AI ELECTION ACCORD

CAI = Content Authority Initiative
C2PA = Coalition for Content Provenance and Authenticity
DAIEC = Deceptive AI election content
PAI = Partnership on AI
NIST = National Institute of Standards and Technology

1. Technology to mitigate risk	**2. Assess existing models for deceptive AI election content risk**	**3. Detect distribution of deceptive AI election content**	**4. Appropriately address deceptive AI election content**	**5. Foster cross-industry resilience**	**6. Transparency to the public**	**7. Continued engagement with civil society organizations and subject matter experts**	**8. Foster public awareness and resilience against deceptive AI election content**
Develop and implement technology to mitigate deceptive AI election content risk By: - Identifying generated content or certifying authenticity. This could include developing classifiers or robust provenance methods like watermarking or signed metadata - Continue to invest in advancing provenance technology for audio, video, and images - Working toward attaching machine-readable information to AI-generated content	- Assessing models to understand the risks they may present regarding deceptive AI election content - Risk evaluations - Existing identification methods	- Seeking to detect distribution of deceptive AI election content This might include: - Using detection technology, - Ingesting standards-based identifiers, - Using content moderation services - Creator disclosure of AI content - Reporting mechanism for deceptive AI election content	- Executing on policy and labeling policy - Removal and consequences This may include: - Adopting and publishing policies - Providing contextual information on realistic AI-generated content	By: - Sharing best practices - Exploring pathways to sharing best tools about deceptive AI election content	- Publishing reports, model cards - Publishing policies about how deceptive AI election content is addressed This may include: - Publish deceptive AI election content-specific policies - Publish deceptive AI election content removals - Publish updates on provenance research - Informing the public about actions taken around commitments	Through: - Engage diverse set of civil society organizations, academics, subject matter experts, and governments through established channels or events - This will inform companies' understanding of global risk landscape	For instance, - Public education campaigns to protect against manipulation and deception - Tools and procedures for providing context about content - Developing and releasing open-source tools for others who are responding - Support organizations and communities responding to these risks

1. Adobe

1. Warner letter:	- Cofounded Content Authority Initiative (CAI) and Coalition for Content Provenance and Authenticity (C2PA) in 2019. Also offers authenticity verification functions. - Content generated with Generative Fill, Adobe Firefly, or other Cloud apps have content credentials applied to them. - Generated content downloaded through Adobe stock also has credentials. - Doesn't believe that detection tools are accurate enough yet - instead, focusing on provenance - Support the R&D of detection tech and will employ once reliable ones are available	- Not addressed	- Detect AI-generated content on Adobe Stock through ingesting standard identifiers - Requires Stock creators to label for their "audit" systems - Report products or services that violate terms of service	- Requires Stock creators to label for AI-generated content. - Content credentials attached to GenAI content downloaded from Stock - Removal of the profile if there is a violation - Reporting features for Terms of Service violations	- Leads C2PA and CAI collaborations to drive awareness on provenance technology - Participates in C2PA steering committee and technical working groups - CAI Open Source software development kit for Content Credentials at C2PA standard	- Adobe transparency center and yearly transparency reports about government requests	-Through CAI, worked with various organizations as "original participants" to gain their viewpoints - Helping these orgs implement content credentials through engineering support -worked with policymakers to help them understand CC - worked directly with the FEC to recommend national standards - Worked with the White House to place content provenance on official documentation - DNC and RNC to promote content credentials	- With CAI, made media literacy resources for the public - Adobe Education Exchange promoting media literacy lesson plans and newsletter - CAI open-sources tools for Guardian Project human rights
Evaluation	Commitment met	Commitment not met	Partially met	Partially met	Commitment met	Partially met	Commitment met	Commitment met
reasoning:		- No mention of existing impact evaluations	- Does not mention generation detection Firefly or other applications - Does not mention audit system capabilities beyond creator self-labeling - Does not mention content moderation or other detection methods for distribution channels - Company view is that detection technology is not effective as it currently exists; instead focuses on provenance	- Notably, no policy rules that directly address elections or misinformation - Doesn't mention any content policy or enforcement for any other applications, such as Firefly		- Transparency center is mostly focused on displaying content policies and documenting government requests - Generally lacks information or analysis on the safety of their GenAI applications	- Does not list specific collaborators and remains indirect engagement, i.e. "through CAI"	
2. Progress update:	- Has increased CAI membership to more than 3,300 members. - Worked with "party conventions" to drive awareness on CC. - Supports development of CAI media literacy curriculum - Discusses investments in Content Credentials - Attaches content credentials to Firefly - And allows users to apply CC to their own work in other Adobe apps	- Not addressed	- Not addressed	- Not addressed	- Co-founder of CAI and co-founder and steering committee member of C2PA - Invested in the development and adoption of Content Credentials - Has increased CAI membership to more than 3,300 members.	- Not addressed	- C2PA has a working group of civil society organizations -Worked with party conventions ahead of the election to drive awareness and adoption of Content Credentials	- Helped develop CAI's media literacy curriculum
Evaluation	Commitment met	Commitment not met	Commitment not met	Commitment not met	Commitment met	Commitment not met	Partially met	Commitment met
reasoning:		- Not addressed	- Not addressed	- Not addressed		- Not addressed	- Does not mention direct engagement with civil society or subject matter experts	
3. Brennan Email:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Evaluation	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met
reasoning:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Composite Score:	Commitment met	Commitment not met	Partially met	Partially met	Commitment met	Partially met	Commitment met	Commitment met

2. Amazon

1. Warner letter:	- Deployed invisible watermarks on Amazon Titan Image Generator - Has Amazon Web Services terms of service that prohibit removal - Watermark validation tool available through Amazon Bedrock, but only works for Titan content - Working to implement C2PA	- Cites scanning and ID technology, but does not show evidence of having evaluated each service for capabilities of generating or distributing DAIEC specifically	- Watermark validation tool available through Amazon Bedrock, but only works for Titan content - All services have ML detection and Trust and Safety teams to review flagged content that violates ToS - reporting mechanism for users to flag content	- "AWS's Responsible AI Policy prohibits use of our AI tools to depict a person's voice or likeness without their consent or other appropriate rights, including unauthorized impersonation" - Has a number of policies across platforms that prevent Misinformation and disallow the use of AI for deceptive content or impersonation. This extends to Alexa, AWS, Twitch, and Amazon. - Remove or disable violative content	- Member of Frontier Model Forum - Participant in Partnership on AI (PAI) - Also a member of National Institute of Standards and Technology (NIST) Consortium working group and AI Safety Institute Consortium	- Not addressed	- Works with Global Challenge to Build Trust in the Age of Generative AI, a G7 project with the UN, OECD, and Global PAI.	- Not addressed
Evaluation	Partially met	Partially met	Commitment met	Commitment met	Partially met	Commitment not met	Commitment met	Commitment not met
reasoning:					- Does not mention specific efforts or actions taken beyond being a member or participant	- Not addressed		- Not addressed
2. Progress update:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Evaluation	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met
reasoning:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
3. Brennan Email:	- Incorporate tamper-resistant invisible watermarks directly into image and video generation processes - add Metadata to Nova Canvas based on C2PA standards - Steering committee member of C2PA.	- Benchmarks used throughout training Nova foundation models - Adversarial techniques tested - red-teaming and testing across threats	- Not addressed	- Not addressed	-"industry partnerships and collaboration" - Fronter model forum member - PAI participant - Cybersecurity and Infrastructure Security Agency's initiative on AI	- Not addressed	- Vulnerability reporting for independent researchers	- Not addressed
Evaluation	Commitment met	Commitment met	Commitment not met	Commitment not met	Partially met	Commitment not met	Partially met	Commitment not met
reasoning:			- Not addressed	- Not addressed	- Does not mention specific efforts or actions taken beyond being a member or participant	- Not addressed		- Not addressed
Composite Score:	Commitment met	Commitment met	Commitment met	Commitment met	Partially met	Commitment not met	Commitment met	Commitment not met

3. Anthropic

1. Warner letter:	- Deployed automated detection systems to prevent misuse, misinformation, or influence operations	- Red-teaming existing systems, covering misinformation and bias, adversarial abuse - in-house technical evaluations for election-related risks, including political parity model, refusal, robustness in preventing production disinformation, voter profiling, targeting tactics	"- Deployed automated detection systems to prevent misuse, misinformation, or influence operations" - Flagging of ai-election-related results - classifier to ID election-related queries and redirect users -user report mechanism for UP violations	- Usage policy that prohibits the use of products for political campaigns or lobbying, targeting political campaigns, and campaign chatbots. Administers suspensions for violators - Usage policy also prevents impersonation and mandates disclosures for organizations using Claude wrappers - Account suspension upon serious violation - Redirect users to official election publications upon election-related query	- mentions being in touch with 'other companies' to share election integrity work - "we are engaging with policymakers, other companies, and civil society organizations to share Anthropic's election integrity work"	- "Additionally, we have been in touch with US civil society organizations and policymakers to share our election intervention updates and will continue to share relevant updates in the months leading up to the election. "	- Engaging with policymakers, other companies, and civil society organizations to share Anthropics election integrity work to spread awareness - In touch with the Euro Commission on election integrity research - in touch with US civil society organizations and policymakers, sharing election intervention updates	- Not addressed
Evaluation	Partially met	Commitment met	Commitment met	Commitment met	Partially met	Partially met	Partially met	Commitment not met
reasoning:	- Does not go into detail about these systems and how they mitigate deceptive AI election content risk. Also does not mention any tools that can identify the origin of content				- Does not mention specifics about their efforts with other companies	- Lacks detail about what information is being shared - Does not mention if this is also shared to the public	- Lacks detail about how they are engaging	- Not addressed
2. Progress update:	- Not addressed	- Conducted policy vulnerability testing and published methodology	- Expanded testing methodology informed improvements for detecting election-related misuse	- Updated usage policy to further prohibit the use to interfere with the electoral process, generate misleading information on elections or use for political lobbying - Linked blog mentions detect and redirect	- Not addressed	- public quantitative evaluations in June 2024	-briefed Euro Commission on election integrity research - informed multiple state and federal policymakers in the US of the work - "In touch" with US civil society organizations	- Not addressed
Evaluation	Commitment not met	Commitment met	Partially met	Commitment met	Commitment not met	Commitment met	Partially met	Commitment not met
reasoning:	- Not addressed		- Does not provide detail on detection efforts or methods - Linked blog discusses "detect and re-direct" for "time-sensitive" election queries		- Not addressed		- Lacks detail about how they are engaging with civil society and governments	- Not addressed
3. Brennan Email:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Evaluation	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met

reasoning:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Composite Score:	Partially met	Commitment met	Commitment met	Commitment met	Partially met	Commitment met	Partially met	Commitment not met

4. Arm

1. Warner letter:	- Founding member of C2PA - Developed Arm Security Manifesto to survey the cybersecurity threats landscape - Offers encryption during digital content creation (helps with secure provenance and allows alteration to be detected) -C2PA engages with governments and other orgs to raise awareness about prov tech	N/A	N/A	N/A	- Cites itself as a founding member of the Cybersecurity Tech accord in 2018, which collaboratively works on security. - Founding member of C2PA	N/A	- cites that C2PA engages with governments - Cites itself as a founding member of the Cybersecurity Tech accord in 2018, which also engages with governments, UN, EU, etc. - Worked w/ Center For Strategic & International Study	- Not addressed
Evaluation	Commitment met	N/A	N/A	N/A	Partially met	N/A	Partially met	Commitment not met
reasoning:		N/A	N/A	N/A	- Lacks detail about how Arm was involved in cited cross-industry efforts	N/A	- Lacks detail about how Arm was involved	- Not addressed
2. Progress update:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Evaluation	Commitment not met	N/A	N/A	N/A	Commitment not met	N/A	Commitment not met	Commitment not met
reasoning:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
3. Brennan Email:	- Arm continues to work with C2PA - continued progress on "enabling the hardware and compute technology required to establish the provenance of digital content," - "remain committed to partnering with the wider coalition and Arm ecosystem to drive ongoing progress"	N/A	N/A	N/A	- Continued work with C2PA	N/A	- Not addressed	- Not addressed
Evaluation	Commitment met	N/A	N/A	N/A	Partially met	N/A	Commitment not met	Commitment not met
reasoning:		N/A	N/A	N/A	- Lacks detail about how Arm was involved, specifically regarding collaborative efforts	N/A	- Not addressed	- Not addressed
Composite Score:	Commitment met	N/A	N/A	N/A	Partially met	N/A	Partially met	Commitment not met

5. ElevenLabs

1. Warner letter:	- Released AI speech classifier for ID of Eleven-generated content - Cited general support for downstream detection tools (metadata, watermarks, and fingerprinting solutions) like C2PA - Member of C2PA	- Through automated scans and a growing team of human moderators, they actively monitor content generated to identify and block the creation of audio that violates their terms	- Released AI speech classifier for ID of Eleven-generated content - "no go" voices safeguard prevent certain generations, such as active political candidates - "actively monitor content generated using our technology to identify and block the creation of audio that violates these terms" - Partnered with with Reality Defender and Loccus, which both specialize in deepfake detection - Reporting mechanism for users	- Usage policy generally prohibits election misinformation, interference, and targeting. - Specifically mention they prohibit the use of their tools for voter suppression and disruption of the electoral processes - blocks content that is detected -reporting violations through web form - All accounts are traceable, so misinformation that hits the public can be backtraced and pinned to specific accounts. Removal and report to authorities	- Partners with Reality Defender who does Cybersecurity and deepfake detection, to help with their detection efforts - partner with Loccus, doing similar work -Member of C2PA and support adoption of downstream detection tools - Member of CAI - "collaborating with civil society organizations and academic institutions in the US and UK on the development of our prohibited content and use policies." - Work with companies to spread awareness of deepfakes and the importance of safeguards to protect from their harm	- Usage policy generally prohibits election misinformation, interference, and targeting. - Specifically mention they prohibit the use of their tools for voter suppression and disruption of the electoral processes - blocks content that is detected	- Will report to authorities when applicable - collaborates with other organizations for developing content policies - member of NIST consortium, National Intelligence Office for AI audio - supports Deceptive AI Act - backs specific DAIEC protection bills	- Not addressed
Evaluation	Commitment met	Commitment met	Commitment met	Commitment met	Commitment met	Partially met	Commitment met	Commitment not met
reasoning:						- Does not address publishing activity around deceptive AI election content removals - Does not address public-facing updates on provenance research		- Not addressed

2. Progress update:	- Released AI speech classifier - Cited general support for 'downstream' detection tools like C2PA - Member of C2PA	- Continuous monitoring for usage policy violations, resulting in removals	- Released AI speech classifier for ID of Eleven-generated content Continuous monitoring for usage policy violations, resulting in removals - reporting violations through web form	- Usage policy generally prohibits election misinformation, interference, and targeting. Specifically they prohibit the use of their tools for voter suppression and disruption of the electoral processes - blocks content that is detected - Continuously monitor for and remove content that violates policies - All accounts are traceable, so misinformation that hits the public can be backtraced and pinned to specific accounts. Removal and report to authorities	- Partners with Reality Defender who does Cybersecurity and deepfake detection, to help with detection efforts - Member of C2PA and support adoption of downstream detection tools - Member of CAI - Member of C2PA and support adoption of downstream detection tools	- Usage policy generally prohibits election misinformation, interference, and targeting. - Specifically mention they prohibit the use of their tools for voter suppression and disruption of the electoral processes - blocks content that is detected	- Will report to authorities when applicable - collaborates with other organizations for developing content policies - member of NIST consortium, National Intelligence Office for AI audio - supports Deceptive AI Act - backs specific DAIEC protection bills - Working with congress	- Not addressed
Evaluation	Commitment met	Commitment met	Commitment met	Commitment met	Commitment met	Partially met	Commitment met	Commitment not met
reasoning:						- Does not address publishing activity around deceptive AI election content removals - Does not address public-facing updates on provenance research		- Not addressed
3. Brennan Email:	-AI speech classifier -Member of CAI and C2PA - employing metadata identifiers in their generated content	- external threat intelligence team that advises on potential misuse	- no-go voices blocks specific generations - increased monitoring for political content - external threat intelligence team that advises on potential misuse - reporting mechanism	- Revised usage policy to include banning campaigning, impersonation of candidates, voter suppression and other disruptions - All accounts are screened and information collected	-Member of CAI and C2PA - Partners with Reality Defender who does Cybersecurity and deepfake detection, to help with detection efforts	- Revised usage policy to include banning campaigning, impersonation of candidates, voter suppression and other disruptions	- Support for government action from congress - Partners with UC Berkeley School of Information on content detection	- Not addressed
Evaluation	Commitment met	Commitment met	Commitment met	Commitment met	Commitment met	Partially met	Commitment met	Commitment not met
reasoning:						- Does not address publishing activity around deceptive AI election content removals - Does not address public-facing updates on provenance research		- Not addressed
Composite Score:	Commitment met	Commitment met	Commitment met	Commitment met	Commitment met	Partially met	Commitment met	Commitment not met

6. Gen

1. Warner letter:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Evaluation	N/A	N/A	N/A	Commitment not met	Commitment not met	N/A	Commitment not met	Commitment not met
reasoning:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
2. Progress update:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Evaluation	N/A	N/A	N/A	Commitment not met	Commitment not met	N/A	Commitment not met	Commitment not met
reasoning:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
3. Brennan Email:	N/A	N/A	N/A	- Publish recommended AI policies in the company blog, including 5 principles they follow	- Not addressed	N/A	- Not addressed	- Not addressed
Evaluation	N/A	N/A	N/A	Partially met	Commitment not met	N/A	Commitment not met	Commitment not met
reasoning:	N/A	N/A	N/A	- Published materials do not specifically cover deceptive AI election content	- Not addressed	N/A	- Not addressed	- Not addressed
Composite Score:	N/A	N/A	N/A	Partially met	Commitment not met	N/A	Commitment not met	Commitment not met

7. GitHub

1. Warner letter:	- Not addressed	- Not addressed	- Not addressed	- Github implemented a policy change to disallow projects that involve producing nonconsensual intimate images or disinformation	- Not addressed	- Not addressed	- Not addressed	- Not addressed
Evaluation	Commitment not met	Commitment not met	Commitment not met	Partially met	Commitment not met	Commitment not met	Commitment not met	Commitment not met

reasoning:	- Not addressed	- Does not address any evaluation methods employed for updated acceptable use policy	- Does not address any detection methods employed for updated acceptable use policy	- Policy does not specifically cover deceptive AI election content	- Not addressed	- Not addressed	- Not addressed	- Not addressed
2. Progress update:	- Not addressed	- Not addressed	- Not addressed	- updated use policies to address nonconsensual intimate imagery and disinformation - Does not allow projects that promote these or imply them - enforces with removal - Have actioned on violative repositories	- Not addressed	- Not addressed	- Not addressed	- Not addressed
Evaluation	Commitment not met	Commitment not met	Commitment not met	Commitment met	Commitment not met	Commitment not met	Commitment not met	Commitment not met
reasoning:	- Not addressed	- Does not address any evaluation methods employed for updated acceptable use policy	- Does not address any detection methods employed for updated acceptable use policy		- Not addressed	- Not addressed	- Not addressed	- Not addressed
3. Brennan Email:	- Not addressed	- Not addressed	- Not addressed	- Not addressed	- Not addressed	- Not addressed	- Not addressed	- Not addressed
Evaluation	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met
reasoning:	- Not addressed	- Not addressed	- Not addressed	- Not addressed	- Not addressed	- Not addressed	- Not addressed	- Not addressed
Composite Score:	Commitment not met	Commitment not met	Commitment not met	Commitment met	Commitment not met	Commitment not met	Commitment not met	Commitment not met

8. Google

1. Warner letter:	- SynthID watermarking for Gemini Chat, ImageFX, and other services - GenAI products recent introduction to restrict election-relate outputs - Part of C2PA - Investing in provenance detection technology	- AI-assisted red-teaming existing models, training AI agents to red-team - Thousands of internal safety specialists giving feedback to improve models	- Investing in provenance detection technology - creator disclosure on youtube - priority flagger program for specific organizations to have impactful reporting schemes - "prebunking" campaign - fact check force in India to early detect misinformation	- Required ad disclosures for generated content - Prohibit impersonation content on YouTube - reporting mechanism for AI-generated content - Google ads also enforces misrepresentation rules - AI use policies prohibit misinformation - restrictions on election-related queries for Gemini and AI overviews -Additional context via 'About this image' feature and double-check feature in Gemini - Content context messages for breaking news stories on YouTube and other sources of election news	- Engages with a number of civil society and research organizations - PAI, ML Commons, Frontier Model Forum - Steering member of the C2PA coalition - Help other companies develop threat management systems	- A large number of DAIEC-specific policies across platforms - Google threat intelligence publishes findings for government-backed attacker groups - Publish updates about Synth ID research	- works with state election offices and govt election commissions across U.S, india, and the European parliament - engages with a number of civil society and research organizations on fact checking and media literacy - Provides education around cybersecurity for high-risk individuals and orgs	- Google civic outreach holds trainings for google tools, including AI, for political campaigns - election info panels on multiple services - invests 25 Million euro to media literacy educational programs - European hub for resources and training for campaigns - Support Global Fact Check Fund and other organizations
Evaluation	Commitment met	Commitment met	Commitment met	Commitment met	Commitment met	Commitment met	Commitment met	Commitment met
reasoning:								
2. Progress update:	- Developed model cards to promote transparency - expanding SynthID to text, audio, images, video -CP2A member and looking to integrate into platforms	- Mentions red-teaming efforts - Developed model cards to promote transparency	- Disclosure requirement for electi	- AI prohibited use policy - election ad disclosures for synthetic content - Youtube synthetic content disclosures as well	- PAI, ML Commons, Frontier Model Forum - Joined the C2PA coalition	- Shares research papers from internal teams - Shares research into provenance solutions - Developed model cards to promote transparency	- Supports the Global Fact Check Fund - Mentions they support numerous civil society, research, and media literacy efforts	- Mentions supporting media literacy efforts
Evaluation	Commitment met	Commitment met	Commitment met	Commitment met	Commitment met	Commitment met	Commitment met	Partially met
reasoning:								- Does not mention specifics, or how they are supporting these efforts
3. Brennan Email:	- Cites progress report as most recent update	- Ongoing red-teaming to test boundaries on Gemini and frontier models - Publish model cards for frontier models - safety evaluations of existing models	- Cites progress report as most recent update	- Cites progress report as most recent update	- Contributed to PAI framework - Responsible AI toolkit for developers	- Cites progress report as most recent update	-$10 million dollar AI Safety Fund to academia, research orgs, startups, etc. - Working with governments to contribute tools, for example, NSF AI Research Resource pilot for AI research - Signed Seoul Frontier AI Safety Commitments - Adversarial Nubbler Challenge collaborative project - Vulnerability rewards programs	- Cites progress report as most recent update
Evaluation	Commitment met	Commitment met	Commitment met	Commitment met	Commitment met	Commitment met	Commitment met	Commitment met
reasoning:								
Composite Score:	Commitment met	Commitment met	Commitment met	Commitment met	Commitment met	Commitment met	Commitment met	Commitment met

9. IBM

1. Warner letter:	- IBM is one of 19 cross-industry companies co-creating data provenance standards to help organizations determine if data is suitable and trusted for use	- Not addressed	- Watsonx.governance platform has detection for IBM AI tooling	- Not addressed	- released AI 360 toolkits as open-source - member of data and Trust Alliance - member of AI Alliance	- released AI 360 toolkits as open-source	-Cofounded Notre Dame-IBM Technology Ethics Lab to promote interdisciplinary research - member of NIST - member of AI Alliance	- IBM SkillsBuild to educate the public on AI skills - released AI 360 toolkits as open-source
Evaluation	Partially met	Commitment not met	Partially met	Commitment not met	Commitment met	Partially met	Commitment met	Partially met
reasoning:	- Does not mention provenance in reference to their own models - No reference on the 19 companies	- Mentions mitigating risk, but does not discuss risk of model, and does not reference deceptive AI election content	- Does not mention specifics of moderation, ingesting, or reporting of deceptive AI election content	- Not addressed		- Does not mention transparency efforts for their platform		- New goal announced, but does not reference existing efforts
2. Progress update:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Evaluation	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met
reasoning:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
3. Brennan Email:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Evaluation	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met
reasoning:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Composite Score:	Partially met	Commitment not met	Partially met	Commitment not met	Commitment met	Partially met	Commitment met	Partially met

10. Inflection

1. Warner letter:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Evaluation	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met
reasoning:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
2. Progress update:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Evaluation	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met
reasoning:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
3. Brennan Email:	- Actively researching and evaluating various methods like data embedding	- continuous expansion of of safety benchmarks - evaluation of models against misuse scenarios - Internal red-teaming - But bounty program	- Reporting mechanism for users	- Publish user guidelines for prohibited uses	- Publically release safety benchmark dataset - Participate in industry consortia and standards bodies	- Publically release safety benchmark dataset - "plan to" provide regular public reports and technical papers - Publish user guidelines for prohibited uses	-Host and participate in workshops, webinars, and panels with government, academia, and non-profits -Invite leading experts to speak with internal teams about AI risks - "frequently collaborate with advocacy groups, NGOs, and academic institutions to ensure our research and models align with broader societal values and legal frameworks"	- Not addressed
Evaluation	Partially met	Commitment met	Partially met	Partially met	Partially met	Partially met	Partially met	Commitment not met
reasoning:	- Does not include specific efforts	- Does not include specific efforts related to deceptive AI election content	- Does not include specific efforts related to deceptive AI election content	- Does not include specific efforts related to deceptive AI election content	- Does not include reference for these efforts	- Does not include specifics about efforts - References future projects that are not yet completed	- Does not include reference for these efforts	- Not addressed
Composite Score:	Partially met	Commitment met	Partially met	Partially met	Partially met	Partially met	Partially met	Commitment not met

11. Intuit

1. Warner letter:	- Not addressed	- Not addressed	- Mailchimp GenAI tool has automated abuse prevention system and a human review process - "We use a combination of human review and automated content moderation to try to detect and prevent any form of deceptive content" - Content moderation teams - Online complaint center to report issues	- Responsible AI program, content moderation principles, terms of service, acceptable use policy, etc. - prohibit selling in mailchimp store any material that would violate someone's rights	- Actively participate in the NIST AI Safety Institute Consortium	- Not addressed	- NIST consortium - Mentions they engage with trade groups and coalitions to enhance awareness and share information	-"we educate Mailchimp users on industry best practices through our guides and tutorials and application messaging"
Evaluation	Commitment not met	Commitment not met	Partially met	Partially met	Partially met	Commitment not met	Partially met	Commitment not met
reasoning:	- Not addressed	- Not addressed	- Lacks specificity related to detection of generative content and deceptive AI election content specifically	- Lacks specificity related to policy of deceptive AI election content specifically	- Efforts are not as robust as compared to other companies	- Not addressed		- Does not provide enough detail to evaluate. Industry best practices does not imply it is media literacy or DAIEC
2. Progress update:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Evaluation	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met
reasoning:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
3. Brennan Email:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Evaluation	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met
reasoning:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Composite Score:	Commitment not met	Commitment not met	Partially met	Partially met	Partially met	Commitment not met	Partially met	Commitment not met

12. LG AI Research

1. Warner letter:	- A form of written watermarking for synthetic content, actual text on image - Considering participation in C2PA - researching own watermarking technology	- Not addressed	- Developing detection models for LLM-generated text - Mention that AI model development and hosting companies in South Korea have proactively taken measures to ensure compliance with the ban on election campaigning using deepfakes - reporting feature for users	- Generally prohibits all election related content in GenAI models - reporting feature for b2b applications - Usage policy applies here as well	- Participates in various government forums led by South Korean government bodies - Share information on AI-generated content detection technologies through the AI Ethics Policy Forum and the Council for Promoting HyperScale AI, both led by the Ministry of Science and ICT	- Share information on detection at various government forums	- Participates in various government forums led by South Korean government bodies - UNESCO AI Ethics MOOC launch in 2026	- Offer AI education to over 30,000 students - UNESCO AI Ethics MOOC launch in 2026
Evaluation	Partially met	Commitment not met	Partially met	Partially met	Partially met	Partially met	Partially met	Partially met
reasoning:	- Does not include details about advanced provenance efforts	- Not addressed	- Does not include details about what LG is doing specifically	- Lacks detail specifically about deceptive AI election content	- Lacks detail around the specific actions taken	- Does not address if this is available to the public	- Does not include details on contributions or citations to forums	- Current offering does not appear to be deceptive AI election content-related, or related to educating the general public about deceptive AI election content risks
2. Progress update:	- Not addressed	- Not addressed	- Not addressed	- ChatEXAONE, an Enterprise AI Agent, cites its sources and presents evidence to improve accuracy of information	- Not addressed	- Not addressed	- UNESCO AI Ethics MOOC launch in 2026	- Developing AI ethics curriculum with UNESCO that will launch in 2026 - LG Discovery Lab and AI Aimers offer practical education to younger generations about the ethical use of AI technologies
Evaluation	Commitment not met	Commitment not met	Commitment not met	Partially met	Commitment not met	Commitment not met	Commitment not met	Partially met
reasoning:	- Not addressed	- Not addressed	- Not addressed	- Does not mention policies related to deceptive AI election content and this model	- Not addressed	- Not addressed	- Not released yet	- Listed efforts do not appear to be deceptive AI election content-related
3. Brennan Email:	- Not addressed	- Not addressed	- Developing a novel detection technology for AI generated images	- Not addressed	- Not addressed	- Publishing LG AI Accountability Report on AI Ethics in February, 2025	- Not addressed	- LG Discovery Lab and AI Aimers offer practical education to younger generations about the ethical use of AI technologies
Evaluation	Commitment not met	Commitment not met	Partially met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Partially met
reasoning:	- Not addressed	- Not addressed	- Does not discuss LG generative models	- Not addressed	- Not addressed	- Not released yet	- Not addressed	- Listed efforts do not appear to be deceptive AI election content-related
Composite Score:	Partially met	Commitment not met	Partially met	Partially met	Partially met	Partially met	Partially met	Partially met

13. Linkedin

1. Warner letter:	- Automatically label content based on C2PA metadata	- Not addressed	- Develops own models to detect generated content, particularly deep fakes of human faces - Ingests C2PA metadata and labels content as AI generated - Reporting features for users - Automatically label content based on C2PA metadata	- Not addressed	- Not addressed	- Not addressed	- Not addressed	- Not addressed
Evaluation	Commitment met	Commitment not met	Commitment met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met
reasoning:		- Not addressed	- Does not discuss deceptive AI election content-related content moderation	- Not addressed	- Not addressed	- Not addressed	- Not addressed	- Not addressed
2. Progress update:	- Automatically label content based on C2PA metadata	- Not addressed	- Not addressed	- Labels synthetic content with CP2A - Cr icon labeling on posts - not DAIEC-specific - prohibits synthetic or manipulated media, misleading content, or photo/audio-realistic content that misrepresents an individual without disclosing the nature of the content	- Not addressed	- Not addressed	- Not addressed	- Not addressed
Evaluation	Commitment met	Commitment not met	Commitment not met	Partially met	Commitment not met	Commitment not met	Commitment not met	Commitment not met
reasoning:		- Not addressed	- Not addressed	- Efforts are not specific to deceptive AI election content	- Not addressed	- Not addressed	- Not addressed	- Not addressed
3. Brennan Email:	- Linkedin became the first major platform to display content credentials	- Not addressed	- Not addressed	- Not addressed	- Not addressed	- Not addressed	- Not addressed	- Not addressed
Evaluation	Commitment met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met
reasoning:		- Not addressed	- Not addressed	- Not addressed	- Not addressed	- Not addressed	- Not addressed	- Not addressed
Composite Score:	Commitment met	Commitment not met	Commitment met	Partially met	Commitment not met	Commitment not met	Commitment not met	Commitment not met

14. McAfee

1. Warner letter:	N/A	N/A	- Working on Project Mockingbird deepfake detector for audio	N/A	- Not addressed	N/A	- Research on election site spoofing, engagement with NSA and CISA - research on public capabilities to spot AI generated content	-educate public on cyber and election harms through blogs and active news hub - Developing more detailed content to alert the public to high-profile deepfakes - Engage with news outlets to provide expertise and spread awareness of issues
Evaluation	N/A	N/A	Partially met	N/A	Commitment not met	N/A	Partially met	Commitment met
reasoning:	N/A	N/A	- Not released at time of letter	N/A	- Not addressed	N/A	- Does not mention specific efforts where engaged with civil society or academia	
2. Progress update:	N/A	N/A	-Deployed McAfee audio deepfake detector in August 2024 - Looking to expand capabilities	N/A	- Not addressed	N/A	- Not addressed	- released mcafee.ai with resources to build awareness of AI-driven scams - blogs and social media posts about DAIEC
Evaluation	N/A	N/A	Commitment met	N/A	Commitment not met	N/A	Commitment not met	Commitment met
reasoning:	N/A	N/A		N/A	- Not addressed	N/A	- Not addressed	
3. Brennan Email:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Evaluation	N/A	N/A	Commitment not met	N/A	Commitment not met	N/A	Commitment not met	Commitment not met
reasoning:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Composite Score:	N/A	N/A	Commitment met	N/A	Commitment not met	N/A	Partially met	Commitment met

15. Microsoft

1. Warner letter:	- Founding member of C2PA - CP2A employed in most popular consumer-facing AI image generation tools - piloting an in-camera provenance tech with TruePic - launched Content Integrity Check ID tool - election authorities were able to sign their content with provenance information	- Not addressed	- linkedin develops own model for synthetic face ID - auto labeling on LinkedIn images/videos with C2PA data - Partners with True Media to provide detection tools to civil society and the government - Content Integrity Check public tool released to scan for content credentials - specialized reporting portal for DAIEC - in-product reporting for many products - “Content Integrity Certify” allows users to add content credentials to their own authentic content - user reporting in services, reporting website	- Bing has links to voting resources in US/EU/UK - policy prohibits deceptive activity, fraud, or misleading content -explicitly prohibits DAIEC	- Grants to PAI -hash sharing initiatives for CSAM and counter-terrorism - Partnerships with TruePic and True Media - founding member of C2PA - Social Resilience Grants with OpenAI	- policy prohibits deceptive activity, fraud, or misleading content -explicitly prohibits DAIEC - Threat Analysis Center	-content integrity tool available to political entities, news and media organizations - briefings with world governments to raise awareness on DAIEC issues - Partnership with True Media to allow access to relevant Microsoft AI operations - Bing worked with NASED to receive the websites for authoritative election information - Mentioned participating in the AZ TTX - Spoke with NASS about the risk environment and possible impacts of AI - reporting portal for candidates/campaigns to report deceptive AI targeting them - Election Communications Hubs for election official support	- societal resilience grants to further AI education AARP, C2PA, International IDEA, and PAI - Media literacy and information literacy partnerships with News Literacy Project - Children's education projects as well -deepfake awareness campaign in EU - funding and research for PSA campaigns
Evaluation	Commitment met	Commitment not met	Commitment met	Commitment met	Commitment met	Partially met	Commitment met	Commitment met
reasoning:		- Does not mention risk evaluations for existing models				- no mention of removal reports or published reports for any product or services relating to deceptive AI election content		
2. Progress update:	-C2PA integrated across most platforms - Pilot program to help political campaigns and news organizations apply these standards to their own authentic media - AI for Good Lab has developed detection models	- Not addressed	- Partnered with True Media to provide governments, civil society, and journalists with access to free tools that help verify media authenticity - Microsoft’s AI for Good Lab has developed detection models - reporting portal for candidates	- Not addressed	- Grants to industry collaboration orgs like C2PA and PAI - Partnered with True Media to provide governments, civil society, and journalists with access to free tools that help verify media authenticity - Partner with TruePic	- Not addressed	- Partnered with True Media to provide governments, civil society, and journalists with access to free tools that help verify media authenticity - Partnership with TruePic	- 150 training sessions for political stakeholders, 4,700 participants - Public awareness campaign about potential AI election risks - 350 million impressions outside of the U.S - 30,000 engagements for the recently begun U.S campaign - societal resilience grants to further AI education AARP, C2PA, International IDEA, and PAI
Evaluation	Commitment met	Commitment not met	Commitment met	Commitment not met	Commitment met	Commitment not met	Partially met	Commitment met
reasoning:		- Does not mention risk evaluations for existing models		- Does not discuss DAIEC policy or how they address DAIEC		- no mention of removal reports or published reports for any product or services relating to DAIEC - Does not discuss policy or actions taken again DAIEC	- Does not mention civil society organizations for DAIEC-related partnerships	
3. Brennan Email:	- Secretary of State Adrien Fontes is able to have content credentials for official images	- Not addressed	- Not addressed	- Not addressed	- Part of the societal resilience grant went to PAI to refine practices on synthetic media framework	- Not addressed	- WITNESS training for journalists and fact-checkers - AI deepfake training with many major political parties	- societal resilience grants to further AI education AARP, C2PA, International IDEA, and PAI - Grant to WITNESS - "Ran a series of public messages and stood up a AI and Elections website focused on engaging voters about the risks of deceptive AI and where to find authoritative election information." - Education campaign for older adults - AI deepfake trainings for politicians and campaigns around the world (including in-person at DNC/RNC)
Evaluation	Partially met	Commitment not met	Commitment not met	Commitment not met	Partially met	Commitment not met	Commitment met	Commitment met
reasoning:		- Does not mention risk evaluations for existing models	- Not addressed	- Does not discuss DAIEC policy or how they address DAIEC		- no mention of removal reports or published reports for any product or services relating to DAIEC - Does not discuss policy or actions taken again DAIEC		
Composite Score:	Commitment met	Commitment not met	Commitment met	Commitment met	Commitment met	Partially met	Commitment met	Commitment met

16. Meta

1. Warner letter:	- employs content labeling across platforms, "AI Info" - For content generated with Meta models, there are both visible and invisible watermarks in the metadata - investigating ways to make the watermarks harder to remove, developing new watermarking technology	- Not addressed	- employ content detection that IDs provenance using metadata - widespread reporting and human evaluations program for IDing and labeling synthetic content - Creator disclosure required for photorealistic generated content, penalties if they do not - Advertiser disclosure also required - Reporting mechanism for users, businesses, and politicians - Automated system flags violative content	- Increasing breadth of labeling across platforms, "AI Info" on video, audio, and image - more prominent label for high-risk content like DAIEC - Content that is IDed as misleading is deprioritized in the feed - Advertisers have to disclose when they use AI to create or alter a political or social issue ad in certain cases - Mention there are numerous ways to report policy violations - Largest independent fact-checking network - Provide official election links as contextual information; Voting Information Center - Country-specific Election Operations Centers used ahead of major elections to triage urgent threats - content that violates policies against voter interference, harassment, violence etc will be removed	- Works with PAI on provenance standards - "we work with industry peers to align on technologies that can make it easier for us and other platform providers to detect when someone shares content that has been AI-generated."	- Adversarial Threat reports - Meta AI Research lab FAIR shares research on Stable Signature - political ads labeled with "paid for by"; ads stored in publicly available library for 7 years	- Advanced Protection program on Facebook for high-risk accounts such as candidates and election officials - Founding member of PAI - work with election officials to issue Voting Alerts with latest voting info	- Working with governments to prevent election official harassment, issue voting alerts - Working with the European Fact-Checking Standards Network on media literacy campaigns
Evaluation	Commitment met	Commitment not met	Commitment met	Commitment met	Partially met	Commitment met	Partially met	Commitment met
reasoning:		- Does not mention risk evaluations for existing models			- Does not include details about specific actions taken		- Does not address ways that collaborations improve deceptive AI election content risks	
2. Progress update:	- Sep 2024 joined C2PA Steering committee - Open-sourced Stable Signature watermarking - Released AudioSeal, audio watermarking technique - For content generated with Meta models, there are both visible and invisible watermarks in the metadata - employs content labeling across platforms, "AI Info"	-Extensive red-teaming on existing models	- Detect industry standard watermarks and appropriately label across platforms - Self-disclosure for content posters - Public reporting - LlamaGuard LLM-powered content moderation - Ingesting standard identifiers and labeling "AI info"	- Required disclosure for AI used on social issue and political ads - Same requirements for Ads that have photorealistic synthetic content - deprioritize synthetic content in feed - Cited extensive fact checking initiatives - Large fact-checking network - remove content that violates community standards (including election content)	- Open sourcing Stable Signature - Work with PAI on the synthetic media transparency methods - Joined C2PA steering committee	- Cited blog post about checking sensitive responses for specific responses - Cited adversarial threat report	- Founding member of PAI	- Not addressed
Evaluation	Commitment met	Partially met	Commitment met	Commitment met	Commitment met	Commitment met	Partially met	Commitment not met
reasoning:		- Does not details of examples of red teaming for deceptive AI election content					- Does not address ways that collaborations improve deceptive AI election content risks	- No mention of efforts to educate the public
3. Brennan Email:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Evaluation	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met
reasoning:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Composite Score:	Commitment met	Partially met	Commitment met	Commitment met	Commitment met	Commitment met	Partially met	Commitment met

17. NetApp

1. Warner letter:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Evaluation	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met
reasoning:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
2. Progress update:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Evaluation	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met
reasoning:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission

3. Brennan Email:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Evaluation	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met
reasoning:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Composite Score:	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met

18. Nota

1. Warner letter:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Evaluation	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met
reasoning:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
2. Progress update:	- Employs C2PA Compliant AI Generation system and "tagging structure"	- Not addressed	- Not addressed	- Not addressed	- Not addressed	- Not addressed	- Not addressed	- Not addressed
Evaluation	Partially met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met
reasoning:	- Does not include details about use of C2PA or examples	- Not addressed	- Not addressed	- Not addressed	- Not addressed	- Not addressed	- Not addressed	- Not addressed
3. Brennan Email:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Evaluation	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met
reasoning:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Composite Score:	Partially met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met

19. OpenAI

1. Warner letter:	- Implements C2PA for Dalle generation - Implementing C2PA for Sora (video model) - Joined C2PA steering committee - experimenting with tamper-proof watermarks - incorporated audio watermarking into their voice model	- Not addressed	- Developing a detection image classifier	- Prohibit the use of tools for deception involving elections - prohibit developers from using chatbots to impersonate real people - directing users to reliable sources for election-related information	- launched a societal resilience fund in partnership with Microsoft - Grant to PAI - Joined C2PA steering committee	-opened applications for access to image classifier to a group of testers from civil society - prohibit developers from using chatbots to impersonate real people - -Prohibit the use of tools for deception involving elections	- Partner with the National association of secretaries of state to direct users to official voting information - Similar partnership in europe - Partner with Cybersecurity and Infrastructure Security Agency and other government stakeholders - organized roundtables with civil society groups on election threats - Outside the US, met with policymakers and EOs in Europe, India, South Korea, and Mexico - Open special access to provenance tools to nonprofits and journalists for	- Direct users to reliable election information - building a number of public engagement and education initiatives through the Societal Resilience Fund with Microsoft with the aim of fostering public trust and improving societal resilience to AI-generated media. - launch societal resilience fund with Microsoft to conduct global trainings for elections officials
Evaluation	Commitment met	Commitment not met	Partially met	Commitment met	Commitment met	Partially met	Commitment met	Commitment met
reasoning:		- No mention of risk assessments for existing models or red-teaming	- Detection classifier is not yet deployed			- Does not provide details about deceptive AI election content removals or how deceptive AI election content is being addressed		
2. Progress update:	- Furthering C2PA, integrated into Dalle generation - Joined C2PA steering committee - Provenance technology for Sora - incorporated watermarks into Voice Engine (voice generation)	- Not addressed	- Developed an image detection classifier for DALLE 3 and shared it with research labs and journalism nonprofits - Systems to prevent misuse, reject prohibited requests	- Usage policy prohibits the use of tools for deceptive political-related activities - Directing users to reliable sources for election-related information	- "regularly publish research" to foster growth in C2PA and provenance technology - joined Microsoft in launching a societal resilience fund	- Publish covert operation disruptions and shared threat intelligence widely	- Endorsed a few deepfake acts: Protect Elections from Deceptive AI Act, No FAKES Act and the California Digital Content Provenance Standards - Shared threat intelligence with government, civil society, and industry stakeholders - Developed an image detection classifier for DALLE 3 and shared it with research labs and journalism nonprofits	- 2 million dollar societal resilience fund for educating the public on these issues - Developed an image detection classifier for DALLE 3 and shared it with research labs and journalism nonprofits - works with NASS and EU parliament to - direct users to reliable sources for election-related information
Evaluation	Commitment met	Commitment not met	Partially met	Commitment met	Partially met	Commitment met	Commitment met	Commitment met

reasoning:		- No mention of risk assessments for existing models or red-teaming	- Does not mention specifics about detection methods in relation to deceptive AI election content		- No further details on this research			
3. Brennan Email:	- Incorporated audio watermarking into Voice Engine - C2PA metadata embedded in image outputs by Dall-E 3 and Sora - Additional research on provenance technology with input through researcher access program	- Red teaming network across organizations and experts - 3rd party red teaming with METR and Apollo research - Researcher access program - Open-source evaluations - Regularly publish system cards	- Responsible disclosure for sensitive vulnerabilities - Automated rejections after violative prompt is detected	- Redirect to authoritative election information	- 10 million dollar AI Safety fund to support new model evaluations and red-teaming strategies - Frontier model forum - Signed AI Seoul summit - UN General Assembly member - C2PA steering committee	- Regularly publish system cards - Publically share disruption of threat actors	- Collaboration with US AI Safety Institute on AI safety research and testing - Democratic input to AI grant program for democratic governance of AI - Partner with Thron to detect and report CSAM - OpenAI Academies investment fund - attendance at AI convenings in the UK, Seoul, Munich, the UN	- Societal resilience fund - Common Sense Media partnership on AI literacy for children and families - Democratic Inputs to AI Grant Program
Evaluation	Commitment met	Commitment met	Commitment met	Partially met	Commitment met	Commitment met	Commitment met	Commitment met
reasoning:				- Does not mention content policy				
Composite Score:	Commitment met	Commitment met	Commitment met	Commitment met	Commitment met	Commitment met	Commitment met	Commitment met

20. Snap Inc

1. Warner letter:	- content watermarks on the surface of generated content, not metadata watermarks - Has content ID for snap-generated content that stays on-platform - Considering ingesting C2PA metadata reading standard, but it is not implemented yet	- All distribution channels are moderated for regular content violations regarding elections - Increased stringency for algorithmically-recommended streams - "pre-moderation systems" - Adversarial testing for AI-specific content policies	- Doesn't do detection for AI-generated content, enforces all content the same and evaluates for election risk - Trusted Flagger program allows trusted orgs and government authorities to escalate high-priority concerns - Content moderation team to evaluate reports - reporting mechanisms	- Pre-moderate content before it is spread at scale - Community Guidelines prohibit harmful false information, hate speech, and threats or calls to violence. - civic integrity policies prohibit election-related misinformation and harmful content, procedural interference (i.e., misinformation related to actual election or civic procedures) - All political ads are human-approved before made public - any content that is achieving large-scale reach is human-reviewed	- collaborative engagements AI Futures Initiative (convened by the Information Technology Industry Council) and the Tech Coalition committee on GenAI Content. - information sharing agreements with other companies - member of the Tech Coalition	- Publish a transparency report every 6 months - Publish blogs discussing their approach to DAIEC	- partners with election officials to prepare official information- UK Electoral Commission, Vote.Org, European Parliament - Worked with civil society, elections authorities, and industry stakeholders to help shape the Voluntary Election Integrity Guidelines for Technology Companies - Provided formal briefings for election authorities in the US and Europe upon request - publicly-released letter to global civil society organizations regarding approach to election integrity	- blog posts about election integrity - Gives funding and ad credits to a cohort of non-partisan, non-profit organizations who are developing a public awareness campaign to educate Americans regarding the risks of deceptive synthetic media in advance of the 2024 elections - Pledged financial and advertising support for a PSA to be produced by a cohort of nonpartisan, non-profit orgs to raise awareness of the risks of deceptive, political deepfakes
Evaluation	Partially met	Partially met	Partially met	Commitment met	Partially met	Commitment met	Commitment met	Commitment met
reasoning:	- Does not mention further provenance research or support	- No mention of Snap Chatbot	- Does not mention detection or ingesting for external content		- Does not mention details about Snap's efforts			
2. Progress update:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Evaluation	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met
reasoning:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
3. Brennan Email:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Evaluation	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met
reasoning:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Composite Score:	Partially met	Partially met	Partially met	Commitment met	Partially met	Commitment met	Commitment met	Commitment met

21. Stability.ai

1. Warner letter:	- content credentials in content made through application programming interface - Implement C2PA standards - investigating ways to implement imperceptible watermarks through their application programming interface	- Review existing model licenses and find that SD2 and SDXL are still protected under OpenRAIL ethical limitations, and SVD and SD3 are covered by acceptable use policy	- After signing tech accord, now have interception technique for Stable Assistant where high-risk prompts are stopped before the system generates something related to political disinformation - Still investigating content detection methods - Safety hotline for user reports	- Earlier image models and latest image and video models are subject to various use policies - Updated policy specifically prohibits the use of generating or promoting disinformation, unlawful impersonation, the production of defamatory content, generating political advertisements, or misrepresenting AI outputs as human-generated	- Work with CAI - Mention they work with other actors in the supply chain, including Google and Meta - Partnerships with AISIC and the AI Alliance, stability has contributed resources and expertise to AISEIC	- Not addressed	- engage with authorities in US, European Union, United Kingdom, Singapore, and Australia - National Center for Missing and Exploited Children, Thorn, and the Internet Watch Foundation (UK) - Partnerships with AISIC and the AI Alliance - testimony at multiple Senate hearings	- Not addressed
Evaluation	Partially met	Partially met	Partially met	Partially met	Partially met	Commitment not met	Partially met	Commitment not met
reasoning:	- Does not mention identification or non-application programming interface generation use cases	- Does not discuss risk evaluations in detail	- Does not discuss detection or ingestion	- Mentions acceptable use policy for SVD and SD3 but does not go into detail	- Does not discuss specific efforts or contributions by Stability as it relates to deceptive AI election content	- Not addressed	- Does not mention engagement with civil society, academia, or subject matter experts in ways that increase understanding of deceptive AI election	- Not addressed
2. Progress update:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Evaluation:	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met
reasoning:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
3. Brennan Email:	- Implements content credentials through the application programming interface - C2PA standards	- red teaming models internally and externally	- Prompt text classifiers to detect prohibited text in API applications - Image classifiers to detect public figures uploaded to models - Content moderation team identifies election misinformation prompts and takes action - New product integrity team working on monitoring, detecting, and removing - Launched content moderation program as well as automation	- Updated Acceptable use policy to prohibit disinformation - Removes violators from API	- Member of CAI	- Stable Safety public resource about Stabilities policies	- Attend NCMEC roundtable to combat CSAM - Partnership with Internet Watch Foundation - Tech Coalition program to engage expert advice -Contributions to CISA and JCDC - Joined Thorn and All Tech Is Human to enact child safety commitments for Gen AI through Safety by Design	- Not addressed
Evaluation	Partially met	Partially met	Commitment met	Partially met	Partially met	Partially met	Partially met	Commitment not met
reasoning:	- Does not mention identification or non-application programming interface generation use cases or continued advancement	- Does not discuss risk evaluations in detail		- Does not mention enforcement for downloadable products		- Does not discuss deceptive AI election content removals	- Does not mention deceptive AI election content-related partnerships	- Not addressed
Composite Score:	Partially met	Partially met	Commitment met	Partially met	Partially met	Partially met	Partially met	Commitment not met

22. TikTok

1. Warner letter:	- First video platform to implement C2PA ingesting - In coming months will start attaching C2PA metadata to AIGC made on TikTok	- "TikTok continues to develop and test models to assist in the detection and moderation of AIGC uploaded to the platform."	- Detection technology from C2PA - Ingesting C2PA - required creator self-labeling for AIGC - Global fact-checking network working with 18 different organizations that inform moderation decisions - Community partner program has special reporting channels - General user reporting in-app	- Auto-labeling from C2PA technology - work with fact-checking organizations - select organizations have heightened reporting features - platform does not allow political advertising - Prohibit misleading AIGC about public figures - community guidelines prevent political false representation - Remove content that violates AI policies - Provide Election Centers contextual information in partnership with Democracy Works	- Member of CAI - Member of PAI, used as a case study for what a labeling implementation looks like - Partner with stopNCII.org on hashes	- Transparency center sharing latest efforts related to GAIC and DAIEC -harmful misinformation guide on Safety Center that describes popular tactics	- Release research tools that allow third-parties to use platform data to complete research - partners with election commissions and nonprofits to provide trustworthy voting information	- Will release 12 videos with MediaWise to increase awareness of AIGC labeling - Partner with WITNESS to increase user knowledge and context on AI labeling with a series of videos - localized media literacy campaigns by region to raise awareness of AIGC -harmful misinformation guide on Safety Center that describes popular tactics
Evaluation	Commitment met	Partially met	Commitment met	Commitment met	Commitment met	Commitment met	Partially met	Commitment met
reasoning:		- Does not discuss specific actions					- Does not address specific actions with academia, civil society organizations, or subject matter	
2. Progress update:	- First video platform to implement C2PA ingesting - Started attaching C2PA metadata to AIGC made on TikTok - Automatically disclose TikTok effects	- Not addressed	- Ingesting C2PA standards	- Tightened policies of harmful content - prohibit certain political misrepresentations	- Not addressed	- Transparency center - US Elections Integrity hub - Community Guidelines Enforcement reports	- US Elections Integrity Advisory Group SME engagement	- Harmful misinformation guide in TikTok Safety Center - A number of public content literacy initiatives through MediaWise and WITNESS - "supporting" AI Literacy initiative with National Association for Media Literacy Education
Evaluation	Commitment met	Commitment not met	Partially met	Commitment met	Commitment not met	Commitment met	Partially met	Commitment met
reasoning:		- Not addressed	- Does not mention detection methods or content moderation methods		- Not addressed		- Unclear if this is internal or external experts - Does not mention civil society or academia	

3. Brennan Email:	- First video platform to implement C2PA ingesting - Started attaching C2PA metadata to AIGC made on TikTok - Automatically disclose TikTok effects	- Not addressed	- Ingesting C2PA standards - Required creator self-labeling -"This investment in detection and labeling has paid off – of the videos we removed for violating our Integrity & Authenticity policies in Q3 2024, over 97% were removed proactively, meaning we identified and removed the video before it was reported."	- Safety and Content Advisory Council SME engagement to tighten relevant policies related to AIGC - Prohibit misleading AIGC about public figures - maintained policies against misleading or manipulated content - updated our Election Center, search banners, and content labels to make election results from the Associated Press	- Not addressed	- Share removals of AI-generated content that violates our policies in the US Election Integrity Hub in Transparency Center - Share data on our content moderation efforts in Community Guidelines Enforcement reports.	- Safety and Content Advisory Council SME engagement to tighten relevant policies related to AIGC - Elections Integrity Advisory Group composed of experts in AI	- Harmful Misinformation Guide in Safety Center - AI Literacy initiative from the National Association for Media Literacy Education - new media literacy videos with guidance from experts that accumulated over 80M views - Election Center has received more than 60 million views since we launched it in January 2024.
Evaluation	Commitment met	Commitment not met	Commitment met	Commitment met	Commitment not met	Commitment met	Partially met	Commitment met
reasoning:		- Not addressed			- Not addressed		- Does not mention civil society collaborations that are focused on deceptive AI election content risk beyond public awareness	
Composite Score:	Commitment met	Partially met	Commitment met	Commitment met	Commitment met	Commitment met	Partially met	Commitment met

23. Trend Micro

1. Warner letter:	N/A	N/A	- Trend is soon releasing a deepfake detection tool	N/A	-:"potentially open to participation in information sharing programs concerning machine generated false and/or malicious content."	N/A	- "Due to our cybersecurity specific mission, we have not created resources specifically for independent media and civil society organizations."	- Presents blogs about threat actors using AI
Evaluation	N/A	N/A	Partially met	N/A	Commitment not met	N/A	Commitment not met	Partially met
reasoning:	- No generative content or content distribution	- No generative content or content distribution	- Tool was not yet released	- No generative content or distribution	- No evidence of existing information sharing efforts	N/A		- Does not provide detail around deceptive AI election content-specific initiatives
2. Progress update:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Evaluation	N/A	N/A	Commitment not met	N/A	Commitment not met	N/A	Commitment not met	Commitment not met
reasoning:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
3. Brennan Email:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Evaluation	N/A	N/A	Commitment not met	N/A	Commitment not met	N/A	Commitment not met	Commitment not met
reasoning:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Composite Score:	N/A	N/A	Partially met	N/A	Commitment not met	N/A	Commitment not met	Partially met

24. True Media

1. Warner letter:	- They are a content ID platform, so they have their own AI that does content ID. - Have tools specifically designed to detect disinformation	N/A	- Run 10 different AI models to determine content ID	N/A	- Works with tech companies	- Not addressed	- Work with candidates elected officials - Tools are offered for free to journalists, fact-checkers, and other relevant demographics - Integrate AI tech from partners in academia and industry	- Run awareness campaigns - speak publicly about these topics
Evaluation	Partially met	N/A	Partially met	N/A	Partially met	Commitment not met	Commitment met	Commitment met
reasoning:	- Does not mention C2PA or metadata standards		- Does not specifically discuss deceptive AI election content		- Does not include specifics about working with other companies	Does not address any kind of transparency reporting		
2. Progress update:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Evaluation	Commitment not met	N/A	Commitment not met	N/A	Commitment not met	Commitment not met	Commitment not met	Commitment not met
reasoning:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
3. Brennan Email:	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Evaluation	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

reasoning:	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Composite Score:	Partially met	N/A	Partially met	N/A	Partially met	Commitment not met	Commitment met	Commitment met

25. Truepic

1. Warner letter:	- Truepic enables C2PA adoption, tamper-evident -100 million images and videos created with Truepic's technology - "powers" OpenAI's C2PA implementation in DALLE 3. - Truepic Vision uses provenance tech to do virtual inspections - Member of C2PA - Have implemented C2PA on multiple fronts - All content created with Truepic technology implements C2PA - Supports social media companies with ingesting C2PA metadata	N/A	- Working with social media companies on C2PA ingestion, offering ingestion tools	N/A	- Part of Partnership on AI's Responsible Practices Framework for Synthetic Media and the Content Authenticity Initiative - Collaborates with tech companies for provenance initiatives - Partnered with Qualcomm to develop C2PA at chipset level in new Snapdragon chipset - Partner with Microsoft on Content Integrity Tools supporting authenticated mobile capture	- Not addressed	- "regularly engage with civil society organizations in the US and overseas" - Works with newspapers to be a source on C2PA - Work with Project Origin to make a list of C2PA verified news publishers - Works with Ballotopedia on Candidate Connection Program - Works with Microsoft Content integrity Tools - Issued over $500,000 in Social Impact Grant program to 12 partner orgs around the world - Seek to educate government and election officials on C2PA	- Educational newsletter on key industry trends and developments - Educational workshop for Mexican photojournalists ahead of their election - Led and participated in workshops for journalists around the world
Evaluation	Commitment met	N/A	Commitment met	N/A	Commitment met	Commitment not met	Commitment met	Commitment met
reasoning:		N/A		N/A		- Not addressed		
2. Progress update:	-Enables C2PA deployment across various organizations	N/A	- Not addressed	N/A	- Work with Microsoft for Content Integrity tools, working capture tech into workflow for verified media related to elections	- Not addressed	- Works with Ballotpedia to verify 8,000 candidates - Worked with the Carter Center monitoring team during the elections in Venezuela	- Not addressed
Evaluation	Commitment met	N/A	Commitment not met	N/A	Commitment met	Commitment not met	Commitment met	Commitment not met
reasoning:		N/A	- Not addressed	N/A		- Not addressed		- Not addressed
3. Brennan Email:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Evaluation	Commitment not met	N/A	Commitment not met	N/A	Commitment not met	Commitment not met	Commitment not met	Commitment not met
reasoning:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Composite Score:	Commitment met	N/A	Commitment met	N/A	Commitment met	Commitment not met	Commitment met	Commitment met

26. X

1. Warner letter:	- Not addressed	- Not addressed	- Employs safety team to "remain alert to any attempt to manipulate the platform by bad actors and networks" - Special reporting mechanism for impersonation - Direct escalation pathways for election officials	- Policy in place to prevent spam and platform manipulation (Bulk, deceptive, or aggressive activity) - Prohibits synthetic, manipulated, or out-of-context media that could mislead viewers, including deceptively altered media, shared in a deceptive manner, or likely to cause widespread confusion - Action threshold depends on the context - civic integrity policy prohibits using X to interfere with elections or election info - Prohibit misleading or deceptive identities and impersonation - Community notes functionality uses Contributors to crowd-source contextual information and present the highest-voted options - Community notes shown9 billion times in 2024	- Not addressed	- Ranking algorithm for community notes can be inspected by anyone - Offer external researcher access for evaluating community notes	- Work with election officials on education sessions, direct reporting pathways, and tabletop exercises - In contact with European Commission about X's efforts to protect the EU elections - Offered NGO partners refresher training on X's safety tools	- Employs community notes feature and encourages diverse perspectives
Evaluation	Commitment not met	Commitment not met	Partially met	Commitment met	Commitment not met	Partially met	Partially met	Partially met
reasoning:	- Not addressed	- Not addressed	- Does not mention detection technologies or ingestion		- Not addressed	- Does not address deceptive AI election content specifically - Does not address removals or provenance research	- Does not address collaboration with civil society or academia	- Does not discuss education campaigns or other initiatives

2. Progress update:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Evaluation	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met
reasoning:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
3. Brennan Email:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Evaluation	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met
reasoning:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Composite Score:	Commitment not met	Commitment not met	Partially met	Commitment met	Commitment not met	Partially met	Partially met	Partially met

27. **Zefr**

1. Warner letter:	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Evaluation	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
reasoning:	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
2. Progress update:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Evaluation	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met
reasoning:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
3. Brennan Email:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Evaluation	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met
reasoning:	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission	- No submission
Composite Score:	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met	Commitment not met