Google

2021 AI Principles Progress Update

Table of contents

Overview

AI continues to help people around the world with everyday tasks, like real-time translation, driving directions, and optimized video lighting so everyone is visible in virtual meetings. AI can also be used to address some of the world's most challenging problems. It is already helping pathologists grade prostate cancer more quickly and consistently, optimizing the efficiency of traffic lights, predicting floods and other impacts of climate change, and informing COVID-19 policy decisions and drug discovery.

Google's approach to AI is a result of more than two decades of responsible innovation at the company. We began machine learning (ML) work in 2001, deep learning research in 2011, user research on algorithmic fairness in 2014, and a cross-functional ML Fairness initiative in 2016, supporting a number of projects in ML fairness, explainability, privacy, and safety. In 2017 we began developing an ethical charter to guide the development and use of AI, resulting in publication of the Google AI Principles in 2018. Since then we have been working to improve implementation, operationalization, and governance of the AI Principles at Google.

The fact that AI carries risks, alongside enormous benefits, is increasingly recognized by the private sector as well as policymakers and civil society. AI principles, governance proposals, and regulations have proliferated around the world, with a growing focus on issues of fairness, safety and privacy of AI systems. The OECD, G20, and G7 have adopted AI principles, and AI governance standards are being developed by IEEE, ISO, and national standards bodies such as NIST in the US. In April 2021, the EU released the first horizontal AI regulatory proposal, the AI Act, which outlines detailed requirements and a regulatory structure to manage high-risk AI systems. Korea's Basic Act on Intelligent Information Society, which includes protections against potential risks of AI systems, went into effect in December 2020, and in September 2021, Brazil's House of Representatives introduced draft legislation on AI governance. Other AI governance proposals are being developed in many countries, including India, Israel, Colombia, China, and the UK.

Within Google, we've built a three-tiered governance program underpinned by industry-leading research and a growing library of resources, tools and recommended practices. We've launched new features and options to empower our users with more control of their data, and explanations of how our products work. We've developed policies and frameworks to guide internal teams, and made tough decisions, for example in 2018 affirming not to release general-purpose facial recognition APIs despite potential business opportunities. And we continue to work with governments and international organizations on AI standards and rules, recognizing the enormous importance of AI innovation and experimentation together with the need for regulation.

While we have made a lot of progress, we are committed to continuing to learn and improve. For example, after a prominent researcher left Google in the past year, we updated our research review process to increase consistency and transparency, appointed HR specialists to review certain sensitive employee exits, and as part of our

ongoing racial equity commitments[1] more than doubled the retention team to ensure we are meaningfully addressing experiences in the workplace, including relationships with managers and peers. In order to better coordinate and execute on advancing progress in this space, we also consolidated dozens of research and technical teams working in responsible and ethical AI into the same organization.

There are a number of open technical questions. For example, we are still learning how to improve the development of machine vision systems that work well for a variety of skin tones. Earlier this year we launched our DermAssist tool to help users better understand skin conditions. While initial experiments demonstrate the tool performs well across self-reported ethnicities, more work needs to be done to understand its performance on the darkest and lightest skin tones. This year we also launched Real Tone, which helps our imaging products like the Pixel phone camera and Google Photos more accurately and beautifully represent a diverse range of skin tones. We developed Real Tone in collaboration with photographers, cinematographers, and directors known for their understanding of the challenges in lighting and capturing a spectrum of skin tones, and who also represented communities of people with darker skin tones. We will continue work on alternative and more inclusive measures that could be useful in the development of our products, in collaboration with scientific and medical experts as well as groups working with communities of color.

While we still have a lot to learn—and will continue learning given the dynamic and evolving nature of technology and society—we remain committed to sharing our progress and findings. In this year's report, we outline our progress in AI Principles implementation to date, highlighting advances in internal governance and operations; resources, research, tools, and responsible practices; product impact; and supporting global dialogue, standards, and policy.

Google AI Principles

We will assess AI in view of the following objectives. We believe AI should:

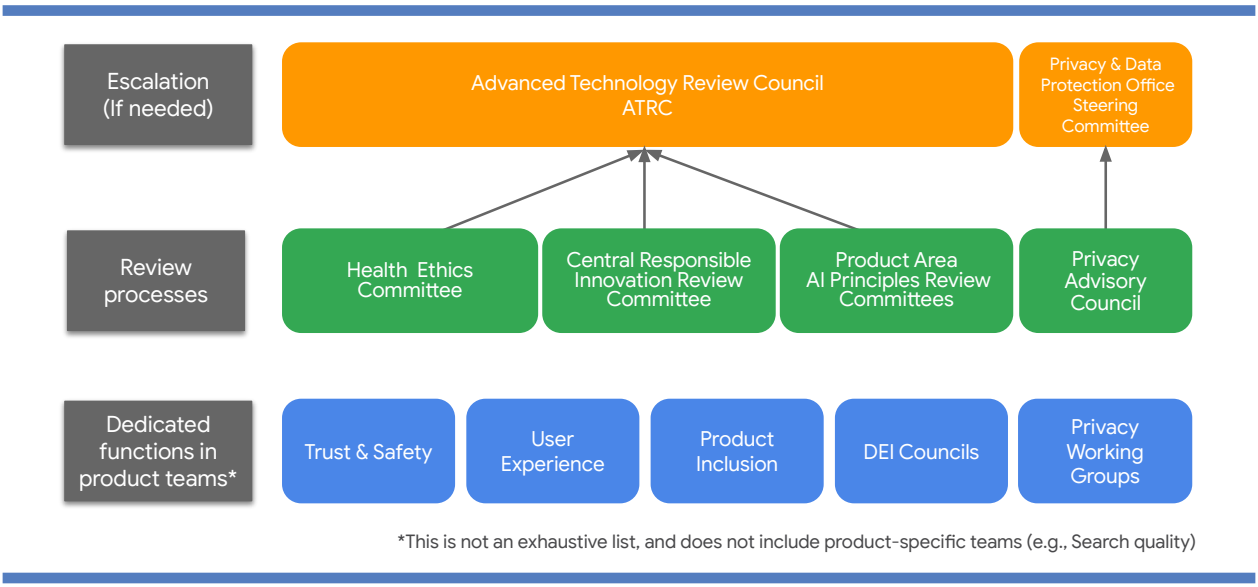
1. **Be socially beneficial:** with the likely benefit to people and society substantially exceeding the foreseeable risks and downsides.
2. **Avoid creating or reinforcing unfair bias:** avoiding unjust impacts on people, particularly those related to sensitive characteristics such as race, ethnicity, gender, nationality, income, sexual orientation, ability and political or religious belief.
3. **Be built and tested for safety:** designed to be appropriately cautious and in accordance with best practices in AI safety research, including testing in constrained environments and monitoring as appropriate.
4. **Be accountable to people:** providing appropriate opportunities for feedback, relevant explanations and appeal, and subject to appropriate human direction and control.
5. **Incorporate privacy design principles:** encouraging architectures with privacy safeguards, and providing appropriate transparency and control over the use of data.
6. **Uphold high standards of scientific excellence:** Technology innovation is rooted in the scientific method and a commitment to open inquiry, intellectual rigor, integrity and collaboration.
7. **Be made available for uses that accord with these principles:** We will work to limit potentially harmful or abusive applications.

In addition to the above objectives, we will not design or deploy AI in the following application areas:

1. Technologies that cause or are likely to cause overall harm. Where there is a material risk of harm, we will proceed only where we believe that the benefits substantially outweigh the risks, and will incorporate appropriate safety constraints.
2. Weapons or other technologies whose principal purpose or implementation is to cause or directly facilitate injury to people.
3. Technologies that gather or use information for surveillance violating internationally accepted norms.
4. Technologies whose purpose contravenes widely accepted principles of international law and human rights.

Internal governance and operations

At the same time as we announced the Google AI Principles in 2018, we founded a central Responsible Innovation team, drawing from existing specialists across the company to operationalize implementation. Starting with half a dozen full time employees, engagement in AI Principles governance has since grown in scale; today, hundreds of Google employees across dozens of teams with a wide range of expertise—human rights, user experience research, ethics, trust and safety, privacy, public policy, machine learning, and more—make up an internal AI Principles ecosystem that supports employees in incorporating responsible practices into their work.



Google operationalizes responsible innovation practices via a three-tiered internal AI Principles Ecosystem.

At the core of this internal ecosystem is a three-tiered governance structure. It starts with our product teams themselves, which include dedicated user experience (UX), privacy, and trust and safety (T&S) experts who provide deep functional expertise consistent with the AI Principles.

The second tier is a set of dedicated review bodies and expert teams. The central Responsible Innovation team is available to support implementation across the company, and all Google employees are encouraged to engage with the AI Principles review process throughout the project development lifecycle. Some product areas have set up review bodies to address specific audiences and needs, such as enterprise offerings in Google Cloud, hardware in Devices and Services, and medical expertise in Health. In addition, the Privacy Advisory Council (PAC) reviews all projects for potential privacy concerns, including (but not exclusively) issues related to AI.

One example of a body in this second tier is the Health Ethics Committee. This is a forum for guidance and decision-making regarding ethical issues arising within the context of health products, health research, or health-related organizational decisions. It was created in 2020 to provide a forum for moral deliberation and decision-making on health-related initiatives to keep Google's users and products safe with expertise across multiple domains. The Health Ethics Committee is a multidisciplinary forum that includes subject matter experts in bioethics, clinical medicine, policy, legal, privacy, compliance, research, and business. In 2021 the Google Bioethics Program created the Health Ethics Cafe, an informal forum to discuss bioethical questions with anyone in the company and at any stage of project development. The most vexing cases from the Health Ethics Cafe are escalated to the Health Ethics Committee for review.

The second tier also includes review committees tailored for specific product areas. This includes Google Cloud's Responsible AI Product and Deal Review Committees, which are designed to ensure Cloud's AI products and projects align with the Google AI Principles in a systematic, repeatable way, and are built with ethics and responsibility by design, across products and geographies. The first committee focuses on the products built by Cloud AI & Industry Solutions. These reviews undertake a comprehensive analysis that includes an evaluation of the sociotechnical landscape, opportunity and harm assessments across each of the AI Principles, and live discussion with a cross-functional and diverse committee, resulting in an actionable alignment plan. The second review committee covers early stage customer engagements leveraging custom AI solutions beyond our generally available products. It is a committee composed of four cross-functional, senior executive members. All decisions require full agreement from all four committee members, and are escalated as needed. Relevant stakeholders across the AI Principles ecosystem at Google help inform these discussions; this input and involvement is critical to ensure decisions are not made in a vacuum.

As one example of how these review processes combine to make and build on decisions, credit risk and worthiness have been noted as an area of concern for algorithmic unfairness at Google since 2017. In 2019, Cloud's Responsible AI Product Review Committee evaluated product opportunities in this space. While our hope is that one day AI can provide access to credit and play a role in increasing financial inclusion and health, the Committee ultimately determined that a creditworthiness product—built with today's technologies and data—could create disparate impact related to gender, race and other marginalized groups, and conflict with Google's AI principle to "avoid creating or reinforcing unfair bias." In mid-2020 the Product Review Committee re-evaluated and reaffirmed this decision. Over the course of last year, Cloud's Responsible AI Deal Review Committee evaluated multiple proposed custom AI engagements related to the assessment of creditworthiness. Each engagement is evaluated for its particular use case, and the Deal Review Committee determined not to pursue many of these. These learnings over multiple years built on each other, leading to a decision to pause development of custom AI solutions related to creditworthiness until risks can be appropriately mitigated. This Cloud-wide policy went into effect last year, and remains in place today.

The third tier of our AI governance structure is the Advanced Technology Review Council (ATRC), a rotating committee of senior product, research, and business executives representing a broad cross section of Google. The ATRC addresses escalations and the most complex and precedent-setting cases, and establishes policies impacting multiple product areas. It has made a number of tough decisions, weighing potential business opportunities against the ethical risks of certain applications. In 2018 this executive council affirmed Google Cloud's Product Review proposal to not offer general-purpose facial recognition APIs before working through important technology and policy questions, and advised the team to focus on narrowly focused solutions. Drawing on this decision, and following a multi-year effort with significant input from internal and external stakeholders, Google Cloud developed, sought approval from the ATRC, and then released a highly constrained and specifically designed Celebrity Recognition API[2]. One topic considered by the ATRC this year was the development of large language models. Following review, the ATRC determined that research involving large language models could continue cautiously, but that no such model should be launched without a full AI Principles review.

Resources, research, tools, and responsible practices

Implementation of the AI Principles is underpinned by a growing body of company-wide input, training and education resources, technical tools, and recommended practices that equip Googlers across various functions and roles to develop and deploy AI responsibly.

We've created more formalized channels for taking into account a wider variety of perspectives throughout the product development lifecycle and in our review processes. Our Responsible AI and Human Centered Technology team, which is growing to 200 experts, is a consolidation of dozens of research and technical teams working in responsible and ethical AI. The team works with hundreds of partners in product, privacy, security, and other teams in the company, providing insights through research, tools, and other technical solutions. Our Product Inclusion team helps product teams tap into Google's Employee Resource Groups to test for equitable product experiences with a range of different users and audiences. We've also created an AI Principles Ethics Fellows program, in which fellows share their perspectives on the responsible development of future technologies and develop hypothetical case studies to inform how Google prioritizes socially beneficial applications. In 2021, the fellows created timely case studies on topics including the responsible development of genome datasets and a COVID-19 content moderation workflow. In late 2021 we complemented this fellowship effort with another internal program focused on scaling product fairness testing across Google. This product fairness (ProFair) program consists of more than 70 trusted and trained AI Principles advisors in 40 offices around the world. Each advisor has received dedicated training on tech ethics and algorithmic fairness, and engaged in projects such as identifying fairness concerns in an image dataset for a Next Billion Users product. We also continue collaborating with human rights experts to conduct Human Rights' Impact Assessments for understanding new and emerging technologies.

Our educational resources include the foundational Technology Ethics Training, which guides Googlers through the philosophy of technology ethics and how to assess potential benefits and harms, and a suite of courses that explain the Google AI Principles and internal governance practices. This year, we launched an AI Principles and responsible innovation training course for new employees in a variety of engineering and UX roles to understand Google's ethical charter and available resources. In the two years since we began internal AI Principles training, these courses have been taken by more than 19,000 Googlers in a variety of roles, including engineers, researchers, product managers, and account representatives. We continue to expand education to more employees.

In 2021 we also launched interactive online puzzles[3] designed to help employees build awareness of, and test their recall of, the AI Principles. More than 6,000 employees have engaged with the puzzles in the six months since launch. We continue to experiment with gamification of our education efforts to improve engagement and retention of key

takeaways. We have also piloted a two-day, live-video immersive "Moral Imagination" workshop for 15 product teams of 6-15 people who are working on a project that may lead to a potential AI product or feature. Through semi-structured conversation and various practical exercises, facilitators make new concepts concrete and actionable for the participating teams by walking them through the ethical implications of their project and coaching them to see their work from various points of view. The goal is to provide a pragmatic and constructive reflection of these often unexamined parts of the role of an engineer, researcher, or product manager. Teams are asked to articulate and interpret key ethics concepts and processes that relate to their project throughout the workshop, helping to instill these learnings. Feedback from the pilots has been positive and we will scale the program next year.

Technical research informs our practices and tooling, and is another avenue for progress. Since 2018 Google has published more than 3300 research papers on a variety of topics, including more than 500 on responsible innovation. This year, publications included the impact of data cascades[4] (downstream effects from poorly managed data used to train high-stakes AI applications), the environmental impact of large models[5], and accountability frameworks[6] for the data used in AI systems. In parallel we have developed a growing library of responsible AI tools, informed by how we put our AI Principles into practice. We also continue to expand our Responsible AI Toolkit[7], launched in 2020; this year we added hands-on technical tutorials on topics such as privacy in machine learning[8], Know Your Data (KYD)[9] (a tool to better understand datasets and improve data quality), and an updated People + AI Research (PAIR) Guidebook[10]. We've also continued to improve the Language Interpretability Tool (LIT), designed to help identify and mitigate bias in language models. LIT Version 0.4[11], released in November 2021, adds a number of new features including support for Google Cloud tools and the ability for developers to explore tabular and image data.

Researchers are continuing to develop a technical infrastructure approach to supporting how fairness research is applied in Google products directly. For example, in the last year, we have been able to apply new, state-of-the-art tools developed and announced over the last couple of years, such as Fairness Indicators[12] (launched in 2019[13]) for model evaluation and our Min-Diff[14] technique (launched in 2020[15]) for remediation to a growing number of product use cases, enabling the scale of our learned best practices to proactively address fairness considerations. We are just at the beginning of the journey to scale these approaches, and are committed to doing so and sharing results responsibly.

→CAPTIONS_WORDS_AGE ↓CAPTIONS_WORDS_MOVEMENT	elderly 316	old 4,936	older 1,460	teenage 133	young 12,701	younger 108
NONE 12,090	↗ 1.19x 245 (206)	↗ 1.37x 4,419 (3,217)	↗ 1.2x 1,141 (952)	↘ 0.73x 63 (86.7)	↘ 0.84x 6,982 (8,278)	↗ 1.14x 80 (70.4)
catching 176	↘ 0.33x 1 (3)	↘ 0.09x 4 (46.8)	↘ 0.51x 7 (13.9)	↗ 2.38x 3 (1.3)	↗ 1.38x 166 (121)	↘ 0 (1)
dancing 23	↗ 2.55x 1 (0.4)	↘ 0.33x 2 (6.1)	↘ 0 (1.8)	↗ 6.06x 1 (0.2)	↗ 1.21x 19 (15.7)	↘ 0 (0.1)
jogging 2	↘ 0 (0)	↘ 0 (0.5)	↘ 0 (0.2)	↘ 0 (0)	↗ 1.46x 2 (1.4)	↘ 0 (0)
jumping 528	↘ 0 (9)	↘ 0.03x 4 (140)	↘ 0.05x 2 (41.6)	↗ 1.58x 6 (3.8)	↗ 1.44x 522 (362)	↘ 0 (3.1)
playing 2,914	↘ 0.62x 31 (49.6)	↘ 0.13x 103 (775)	↘ 0.64x 146 (229)	↗ 1.15x 24 (20.9)	↗ 1.37x 2,724 (1,995)	↘ 0.77x 13 (17)
riding 2,146	↘ 0.44x 16 (36.6)	↘ 0.33x 191 (571)	↘ 0.49x 82 (169)	↗ 1.95x 30 (15.4)	↗ 1.28x 1,885 (1,469)	↘ 0.4x 5 (12.5)
running 290	↘ 0 (4.9)	↘ 0.34x 26 (77.2)	↘ 0.39x 9 (22.8)	↘ 0.96x 2 (2.1)	↗ 1.29x 257 (199)	↗ 1.18x 2 (1.7)
skating 209	↘ 0 (3.6)	↘ 0.02x 1 (55.6)	↘ 0 (16.4)	↗ 4x 6 (1.5)	↗ 1.45x 207 (143)	↘ 0.82x 1 (1.2)
swimming 54	↘ 0 (0.9)	↘ 0.28x 4 (14.4)	↘ 0.47x 2 (4.3)	↗ 2.58x 1 (0.4)	↗ 1.35x 50 (37)	↘ 0 (0.3)
throwing 285	↘ 0 (4.9)	↘ 0.08x 6 (75.8)	↘ 0.45x 10 (22.4)	↗ 2.45x 5 (2)	↗ 1.41x 275 (195)	↘ 0 (1.7)
walking 1,012	↗ 1.39x 24 (17.2)	↘ 0.76x 204 (269)	↗ 1.05x 84 (79.7)	↗ 1.24x 9 (7.3)	↗ 1.08x 748 (693)	↗ 1.7x 10 (5.9)

This screenshot from Know Your Data shows a relationship between words associated with age and movement. KYD analysis can be used to identify and address potential unfair biases in data labels.

To do so, we continue to develop practices for creating transparency documents around datasets used for fairness evaluation and training; this year we released new datasets and data cards—including for avoiding unfair bias in Translate[16] and more inclusive annotations of people[17]—together with guidance on how to document these datasets responsibly via the Data Cards Playbook[18], launched this year with an interactive external workshop[19] at the 2021 Association for FAccT (Fairness, Accountability and Transparency) conference.

Important progress has also been made on privacy-preserving technologies. For example, federated learning[20], used in products like Gboard, allows models to be centrally trained and updated based on real user interactions without collecting centralized data from individual users. Since releasing federated learning in 2017, Google researchers have developed federated analytics[21], which uses similar techniques to get insight into how product features and models perform for different users without collecting centralized data. Among other things, federated analytics allows model developers to conduct certain types of fairness testing on federated systems despite not having access to raw user data,

overcoming a significant challenge with federated systems in the past. This year, the team released federated reconstruction[22], a model-agnostic approach to faster, large-scale federated learning and personalization under privacy and communication constraints.

Researchers have also continued to build on the 2020 GShard results[23] for more compute- and energy-efficient models. This year the team developed GSPMD[24], an automatic, compiler-based, and scalable parallelization system: compilation time stays constant with an increasing number of devices. The team has also used the combination of pre-training, self-training, and scaling up model size to enable significant efficiency improvements in other models[25], matching state-of-the-art performance in an automatic speech recognition model with only 3% of the training data.

Researchers and practitioners at Google continue to use and improve existing tools for examining fairness. For example, the research team developing novel techniques for an AI-based model for COVID-19 epidemiology[26] published new findings[27] this year on using the What-If Tool (WIT)[28] and fairness analysis techniques to identify and address potential challenges that could impact public health. Another team used KYD to explore gender bias[29] in the COCO Captions dataset, which includes over 300K human-captioned images used to train models in image labeling and classification tasks.

Finally, we note the importance and responsible practice of continuing to build a strong and diverse team. Google is committed[30] to building a workforce that is more representative of our users and a workplace that creates a sense of belonging for everyone. One goal is to improve leadership representation of Black+, Latinx+, and Native American+ Googlers in the US by 30% by 2025. We've already reached this goal and are on track to double the number of Black+ Googlers at all other levels by 2025. In addition, all VPs and above at Google are now evaluated on progress in diversity, equity, and inclusion. These efforts are crucial if we're going to have a more representative set of researchers and engineers building future technologies. With respect to representation in AI, while there are dozens of Black+ and Hispanic/Latinx+ Googlers and hundreds of female Googlers on our AI teams, we continue to work to expand this representation. This includes attracting talent and building partnerships in regions with diverse talent pools, and investing in the external research community through grants, workshops, and other initiatives to support new voices entering into the field of computer science and more equitable outcomes.

Product impact

Our progress in responsible innovation is reflected in improvements in our products. We are focused on user-centric design and our mission of universal accessibility and utility. This goes back to our earliest efforts to use AI to improve our flagship Search product. For example, in 2011, we announced Panda[31], a machine learning algorithm that allowed us to take the overall content quality of a website into account and adjust its Search rank accordingly. Ten years later, our Search Quality team continues to leverage AI tools to improve results for users, address misleading information, and minimize user exposure to offensive query predictions.

We continue to develop and improve Search responsibly, carefully testing and evaluating new features to ensure that they are beneficial to our users. A big advance this year was the announcement of our Multitask Unified Model (MUM)[32], which will allow Search to understand information across a wide range of formats, like text, images and video, and draw implicit connections between concepts, topics and ideas about the world around us. Applying MUM will not only improve how people around the world find the information they need, but will also play a role in increasing economic opportunity for creators, publishers, startups and small businesses. Ahead of MUM deployment in Search, the team is carefully analyzing quality gains and losses and examining the impact on a broad set of queries and specialized slices to identify any unexpected or concerning performance.

Another area of progress in 2021 was improving the performance of our products for darker skin tones. The Pixel and Photos teams partnered with a diverse range of renowned image makers who are celebrated for their beautiful and accurate depictions of communities of color to help our teams understand where we needed to do better. With their help, we significantly increased the number of portraits of people of color in the image datasets that train our camera models. This culminated in our announcement of Real Tone[33], a feature in the Pixel 6 and Pixel 6 Pro, enabling better performance of features like face detection, auto-exposure, and auto-enhance for users with darker skin tones.

Accessibility in our products has also improved with the application of AI. In 2021, the Android team launched an updated version of Lookout[34], an Android app developed for blind and low vision individuals which uses computer vision to provide information about a person's surroundings. In June, Lookout v2.3 launched with a much-improved Explore mode: object identification is now faster and more accurate. At the same time, the Android team also released the latest version of Voice Access[35], including improvements for entering passwords and the gaze detection beta. This year we also announced Project Relate[36], which uses machine learning to help people with speech impairments communicate and use technology more easily.

We have continued to refine Privacy Sandbox[37], a collaboration with the digital ads industry to enhance privacy through the addition of AI-related techniques. For example, one Privacy Sandbox proposal is to use a technology called Federated Learning of Cohorts (FLoC) to enable the digital ads ecosystem to serve relevant ads to users without tracking

user identities across the web. While still a work in progress, Sandbox aspires to take a significant step forward in delivering a more private user experience while supporting publishers, advertisers and content creators and preserving the vitality of the open internet.

Expert teams provide dedicated resources and consultations to product teams on AI Principles-related issues, so that product teams can learn about—and proactively avoid or mitigate—common failure modes. The product fairness (ProFair) team, for example, provides socio-technical advice and conducts proactive algorithmic fairness testing to ensure new AI technologies do not reflect or perpetuate sociological or socioeconomic inequalities. The team has also taken on responsibilities for investigating inclusion and equity of language in data labeling and diversity and representation in training data for AI. For instance, before launching a new feature called Style AI in Google Shopping[38] in countries like India and Brazil, ProFair held local focus groups in collaboration with Google's Product Inclusion team. The process helped the Style AI team find diverse images and clothing for these specific demographics. In the past year, a variety of teams, including YouTube, Meet, Translate, and Lookout improved projects related to potential new product features consulted with ProFair testing.

We also continue to evolve internal policies and frameworks to support product teams, gathering lessons learned from reviews and product engagements to guide teams on how to think about complex AI Principles challenges. For example, over the past year we have seen increases in reviews and consultations on surveillance, synthetic media, and affective technologies. To ensure a consistent and higher-throughput review process we are piloting a few frameworks on these topics with assessment guidelines, including objective criteria drawn from precedent centered around the concepts of dignity, autonomy and consent, guidance from human rights experts, and sample case studies developed by Google's AI Principles Ethics Fellows.

Supporting global dialogue, standards, and policy

Google remains committed to engaging with external experts and civil society stakeholders around the world. Establishing norms and best practices for responsible AI development will succeed only as a community endeavor, and it is vital to share learnings and receive feedback on our efforts from the wider community.

This year, we have continued to support programs on inclusion, diversity, and equity in AI research and product development. We sponsored 114 conferences in computer science and related fields, including NeurIPS and ICML. And we continue to partner with the Algorithms for Opacity Group (AFOG) at UC Berkeley, an interdisciplinary research group that bridges disciplinary boundaries to support the equitable development of AI systems. We continue to host quarterly Equitable AI Research Roundtables (EARR), launched in 2020 to focus on the potential downstream harms of AI with experts from the Othering and Belonging Institute at UC Berkeley, PolicyLink, and Emory University School of Law. In 2021, we co-created with our EARR partners a set of internal exercises for product teams to apply equitable research practices.

We're developing new engagements with students who may represent communities not currently represented in the technology industry. In fall 2021, for example, in the US, we've added sessions focused on responsible innovation topics including ethics in ML, applying AI principles, algorithmic unfairness, and bias in technology into two of Google's classroom programs for Historically Black Colleges and Universities (HBCU), Google In Residence (GIR) and TechExchange. We continue to expand our engagements with a growing variety of educational institutions, through ongoing partnerships with schools such as Berea College, where we host hands-on app-development workshops focused on building critical thinking skills toward responsible development of advanced technology. Our Research Scholar[39] Program, started this year, gave grants to more than 50 universities in 15+ countries — and 43% of the principal investigators identify as part of a group that's been historically marginalized in tech. Similarly, our exploreCSR[40] and CS Research Mentorship[41] programs support thousands of undergrads from marginalized groups.

We've also made a concerted effort to continue hands-on learning opportunities globally. Our global community educational outreach efforts have included partnerships[42] with dozens of universities around the world to support development of new ML courses, diversity, and inclusion. We continue to host two-hour interactive digital ML for Policy Leaders workshops and other related trainings. Since launch in May 2020, we have provided ML education to more than 450 policymakers and 150 organizations across North America, Latin America, Europe, Africa, and the Asia Pacific. And we continue to share recommended responsible AI practices via TensorFlow[43], Cloud[44], and Google AI[45] to a range of audiences.

Governments have an important role to play in the responsible development of AI, and we remain committed to sharing our learnings and responsible practices and partnering with governments to support responsibility across the world's AI ecosystem. For example, in addition to providing input on the OECD AI Principles, which were adopted by member countries in 2019, Google participates in the OECD Network of Experts on AI (ONE AI) trustworthy AI working group, developing practical guidance around policies that lead to trustworthy AI. Google also participates in the Global Partnership on AI (GPAI), a multi-stakeholder initiative established by the G7 which aims to bridge the gap between theory and practice on AI by supporting cutting-edge research and applied activities on AI-related priorities.

In addition to international organizations, Google works with national governments around the world on AI policy. Google partners with governments to support small and medium enterprises and academics developing responsible AI techniques and services, including $5M support[46] for the US National Science Foundation's Human-AI Interaction and Collaboration research center and $2M in Google Cloud Platform credits[47] to Portugal's Foundation for Science and Technology (FCT) to support research in natural language understanding and responsible AI.

AI is too important not to regulate, and we are working closely with governments to establish policies that support innovation while managing risk. This year we provided submissions[48] to consultations from the EU on the draft AI Act and proposed reforms to the Product Liability Directive for AI, provided feedback on proposed AI legislation in Brazil, responded to a request for information on AI regulation in Israel, worked with NITI Aayog to develop AI policy proposals for India, submitted comments on the National AI Research Resource Task Force, and provided input on NIST's growing body of responsible AI frameworks, including the AI Risk Management framework and Explainable AI framework, among many others. Standards is another key area of partnership and Google is also an active participant in a variety of efforts, including serving as a founding member of the International Organization for Standardization (ISO)'s SC42 working group[49], and contributing to the Institute of Electrical and Electronics Engineers (IEEE) AI standards[50].

Conclusion

Responsible AI at Google has come a long way since we launched our AI Principles in 2018. We remain committed to these principles, and in the last three years we have expanded our AI Principles governance and support functions from a small internal "start-up" to a cross-functional ecosystem.

There is still a lot we don't know. Foundational questions like how to define fairness for AI systems are not yet fully answered, and techniques to assess and manage AI risks are still being refined. AI technology also continues to evolve, introducing new potential benefits and challenges. Synthetic data, for instance, has the potential to replace the use of sensitive data in model training, but we are still learning how to generate and use synthetic data responsibly and safely. Similarly, emergent areas like large multipod models and affective computing promise exciting new capabilities, but also introduce new risks that we must learn to understand and manage.

We are improving our ability to implement responsible AI through research, tools, frameworks, recommended practices, and engagement with a range of experts and institutions, and will continue to share progress reports (as we have in 2019,[51] 2020[52]) and product case studies at ai.google/responsibilities[53].

End Notes

1. https://blog.google/outreach-initiatives/diversity/racial-equity-update-nov-2021/
2. https://cloud.google.com/blog/products/ai-machine-learning/celebrity-recognition-now-available-to-approved-media-entertainment-customers
3. https://blog.google/technology/ai/crossword-puzzle-big-purpose/
4. https://dl.acm.org/doi/10.1145/3411764.3445518
5. https://arxiv.org/abs/2104.10350
6. https://arxiv.org/pdf/2010.13561.pdf
7. https://www.tensorflow.org/responsible_ai?hl=ro
8. https://www.tensorflow.org/responsible_ai/privacy/guide?hl=ro
9. https://knowyourdata.withgoogle.com/
10. https://pair.withgoogle.com/guidebook/
11. https://github.com/PAIR-code/lit/blob/main/RELEASE.md
12. https://www.tensorflow.org/responsible_ai/fairness_indicators/guide
13. https://ai.googleblog.com/2019/12/fairness-indicators-scalable.html
14. https://www.tensorflow.org/responsible_ai/model_remediation
15. https://ai.googleblog.com/2020/11/mitigating-unfair-bias-in-ml-models.html
16. https://storage.googleapis.com/gresearch/translate-gender-challenge-sets/Data%20Card.pdf
17. https://storage.googleapis.com/openimages/open_images_extended_miap/Open%20Images%20Extended%20-%20MIAP%20-%20Data%20Card.pdf
18. https://pair-code.github.io/datacardsplaybook/
19. https://facctconference.org/2021/acceptedcraftsessions.html#data_cards
20. https://federated.withgoogle.com/
21. https://ai.googleblog.com/2020/05/federated-analytics-collaborative-data.html
22. https://arxiv.org/abs/2102.03448
23. https://arxiv.org/abs/2006.16668
24. https://arxiv.org/abs/2105.04663
25. https://arxiv.org/abs/2109.13226
26. https://cloud.google.com/blog/products/ai-machine-learning/google-cloud-is-releasing-the-covid-19-public-forecasts
27. https://ai.googleblog.com/2021/10/an-ml-based-framework-for-covid-19.html
28. https://pair-code.github.io/what-if-tool/
29. https://ai.googleblog.com/2021/08/a-dataset-exploration-case-study-with.html
30. https://blog.google/outreach-initiatives/diversity/racial-equity-update-nov-2021/
31. https://googleblog.blogspot.com/2011/02/finding-more-high-quality-sites-in.html
32. https://www.blog.google/products/search/introducing-MUM/
33. https://store.google.com/intl/en/discover/realtone/
34. https://support.google.com/accessibility/android/answer/9031274?hl=en
35. https://support.google.com/accessibility/android/answer/6151848?hl=en&ref_topic=6151842
36. https://blog.google/outreach-initiatives/accessibility/project-relate/

37. https://privacysandbox.com/
38. https://blog.google/technology/ai/this-googlers-team-is-making-shopping-more-inclusive/
39. https://ai.googleblog.com/2021/04/announcing-2021-research-scholar.html
40. https://research.google/outreach/explore-csr/
41. https://research.google/outreach/csrmp/
42. https://blog.tensorflow.org/2021/06/2021-request-for-proposals-ml-faculty-awards.html
43. https://www.tensorflow.org/responsible_ai
44. https://cloud.google.com/responsible-ai
45. https://ai.google/responsibilities/
46. https://blog.google/technology/ai/partnering-nsf-human-ai-collaboration/
47. https://portugal.googleblog.com/2021/07/2-milhoes-de-dolares-para-apoiar.html
48. https://ai.google/responsibilities/public-policy-perspectives/
49. https://www.iso.org/committee/6794475.html
50. https://standards.ieee.org/initiatives/artificial-intelligence-systems/index.html
51. https://ai.google/static/documents/ai-principles-2019-progress-update.pdf
52. https://ai.google/static/documents/ai-principles-2020-progress-update.pdf
53. https://ai.google/responsibilities

Google