

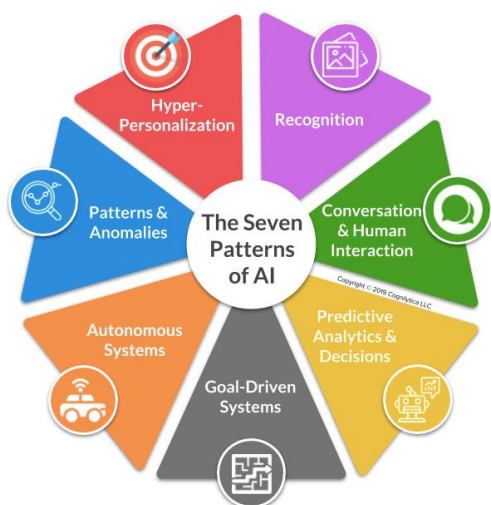
# ARTIFICIAL INTELLIGENCE & RESPONSIBLE BUSINESS CONDUCT

Over the last decade, rapid advancements in artificial intelligence (AI) and machine learning have opened up new opportunities for productivity, economic development, and advancements in various sectors, from agriculture to healthcare. While current and future AI applications have the potential to advance responsible business, they can also pose risks to human rights, the environment and other important elements of responsible business conduct as addressed in the OECD Guidelines for multinational enterprises.

For example, AI can unlock significant improvements in occupational health and safety through automation of dangerous tasks. The use of AI in smart grids, smart cities and connected devices can help reduce greenhouse gas emissions and aid in the adaptation to climate change.

On the other hand, the use of AI in hiring, law enforcement, lending, and other fields could lead to discriminatory outcomes through the reliance on inappropriately biased data or algorithms. Relying on and collecting increasing amounts of personal data, there is a risk of AI adversely impacting privacy. When used in autonomous weapons systems, AI could lead to impacts on the human right to life, personal security, and due process.

This background paper provides an overview of the different types of AI applications, the ways in which humans and AI can interact, and potential adverse human rights and societal impacts that AI technology may introduce.



## Definition and Overview

The OECD defines an *Artificial Intelligence (AI) System* as a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments.<sup>1</sup> When applied, AI has seven different use cases, also known as patterns, that can coexist in parallel within the same AI system.<sup>2</sup>

1. *Hyper-personalisation* uses AI to develop a profile of each individual, and then having that profile learn and adapt over time for a wide variety of purposes including

<sup>1</sup> OECD, "Recommendation of the Council on Artificial Intelligence" (2019) Available at: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>

<sup>2</sup> Source of diagram: Cognilytica, "The Seven Patterns of AI" (2019) Available at: <https://www.cognilytica.com/2019/04/04/the-seven-patterns-of-ai/>

displaying relevant content in online search, recommending relevant products, providing personalized recommendations and so on. Hyper-personalisation is often used in targeted marketing, for example, Netflix suggesting other shows and movies on its platform based on relevant content the user already watched.<sup>3</sup> Another example of hyper-personalisation is in finance, with the movement away from traditional financial credit scoring systems, which work by “bucketing” people into credit worthiness scores based on an individual’s past data to a hyper-personalised 1:1 credit scoring system using algorithms that look at the credit histories of individuals with similar characteristics in order to produce targeted scores for an individual user.

2. *Conversation and human interaction* use cases involve machines and humans interacting with each other through conversational content across a variety of methods including voice, text, and image forms. For example, applications such as Wysa, Joyable and Talkspace use chatbots to provide users with automated mental health care and conduct mood and intent analysis. Other examples include digital assistants such as Siri and Alexa.
3. Applications using *pattern and anomaly detection* identify patterns in the data and higher order connections between data points to see if they fit an existing pattern or if they are an outlier or anomaly. For example, this type of AI is used in law enforcement to detect financial fraud or money laundering, flagging purchases in unusual amounts or locations.<sup>4</sup>
4. *Recognition* uses machine learning and other cognitive approaches to identify and determine objects or other data points within image, video, audio, text, or other primarily unstructured data. This type of AI is used in facial recognition to verify someone’s identity, for example in border control, to monitor the movement of people or to unlock a smart phone. It can also be used in healthcare, in combination with anomaly detection, to review pictures of skin conditions and provide potential diagnoses.
5. *Goal Driven Systems* use machine learning and other cognitive approaches to give computer systems the ability to learn through trial and error. The primary objective of these systems is to find the optimal solution to a problem. Use cases include bidding and real time auctions. This type of AI has also been tested in games, for example when IBM’s Deep Blue computer beat world champion Gary Kasparov in chess in 1997.<sup>5</sup>
6. *Predictive analytics and decision support* are used to understand how past or existing conditions or behaviours can help predict future outcomes to help humans make better decisions. An example of this is the application of an AI system to weather forecasting, using data to assess past forecasts and improve predictions over time. With these algorithms, humans still make the final decisions about what to do with the AI system’s

---

<sup>3</sup> Forbes, “The Seven Patterns of AI” (2019) Available at: <https://www.forbes.com/sites/cognitiveworld/2019/09/17/the-seven-patterns-of-ai/#2e384e2c12d0>

<sup>4</sup> Forbes, “AI Is Predicting The Future Of Online Fraud Detection” (2019) Available at: <https://www.forbes.com/sites/louiscolombus/2019/08/01/ai-is-predicting-the-future-of-online-fraud-detection/#736abeac74f5>

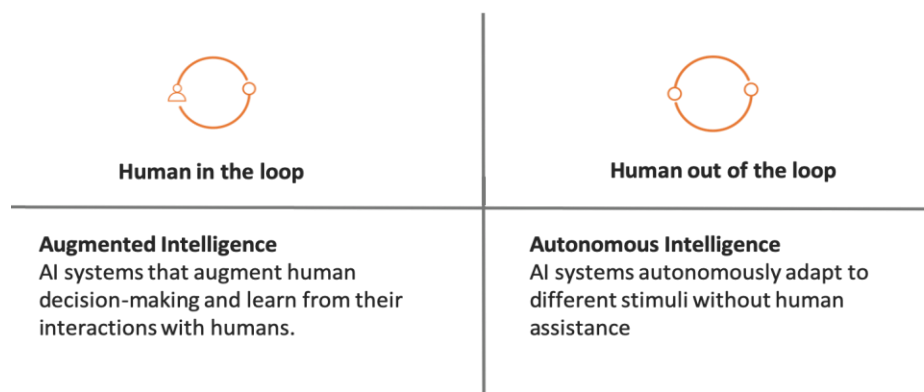
<sup>5</sup> Scientific American, “20 Years after Deep Blue: How AI Has Advanced Since Conquering Chess” (2017) Available at: <https://www.scientificamerican.com/article/20-years-after-deep-blue-how-ai-has-advanced-since-conquering-chess/>

predictions. This is called augmented intelligence, as opposed to other forms of AI which are fully autonomous.

7. *Autonomous systems* are physical and virtual software and hardware systems that are able to accomplish a task, reach a goal, interact with their surroundings, and achieve an objective with varying degrees of human involvement; they require machine learning capability that can without further human interaction i receive information about the real or virtual environment , and predict the future behaviour of all elements in the system, and plan for how to deal with those changes. An example of this is autonomous vehicles.

### Human Interactions with AI

Humans interact with AI enabled technologies at different levels. Even where an AI system is designed to operate autonomously there is a role for humans in the responsible development of the system:

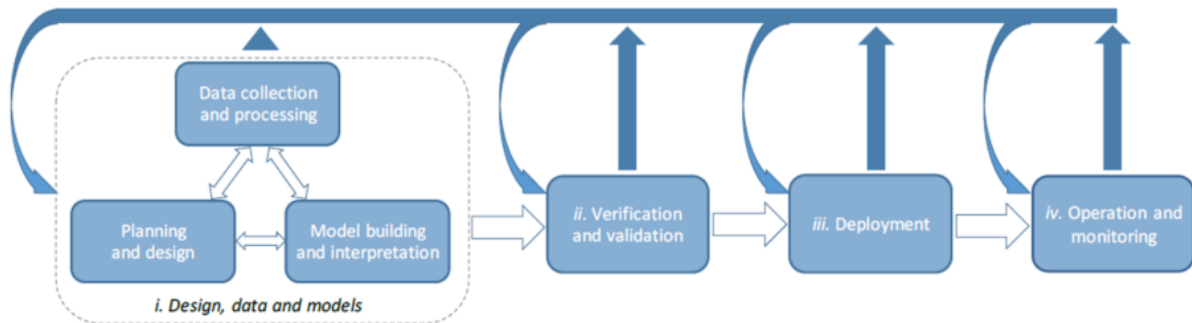


We can understand the different levels of human involvement with the example of an AI application that helps doctors more efficiently diagnose patients. If the app provides *augmented intelligence*, it may use hyper-personalisation, predictive analytics and decision support AI patterns to inform the doctor on the “best” diagnosis, using the patient’s symptoms and medical history for data, as well as verified medical sources. The app would suggest the most statistically likely diagnoses. Based on the doctor’s feedback on correct and incorrect diagnoses the AI system would learn to improve, becoming more accurate over time. Due to the doctor’s contribution to how the app will synthesise the data moving forward, the human is in in the loop.

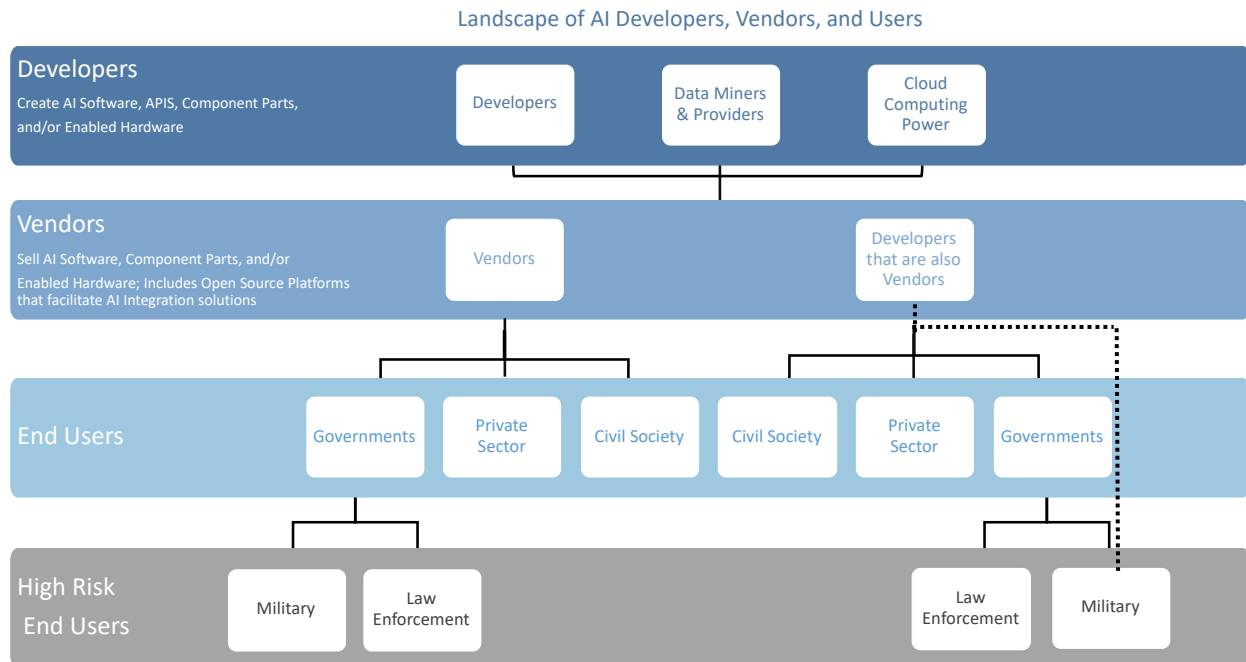
On the other hand, if the app was designed to use *autonomous intelligence*, it could continuously review the patient’s profile and intake other new relevant information, adapting along the way to improve the accuracy of its diagnosis without the help of a doctor. Apart from the design of the app, humans are not in the loop in this process.

## AI Actor Landscape

The ability of an AI system to augment human decision-making and to make changes autonomously that impact how an end user interacts with an AI enabled product requires new approaches to performing human rights due diligence around AI technologies. Different from a standard physical product, where the relationship between the makers, supply chain, retailers and consumer is linear, there is significant overlap and exchange in the AI landscape between developers, vendors and end-users. Indeed, there remains a question of assigning responsibility to different AI actors who either develop, sell, or deploy different technologies. The OECD defines AI actors as “those who play an active role in the AI system lifecycle, including organisations and individuals that deploy or operate AI”.



Source: OECD (2019) *AI in Society: AI system lifecycle as defined and approved by the OECD Expert Group on AI (AIGO) in February 2019*



An example landscape of different AI actors across an AI system value chain provided as an illustration. It may not include all actors in the value chain.

### Developers

AI products are created by developers, through the following process:<sup>6</sup>

1. Ideation: Identifying a problem and priorities for the AI
2. Data Gathering: Selecting the appropriate data set for the AI to learn from
3. Method Selection: Selected the method for teaching the AI what to do with the data
4. Performance Testing: Testing the AI's ability to perform the task it was taught

While due diligence should cover all stages of the product lifecycle, companies have the greatest amount of leverage during the product development of AI technologies. By applying a “human rights by design strategy,” developers can mitigate potential risks of technologies at every step of development. Developer due diligence could include asking questions such as:

- Who will likely use the product and for what purpose?
- Is there the potential for misuse, poor handling or lack of enforcement of respective rules and standards?
- Is there a chance that vulnerable groups will be especially impacted by the use of the technology?

### Vendors

Once a product is developed, it is sold by vendors to end-users, who deploy and operate the technology. It is the responsibility of the vendor to conduct RBC due diligence at the point of sale on the risks associated with the use of the product. Importantly, most AI developers also sell their own products, for example, Microsoft, Apple, IBM, Amazon, and Google.

These companies may be contracted by governments or militaries directly to create specific products, and as such, take on extended responsibilities of due diligence. Other companies, such as Cisco, develop AI products that are distributed by its partners or third-party retailers.

Vendor due diligence could include asking questions such as:

- Was the product designed and assembled according to RBC standards?
- Is the product being sold directly to the end-user or to another distributor?
- Does the product come with an end-user agreement or training on AI limitations?

### End Users

End users can be anyone, ranging from a government, to a government contractor to another company, to an NGO to a member of civil society. When end users are government agencies or government contractors, particularly militaries or private military and security companies, they present higher risks of the technology being used for harm and require more stringent due diligence.

---

<sup>6</sup> Adapted from Data Robot, (2019) “Machine Learning Life Cycle” Available at: <https://datarobot.com/wiki/machine-learning-life-cycle/> and Brook, Adrien (2019) Towards Data Science, “10 Steps to Create Your Very Own Corporate AI Project” Available at: <https://towardsdatascience.com/10-steps-to-your-very-own-corporate-a-i-project-ced3949faf7f>

For many AI technologies that are licensed to end users, developers have the ability to monitor the product, creating opportunities for human rights due diligence directly between the developer and the end user. For example, developers and vendors can limit licensing renewals with end users. End user due diligence could include asking questions such as:

- Does the product have a dual-use that is harmful?
- Was the product accompanied by guidance or training on its limitations?
- Has the product been altered in any way that may increase its potential RBC risks through resale channels?

## AI and RBC

RBC includes all interfaces between business and society with human rights playing a particularly important role in the context of AI. The use of AI has the potential to be affiliated or linked with various human rights harms, including the following:



**Article 2 – Right to Non-Discrimination:** Risk of AI incorporating human biases through incomplete or inappropriately biased data sets or through the algorithm design itself, thereby infringing on the right to non-discrimination.



**Article 12 – Right to Privacy:** AI applications require large amounts of data, creating a risk that:

- Law enforcement and intelligence agencies make illegitimate requests or demands for personal information.
- Companies collect, use, or share a person's personally identifiable information without informed consent.

**Article 3 – Right to Life and Personal Security:** AI technology will be used to aid, and potentially to replace, human decision-making on issues directly affecting human life (e.g. autonomous weapons, self-driving cars).



**Article 19 – Freedom to Opinion and Expression:** AI might adversely impact these rights in several ways:

- Risk that human rights defenders self-censor their expression if they fear being surveilled.
- Risk of AI bots influencing social media with misinformation or biased views and opinions.



**Article 20 – Right to Freedom of Association:** The right to freedom of assembly and association can be adversely impacted if governments use AI to monitor and repress individuals' civic engagement.



**Article 23 – Right to Work:** Risk that broad adoption of AI across industrial sectors, job types, and skill-levels could lead to widespread job losses to automation.

**Article 21 – Right to Take Part in Government:** False information perpetuated on internet platforms may delegitimize electoral processes and threaten the human right to participate in government and free elections.



**Article 25 – Right to Adequate Standard of Living:** Widespread displacement of jobs due to automation may adversely impact the human right to an adequate standard of living, if policy solutions are not developed.



Source: <https://visual.ly/community/infographic/human-rights/30-universal-declarations-human-rights>

Societal harm stemming from the use of AI can be the result of a business or government's own activities or can be directly linked to its business operations. As an element of due diligence there

are opportunities for risk mitigation at all stages of AI development. The table below outlines different types of harm:

TYPE OF HARM	EXAMPLES	OPPORTUNITIES FOR RISK MITIGATION	OPEN QUESTIONS
Purposeful “harm by design”	Deepfake video designed to harm an individual’s reputation and right to privacy.	Investments in detection technology	How can harmful deepfakes be prevented without curtailing the freedom of expression?
Harm caused by inherent “side effects”	Biased training data leading to discrimination. Social media algorithm promoting hate speech or false information.	Risk review during product development  Increased user transparency	What checks and balances are necessary to reduce risk of bias transfer or new biases in AI systems?  How should business AI actors provide greater transparency on bias or hate speech risks? What is the role of governments?
Harm caused by failure rates	False positives by facial recognition in law enforcement	Product development Customer human rights due diligence	How can the risks of harm caused by AI failure be mitigated? What tools and systems are needed to ensure appropriate remediation and ensure “lessons learnt”?
Harm caused by intentional misuse	Political disinformation using data / AI systems	Customer and user due diligence	How should business AI actors provide greater transparency on use of AI to spread disinformation or other harms? What is the role of governments?
Harm caused by security breach	Hackers taking control of autonomous vehicles	Product development Public policy	What is the appropriate level of care to ensure security?

## Collaborative Solutions

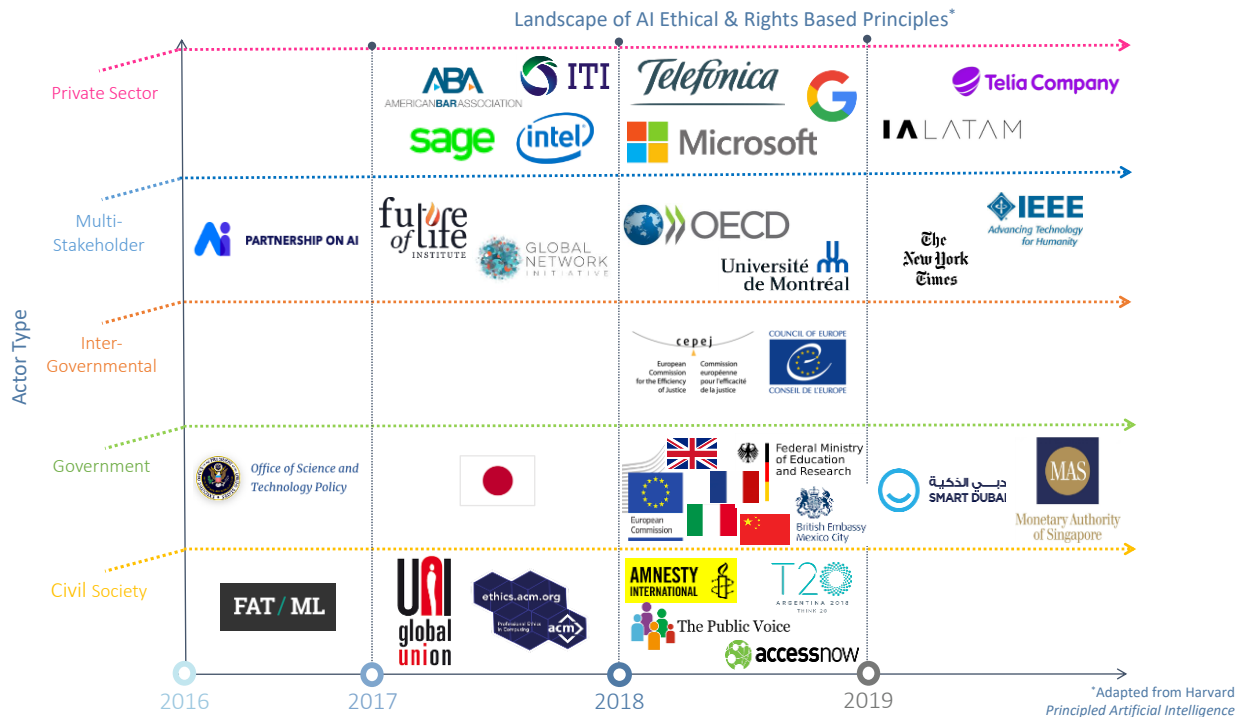
In response to the way AI is reshaping society, a number of companies, governments, and civil society actors are putting in place risk mitigation measures. The most common approach is developing and implementing principles for ethical and rights respecting AI. [Companies such as Google, Workday, TeliaSonera](#) have issued ethical AI principles. Organisations such as the [OECD](#) have put forward guidance, while civil society groups have put forward principles including the [Toronto Declaration](#) and [IEEE Principles](#). A landscape of ethical and rights based AI guiding principles that have come out since 2016 is mapped in the chart on the next page.

To implement ethical AI principles, companies have taken steps including:

- Conducting human rights impact assessments on emerging technologies
- Driving collaboration and dialogue through industry and multi-stakeholder platforms
- Putting in place governance structures, such as internal review committees
- Advocating for public policy aimed at mitigating human rights risks related to AI

[Intel](#) and [Microsoft](#) are among the companies that have conducted human rights impact assessments on emerging technologies. For ongoing risk review, Microsoft has created an internal governance structure to identify and review risks, known as the [AETHER Committee](#).

Collaborative platforms such as the [Partnership on AI](#) have emerged to connect companies with external stakeholders, and bridge gaps to implement rights respecting mitigation at all stages of the product lifecycle, spanning from development to deployment and including both public and private sector actors.



The [OECD Directorate for Science, Technology, and Innovation \(STI\)](#), with the Committee for Digital Economy Policy, supports governments by measuring and analysing the economic and social impacts of AI technologies and applications.



## Further Reading

1. OECD (2019) “Artificial Intelligence in Society” Available at: <http://www.oecd.org/going-digital/artificial-intelligence-in-society-eedfee77-en.htm>
2. OECD (2019) “Recommendation of the Council on Artificial Intelligence” : <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>
3. BSR (2018), “Artificial Intelligence: A Rights-Based Blueprint for Business: <https://www.bsr.org/reports/BSR-Artificial-Intelligence-A-Rights-Based-Blueprint-for-Business-Paper-01.pdf>
4. Data & Society, (2018) “Governing Artificial Intelligence: Upholding Human Rights and Dignity” : [https://datasociety.net/wpcontent/uploads/2018/10/DataSociety\\_Governing\\_Artificial\\_Intelligence\\_Upholding\\_Human\\_Rights.pdf](https://datasociety.net/wpcontent/uploads/2018/10/DataSociety_Governing_Artificial_Intelligence_Upholding_Human_Rights.pdf)
5. AI for Good Global Summit (2017), “AI For Good Global Summit Report,”: Available at: [https://www.itu.int/en/ITU-T/AI/Documents/Report/AI\\_for\\_Good\\_Global\\_Summit\\_Report\\_2017.pdf](https://www.itu.int/en/ITU-T/AI/Documents/Report/AI_for_Good_Global_Summit_Report_2017.pdf)
6. Ada Lovelace Institute (2019), “Ethical and Societal Implications of Algorithms, Data, and AI: a roadmap for research” Available at: <https://www.nuffieldfoundation.org/sites/default/files/files/Ethical-and-Societal-Implications-of-Data-and-AI-report-Nuffield-Foundat.pdf>
7. Cognilytica (2018), “What is Artificial Intelligence?” Available at: <https://www.cognilytica.com/2018/10/08/white-paper-what-is-artificial-intelligence-for-consumer-technology-association/>

This paper was prepared by the OECD Centre for Responsible Business Conduct in the Directorate for Financial and Enterprise Affairs in collaboration with Article One Advisors. It was produced with support from the Government of the Netherlands in order to inform discussions at the OECD workshop on Responsible Business Conduct and Digitalisation taking place on 4 November 2019. The authors are solely responsible for any remaining errors.

This document is published under the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of OECD member countries. This document, as well as any data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

© OECD 2019