The Deepfake Detection Challenge: Insights and Recommendations for AI and Media Integrity

March 12, 2020

PARTNERSHIP ON AI

Contents

Executive Summary

In 2019, The Partnership on AI (PAI) mobilized its diverse Partner community towards addressing AI's impact on media integrity and public discourse - a timely topic that requires coordinated, collective attention. In this document, PAI shares insights from our work with the AI and Media Integrity Steering Committee and its guidance of the Deepfake Detection Challenge (DFDC).

We present six key insights and associated recommendations that can inform future work on synthetic media[1] detection, many of which extend to AI and its impact on media integrity more broadly. In doing so, we also document PAI's involvement with the DFDC and share our learnings for conducting meaningful multistakeholder work for AI's development.

These insights and recommendations highlight the importance of coordination and collaboration among actors in the information ecosystem. Journalists, fact-checkers, policymakers, civil society organizations, and others outside of the largest technology companies who are dealing with the potential malicious use of synthetic media globally need increased access to useful technical detection tools and other resources for evaluating content. At the same time, these tools and resources need to be inaccessible to adversaries working to generate malicious synthetic content that evades detection. Overall, detection models and tools must be grounded in the real-world dynamics of synthetic media detection and an informed understanding of their impact and usefulness.

1 *Note:* Synthetic media includes any media (e.g., audio, image, text) that has been AI-generated or synthesized. Deepfakes are a specific type of synthetic media: AI-generated videos. We use the terms somewhat interchangeably over the course of the document, since many of the insights and recommendations below extend beyond audio-visual content and can apply to synthetic content more generally.

Insights and Recommendations

INSIGHT 1:

Detecting synthetic media alone does not solve information integrity challenges.

Recommendation:

- Ensure that solutions for detecting synthetic media are integrated with solutions for countering non-AI generated, low-tech manipulations (e.g., shallowfakes, mis-contextualized content), and consider techniques for confirming content authenticity (not just inauthenticity).

INSIGHT 2:

The ability to detect synthetic media should extend to journalists, fact-checkers, and those working in civil society organizations around the world.

Recommendation:

- Design access and licensing agreements that allow models to be shared outside technology platforms, in interfaces that enable global stakeholders to meaningfully engage with the models and their outputs.

INSIGHT 3:

It is vital to think about the real-world incidence and dynamics of synthetic media when constructing the datasets and scoring rules for synthetic media detection challenges.

Recommendation:

- Datasets and scoring rules for future detection challenges should reflect real-world trends and instances of synthetic media in order to inspire the most robust technical detection tools.

INSIGHT 4:

Synthetic media detection solutions involve trade-offs between open access to models and the need to deter adversaries generating synthetic media.

Recommendation:

- Consider possible malicious uses when devising and evaluating licensing structures for synthetic media detection models; licenses should enable adversarial deterrence while ensuring access for pro-social uses.
- Invest in technical experimentation and "red-teaming" or ethical hacking/intentional pressure testing of winning models.
- Promote more frequent, possibly ongoing, detection challenges that attend to the fact that deepfake detection is a constantly evolving and adversarial exercise.

INSIGHT 5:

The results of synthetic media detection need to be understandable to those making sense of digital content.

Recommendation:

- Explore options for embedding explanations research in (or adjacent to) detection challenges.
- Promote research on trusted explanations that meaningfully convey media manipulations and inauthentic content to the public.

INSIGHT 6:

Synthetic media detection challenges benefit from meaningful multistakeholder input, which takes substantial time and energy.

Recommendation:

- Leverage global multistakeholder input when developing synthetic media detection challenges, and ensure you have enough time to do so meaningfully.

About PAI's AI & Media Integrity Steering Committee and the Deepfake Detection Challenge

PAI created the AI and Media Integrity Steering Committee in late 2019 as a formal body of Partners for developing and advising projects that strengthen mis/disinformation solutions, including detection of manipulated and synthetic content.

The Steering Committee builds upon PAI's work in AI and Media Integrity throughout 2019, including convenings with representatives from media entities, civil society, and technology companies. In particular, it draws on key themes identified at a Media Preparedness Workshop hosted by PAI in collaboration with WITNESS and the BBC in May 2019. Workshop participants surfaced key aspects of AI-generated mis/disinformation and synthetic media challenges that are ripe for coordinated attention - content authentication and provenance, technical detection of synthetic and manipulated media, mechanisms for communicating inauthentic and synthetic content to audiences, and broader tactics for coordinating and sharing amongst actors working in the information ecosystem.

PAI's AI and Media Integrity Steering Committee was convened to enable such coordinated action. It serves as a venue for organizations to present initiatives and technical projects related to AI and media integrity that have social and political implications and a need for broader, community-wide contributions.

The AI and Media Integrity Steering Committee consists of formal representatives from nine PAI Partner organizations spanning civil society, media, and technology - Amazon, the BBC, CBC/Radio-Canada, Facebook, First Draft, Microsoft, The New York Times, WITNESS, and XPRIZE. To ensure a balance of expertise on the committee, ⅓ of organizations were from civil society, another ⅓ from technology companies, and the final ⅓ from media entities. PAI strategically chose a small group of actors, though it is crucial to note that the benefits of working with a small group traded off with opportunities for an increasingly global and diverse cohort.

These representatives reflect a cross-section of PAI Partners with varied backgrounds, roles, and expertise in information integrity challenges and solutions. While some members of the group have deep, technical expertise in machine learning (e.g., a computer vision scientist, a computer security researcher), others work more specifically on the societal implications of the technology - for example, mis/disinformation, human rights, and information quality as it relates to the needs of journalistic entities and newsrooms. This diverse group meets weekly for hour-long, virtual Steering Committee meetings that are organized, scoped, and facilitated by PAI staff.

The AI and Media Integrity Steering Committee's first project was the Deepfake Detection Challenge (DFDC), a machine learning challenge designed to promote development of technical solutions that detect AI-generated and manipulated videos. The challenge is rooted in the concern that AI-generated videos, or deepfakes, have the potential to significantly impact the legitimacy of online information.[2]

2 https://ai.facebook.com/blog/deepfake-detection-challenge/

Launched initially by Facebook, and developed in collaboration with Microsoft, Amazon Web Services[3], PAI, and a group of academics, the DFDC is a machine learning competition, hosted on Kaggle[4], and created to incentivize the development of technical solutions that detect AI-generated videos. The winners of the challenge will be those researchers who develop state-of-the-art models that best predict whether or not videos have been manipulated with AI-techniques.

In order to mobilize the research community towards technical synthetic media detection, Facebook created a unique dataset of over 100,000 videos using paid actors in different settings, poses, and backgrounds. They used many techniques to generate both face swaps and voice alterations from the original videos. For a subset of the videos, Facebook applied augmentations that approximate actual degradations seen in real-life videos shared online. See the Facebook AI team's paper and challenge launch blog for more details on dataset construction and challenge design.

Different organizations could use deepfake detection technology in different ways. Platforms like Facebook could use deepfake detection technology to scan uploaded content at scale and automatically flag synthetic videos for human review. Journalists and those working in civil society can use detection results on specific videos as a signal in media verification work. It is important to note, however, that simply identifying manipulated content does not increase the perceived legitimacy of unmanipulated media or establish truth. Technical detection is one important aspect of addressing information integrity challenges, but it is not a panacea.

While Facebook could have developed the DFDC completely independently, they turned to the multistakeholder AI and Media Integrity Steering Committee, as well as formal partnerships with leading academics working on technical detection, for advice on certain challenge elements. This collaborative gesture reflects the recognition that questions around the technical detection of synthetic media must be situated within the complex dynamics of how information is generated, spread, and weaponized in the 21st century and therefore requires cross-sector and multidisciplinary collaboration. While a machine learning challenge like the DFDC and others hosted on Kaggle may seem to be a purely technical endeavor, ensuring that the challenge is valuable as one of many solutions for furthering information integrity requires the input of a spectrum of actors.

The Steering Committee's work on the DFDC took place over a series of ten meetings, from October 2019 until the launch of the challenge at the Neural Information Processing Systems Conference (NeurIPS) in December 2019. To supplement these official meetings, PAI staff worked to ensure that all participants had a shared background in both the mechanics of a machine learning challenge and the dynamics of mis/disinformation in order to meaningfully contribute to complex, weekly discussions about specific DFDC challenge features.

Weekly meetings were typically scoped to one or more DFDC decision-points, anchored on a pre-read document prepared by PAI staff. While we envisioned that weekly meetings would resolve with a vote, they typically ended with follow ups for PAI staff, in consultation with Facebook staff driving the day to day operations of the challenge, to consider how to meaningfully operationalize the input derived from the committee at a

3 Note: Amazon, Facebook, and Microsoft were on the Steering Committee. Facebook recused itself from voting in the committee. The extent of Amazon and Microsoft's formal partnership did not warrant their recusal from voting on the committee.

4 Note: An Alphabet company.

time frame and point of entry for feasibly shaping the DFDC. The committee's input into the DFDC, was honed in collaboration with Facebook, PAI staff, and other Steering Committee members in the time between meetings.

The AI and Media Integrity Steering Committee was initially tasked with contributing to challenge elements including:

- Eligibility criteria to enter the challenge
- Terms for how participants' entries are shared and distributed (open source, implementation rights)
- Scoring and judging criteria (including potential subjective scoring criteria)
- Award distribution process

Meetings were not distributed evenly across these challenge elements and based on the complexity and interconnectedness of specific topic areas, PAI adapted the distribution of meeting agendas from the original plan. Meetings focused on challenge scoring and judging criteria, as well as the terms for how participants' entries are shared and distributed.

Importantly, PAI got involved in and inherited elements of DFDC guidance at a late stage in the project's development, once the videos in the dataset had already been developed and filmed, certain amounts of the announced 10 million dollar budget allocated[5], and the project timeline from Facebook already prescribed. This restricted the extent to which PAI could offer any design input that would have required implementation in the dataset construction phase. It also made the time frame for sourcing meaningful multistakeholder input too tight for ensuring implementation of all Steering Committee recommendations. While these constraints were an impediment to PAI's ability to impact the DFDC, cross-sector collaboration for the DFDC design process elicited many best practices and next steps for future efforts to protect the integrity of digital media.

The AI and Media Steering Committee's involvement in the DFDC challenge serves as a case study in effective cross-sector, multistakeholder contribution to AI development. This multistakeholder process is one of the first examples of a model for engaging diverse, cross-sectoral stakeholders around the construction and launch of a technical challenge. Future challenges should draw on our insights to evolve the model; doing so will help ensure that machine learning challenges that spur technical innovation are rooted in their global social and political realities.

5 https://ai.facebook.com/blog/deepfake-detection-challenge/

Insights and Recommendations

Insight 1: Detecting synthetic media alone does not solve mis/disinformation challenges.

Merely developing AI models to detect deepfakes does not protect against audio-visual mis/disinformation. First of all, AI-generated videos are only one example of manipulated media and the majority of image manipulations used today involve low-tech, non-AI based tactics for manipulation. For example, the May 2019 video of Nancy Pelosi appearing drunk is a shallowfake, or low-tech manipulation, since the video was simply slowed down. In fact, many pieces of video-based mis/disinformation merely feature edited or mis-contextualized content, like an authentic video with an inaccurate caption, rather than image or audio manipulations of the video itself.

Second, once deepfake videos are detected, it is important to understand the meaning of that detection signal. Technical detection conveys the way in which a video is generated or manipulated, but does not provide other details about the specific content, let alone whether or not the content is mis/disinformation. While a technology platform might want to eliminate synthetic videos that are misleading or deceptive, they might not have any issue with AI-generated satire or value-neutral videos on their platform. Additionally, the nature of the content does not necessarily indicate the creator's intent (e.g., to deceive or inform).

The value of technical detection for understanding the nature of produced and disseminated media needs to be understood within the context of these broader mis/disinformation dynamics. Despite the broader dynamics of mis/disinformation, the committee still recognized the value of technical detection as part of an integrated strategy for addressing mis/disinformation. Technical detection is particularly attractive as a triage tool, in contexts where only a small percentage of videos can undergo human review. This is particularly important as techniques for generating synthetic media become increasingly accessible to a large spectrum of actors, including actors making use of such tools for malicious purposes.

Recommendation:

Ensure that solutions for detecting synthetic media are integrated with solutions for countering non-AI generated, low-tech manipulations (e.g., shallowfakes, mis-contextualized content), and consider techniques for confirming content authenticity (not just inauthenticity).

Insight 2: The ability to detect synthetic media should extend to journalists, fact-checkers, and those working in civil society organizations around the world.

The DFDC's original intent was to produce models that can accurately detect when face swapping has been used to alter videos. The winning submissions would essentially be pieces of code, which might be suitable for integration by platforms like Facebook, but would not on their own serve other needs. The committee voted unanimously to advocate for the creation of tools for journalists, fact-checkers, and others working in civil society to make the winning models useful to other actors in the broader information ecosystem. PAI plans to undertake work that furthers creation of such tools in 2020.

Although platforms have a major role to play in stopping harmful synthetic content from spreading, these tools should also be accessible to those in the broader media ecosystem working to uphold media integrity. This includes many organizations that do not have the resources and technical capacity to do deepfake detection. PAI's Media and Integrity Steering Committee has begun exploring the possibility of producing a tool or app that journalists, fact-checkers, and those working in civil society can use to identify suspect videos, and working with others in the community to think about how we can collaboratively produce more accessible tools and media forensics resources. Such a process should be incorporated into existing video verification and mis/disinformation workflows already used by the journalistic and fact-checking community and developed in deep consultation with these communities. This approach would ensure that the best models produced from the technical challenge are useful beyond the largest technology platforms.

Early discussions about such a tool emphasize the need for: 1) extending access to tools beyond the largest media entities to smaller, more local and community-based media, including those in the Global South, (2) designing an interface that makes it intuitive and useful for the journalistic, fact-checking, and civil society community to interpret the results, and (3) ensuring appropriate restrictions on access to prevent adversaries or mis/disinformation generators from gaining access to such a tool. The group highlighted the tension between a desire to increase access to synthetic media detection tools, and constraints on access so that malicious actors do not make use of such tools to strengthen their content generation. This tension is explored further in Insight 4.

Recommendation:

Design access and licensing agreements that allow models to be shared outside technology platforms, in interfaces that enable global stakeholders to meaningfully engage with the models and their output.

Insight 3: It is vital to think about the real-world incidence and dynamics of synthetic media when constructing the datasets and scoring rules for synthetic media detection challenges.

Building on insights first generated at the Media Preparedness Workshop hosted by PAI, the BBC, and WITNESS in May 2019, the Steering Committee emphasized the importance of considering the real-world impact of synthetic media when developing technical tools for its detection. Ideally, the demographic distribution of contest data would reflect what we see in synthetic videos in the wild; the same can be said of other aspects of the content in those videos.

While the dataset Facebook created includes individuals in domestic environments, future datasets should visually reflect the spectrum of content -- situations and environments -- typically seen in the most frequently disseminated and, ideally, the most misleading and harmful content. Of course, robust datasets of naturally occurring, synthetic content are not publicly available, though some have written on the types of deepfakes that are most frequently created and spread on the internet (the vast majority of deepfakes on the web today feature pornographic content and non-consensual sexual imagery).[6]

The group also examined how the process of scoring could better incentivize the creation of winning models that reflect real-world deepfake phenomena. The choice of a scoring rules is one of the central decisions in designing a machine learning competition, and the Steering Committee explored scoring rules that:

1. Encourage accurate detection, but discourage over-fitting and overconfidence
2. Take into account the enormously different frequencies of genuine and synthetic media in any real-world setting where a detector would be deployed
3. Discourage algorithmic bias (i.e., different detection rates in different subpopulations) with regard to either subject class, or potentially, types of deepfake generation
4. Are unambiguous, straightforward to implement, and difficult to game
5. Encourage disparate types of detection algorithms that are collectively more difficult to circumvent than any individual detector

The Steering Committee advocated for a scoring rule balancing these requirements. The scoring rule Facebook originally proposed assumed a very low incidence of deepfakes and thus was designed to heavily penalize false positives (erroneously indicating that a video is fake.) This is appropriate for automated use by platforms, but other actors are likely to use detection tools to gain additional information about videos which are already suspect, due to their source, content, timing, or other signals. PAI staff proposed a simple scoring rule designed to incentivize models that can be adapted to contexts with widely differing incidences of manipulation, and this was adopted by the Steering Committee and reflected in the final challenge scoring. This is a concrete example of multistakeholder input shaping detailed technical choices. As the

6 https://blog.witness.org/2018/07/deepfakes/

adversarial dynamics and use cases of deepfake detection continue to develop, the consensus scoring recommendations may shift over time.

Recommendation:

Datasets and scoring rules for future detection challenges should reflect real-world trends and instances of synthetic media in order to inspire the most robust technical detection tools.

Insight 4: Synthetic media detection solutions involve trade-offs between open access to models and the need to deter adversaries generating synthetic media.

The Steering Committee advocated for increased access to synthetic media detection models and tools. However, this call for increased access was often in tension with the recognized adversarial dynamics associated with deepfake detection.

There is a precedent of openness in the computer science community associated with freely sharing code.[7] However, there has recently been increasing introspection on this default to openness, as some believe that openness around certain AI developments can have negative consequences due to malicious use by those able to access the code. As an inherently adversarial challenge, deepfake detection is only able to provide value to platforms, journalists, policymakers, and fact-checkers if those generating malicious deepfakes do not learn how to evade detection methods, rendering detection useless in the fight against mis/disinformation. Therefore, some have advocated for responsible disclosure mechanisms[8] that limit the release of synthetic media detection technology, in order to ensure that the detection capabilities of those working to prevent mis/disinformation outpace the tactics deployed by those generating malicious, synthetic content.

The Steering Committee explored options for releasing and licensing the winning code from the DFDC that would attend to concerns about adversarial uses, while allowing enough openness to advance deepfake detection research and facilitate the creation of useful tools for journalists, fact-checkers, policymakers, and individuals working in civil society.

PAI staff suggested a hybrid model for openness that encouraged publication of papers and other materials discussing the chosen methods, while delaying the open release of code and models for a certain time, perhaps six months to a year. Meanwhile, trusted parties would be granted immediate non-open source licenses to use the technology for their own work. This approach would deliver the eventual benefits of openness and increased access, while the lag would provide some detector advantage over those trying to generate evasive, adversarial synthetic content.

7 https://www.partnershiponai.org/when-is-it-appropriate-to-publish-high-stakes-ai-research/

8 https://www.partnershiponai.org/case-study/publication-norms/

Experimenting with and evolving the traditional technical challenge structure may also help address these inherent adversarial dynamics. Steering Committee members suggested an ongoing challenge, with winners declared at various milestones. This would help respond to the reality of a constantly moving target for successful deepfake detection as malicious actors work to evade detection tactics. Additionally, an ongoing program of friendly adversarial "red-teaming" of detection models would provide critical information on important weaknesses and current detector advantage.

Recommendations:

Consider possible malicious uses when devising and evaluating licensing structures for synthetic media detection models; licenses should enable adversarial deterrence while ensuring access for pro-social uses.

Invest in technical experimentation and "red-teaming" or ethical hacking/intentional pressure testing of winning models.

Promote more frequent, possibly ongoing, detection challenges that attend to the fact that deepfake detection is a constantly evolving and adversarial exercise.

Insight 5: The results of synthetic media detection need to be understandable to those making sense of digital content.

Steering Committee conversations reinforced the need for research on explanations of inauthenticity and synthetic content that initially emerged from the Media Preparedness Workshop. This is especially important if the output of detection models is intended to be used by those who are not technical experts. Journalists, fact-checkers, policymakers, and the general public all require comprehensible and useful detection explanations.

The group explored whether or not explainability should be included as part of the challenge criteria, including whether or not they should reallocate prize money to a separate event for building UX/UI around the winning models. One potential example proposed labeling manipulated regions of the video with bounding boxes. However, the DFDC does not present a sufficient variety of manipulation types to evaluate these systems (only faces), so systems explaining the results of detector models should be the subject of research and development following the competition.

The Steering Committee also considered explanation methods that are more difficult to evaluate quantitatively. Qualitative approaches to explaining the results of deepfake detection must consider how audiences making sense of audio-visual content online interpret signals of inauthenticity. PAI and First Draft[9] have just begun work on how to best publicly explain and label new forms of audio-visual manipulations that are increasingly difficult to detect with the naked eye, with hopes of sharing insights by the last few months of 2020. It is important that the ways in which journalists and fact-checkers label audio-visual manipulations are consistent with effective methods for labeling content for platform audiences.

9 https://www.partnershiponai.org/media-manipulation-research-fellowship/

Recommendations:

Explore options for embedding explanations research in (or adjacent to) detection challenges.
Promote research into trusted explanations that meaningfully convey media manipulations and inauthentic content to the public.

Insight 6: Synthetic media detection challenges benefit from meaningful multistakeholder input, which takes substantial time and energy.

Future efforts should ensure that multistakeholder input is sought from project inception, not after significant work has already begun. The Steering Committee was constrained in its ability to impact DFDC components which had already been developed. This was particularly true in the case of dataset construction and licensing agreements. The selection of dataset cases greatly affects the applicability of detection models in different contexts and on different types of videos; it also has implications for fairness and bias outcomes. Unfortunately, dataset creation was mostly complete by the time the Steering Committee was involved in the DFDC. The exploration of a hybrid open-source license that involves delayed release was also constrained by the tight timeline, as the successful deployment of such models would require substantial multistakeholder legal consultation to draft novel contest rules and software licenses. As a result, the Steering Committee was not able to meaningfully implement these particular changes.

While the ambitious DFDC timeline allowed for only about 10 weeks to successfully learn about, interrogate, agree, and operationalize decision-points for the challenge design and structure, future efforts should likely allow 15-20 weeks of collective multistakeholder engagement on challenge design. This projection assumes that, like in the case of the DFDC, there is a full time staff working on the effort and many group members had already worked on information integrity issues together. Future efforts around synthetic media detection can also be strengthened by sourcing collective input from the broader media ecosystem and a larger, more global cohort.

Recommendation:

Leverage global multistakeholder input when developing synthetic media detection challenges, and ensure you have time to do so meaningfully.

Reflections on Multistakeholder Collaboration

The AI and Media Steering Committee serves as a case study in effective cross-sector, multistakeholder contribution to AI development. This process is one of the first examples of a model for engaging diverse, cross-sector stakeholders around the construction and launch of a technical challenge. Future challenges should draw on our insights to evolve the model in order to help ensure that machine learning challenges that spur technical innovation are rooted in socio-political realities, and informed by experts from across the media and information ecosystem.

Key elements that enabled the Steering Committee to bring together a diverse group of actors around a complex socio-technical challenge to generate actionable input included:

1. Strong, trusting relationships with individuals and organizations

   The majority of Steering Committee members had been working with each other on media integrity for over 6 months through participation in a PAI Working Group. Several committee members also participated in the May 2019 Media Preparedness Workshop, and many were able to interface and connect with PAI staff and other Partners at the PAI All Partner's Meeting in September 2019, directly preceding formal launch of the Steering Committee. These meaningful exchanges ensured that trust was built between Partners and PAI staff. This foundation ensured that the proper tone was set for complex discussion and debate.

2. Pre-reads and clearly scoped meeting goals

   Each week, PAI staff spent a significant amount of time (~6 hours) preparing pre-read documents that were clear, concise, and scoped narrowly. We also provided time to read these documents in the meetings themselves, so that busy Steering Committee members could be briefed. These documents, alongside detailed agendas, ensured that participants had the context and understanding of meeting goals to participate meaningfully each week.

3. Time devoted to goal alignment

   Each individual/organization came to the committee with a different vision for what it means to shape a technical challenge around synthetic media. Some wanted to explicitly drive towards the creation of models for technical detection, whereas others wanted to extend that goal to also include connections to other deliverables beyond the technical challenge. Spending time scoping and coming to clearer agreement about projects goals and deliverables at the beginning of the process was vital to its success.

Conclusion

Projects like the DFDC demonstrate how the technology industry can mobilize collective attention and cross-sectoral involvement on issues of media integrity, and that situating these efforts within the broader global community is integral to upholding the quality of public discourse. Future efforts and conversations around synthetic media detection should seek to include communities that are most vulnerable to mis/disinformation, as well as actors best situated to confront mis/disinformation challenges at the local and national level.

PAI will continue to iterate on initiatives that touch on synthetic media detection while also attending to the other aspects of AI and media integrity that warrant collective attention and multistakeholder engagement. In 2020, we seek to ensure more coordinated, diverse governance of technical tools and systems built for and around AI-generated mis/disinformation, and to increase access to such tools and knowledge. The complex socio-technical challenges around synthetic media detection, and AI's connections to media integrity challenges and potential benefits in general, require the type of multistakeholder input that PAI's AI and Media Integrity Steering Committee brought to the DFDC.

Acknowledgements

PAI is deeply grateful to the AI and Media Integrity Steering Committee. Their comments, insights, and supplemental information enriched the paper, and we appreciate the time and expertise they shared with us.

About the Partnership on AI

The Partnership on AI (PAI) is a global multi-stakeholder nonprofit committed to the creation and dissemination of best practices in artificial intelligence through the diversity of its Partners. By gathering the leading companies, organizations, and people differently affected by artificial intelligence, PAI establishes a common ground between entities which otherwise may not have cause to work together – and in so doing – serves as a uniting force for good in the AI ecosystem. Today, PAI convenes more than 100 partner organizations from around the world to realize the promise of artificial intelligence. Find more information about PAI at partnershiponai.org.

PARTNERSHIP ON AI

PartnershiponAI.org
115 Sansome Street, Suite 1200
San Francisco, CA 94104