[redacted]@salesforce.com>

Salesforce | Vielen Dank für das heutige Gespräch + Einladung zum Agentforce Breakfast für einen Mitarbeitenden aus dem Büro

1 message

[redacted]@salesforce.com> Tue, Apr 8, 2025 at 3:04 PM
To: [redacted]@bundestag.de>, [redacted]@bundestag.de>

[redacted]

habt vielen Dank, dass Ihr Euch heute Zeit für den Austausch mit mir genommen habt und mir somit die Möglichkeit gegeben habt, Salesforce und mich vorzustellen.

Im Verlauf unseres Gesprächs kamen wir sowohl auf die Rahmenbedingungen für die agentische KI sowie auf die Auswirkungen auf die Umwelt und Emissionen von KI Modellen zu sprechen. Gerne reiche ich ein paar weitere Informationen zu den beiden Themenkomplexen mit dieser Nachricht hinterher (unten eingefügt).

Zudem füge ich unten eine Einladung zu dem von Salesforce ausgerichteten "Agentforce Breakfast" am 15.4. in Berlin ein. Diese praxisnahe Veranstaltung zur agentischen KI richtet sich an die Mitarbeitenden der Bundestagsbüros und Verbände. Wir würden uns sehr freuen, eine Person aus Euren Büros vor Ort begrüßen zu dürfen. Eine Anmeldung kann gerne direkt an mich erfolgen unter Angabe des vollständigen Namens und der E-Mail Adresse.

Ich wünsche für den weiteren Start im Bundestag weiterhin alles Gute und viel Erfolg. Sollte der Themenzuschnitt für Digitalisierung in dem Zuständigkeitsbereich von einem von Euch landen, werden wir uns sicher in Berlin das ein oder andere Mal über den Weg laufen. Bei Rückfragen sowie als Ansprechpartnerin für Salesforce in Deutschland stehe ich Euch und Euren Teams gerne zur Verfügung.

Mit besten Grüßen
[redacted]

1. Informationen - Vertrauenswürdige Agentische KI:
Bei Salesforce investieren wir seit 2014 in KI-Forschung. 2016 haben wir unsere eigene vertrauenswürdige KI für Unternehmen, die Einstein-KI, eingeführt. Vor kurzem haben wir unsere neuste Innovation “Agentforce” auf den Markt gebracht, mit der unsere Kunden autonome KI-Agenten für zahlreiche Unternehmensfunktionen entwickeln und einsetzen können. Auf dieser Plattform arbeiten Menschen und KI-Agenten Hand in Hand für maximalen Kundenerfolg.

Dabei unterstützt Salesforce eine gezielte, risikobasierte KI-Regulierung, die zwischen verschiedenen Anwendungsfällen der Technologie unterscheidet. Ein maßgeschneiderter Ansatz ist hier entscheidend, um Vertrauen in die agentenbasierte KI aufzubauen und ihr volles Potenzial auszuschöpfen. Ziel sollte es aus Sicht von Salesforce sein, die Belegschaften zu schulen, um KI-Agenten zu verwalten und zu führen. Darüber hinaus sind zuverlässige Tests und Leitplanken für den Einsatz autonomer KI-Agenten ebenso wie eine klare Kommunikation über KI-Interaktionen und Verantwortlichkeiten und klare Regeln zur Data Governance unerlässlich.

- Salesforce Positionspapier zu Unternehmens-KI: Shaping the Future: A Policy Framework for Trusted Enterprise AI
- Positionspapier zu agentischer KI: Shaping Public Policies for Trusted AI Agents

2. Informationen - AI Energy Score:
Kürzlich hat Salesforce gemeinsam mit der Hochschule Carnegie Mellon University sowie den Partnern Hugging Face und Cohere den „AI Energy Score“ ins Leben gerufen, der die Energie-Effizienz von derzeit 166 Sprachmodellen bewertet und vergleichbar macht. Dabei werden sowohl Open-Source- als auch proprietäre Modelle einbezogen.

Der AI Energy Score misst die Energie-Effizienz von KI-Modellen während der Inferenz, also der Phase, in der ein trainiertes Modell zur Verarbeitung neuer Daten eingesetzt wird. Er bewertet den Energieverbrauch anhand von standardisierten Benchmarks für gängige KI-Aufgaben wie Text- und Bildgenerierung und berücksichtigt dabei Faktoren wie Hardware-Effizienz, Modellarchitektur und Ressourcennutzung.

- Auf einem öffentlich zugänglichen Leaderboard können die aktuellen Bewertungen eingesehen werden.

3. Salesforce | Herzliche Einladung zum Agentforce Breakfast am 15. April 2025

Sehr geehrter Damen und Herren,

Deutschland steht am Anfang einer neuen Legislaturperiode: die Weichen für die politische Zukunft der nächsten Jahre werden gestellt. Dabei geht es auch um nichts Geringeres als die Sicherung der Wettbewerbsfähigkeit der deutschen Wirtschaft und unseres gesamtwirtschaftlichen Wachstums.

Wir möchten dies zum Anlass nehmen, uns mit ausgewählten Gästen – darunter einen/eine Vertreter:in aus Ihrem Team – zu einem unserer Kernthemen auszutauschen: **KI-Agenten – die dritte Welle der künstlichen Intelligenz** (nach der prädiktiven und generativen KI).

Denn bei Salesforce ist agentische KI keine Zukunftsmusik, sondern mit *Agentforce* bereits Realität!

Im Mittelpunkt stehen dabei die Fragen: Wie werden Menschen und KI-Agenten zukünftig zusammenarbeiten, welche Chancen, aber auch Herausforderungen sind damit verbunden und wie gelingt vertrauenswürdige KI?
Diese und Ihre Fragen möchten wir beantworten und danach auch praktisch werden: Nutzen Sie die Gelegenheit, mit Hilfe unserer technischen Experten Ihren ersten eigenen KI-Agenten zu bauen!

Sie möchten dabei sein? Dann würde ich mich über eine kurze Anmeldungsmail mit Angabe von Namen, Organisation und E-Mail freuen.

Mit freundlichen Grüßen,
[redacted]

Datum und Zeit

Dienstag, 15. April 2025

09:00 - 10:30 Uhr

Programm

09:00 Uhr – Ankunft & Kaffee

09:10 Uhr – Einführung Salesforce & Vorstellung Agentforce White Paper

09:30 Uhr – Agentforce Demo

09:45 Uhr – Build your own Agent!

10:00 Uhr – Offener Austausch bei Frühstück

10:30 Uhr – Ausblick und Verabschiedung

📍 **Addresse**

Salesforce

Kurfürstendamm 194

10707 Berlin

--

[redacted]

[redacted] Germany | Salesforce

[redacted] | Salesforce

Kurfürstendamm. 194, D- 10707 Berlin

Mobile: [redacted]

E-Mail: [redacted]

[redacted]

--
[redacted]
[redacted] Germany
[redacted]
Kurfürstendamm. 194, D- 10707 Berlin
Mobile: [redacted]
E-Mail: [redacted]

[redacted]

Shaping the Future:
A Policy Framework for Trusted Enterprise AI

Contents

03 Introduction

04 How Enterprise AI is different from Consumer AI

06 Salesforce's Trusted Enterprise CRM AI

09 Recommendations to Policymakers to Foster Responsibility, Innovation, and Competition in Enterprise CRM AI

13 Conclusion

Introduction

Motivated by the emergence of generative artificial intelligence ("GenAI"), governments worldwide are prioritizing regulatory and policy frameworks for artificial intelligence ("AI").

AI is not a monolithic technology and is used in many different ways and contexts, creating different types of risk. Furthermore, AI ecosystems have increased in complexity with a number of actors involved in various stages of the lifecycle of AI products and services. Some of the main elements of the AI ecosystems are:

Data: **All AI starts with data and there are varying sources to gain the data that powers AI tools such as data aggregators, data brokers, and consumers themselves.**

AI models, including Large Language Models (LLMs): **These models are carefully crafted using research techniques and trained using a combination of public and privately curated data.**

Compute Hardware Providers: **The providers host several layers of data, allowing their customers to both train Large Language Models ("LLM") and process requests to use the model after it's been trained.**

Infrastructure optimization providers: **Entities that provide tools and services that make for more efficient and higher-quality model training such as fine-tuning with specialized, proprietary data to better meet the needs of a particular company.**

Cloud Platforms: **These digital spaces allow developers to tap into the computing power in a cloud deployment model that can also provide applications to help customers organize their data.**

Application Creators: **Entities that build applications on top of models that cater to the unique needs of users and clients to provide services like app building or enhanced business intelligence.**

Denotes a role that Salesforce fulfills.

As governments are working to establish their approach to AI, they should acknowledge these nuances, **focusing on high-risk AI, while clearly delineating appropriate responsibilities** at every layer of the AI value chain to ensure that entities can meaningfully and correctly engage in the development of trusted AI.

While there is diversity in both the AI value chain and its use cases, the current regulatory and policy environment is narrowly focused on a few aspects of AI. The purpose of this paper is to bring attention to the Enterprise AI perspective and its particular relevance to this conversation. Enterprise AI, that is AI developed for and deployed in business settings, has several inherent characteristics that differentiate it from Consumer AI, including: the business model that generally does not monetize customer data; the use-specific contexts for which it is developed and in which it is deployed as opposed to the open-ended nature of Consumer AI; and the higher levels of privacy, security, fairness and accuracy as a result of customer expectations and contractual commitments. Salesforce, as an enterprise company specializes in AI customer relationship management (CRM) solutions and we are taking additional measures that set us apart both from Consumer AI and other Enterprise AI companies.

> **Enterprises need to have the same capabilities that are captivating consumers, but they need to have it with trust, and they need to have it with security.**
>
> **- Marc Benioff**

How Enterprise AI is different from Consumer AI

Context-specific

Enterprise AI applications are usually use-case specific and created for particular work contexts, e.g. productivity or CRM tools, as opposed to the open-ended nature of Consumer AI. Enterprise AI applications often operate in a more closed work environment with limited potential inputs and outputs. Even in the cases of enterprise uses of generative AI, both the prompt and the data that grounds the prompt have been developed to ensure an optimized output for the customer. Consumer AI is usually asked to perform general tasks that can greatly vary depending on the user, including creating images, a real human's voice and likeness, or creating lifelike videos. The very broad contexts in which Consumer AI can be used make it generally more prone to misuse and potentially harmful effects.

Grounded on trusted data

Enterprise AI systems are grounded on and operate on curated data, which generally is consensually obtained from enterprise customers, and are deployed in more controlled environments. This limits the risk of hallucinations and increases accuracy. In contrast, in Consumer AI, the data can come from an array of sources such as the users themselves or, more broadly, the public Internet.

High levels of data privacy, security, and accuracy

Depending on their place in the world, consumers may be covered by data protection and content laws. Enterprise AI companies often go beyond legislative requirements, for instance to be able to service their customers that operate in highly regulated industries, like the government, financial services, and healthcare. These organizations, by virtue of their own regulatory requirements, demand service providers that can from day one, ensure robust privacy, security and accountability controls to prevent bias, toxicity, and hallucinations. Customers also want to have the ability to audit their operations. Enterprise AI companies offer these types of safeguards to their customers, as their reputation and competitive advantage rely on it.

Contractual obligations

The relationship between an Enterprise AI provider and its customers is underpinned by contracts or even procurement rules, which clearly describe the rights and obligations of each party, including third-party vendors, all to provide more reassurance across the value chain. In general, the relationship also means that Enterprise AI companies are handling data in line with the contractual obligations and their ethical guidelines. Further, these same contracts are regularly reviewed to remain aligned with the high standards of the business customers and responsive to the risk environment. In contrast, Consumer AI companies have created terms of service that consumers can read to understand what data will be collected and its use, but lack the ability to negotiate or tailor these terms to all their specific preferences.

Salesforce's Trusted Enterprise CRM AI

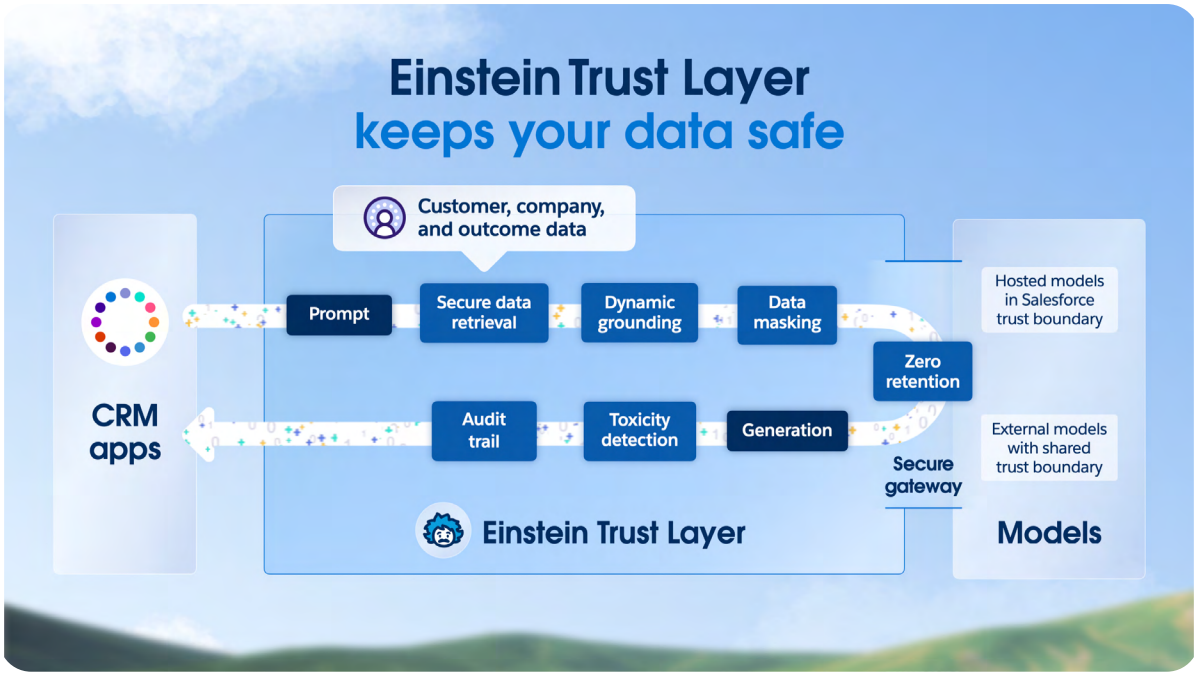
Founded in 1999, Salesforce is a global leader in cloud enterprise software for customer relationship management (CRM), providing software-as-a-service and platform-as-a-service offerings to businesses, governments, and other organizations around the world. Our customers are companies of all sizes and across all sectors that use our tools to connect in new ways with their own customers, employees and citizens.

Salesforce has been active in the research and development of AI technologies for almost a decade. In 2014, we established Salesforce AI Research, and in 2016 we introduced our first AI functionalities into our products under the "Salesforce Einstein" brand. Salesforce Research has published 200+ research papers and registered 300+ AI patents. In 2018, we established our Office of Ethical and Human Use of Technology.

Salesforce is not Consumer AI. **We are trusted Enterprise CRM AI.** We provide our customers with Enterprise AI applications that are highly specialized and relevant to their needs. We are focused on producing specialized AI models that perform everyday work tasks like email generation, summaries of sales or service calls, or surfacing of relevant information during customer support interactions. We also introduced "Einstein Copilot", a conversational AI assistant that our customers can converse with directly to solve issues faster. We also act as an intermediary between other GenAI providers and our customers by integrating their models into our products and services.

At Salesforce, **trust is our #1 value**, which means we develop and deploy AI with trust, security and ethics at the center. We take measures to ensure our AI tools are developed responsibly and we also establish further guardrails to assist our customers in the deployment of trusted AI, including:

- **Our customers' data is not our product.** We have strict rules around the viewing, processing, and disclosure of our customers' data. Salesforce and our partners adhere to our strict data policies and controls, which are also outlined in our contracts and data processing addendum.
- Our Einstein Trust Layer, which is a secure architecture natively built into the Salesforce platform, addresses the concerns of our customers associated with using GenAI. The Einstein Trust Layer is equipped with security guardrails that allow Salesforce customers to benefit from GenAI without compromising their customer data. The Einstein Trust Layer prevents third-party LLMs from retaining sensitive customer data, and masks the data when the prompt is shared with LLMs. Finally, it allows for our customers to audit their AI-generated outputs.



- Our AI Acceptable Use Policy (AI AUP) **guides our customers** in the responsible use of our products with guidelines around prohibited actions such as automated decision-making processes with legal effects, mandating human control, and disclosures.
- We design powerful system-wide controls that put **humans in control** of AI outcomes (what we call "human at the helm"). These controls include our Prompt Builder, audit trails, and robust data controls in our Data Cloud.
- We provide our customers with "**mindful friction**" as they interact with AI, for example, we flag for users when the sorting or evaluation of their data based on certain categories such as zip codes could introduce bias.
- We create, test, and improve the **prompt templates** used by Einstein to provide consistently useful, high-quality responses.
- We employ **red teaming**, a process that involves intentionally trying to find vulnerabilities in a system by anticipating and testing how users might use and misuse it, to make sure that our gen AI products hold up under malicious input.
- We maximize choice for our customers by being **open, extensible, and model agnostic** in our integrations with the emerging AI ecosystem. We aim to enable customers to use the LLMs of their choice with Salesforce's services – whether that be a model provided by a third party, Salesforce, or the customer itself.

Recommendations to Policymakers to Foster Responsibility, Innovation, and Competition in Enterprise CRM AI

Salesforce is committed to building trusted, transparent, open, interoperable and accountable systems. As a service provider servicing organizations of all sizes, in multiple jurisdictions, and across many sectors, we are in a unique position to observe global trends in AI technology and to identify developing areas of risk and opportunity.

We believe that harnessing the power of AI in a trusted way will require governments, businesses, and civil society to work together to advance responsible, safe, risk-based and globally interoperable AI policy and regulatory frameworks.

In this spirit, we offer the following recommendations to convey the Enterprise AI viewpoint and provide some suggestions to policymakers to ensure the development of trustworthy AI.

Definitions of AI actors

The diversity and rapid evolution of the AI ecosystem calls for a nuanced approach to assigning responsibilities to different actors. Therefore, it is important for policymakers to have a clear understanding of the different roles these actors play. In the enterprise space, creators of AI ("AI Developers") often build general customizable AI tools, of which the intended purpose can be low-risk, e.g. Enterprise AI CRM. It is then up to the customer ("AI Deployer") to decide how these tools are employed. The customer also controls the data that is submitted to the AI system. Further, the advent of generative AI has introduced a new role: that of the distributor. This is an entity that is neither developing, nor deploying an AI system, but rather facilitating access. These roles are not mutually exclusive, meaning that one company could operate as a developer, a deployer, and a distributor. Overbroad definitions subject companies to a series of obligations with which they may be unable to comply.

While the AI landscape has and will continue to evolve, at this time, we support narrow, and targeted definitions of developer, deployer, and distributor.

- Developers should be defined as entities that design, code, or produce AI systems. This definition accounts for companies making both predictive and generative AI systems.
- Deployers should be defined as entities that are using or modifying an AI system under their authority. This definition is important because while developers make AI systems, some of these systems are customizable and become specific to the deployer once the deployer inputs its data.
- Distributors should be defined as entities other than the developer or deployer that integrates an AI system into a downstream application or system without substantial or intentional modification. Distributors provide customers with a platform or interface that allows general-purpose systems to be tailored to fulfill more narrow business applications of AI.

Risk-Based Approach

AI policy frameworks should be risk-based and appropriately address the full spectrum of potential harms caused by AI systems. Definitions of high-risk AI should be narrow and take into account the following considerations:

- Activity that has a high-risk of physical impact such as management or operation of critical infrastructure in energy, transportation, and water;
- Economic impact, including automated determinations of eligibility for credit, employment, educational institutions, or public assistance services;
- Government decision-making such as law enforcement/criminal justice and migration/asylum;
- Impact on democracy and the rule of law, for example, the spread of disinformation at scale; and
- Violations of internationally recognized human rights.

Transparency

Salesforce believes that humans and technology work best together. To facilitate human oversight of AI technology, transparency is critical. This means that humans should be in control, and equipped with the documentation to understand the genesis, limitations, and proper use of the AI system. Salesforce is advocating for transparency provisions that are responsive to the nuances of various roles in the AI value chain as described above.

- Documentation: Developers should provide their deployers and distributors with information such as model cards and a document outlining the proper use of the system to help deployers or end-users correctly utilize the system.

  Deployers know both the context and data inputs powering the AI system. They should provide end-users with information about the proper use of the AI system and perform assessments of the AI model. Deployers should also ensure there are clear terms of use for end-users. There should be some reasonable expectation of transparency for end-users and governments if high-risk systems are being utilized.

  Distributors should provide information on their data governance program to both the developers and deployers that interact with their platform. Details should include policies on data retention, data minimization efforts, and audit procedures.

- Human Control: Deployers should ensure that a human is making the final determination when a model is being used in a high-risk situation. These users should also be encouraged to consult other factors beyond the system's recommendations.

- Notice: In instances where deployers are using AI to make high-risk evaluations of individuals, deployers should publish the model's decision-making framework, and provide individuals with the notice that their data is being processed using an AI tool.

- Disclosures: Deployers should make a disclosure when end-users or consumers are interacting directly with automated systems and it is not obvious it is an AI system. Further, AI outputs should be labeled as such to inform consumers where it appears to be human-generated or original content.

Data Governance

As a service provider entrusted with the data of companies large and small, in multiple jurisdictions, and across many sectors, Salesforce understands the importance of data. As a first principle, Salesforce believes that a comprehensive data protection law and other sound data governance practices are foundational to responsible AI.

- Data Minimization and Storage Limitation: Everyone in the AI value chain should endeavor to only store personal data for as long as it's required and for the originally intended purpose of that data. Developers, deployers, and distributors should all have an external policy outlining clear rationales for the retention of data as well as clear timeframes for its deletion.
- Chain of Custody/Data Provenance: Developers, deployers, and distributors, should be clear with users about what is being done with the data with which they are entrusted. For example, Salesforce utilizes changelog abilities which track information on what was created by AI, when, by which system, and how that AI-generated item (action, content, etc.) flowed through the system.

Globally Interoperable and Inclusive AI Policy Frameworks

Globally interoperable AI policy frameworks based on common principles will create more durable, robust and, eventually, long-standing AI norms. Global consistency of AI policy frameworks will further ensure that the challenges presented by AI can be tackled collectively, whereas the benefits can be shared by many. It is also important that global efforts on AI governance remain inclusive, incorporating views from diverse geographies, economic sectors, and disciplines.

- Salesforce supports the multilateral consensus-driven work that is occurring in spaces like the G7, the United Nations, and the Organization for Economic Co-operation and Development (OECD).
- To complement robust AI policy frameworks, Salesforce supports the free flow of data to and from countries provided it is subject to appropriate safeguards. The ability to transfer data between jurisdictions in a seamless and responsible manner supports both a high level of data protection and continued innovation.

Conclusion

Salesforce believes in the tremendous opportunities that AI can bring to individuals and businesses alike - with proper governance. To that end, Salesforce supports a multi-stakeholder approach to AI policymaking, prioritizing the design of flexible, nuanced, and adaptive policies that respond to the rapid pace of AI innovation. Enterprise AI companies have unique perspectives on how to tackle some of the most pressing concerns policymakers are grappling with. We look forward to sharing our expertise on trusted Enterprise AI CRM with governments, industry, and civil society as they are debating AI governance efforts.

Thank You

Learn more about Salesforce Public Policy here.

The Next Frontier in Enterprise AI: Shaping Public Policies For Trusted AI Agents

Contents

03 Introduction: AI agents – Boosting productivity, empowering humans

05 Key considerations for AI agents

06 Agents powered by trusted enterprise AI

08 Looking Ahead: Fostering an agent-ready digital ecosystem

- Government services as a role model for agentic AI and humans together
- Public policy that enhances stakeholder trust
- Workforce skilling for agentic AI and humans together

15 Conclusion

Introduction

AI agents are a technological revolution - the third wave of artificial intelligence after predictive and generative AI. They go beyond traditional automation, being capable of searching for relevant data, analyzing it to formulate a plan, and then putting the plan into action. Users can configure agents with guardrails that specify what actions they can take and when tasks should be handed off to humans.

AI agents are also an exciting next step in the economic revolution powered by AI. With their ability to independently tackle complex problems, AI agents promise to deliver the agility, efficiency, and competitiveness that organizations of all sizes seek. Unlike previous tech transformations that demanded years of expensive infrastructure development, AI agents can be quickly and easily built and deployed, significantly increasing capacity.

It is estimated that **75% of AI's ultimate value** lies in the "front office," where businesses directly interact with their customers. Yet 41% of employee time **is currently lost** to repetitive, low-value tasks that have little to do with this essential front-office work. AI agents have the potential to ease these burdens, allowing workers to focus on more meaningful and strategic tasks, making work not only more productive, but more fulfilling.

In surveys conducted for the OECD, four in five workers said that AI had improved their performance at work and three in five said it had increased their enjoyment of work. Workers were also positive about the impact of AI on their physical and mental health, as well as its usefulness in decision making.

Agentic AI is already here. Innovative organizations have started deploying and benefitting from AI agents in concrete ways. However, for governments, enterprises, and the workforce to be able to harness the full potential of AI agents, they will need more than just the right technology; they will need the right public policies to help them become agent-ready. At a time when global policymakers seek pathways to economic growth, the advent of AI agents offers a unique opportunity to think deeply about the policies that will enable the diffusion of trusted AI in government and the enterprise, and equip the workforce with the necessary skills to unlock AI's full potential for enhanced productivity and more meaningful work.

This paper outlines key considerations for designing and using AI agents, and describes how enterprise AI can help ensure this is done in a safe and responsible way. It also provides a series of recommendations for policymakers who want to accelerate adoption of these new, productivity enhancing tools.

With Agentforce to handle routine inquiries, publisher Wiley's employees have more time to focus on complex cases, outperforming the company's previous chatbot by 40% case resolution in the first few weeks of use.

What are AI agents?

AI agents build on earlier innovations, including chatbots and AI assistants. These all play a role in task automation; however, there are differences in their levels of sophistication and personalization in serving their users and the range of tasks they can perform.

Grounded on trusted data

Chatbots are relatively simple algorithms that provide answers to simple questions, often based on predefined rules. Some chatbots may use basic natural language processing, while others may give pre-programmed answers.

AI assistants or copilots use more advanced natural language processing and large language models (LLMs) to address a wider range of questions that the AI system may not be explicitly programmed to answer in advance. When prompted, they produce outputs based on data used to train or fine tune the LLM, or may be able to retrieve information from integrated data sources.

AI agents do not just answer questions in response to user prompts - they can also initiate, plan, and execute multi-step processes to achieve user-defined objectives. AI agents can adapt to the evolving needs of the enterprise, continuously learning from every interaction, becoming more effective and accurate over time. They can integrate with different underlying models and data streams, and be configured to handle diverse tasks and workflows. While chatbots and assistants might still be useful for simpler interactions, agents are the better choice for more complex scenarios that require reasoning.

Using Agentforce, reps at leading ed tech company Carnegie Learning were able to focus on high-value prospects, reducing account research time from up to an hour to just five to ten minutes–a reduction up to 92%

Key considerations for AI agents

- **Humans working with agents:** Employees, from business leaders to frontline workers, will need new skills to configure, task, manage, and oversee AI agents. Agents will need to be easy to program and use in a variety of contexts.
- **Reliability:** AI agents must be carefully designed and equipped with guardrails to ensure clear and smooth handoffs to humans, as well as to minimize, flag, and correct hallucinations. Careful engineering and robust testing are required to ensure the accuracy and reliability of agents.
- **Fluency across multiple domains:** AI agents will interact with users and third parties within and outside of the enterprise, retrieving, interpreting, and acting on different types of information across these domains. This requires advanced programming and thoughtful integration of business processes and data systems.
- **Transparency and explainability:** Users need to know when they are interacting with an AI agent instead of a human. Customers, regulators, and the public will also expect AI agents' outputs and the sequence of steps they follow in taking an action to be transparent and explainable.
- **Accountability:** It will be important to provide clarity around who is responsible for ensuring that the agent is working properly and that its output is accurate.
- **Data governance and privacy:** AI agents may require access to personal or other sensitive data to complete their assigned tasks. For users and enterprises to trust AI agents, they will have to operate with high standards of privacy and data security.
- **Security:** Like other AI applications, agents may be vulnerable to adversarial attacks, where malicious inputs are designed to deceive the AI into producing bad outputs. As AI agents take on increasingly complex tasks, adhering to best practices for AI safety and quality control will be essential.
- **Ethics:** Companies that use AI agents should ensure they follow ethical guidelines consistently and reliably. This will require developing new protocols and norms for autonomous AI systems, fostering effective human-AI collaboration, and building consensus and confidence in decision-making processes.
- **Agent-to-agent interactions:** Common protocols and standards will be important to instill trust and help ensure controlled, predictable, and accountable AI agent behavior. Fundamental to this is a secure information exchange environment and, when relevant, audit trails of agent-to-agent interactions.

Agents powered by trusted enterprise AI

Motivated by the emergence of generative artificial intelligence (“GenAI”), governments worldwide are prioritizing regulatory and policy frameworks for artificial intelligence (“AI”).

At Salesforce, trust is our #1 value. From the outset, we have been 100% focused on ensuring that we develop, deploy, and distribute AI with trust, security and ethics at the center. Enterprise AI demonstrates higher levels of privacy, security, and accuracy, and Salesforce implements a number of measures to address business concerns with AI.

Agentforce is our newest AI innovation, bringing together humans with agents to help enterprises connect to their customers in a whole new way. Agentforce is a suite of out-of-the-box AI agents and a set of tools to build, customize, and deploy them across any industry. Agents built with Agentforce can complete tasks such as analyzing data, making decisions, and taking action to answer customer service inquiries, qualify sales leads, and optimize marketing campaigns. All this is built on our trusted Salesforce platform, which means that customers do not have to bear the high technical complexity and large financial cost of building or training their AI from scratch - the same way that customers do not have to build their own cloud infrastructure. Agentforce provides secure access to third-party LLMs through the Einstein Trust Layer and uses retrieval-augmented generation (RAG) to apply customers’ trusted data sources to those LLMs to ensure reliable outputs.

Watch Video

Our core values and trusted AI principles remain central to Agentforce.

- **Trust:** The core tenet of our business model remains unchanged: our customers' data is not our product, meaning that, when it comes to AI, customers control how their data is used.
- **Data:** Integrity and quality of data are critical to ensure effective problem-solving. Salesforce's Data Cloud and our new retrieval-augmented generation (RAG) functionality allow organizations to securely leverage their own trusted data to reduce the risk of hallucinations and improve the performance and reliability of the agents.
- **Accuracy:** The Atlas Reasoning Engine is the system that acts as the "brain" inside Agentforce. It starts by evaluating user queries and refining them for clarity and relevance. Then, it retrieves the most relevant data and builds a plan for execution. Our research showed that the results Agentforce could deliver were twice as relevant and 33% more accurate than other available solutions.
- **Privacy & Security:** Every Agentforce interaction runs through the Salesforce Einstein Trust Layer, a comprehensive set of security measures and protocols designed to protect the privacy and security of customer data. The bedrock principle of the Trust Layer is zero data retention, meaning that data is used to generate outputs but never to improve the underlying models. In addition, the Trust Layer offers auxiliary features like dynamic grounding, toxicity detection at the input and output levels, and an audit trail to track AI agent actions and outputs.
- **Guardrails:** Agentforce comes with a set of features and controls that reinforce trusted behavior, and prevent deviations from the intended behavior of AI agents. Customers can use natural language to create instructions and guardrails for their agents, including which actions an agent can and cannot take or when to escalate or hand off a task to a human. Customers can also easily adjust the guardrails to fit their specific needs.
- **Open ecosystem:** Agentforce can easily connect to Salesforce-provided as well as third-party enterprise systems, data lakes, and warehouses that our customers use, allowing them to make the most of their technology investments. We have also launched the world's first agent partner network, enabling customers to access a catalog of third-party skills, actions, and agents to increase the capabilities of their agents.
- **Responsible innovation:** Our guiding principles for the responsible development of agentic AI are accuracy, safety, honesty, empowerment, and sustainability. Our Office of Ethical and Humane Use has designed a set of trust patterns, that are additional to the Einstein Trust Layer, to ensure that Agentforce is reliable and transparent. For instance, Agentforce products use standard language to alert administrators and agent managers when they're about to implement or use AI agents. These notes highlight the capabilities and limitations of AI, ensuring a clear understanding of its impact and potential.

- **Testing & evaluation:** To ensure the reliability of our AI agents and relevant safeguards, we conduct rigorous testing and **red teaming**, including adversarial testing. Before launching Agentforce, we subjected our AI agents to over 8,000 adversarial inputs to pressure-test their boundaries. We have also published the world's first **LLM benchmark** for customer relationship management (CRM) models.
- **Sustainability:** Our **Sustainable AI Policy Principles** prioritize managing and mitigating the environmental impact of AI models. Rather than using a single large, energy-intensive model for every task, Agentforce leverages a variety of optimized models specifically tailored to each use case. This approach enables high performance with a fraction of the environmental impact.
- **Government partnerships:** We continue to work with governments around the world to advance responsible AI. In 2024, we **reported** on our progress on the White House AI Voluntary Commitments and signed onto the **EU AI Pact** and **Canada's voluntary AI Code of Conduct**. We also participate in a number of government-led initiatives including **UNESCO's** Global Business Council for the Ethics of AI, the **US National AI Advisory Committee**, and **Singapore's AI Verify Foundation**.

Looking Ahead: Policies to foster an agent-ready digital ecosystem

Policymakers around the world are grappling with questions about the risks and opportunities of AI while trying to keep pace with rapidly advancing innovation. Although agents are the latest technology breakthrough, the fundamental principles of sound AI public policy that protects people and fosters innovation remain unchanged: risk-based approaches, with clear delineation of the roles of different actors in the ecosystem, supported by robust privacy, transparency, and safety rules.

With this in mind, it is time to think beyond regulating how AI is built and used. Policymakers should focus on creating the right conditions to support **wide adoption of trusted agentic AI across industries and geographies.** It is imperative that no nation or community is left behind. **Equipping the workforce** with the necessary skills to harness the potential of AI agents for less tedious, more meaningful and productive work should be a cornerstone of any governmental policy aiming to advance AI development and diffusion.

Government services as a role model for agentic AI and humans together

Governments today face ever-increasing pressures to serve citizens with constrained resources. Research surveying people in over 40 countries has shown that 75% expect government service quality to be on par with leading private sector companies, while 72% are comfortable with personalized government digital services. Amidst these pressures, there is clear demand for trusted innovative tools from government workers themselves - a recent survey across 14 countries estimated that 49% of government workers have used unapproved generative AI tools at work.

Governments have enormous opportunities to safely leverage AI agents - in particular their speed, responsiveness, personalization - to enhance citizen-facing services. For example, if a citizen wants to check on the status of their application for a license or a public benefit, an agent-powered interface could at any hour assist them in ascertaining the status of their application, locate public information on policies and application procedures, or schedule an appointment to access a related service. These interactions would have previously taken the citizen far longer, required valuable government employee time, may have only been available during business hours, or been left undone altogether. By deploying agents in these contexts governments can make the most of their limited resources while building citizen satisfaction with and trust in services.

In addition to leading by example in digital transformation, **governments can play a pivotal role in boosting diffusion of trusted AI solutions, including AI agents, in the private sector.** As governments search for new ways to promote economic growth and combat inflation, this is a key moment for national economies. The rapid development and commercialization of AI tools have created opportunities for businesses large and small around the world to rapidly become more productive and gain a competitive edge by adopting frontier technologies. But to achieve the benefits, businesses need to overcome resource constraints, rapidly acquire new commercial knowledge and practical skills, and have regulatory clarity. For governments to position their businesses for success, they need to set ambitious goals and identify and remedy the blockers that hold back the uptake of new technologies by enterprises.

A recent report of the European Commission showed that the uptake of AI, cloud and big data by European companies is well below the Digital Decade target of 75%, with only 17% of businesses projected to use AI by 2030.

Recommendations:

- **Ambitious AI adoption strategies:** Many governments worldwide have established AI strategies, but in the face of rapidly developing applications of AI these may no longer reflect cutting-edge opportunities for economies and lack clear targets for industry adoption. Governments should conduct comprehensive analyses of potential blockers, like red tape, lack of trust, or legal clarity, etc. that stall AI adoption. They should also work with private sector stakeholders to foster the diffusion of AI tools and solutions within industries, for example through AI Showcases at industry and trade convenings. Earmarking funding, in the form of a National AI Adoption Fund, to strengthen AI diffusion in the private sector would also be a positive measure, particularly for smaller and medium-sized enterprises that often face more challenges in their digital transformation journeys.

- **Agent-ready governments:** Governments should adopt AI-first approaches to transform public administration, with clear measures to encourage AI adoption for all government agencies, including by ensuring that government technology systems are modernized and agent-ready. Government IT modernization strategies should support the preconditions that smooth the path to adoption of AI agents, including by redoubling efforts to ensure internally coherent and interoperable data systems and integration strategies.

- **Public procurement modernization:** Governments should review procuring procedures to ensure flexible, outcome-based standards that do not inadvertently hinder the procurement of innovative solutions like AI agents. Memoranda of Understanding or similar framework agreements with technology suppliers can help streamline the adoption of new solutions at scale by easing the path for procuring agencies to verify compliance of these solutions with procurement guidelines or other requirements. To help procurement officers traverse the AI learning curve and gain confidence, governments should establish processes to share information and best practices in government AI agent use cases, contracting, and management. Senior government officials should also be provided with "AI for Executives" courses to ensure that those accountable for government technology decisions understand the benefits and can appropriately oversee AI procurement.

- **Chief AI Officer:** Governments should coordinate their AI adoption efforts to ensure best practices are shared across different agencies and departments. While each government agency will benefit from having its own AI experts who assess AI-related rules for the field of activity that the agency supervises, a Chief AI Officer could promote government-wide approaches to AI adoption, advocating for an AI-first, and even AI agent-first, approach to public administration. This role could also be an extension of the remit of existing government Chief Technology Officers, Chief Data Officers, and/or Chief Information Officers.

Public policy that enhances stakeholder trust

Public policy frameworks should aim to establish and maintain trust with stakeholders, creating the right conditions for the safe and responsible development and use of AI agents, while promoting innovation and competition. Fundamental principles, such as a risk-based approach, appropriate allocation of responsibilities across the supply chain, and predictable and proportional privacy laws remain central to the proper governance of AI agents.

International AI governance efforts should include diverse geographies (Global South), sectors (developers, deployers, distributors, users, large and small companies), and disciplines (scientists, ethicists, technical experts, policymakers, academia, privacy practitioners, and civil society).

Recommendations:

- **Risk-based policy frameworks:** To build on the benefits that enterprises are already seeing from AI adoption, governments should pursue nuanced policies that establish effective guardrails around AI while also enabling creative experimentation and innovation. In our white paper "**Shaping the future: A policy framework for trusted Enterprise AI**", we outlined why policymakers should adopt tailored approaches based on appropriate definitions of risk. These should focus on the applications or contexts of use of AI that are likely to cause significant harm to the rights and freedoms of an individual, including economic and physical impact, and also to the rights to privacy and to be free from discrimination. Enterprise AI solutions such as Agentforce are generally lower risk than consumer-facing AI technologies, because they are context-specific, are grounded on trusted customer data, and comply with higher levels of data privacy, security, and accuracy.
- **Appropriate allocation of responsibilities:** Agentic AI is made possible by different entities in the AI value chain. The data, LLM, platform, application, and fine-tuning could all be performed by different entities. To effectively address risk, policy frameworks should acknowledge this nuance and assign responsibilities carefully and in relation to the amount of information and level of control different parties have over the distinct elements comprising an agent. In the enterprise space, the platform provider may be best positioned to recommend and evaluate responsible configurations; the LLM provider to describe the underlying model and explain its decision-making algorithm; and the customer providing the data that goes into the AI agent to control how such data is used as well as the context in which the agent is utilized. Our **proposed definitions of AI developer, deployer, and distributor** take the foregoing considerations into account. Appropriate allocation of responsibilities is also a critical step in avoiding the privacy harms that may result by assigning to entities inconsistent (or even conflicting) responsibilities under AI and data protection regulatory schemes. For example, requiring that data processors/service providers take an active hand in monitoring or modifying data under AI laws – especially in ways typically reserved for data controllers/businesses under privacy laws – will only increase the number of parties that have access to and can manipulate consumers' personal data, yielding more vectors of attack.

- **Supportive data privacy rules:** Because data quality is so important to effective AI agents, care should be taken to ensure that rules facilitate productive, secure, and transparent uses of data, including personal data. Comprehensive privacy protections and clear rights and obligations will enable both enterprise users and consumers to feel confident managing their sensitive information in AI systems incorporating agents. Alignment on core concepts such as a delineation between data controllers and data processors is critical to achieve sound risk-based and proportionate regulation. Privacy rules should also facilitate safe and compliant uses of data, including through the deployment of Privacy Enhancing Technologies (PETs).
- **Documentation:** Creators of agentic AI should provide proper documentation on both the responsible configurations and acceptable uses of agents to their customers.
- **Transparency:** Where AI agents are interacting with end-users, policy frameworks should require that it is made clear that the interaction is with an AI tool and not a human. Furthermore, content created and autonomously delivered by AI should be clearly marked as such (e.g. disclosures in agent responses to consumers, or use of watermarks on an AI-generated image).
- **International alignment:** Globally interoperable rules and norms will help with reinforcing trust in and diffusion of agentic AI. Coordination among National AI Safety Institutes should continue and be further strengthened with regular meetings including broad industry participation as appropriate. Ongoing efforts by the G7, the OECD, and the UN are examples of promising initiatives to advance international cooperation on AI governance. Inclusion of diverse voices in these conversations, including from the Global South, will ensure a holistic approach to dealing with the risks and opportunities of AI.

Workforce skilling for agentic AI and humans together

The opportunities of AI agents to enhance productivity and creativity, augment the work of humans, and create new jobs, are limitless. However, the rapid pace of innovation is raising concerns about AI replacing human workers. AI isn't the first automated technology to stir fears about job displacement. Our societies have experienced similar shifts before, such as when the Internet revolution changed the way we work and interact forever. Salesforce itself was born out of the movement to the cloud, and our ecosystem alone has created its own job surge.

As AI reshapes the way we work, the focus should be on the training, creativity, and critical thinking skills that are uniquely human. The global workforce must be equipped to use this technology safely and confidently. Equitable access to AI is essential so that all individuals, regardless of their background or location, can benefit from these advancements and contribute to a more inclusive and prosperous future.

> **While AI's potential is vast, its highest purpose is realized when it enhances — rather than replaces — the unique judgment, creativity, and empathy that only people can provide.**
>
> **- from "State of the AI Connected Customer" report**

Understanding the future of work means understanding which tasks, as opposed to whole jobs, should be automated, and which should not. AI itself can help us identify the tasks it can best handle and provide better tools to understand the skills workers might need in the future, designing personalized learning paths to assist with the transition.

Governments should prioritize public policies and collaborate with civil society and industry to equip workers with the right skills to thrive in the new opportunities created by agentic AI. While each jurisdiction will adopt capacity-building policies tailored to their national or regional needs, there is already a robust set of recommendations and approaches from international organizations like **UNESCO**, the **OECD**, and the **IMF** that can serve as a basis to build on.

Organizations must focus on training their employees to use and manage AI agents, but also to be strategic, make good decisions, and be creative, ensuring that technology remains in the service of humans. **Trailhead**, our free online platform, has expanded its courses to offer AI-specific skills training, including AI fundamentals, ethical AI use and prompting. Salesforce also **announced** that it will offer its existing premium AI courses and AI certifications free of charge and available to anyone on Trailhead through the end of 2025.

Additionally, we make our spaces around the world available for on-site training sessions. In June 2024, Salesforce opened its first **AI Center in London**, and will unveil a new pop-up AI Center at its headquarters in San Francisco in 2025, with plans to roll out additional training centers in key hubs around the world like Chicago, Tokyo, and Sydney. These centers will bring together industry experts, partners, and customers to advance AI innovation alongside providing critical upskilling opportunities.

The **Salesforce Accelerator – Agents for Impact** is a new initiative that provides nonprofits with technology, funding, and pro bono expertise to help them innovate and develop AI-powered agent solutions to scale community impact.

Recommendations:

- **Study the impact of AI agents on jobs:** Much is not yet known about how AI agents will impact the workforce and how this will vary among different countries, communities, sectors, and types of work. Understanding the specific contours of this impact for stakeholders is necessary for national, regional, and local governments to shape informed policies that support their workers and industries during this technological shift. Policymakers should evaluate how AI agents influence job creation, task transformation, and skill requirements in order to design targeted interventions.
- **Expand and invest in public-private partnerships for workforce reskilling:** Governments, industry, NGOs, and civil society each bring necessary perspectives and capabilities. Policymakers should focus on public-private partnerships specifically geared towards reskilling the workforce to adapt to the unique challenges and opportunities posed by AI agents. These can include:
  - Industry-specific reskilling initiatives that bring together workers from sectors where AI agents are more intensively transforming job roles, along with educational specialists and technologists.
  - Collaborative programs involving industry, universities, and NGOs to connect AI-ready talent with suitable job opportunities and facilitate the transition into the workforce.
  - Workshops aimed at providing businesses with practical guidance on adopting trusted enterprise AI solutions, including the integration of AI agents.
  - Collaboration among AI companies, industry associations, and labor unions to develop AI agent skills frameworks that are transferable within and between industries.

Salesforce has joined UNESCO's call to action to build ethical AI, prioritizing responsible skills training and governance for a fair AI future.

Salesforce collaborates with governments to equip businesses of all sizes with the right AI skills and innovation, with programs like Data + AI Boost SME Program in Singapore.

- **Expand funding for AI skills training:** Funding programs should focus on upskilling workers in industries at risk of AI disruption, helping them transition to higher-value roles that involve managing and collaborating with AI systems. Key skills for AI preparedness in ICT roles include AI literacy, data fundamentals, and prompt engineering. As AI agents become integral to human-agent teams, success will require more than just technical engineering expertise. Equally important will be strong communication, collaboration, and problem-solving skills that enable individuals to work seamlessly with both AI systems and humans, fostering effective teamwork in AI-driven environments.

- **Integrate AI skills into the curriculum:** Preparing the next generation for an AI-first future is essential. Education and vocational curriculums should include ongoing training in AI concepts, skills, ethical considerations, and its potential future integration as the discipline evolves. Educators should also be enabled and reskilled to integrate AI into the curriculum and meet the needs of future classrooms.

Conclusion

For the past 25 years, Salesforce has led our customers through every major technological shift: from cloud, to mobile, to predictive and generative AI, and, today, agentic AI. We are at the cusp of a pivotal moment for enterprise AI that has the opportunity to supercharge productivity and change the way we work forever. This will require governments working together with industry, civil society, and all stakeholders to ensure responsible technological advancement and workforce readiness. We look forward to continuing our contributions to the public policy discussions on trusted enterprise AI agents.

Thank You

Learn more about Salesforce Public Policy here.