Ethics in digital technologies

January 2021

Aude Schoentgen, Laura Wilkinson

plumconsulting.co.uk

About Plum

Plum is a leading independent consulting firm, focused on the telecommunications, media, technology and adjacent sectors. We apply extensive industry knowledge, consulting experience and rigorous analysis to address challenges and opportunities across regulatory, radio spectrum, economic, commercial, and technology domains.

About this paper

This paper explores how ethical frameworks and concerns arise within digital technologies, and examines the implementation of ethics by governments and private companies. It concludes by proposing a forward-looking approach and identifying potential challenges to implement ethics within the design and use of digital technologies.

Plum Consulting
10 Fitzroy Square
London
W1T 5HP

T +44 20 7047 1919
E info@plumconsulting.co.uk

Contents

Introduction

The omnipresence of digital technology in our daily life, its use and its impact on organisations and individuals, raises ethical questions about its role in our society. These concerns include consent and privacy, security, inclusion and fairness, protection from online harm, transparency and accountability. Notable examples include the Cambridge Analytica scandal and concerns about racial discrimination in the design of facial recognition algorithms. Taking into account ethical considerations is essential to ensure that digital technologies benefit as many people as possible.

The maturity of the ethical approach depends largely on the type of technology. The ethics of blockchain and cloud services seem globally at the stage of academic research, while the ethics of technologies such as smartphones, robots and artificial intelligence are more subject to recommendations from States or to enterprise self-regulation. Certain regulatory frameworks exist or are under development concerning platforms (e.g. Digital Services Legislation, European Union) and their related subjects such as data privacy (e.g. GDPR) and online harms (e.g. NetzDG in Germany).

This Plum Consulting white paper explores the levels of ethical considerations according to the categories of digital technologies, examines the approaches adopted by the international community and details the different tools put in place by private companies. It concludes by proposing a balanced prospective approach based on one hand, on the social value of ethics from a business perspective and on the other hand, on the individual responsibility of users.

1 Ethical challenges in digital technologies

The use of technology can deliver significant benefit to citizens and wider society. Digital technologies provide a direct contribution to the economy, in terms of employment and value-added, and delivering productivity improvements where used in other sectors. Users benefit from digital technologies and supported services, and social value from communication networks.

In some circumstances, however, technology can have harmful side effects or be used for harmful objectives. Notable examples include the Cambridge Analytica scandal, when 50 million Facebook users' personal data was acquired by Cambridge Analytica without users' consent and profiles then harvested, analysed and targeted during political elections[1]; and concerns around facial recognition, owing to potential (racial) bias in algorithms and targeted use of the technology which may be discriminatory to particular racial groups[2]. The use of migration control technologies within immigration and deportation decisions has been heavily condemned, following potential human rights infringements due to faulty algorithms or mandatory iris scanning to access refugee services (which does not meet voluntary and informed consent)[3].

It is important to identify that the design and application of technology in some business models may be anti-ethical. For example, firms may gather and analyse data in order to tailor its services or products to users. Social media and video sharing platforms (VSPs) have been criticised for creating political echo chambers or filter bubbles, where the algorithms used for recommendation systems, to curate homepage content or to queue videos are based on the user profile data, network and previous activity. These features were developed to keep users engaged the platform, but this may inadvertently lead to a situation where users are only shown content that supports a particular narrative (which is not considered socially optimal)[4]. Similarly, misinformation and fake news create a societal harm whilst generating ad revenue. Technology is responsible for both exacerbating and tackling misinformation; artificial intelligence and machine learning (AI/ML) content creation and content is hosted by digital platforms, whilst also being used to audit and remove misinformation.[5]

In some cases, firms may use data (intentionally or unintentionally) to discriminate against particular individuals. In November 2019, Apple Card was investigated after complaints of gender discrimination after offering male customers significantly higher credit limits than women with similar or equivalent credit histories. This was linked to an algorithm used by Apple and Goldman Sachs to determine applicants' credit worthiness[6].

The leveraging of ICTs and the emergence of new applications has been accompanied by growing ethical concerns and dilemmas. The pervasiveness of technology in everyday life – its use, design and impact on organisations and individuals – leads to moral questions regarding its role within our society. There is increasing recognition that incorporating ethics into ICTs – either through regulation, by design, or other means – will be necessary to maximise the value delivered to all stakeholders.

[1] Cadwalladr, C., and Graham-Harrison, E. (2019), 'Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach', *The Guardian,* 17 March [online]. Available at: https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election

[2] The Liberty UK paper identifies potential discriminatory use of facial recognition technology at the Notting Hill Carnival, a festival celebrating West Indies culture in the UK, and in the London Borough of Newham, one of the most ethnically diverse areas of the UK where the White British population accounts for 16.7%. Couchman, H., and Bradley, G. (2019), 'Liberty's Briefing on Police Use of Live Facial Recognition Technology', Liberty [policy paper], November [online]. Available at: http://www.libertyhumanrights.org.uk/wp-content/uploads/2020/02/LIBERTYS-BRIEFING-ON-FACIAL-RECOGNITION-November-2019-CURRENT.pdf

[3] EDRi (2020), *The human rights impacts of migration control technologies,* 12 February [online]. Available at: https://edri.org/our-work/the-human-rights-impacts-of-migration-control-technologies/

[4] VOX-Pol (2019), *Chambers of Secrets? Cognitive echo chambers and the role of social media in facilitating them,* 2 October [online]. Available at: https://www.voxpol.eu/chambers-of-secrets-cognitive-echo-chambers-and-the-role-of-social-media-in-facilitating-them/

[5] Center for Digital Ethics & Policy (2017), *Why we should hold Facebook responsible for fake news,* 22 March [online]. Available at: https://www.digitalethics.org/essays/why-we-should-hold-facebook-responsible-fake-news

[6] Vigdor, N. (2019), 'Apple Card Investigated After Gender Discrimination Complaints', *The New York Times,* 10 November [online]. Available at: https://www.nytimes.com/2019/11/10/business/Apple-credit-card-investigation.html

1.1 Types of ethical values

Transparency and definition of the underlying moral values and perspectives is an essential step to develop and implement the ethical process. Ethical theories in moral philosophy include metaethics (nature of ethics and moral reasoning), normative ethics (seeks to determine content of moral behaviour, including deontological, consequentialist and virtue ethics) and applied ethics[7]. Applying ethics to digital technology relates to applied ethics, as it uses normative ethics to assess concrete and real-life examples.

There are other approaches to ethics that can be applied to digital technologies. For example, ethics by design, ethics of usage, or societal which focuses on promoting societal benefits from the use of technology. Each approach and the timing of the evaluation (i.e. at point of design, point of usage, or at end point to assess overall societal impact) may affect the ethical issues considered and the conclusions from the ethical evaluation. This is also likely to be embedded in the design and application of the technology along with other underlying moral views (for example, the societal or individual view of 'good'). As a result, these values may be embedded within the technology – to promote or demote particular moral value – and will manifest from using the technology[8]. For example, HTTP cookies allow internet users to be tracked and for content to be personalised though this demotes the value of online privacy. Therein lies a tension between the designer/technologist and the end-user/society; the advantages (and costs) face by one party may be contrary to the others.

Moral dilemma is a prevailing challenge for ICTs. Technologists (and users) will decide between competing ethical values. There is always a tension when dealing with ethics in digital technologies which is due to the potential positive and negative impacts that can vary across stakeholders and vary depending on the contextual use of specific technologies. Perhaps the most prevalent example is data privacy. Analysing users' data may improve overall end-user experience of a digital platform or data gathering may be used to track and monitor individuals, for example by public health Covid-tracing apps, which ultimately will deliver a societal benefit. However, this may be at a cost to the end-user (i.e. individual citizens) and concerns about abuse of private data[9]. This leads to further questions regarding data ownership and access. Accuracy poses a further dilemma. Automated decision making based on collected data can improve speed and reduce human bias; however, inaccurate data (from input errors or incorrectly inferred data) may have an adverse impact on an individual or impair the reliability of data-based decisions. However, in most cases, the use of data or AI can improve decision-making.

1.2 Assessing ethical concerns

These examples highlight the trade-off between the interests and moral values of different stakeholders. Further ethical concerns arising from ICTs are summarised in Figure 1.1. These ethical concerns place the human (and the societal impact) at the centre of technology. However, to address this requires explicit human control of technology and applying ethical values onto technologies[10].

[7] This classification is quite commonly accepted in ethical philosophy. This link is one of the sources: https://iep.utm.edu/ethics/

[8] European Parliamentary Research Service (2020), *Artificial intelligence: From ethics to policy*, study for the Panel for the Future of Science and Technology, European Parliament.

[9] ICO (2020), *Blog: Data protection considerations and the NHS COVID-19 app,* 18 September [online]. Available at: https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2020/09/blog-data-protection-considerations-and-the-nhs-covid-19-app/

[10] Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., and Srikumar, M. (2020), 'Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI', *Berkman Klein Center Research Publication,* No. 2020-1 [online]. Available at: http://dx.doi.org/10.2139/ssrn.3518482

Figure 1.1: Framework for assessing ethical concerns in digital technologies

Ethical concern	Examples	Platforms & apps	AI/algorithm design/big data/facial recognition	Personal & IoT devices Robotics	Cloud services	DLT
Consent & Privacy	Data collection from online platforms & mobile apps Data sharing between entities Anonymisation of personal data	●	◕	◕	◑	◔
Security	Data hack (stolen) from firms & individuals Access to data (personal data) Storage of data (personal data) Access to devices through networks	◕	◔	◕	◕	◔
Accuracy	Data collection from online platforms & mobile apps Data processing Inferred data e.g. from algorithms	◑	●	◔	◑	●
Inclusion, Fairness & Non-discrimination	Inferred data based on characteristics Data processing & algoirthms in decision making Protection or fairness towards demographi groups	◑	●	◔	◔	◔
Protection from or minimisation of online harms	Illegal or harmful content/behaviour accessible online Inappropriate targeting Protection of vulnerable groups	●	◕	◑	◔	◑
Transparency	Purpose & method for data collection, storage, and processing Algorithm purpose & design	◕	●	◑	◑	◑
Accountability	Responsibility for oversight of data collection, storage, & design Accountability of robotics and AI	◕	●	◑	◑	◑

Key: ● very high ethical concern or relevance, ◕ high, ◑ moderate, ◔ limited, ○ none.

There has been significant growth in the number of ethical frameworks for ICTs in recent years, particularly relating to AI and data privacy[11]. Proponents of an ethical approach consider ethics to provide a framework in which to assess the purpose, design and impact of technology upon society. However, some critics argue that the ethical frameworks are a 'soft' option to formal regulation, providing structure and weight to self-regulatory initiatives[12]. A paper for the European Parliament identifies that ethics cannot be "reduced to codes of conduct, guidelines or principles exclusively. Rather, ethics should also be understood as a continuous process" [13] which requires ongoing reflection and refinement. The authors argue that the ethical process is to uncover ethical issues and identify the governance mechanism required (i.e. 'soft' frameworks or 'hard' regulation) to address emerging issues for digital ethics in a timely manner.

[11] These being the ethical values promoted by society in general and being held in a non-technological environment.
[12] Wagner, B. (2018), 'Ethics as an escape from regulation. From "ethics-washing" to ethics-shopping?', In Bayamlioğlu, E., Baraliuc, I., Janssens, L. (eds.): *Being Profiled: Cogitas Ergo Sum. 10 Years of 'Profiling the European Citizen'*. Amsterdam: Amsterdam University Press 2018, S. 84–88
[13] European Parliamentary Research Service (2020).

2 From academic research to regulation

Protecting individuals, both as citizens and as consumers, while promoting the national economy and businesses are states' main aims regarding ethics in digital technologies. What is meant by these issues nevertheless varies across countries. In China, the culture of common good makes private data a natural resource of the digital economy, and sharing personal data to contribute to the common good is accepted. On the contrary, European countries have had a strongly embedded culture of individual and privacy rights[14] for centuries: protecting citizens does not have the same meaning across countries and cultures. The USA navigates between commitment to performance (through more personal data collected) and commitment to freedom (through less personal data collected), which may explain the existing disparities between individual American states.

In this context of strong cultural differences, numerous guidelines and frameworks have been published in the past few years by countries and by international organisations. The majority have a technological focus (data, AI, platforms) and include a section that covers the related ethical issues.

The maturity of the ethics approach, at international and state level, largely depends on the type of technology.

- Ethics in **blockchain** and in **cloud services** have started to be analysed in academic research.
- Technologies like **devices**, **robots** and **AI** seem to attract a greater degree of consideration, being subject to recommendations, guidelines, and self-regulation. It is important to note here that national and international frameworks presented below promote ethics in the design, development and/or use of digital technologies, but most of them do not provide for any monitoring or enforcement mechanism.
- Only laws, regulations and court decisions would enable oversight and enforcement of public and private stakeholders, whether it be a general rule (e.g. civil code), a digital specific rule (e.g. GDPR), or a sector-specific law or regulation (e.g. health). Thus, at a third level, **platforms** and their related topics of data privacy and online harms have started to reach regulatory or legal levels.

We are still at a nascent stage of ethical considerations in digital technology overall; and geographies are not at the same level of consideration and implementation on the subject. The sections below offer a non-exhaustive review of tools implemented by states and international organisations, to promote ethics in digital technologies.

Ethical issues identified for each technology are not necessarily unique to this technology. Therefore, overlaps appear. For instance, privacy and data ownership are a concern for cloud services, robots, IoT, AI as well as platforms. It is also important to note that – because it uses algorithms and data while being used as a background technology to make other technologies work or make them more efficient (e.g. in platforms, apps, IoT) – AI is the central subject of many recommendations, and guidelines described below.

2.1 Cloud services and blockchain: Academic research on ethics

Cloud services and Software as a Service (SaaS) enable a flexible usage of a service based on the following elements: computing, network, and storage. This technology has been studied lately from an ethical perspective, mostly by academia[15], through the themes of privacy, data ownership, data control, security (related to intellectual property for instance) and thus trust (of user to the cloud service supplier). Some conclusions drawn cover the mutual informational obligations between hosting companies and service providers about consumer

[14] For food for thought on cultural and philosophical impact on tech, read: Koenig, G. (2019), *La fin de l'individu, Voyage d'un philosophe au pays de l'intelligence artificielle*.
[15] Faragardi, H. R. (2017), 'Ethical considerations in cloud computing systems', Summit of digitization for a sustainable society, Sweden; de Bruin, B., and Floridi, L. (2017), 'The ethics of cloud computing', Science and Engineering Ethics.

privacy, service reliability, data mining and data ownership. Potential restrictions from governments would not be necessary, but some specific companies (banks, law firms, hospitals) would have to follow more stringent regulations.

Furthermore, the word **"blockchain"** has very recently appeared next to the word "ethics" in academic papers. Blockchain has first been known as being adopted in finance with cryptocurrencies. Coupled with IoT, the smart contract feature of blockchain leads to more transparency in international trade of goods and supply chain management (environmental sustainability, respect of human rights in the supply chain). But ethically unfavourable applications of blockchain have been identified, like blockchain-based cryptocurrencies enabling illegal transactions (money laundering, black markets for drugs, illegal pornography, weapons), assassination markets where blockchain enables anonymous bets or *cryptojacking*[16]. Because it creates decentralized organisations, it raises the issue of who should be responsible. There is a void of research and debate on ethics of blockchain and understanding its potential applications and its socio-technical challenges is the current key topic.

2.2 Devices and robots: From recommendations to laws

In 2018, the French regulator ARCEP raised awareness among policymakers and industry players regarding smartphones acting to limit freedom of choice for users and thus imposing limits to an open internet[17,18]. The Digital Services Act package, an EU draft law published in December 2020, that looks to monitor what is referred as "structuring platforms" and could ban for instance:

- the binding pre-installation of applications on **mobile phones and computers**,
- the automatically imposed applications stores, like Apple Store or Google Play.

Academia had paved the way in the ethics of building and using **robots** as from 2006-2008. In 2017, the UNESCO's "Report of COMEST[19] on robotics ethics"[20], written on the usage of robots in the society (industry, military and civilian, transportation, health and welfare, education, household, agriculture and environment), raised the related ethical challenges, and made recommendations on robotics ethics. This included, for instance, the retraining of the workforce, the importance of public debate and of environmental impact assessment, as well as the need to extend the work to study IoT ethics. Autonomy, decision ability, learning, legal status of robots, security and data protection are ethical subjects related to robotics that have been under study in the past few years. The European Parliament resolution of 12 February 2019 covers both AI and robotics. It underlines that both should respect national, European and international laws. It also calls for the creation of a European code of ethics and for the EC to build a framework in which users' consent would be necessary for their personal data to be used[21].

[16] Cryptojacking is a malicious form of cryptomining or cryptocurrency theft. It is the unauthorised use of someone else's computer to mine cryptocurrency.

[17] ARCEP (no date), *Nos sujets: L'ouverture des terminaux (smartphones, assistants vocaux…)* [online]. Available at: https://www.arcep.fr/nos-sujets/nos-sujets-terminaux-ouverts.html

[18] ARCEP (2018), 'Les terminaux maillon faible de l'ouverture d'Internet', ARCEP report, February [online]. Available at: https://www.arcep.fr/uploads/tx_gspublication/rapport-terminaux-fev2018.pdf

[19] World Commission on the Ethics of Scientific Knowledge and Technology (COMEST)

[20] COMEST (2017), *Report of COMEST on Robotics Ethics,* SHS/YES/COMEST-10/17/2 REV., Paris, 14 September [online]. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000253952

[21] *Résolution du Parlement européen du 12 février 2019 sur une politique industrielle européenne globale sur l'intelligence artificielle et la robotique,* European Parliament, Strasbourg, 12 February [online]. Available at: https://www.europarl.europa.eu/doceo/document/TA-8-2019-0081_FR.html

2.3 Artificial Intelligence: Guidelines and recommendations

Following its 2018 "AI for Europe" report, the European Commission has released in April 2019 its "Ethics Guidelines for Trustworthy Artificial Intelligence", developed by a group of independent experts based the feedback of 500 contributors. These Guidelines' objective is to "promote Trustworthy AI", which would be a lawful, ethical, and robust AI: "These guidelines are addressed to all AI stakeholders designing, developing, deploying, implementing, using or being affected by AI, including but not limited to companies, organisations, researchers, public services, government agencies, institutions, civil society organisations, individuals, workers and consumers. Stakeholders committed towards achieving Trustworthy AI can voluntarily opt to use these Guidelines as a method to operationalise their commitment, in particular by using the practical assessment list of Chapter III when developing, deploying or using AI systems. This assessment list can also complement – and hence be incorporated in – existing assessment processes."

Figure 2.1: Seven requirements of EU Ethics Guidelines for Trustworthy Artificial Intelligence[22]

In May 2019, the OECD adopted the first international standards for a responsible AI, the "OECD Principles on AI"[23]. These principles were validated by OECD countries, along with Argentina, Brazil, Costa Rica, Malta, Peru, Romania, and Ukraine. The OECD recommendations coming out from these principles are not legally binding but may influence national legislation. In June 2019, G20 countries, in their Ministerial Statement on Trade and Digital Economy, stated their support to the OECD Principles on AI.

More recently in June 2020, the Global Partnership on AI (GPAI), an international initiative was launched by France and Canada, and supported by Australia, the European Union, Germany, India, Italy, Japan, Mexico, New Zealand, the Republic of Korea, Singapore, Slovenia, the United Kingdom and the United States of America. The GPAI will involve experts from civil society, industry, governments, and academia to discuss responsible use of AI and data governance, as well as collaborate and fund each other's ideas.

In the past few years, numerous guidelines and frameworks on AI have been published by various countries and groups of countries. Europe and the US have been early adopters to the inclusion of ethics in digital technologies. In 2016, the US National Science and Technology Council published "Preparing for the Future of

[22] European Commission (2019), *EU Ethics Guidelines for Trustworthy Artificial Intelligence*, policy paper from the High-Level Group on AI, 8 April [online]. Available at: https://ai.bsa.org/wp-content/uploads/2019/09/AIHLEG_EthicsGuidelinesforTrustworthyAI-ENpdf.pdf
[23] OECD (no date), *What are the OECD Principles on AI?* [online]. Available at? https://www.oecd.org/going-digital/ai/principles/

AI" which promoted ethics by design[24]. In Europe, the French Prime Minister assigned the mission "For a meaningful AI" in March 2018 and the government is working toward greater explicability of algorithms[25]; Germany covered ethics in its November 2018 AI strategy[26] and the UK Centre of Data Ethics and Innovation created the AI Barometer[27] with the objective to maximise benefits and minimise risks of AI, as well as more recently, the ICO (UK data regulator) has released its "Guidance on AI and data protection"[28].

China has entered the debate on ethics later, despite its huge investments and technical advance in AI[29]. Its "New generation AI development plan" identified, in 2017, the need to carry out ethics research and to formulate ethical norms to promote AI development. Its "Governance principles for a new generation of AI", released in June 2019, provides action guidelines[30].

The matter is gaining traction in other geographic areas, with, for instance, the "AI in Mexico" national strategy (2018), the "AI Principles and Ethics", in UAE (2019), the "Principles to promote FEAT AI in the financial sector" in Singapore (2019) and the "Social principles of human-Centric AI"[31] in Japan (2019).

2.4 Platforms, data privacy and online harm

Although they are not explicitly specific to ethics, some regulations have been or are being implemented regarding data privacy and online harms related to platform use.

Data privacy. The 2018 European Union General Data Protection Regulation (GDPR) relates to data protection. It gives no clear guarantees on any ethical groundwork but offers some explainability to consumers. It has influenced other regulations, such as the California Consumer Privacy Act (January 2020).

Online harm. Introduced in September 2017, the German law "NetzDG" requires social networks to remove hateful content within 24 hours after reporting. In cases where the illegality is not obvious, the platform has a week to react. Offenders are exposed to a fine of up to 50 million euros. A similar law was submitted in France in March 2019, but a key requirement that platforms should remove hateful content within 24 hours was eventually removed, in order to comply with freedom of expression. More recently, the French ministry proposed in October 2020 a collaboration between internal security services (Pharos platform) and platforms to make fight against cyber terrorism more efficient.

Platforms under watch of governments. Australia has an "e-safety commissioner"[32] which helps Australians who are experiencing online abuse to make a complaint, and can order social media firms (with legal force) to take down certain content. In mid-2019, the American Ministry of Justice launched an anti-trust investigation against big tech companies, which recently concluded on the unprecedented financial and influence power, and anti-competitive practices of these platforms, and asked for strong measures against big tech companies, up to their dismantling. In October 2020, France and Netherlands sent a joint document[33] to the European

[24] Ethics can help practitioners understand their responsibilities to all stakeholders, but ethical training needs to be augmented with the technical capability to put good intentions into practice by taking technical precautions as a system is built and tested.
[25] France Government (no date), Expliquer les algorithms publics [online]. Available at: https://guides.etalab.gouv.fr/algorithmes/
[26] German Federal Ministries of Education, Economic Affairs, and Labour and Social Affairs
[27] UK Government (2020), *CDEI AI Barometer,* independent report from the Centre for Data Ethics and Innovation, published 18 June, last updated: 23 June [online]. Available at: https://www.gov.uk/government/publications/cdei-ai-barometer
[28] ICO (no date), *Guidance on AI and data protection* [online]. Available at: https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/guidance-on-ai-and-data-protection/
[29] Sharma, Y. (2017), 'Robots bring Asia into the AI research ethics debate', *University World News,* 24 November [online]. Available at: https://www.universityworldnews.com/post.php?story=20171124143449109
[30] Zhang, L. (2019), 'China: AI Governance Principles Released', *Library of Congress,* 9 September [online]. Available at: https://www.loc.gov/law/foreign-news/article/china-ai-governance-principles-released/
[31] Government of Japan, Cabinet Office, Council of Science, Technology and Innovation.
[32] Australian Government and eSafety Commissioner: https://www.esafety.gov.au/
[33] Government of the Netherlands (2020), *Considerations of France, Belgium and the Netherland regarding intervention on platforms with a gatekeeper position,* 15 October [online]. Available at: https://www.government.nl/documents/publications/2020/10/15/considerations-of-france-and-the-netherlands-regarding-intervention-on-platforms-with-a-gatekeeper-position

Commission to ask for stricter rules to monitor *'gatekeepers'*. The EU, through the Digital Services Act, is planning to ask big tech companies for more transparency in users' data collected and to implement new regulations on fake news and harmful content. Regarding data breaches, the most recent example at the time of writing is the £18.4 million imposed by the UK Information Commissioner's Office to Marriott International Hotels for failing to keep customers' data secure following a cyberattack that impacted 339 million guest records[34].

[34] ICO (2020), *ICO fines Marriott International Inc £18.4 million for failing to keep customers' personal data secure,* 30 October [online]. Available at: https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2020/10/ico-fines-marriott-international-inc-184million-for-failing-to-keep-customers-personal-data-secure/

3 Companies' internal governance and self-regulation

More and more companies are tackling the issue of ethics in digital technologies internally. One of their challenges in this area is to meet customers' expectations on the product or service delivered while protecting trust of customers, business partners and governments. For example, consumers may wish to be shown their location on a map application as soon as it opens but may not wish to be tracked constantly. This is about finding a trade-off between innovation, customer experience and ethics. A plethora of tools have been developed both in tech and in non-tech companies, to share knowledge and implement ethics and responsibility in the design, the building, the use and/or the management of their tech products and services. The types of technologies that have been approached by companies with an ethical view are mostly platforms, data and AI, with obvious strong overlaps between them.

Companies differ more on the level of detail of their guidelines, of their recommendations and implementation, than on the type of ethics subjects they cover. In terms of the latter, they seem to define their ethical standards from national and international frameworks and guidelines. As illustrated in Figure 3.1 below, most of the OECD AI principles for instance are covered by companies' internal frameworks.

3.1 Principles and guidelines

Companies have developed internal codes of conducts, providing recommendations and actions to take to their employees. Google's "AI principles"[35] (2018) and its "Responsible AI Practices"[36] look for sharing research results, incorporating findings in their practices, and adapting over time. Microsoft AI Principles[37] (2018) have paved the way for operationalizing responsible AI. ARM focuses on the ethics by design in their "Trust manifesto"[38] (2019).

Figure 3.1: Comparison of OECD AI principles with AI principles of selected companies

OECD	Telefonica[39]	IBM[40]	Telia[41]	ARM[42]	Microsoft
AI should benefit people and the planet by driving inclusive growth, sustainable development and well-being.	"human-centric AI"	"fairness"	"responsible & value centric" "human centric"	5/"human safety must be the primary consideration in the design of any AI system"	"inclusiveness"

[35] Google blog (2018), *Google AI Principles updates, six months in*, 18 December [online]. Available at: https://www.blog.google/technology/ai/google-ai-principles-updates-six-months/
[36] Google AI (no date), *Responsible AI Practices* [online]. Available at: https://ai.google/responsibilities/responsible-ai-practices/
[37] Microsoft (no date), *Responsible AI* [online]. Available at: https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1:primaryr6
[38] arm AI (2019), 'Trust Manifesto', 6 November [online]. Available at: https://www.arm.com/blogs/blueprint/wp-content/uploads/2019/11/Arm-AI-Trust-Manifesto-2019.pdf
[39] Telefonica (2018), 'AI Principles of Telefonica', October [online]. Available at: https://www.telefonica.com/documents/364672/143939197/principios-ai-eng-2018.pdf/f3e86fb8-d0c3-a8ac-a2c5-6aae553e71a3
[40] IBM (2019), 'Everyday Ethics for Artificial Intelligence' [online]. Available at: https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf
[41] Telia Company (2019), 'Guiding Principles on Trusted AI Ethics', January [online]. Available at: https://www.teliacompany.com/globalassets/telia-company/documents/about-telia-company/public-policy/2018/guiding-principles-on-trusted-ai-ethics.pdf
[42] arm AI (2019).

OECD	Telefonica[39]	IBM[40]	Telia[41]	ARM[42]	Microsoft
AI systems should be designed in a way that respects the rule of law, human rights, democratic values and diversity, and they should include appropriate safeguards – for example, enabling human intervention where necessary – to ensure a fair and just society.	"fair AI"	"fairness"	"right respecting" "fair & equal"	2/"every effort should be made to eliminate discriminatory bias in designing and developing AI decision systems" 6/"we will support efforts to retrain people from all backgrounds to develop the skills needed for an AI world"	"fairness"
There should be transparency and responsible disclosure around AI systems to ensure that people understand AI-based outcomes and can challenge them	"transparent and explainable AI"	"explicability" "user data rights"	"transparent & explainable"	3/"AI should be capable of explaining itself as much as possible" 4/"users of AI systems have a right to know who is responsible for the consequences of AI decision making"	"transparency"
AI systems must function in a robust, secure and safe way throughout their life cycles and potential risks should be continually assessed and managed.	"privacy & security by design"	"user data rights"	"control" "safe & secure"	1/"we believe all AI systems should employ state-of-the-art security" 5/"human safety must be the primary consideration in the design of any AI system"	"reliability & safety" "privacy & security"
Organisations and individuals developing, deploying or operating AI systems should be held accountable for their proper functioning in line with the above principles	Not specified	"accountability"	"accountable"	Not specified	"accountability"

In a document aimed at AI designers and developers[43], IBM focuses on five areas of ethics: accountability, value alignment, explicability, fairness, and user data rights, through the idea that "ethical decision-making is not just another form of technical problem solving" and proposes internal recommendations. Telecommunication operators have also followed the trend, such as Telefonica (Spain) and its "AI Principles"[44] (2018) and Telia (Sweden) (2019) and its "Guiding principles on trusted AI ethics"[45]. Most of these companies regularly review and develop their principles; continuous review is part of Telia's guidelines for instance (Figure 3.2).

Figure 3.2: Telia company guiding principles on trusted AI ethics

An example of how platforms tackle harmful content issues is TikTok[46]. The Chinese video sharing platform, created by Douyin for non-Chinese markets and launched in 2016, has published Community Guidelines[47] on violence, hate speech, harassment among others, and its Trust & Safety teams moderate and remove content that is not compliant to these guidelines. TikTok also works on increasing transparency about the reasons why content has been removed in order to inform users "how their actions could be hurtful to others". TikTok recently joined other platform by signing the EU Code of Practice on Disinformation.

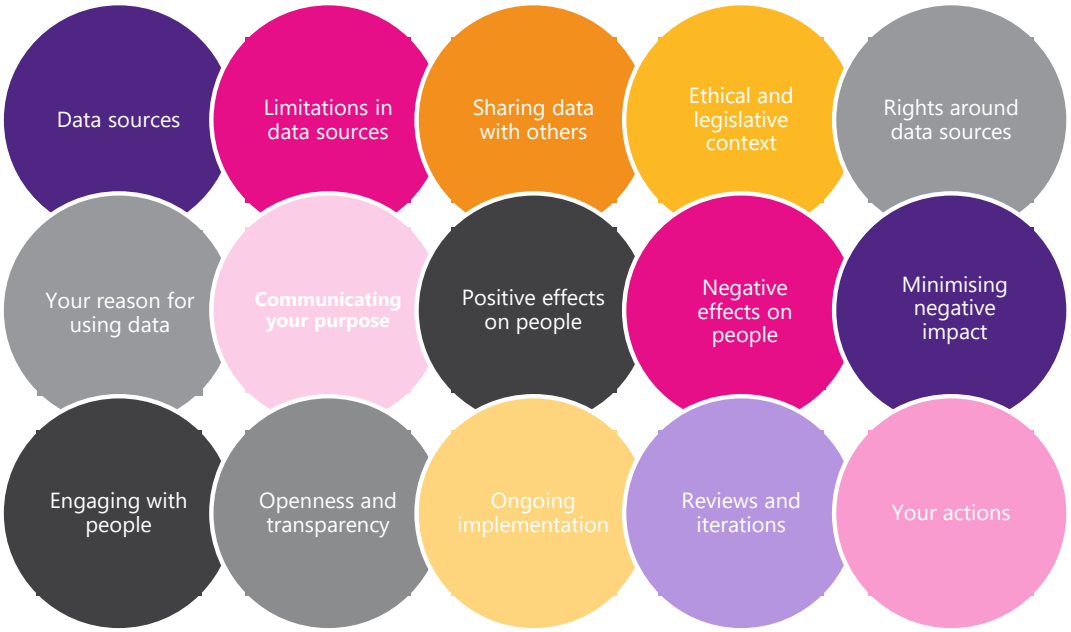
3.2 Toolkits

Open-source toolkits aimed to developers and data scientists to test fairness of their systems and mitigate bias, have been created by companies (e.g. IBM's AI Fairness 360[48], Microsoft's Fairlearn[49]) as well as by academia

[43] IBM (2019).
[44] Telefonica (2018).
[45] Telia Company (2019).
[46] TikTok (2020), *Countering hate on TikTok.* 21 October [online]. Available at: https://newsroom.tiktok.com/en-gb/countering-hate-on-tiktok-gb
[47] TikTok (no date), *Community Guidelines* [online]. Available at: https://www.tiktok.com/community-guidelines?lang=en
[48] IBM Research (no date), *AI Fairness 360* [online]. Available at:https://aif360.mybluemix.net/
[49] Microsoft (2020), *Fairlearn: A toolkit for assessing and improving fairness in AI*, May [online]. Available at: https://www.microsoft.com/en-us/research/publication/fairlearn-a-toolkit-for-assessing-and-improving-fairness-in-ai/

(e.g. Stanford Policy Lab's Fair ML[50]). The ODI (Open Data Institute) network has created a Data Ethics Canvas[51] (Figure 3.3). Its objective is to help identify and manage ethical issues for anyone collecting, sharing, or using data. Among non-tech companies, the case of Rolls-Royce, developing an open-source framework on ethical AI for industry is notable. The objective is twofold: ensuring that decisions taken to use AI in the company are ethical and checking that the outcomes of algorithms can be trusted and are not biased. The system has been peer-reviewed by tech- and non tech-company experts and is one of the first cases of practical application of AI ethics in industry. Regarding online harms, YouTube has for instance has launched a fact-checking tool that relies on an open network of third-party publishers.[52]

Figure 3.3: ODI Data Ethics Canvas



3.3 Training and awareness

Training programs, designed for managers and developers for instance, are set up to raise internal awareness about ethical issues related to technologies. Telefonica has a 6 module-course on AI and Ethics available in three languages (Spanish, English, and Portuguese).

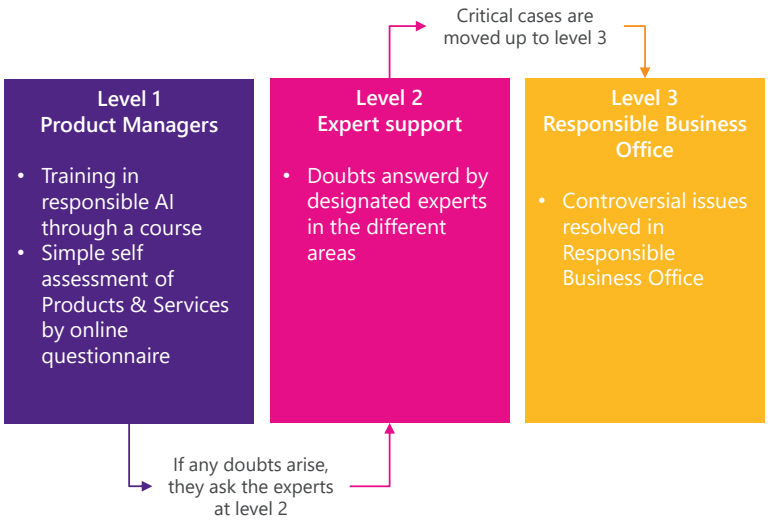
3.4 Governance

Ethics committees have been implemented in some companies. Google's ethics committee[53] – Advanced Technology External Advisory Council (ATEAC) – closed in 2019[54] in a publicised failure. Microsoft has created its internal committee "AETHER – AI, Ethics, and Effects in Engineering and Research" and an Office of Responsible AI. Other companies appoint a person in charge of developing and ensuring ethics in the development and use

[50] Corbett-Davies, S., and Goel, S. (no date), 'Defining and Designing Fair Algorithms', *Fair ML: An ICML tutorial,* Stanford Policy Lab [online]. Available at: https://stanford-policylab.github.io/fairML/
[51] The ODI (2019), 'Data Ethics Canvas', May [online]. Available at: https://theodi.org/wp-content/uploads/2019/07/ODI-Data-Ethics-Canvas-2019-05.pdf
[52] YouTube (2020), *Expanding fact checks on YouTube to the United States,* 28 April [online]. Available at: https://blog.youtube/news-and-events/expanding-fact-checks-on-youtube-to-united-states
[53] Lee, D. (2019), 'Google announces AI ethics panel', *BBC News,* 26 March [online]. Available at: https://www.bbc.com/news/technology-47714921
[54] Johnson, B., and Lichfield, G. (2020), 'Hey Google, sorry you lost your ethics council, so we made one for you', *MIT Technology Review,* 6 April [online]. Available at: https://www.technologyreview.com/2019/04/06/65905/google-cancels-ateac-ai-ethics-council-what-next/

of digital technologies, like at CNP Insurances[55]. Telefonica has a different approach to the question with a self-responsibility approach with a three step-escalation, shown below[56].

Figure 3.4: Telefonica escalation process on AI ethical issues



3.5 Employees' initiatives

There have been some cases of employees protesting against their employer for ethical reasons. In 2018, thousands of Google employees have signed a letter to their CEO protesting company's work in a Pentagon program ("Project Maven"[57]) that would have weaponized the use of AI. This issue has raised based on the tension between the increasing demand for AI technology for military purposes and the ethos of some of Google's employees, "don't be evil" being the company's motto.

3.6 Cross-companies initiatives, PPPs, and support to research

The "Partnership on AI" brings together major technology companies[58], start-ups, academics and specialists in policy and ethics, in a non-profit consortium "on a mission to shape best practices, research, and public dialogue about AI's benefits for people and society to come up with ethical standards for researchers in AI. The ITI (Information Technology Industry Council), a trade association representing the ICT industry, has published its AI Policy Principles[59] which promote, among other things, the use of PPP to "democratize access" and "prioritize diversity and inclusion". Private companies contribute to independent AI Ethics research. For example, Facebook collaborates with the University of Hong Kong to support research on AI ethics in Asia Pacific[60].

[55] Telecom Paris (2020), *Une IA éthique au service du secteur de l'assurance,* 24 July [online]. Available at: https://www.telecom-paris.fr/ia-ethique-service-secteur-assurance

[56] Telefonica (2020), 'Telefónica's Approach to the Responsible Use of AI' [online]. Available at: https://www.telefonica.com/documents/1258915/3538310/ia-responsible-governance.pdf/2e6fe3e7-5a66-718d-177d-430f5a12b963

[57] Shane, S., and Wakabayashi, D. (2018) ''The Busines of War': Google Employees Protest Work for the Pentagon', *The New York Times,* 4 April [online]. Available at: https://www.nytimes.com/2018/04/04/technology/google-letter-ceo-pentagon-project.html

[58] Partnership on AI (no date), *Meet the Partners* [online]. Available at;: https://www.partnershiponai.org/partners/

[59] ITI (no date), 'AI Policy Principles Executive Summary' [online]. Available at: https://www.itic.org/dotAsset/50ed66d5-404d-40bb-a8ae-9eeeef55aa76.pdf

[60] Facebook Research (no date), *Facebook announced award recipient of the Ethics in AI Research Initiative for the Asia Pacific* [ONLINE]. Available at: https://research.fb.com/blog/2020/06/facebook-announces-award-recipients-of-the-ethics-in-ai-research-initiative-for-the-asia-pacific/

Regarding fact checking and misinformation, Whatsapp has partnered with World Health Organization to create the Whatsapp Coronavirus Information Hub to advise people on checked sources on the virus[61].

3.7 Ethics-washing

These approaches, particularly when taken by billion-dollar tech companies, can sound hypocritical. Their attempts to appear as concerned by the ethical issue may be an attempt to push back from any new regulation by showing customers and regulators that their products or services are trustworthy. This is called ethics-washing and it usually comes from a lack of willingness to tackle ethical issuesor from public relations concerns where the overriding objective is to monitor the reputational risk of the company. There is also the misconception that the management of ethical issues could conflict with innovation and competitiveness. However, it seems that tech consumers and users are becoming more and more sensitive to these issues and businesses would have interests in acting authentically in that matter to remain competitive.

Furthermore, if platforms have signed the EU Code of Practice on Disinformation in late 2018, some studies argue that the Code should be strengthened, particularly on implementation and monitoring[62].

[61] Rochefort, M. (2020), 'Covid-19 : WhatsApp annonce un partenariat inédit pour lutter contre les fake news', *SiecleDigital,* 19 March [online]. Available at: https://siecledigital.fr/2020/03/19/whatsapp-coronavirus-oms-unicef/

[62] https://ec.europa.eu/digital-single-market/en/news/study-assessment-implementation-code-practice-disinformation

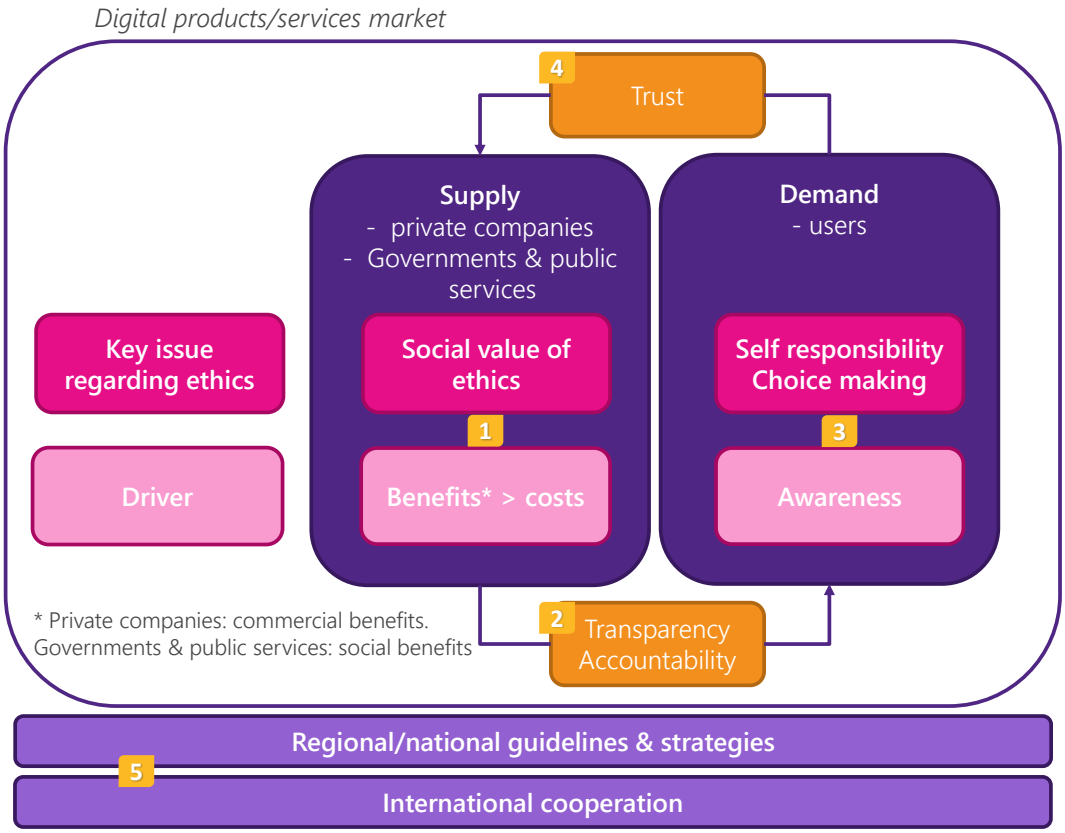
4 Proposed approach for and challenges to ethics in digital technologies

In this concluding section, we propose a two-fold approach to tackle public concern around ethics and to address the social impact of these technologies use by private enterprise and states. We then highlight the key challenges that policy makers and private companies face when addressing ethical concerns relating to digital technologies.

4.1 Proposed approach

We set out a two-fold approach to tackle public concern around ethics and to address the social impact of these technologies use by private enterprise and states. This approach, set out below and illustrated in Figure 4.1, depends on supply, which includes private companies as well as public services and Government putting digital tools available to citizens (e.g., Covid tracing apps), and demand, which includes all types of users (for instance, consumers of digital services, citizens using public e-services, businesses using digital services from other businesses, governments using services from private companies).

Figure 4.1: Two-fold approach of recommendations



Digital technologies must service individual people and society; this is the central principle of our two-fold approach. Nevertheless, the majority of digital technologies are designed and used for commercial benefits rather than for social good. Transitioning to a scenario which places ethics at the centre of the development and use of digital technologies relies on feedback effects, as explained below.

1. Supply of digital technologies (and/or services underpinned by these technologies) is done by private companies to consumers and by Governments and public organisations to citizens. Incorporating ethics – through frameworks, governance and self-regulation – will benefit consumers and deliver broader societal benefits. It is, however, essential that companies are rewarded for this behaviour (i.e. their private benefit must exceed the cost from integrating ethics).

2. The most effective way for private companies (or public actors) to communicate this to consumers (citizens) is by increased transparency and accountability. Making ethics visible will build trust. This will also require for ethics to be 'measurable' and allows for comparison between suppliers.

3. Users of digital services and (underpinning) technologies also face ethical dilemmas from self-responsibility and choice making. Driving awareness of ethics – through education and data literacy – is key in empowering users. Users consumption and interaction with specific digital technologies/services will likely be influenced by the ethics of the private companies and public organisations.

4. User trust in ethical companies will increase engagement and consumption of digital technologies/services. This will reward the companies and organisations most invested in the ethics approach and that effectively communicate this (via transparency and accountability).

5. The regulatory environment and international cooperation on ethics in digital technologies underpin this approach. These can affect the drivers to implement ethics for private companies and users. For example, subjecting private companies to GDPR regulation will increase the social value delivered from better handling of personal data.

4.2 Challenges to implementing ethics

To conclude, we highlight four key considerations to this or any alternate ethics approach.

Coping with rapid evolution of digital technology. Ensuring that ethical frameworks and tools are adaptable and remain relevant to evolving technologies is a central on-going challenge. A flexible and forward-looking approach should be central in the development of governance tools – from formalised legislation and regulation to self-regulatory initiatives, such as oversight by industry bodies and self-certification. Nonetheless, development of technology will outpace these governance tools, pose new ethical dilemmas, and will require continuous review of governance tools' efficacy and effectiveness.

Different cultural perspectives. Social norms and ethical values vary across countries and cultures. Given the very nature of the Internet and digital technologies, implementing ethics necessitates international cooperation on standards and governance whilst also allowing for diverse cultural perspectives.[63] This inherent tension, which is exacerbated by mistrust and practical coordination challenges, can only be overcome through cross-cultural cooperation and exchange of best practices. ÓhÉigeartaigh *et al* (2020) argue that the potential barriers to cross-cultural cooperation tend to arise from cultural mistrust, rather than fundamental disagreements.

Managing different cultural values is a dilemma for multinational tech companies that provide services and products in different geographies. Multinationals can either adopt a fragmented approach, whereby the ethical principles and decisions are different across geographies, or they may choose to adopt the highest standard. For example, Microsoft have adopted the latter approach with regards to personal data and the EU's GDPR law, including lobbying for similar legislation in the USA.

[63] ÓhÉigeartaigh, S., and Whittlestone, J., and Liu, Y., Zeng, Y., and Liu, Z. (2020), 'Overcoming Barriers to Cross-cultural Cooperation in AI Ethics and Governance', *Philosophy & Technology*, 33, pp571-591, May [online]. Available at: https://link.springer.com/article/10.1007/s13347-020-00402-x

States and enterprises are both judge and party. Ethics frameworks and governance tools (for examples, principles, initiatives and legislation) for digital technologies are developed often by the actors to whom they apply. As highlighted by the initiatives above, ethics has been a growing preoccupation for several large-scale private companies. Although this may reflect a growing sense of corporate social responsibility, with companies signalling that they are trustworthy and responsible to consumers and states, critics have raised two potential issues with this approach.[64]

- Industry support for ethics frameworks and self-regulatory initiatives may be seen as an attempt to avoid more formal regulation or stringent requirements.

- Potential ethics washing, where companies exaggerate their interest in ethics and social impact in order to extract greater public trust (and profit). Ethics washing is a marketing strategy, with companies failing to implement these policies internally efficiently and effectively.

Any possible solutions to these concerns need to ensure real oversight and accountability, for both private and public actors. Top-down approaches may include ethics boards or introduction of ethics audits,[65,66] whereas a bottom-up approach would be to investigate and communicate the commercial, social and environmental benefits of enterprises implementing ethics. There is a research gap in the latter, but it is likely to substantially accelerate progress towards more ethical digital technologies.

Engaging individuals in ethics. A key challenge will be to educate and empower individuals in their use of digital technologies. This is an essential step in aligning incentives for states and private companies to implement ethics. In most instances, the cost of ethics is borne by private companies and will benefit from ethical consumers choice of services and products. Engaged citizens and consumers will also ensure effective public accountability and scrutiny of State and private actors.

[64] Wagner (2018).
[65] MAIEI (2020), *Why We Need to Audit Government AI*, 14 September [online]. Available at: https://montrealethics.ai/why-we-need-to-audit-government-ai/
[66] Arborus (2020), *Lancement du ler label international pour une IA Inclusive*, 7 September [online]. Available at: https://arborus.org/lancement-du-1er-label-international-pour-une-ia-inclusive/

