Travis Hall,
National Telecommunications and Information Administration, U.S. Department of Commerce,
1401 Constitution Avenue NW,
Room 4725,
Washington, DC 20230

Re: AI Accountability Policy Request for Comment
Docket No. 230407-0093

Comments of Salesforce, Inc.

Salesforce, Inc. ("we," "us," or "Salesforce") appreciates the opportunity to respond to the National Telecommunications and Information Administration's ("NTIA") request for comments on Artificial Intelligence ("AI") system accountability measures and policies ("RFC"). We welcome the efforts of the Administration to address Artificial Intelligence policies.

About Salesforce

Founded in 1999, Salesforce is a global leader in cloud enterprise software for customer relationship management (CRM); providing software-as-a-service ("SaaS") and platform-as-a-service ("PaaS") offerings to businesses, governments and other organizations around the world. Our customers represent companies of all sizes and across all sectors. Our business model is cloud-based and low code, allowing for faster deployment of technologies and greater agility. We help our customers connect with their customers – or employees or citizens – in a whole new way using cloud, data and AI technologies.

Salesforce & AI

Salesforce's AI capability, called "Einstein", is built into the Salesforce platform, and is designed to combine artificial intelligence with Salesforce's suite of enterprise services, democratizing the power of AI for every Salesforce user. Einstein enables organizations of all sizes to deliver smarter, more personalized customer experiences by automatically discovering relevant insights, predicting future behavior, proactively recommending best next actions and automating tasks. And now, generative AI (GAI) has brought a new dynamic to these capabilities. Salesforce's Einstein GPT will augment our users' capabilities by providing AI-generated content across our different products and IT interactions.

415 Mission Street
San Francisco,
California 94105
Salesforce, Inc.
/salesforce
@salesforce

At Salesforce, we have determined that the ethical use of technology must be clearly addressed. Technology is not inherently good or bad. It's what we do with it that matters. That is why we made ethical use of technology a strategic initiative at Salesforce through the establishment of the Office of Ethical and Humane Use of Technology (OEHU) in 2018. The office works across product, law, policy, and ethics to develop and implement a strategic framework for the ethical and humane development, deployment and use of technology. Through this work, we have established a set of guiding principles on the ethical use of technology that protects human rights, data privacy, and human safety.

We believe that the tremendous benefits of AI should be accessible to everyone, while ensuring that those technologies remain safe and inclusive. At Salesforce, we are committed to providing our employees, customers, and partners with the tools they need to develop and use AI safely, accurately, and ethically. Our commitment to ethical AI consists of adherence to the following principles:

- **Responsible**: To safeguard human rights and protect the data we are entrusted with, we work with human rights experts, and educate, empower and share our research with customers and partners to enable them to use AI responsibly. We strive to comply with the laws and values of the markets in which we operate. We strive to adhere to the highest security and safety protocols.
- **Accountable**: Accountability to customers, partners and society is essential. Independent feedback should be sought for continuous improvement of practices and policies and work to mitigate against harm to customers and consumers. We seek stakeholders' feedback, take guidance from our Ethical Use Advisory Council, and conduct our own data science review board.
- **Transparent**: Our customers should be able to understand the "why" behind each AI-driven recommendation, output and prediction so they can make informed decisions, identify unintended outcomes, and mitigate harm. We strive for model explainability and clear usage terms, and ensure customers control their own data.
- **Empowering**: AI is best utilized when paired with human ability, effectively augmenting people and enabling them to make better decisions. Accessible AI promotes growth and efficiency, and benefits society as a whole.
- **Inclusive:** AI should respect the values of all those impacted, not just of its creators. To achieve this, we test models with diverse data sets, seek to understand their impact, and build inclusive teams.

Because of the rapid evolution of the technology, along with the opportunities and challenges emerging from the use of generative AI, we have gone one step further and articulated an additional set of guidelines meant to guide the development of trusted GAI, at Salesforce and beyond.

415 Mission Street
San Francisco,
California 94105
Salesforce, Inc.
/salesforce
@salesforce

- **Accuracy:** We need to deliver verifiable results that balance accuracy, precision, and recall in the models by enabling customers to train models on their own data. We should communicate when there is uncertainty about the veracity of the AI's response and enable users to validate these responses.
- **Safety:** As with all of our AI models, we should make every effort to mitigate bias, toxicity, and harmful output by conducting bias, explainability, and robustness assessments, and red teaming. We must also protect the privacy of any personal data present in the data used for training and create guardrails to prevent additional harm.
- **Honesty**: When collecting data to train and evaluate our models, we need to respect data provenance and ensure that we have consent to use data. We must also be transparent that an AI has created content when it is autonomously delivered.
- **Empowerment:** There are some cases where it is best to fully automate processes but there are other cases where AI should play a supporting role to the human – or where human judgment is required. We need to identify the appropriate balance to "supercharge" human capabilities and make these solutions accessible to all (e.g., generate ALT text to accompany images).
- **Sustainability:** As we strive to create more accurate models, we should develop right-sized models where possible to reduce our carbon footprint. When it comes to AI models, larger doesn't always mean better: In some instances, smaller, better-trained models outperform larger, more sparsely trained models.

As AI becomes ubiquitous in a modern economy, policy makers and industry should work together to establish guardrails ensuring the ethical development and utilization of this powerful tool.

AI & Regulation

Salesforce is committed to building trusted, transparent, and accountable AI systems that prioritize fairness, accuracy, privacy, and positive societal impact. We are committed to responsibly building and deploying trusted AI for all our customers and for all anticipated uses that we allow.

Emerging technologies like AI and generative AI hold great potential for consumers, businesses and society as a whole. However, they have also resulted in a wide range of risks including issues of accuracy, bias and inequality, privacy and security, and sourcing of content. The context in which technology is used matters, and some business models are more likely to create higher risks for their users than others.

Regulation has a critical role both to protect people and also to foster innovation. AI regulation should apply a risk based framework to proportionately address a full spectrum

of harms that might be caused by AI. In a risk-based framework, the more rigorous AI regulatory obligations should focus on the high-risk AI applications that are more likely to cause the most significant impacts or harms on individuals. Similarly in a risk-based framework, AI regulations should have less-intensive obligations for low-risk applications.

We have long held concerns that certain technologies, like facial recognition, currently pose a high risk of harm and discriminatory impacts, particularly for underserved communities. Because of these concerns, we don't offer facial recognition capabilities in our products. But we continue to engage with certain categories of emerging technologies; for example, the aforementioned guidelines for the responsible development of generative AI, which we're committed to following as the technology continues to advance.

However, we also believe that with additional research and thoughtful regulation, the industry can develop appropriate safeguards that may allow companies to responsibly unlock the high potential that advanced technologies offer. This involves learning from each other across disciplines including technologists, ethicists, engineers, academics, and members of historically marginalized communities. We welcome the opportunity to share our thoughts about critical elements for creating trusted AI.

AI Accountability

- What is the purpose of AI accountability mechanisms such as certifications, audits, and assessments?

Accountability to our customers, partners, and society is essential. As a company, Salesforce works to invite feedback from different stakeholders (e.g. customers, employees, and both external human rights and ethics experts) through advisory boards and open dialog, then incorporating that feedback into the deliberative process of building products and features. We also collaborate with peers through industry groups and civil society forums to continuously improve practices and policies, as well as share insights and give back to the community of practice. Just as thoughtful regulation will take collaboration across different disciplines, the creation of trusted AI takes a range of different tools and input from the whole of the AI ecosystem.

Certifications, audits, and assessments can help with the creation and demonstration of trusted AI. To create these systems, tools like audit reports and certifications can provide information about the system, which can be critical in diagnosing potential or real security issues as well as monitor for unexpected changes, usage trends, and detect potential abuse. However, these tools need to be built on accepted AI definitions, thresholds, and norms that are not yet established in the United States.

415 Mission Street
San Francisco,
California 94105
Salesforce, Inc.
/salesforce
@salesforce

Another tool for accountability is an impact assessment. We believe there's substantial value in performing data impact assessments and/or algorithmic impact assessments when there's a high potential or risk of consequential harm, like those included under the American Data Privacy and Protection Act (ADPPA). Data sets used to train AI models can be discriminatory, and discriminatory training data will likely yield discriminatory model outputs. To counter this, industry should purposefully and proactively work to identify safe and representative datasets that can then be used as a means to counter any latent discrimination or bias that may exist. These types of assessments have led us to our conclusions on facial recognition and allowed us to more safely develop appropriate safeguards and controls in other AI products we do offer.

Salesforce recognizes the need for AI accountability regimes, but they need to be risk-based, proportionate and constructive and comport with the different roles in the AI ecosystem.

- Given the likely integration of generative AI tools such as large language models (*e.g.,* ChatGPT) or other general-purpose AI or foundational models into downstream products, how can AI accountability mechanisms inform people about how such tools are operating and/or whether the tools comply with standards for trustworthy AI?

As previously outlined in our principles, we believe new rules and guidance should require automated systems (including conversational AI and chatbots) to disclose to users that they are interacting with an AI system. Products that simulate another person must either have that person's explicit consent or be clearly labeled as "simulated" or "parody." Further, regulation should enable humans to have sufficient information to truly know if the AI recommendation is fair/safe/accurate to empower them to be a meaningful check on the system. Consistent with the risk-based approach, human intervention in respect of automated decisions should be mandated only in the instances where the decision involves legal or similarly significant effects.

- Are there ways in which accountability mechanisms are unlikely to further, and might even frustrate, the development of trustworthy AI? Are there accountability mechanisms that unduly impact AI innovation and the competitiveness of U.S. developers?

As with other elements of AI regulation, the use of accountability mechanisms should be assessed in a risk-based framework. There are several points in the AI development lifecycle where accountability checks can be helpful. However, whether it is assessing the extent to which a dataset may be considered biased or the manner in which fairness is built into a system, making such audits compulsory, widespread, or systematic without any regard to

415 Mission Street
San Francisco,
California 94105
Salesforce, Inc.
/salesforce
@salesforce

the system (and risk level) itself could create burdensome obligations that do not convey meaningful protections for consumers.

Existing Resources and Models

- What AI accountability mechanisms are currently being used? Are the accountability frameworks of certain sectors, industries, or market participants especially mature as compared to others? Which industry, civil society, or governmental accountability instruments, guidelines, or policies are most appropriate for implementation and operationalization at scale in the United States? Who are the people currently doing AI accountability work?

As outlined in our principles, independent feedback should be sought for continuous improvement of practices and policies. Without AI regulations, companies like Salesforce have worked to build methods to ensure the development and deployment of trusted AI. We continuously seek feedback and engage stakeholders both internally and externally to help us uphold our commitment to trusted AI as well as conduct bias, explainability, and robustness assessments, and red teaming. We have committed to working with external human rights and technology ethics experts through our Ethical Use Advisory Council. We also collaborate with our peers in industry groups and in civil society forums to continuously improve our practices. An example of this is our partnership with the World Economic Forum (WEF) and our integration of their best practices.

Internally we have several mechanisms to ensure the ethical development and deployment of our products. OEHU was established to help facilitate wider and systemic accountability in our products. At the time, the office was a new kind of department in a major technology company, with people dedicated not only to thinking about how our products are used but also to how our products are built in the first place. When examining Salesforce products the OEHU works in partnership with teams to conduct product reviews to evaluate if an AI application could, a) be used in life-altering decision-making (e.g., health, access to benefits, justice), b) have a risk of violating human rights and or impact vulnerable populations, c) be used to spread disinformation, or d) use special categories of personal data (e.g., biometric, SSN, financial data, passport number, health data). OEHU has created a number of policies and processes in place to ensure the ethical development and deployment of AI including but not limited to:

- Trusted AI Principles: The trusted AI principles became the basis for building trusted AI around three pillars: employee engagement, product development, and empowering customers.
- Responsible AI Development Lifecycle: Our AI development lifecycle aims to provide good technical practices in order to avoid results contrary (often unexpected) to the objectives of our customers. Our responsible development lifecycle is performed by

415 Mission Street
San Francisco,
California 94105
Salesforce, Inc.
/salesforce
@salesforce

the Office of Ethical and Humane Use in partnership with other organizations such as product legal.
- Model Cards: They seek to standardize documentation procedures to communicate the performance characteristics of trained machine learning (ML) and AI models. They are a sort of nutrition label, designed to provide critical information about how our models work–including inputs, outputs, the conditions under which the model works best, and ethical considerations in their use. While not required, SFDC encourages its product teams to create model cards for their AI models. Published model cards can be found here.

We also work to ensure contractual accountability through Salesforce's Acceptable Use Policy (AUP), which articulates what our customers cannot do with our products. Our AUP prohibits customers from using our technology to take actions such as predicting sensitive categories of personal data (an individual's racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, age, gender, sex life, sexual orientation, criminal convictions, disability, health status, financial status, or medical condition) or when deploying a bot, communicating with any third party without clearly communicating that the individual is speaking with a bot. The AUP is a living document that is evolving to address emerging technology and the resulting use cases like GAI.

Salesforce has also engaged with government frameworks such as NIST's AI Risk Management Framework (RMF) and Playbook, which includes many of the measures we have taken. Salesforce supports the NIST AI RMF as it provides a flexible yet robust approach to this emerging technology. The rate of AI innovation means that providing entities with a framework to help their organization manage AI risks will help these same organizations to better manage unforeseen issues. The RMF came as a result of a highly collaborative effort and therefore is a good summation of the efforts industry and other groups found to be important to ensure robust AI. In the face of rapid innovation, focusing solely on outcomes of AI does not holistically address the problems raised by AI. Instead, companies should have documentation outlining items like the intended use of the system, risk controls, and risk mitigation tools and these documents should be accompanied by transparency tools like model cards. Salesforce has, and will continue to ensure that the NIST AI RMF is integrated to our AI development lifecycle.

- Which non-U.S. or U.S. (federal, state, or local) laws and regulations already requiring an AI audit, assessment, or other accountability mechanism are most useful and why? Which are least useful and why?

Salesforce supports discussions like those made at the Organization for the Economic Cooperation and Development (OECD) and European Union (EU) level to develop a risk-based approach to AI. In the absence of global consensus on AI, it is helpful to have governments driving the debate toward a regulatory approach that is in line with

415 Mission Street
San Francisco,
California 94105
Salesforce, Inc.
/salesforce
@salesforce

established best practices. In this current environment, regulations regarding AI seem inevitable. Therefore, a clear and consistent approach would be welcomed to bring surety to the AI ecosystem.

Although there are already various frameworks and guidelines (UK ICO guidelines and Singapore's Model AI Governance Framework), the EU has crafted a risk-based legislative approach that acknowledges the different roles and responsibilities across the AI ecosystem, which is in line with those guidelines first established by the OECD. While the EU AI Act is still being negotiated, we are supportive of this approach and efforts to create a durable approach to developing trusted AI.

Accountability Subjects

- The AI value or supply chain is complex, often involving open source and proprietary products and downstream applications that are quite different from what AI system developers may initially have contemplated. Moreover, training data for AI systems may be acquired from multiple sources, including from the customer using the technology. Problems in AI systems may arise downstream at the deployment or customization stage or upstream during model development and data training.

AI is best utilized when paired with human ability, effectively augmenting people and enabling them to make better decisions. However, when it comes to general-purpose AI that can be customized by the user, there needs to be a clear designation of responsibility between AI providers and AI users.

AI developers like Salesforce often create general customizable AI tools, whose intended purpose is low-risk, and it is the customer's responsibility (i.e., the AI deployer) to decide how these tools are employed. This means that it is the customer who ultimately controls when to use the products, which data is submitted to the AI and when, how the AI is configured, and, most critically, how the resulting predictions are used. It is the customer, and not Salesforce, that knows what has been disclosed to the affected individual, and what the risk of harm is to the affected individual.

While every member of the AI ecosystem has some role to play around accountability, companies that have a direct relationship with this end user should be the party that communicates all relevant transparency and accountability information. For our part, Salesforce has developed documents like model cards, our AUP, and principles to help our customers understand how the system was developed. Further, Salesforce and our customers have a contractual relationship that outlines the support that we will provide in order to help our customers better use our products and services.

Barriers to Effective Accountability - Lack of Comprehensive U.S. Privacy Law

415 Mission Street
San Francisco,
California 94105
Salesforce, Inc.
/salesforce
@salesforce

- Is the lack of a general federal data protection or privacy law a barrier to effective AI accountability?

Data protection laws that secure the fundamental human right to privacy are a foundation of responsible AI regulation. Although comprehensive AI regulation is still being developed, privacy laws such as Europe's General Data Protection Regulation (GDPR) and comprehensive US state laws that have passed in California, Virginia, Colorado, Utah, Iowa, Indiana, Connecticut, Florida, Washington and Montana are examples of government regulation that still have an impact on trusted AI. These laws hold companies accountable for providing individuals with the right to control and access their personal data, and for providing more transparency and explanation into how and why an individual's information is being used. The lack of an overarching Federal standard means that the data which powers AI systems could be collected in a way that prevents the development of trusted AI. Further, we believe that any comprehensive federal privacy legislation in the United States should include provisions prohibiting the use of personal data to discriminate on the basis of protected characteristics.

Salesforce welcomes the role of regulators in shaping responsible innovation. Presently, the world is looking to EU regulators and GDPR to write the rules of the road for AI. In fact, GDPR has been used by EU regulators to demand and achieve additional protections in OpenAI's generative AI services. The U.S. has introduced important ideas in ADPPA and laws passed at the state level that would advance AI regulation and should also be part of the global conversation through incorporation in a comprehensive privacy law.

AI Accountability Policies

- What role should government policy have, if any, in the AI accountability ecosystem? For example: a. Should AI accountability policies and/or regulation be sectoral or horizontal, or some combination of the two?

AI and now, GAI are becoming more present in our everyday life. Moreover, technology now underpins most of our lives, which means AI is used in settings across the different sectors as well as the government. Foremost, AI regulations should be risk-based, and differentiate contexts and uses of the technology, assigning responsibilities based on the different roles that various entities play in the AI ecosystem. AI regulations also should be harmonized and consistent with existing rules that may exist in different sectors. Therefore, AI rules should have a strong degree of horizontal consistency on core areas like definitions as well as roles and responsibilities, while recognizing that within a risk-based framework, some sectoral use cases will require different treatment based on the underlying activity (e.g. AI used to sell shoes vs AI used to operate power plants).

415 Mission Street
San Francisco,
California 94105
Salesforce, Inc.
/salesforce
@salesforce

- If a federal law focused on AI systems is desirable, what provisions would be particularly important to include? Which agency or agencies should be responsible for enforcing such a law, and what resources would they need to be successful?

At the foundational level, regulation should be context specific and aim to mitigate the potential impacts of high-risk AI applications that may negatively impact the human rights and freedoms of individuals. Regulation should also include a light-touch approach to regulating low-risk applications. A one-size-fits-all approach, which requires a blanket mechanism to provide AI or GAI systems would lack the nuanced approach that is necessary to maintain an AI ecosystem that includes actors and entities of different sizes.

Definitions of "High Risk" AI should be narrow and focused on applications that could do consequential harm. Another key aspect is provisions that differentiates contexts and uses of the technology (e.g. Business to Consumer v. Business to Business) and assigns responsibilities based on the different roles that various entities play in the AI ecosystem. Finally, the responsible development and use of AI goes hand-in-hand with privacy and sound data governance, including security, confidentiality, and preserving the integrity of data.

At the obligations levels, regulation should require automated systems (including conversational AI and chatbots) to disclose to users that they are interacting with an AI system. Products that simulate another person must either have that person's explicit consent or be clearly labeled as "simulated" or "parody." Further, regulation should enable humans to have sufficient information to truly know if the AI recommendation is fair/safe/accurate to empower them to be a meaningful check on the system.

Already, governments have passed and amended laws around privacy, safety, discrimination, and intellectual property and these laws have in some cases been refined by judicial action. As regulators look at building trusted AI, they should keep in mind that there are already laws and policies that provide some guardrails around the harmful aspects of AI systems. New rules should be considered alongside existing rules to ensure that there is harmonization and consistency in the approach.

- Is it important that there be uniformity of AI accountability requirements and/or practices across the United States? Across global jurisdictions? If so, is it important only within a sector or across sectors? What is the best way to achieve it? Alternatively, is harmonization or interoperability sufficient and what is the best way to achieve that?

Among industry experts there are few commonly agreed upon definitions for the foundational elements of AI. NIST has made a concerted effort to standardize many of these terms and definitions but fractures are already appearing in regulatory approaches.

415 Mission Street
San Francisco,
California 94105
Salesforce, Inc.
/salesforce
@salesforce

There should be an effort to prioritize work around global standardization for regulatory consistency, efficiency and interoperability. In some cases, we may want to require AI creators to provide multiple measures of a single construct like fairness or explainability. Another benefit of standardization would be clear and robust auditing. If definitions and methods were standardized, audits would be more consistent and lead to more confidence. This will also be necessary if third party certifications are included in future regulations. For these reasons, Salesforce is pleased to see further conversations around AI/GAI standards at the OECD and G7 meetings.

Conclusion

Salesforce is committed to building trusted, transparent, and accountable AI systems that prioritize fairness, accuracy, privacy, and positive societal impact. Salesforce welcomes that AI systems, including generative AI, are a priority focus for legislators and regulators. These are critical and rapidly evolving issues for society, and Salesforce is proactively engaging with governments and all stakeholder groups to advance responsible AI public policies. Governments and industry working together with other stakeholders to establish a common approach on definitions and obligations will create more durable, robust, and interoperable AI norms.

We are encouraged by the Administration's interest and look forward to further engagement with the NTIA. Salesforce remains committed to the success of our customers and we view our active participation in this important national discussion as advancing that success. We would be pleased to serve as a resource to the Administration as it further develops its approach on Trusted AI.

###

415 Mission Street
San Francisco,
California 94105
Salesforce, Inc.
/salesforce
@salesforce