

United States Senate

WASHINGTON, DC 20510

December 11, 2025

Eugenia Kuyda
Chief Executive Officer
Luka, Inc. (Replika)
600 Arkansas St
San Francisco, CA 94107

Dear Ms. Kuyda,

We write to request additional information about your company's public reporting practices regarding its artificial intelligence (AI) models. As a market leader for consumer AI chatbots, the steps that you take to increase transparency around your capabilities and risk evaluations have direct impacts on the wellbeing of Americans who use your products. We are already seeing how increasingly powerful models have become integrated into many aspects of personal life for users. As demonstrated harms to users' psychological well-being have emerged, it is critical that, in addition to improving safety design for users, your company takes necessary steps to foster public transparency to promote these goals.

In the past few years, reports have emerged about chatbots that have engaged in suicidal fantasies with children, drafted suicide notes, and provided specific instructions on self-harm.¹ These incidents have exposed how companies can fail to adequately evaluate models for possible use cases and inadequately disclose known risks associated with chatbot use. Additionally, we have seen how companies can struggle to prevent known model risks or unwanted behaviors prior to deployment.² If AI companies struggle to predict and mitigate relatively well-understood risks, it raises concerns about their ability to manage more complex risks. While we have been encouraged by the arrival of overdue safety measures for certain chatbots, these must be accompanied by improved public self-reporting so that consumers, families, educators, and policymakers can make informed decisions around appropriate use. More detailed information is

¹ Nitasha Tiku. "A Teen Contemplating Suicide Turned to a Chatbot. Is It Liable for Her Death?" The Washington Post. September 16, 2025. <https://www.washingtonpost.com/technology/2025/09/16/character-ai-suicide-lawsuit-new-juliana/?ref=dispatch.techoversight.org>;

Tiffany, Kaitlyn and Matteo Wong. "AI's Emerging Teen-Health Crisis" The Atlantic. September 18, 2025.

<https://www.theatlantic.com/technology/2025/09/openai-teen-safety/684268/>;

Andoh, Efua. "Many teens are turning to AI chatbots for friendship and emotional support." American Psychological Association. October 1, 2025. <https://www.apa.org/monitor/2025/10/technology-youth-friendships>.

² Sharma, Mrinank, Meg Tong, Tomasz Korbak, David Duvenaud, Amanda Askell, Samuel R. Bowman, Newton Cheng, et al. "Towards Understanding Sycophancy in Language Models." ArXiv. October 20, 2023.

<https://doi.org/10.48550/arXiv.2310.13548>;

Sample, Ian. "'Sycophantic' AI Chatbots Tell Users What They Want to Hear, Study Shows." The Guardian.

October 24, 2025. <https://www.theguardian.com/technology/2025/oct/24/sycophantic-ai-chatbots-tell-users-what-they-want-to-hear-study-shows>.

still needed for age-restricted models; even chatbots aimed at children *with* additional safeguards have produced pro-eating disorder, violent, and sexual content.³

In addition to impacts on mental health and risks to vulnerable users, accelerating model capabilities necessitate greater transparency around other potential risks involving public safety and national security. In particular, companies have disclosed how advanced models may pose misuse risks in areas including cybersecurity and biosecurity.⁴ Many frontier AI companies made voluntary commitments at the Seoul AI Summit, or in support of the G7 Code of Conduct, to provide transparency into their efforts to assess risks to national security.⁵ We are supportive of ongoing disclosures for these risks, and request that companies adhere to their prior commitments.

Public disclosure reports, such as AI model and system cards, serve as the closest equivalent to nutrition labels for AI models. While they are essential public transparency tools, today's changed landscape calls for assessing current best practices and how they can be better responsive to user risks. Current public disclosure practices can be inconsistent or insufficient, may not be released alongside product launches, and can lack standardization.⁶ The distinction between major and minor releases is left to the discretion of developers, sometimes without explanation. Model and system cards may also fail to incorporate new or updated information about existing models while models are deployed to the public, including information about user safeguards.⁷ Companies must continue to monitor their model performance and publicly disclose new developments as they relate to security and user safety. This information enables third-party evaluators to assess a model's risks and supports organizations, governments, and consumers in making more informed decisions.

It is critical that public disclosures and risk evaluations are comprehensive, consistent, timely, and responsive to emerging risks. We therefore request responses to the following questions by January 8, 2026:

1. Do you commit to publishing model and system cards in a standardized location concurrently or prior to future model releases?

³ Baker-White, Emily. "AI Tutors for Kids Gave Fentanyl Recipes and Dangerous Diet Advice." *Forbes*. May 12, 2025. <https://www.forbes.com/sites/emilybaker-white/2025/05/12/these-ai-tutors-for-kids-gave-fentanyl-recipes-and-dangerous-diet-advice/>.

⁴ Pannu, Jaspreet et al. "Dual-use capabilities of concern of biological AI models." *PLOS Computational Biological*. May 8, 2025. <https://pmc.ncbi.nlm.nih.gov/articles/PMC12061118/>.

⁵ U.K. Department for Science, Innovation & Technology. "Frontier AI Safety Commitments, AI Seoul Summit 2024." February 7, 2025. <https://www.gov.uk/government/publications/frontier-ai-safety-commitments-ai-seoul-summit-2024/frontier-ai-safety-commitments-ai-seoul-summit-2024>;

⁶ Perrigo, Billy. "Top AI Companies Have 'Unacceptable' Risk Management, Studies Say." *TIME*. July 17, 2025. <https://time.com/7302757/anthropic-xai-meta-openai-risk-management-2/>.

The Midas Project. "Seoul Commitment Tracker." Seoul Commitment Tracker. Accessed December 2025. <https://www.seoul-tracker.org/>.


⁷ Nolan, Beatrice. "AI chatbots are harming young people. Regulators are scrambling to keep up." *Fortune*. September 14, 2025. <https://fortune.com/2025/09/14/ai-chatbots-teens-children-mental-health-suicide-openai-chatgpt-regulation-lawsuit/>.

2. Do you commit to maintaining a consistent schema for evaluating potential harms in your public disclosures and risk frameworks, and to providing transparent justifications for changes made to them, or for when a released model evaluation deviates from the established schema?
3. Do you commit to clearly defining the criteria for determining which model releases require updated public disclosures and new risk evaluations, and which descriptors of models exempt them from such evaluations, such as 'experimental,' 'preview,' 'minor,' or other similar terms? Please elaborate on the rationale for why such models may be considered sufficiently low-risk to not warrant full reassessment.
4. Do you commit to flagging significant patterns of violations of risk frameworks for your models and providing explanations for their continued deployment despite these issues?
5. Do you commit to collaborating with external partners—including CAISI, academic researchers, and experts—to establish appropriate timelines for conducting sufficiently robust pre-deployment evaluations?
6. Do you commit to researching the short-term and long-term emotional and psychological impacts to your chatbots' users across various populations—including vulnerable groups such as children and the elderly—and including their results in your public disclosures? If so, please elaborate on how this research will be conducted in consultation with relevant third-party researchers.
7. Do you commit to disclosing how safeguards for vulnerable groups – such as children and senior citizens, and those experiencing mental or emotional distress – including but not limited to age-estimation technologies and mental health referrals? If so, please elaborate on how they are informed by best practices to ensure user safety and wellbeing, data privacy, and efficacy.
8. Do you commit to disclosing whether information from your company's chatbot conversations is used for targeted advertising? If your company already discloses this information, how do you disclose that fact to users? If conversation data is used for targeted advertising or otherwise shared with third parties, will you commit to conducting ongoing analysis of potential data privacy, cybersecurity, and other risks from such access?
9. Do you commit to publicly disclosing new information about other potential risks as they are made known to you in a timely manner, including categories of harm not previously recognized or accounted for within your existing evaluation frameworks?
10. Do you commit to ensuring that public disclosure information that is relevant for consumers is shared in a manner accessible and understandable to consumers? If so, please disclose where you will host this information.

11. Do you commit to disclosing whether your company withdraws from, or modifies its full participation in, voluntary agreements or frameworks?

Thank you for your attention to these matters.

Sincerely,

A handwritten signature in blue ink, appearing to read "Brian Schatz", written over a horizontal line.

Brian Schatz
United States Senator

A handwritten signature in blue ink, appearing to read "Katie Boyd Britt", written over a horizontal line.

Katie Boyd Britt
United States Senator

A handwritten signature in blue ink, appearing to read "James Lankford", written over a horizontal line.

James Lankford
United States Senator

A handwritten signature in blue ink, appearing to read "Chris Coons", written over a horizontal line.

Christopher A. Coons
United States Senator

United States Senate

WASHINGTON, DC 20510

December 11, 2025

Karandeep Anand
Chief Executive Officer
Character Technologies
1152 700 El Camino Real Suite 120
Menlo Park, CA 94025

Dear Mr. Anand,

We write to request additional information about your company's public reporting practices regarding its artificial intelligence (AI) models. As a market leader for consumer AI chatbots, the steps that you take to increase transparency around your capabilities and risk evaluations have direct impacts on the wellbeing of Americans who use your products. We are already seeing how increasingly powerful models have become integrated into many aspects of personal life for users. As demonstrated harms to users' psychological well-being have emerged, it is critical that, in addition to improving safety design for users, your company takes necessary steps to foster public transparency to promote these goals.

In the past few years, reports have emerged about chatbots that have engaged in suicidal fantasies with children, drafted suicide notes, and provided specific instructions on self-harm.¹ These incidents have exposed how companies can fail to adequately evaluate models for possible use cases and inadequately disclose known risks associated with chatbot use. Additionally, we have seen how companies can struggle to prevent known model risks or unwanted behaviors prior to deployment.² If AI companies struggle to predict and mitigate relatively well-understood risks, it raises concerns about their ability to manage more complex risks. While we have been encouraged by the arrival of overdue safety measures for certain chatbots, these must be accompanied by improved public self-reporting so that consumers, families, educators, and policymakers can make informed decisions around appropriate use. More detailed information is

¹ Nitasha Tikku. "A Teen Contemplating Suicide Turned to a Chatbot. Is It Liable for Her Death?" The Washington Post. September 16, 2025. <https://www.washingtonpost.com/technology/2025/09/16/character-ai-suicide-lawsuit-new-juliana/?ref=dispatch.techoversight.org>;

Tiffany, Kaitlyn and Matteo Wong. "AI's Emerging Teen-Health Crisis" The Atlantic. September 18, 2025.

<https://www.theatlantic.com/technology/2025/09/openai-teen-safety/684268/>;

Andoh, Efua. "Many teens are turning to AI chatbots for friendship and emotional support." American Psychological Association. October 1, 2025. <https://www.apa.org/monitor/2025/10/technology-youth-friendships>.

² Sharma, Mrinank, Meg Tong, Tomasz Korbak, David Duvenaud, Amanda Askill, Samuel R. Bowman, Newton Cheng, et al. "Towards Understanding Sycophancy in Language Models." ArXiv. October 20, 2023.

<https://doi.org/10.48550/arXiv.2310.13548>;

Sample, Ian. "'Sycophantic' AI Chatbots Tell Users What They Want to Hear, Study Shows." The Guardian.

October 24, 2025. <https://www.theguardian.com/technology/2025/oct/24/sycophantic-ai-chatbots-tell-users-what-they-want-to-hear-study-shows>.

still needed for age-restricted models; even chatbots aimed at children *with* additional safeguards have produced pro-eating disorder, violent, and sexual content.³

In addition to impacts on mental health and risks to vulnerable users, accelerating model capabilities necessitate greater transparency around other potential risks involving public safety and national security. In particular, companies have disclosed how advanced models may pose misuse risks in areas including cybersecurity and biosecurity.⁴ Many frontier AI companies made voluntary commitments at the Seoul AI Summit, or in support of the G7 Code of Conduct, to provide transparency into their efforts to assess risks to national security.⁵ We are supportive of ongoing disclosures for these risks, and request that companies adhere to their prior commitments.

Public disclosure reports, such as AI model and system cards, serve as the closest equivalent to nutrition labels for AI models. While they are essential public transparency tools, today's changed landscape calls for assessing current best practices and how they can be better responsive to user risks. Current public disclosure practices can be inconsistent or insufficient, may not be released alongside product launches, and can lack standardization.⁶ The distinction between major and minor releases is left to the discretion of developers, sometimes without explanation. Model and system cards may also fail to incorporate new or updated information about existing models while models are deployed to the public, including information about user safeguards.⁷ Companies must continue to monitor their model performance and publicly disclose new developments as they relate to security and user safety. This information enables third-party evaluators to assess a model's risks and supports organizations, governments, and consumers in making more informed decisions.

It is critical that public disclosures and risk evaluations are comprehensive, consistent, timely, and responsive to emerging risks. We therefore request responses to the following questions by January 8, 2026:

1. Do you commit to publishing model and system cards in a standardized location concurrently or prior to future model releases?

³ Baker-White, Emily. "AI Tutors for Kids Gave Fentanyl Recipes and Dangerous Diet Advice." *Forbes*. May 12, 2025. <https://www.forbes.com/sites/emilybaker-white/2025/05/12/these-ai-tutors-for-kids-gave-fentanyl-recipes-and-dangerous-diet-advice/>.

⁴ Pannu, Jaspreet et al. "Dual-use capabilities of concern of biological AI models." *PLOS Computational Biological*. May 8, 2025. <https://pmc.ncbi.nlm.nih.gov/articles/PMC12061118/>.

⁵ U.K. Department for Science, Innovation & Technology. "Frontier AI Safety Commitments, AI Seoul Summit 2024." February 7, 2025. <https://www.gov.uk/government/publications/frontier-ai-safety-commitments-ai-seoul-summit-2024/frontier-ai-safety-commitments-ai-seoul-summit-2024>;

⁶ Perrigo, Billy. "Top AI Companies Have 'Unacceptable' Risk Management, Studies Say." *TIME*. July 17, 2025. <https://time.com/7302757/anthropic-xai-meta-openai-risk-management-2/>.

The Midas Project. "Seoul Commitment Tracker." Seoul Commitment Tracker. Accessed December 2025. <https://www.seoul-tracker.org/>.


⁷ Nolan, Beatrice. "AI chatbots are harming young people. Regulators are scrambling to keep up." *Fortune*. September 14, 2025. <https://fortune.com/2025/09/14/ai-chatbots-teens-children-mental-health-suicide-openai-chatgpt-regulation-lawsuit/>.

2. Do you commit to maintaining a consistent schema for evaluating potential harms in your public disclosures and risk frameworks, and to providing transparent justifications for changes made to them, or for when a released model evaluation deviates from the established schema?
3. Do you commit to clearly defining the criteria for determining which model releases require updated public disclosures and new risk evaluations, and which descriptors of models exempt them from such evaluations, such as 'experimental,' 'preview,' 'minor,' or other similar terms? Please elaborate on the rationale for why such models may be considered sufficiently low-risk to not warrant full reassessment.
4. Do you commit to flagging significant patterns of violations of risk frameworks for your models and providing explanations for their continued deployment despite these issues?
5. Do you commit to collaborating with external partners—including CAISI, academic researchers, and experts—to establish appropriate timelines for conducting sufficiently robust pre-deployment evaluations?
6. Do you commit to researching the short-term and long-term emotional and psychological impacts to your chatbots' users across various populations—including vulnerable groups such as children and the elderly—and including their results in your public disclosures? If so, please elaborate on how this research will be conducted in consultation with relevant third-party researchers.
7. Do you commit to disclosing how safeguards for vulnerable groups – such as children and senior citizens, and those experiencing mental or emotional distress – including but not limited to age-estimation technologies and mental health referrals? If so, please elaborate on how they are informed by best practices to ensure user safety and wellbeing, data privacy, and efficacy.
8. Do you commit to disclosing whether information from your company's chatbot conversations is used for targeted advertising? If your company already discloses this information, how do you disclose that fact to users? If conversation data is used for targeted advertising or otherwise shared with third parties, will you commit to conducting ongoing analysis of potential data privacy, cybersecurity, and other risks from such access?
9. Do you commit to publicly disclosing new information about other potential risks as they are made known to you in a timely manner, including categories of harm not previously recognized or accounted for within your existing evaluation frameworks?
10. Do you commit to ensuring that public disclosure information that is relevant for consumers is shared in a manner accessible and understandable to consumers? If so, please disclose where you will host this information.

11. Do you commit to disclosing whether your company withdraws from, or modifies its full participation in, voluntary agreements or frameworks?

Thank you for your attention to these matters.

Sincerely,

A handwritten signature in blue ink, appearing to read "Brian Schatz", written over a horizontal line.

Brian Schatz
United States Senator

A handwritten signature in blue ink, appearing to read "Katie Boyd Britt", written over a horizontal line.

Katie Boyd Britt
United States Senator

A handwritten signature in blue ink, appearing to read "James Lankford", written over a horizontal line.

James Lankford
United States Senator

A handwritten signature in blue ink, appearing to read "Chris Coons", written over a horizontal line.

Christopher A. Coons
United States Senator

United States Senate

WASHINGTON, DC 20510

December 11, 2025

Elon Musk
Chief Executive Officer
xAI Corp.
216 Park Road
Burlingame, CA 94010

Dear Mr. Musk,

We write to request additional information about your company's public reporting practices regarding its artificial intelligence (AI) models. As a market leader for consumer AI chatbots, the steps that you take to increase transparency around your capabilities and risk evaluations have direct impacts on the wellbeing of Americans who use your products. We are already seeing how increasingly powerful models have become integrated into many aspects of personal life for users. As demonstrated harms to users' psychological well-being have emerged, it is critical that, in addition to improving safety design for users, your company takes necessary steps to foster public transparency to promote these goals.

In the past few years, reports have emerged about chatbots that have engaged in suicidal fantasies with children, drafted suicide notes, and provided specific instructions on self-harm.¹ These incidents have exposed how companies can fail to adequately evaluate models for possible use cases and inadequately disclose known risks associated with chatbot use. Additionally, we have seen how companies can struggle to prevent known model risks or unwanted behaviors prior to deployment.² If AI companies struggle to predict and mitigate relatively well-understood risks, it raises concerns about their ability to manage more complex risks. While we have been encouraged by the arrival of overdue safety measures for certain chatbots, these must be accompanied by improved public self-reporting so that consumers, families, educators, and policymakers can make informed decisions around appropriate use. More detailed information is

¹ Nitasha Tikun. "A Teen Contemplating Suicide Turned to a Chatbot. Is It Liable for Her Death?" The Washington Post. September 16, 2025. <https://www.washingtonpost.com/technology/2025/09/16/character-ai-suicide-lawsuit-new-juliana/?ref=dispatch.techoversight.org>;

Tiffany, Kaitlyn and Matteo Wong. "AI's Emerging Teen-Health Crisis" The Atlantic. September 18, 2025.

<https://www.theatlantic.com/technology/2025/09/openai-teen-safety/684268/>;

Andoh, Efua. "Many teens are turning to AI chatbots for friendship and emotional support." American Psychological Association. October 1, 2025. <https://www.apa.org/monitor/2025/10/technology-youth-friendships>.

² Sharma, Mrinank, Meg Tong, Tomasz Korbak, David Duvenaud, Amanda Askeel, Samuel R. Bowman, Newton Cheng, et al. "Towards Understanding Sycophancy in Language Models." ArXiv. October 20, 2023.

<https://doi.org/10.48550/arXiv.2310.13548>;

Sample, Ian. "'Sycophantic' AI Chatbots Tell Users What They Want to Hear, Study Shows." The Guardian.

October 24, 2025. <https://www.theguardian.com/technology/2025/oct/24/sycophantic-ai-chatbots-tell-users-what-they-want-to-hear-study-shows>.

still needed for age-restricted models; even chatbots aimed at children *with* additional safeguards have produced pro-eating disorder, violent, and sexual content.³

In addition to impacts on mental health and risks to vulnerable users, accelerating model capabilities necessitate greater transparency around other potential risks involving public safety and national security. In particular, companies have disclosed how advanced models may pose misuse risks in areas including cybersecurity and biosecurity.⁴ Many frontier AI companies made voluntary commitments at the Seoul AI Summit, or in support of the G7 Code of Conduct, to provide transparency into their efforts to assess risks to national security.⁵ We are supportive of ongoing disclosures for these risks, and request that companies adhere to their prior commitments.

Public disclosure reports, such as AI model and system cards, serve as the closest equivalent to nutrition labels for AI models. While they are essential public transparency tools, today's changed landscape calls for assessing current best practices and how they can be better responsive to user risks. Current public disclosure practices can be inconsistent or insufficient, may not be released alongside product launches, and can lack standardization.⁶ The distinction between major and minor releases is left to the discretion of developers, sometimes without explanation. Model and system cards may also fail to incorporate new or updated information about existing models while models are deployed to the public, including information about user safeguards.⁷ Companies must continue to monitor their model performance and publicly disclose new developments as they relate to security and user safety. This information enables third-party evaluators to assess a model's risks and supports organizations, governments, and consumers in making more informed decisions.

It is critical that public disclosures and risk evaluations are comprehensive, consistent, timely, and responsive to emerging risks. We therefore request responses to the following questions by January 8, 2026:

1. Do you commit to publishing model and system cards in a standardized location concurrently or prior to future model releases?

³ Baker-White, Emily. "AI Tutors for Kids Gave Fentanyl Recipes and Dangerous Diet Advice." *Forbes*. May 12, 2025. <https://www.forbes.com/sites/emilybaker-white/2025/05/12/these-ai-tutors-for-kids-gave-fentanyl-recipes-and-dangerous-diet-advice/>.

⁴ Pannu, Jaspreet et al. "Dual-use capabilities of concern of biological AI models." *PLOS Computational Biological*. May 8, 2025. <https://pmc.ncbi.nlm.nih.gov/articles/PMC12061118/>.

⁵ U.K. Department for Science, Innovation & Technology. "Frontier AI Safety Commitments, AI Seoul Summit 2024." February 7, 2025. <https://www.gov.uk/government/publications/frontier-ai-safety-commitments-ai-seoul-summit-2024/frontier-ai-safety-commitments-ai-seoul-summit-2024>;

⁶ Perrigo, Billy. "Top AI Companies Have 'Unacceptable' Risk Management, Studies Say." *TIME*. July 17, 2025. <https://time.com/7302757/anthropic-xai-meta-openai-risk-management-2/>.

The Midas Project. "Seoul Commitment Tracker." Seoul Commitment Tracker. Accessed December 2025. <https://www.seoul-tracker.org/>.


⁷ Nolan, Beatrice. "AI chatbots are harming young people. Regulators are scrambling to keep up." *Fortune*. September 14, 2025. <https://fortune.com/2025/09/14/ai-chatbots-teens-children-mental-health-suicide-openai-chatgpt-regulation-lawsuit/>.

2. Do you commit to maintaining a consistent schema for evaluating potential harms in your public disclosures and risk frameworks, and to providing transparent justifications for changes made to them, or for when a released model evaluation deviates from the established schema?
3. Do you commit to clearly defining the criteria for determining which model releases require updated public disclosures and new risk evaluations, and which descriptors of models exempt them from such evaluations, such as 'experimental,' 'preview,' 'minor,' or other similar terms? Please elaborate on the rationale for why such models may be considered sufficiently low-risk to not warrant full reassessment.
4. Do you commit to flagging significant patterns of violations of risk frameworks for your models and providing explanations for their continued deployment despite these issues?
5. Do you commit to collaborating with external partners—including CAISI, academic researchers, and experts—to establish appropriate timelines for conducting sufficiently robust pre-deployment evaluations?
6. Do you commit to researching the short-term and long-term emotional and psychological impacts to your chatbots' users across various populations—including vulnerable groups such as children and the elderly—and including their results in your public disclosures? If so, please elaborate on how this research will be conducted in consultation with relevant third-party researchers.
7. Do you commit to disclosing how safeguards for vulnerable groups – such as children and senior citizens, and those experiencing mental or emotional distress – including but not limited to age-estimation technologies and mental health referrals? If so, please elaborate on how they are informed by best practices to ensure user safety and wellbeing, data privacy, and efficacy.
8. Do you commit to disclosing whether information from your company's chatbot conversations is used for targeted advertising? If your company already discloses this information, how do you disclose that fact to users? If conversation data is used for targeted advertising or otherwise shared with third parties, will you commit to conducting ongoing analysis of potential data privacy, cybersecurity, and other risks from such access?
9. Do you commit to publicly disclosing new information about other potential risks as they are made known to you in a timely manner, including categories of harm not previously recognized or accounted for within your existing evaluation frameworks?
10. Do you commit to ensuring that public disclosure information that is relevant for consumers is shared in a manner accessible and understandable to consumers? If so, please disclose where you will host this information.

11. Do you commit to disclosing whether your company withdraws from, or modifies its full participation in, voluntary agreements or frameworks?

Thank you for your attention to these matters.

Sincerely,

A handwritten signature in blue ink, appearing to read "Brian Schatz", written over a horizontal line.

Brian Schatz
United States Senator

A handwritten signature in blue ink, appearing to read "Katie Boyd Britt", written over a horizontal line.

Katie Boyd Britt
United States Senator

A handwritten signature in blue ink, appearing to read "James Lankford", written over a horizontal line.

James Lankford
United States Senator

A handwritten signature in blue ink, appearing to read "Chris Coons", written over a horizontal line.

Christopher A. Coons
United States Senator

United States Senate

WASHINGTON, DC 20510

December 11, 2025

Satya Nadella
Chief Executive Officer
Microsoft Headquarters
1 Microsoft Way
Redmond, WA 98052

Dear Mr. Nadella,

We write to request additional information about your company's public reporting practices regarding its artificial intelligence (AI) models. As a market leader for consumer AI chatbots, the steps that you take to increase transparency around your capabilities and risk evaluations have direct impacts on the wellbeing of Americans who use your products. We are already seeing how increasingly powerful models have become integrated into many aspects of personal life for users. As demonstrated harms to users' psychological well-being have emerged, it is critical that, in addition to improving safety design for users, your company takes necessary steps to foster public transparency to promote these goals.

In the past few years, reports have emerged about chatbots that have engaged in suicidal fantasies with children, drafted suicide notes, and provided specific instructions on self-harm.¹ These incidents have exposed how companies can fail to adequately evaluate models for possible use cases and inadequately disclose known risks associated with chatbot use. Additionally, we have seen how companies can struggle to prevent known model risks or unwanted behaviors prior to deployment.² If AI companies struggle to predict and mitigate relatively well-understood risks, it raises concerns about their ability to manage more complex risks. While we have been encouraged by the arrival of overdue safety measures for certain chatbots, these must be accompanied by improved public self-reporting so that consumers, families, educators, and policymakers can make informed decisions around appropriate use. More detailed information is

¹ Nitasha Tikku. "A Teen Contemplating Suicide Turned to a Chatbot. Is It Liable for Her Death?" The Washington Post. September 16, 2025. <https://www.washingtonpost.com/technology/2025/09/16/character-ai-suicide-lawsuit-new-juliana/?ref=dispatch.techoversight.org>;

Tiffany, Kaitlyn and Matteo Wong. "AI's Emerging Teen-Health Crisis" The Atlantic. September 18, 2025. <https://www.theatlantic.com/technology/2025/09/openai-teen-safety/684268/>;

Andoh, Efua. "Many teens are turning to AI chatbots for friendship and emotional support." American Psychological Association. October 1, 2025. <https://www.apa.org/monitor/2025/10/technology-youth-friendships>.

² Sharma, Mrinank, Meg Tong, Tomasz Korbak, David Duvenaud, Amanda Askill, Samuel R. Bowman, Newton Cheng, et al. "Towards Understanding Sycophancy in Language Models." ArXiv. October 20, 2023. <https://doi.org/10.48550/arXiv.2310.13548>;

Sample, Ian. "'Sycophantic' AI Chatbots Tell Users What They Want to Hear, Study Shows." The Guardian. October 24, 2025. <https://www.theguardian.com/technology/2025/oct/24/sycophantic-ai-chatbots-tell-users-what-they-want-to-hear-study-shows>.

still needed for age-restricted models; even chatbots aimed at children *with* additional safeguards have produced pro-eating disorder, violent, and sexual content.³

In addition to impacts on mental health and risks to vulnerable users, accelerating model capabilities necessitate greater transparency around other potential risks involving public safety and national security. In particular, companies have disclosed how advanced models may pose misuse risks in areas including cybersecurity and biosecurity.⁴ Many frontier AI companies made voluntary commitments at the Seoul AI Summit, or in support of the G7 Code of Conduct, to provide transparency into their efforts to assess risks to national security.⁵ We are supportive of ongoing disclosures for these risks, and request that companies adhere to their prior commitments.

Public disclosure reports, such as AI model and system cards, serve as the closest equivalent to nutrition labels for AI models. While they are essential public transparency tools, today's changed landscape calls for assessing current best practices and how they can be better responsive to user risks. Current public disclosure practices can be inconsistent or insufficient, may not be released alongside product launches, and can lack standardization.⁶ The distinction between major and minor releases is left to the discretion of developers, sometimes without explanation. Model and system cards may also fail to incorporate new or updated information about existing models while models are deployed to the public, including information about user safeguards.⁷ Companies must continue to monitor their model performance and publicly disclose new developments as they relate to security and user safety. This information enables third-party evaluators to assess a model's risks and supports organizations, governments, and consumers in making more informed decisions.

It is critical that public disclosures and risk evaluations are comprehensive, consistent, timely, and responsive to emerging risks. We therefore request responses to the following questions by January 8, 2026:

1. Do you commit to publishing model and system cards in a standardized location concurrently or prior to future model releases?

³ Baker-White, Emily. "AI Tutors for Kids Gave Fentanyl Recipes and Dangerous Diet Advice." *Forbes*. May 12, 2025. <https://www.forbes.com/sites/emilybaker-white/2025/05/12/these-ai-tutors-for-kids-gave-fentanyl-recipes-and-dangerous-diet-advice/>.

⁴ Pannu, Jaspreet et al. "Dual-use capabilities of concern of biological AI models." *PLOS Computational Biological*. May 8, 2025. <https://pmc.ncbi.nlm.nih.gov/articles/PMC12061118/>.

⁵ U.K. Department for Science, Innovation & Technology. "Frontier AI Safety Commitments, AI Seoul Summit 2024." February 7, 2025. <https://www.gov.uk/government/publications/frontier-ai-safety-commitments-ai-seoul-summit-2024/frontier-ai-safety-commitments-ai-seoul-summit-2024>;

⁶ Perrigo, Billy. "Top AI Companies Have 'Unacceptable' Risk Management, Studies Say." *TIME*. July 17, 2025. <https://time.com/7302757/anthropic-xai-meta-openai-risk-management-2/>.

The Midas Project. "Seoul Commitment Tracker." Seoul Commitment Tracker. Accessed December 2025. <https://www.seoul-tracker.org/>.


⁷ Nolan, Beatrice. "AI chatbots are harming young people. Regulators are scrambling to keep up." *Fortune*. September 14, 2025. <https://fortune.com/2025/09/14/ai-chatbots-teens-children-mental-health-suicide-openai-chatgpt-regulation-lawsuit/>.

2. Do you commit to maintaining a consistent schema for evaluating potential harms in your public disclosures and risk frameworks, and to providing transparent justifications for changes made to them, or for when a released model evaluation deviates from the established schema?
3. Do you commit to clearly defining the criteria for determining which model releases require updated public disclosures and new risk evaluations, and which descriptors of models exempt them from such evaluations, such as 'experimental,' 'preview,' 'minor,' or other similar terms? Please elaborate on the rationale for why such models may be considered sufficiently low-risk to not warrant full reassessment.
4. Do you commit to flagging significant patterns of violations of risk frameworks for your models and providing explanations for their continued deployment despite these issues?
5. Do you commit to collaborating with external partners—including CAISI, academic researchers, and experts—to establish appropriate timelines for conducting sufficiently robust pre-deployment evaluations?
6. Do you commit to researching the short-term and long-term emotional and psychological impacts to your chatbots' users across various populations—including vulnerable groups such as children and the elderly—and including their results in your public disclosures? If so, please elaborate on how this research will be conducted in consultation with relevant third-party researchers.
7. Do you commit to disclosing how safeguards for vulnerable groups – such as children and senior citizens, and those experiencing mental or emotional distress – including but not limited to age-estimation technologies and mental health referrals? If so, please elaborate on how they are informed by best practices to ensure user safety and wellbeing, data privacy, and efficacy.
8. Do you commit to disclosing whether information from your company's chatbot conversations is used for targeted advertising? If your company already discloses this information, how do you disclose that fact to users? If conversation data is used for targeted advertising or otherwise shared with third parties, will you commit to conducting ongoing analysis of potential data privacy, cybersecurity, and other risks from such access?
9. Do you commit to publicly disclosing new information about other potential risks as they are made known to you in a timely manner, including categories of harm not previously recognized or accounted for within your existing evaluation frameworks?
10. Do you commit to ensuring that public disclosure information that is relevant for consumers is shared in a manner accessible and understandable to consumers? If so, please disclose where you will host this information.

11. Do you commit to disclosing whether your company withdraws from, or modifies its full participation in, voluntary agreements or frameworks?

Thank you for your attention to these matters.

Sincerely,

A handwritten signature in blue ink, appearing to read "Brian Schatz", written over a horizontal line.

Brian Schatz
United States Senator

A handwritten signature in blue ink, appearing to read "Katie Boyd Britt", written over a horizontal line.

Katie Boyd Britt
United States Senator

A handwritten signature in blue ink, appearing to read "James Lankford", written over a horizontal line.

James Lankford
United States Senator

A handwritten signature in blue ink, appearing to read "Chris Coons", written over a horizontal line.

Christopher A. Coons
United States Senator

United States Senate

WASHINGTON, DC 20510

December 11, 2025

Dario Amodei
Chief Executive Officer
Anthropic PBC
548 Market Street, PMB 90375
San Francisco, CA 94104

Dear Mr. Amodei,

We write to request additional information about your company's public reporting practices regarding its artificial intelligence (AI) models. As a market leader for consumer AI chatbots, the steps that you take to increase transparency around your capabilities and risk evaluations have direct impacts on the wellbeing of Americans who use your products. We are already seeing how increasingly powerful models have become integrated into many aspects of personal life for users. As demonstrated harms to users' psychological well-being have emerged, it is critical that, in addition to improving safety design for users, your company takes necessary steps to foster public transparency to promote these goals.

In the past few years, reports have emerged about chatbots that have engaged in suicidal fantasies with children, drafted suicide notes, and provided specific instructions on self-harm.¹ These incidents have exposed how companies can fail to adequately evaluate models for possible use cases and inadequately disclose known risks associated with chatbot use. Additionally, we have seen how companies can struggle to prevent known model risks or unwanted behaviors prior to deployment.² If AI companies struggle to predict and mitigate relatively well-understood risks, it raises concerns about their ability to manage more complex risks. While we have been encouraged by the arrival of overdue safety measures for certain chatbots, these must be accompanied by improved public self-reporting so that consumers, families, educators, and policymakers can make informed decisions around appropriate use. More detailed information is

¹ Nitasha Tikun. "A Teen Contemplating Suicide Turned to a Chatbot. Is It Liable for Her Death?" The Washington Post. September 16, 2025. <https://www.washingtonpost.com/technology/2025/09/16/character-ai-suicide-lawsuit-new-juliana/?ref=dispatch.techoversight.org>;

Tiffany, Kaitlyn and Matteo Wong. "AI's Emerging Teen-Health Crisis" The Atlantic. September 18, 2025.

<https://www.theatlantic.com/technology/2025/09/openai-teen-safety/684268/>;

Andoh, Efua. "Many teens are turning to AI chatbots for friendship and emotional support." American Psychological Association. October 1, 2025. <https://www.apa.org/monitor/2025/10/technology-youth-friendships>.

² Sharma, Mrinank, Meg Tong, Tomasz Korbak, David Duvenaud, Amanda Askeel, Samuel R. Bowman, Newton Cheng, et al. "Towards Understanding Sycophancy in Language Models." ArXiv. October 20, 2023.

<https://doi.org/10.48550/arXiv.2310.13548>;

Sample, Ian. "'Sycophantic' AI Chatbots Tell Users What They Want to Hear, Study Shows." The Guardian.

October 24, 2025. <https://www.theguardian.com/technology/2025/oct/24/sycophantic-ai-chatbots-tell-users-what-they-want-to-hear-study-shows>.

still needed for age-restricted models; even chatbots aimed at children *with* additional safeguards have produced pro-eating disorder, violent, and sexual content.³

In addition to impacts on mental health and risks to vulnerable users, accelerating model capabilities necessitate greater transparency around other potential risks involving public safety and national security. In particular, companies have disclosed how advanced models may pose misuse risks in areas including cybersecurity and biosecurity.⁴ Many frontier AI companies made voluntary commitments at the Seoul AI Summit, or in support of the G7 Code of Conduct, to provide transparency into their efforts to assess risks to national security.⁵ We are supportive of ongoing disclosures for these risks, and request that companies adhere to their prior commitments.

Public disclosure reports, such as AI model and system cards, serve as the closest equivalent to nutrition labels for AI models. While they are essential public transparency tools, today's changed landscape calls for assessing current best practices and how they can be better responsive to user risks. Current public disclosure practices can be inconsistent or insufficient, may not be released alongside product launches, and can lack standardization.⁶ The distinction between major and minor releases is left to the discretion of developers, sometimes without explanation. Model and system cards may also fail to incorporate new or updated information about existing models while models are deployed to the public, including information about user safeguards.⁷ Companies must continue to monitor their model performance and publicly disclose new developments as they relate to security and user safety. This information enables third-party evaluators to assess a model's risks and supports organizations, governments, and consumers in making more informed decisions.

It is critical that public disclosures and risk evaluations are comprehensive, consistent, timely, and responsive to emerging risks. We therefore request responses to the following questions by January 8, 2026:

1. Do you commit to publishing model and system cards in a standardized location concurrently or prior to future model releases?

³ Baker-White, Emily. "AI Tutors for Kids Gave Fentanyl Recipes and Dangerous Diet Advice." *Forbes*. May 12, 2025. <https://www.forbes.com/sites/emilybaker-white/2025/05/12/these-ai-tutors-for-kids-gave-fentanyl-recipes-and-dangerous-diet-advice/>.

⁴ Pannu, Jaspreet et al. "Dual-use capabilities of concern of biological AI models." *PLOS Computational Biological*. May 8, 2025. <https://pmc.ncbi.nlm.nih.gov/articles/PMC12061118/>.

⁵ U.K. Department for Science, Innovation & Technology. "Frontier AI Safety Commitments, AI Seoul Summit 2024." February 7, 2025. <https://www.gov.uk/government/publications/frontier-ai-safety-commitments-ai-seoul-summit-2024/frontier-ai-safety-commitments-ai-seoul-summit-2024>;

⁶ Perrigo, Billy. "Top AI Companies Have 'Unacceptable' Risk Management, Studies Say." *TIME*. July 17, 2025. <https://time.com/7302757/anthropic-xai-meta-openai-risk-management-2/>.

The Midas Project. "Seoul Commitment Tracker." Seoul Commitment Tracker. Accessed December 2025. <https://www.seoul-tracker.org/>.


⁷ Nolan, Beatrice. "AI chatbots are harming young people. Regulators are scrambling to keep up." *Fortune*. September 14, 2025. <https://fortune.com/2025/09/14/ai-chatbots-teens-children-mental-health-suicide-openai-chatgpt-regulation-lawsuit/>.

2. Do you commit to maintaining a consistent schema for evaluating potential harms in your public disclosures and risk frameworks, and to providing transparent justifications for changes made to them, or for when a released model evaluation deviates from the established schema?
3. Do you commit to clearly defining the criteria for determining which model releases require updated public disclosures and new risk evaluations, and which descriptors of models exempt them from such evaluations, such as 'experimental,' 'preview,' 'minor,' or other similar terms? Please elaborate on the rationale for why such models may be considered sufficiently low-risk to not warrant full reassessment.
4. Do you commit to flagging significant patterns of violations of risk frameworks for your models and providing explanations for their continued deployment despite these issues?
5. Do you commit to collaborating with external partners—including CAISI, academic researchers, and experts—to establish appropriate timelines for conducting sufficiently robust pre-deployment evaluations?
6. Do you commit to researching the short-term and long-term emotional and psychological impacts to your chatbots' users across various populations—including vulnerable groups such as children and the elderly—and including their results in your public disclosures? If so, please elaborate on how this research will be conducted in consultation with relevant third-party researchers.
7. Do you commit to disclosing how safeguards for vulnerable groups – such as children and senior citizens, and those experiencing mental or emotional distress – including but not limited to age-estimation technologies and mental health referrals? If so, please elaborate on how they are informed by best practices to ensure user safety and wellbeing, data privacy, and efficacy.
8. Do you commit to disclosing whether information from your company's chatbot conversations is used for targeted advertising? If your company already discloses this information, how do you disclose that fact to users? If conversation data is used for targeted advertising or otherwise shared with third parties, will you commit to conducting ongoing analysis of potential data privacy, cybersecurity, and other risks from such access?
9. Do you commit to publicly disclosing new information about other potential risks as they are made known to you in a timely manner, including categories of harm not previously recognized or accounted for within your existing evaluation frameworks?
10. Do you commit to ensuring that public disclosure information that is relevant for consumers is shared in a manner accessible and understandable to consumers? If so, please disclose where you will host this information.

11. Do you commit to disclosing whether your company withdraws from, or modifies its full participation in, voluntary agreements or frameworks?

Thank you for your attention to these matters.

Sincerely,

A handwritten signature in blue ink, appearing to read "Brian Schatz", written over a horizontal line.

Brian Schatz
United States Senator

A handwritten signature in blue ink, appearing to read "Katie Boyd Britt", written over a horizontal line.

Katie Boyd Britt
United States Senator

A handwritten signature in blue ink, appearing to read "James Lankford", written over a horizontal line.

James Lankford
United States Senator

A handwritten signature in blue ink, appearing to read "Chris Coons", written over a horizontal line.

Christopher A. Coons
United States Senator

United States Senate

WASHINGTON, DC 20510

December 11, 2025

Sam Altman
Chief Executive Officer
OpenAI
3180 18th Street
San Francisco, CA 94110

Dear Mr. Altman,

We write to request additional information about your company's public reporting practices regarding its artificial intelligence (AI) models. As a market leader for consumer AI chatbots, the steps that you take to increase transparency around your capabilities and risk evaluations have direct impacts on the wellbeing of Americans who use your products. We are already seeing how increasingly powerful models have become integrated into many aspects of personal life for users. As demonstrated harms to users' psychological well-being have emerged, it is critical that, in addition to improving safety design for users, your company takes necessary steps to foster public transparency to promote these goals.

In the past few years, reports have emerged about chatbots that have engaged in suicidal fantasies with children, drafted suicide notes, and provided specific instructions on self-harm.¹ These incidents have exposed how companies can fail to adequately evaluate models for possible use cases and inadequately disclose known risks associated with chatbot use. Additionally, we have seen how companies can struggle to prevent known model risks or unwanted behaviors prior to deployment.² If AI companies struggle to predict and mitigate relatively well-understood risks, it raises concerns about their ability to manage more complex risks. While we have been encouraged by the arrival of overdue safety measures for certain chatbots, these must be accompanied by improved public self-reporting so that consumers, families, educators, and policymakers can make informed decisions around appropriate use. More detailed information is

¹ Nitasha Tiku. "A Teen Contemplating Suicide Turned to a Chatbot. Is It Liable for Her Death?" The Washington Post. September 16, 2025. <https://www.washingtonpost.com/technology/2025/09/16/character-ai-suicide-lawsuit-new-juliana/?ref=dispatch.techoversight.org>;

Tiffany, Kaitlyn and Matteo Wong. "AI's Emerging Teen-Health Crisis" The Atlantic. September 18, 2025. <https://www.theatlantic.com/technology/2025/09/openai-teen-safety/684268/>;

Andoh, Efua. "Many teens are turning to AI chatbots for friendship and emotional support." American Psychological Association. October 1, 2025. <https://www.apa.org/monitor/2025/10/technology-youth-friendships>.

² Sharma, Mrinank, Meg Tong, Tomasz Korbak, David Duvenaud, Amanda Askell, Samuel R. Bowman, Newton Cheng, et al. "Towards Understanding Sycophancy in Language Models." ArXiv. October 20, 2023. <https://doi.org/10.48550/arXiv.2310.13548>;

Sample, Ian. "'Sycophantic' AI Chatbots Tell Users What They Want to Hear, Study Shows." The Guardian. October 24, 2025. <https://www.theguardian.com/technology/2025/oct/24/sycophantic-ai-chatbots-tell-users-what-they-want-to-hear-study-shows>.

still needed for age-restricted models; even chatbots aimed at children *with* additional safeguards have produced pro-eating disorder, violent, and sexual content.³

In addition to impacts on mental health and risks to vulnerable users, accelerating model capabilities necessitate greater transparency around other potential risks involving public safety and national security. In particular, companies have disclosed how advanced models may pose misuse risks in areas including cybersecurity and biosecurity.⁴ Many frontier AI companies made voluntary commitments at the Seoul AI Summit, or in support of the G7 Code of Conduct, to provide transparency into their efforts to assess risks to national security.⁵ We are supportive of ongoing disclosures for these risks, and request that companies adhere to their prior commitments.

Public disclosure reports, such as AI model and system cards, serve as the closest equivalent to nutrition labels for AI models. While they are essential public transparency tools, today's changed landscape calls for assessing current best practices and how they can be better responsive to user risks. Current public disclosure practices can be inconsistent or insufficient, may not be released alongside product launches, and can lack standardization.⁶ The distinction between major and minor releases is left to the discretion of developers, sometimes without explanation. Model and system cards may also fail to incorporate new or updated information about existing models while models are deployed to the public, including information about user safeguards.⁷ Companies must continue to monitor their model performance and publicly disclose new developments as they relate to security and user safety. This information enables third-party evaluators to assess a model's risks and supports organizations, governments, and consumers in making more informed decisions.

It is critical that public disclosures and risk evaluations are comprehensive, consistent, timely, and responsive to emerging risks. We therefore request responses to the following questions by January 8, 2026:

1. Do you commit to publishing model and system cards in a standardized location concurrently or prior to future model releases?

³ Baker-White, Emily. "AI Tutors for Kids Gave Fentanyl Recipes and Dangerous Diet Advice." *Forbes*. May 12, 2025. <https://www.forbes.com/sites/emilybaker-white/2025/05/12/these-ai-tutors-for-kids-gave-fentanyl-recipes-and-dangerous-diet-advice/>.

⁴ Pannu, Jaspreet et al. "Dual-use capabilities of concern of biological AI models." *PLOS Computational Biological*. May 8, 2025. <https://pmc.ncbi.nlm.nih.gov/articles/PMC12061118/>.

⁵ U.K. Department for Science, Innovation & Technology. "Frontier AI Safety Commitments, AI Seoul Summit 2024." February 7, 2025. <https://www.gov.uk/government/publications/frontier-ai-safety-commitments-ai-seoul-summit-2024/frontier-ai-safety-commitments-ai-seoul-summit-2024>;

⁶ Perrigo, Billy. "Top AI Companies Have 'Unacceptable' Risk Management, Studies Say." *TIME*. July 17, 2025. <https://time.com/7302757/anthropic-xai-meta-openai-risk-management-2/>.

The Midas Project. "Seoul Commitment Tracker." Seoul Commitment Tracker. Accessed December 2025. <https://www.seoul-tracker.org/>.


⁷ Nolan, Beatrice. "AI chatbots are harming young people. Regulators are scrambling to keep up." *Fortune*. September 14, 2025. <https://fortune.com/2025/09/14/ai-chatbots-teens-children-mental-health-suicide-openai-chatgpt-regulation-lawsuit/>.

2. Do you commit to maintaining a consistent schema for evaluating potential harms in your public disclosures and risk frameworks, and to providing transparent justifications for changes made to them, or for when a released model evaluation deviates from the established schema?
3. Do you commit to clearly defining the criteria for determining which model releases require updated public disclosures and new risk evaluations, and which descriptors of models exempt them from such evaluations, such as 'experimental,' 'preview,' 'minor,' or other similar terms? Please elaborate on the rationale for why such models may be considered sufficiently low-risk to not warrant full reassessment.
4. Do you commit to flagging significant patterns of violations of risk frameworks for your models and providing explanations for their continued deployment despite these issues?
5. Do you commit to collaborating with external partners—including CAISI, academic researchers, and experts—to establish appropriate timelines for conducting sufficiently robust pre-deployment evaluations?
6. Do you commit to researching the short-term and long-term emotional and psychological impacts to your chatbots' users across various populations—including vulnerable groups such as children and the elderly—and including their results in your public disclosures? If so, please elaborate on how this research will be conducted in consultation with relevant third-party researchers.
7. Do you commit to disclosing how safeguards for vulnerable groups – such as children and senior citizens, and those experiencing mental or emotional distress – including but not limited to age-estimation technologies and mental health referrals? If so, please elaborate on how they are informed by best practices to ensure user safety and wellbeing, data privacy, and efficacy.
8. Do you commit to disclosing whether information from your company's chatbot conversations is used for targeted advertising? If your company already discloses this information, how do you disclose that fact to users? If conversation data is used for targeted advertising or otherwise shared with third parties, will you commit to conducting ongoing analysis of potential data privacy, cybersecurity, and other risks from such access?
9. Do you commit to publicly disclosing new information about other potential risks as they are made known to you in a timely manner, including categories of harm not previously recognized or accounted for within your existing evaluation frameworks?
10. Do you commit to ensuring that public disclosure information that is relevant for consumers is shared in a manner accessible and understandable to consumers? If so, please disclose where you will host this information.

11. Do you commit to disclosing whether your company withdraws from, or modifies its full participation in, voluntary agreements or frameworks?

Thank you for your attention to these matters.

Sincerely,



Brian Schatz
United States Senator



Katie Boyd Britt
United States Senator



James Lankford
United States Senator



Christopher A. Coons
United States Senator

United States Senate

WASHINGTON, DC 20510

December 11, 2025

Mark Zuckerberg
Chief Executive Officer
Meta Platforms, Inc.
1 Meta Way
Menlo Park, CA 94025

Dear Mr. Zuckerberg,

We write to request additional information about your company's public reporting practices regarding its artificial intelligence (AI) models. As a market leader for consumer AI chatbots, the steps that you take to increase transparency around your capabilities and risk evaluations have direct impacts on the wellbeing of Americans who use your products. We are already seeing how increasingly powerful models have become integrated into many aspects of personal life for users. As demonstrated harms to users' psychological well-being have emerged, it is critical that, in addition to improving safety design for users, your company takes necessary steps to foster public transparency to promote these goals.

In the past few years, reports have emerged about chatbots that have engaged in suicidal fantasies with children, drafted suicide notes, and provided specific instructions on self-harm.¹ These incidents have exposed how companies can fail to adequately evaluate models for possible use cases and inadequately disclose known risks associated with chatbot use. Additionally, we have seen how companies can struggle to prevent known model risks or unwanted behaviors prior to deployment.² If AI companies struggle to predict and mitigate relatively well-understood risks, it raises concerns about their ability to manage more complex risks. While we have been encouraged by the arrival of overdue safety measures for certain chatbots, these must be accompanied by improved public self-reporting so that consumers, families, educators, and policymakers can make informed decisions around appropriate use. More detailed information is

¹ Nitasha Tiku. "A Teen Contemplating Suicide Turned to a Chatbot. Is It Liable for Her Death?" The Washington Post. September 16, 2025. <https://www.washingtonpost.com/technology/2025/09/16/character-ai-suicide-lawsuit-new-juliana/?ref=dispatch.techoversight.org>;

Tiffany, Kaitlyn and Matteo Wong. "AI's Emerging Teen-Health Crisis" The Atlantic. September 18, 2025.

<https://www.theatlantic.com/technology/2025/09/openai-teen-safety/684268/>;

Andoh, Efua. "Many teens are turning to AI chatbots for friendship and emotional support." American Psychological Association. October 1, 2025. <https://www.apa.org/monitor/2025/10/technology-youth-friendships>.

² Sharma, Mrinank, Meg Tong, Tomasz Korbak, David Duvenaud, Amanda Askill, Samuel R. Bowman, Newton Cheng, et al. "Towards Understanding Sycophancy in Language Models." ArXiv. October 20, 2023.

<https://doi.org/10.48550/arXiv.2310.13548>;

Sample, Ian. "'Sycophantic' AI Chatbots Tell Users What They Want to Hear, Study Shows." The Guardian.

October 24, 2025. <https://www.theguardian.com/technology/2025/oct/24/sycophantic-ai-chatbots-tell-users-what-they-want-to-hear-study-shows>.

still needed for age-restricted models; even chatbots aimed at children *with* additional safeguards have produced pro-eating disorder, violent, and sexual content.³

In addition to impacts on mental health and risks to vulnerable users, accelerating model capabilities necessitate greater transparency around other potential risks involving public safety and national security. In particular, companies have disclosed how advanced models may pose misuse risks in areas including cybersecurity and biosecurity.⁴ Many frontier AI companies made voluntary commitments at the Seoul AI Summit, or in support of the G7 Code of Conduct, to provide transparency into their efforts to assess risks to national security.⁵ We are supportive of ongoing disclosures for these risks, and request that companies adhere to their prior commitments.

Public disclosure reports, such as AI model and system cards, serve as the closest equivalent to nutrition labels for AI models. While they are essential public transparency tools, today's changed landscape calls for assessing current best practices and how they can be better responsive to user risks. Current public disclosure practices can be inconsistent or insufficient, may not be released alongside product launches, and can lack standardization.⁶ The distinction between major and minor releases is left to the discretion of developers, sometimes without explanation. Model and system cards may also fail to incorporate new or updated information about existing models while models are deployed to the public, including information about user safeguards.⁷ Companies must continue to monitor their model performance and publicly disclose new developments as they relate to security and user safety. This information enables third-party evaluators to assess a model's risks and supports organizations, governments, and consumers in making more informed decisions.

It is critical that public disclosures and risk evaluations are comprehensive, consistent, timely, and responsive to emerging risks. We therefore request responses to the following questions by January 8, 2026:

1. Do you commit to publishing model and system cards in a standardized location concurrently or prior to future model releases?

³ Baker-White, Emily. "AI Tutors for Kids Gave Fentanyl Recipes and Dangerous Diet Advice." *Forbes*. May 12, 2025. <https://www.forbes.com/sites/emilybaker-white/2025/05/12/these-ai-tutors-for-kids-gave-fentanyl-recipes-and-dangerous-diet-advice/>.

⁴ Pannu, Jaspreet et al. "Dual-use capabilities of concern of biological AI models." *PLOS Computational Biological*. May 8, 2025. <https://pmc.ncbi.nlm.nih.gov/articles/PMC12061118/>.

⁵ U.K. Department for Science, Innovation & Technology. "Frontier AI Safety Commitments, AI Seoul Summit 2024." February 7, 2025. <https://www.gov.uk/government/publications/frontier-ai-safety-commitments-ai-seoul-summit-2024/frontier-ai-safety-commitments-ai-seoul-summit-2024>;

⁶ Perrigo, Billy. "Top AI Companies Have 'Unacceptable' Risk Management, Studies Say." *TIME*. July 17, 2025. <https://time.com/7302757/anthropic-xai-meta-openai-risk-management-2/>.

The Midas Project. "Seoul Commitment Tracker." Seoul Commitment Tracker. Accessed December 2025. <https://www.seoul-tracker.org/>.


⁷ Nolan, Beatrice. "AI chatbots are harming young people. Regulators are scrambling to keep up." *Fortune*. September 14, 2025. <https://fortune.com/2025/09/14/ai-chatbots-teens-children-mental-health-suicide-openai-chatgpt-regulation-lawsuit/>.

2. Do you commit to maintaining a consistent schema for evaluating potential harms in your public disclosures and risk frameworks, and to providing transparent justifications for changes made to them, or for when a released model evaluation deviates from the established schema?
3. Do you commit to clearly defining the criteria for determining which model releases require updated public disclosures and new risk evaluations, and which descriptors of models exempt them from such evaluations, such as 'experimental,' 'preview,' 'minor,' or other similar terms? Please elaborate on the rationale for why such models may be considered sufficiently low-risk to not warrant full reassessment.
4. Do you commit to flagging significant patterns of violations of risk frameworks for your models and providing explanations for their continued deployment despite these issues?
5. Do you commit to collaborating with external partners—including CAISI, academic researchers, and experts—to establish appropriate timelines for conducting sufficiently robust pre-deployment evaluations?
6. Do you commit to researching the short-term and long-term emotional and psychological impacts to your chatbots' users across various populations—including vulnerable groups such as children and the elderly—and including their results in your public disclosures? If so, please elaborate on how this research will be conducted in consultation with relevant third-party researchers.
7. Do you commit to disclosing how safeguards for vulnerable groups – such as children and senior citizens, and those experiencing mental or emotional distress – including but not limited to age-estimation technologies and mental health referrals? If so, please elaborate on how they are informed by best practices to ensure user safety and wellbeing, data privacy, and efficacy.
8. Do you commit to disclosing whether information from your company's chatbot conversations is used for targeted advertising? If your company already discloses this information, how do you disclose that fact to users? If conversation data is used for targeted advertising or otherwise shared with third parties, will you commit to conducting ongoing analysis of potential data privacy, cybersecurity, and other risks from such access?
9. Do you commit to publicly disclosing new information about other potential risks as they are made known to you in a timely manner, including categories of harm not previously recognized or accounted for within your existing evaluation frameworks?
10. Do you commit to ensuring that public disclosure information that is relevant for consumers is shared in a manner accessible and understandable to consumers? If so, please disclose where you will host this information.

11. Do you commit to disclosing whether your company withdraws from, or modifies its full participation in, voluntary agreements or frameworks?

Thank you for your attention to these matters.

Sincerely,



Brian Schatz
United States Senator



Katie Boyd Britt
United States Senator



James Lankford
United States Senator



Christopher A. Coons
United States Senator

United States Senate

WASHINGTON, DC 20510

December 11, 2025

Sundar Pichai
Chief Executive Officer
Google LLC
1600 Amphitheatre Parkway
Mountain View, CA 94043

Dear Mr. Pichai,

We write to request additional information about your company's public reporting practices regarding its artificial intelligence (AI) models. As a market leader for consumer AI chatbots, the steps that you take to increase transparency around your capabilities and risk evaluations have direct impacts on the wellbeing of Americans who use your products. We are already seeing how increasingly powerful models have become integrated into many aspects of personal life for users. As demonstrated harms to users' psychological well-being have emerged, it is critical that, in addition to improving safety design for users, your company takes necessary steps to foster public transparency to promote these goals.

In the past few years, reports have emerged about chatbots that have engaged in suicidal fantasies with children, drafted suicide notes, and provided specific instructions on self-harm.¹ These incidents have exposed how companies can fail to adequately evaluate models for possible use cases and inadequately disclose known risks associated with chatbot use. Additionally, we have seen how companies can struggle to prevent known model risks or unwanted behaviors prior to deployment.² If AI companies struggle to predict and mitigate relatively well-understood risks, it raises concerns about their ability to manage more complex risks. While we have been encouraged by the arrival of overdue safety measures for certain chatbots, these must be accompanied by improved public self-reporting so that consumers, families, educators, and policymakers can make informed decisions around appropriate use. More detailed information is

¹ Nitasha Tiku. "A Teen Contemplating Suicide Turned to a Chatbot. Is It Liable for Her Death?" The Washington Post. September 16, 2025. <https://www.washingtonpost.com/technology/2025/09/16/character-ai-suicide-lawsuit-new-juliana/?ref=dispatch.techoversight.org>;

Tiffany, Kaitlyn and Matteo Wong. "AI's Emerging Teen-Health Crisis" The Atlantic. September 18, 2025.

<https://www.theatlantic.com/technology/2025/09/openai-teen-safety/684268/>;

Andoh, Efua. "Many teens are turning to AI chatbots for friendship and emotional support." American Psychological Association. October 1, 2025. <https://www.apa.org/monitor/2025/10/technology-youth-friendships>.

² Sharma, Mrinank, Meg Tong, Tomasz Korbak, David Duvenaud, Amanda Askill, Samuel R. Bowman, Newton Cheng, et al. "Towards Understanding Sycophancy in Language Models." ArXiv. October 20, 2023.

<https://doi.org/10.48550/arXiv.2310.13548>;

Sample, Ian. "'Sycophantic' AI Chatbots Tell Users What They Want to Hear, Study Shows." The Guardian.

October 24, 2025. <https://www.theguardian.com/technology/2025/oct/24/sycophantic-ai-chatbots-tell-users-what-they-want-to-hear-study-shows>.

still needed for age-restricted models; even chatbots aimed at children *with* additional safeguards have produced pro-eating disorder, violent, and sexual content.³

In addition to impacts on mental health and risks to vulnerable users, accelerating model capabilities necessitate greater transparency around other potential risks involving public safety and national security. In particular, companies have disclosed how advanced models may pose misuse risks in areas including cybersecurity and biosecurity.⁴ Many frontier AI companies made voluntary commitments at the Seoul AI Summit, or in support of the G7 Code of Conduct, to provide transparency into their efforts to assess risks to national security.⁵ We are supportive of ongoing disclosures for these risks, and request that companies adhere to their prior commitments.

Public disclosure reports, such as AI model and system cards, serve as the closest equivalent to nutrition labels for AI models. While they are essential public transparency tools, today's changed landscape calls for assessing current best practices and how they can be better responsive to user risks. Current public disclosure practices can be inconsistent or insufficient, may not be released alongside product launches, and can lack standardization.⁶ The distinction between major and minor releases is left to the discretion of developers, sometimes without explanation. Model and system cards may also fail to incorporate new or updated information about existing models while models are deployed to the public, including information about user safeguards.⁷ Companies must continue to monitor their model performance and publicly disclose new developments as they relate to security and user safety. This information enables third-party evaluators to assess a model's risks and supports organizations, governments, and consumers in making more informed decisions.

It is critical that public disclosures and risk evaluations are comprehensive, consistent, timely, and responsive to emerging risks. We therefore request responses to the following questions by January 8, 2026:

1. Do you commit to publishing model and system cards in a standardized location concurrently or prior to future model releases?

³ Baker-White, Emily. "AI Tutors for Kids Gave Fentanyl Recipes and Dangerous Diet Advice." *Forbes*. May 12, 2025. <https://www.forbes.com/sites/emilybaker-white/2025/05/12/these-ai-tutors-for-kids-gave-fentanyl-recipes-and-dangerous-diet-advice/>.

⁴ Pannu, Jaspreet et al. "Dual-use capabilities of concern of biological AI models." *PLOS Computational Biological*. May 8, 2025. <https://pmc.ncbi.nlm.nih.gov/articles/PMC12061118/>.

⁵ U.K. Department for Science, Innovation & Technology. "Frontier AI Safety Commitments, AI Seoul Summit 2024." February 7, 2025. <https://www.gov.uk/government/publications/frontier-ai-safety-commitments-ai-seoul-summit-2024/frontier-ai-safety-commitments-ai-seoul-summit-2024>;

⁶ Perrigo, Billy. "Top AI Companies Have 'Unacceptable' Risk Management, Studies Say." *TIME*. July 17, 2025. <https://time.com/7302757/anthropic-xai-meta-openai-risk-management-2/>.

The Midas Project. "Seoul Commitment Tracker." Seoul Commitment Tracker. Accessed December 2025. <https://www.seoul-tracker.org/>.


⁷ Nolan, Beatrice. "AI chatbots are harming young people. Regulators are scrambling to keep up." *Fortune*. September 14, 2025. <https://fortune.com/2025/09/14/ai-chatbots-teens-children-mental-health-suicide-openai-chatgpt-regulation-lawsuit/>.

2. Do you commit to maintaining a consistent schema for evaluating potential harms in your public disclosures and risk frameworks, and to providing transparent justifications for changes made to them, or for when a released model evaluation deviates from the established schema?
3. Do you commit to clearly defining the criteria for determining which model releases require updated public disclosures and new risk evaluations, and which descriptors of models exempt them from such evaluations, such as 'experimental,' 'preview,' 'minor,' or other similar terms? Please elaborate on the rationale for why such models may be considered sufficiently low-risk to not warrant full reassessment.
4. Do you commit to flagging significant patterns of violations of risk frameworks for your models and providing explanations for their continued deployment despite these issues?
5. Do you commit to collaborating with external partners—including CAISI, academic researchers, and experts—to establish appropriate timelines for conducting sufficiently robust pre-deployment evaluations?
6. Do you commit to researching the short-term and long-term emotional and psychological impacts to your chatbots' users across various populations—including vulnerable groups such as children and the elderly—and including their results in your public disclosures? If so, please elaborate on how this research will be conducted in consultation with relevant third-party researchers.
7. Do you commit to disclosing how safeguards for vulnerable groups – such as children and senior citizens, and those experiencing mental or emotional distress – including but not limited to age-estimation technologies and mental health referrals? If so, please elaborate on how they are informed by best practices to ensure user safety and wellbeing, data privacy, and efficacy.
8. Do you commit to disclosing whether information from your company's chatbot conversations is used for targeted advertising? If your company already discloses this information, how do you disclose that fact to users? If conversation data is used for targeted advertising or otherwise shared with third parties, will you commit to conducting ongoing analysis of potential data privacy, cybersecurity, and other risks from such access?
9. Do you commit to publicly disclosing new information about other potential risks as they are made known to you in a timely manner, including categories of harm not previously recognized or accounted for within your existing evaluation frameworks?
10. Do you commit to ensuring that public disclosure information that is relevant for consumers is shared in a manner accessible and understandable to consumers? If so, please disclose where you will host this information.

11. Do you commit to disclosing whether your company withdraws from, or modifies its full participation in, voluntary agreements or frameworks?

Thank you for your attention to these matters.

Sincerely,

A handwritten signature in blue ink, appearing to read "Brian Schatz", written over a horizontal line.

Brian Schatz
United States Senator

A handwritten signature in blue ink, appearing to read "Katie Boyd Britt", written over a horizontal line.

Katie Boyd Britt
United States Senator

A handwritten signature in blue ink, appearing to read "James Lankford", written over a horizontal line.

James Lankford
United States Senator

A handwritten signature in blue ink, appearing to read "Chris Coons", written over a horizontal line.

Christopher A. Coons
United States Senator