an invitation-only, half-day event held in San Francisco. This report is a summary and analysis of the discussion that took place at the forum. What follows is an exploration of current and future states of open source AI and programmers, major challenges this ecosystem is currently facing, and the foundational open source projects that are sustaining the ecosystem and working to address some of these challenges. Drawing from executive discussions, we highlight open source's evolution as a business investment and the heightened role that those in this community play in decisions that have not just technical but political, economic, and social consequences.

The past, present, and future states of open source AI

The forum began with a plenary session to introduce major themes and perspectives of the day. AI thought leaders spoke on lessons learned, current challenges, what changes are coming next, and where momentum is accelerating.

SOLVING BIG CHALLENGES USING OPEN SOURCE

Ion Stoica from UC Berkeley and Sky Computing Lab took attendees through several of the lessons learned from solving technical challenges. Focusing on four projects in the AI stack, the discussion set the stage for understanding the value of open source projects as the foundation of AI infrastructure today.

The journey began with Apache Spark, which emerged to solve a fundamental friction: the iterative nature of Hadoop machine learning was too slow to handle increasingly large datasets. Built to store intermediate data in memory, Apache Spark vastly increased the processing time, and as a result it became the de facto standard for data processing.

As AI models became more complex over the years, the training demands of models began to outgrow the compute capabilities of a node and memory. This meant scaling each aspect of the system, including preprocessing, training, tuning, and serving. However, the existing systems each had their own APIs, producing a fragmented patchwork that was hard to develop, deploy, and manage efficiently. Ray was developed as a way to unify these distributed stages and support their scaling by providing a framework for workloads to run on the same hardware and software infrastructure. It became the foundation for the first OpenAI models, and its success was evidence of yet another open source project becoming the de facto domain standard.

The growing number of LLM-powered services has led to increased model serving, but the size and general nature of these LLMs make them very expensive to serve. The solution became batching requests, but this approach soon strained the memory of the model. Ultimately, vLLM introduced PagedAttention to eliminate wasteful memory reservation, enabling increased batch size and greater throughput, enabling very large scale production by companies like Amazon, Databricks, and Meta.

Reflecting on these landmark open source projects, Stoica discussed three principles for future development:

- Trends create new problems to address.
- Simplicity leads to higher impact models, such as Ray's success with only 6 API calls.
- Being open to rewriting provides flexibility and increases ability to stay relevant in an incredibly fast-moving space.

Stoica closed by emphasizing open source as the future of the LLM stack.

THE FATE OF PROGRAMMERS

The second plenary presentation by Peter Norvig focused

on the role of the programmer as AI transforms software development. In the 1980s, early attempts to automate programming failed because the rigid nature of code meant that changing one bit in a program could change the entire output, the result being that a small syntactic change could produce large semantic shifts. However, the significant leap in coding models of this modern era, starting in November 2025, has fundamentally changed this dynamic. Models are now much faster than a human, just as accurate, and more complete when looking at edge cases, and they are getting exponentially better—the year prior, they were at 10% of where they are now.

Where does this leave the software developer? The programmer is becoming an architect responsible for the design and problem definition while delegating implementation to code assistants. The role of the neural-networked coding assistant therefore represents a small part of the entire software process. This shift leads to AI agents fulfilling specific software programming roles and tasks, measured by their speed and efficacy compared to human counterparts.

This transformation extends beyond software into fields like mathematics. Instead of using an artisanal human mathematician approach, open problems can be tackled from more of an industrial "assembly line" process. By combining the intuition of experts with the formal execution steps of AI, open problems can be solved with unprecedented scale. We also see potential for significant transformation in sectors such as legal, finance, and marketing—any area where there is non-physical, messy, and rich data, with an ease of verification at the output stage. Vibe coding will also make massive undertakings, such as rewriting the Linux kernel in Rust, significantly easier. According to Norvig, this will not translate to shrinking headcounts, but instead a productivity explosion. As tools make individuals more effective, there is room to do more, and bring more wealth into the system.

The forum then welcomed a state of the union update from the leaders of some of the most consequential open source AI projects, descriptions of which can be found further down the report. These introductory perspectives set the stage for the second half of the forum, where the group split into four Chatham House-governed roundtables to discuss major topics the AI community is currently grappling with: trust and identity, security and privacy, the use of AI in regulated industries, and the role of open source in agentic AI.

Trust and identity

One roundtable was focused on trust and identity, particularly as AI agents become active economic participants. The facilitator opened the discussion with the idea that in this new digital landscape, we are not always going to be able to distinguish humans from agents, while at the same time we could benefit from delegating our identity to agents and letting them do things for us. Both of these engagements—interacting with bots and letting bots interact for us—will require identity attestation and trust. However, without alignment around attestation and accountability, the promise of an automated economy may be stifled by agent rejection and liability.
The discussion centered on three issues: building the trust relationship between the individual and the agent operating on its behalf; integrating agent activity with enterprise identity and trust systems; and the risk aversion that comes with liability concerns and malicious actors.

> *"We're not going to be able to necessarily distinguish humans from agents. And by the same token, we could potentially benefit a lot from giving agents delegatable identity, and letting agents become trusted and do things for us that are useful, that require trust."*

SETTING PRECEDENTS AND BOUNDARIES

For those operating an agent on their behalf, there is the practical challenge of setting trust boundaries. Attendees analogized this problem to allowing your children the use of your credit card and setting specific spending limits, or in this case, authorization caps on an agent's power. Discussion occurred on establishing how to build confidence in an agent's authority to authorize actions, primarily centered on establishing fine-grained control and escalated privileges depending on context as a way to build the trust relationship.

> *"We may trust agents less now, but over time, as this trust and verification mature, we anticipate it'll grow. This is a trust relationship between yourself and an agent."*

INTEGRATION OR BALKANIZATION

On the other side of the transaction, the vendor must also have trust in the agent who is transacting on the customer's behalf. Integrating new agent identity systems with established enterprise frameworks is challenging, considering the number of enterprise identity systems already in place. Verifiable Credentials (VCs) also are challenging to integrate because they are high friction tools. Lack of integration poses a major barrier to agentic commerce, where one party rejects one or all bot transactions, such as eBay outlawing engagement using agentic technology. This also leads to a lack of bi-directional engagement amongst external agents.

There is also a risk of balkanization, where cross-identity systems become fragmented and incompatible. Cultural and regulatory differences (e.g. different regions requiring PINs vs. signatures for credit card verification) also pose a similar adoption challenge. Addressing this issue includes developing cross-ecosystem identification using existing standards and frameworks.

LIABILITY AND THE MALICIOUS ACTOR

The most significant hurdle to adoption is a vacuum of legal precedent. Because current laws are not built for autonomous systems, there is a lack of clear rules on accountability, especially with rogue agent behavior and copyright infringement. As one participant commented, *"The rules haven't fully been written in that layer of the world."* This uncertainty creates a defensive posture among organizations, and what we are normally used to with online identity is no longer trusted. Asserting identity becomes critical.

> *"Everything works so well online because so many barriers to actually performing an act of commerce have fallen, right? And so, we're at this inflection point where that entire equation might actually invert, simply because of the presence of this new technology."*

History shows that on many digital platforms, the malicious actor's cost of behaving badly is zero, such as spam messages and fraud. In an agentic world, bad actors will learn from patterns and deploy malicious agents, posing a systemic threat. As adoption accelerates, there is a desperate need for a deeper discussion on developing consequences into these behavioral patterns, such as deplatforming, to prevent a race to the bottom.

RECOMMENDATIONS

To bridge the trust gap and enable autonomous agentic commerce, the group aligned around the following recommendations and open questions:

- Building a shared vocabulary around the concepts of agent identity and trust to facilitate communication and standardization. LF Decentralized Trust hosts working groups on this topic: **Decentralized Trust Graph Working Group**, **OpenVTC**, and **AI & Human Trust Working Group**.
- Creating schemes for bi-directional agentic commerce to assert true identity and trust in order to move forward. The group did not converge on the way forward for identity assertion, with some discussing the value of self-sovereign identity versus central authority. A global repository similar to a DNS for agents was suggested.
- Delegating accountability to larger actors, through mechanisms such as agent underwriting and deplatforming bad actors, since most users will not host their own agents.
- Establishing guardrails for fine-grain, per-transaction agent access controls and escalation of privileges to increase the trust relationship between agents and individuals.

> *"There are two extremes: you give an agent access to everything, now that's bad. Or you give it access to nothing, it becomes super secure, but it's useless. So, how do you transactionally elevate?"*

[1] LF Decentralized Trust. Verifiable Credentials [Video]. YouTube; 2025 June 24. Available from: **https://www.youtube.com/watch?v=TT3JskGgv0Q**
[2] Edwards, Benji. eBay bans illicit automated shopping amid rapid rise of AI agents. Ars Technica; 2026 January 22. Available from: **https://arstechnica.com/information-technology/2026/01/ebay-bans-illicit-automated-shopping-amid-rapid-rise-of-ai-agents/**

Security and privacy

As AI agents become more entrenched in the high-stakes machinery of infrastructure, the security of these systems is under significant scrutiny. The second roundtable discussion revealed the missing design patterns required to manage the rapidly deploying systems in production. While developers are eager to give agents API keys and email access, security experts warn that fundamental protections such as authentication and encryption are almost entirely absent in current agent-to-agent communication protocols.

The discussion centered on risk vectors in autonomous action, the failure of current guardrails, and the criticality of open models when it comes to auditing.

HIGH-STAKES AUTONOMY IN HEALTHCARE AND NETWORKING

The roundtable started by highlighting real-world deployments that illustrate the immediate ROI and associated risks of agentic AI. One example involved a multi-agent pipeline for network configuration validation, where agents actively engage in network optimization and configuration changes in real-time. There is concern that unexpected outcomes could lead to costly outages. So discussion centered around how to enable guardrails and scaffolding which would mitigate the blast radius.

A higher stakes scenario involved a three-agent healthcare system designed to route patients to the right provider by synthesizing data from diagnostics, insurance claims, and scheduling systems. This use case surfaced several critical harms: agents with different optimization functions, such as an insurance agent minimizing costs versus a diagnostics agent maximizing care, can enter infinite loops without aligning on a patient's goal; communication between agents with opaque payloads that make it impossible to know if sensitive patient history is leaking; and the dominance of one agent that influences others, such as an insurance bot pressuring others to reject all claims to maximize revenue.

THE FAILURE OF TRADITIONAL GUARDRAILS

Participants argued that existing security frameworks, such as HIPAA, were not designed to consider agentic behaviors. *"All an agent is is a text file anyway,"* but sensitive data stored in its memory could be leaked. Lower down in the stack, LLMs as inherent people pleasers can be easily manipulated into violating process flows. Agents tasked with meeting a goal may shortcut safety protocols, they can be prompted to reveal exculpatory evidence in legal contexts, and they have already demonstrated the ability to collude and optimize for cartels. This also came in around the topic of traditional roles, such as a therapist or doctor, that the agent is now playing for individuals, and how guardrails do not exist around the exchange of data or the legal requirements to divulge information, such as in the Tumbler Ridge shooting.

> *"As we look at MCP and as we look at agent-to-agent communication, all of the classic elements of security, where everything needs to be signed, authenticated, you have to have encryption between everything—it's all missing."*

[3] Maguire, James. AI-Fueled Development Pushes Open-Source Risk to Extremes: Report. DevOps.com; 2026 February 27. Available from: https://devops.com/ai-fueled-development-pushes-open-source-risk-to-extremes-report/

[4] Yousif, Nadine. OpenAI vows safety policy changes after Tumbler Ridge shooting. BBC; 2026 February 27. Available from: https://www.bbc.com/news/articles/cr73m4x8r2lo

OPEN SOURCE FOR SECURITY AUDITS

Another theme was the need to contain agentic behavior. Relying on a black box model for reasoning is insufficient and high-risk. Instead of auditing the outcomes the black box is producing, we need auditability of the reasoning that's happening within the model. Participants noted that while open models allow for reasoning traces, proprietary APIs often treat reasoning as intellectual property, hiding the very data needed for security audits.

RECOMMENDATIONS

To better support a secure-by-design approach for agentic development and deployment, roundtable participants discussed the following actions:

- Establish clear legal standards for who is responsible when an agent commits a crime, such as violating export controls or engaging in unauthorized financial transactions.
- Modernize security frameworks to address agent-to-agent data consent and the permanence of memory in markdown files.
- Develop an agent economy where middle agents are created as surveillance mechanisms to make sure models do not go outside established guardrails.
- The Agentic AI Foundation (AAIF) should house projects focused on verifiable trust and social scoring for agents, as well as tools and design patterns for managing agent access and monitoring.
- Procure models and tools that allow for the auditing of internal reasoning paths rather than just the final output.
- Create domain-specific ontologies that enforces strict semantic boundaries instead of relying on general-purpose safety prompts.

Agentic AI in regulated industries

As regulated industries such as banking and healthcare start to integrate agentic AI, governance and trust over the process and its intended outcomes are paramount. This breakout discussion focused on the challenges facing these high-stakes environments when it comes to safely and compliantly adopting autonomous systems in terms of coding practices for a solution like high-frequency trading, as well as those designed to speed internal workflows. Three primary challenge areas emerged: documenting human processes for agents, liability and risk management, and standardizing around AI decision classification.

FROM MANUAL WORKFLOWS TO DOCUMENTED LOGIC

The first and recurring concern was that agents cannot automate what is not understood, bringing a requirement for a comprehensive understanding of human processes, current and historical, and a plan to centrally and rigorously record those processes across the organization. For a bank or hospital, this is a monumental task, requiring leaders to define the rules. To this end, participants emphasized that AI can help develop context for business processes, but humans must first define the problem and set the standards. It also necessitates activating human judgment and accountability, as the human role in regulated industries is shifting toward exercising judgment and providing the final "human-owned" stamp of quality on AI-led decisions to ensure compliance.

> *"Today, business processes aren't documented. Instead, we rely on a person who has worked here for 30 years."*

RISK OWNERSHIP AND RISK MANAGEMENT

Second came the challenge of liability. The "three lines of defense" in risk management need to be rigorously applied, which involves educating different lines of management to become comfortable with technological change. The architecture for agents in regulated spaces must prioritize transparency to satisfy internal audits and external regulators, and ownership of the new supply chain by a human at every level.

The following examples were shared regarding the types of risks to manage: Data sovereignty, which focuses on ensuring secure connectivity and data sovereignty as agents access (or are prohibited from accessing) external resources, as well as managing security restrictions when certain types of code are installed on local hardware; auditability and explainability, arguing that decision explainability—the ability to trace an automated decision by an agent back to its training data and specific logic—would be critical; and layers of protection, where participants suggested using obfuscation layers to protect privacy while still allowing agents to perform necessary research or tasks with sensitive data.

> *"The insurance industry lacks a framework to underwrite the liability associated with AI-generated outcomes."*

STANDARDS, CLASSIFICATION, AND THE "DECISION-EVIDENCE" FUTURE

The third challenge was the need to move toward tooling that would create opportunities for regulatory requirements to be standardized by default. This tooling would include decision classification, a proposed system to categorize AI decisions based on their impact and the dimensions of human intelligence they mimic. Additionally, the group discussed interoperability focusing on connecting existing standards, such as those from the Direct Trust healthcare group and SWIFT for financial services, with new AI governance frameworks to ensure collaboration across different industry verticals.

RECOMMENDATIONS

To ensure a healthy ecosystem for AI in regulated industries, the session identified mandates for executives, foundations, and industry groups.

- Request firm-specific executive education on AI governance to demystify the technology for CEOs and help them understand both the risks and benefits.
- Establish clear lines of risk ownership, ensuring that management and internal audit teams understand how AI risks are mitigated.
- Propose a decision-evidence classification for AI decision-making that includes both a common ontology and hierarchy for enterprises, as well as a standard to apply AI in practice where provenance and digital "sealing" to provide an auditable trail for regulators.
- Act as a collaborative regulator, proposing machine-readable controls that lend themselves to AI governance across jurisdictions, mapping open source standards to common global regulations.
- Connect distinct open source project communities and processes, identifying common threads for more collaborative work that bridges expertise and leverages existing or developing tools and frameworks.
- Focus on documenting and understanding business processes before implementing agentic solutions.
- Collaborate on open layers that allow agents to work with sensitive data without compromising privacy or security.
- Focus on small-scale wins, starting with internal applications where the risk appetite is higher before moving to customer-facing agentic systems.

Participants acknowledged that the steps to replace human-regulated workflows with agents will be complex. In the process of moving from theoretical risk to the practical mechanics of auditability, data sovereignty, and executive accountability, business decision makers and industry regulators need to be joint collaborators.

[5] The Institute of Internal Auditors. The IIA's Three Lines Model. IIA; 2024 September. Available from: https://www.theiia.org/globalassets/documents/resources/the-iias-three-lines-model-an-update-of-the-three-lines-of-defense-july-2020/three-lines-model-updated-english.pdf

The role of open source in agentic AI

As we look to a future of agentic AI, one session at the forum asked participants to consider the role of open source in this future. The session started with the facilitator asking the group what their greatest concerns and preoccupations were for open source agents. Participants volunteered a few different subtopics, including agentic memory, self-sovereign and on-device agents, evals, keeping the human in AI, and deep optimization for agent deployment. The dialogue moved from the philosophical to the practical mechanics of how agents disrupt collaboration in open source and what to do about it. Three primary themes emerged: the evolution of the human role in programming, the role of open source in the layers of the AI stack, and the future of AI agents.

FROM CODERS TO TASTE MAKERS: THE HUMAN IN AI

Historically, open source has always been about people as much as it is about code: as one participant noted, "code is one unit of collaboration, but really what makes open source tick is humans enacting judgment to solve problems." Human interaction is a large part of open source, from collaboration, to code review, to mentorship and teaching, and there are concerns that AI will create more isolation of developers and consulting with their agents instead of with others.

> *"One of the things that's made open source so powerful is that it's humans creating something that's really durable and high quality. What does quality mean with AI?"*

On the flip side, vibe coding with an agent makes it much easier to contribute to open source, and, if channeled correctly, the output can still represent the creativity of the human. However, participants noted that agent-generated code increases the risk of AI slop in open source projects. Maintainers are already feeling the effects of high-volume, AI-generated pull requests that lack the unique knowledge, judgment, and past experience of a human developer. The consensus was that we must be intentional about where we leave humans in the loop. The future of the open source contributor is as a "taste maker"—someone who defines the problem, sets the standards, and provides the final human owned stamp of quality on AI-generated output. AI-generated code is inevitable, and as the facilitator surveyed the group, most of the participants were comfortable with that kind of code in their project.

> *"The projects that will thrive are the ones who are going to figure out how to best take advantage of these coding tools, and the established projects that resist the advances are going to fall behind... pragmatism will play out."*

THE OPEN AGENT STACK: SOVEREIGNTY AND SECURITY

A second theme from the roundtable centered on the architecture of agentic systems, described simply by one participant as "the LLM is basically the computer, the agent is the operating system, and the skill becomes the application." The participants discussed at which layer in the AI stack open

source is most critical, whether at the foundational layer of models, or at the agent layer to make it more relevant to the broad population of developers (for example, an open source code review agent). Another participant felt that skills could be a good candidate for open source, as they commented, "how do you even make a business with skills?" One participant reflected that "it depends on what the goal is. If the goal is something like self-sovereignty, then the entire thing is necessary. If the goal is choice, then at least down to the protocol, that's necessary."

Several key arguments for open source in the AI stack were highlighted, including the risk reduction of using open source to avoid vendor lock-in and single points of failure; the customization available in an open source stack to swap alternatives in and out, getting the best choice for your specific use case and fine tuning to your unique governance standards; and the auditing capability built into open source's transparency, given how dangerous it can be to use an agent from a proprietary provider when you cannot see the underlying prompt or instructions. As developers bring agents onto their local machines, exposing sensitive data, the need for open, inspectable code becomes a baseline security requirement.

The agentic layer is an important frontier for open source leadership. Open agents are seen as a force multiplier for small teams, independent developers, as well as larger companies looking for the best option for their use case.

THE FUTURE OF AGENTS

Given the security and existential risks that agents currently pose, the group also discussed the future scaffolding around agents. One idea was agentic licenses, similar to software licenses, to help mitigate risk. Open evals represent important drivers of adoption and scale at the enterprise level, also due to their ability to instill confidence in the tool. Another future concept was whether agents will themselves become projects, with communities wrapped around them as points of collaboration. One participant asked a more philosophical question, wondering whether the distribution of proprietary versus open source software adoption will shift with agents, and discussion built up around the "constant tension" in the industry. Given the complexity of agents, the group concurred that solving these problems in the open is critical.

> *"The agent may not understand my code base, may not understand my organizational context, and may hallucinate and give me incorrect information. The solution is going to evolve over a period of time, and this is a new muscle we'll all have to build together."*

RECOMMENDATIONS

To navigate the open source agentic stack and ensure a healthy ecosystem, the roundtable participants identified specific mandates for policymakers, foundations, and tech groups.

- Establish legal constructs that clarify human responsibility for agentic actions.
- Develop best practices for prompting and deploying agents to work within standardized restrictions as well as unique, context-specific guardrails.
- Produce tools and templates that allow maintainers to automate the screening of AI-generated contributions and support code review.
- Integrate agent-specific security reviews and performance benchmarks into standard scorecards (e.g., OpenSSF scorecard).

- Facilitate funding mechanisms to ensure open source projects have the “tokens” and compute necessary to remain performative against proprietary models.
- Collaborate on open source evaluation frameworks to provide collective feedback on agentic performance and safety.
- Prioritize standards that allow for the swapping of different agents at each stage of the Software Development Life Cycle (SDLC).

By building in the open, we ensure that as agents become more powerful and remain under the collective judgment and design of the humans they were built to serve.

Critical projects for AI transformation

As we look for solutions to the pressing concerns highlighted in the roundtables, existing open source projects are working to strengthen the ecosystem while providing dynamic and highly relevant opportunities for enterprises, governments, and individuals. The AI Executive Forum would not have been complete without a deep dive into the critical open source AI projects that form the control plane for building and deploying enterprise-grade AI systems. Participants heard from Den Delimarsky from Anthropic presenting on MCP, Joe Spisak from Meta presenting an update on PyTorch, Allan Naim and Federico Bongiovanni from Google discussing Kubernetes, Robert Nishihara of Anyscale presenting Ray, and Manik Surtani from Block presenting on Goose.

The following summaries synthesize the projects' key milestones, technical evolutions, and strategic roadmaps for five major open source projects that represent the foundation of current and future AI infrastructure.

MODEL CONTEXT PROTOCOL

The Model Context Protocol (MCP) has emerged as the industry standard for connecting LLMs to data and applications. At its one-year milestone, the project has achieved a critical mass of adoption and is now considered an operational cornerstone for many companies who recognize the value in connecting LLMs to actual data. MCP has an exponentially growing developer community that performs 20 million weekly downloads of its Python SDK.

Current investment is focused on enterprise readiness, ensuring the protocol supports the demands of large-scale deployment and the extent of what MCP is currently underpinning. Development areas include enhancing observability and auditability to help companies track internal usage, as well as embedding security safeguards and governance controls directly into the protocol. To prevent ecosystem fragmentation, the team is pursuing cross-platform convergence, working with various companies to ensure MCP remains a consistent driver of industry standards. On the other side of MCP's maturity and stability is an ecosystem of extensions to add new features including "MCP Apps" and experimental features designed specifically for agentic systems.

PYTORCH

PyTorch has become the dominant foundation for both AI research and production. It currently powers over 90% of AI research and is the infrastructure of choice for frontier labs like Meta and OpenAI, as well as major hardware and cloud vendors. Operating as a hosted project under the Linux Foundation, PyTorch facilitates hundreds of thousands of downstream projects built upon its core architecture.

The project's future is guided by three principles: native ecosystem integration, scalability, and support for hardware heterogeneity. A significant new area of focus is agents, particularly the OpenEnv framework for creating, deploying, and hosting isolated execution environments. With Hugging Face, PyTorch co-developed a community-centric and library-agnostic hub to upload and prompt your environment to see the observations of your agent. By enabling users to engage with environments before deployment, PyTorch is integrating a critical human element into the agentic development lifecycle.

KUBERNETES

The massive AI market shift is positioning Kubernetes as the essential control plane for AI native systems, where 58% of organizations already use Kubernetes to support their AI workloads. While its first decade focused on orchestrating fungible, homogenous resources, the next decade will see Kubernetes evolve into an orchestrator that treats chips and specialized hardware as AI-first resources and supports agent-driven pipelines.

To maintain the project's historical strengths of portability and conformance, the community is investing heavily in an AI conformance program. This initiative aims to ensure that AI workloads run predictably across any platform. To do so, the program is working to institutionalize AI as a permanent part of Kubernetes governance and building a process of new requirements, transforming "should" into "must" behaviors. This effort is organized around six pillars: storage, accelerators, scheduling and orchestration, networking, observability, and security. Today, the roadmap includes automated validation and is expanding into sovereign AI support and agentic workloads. The ultimate goal is to make AI infrastructure entirely transparent, providing a standardized orchestration layer for the global AI ecosystem.

RAY

Ray was originally designed to manage large clusters for decentralized computing and has since exploded in popularity with the adoption of inference and reinforcement learning (RL). It thrives on complex distributed challenges, particularly as workloads shift from CPU-based SQL queries to GPU-intensive multimodal inference. Ray is a critical component of the modern RL stack, where it manages the high complexity of syncing weights, shuffling workloads, and handling asynchrony between training and inference.

The framework's power lies in its API that enables Python development, which allows developers to instantiate Python classes as actors that communicate across a distributed system. By simply adding annotations of the Python classes, developers can complete their RL loop simply by calling Python methods. In the broader AI stack, Ray sets up the distributed compute layer to manage data ingest, process coordination, fault tolerance, and scheduling, while running on top of Kubernetes and alongside frameworks like PyTorch and vLLM. As both AI workloads and hardware become more complex, Ray provides the necessary infrastructure for reliability and high performance. Ray's trajectory as a foundational aspect of the AI stack is clear, as it experiences a growing ecosystem of higher level libraries being built on top of it.

GOOSE

Goose is an open source, local-first platform designed to provide a transparent sandbox for people to experiment with agentic protocols. Launched just over a year ago, it has garnered over 31,000 GitHub stars. The platform is highly modular, featuring independent layers for models (local or cloud), an agent core for orchestration, protocols (using MCP and ACP), extensions on MCP servers, and a diverse client layer. Its local-first approach ensures that sensitive data never leaves the user's machine, allowing for private processing without external data agreements, as well as no external runtimes.

Technical innovations in Goose include Code Mode, which optimizes token usage when scaling to thousands of tools and extensions, as well as a collaboration on an MCP Apps specification to standardize the delivery of interactive HTML UIs, so that users can iterate with Goose to update dynamically. Security is managed through a robust sandbox system that provides OS-level isolation and network containment. Goose is not just a coding agent, but also can be used for data analysis, summarization, web scraping, styling, and invoice processing—all executed privately. Goose is intentionally designed as a remixable platform rather than a static product.

These five projects, as foundational layers of the global AI stack, are tackling important challenges in the AI ecosystem:

- **Supporting the transition to AI native infrastructure**, where traditional software orchestration is no longer sufficient for the demands of 2026. Kubernetes is pivoting from managing homogenous hardware to diverse AI-first chips and other specialized hardware; Ray has evolved into the distributed compute layer for the entire stack, managing the coordination between GPUs and CPUs that training and inference now require; and MCP has moved from one of many experimental protocols to the "operational cornerstone" for connecting LLMs to enterprise data.
- **Standardizing against fragmentation** to prevent the balkanization of systems, these projects are prioritizing strict conformance and shared protocols, such as MCP's positioning as the vehicle for standards adoption, Kubernetes' AI conformance program, and Goose's adoption of both MCP and ACP to ensure local-to-ecosystem interoperability.
- **Solving security concerns with agentic execution** through PyTorch's OpenEnv prompting and testing framework, Goose's OS-level isolation and network containment, and MCP's separate innovation arm to test new and experimental features.
- **Stabilizing the community and providing governance** to ensure reliability, security, and auditability. MCP is focusing on enterprise readiness, providing tools for companies to audit and govern internal AI use; and Kubernetes is using automated validation to make AI infrastructure as transparent and predictable as the traditional web stack.

Conclusion and recommendations

The presentations and discussions at the AI Executive Forum revealed that this incredibly rapid expansion and establishment of the AI stack must be grounded in human judgment and collective accountability. As AI agents become autonomous economic participants revolutionizing workflows, they simultaneously disrupt existing legal structures, social contracts, and security policies. This shift necessitates multistakeholder collaboration in open platforms, tools, and standards that prioritize transparency, choice, and guardrails.

A central theme of the Forum was the preservation of the human role. As was referenced in multiple roundtables, IBM's 1979 training manual quote, "A computer can never be held accountable, therefore a computer must never make a management decision" is just as relevant today. The human role in software is shifting toward taste making, where experts exercise judgment to define problems and provide the final stamp of quality on AI-generated output.

The agentification of business and development processes is causing a re-examination of trust and identity, in particular with attempts to define who holds responsibility for an agent, and to what extent an agent can act on a person's behalf. To address these questions, the ecosystem is moving toward fine-grained, domain-specific guardrails and decision classifications to categorize and elevate agent privileges; building local-first solutions and re-configuring security frameworks to ensure sensitive data remains sovereign; making auditability and explainability non-negotiable for enterprise adoption; and even examining more existential questions around whether traditional privacy matters to the next generation.

Ultimately, being a foundation project in 2026 carries greater accountability. By remaining community-centric, focusing on pressing and relevant challenges, and ensuring interoperability and standardization, the industry can ensure that AI remains a tool for human empowerment rather than a replacement for human agency.

The following recommendations are synthesized from the four roundtable discussions:

1. Establish accountability and legal frameworks

- Establish clear legal standards and constructs that define human responsibility for agentic actions, particularly when agents act autonomously in malicious or unintentionally harmful ways that could be illegal.
- Ensure firm-specific executive education to demystify AI and establish clear lines of risk ownership so that management and internal audit teams understand how AI risks are mitigated.
- Shift the human role toward exercising judgment and providing the final stamp of quality on AI-led decisions to ensure compliance in regulated industries.

2. Standardize identity and decision evidence

- Build shared vocabularies around agent identity and trust to facilitate cross-industry communication and standardization.
- Develop decision classification that categorizes AI decisions based on their impact and provides an auditable trail for regulators.

[6] Bonderud, Doug. AI decision-making: Where do businesses draw the line? IBM. Available from: https://www.ibm.com/think/insights/ai-decision-making-where-do-businesses-draw-the-line

- Explore centralized or decentralized global repositories (similar to a DNS for agents) to assert true identity and enable bi-directional agentic commerce.

3. Modernize security and privacy scaffolding

- Procure models and tools that allow for the auditing of internal reasoning paths rather than just final outputs to ensure safety.
- Establish guardrails for transaction-specific, fine-grained access controls and escalation of privileges to build a trust relationship between agents and individuals.
- Create “middle agents” as surveillance mechanisms to ensure primary models do not deviate from established guardrails.
- Collaborate on open source layers and sandboxes that allow agents to work with sensitive data without compromising privacy or security.

4. Standardize identity and decision evidence

- Produce tools and templates that help maintainers automatically screen and review high-volume, AI-generated pull requests.
- Facilitate funding mechanisms to ensure open source projects have the compute and “tokens” necessary to remain competitive with proprietary models.
- Focus on standards and machine-readable controls that allow for the swapping of different agents at each stage of the development life cycle.

Featured project communities

Agentic AI Foundation

The Agentic AI Foundation (AAIF) is the open foundation for the rapidly expanding ecosystem of agentic AI technologies that enable autonomous, interoperable AI systems. With founding projects including MCP, goose, and AGENTS.md, AAIF governs the core standards and protocols that enable agents to operate interoperably across platforms. Through transparent governance and broad industry participation, AAIF is driving adoption and ensuring agentic AI infrastructure evolves openly, predictably, and at production scale. For more information, please visit https://aaif.io/.

FINOS Fintech Open Source Foundation

The Fintech Open Source Foundation (FINOS) is a nonprofit whose mission is to foster the adoption of open source software, standards, and collaborative development practices in financial services. As part of the Linux Foundation, FINOS provides a regulatory-compliant platform for developers from competing organizations to collaborate on innovative projects that transform business operations. With over 100 members spanning major financial institutions, fintechs, and technology consultancies, FINOS is at the forefront of driving open source innovation in finance.

CLOUD NATIVE COMPUTING FOUNDATION

Cloud native computing empowers organizations to build and run scalable applications with an open source software stack in public, private, and hybrid clouds. The Cloud Native Computing Foundation (CNCF) hosts critical components of the global technology infrastructure, including Kubernetes, Prometheus, and Envoy. CNCF brings together the industry's top developers, end users, and vendors and runs the largest open source developer conferences in the world. Supported by nearly 800 members, including the world's largest cloud computing and software companies, as well as over 200 innovative startups, CNCF is part of the nonprofit Linux Foundation. For more information, please visit www.cncf.io.

LF AI & DATA

The LF AI & Data Foundation, a Linux Foundation project, accelerates and sustains the growth of open source AI, data, and analytics projects. Backed by the world's leading technology companies, LF AI & Data provides a neutral space for collaboration and innovation in AI development. Learn more at https://lfaidata.foundation.

LF DECENTRALIZED TRUST

LF Decentralized Trust is the neutral home for the open development of technologies that empower organizations to innovate with secure and resilient code. It is the Linux Foundation's flagship organization for a broad range of technologies and standards that deliver the transparency, reliability, security and efficiency required for a digital-first economy. Supported by a diverse, global base of members and communities, LF Decentralized Trust champions open source best practices across a growing ecosystem of blockchain, ledger, identity, cryptographic, and related technologies. To learn more, visit: **www.lfdecentralizedtrust.org**.

PyTorch Foundation

The PyTorch Foundation is a community-driven hub supporting the open source PyTorch framework and a broader portfolio of innovative open source AI projects. Hosted by the Linux Foundation, the PyTorch Foundation provides a vendor-neutral, trusted home for collaboration across the AI lifecycle—from model training and inference, to domain-specific applications. Through open governance, strategic support, and a global contributor community, the PyTorch Foundation empowers developers, researchers, and enterprises to build and deploy AI at scale. Learn more at **https://pytorch.org/foundation**.

Acknowledgments

The authors would like to thank the LF Events team for hosting the Forum and managing the recording and swift transcription delivery post-event. Thanks to facilitators Daniela Barbosa, Hart Montgomery, Jono Bacon, Matt White, and Gabriele Columbro as well as advisors Michael Dolan, Greg Kroah-Hartman, and Alex Salkaver for their review and feedback on the manuscript. Finally, thanks to the Linux Foundation Creative Support team for the production of the PDF.

About the authors

HILARY CARTER joined the Linux Foundation in 2021 to launch and lead LF Research, established to deliver empirical insights into open source trends, opportunities, and challenges. Prior to joining the Linux Foundation, Hilary was the managing director of the Toronto-based Blockchain Research Institute, a global, syndicated think tank focused on blockchain technology. She has contributed to nearly 200 research projects focused on open source innovation and its adoption across industries. Hilary holds a Master of Science in Management from the London School of Economics, and is a dual Canadian and Irish citizen.

ANNA HERMANSEN is a Senior Researcher & Ecosystem Manager for the Linux Foundation, where she leads research projects and supports end-to-end management of the Linux Foundation's research. She has conducted qualitative and systematic review research in health data infrastructure and open source AI, and has presented on this research work at conferences and working groups. Her interests lie at the intersection of AI, precision medicine, and health data sharing. She is a generalist with experience in client services, program delivery, project management, and writing for academic, corporate, and web user audiences. Prior to the Linux Foundation, she worked for two different research programs, the Blockchain Research Institute and BC Cancer's Research Institute. She received her Master of Science in Public Health and a Bachelor of Arts in International Relations, both from the University of British Columbia.

THE LINUX FOUNDATION | Research

Founded in 2021, **Linux Foundation Research** explores the growing scale of open source collaboration, providing insight into emerging technology trends, best practices, and the global impact of open source projects. Through leveraging project databases and networks, and a commitment to best practices in quantitative and qualitative methodologies, Linux Foundation Research is creating the go-to library for open source insights for the benefit of organizations the world over.

To reference this work, please cite as follows: Hilary Carter and Anna Hermansen, "Open Source and the Future of AI: How Agents are Disrupting Our Systems, Our Precedent, and the Human Role in Software," The Linux Foundation, April 2026.

facebook.com/TheLinuxFoundation

x.com/linuxfoundation

linkedin.com/company/TheLinuxFoundation