



Artificial intelligence and the development of new technologies promises to bring huge advantages to businesses and people all over the world, enhancing efficiencies and helping to address complex social problems. We have seen great advances, from AI-assisted robotic surgery to the ability of technology to detect crop disease, predict wildfires, and help prevent blindness. Cognitive computing, artificial intelligence, and deep learning are the next wave of technology transforming how consumers and enterprises work, learn, and play, and global spending on cognitive and AI solutions is expected to reach \$46 billion by 2020 .<sup>1</sup>

Human rights and new technology can intersect in many ways. As noted in the issues paper, AI-supported analysis offers the prospect of extraordinary improvements in humans' data analysis and, especially, in our predictive capabilities, but important questions are being raised about how this powerful technology is developed and used, and there is an increased emphasis on research and ethical standards that drive responsible and beneficial development. While advances in technology have and continue to benefit people and improve human rights, as outlined in the issues paper, freedom of expression, access to information for people of all ages, privacy, safety, and equality are all human rights that can also be restricted by these same advances. As Roman Yampolskiy, an AI researcher at the University of Louisville, said *"It is not enough to simply attempt to design a capable intelligence, it is important to explicitly aim for an intelligence that is in alignment with goals of humanity"*.<sup>2</sup>

Questions around bias, accessibility, mistakes and accuracy compared to human decision making remain. It is clear when looking at both these questions and examples of technological advances that technologies that are designed for similar purposes can generate different outcomes. For example, as referenced in the whitepaper, the 2016 ProPublica investigation into the use of an algorithm called 'COMPAS' by US judges claimed that it was biased against African Americans. Yet a 2017 study found that algorithms have the potential to make more equitable decisions than the judges who currently make bail decisions while simultaneously reducing racial disparities<sup>3</sup>. Similarly, the use of predictive technology to help children and family services has proven to be difficult to build and implement<sup>4</sup>, while also successful when executed carefully<sup>5</sup>. While AI has delivered great benefits in a short period of time, it is crucial to ensure that technology does not cause or contribute to human rights abuses. The next developments must come with investment in

---

<sup>1</sup> <https://www.idc.com/getdoc.jsp?containerId=prUS42439617>

<sup>2</sup> <https://futureoflife.org/2017/12/27/research-for-beneficial-artificial-intelligence/?cn-reloaded=1>

<sup>3</sup> <http://www.nber.org/papers/w23180>

<sup>4</sup> <http://www.chicagotribune.com/news/watchdog/ct-dcfs-eckerd-met-20171206-story.html>

<sup>5</sup> <https://www.nytimes.com/2018/01/02/magazine/can-an-algorithm-tell-when-kids-are-in-danger.html>

research into the ethical development and future outcomes of new technologies, including considerations of equality and accountability.

We support the goals of this project and welcome the opportunity to engage with the Australian Human Rights Commission on innovative ways to ensure human rights are prioritised in the design and governance of emerging technologies. For example, the suggested voluntary trustmark for ethically compliant AI design is an interesting idea. Further research would be useful to provide more insight into how this could work including a comparative analysis of benefits versus alternative approaches. In particular, an approach that provides greater context around ML models should also be explored, such as globally standardised documentation highlighting how well such models and applications work by detailing their performance characteristics, intended use cases, and limitations to allow users to evaluate the suitability of these systems to their context. While a trustmark might be a recognised symbol, more detailed alternatives may be more valuable in helping people understand how models work in practice.

At Google, our mission is to make sure that information serves everyone, not just a few. We develop our products and tools for a variety of reasons - to help people explore the world and find information; to help people build new skills; to help people connect with each other; to help businesses succeed. We build products with inclusiveness in mind and are always seeking to make sure the opportunities created by new technology are available for everyone, in any city, in any country.

As technology plays a bigger role in our lives, industry, policymakers and civil society have a shared responsibility to explore the benefits it can bring to society. We are committed to respecting the rights enumerated in the [Universal Declaration of Human Rights](#) and its implementing treaties, as well as upholding the standards established in the [United Nations Guiding Principles on Business and Human Rights](#). States have a duty to protect against human rights abuses and business has a responsibility to respect human rights. Responsible development and use of new technologies is crucial, and research into fairness, accountability, explainability, and privacy and security is being undertaken. Equality of access and accessibility, along with education and awareness building are also essential. But we must remember that this area is evolving. It requires an adaptive approach that involves multiple stakeholders across government, industry, civil society, and academia. We need to work together to find ways of leveraging technology to promote human rights on a global scale, while also being aware of risks that may arise and taking the necessary steps to mitigate those risks.

## Ethical and Responsible Development

### Principles and Research

The design and development of new technologies has to be ethical, benefit society and protect human rights. Harnessed appropriately, AI, ML, algorithms, and new technologies can deliver great benefits for economies and society, and support decision-making which is fairer, safer and more inclusive and informed. However, this will not be realised without great

care and effort, including addressing the risks of technology being misused and taking steps to minimise those risks.

Google published<sup>6</sup> AI principles<sup>7</sup> in June 2018, committing to assess AI applications in view of a set of objectives, including that AI is socially beneficial, avoids creating or reinforcing unfair bias, and is built and tested for safety. We also set out four application areas where we will not design or deploy AI, including technologies whose purpose contravenes widely accepted principles of international law and human rights. These principles are being integrated into existing launch and deal review processes and we're working with cross-functional teams across Google to help identify issues or concerns that may require review. We are also establishing a review body that will include external experts from diverse disciplines, geographies and points of view to help advise us on how to implement our Principles moving forward.

Given the potential AI has to help solve problems in a range of fields, it is important that everyone has access to the opportunities this technology presents. Google.ai<sup>8</sup> pulls all our AI initiatives into one effort to reduce barriers and accelerate how researchers, developers and companies work in this field, and we have made a range of educational resources for all types of learners available for free. We have published a machine learning crash course<sup>9</sup>, a training that Google Engineers take, that anyone can access for free to learn and apply fundamental machine learning concepts. We have also made available AutoML<sup>10</sup>, a service that allows third parties with limited machine learning expertise to create custom machine learning models. Our goal is to build people-first AI and we have invested in initiatives, like People + AI Research (PAIR),<sup>11</sup> to design frameworks and tools to make AI work well for everyone and to reflect human values. We have developed various PAIR tools to help promote understanding of the relationship between users and technology, the new applications AI enables, and how to make it broadly inclusive. For example, Facets<sup>12</sup> is based on the premise that better data leads to better models. It is an open-source application that gives engineers a clear view of the data they use to train AI systems, providing robust visualisations to aid in understanding and analysing machine learning datasets. What-If<sup>13</sup> is another open-source PAIR tool that allows engineers to probe "what if" scenarios and inspect a machine learning model without the need for writing any further code. Our commitment to open science has also led to the development of the open-source machine learning framework TensorFlow<sup>14</sup>, which is now the largest machine learning community on GitHub. Researchers and developers around the world share deep learning models and datasets designed to make deep learning more accessible and accelerate machine learning research.

---

<sup>6</sup> <https://blog.google/technology/ai/ai-principles/>

<sup>7</sup> <https://ai.google/principles/>

<sup>8</sup> <https://ai.google/education>

<sup>9</sup> <https://developers.google.com/machine-learning/crash-course/>

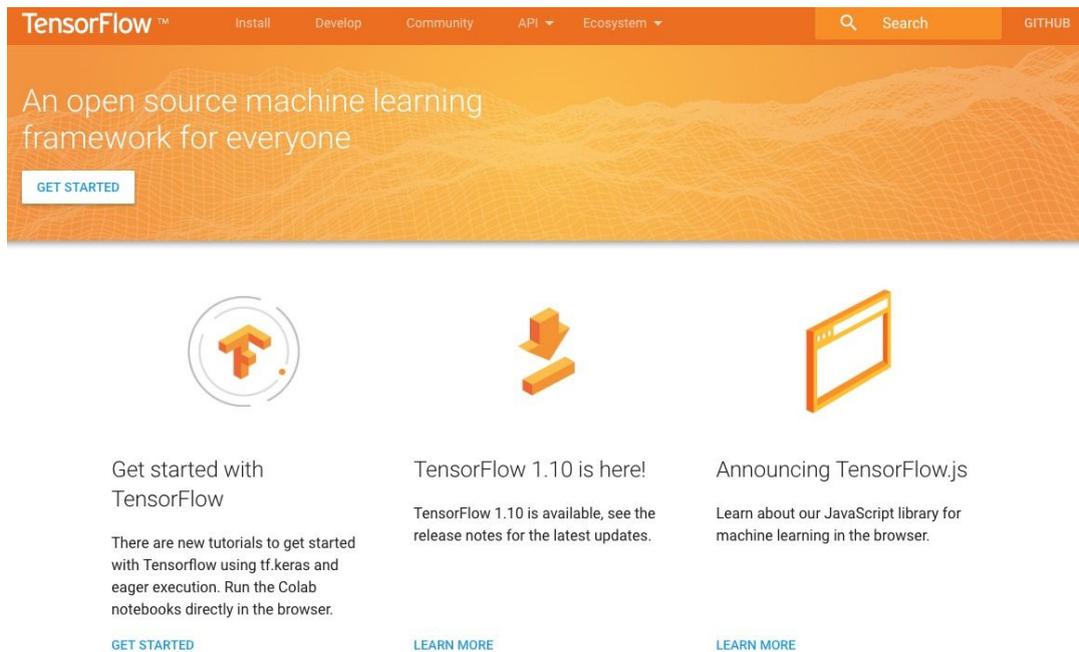
<sup>10</sup> <https://www.blog.google/products/google-cloud/empowering-businesses-and-developers-to-do-more-with-ai/>

<sup>11</sup> <https://ai.google/research/teams/brain/pair>

<sup>12</sup> <https://pair-code.github.io/facets/>

<sup>13</sup> <https://pair-code.github.io/what-if-tool/>

<sup>14</sup> <https://github.com/tensorflow>



Explainability and transparency into research methods are central considerations for our Engineers and product teams. We have developed supporting AI technical practices<sup>15</sup> to share some of our current work and recommended practices, and are continuing to work towards sharing our research methods on how to build more inclusive ML<sup>16</sup>. Google is committed to making progress in the responsible development of new technologies and sharing knowledge, research, tools, datasets, and other resources with the larger community. In 2016, Google, DeepMind, Microsoft, Amazon, Facebook, and IBM founded The Partnership on AI to Benefit People and Society: a non-profit organisation to advance public understanding of AI, promote open collaboration, formulate best practices in the field and support research on the social impact of the technology. This partnership aims to provide diverse viewpoints and support informed dialogue where challenging questions are informed by thoughtful analysis. The governing board has representation from policy and ethics experts drawn from non-profits, academia and user-advocacy groups. Together, the organisation's members will conduct and publish research under an open license on these industry-defining topics.

## New Technologies & Google

### **Balancing access and expression with security and positive experiences**

In 2016, we launched Jigsaw<sup>17</sup> with the goal to tackle the toughest geopolitical challenges to help make people in the world safer - safer from attacks on free speech, online harassment, violent extremism, and injustice and corruption. Significant investment has been made in developing technology to work towards achieving this goal, for example with Perspective<sup>18</sup>,

<sup>15</sup> <https://ai.google/education/responsible-ai-practices>

<sup>16</sup> <https://developers.google.com/machine-learning/fairness-overview/>

<sup>17</sup> <https://jigsaw.google.com/>

<sup>18</sup> <https://www.perspectiveapi.com/#/>

an API that uses machine learning to detect abuse and harassment online. Perspective scores comments based on their similarity to other comments that others have marked as “toxic”. Developers and publishers can use this score to give real time feedback to commenters or help moderators do their job or allow readers to more easily find relevant information.

Given that words can be used in both positive and negative ways, in March this year we also launched Project Respect<sup>19</sup> to expand the diversity of training. It is an open dataset that collects diverse statements from the LGBTIQ+ community and will be made available in an open dataset, which coders, developers and technologists can use to help teach machine learning models. We are continuing to collect input to help make conversations more inclusive and are taking Project Respect on tour to Pride events around the world. We started in Sydney<sup>20</sup> and Auckland and will continue to roll out in other parts of the world.

Google’s mission continues to be to organise the world’s information and make it universally accessible and useful. With this goal comes many benefits to society -- access to art and culture, news and entertainment, and educational materials. We value this openness. For instance, another Jigsaw project is Project Shield<sup>21</sup>, which helps news organisations around the world who suffer crippling digital attacks when they publish something controversial or that questions powerful institutions. Project Shield uses Google’s infrastructure to protect independent news sites, who often depend entirely on their web presence to get information to the public, from DDoS attacks. Project Shield is built on Google’s infrastructure and offers free and unlimited protection.

Although we are a company fundamentally committed to access to information, it is not anything goes on our platforms. We’ve created content policies that set the rules of the road for what we don’t allow. And once we are on notice of content that violates a local law, we act swiftly to restrict access to it from the relevant jurisdiction. We continue to improve on our processes and our technology to enforce those rules, while preserving the open democratic nature of our services. We take these issues seriously and want to be a part of the solution. We work regularly with NGOs and governments to better understand the issues and how we can be of help.

At YouTube, we work hard to maintain a safe and vibrant community, and it is an area we have always been heavily invested in. We have Community Guidelines that set the rules of the road for what we don’t allow on YouTube. We have long used a mix of technology and humans to deal with controversial content. Our community of users notifies us of content that may violate our policies by flagging content to us for review using the various reporting options available on the platform. We have teams based in multiple locations around the world who review flagged content in order to take appropriate action in a timely manner. We also use technology to expedite review of content that may violate our policies and to prevent exact reuploads of content that we have determined to be in violation of our policies.

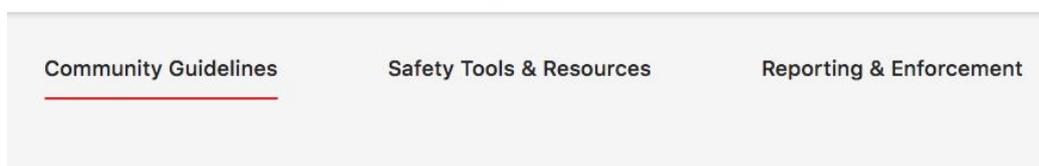
---

<sup>19</sup> <https://g.co/projectrespect>

<sup>20</sup> <https://australia.googleblog.com/2018/02/promoting-pride-and-respect-with.html>

<sup>21</sup> <https://projectshield.withgoogle.com/public/>

As the technology improves, we can use it more to help us detect this content. We're now using machine learning to help detect potentially violative content in certain areas (e.g. extremist content and child exploitation). Between April and June 2018, 87% of the almost 8 million video removed from YouTube were initially flagged by our AI systems. In addition, 76% of video removals (first flagged through our automated flagging systems) occurred before they received any views<sup>22</sup>. Machines can help identify problematic videos, but human experts still play a role in nuanced decisions. Content needs to be reviewed by a human before it is taken down and we do not automatically remove all material that has been identified by machine learning. We believe our ability to address sensitive or controversial content at scale hinges on a variety of different approaches - including investment in technology, human reviewers, and awareness efforts.



Here are some common-sense rules that'll help you steer clear of trouble. Please take these rules seriously and take them to heart. Don't try to look for loopholes or try to lawyer your way around the guidelines—just understand them and try to respect the spirit in which they were created.

## Accessing and Benefiting from Technology

### Product Accessibility

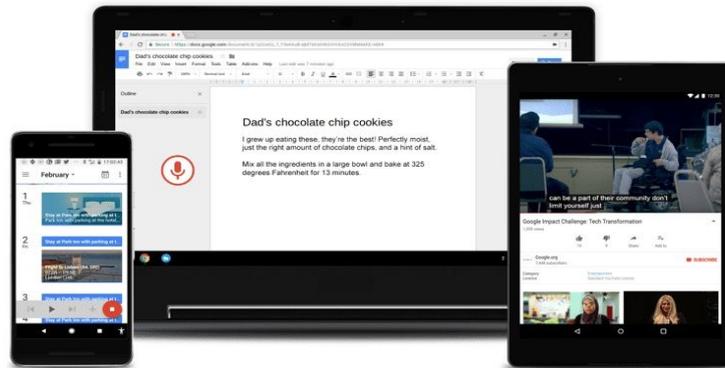
Access and accessibility need to be considered from the beginning of the development cycle in order to produce technologies and products that contribute positively to people's lives and help make them a bit easier. Inclusive user experiences can be built when teams understand and close gaps in user testing by thinking globally, accounting for intersectionality, and identifying differences in experiences and knowledge as well as identity. Whether it's accessible technology, assistive technology or breaking down social or economic barriers, advances in technology, product development and digital literacy efforts are helping to provide more people with the ability to maximise what technology has to offer.

We are committed to making accessibility a core consideration from the earliest stages of product design. We have a central accessibility team with a mandate to monitor the state of accessibility of Google products and coordinate accessibility training, testing, and consulting. Product teams are offered training to help incorporate accessibility principles into the design and release of products.

---

<sup>22</sup> <https://transparencyreport.google.com/youtube-policy/overview?hl=en>

Everyone should be able to access and enjoy the web.  
We're committed to making that a reality.



The Google User Experience team believes that usability and accessibility go together. Researchers work with people from all backgrounds and with a variety of abilities in order to understand what matters to people in their daily lives. They conduct usability studies which enable us to collect direct feedback and provide a better understanding on how usable products are, highlight what the main roadblocks may be, and help to improve our products for everyone. Based on these insights, we aim to design products that are inclusive and consider the needs of all users.

Google actively promotes an accessible web by serving on standards and advisory committees. For example, to make it easier for companies, educational institutions, and government agencies to comply with accessibility standards, we provide transparent information about how our products currently work for people with disabilities. We offer Voluntary Product Accessibility Templates (VPAT) for [Gmail](#), [Drive](#), [Docs](#), [Sheets](#), [Slides](#), [Forms](#), [Sites](#). More information can be found on the Google Accessibility website<sup>23</sup> and The Designer's Guide to Accessibility Research blog post<sup>24</sup>.

At YouTube, features like captions help make video content accessible for more viewers, including those who speak a different language as well as the hearing impaired. YouTube has two kinds of caption files: auto-captions and captions that are manually added. Some channels let viewers contribute subtitles and closed captions to their videos. People can submit content for an entire video or just add what they know. The auto-captions are generated automatically by speech recognition technology and machine learning algorithms. As of 2017, we had captioned more than 1 billion videos, with automatic captioned videos receiving more than 15 million visits per day.

<sup>23</sup> <https://www.google.com/accessibility/>

<sup>24</sup> <https://design.google/library/designers-guide-accessibility-research/>

Google Cloud Speech-to-Text machine is powered by learning and enables developers to convert audio to text by applying neural network models in an easy-to-use API. The API recognises 120 languages and variants to support global users. This can help people overcome challenges in communication, like when trying to communicate with someone who speaks another language or dialect, and it can be helpful for people from minority language groups whose languages often have clunky typing interfaces..

The focus isn't just on our products and services, we're working with others to build a more inclusive world. In 2015, we introduced the Google Impact Challenge: Disabilities<sup>25</sup> with the aim of making the world more accessible for the 1 billion people living with disabilities. Google pledged \$20M in grants to 29 nonprofit organisations using technology to take on a wide range of accessibility challenges, including Wheelmap (creating a global dataset of accessible locations for people with disabilities), Benetech (making more books accessible to the visually impaired), e-Nable (a global community of volunteers working to create 3D printed hands for children with limb differences).

### **User Education & Control**

Products should be designed with safety and security as core principles and with the goal of being accessible to many users. We believe in providing all our users with choice, transparency, and control and invest significant resources to make sure we are making the right engineering decisions (using 'privacy by design' and 'safety by design' principles), providing our users with tools to control how their data is being used. We recently updated our Privacy Policy<sup>26</sup> to make it easier to understand what information we collect and why we collect it. We explain our practices with clearer language and included new explanatory videos and illustrations to help users better understand the information provided.

In Google Account, users can review their Google dashboard, view or delete data, manage ad settings, take a Security Checkup<sup>27</sup> or Privacy Checkup<sup>28</sup>, and use the Activity Controls<sup>29</sup> to choose what activity (e.g. location history, app history, YouTube search history) is saved to their Google Account. Google Account is used extensively: In 2017, nearly 2 billion unique users visited Google Account, with the majority making changes or adjustments to their privacy settings.

'Download Your Data' allows users of Google products, such as Gmail, Calendar, YouTube and Photos, to export their data to a downloadable ZIP file. More recently, a collaboration of organisations (Google, Facebook, Microsoft, and Twitter) teamed up to launch a new open-source, service-to-service data portability platform, called the Data Transfer Project (DTP). While users were previously able to download a copy of their data from most services, DTP aims to make moving data between providers significantly easier for users. The Data Transfer Project uses services' existing APIs and authorisation mechanisms to access data. It then uses service specific adapters to transfer that data into a common

---

<sup>25</sup> <https://www.google.org/impactchallenge/disabilities/index.html>

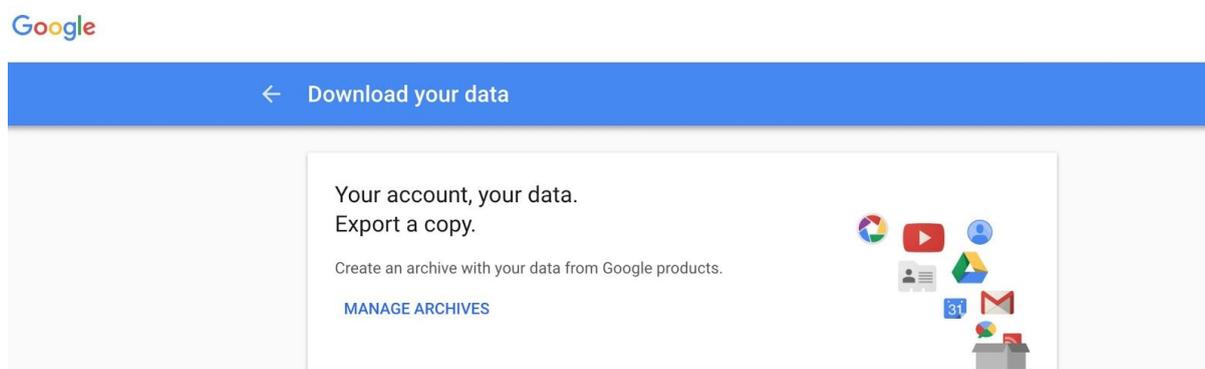
<sup>26</sup> <https://policies.google.com/privacy/update>

<sup>27</sup> <https://myaccount.google.com/security-checkup?continue=https://myaccount.google.com/security>

<sup>28</sup> <https://myaccount.google.com/privacycheckup>

<sup>29</sup> <https://myaccount.google.com/activitycontrols>

format, and then back into the new service's API. It can be used by users who want to back up their data, leave an existing service, or try a new service.



In November 2017, we launched Family Link<sup>30</sup> in Australia. Family Link enables parents to create a Google account for their child under 13, and helps them manage aspects of the account and their child's device. With Family Link, parents can keep an eye on screen time and their child's use of technology. Parents can use Family Link to set a device bedtime and daily limit, see how much time their child is spending on their favourite apps with daily, weekly and monthly app activity reports, and have the ability to approve or block the apps their child wants to download. Be Internet Awesome<sup>31</sup>, Google's curriculum to teach kids the fundamentals of digital citizenship and safety, and Interland<sup>32</sup>, the accompanying game, puts the key lessons of digital safety into four games. We also support The Alannah and Madeline Foundation's Digital Licence<sup>33</sup> that teaches school aged children critical digital skills.

## Looking Ahead

We believe new technologies will be overwhelmingly beneficial, and there are many examples of how they are already improving people's lives. But new technologies can impact society in a variety of ways, which is why Government has such an important role to play alongside industry and civil society to ensure good outcomes with sensible and informed discussion.

As technology advances, governments should expand technological expertise and explore various cooperative frameworks to minimise issues and maximise potential. We believe that consensus-driven best practices and self-regulatory bodies can also contribute to creating flexible and nuanced approaches.<sup>34</sup> Initiatives like the Partnership on AI demonstrate that we have a joint responsibility and that industry, civil society and policymakers can come together to create solutions that keep humans at the core of new technologies. We are also conducting research to understand and address any potential risks early on. We want to build people-first AI and have invested in initiatives like People + AI Research (PAIR), to design frameworks and tools to make AI work well for everyone and to reflect human values.

<sup>30</sup> <https://families.google.com/familylink/>

<sup>31</sup> <https://beinternetawesome.withgoogle.com/en>

<sup>32</sup> <https://beinternetawesome.withgoogle.com/en/interland>

<sup>33</sup> <https://www.digitalllicence.com.au/>

<sup>34</sup> Responsible Development of AI - 2018 Google White Paper