CODE OF ETHICS

for Data-Based Value Creation

RECOMMENDATIONS

www.data-innovation.org

data innovation alliance

OVERVIEW

The “Code of Ethics for Data-Based Value Creation” consists of the following documents: 1) Overview; 2) Basics; 3) Recommendations; 4) Implementation; 5) Context.

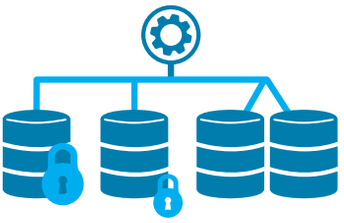
The **RECOMMENDATIONS** document gives concrete recommendations for ethical data-based value creation structured by the four steps of the data life cycle. The recommendations are based on the three basic ethical orientations and the three procedural values, yielding six groups of recommendations for each step. For each group, the intentions of the recommendations are explained first, and the concrete recommendations follow. Each group ends with an example illustrating what implementation of the recommendations might look like. Explanations of the data life cycle and basic orientations can be found in the Basics document. The Recommendations document is available in German, English, French and Italian.

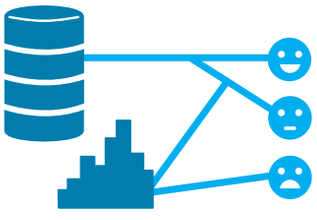
IMPRINT

The “Code of Ethics for Data-Based Value Creation” was drawn up by the “Data Ethics” expert group of the Swiss Alliance for Data-Intensive Services. Editorial team: Markus Christen, Christoph Heitz, Tom Kleiber, Michele Loi (lead editor). French translation: Jean-Gabriel Piguet. Graphics: Ana Nicolasa Caduff. Status: 2020.

ISBN 978-3-9522703-3-2; www.data-service-alliance.ch/codex

TABLE OF CONTENTS





STEP 1: DATA GENERATION AND ACQUISITION

BASIC ETHICAL ORIENTATIONS

HARM AVOIDANCE

Harm avoidance requires ensuring that no damage is caused to the person from which the data is emerging (often the user of a data product) by the way the data are collected. Questions about possible harm may arise if the interaction with the user is designed to encourage the user to reveal more data about himself or herself (such designs may use findings from behavioural research and may be known as 'nudges'). If a user interface is deliberately designed to coerce a user to reveal information that is contrary to his interests, this is called a 'dark pattern'.

Recommendations

- Avoid exploiting the inertia, inattention, or other psychological weaknesses of individuals to obtain more personal information from them than is necessary to provide your service.
- Check web design and user experience for dark patterns and remove them.
- If you use nudges, you should prove in advance or subsequently document that they also serve the interests of the user (e.g. by increasing cyber security or improving the service provided).

EXAMPLE – AVOID DARK PATTERNS

Dark patterns are deployed to mislead the user and – in the worst case – lead the user to take an action he or she did not want to perform. To trigger this 'misbehaviour', users' learned behaviour patterns are exploited: users are manipulated by cleverly placed buttons or misleading drop-down menus. For example, a company wants to sell a travel insurance policy to a customer who has just booked a trip. The option 'No Insurance' is hidden (under 'N') in an alphabetical list of countries to which the insurance could apply. Interfaces should instead support usability by organising interactive features transparently and in accordance with logic. In the travel insurance case, the initial decision ('yes' versus 'no') should be separated from the possible options that a 'yes' implies.

JUSTICE

The recommendations based on justice are intended to ensure reciprocity, so that when data are collected, subjects are offered benefits that are proportionate to the data you request.

Recommendations

- Make clear to your customers the value returned for disclosing their data (e.g. more personalized treatment).
- If the collection of certain data is legally required, make this clear.

EXAMPLE – ELECTRONIC PATIENT DOSSIER (EPD)

A physician with her own practice has registered with the EPD system so that her patients can be treated (even) better in the future. Once a patient has created an individual EPD, the doctor can enter the patient's health and treatment data. However, Patient U. is now unsure whether he wants to open an EPD and is worried that his data could be misused. With reference to the basic orientation of justice, the physician can convince Patient U. as follows:

- She shows the patient the individual benefit: with the EPD, access to documents relevant to treatment is guaranteed at all times, and duplicate or unnecessary treatments can be avoided.
- She makes it clear to the patient that he or she can determine who has access to which health data. Only in an emergency can the EPD be accessed without the patient's consent. The data storage is decentralised and only Swiss servers are used. Information is exchanged only via a secure connection, with all processes being specially certified.
- She points out that patients can easily make their health data available for research if they wish.

AUTONOMY

An orientation toward autonomy means giving users sufficient informational self-determination.

Recommendations

- Provide your customers with highly usable tools for making consent decisions about data usage.
- For this purpose, develop intuitively plausible and, if necessary, user-selectable forms of informed consent.
- Allow users to adapt or revoke their consent later.
- When the circumstances allow it, enable granular preferences regarding the purpose of data processing (e.g. by enabling customers to release certain data for research in other contexts).

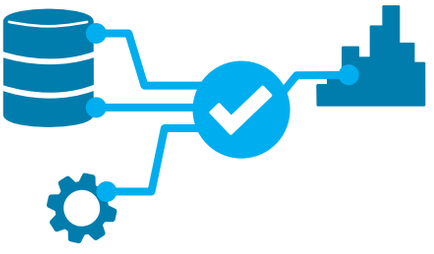
EXAMPLE – DIFFERENT LEVELS OF CONSENT

A web platform delivers information about medical research on specific pathologies, social networking opportunities, and self-tracking services to patient groups. It collects, analyses, and commercialises its user data, but also uses them for its personalised recommendations, which are key to the platform functionality and attractiveness. The business model of the platform is to make both anonymised and personal patient data available to pharmaceutical companies for recruiting patients for trials of new drugs and to partner with both industry and university researchers to do behavioural research (e.g., on attitudes towards therapies and their side-effects). To this end:

- The company provides different levels of consent: patients can consent to their data being used for recommendations without data being shared with the industry.
- Users can review their consent and sharing preferences, at any time. This information is easily reached from the patient dashboard.
- Users can decide to stop sharing data with the industry at any time – while previously shared anonymised data will not be withdrawn, personal data will be locked the moment consent is electronically revoked.

PROCEDURAL VALUES

CONTROL



The recommendations regarding control are intended to ensure that, when data are generated or acquired, all information needed for the ethical assessment of data use is collected.

Recommendations

– Document your data sources and any restrictions on the appropriate use of the data.
– If other companies use your information, provide them with the information they need to assess data origins and restrictions on use.

EXAMPLE – OPEN DATA FOR NEW BUSINESS MODELS

A large transport service provider decides to make the data it collects available to third parties for secondary use free of charge and in machine-readable form (Open Data). The aim is to support creative market participants in developing new, attractive information and service offerings. It is hoped that this will promote new business models with innovative information or service offerings and thus ultimately increase the attractiveness of the company's market environment. The company defines the conditions of use, which may include obligations on the part of the user to state the source of the data, to update the raw data if the company makes such updates available (in order to prevent business models being based on outdated data), and to comply with data protection rules (in particular, to refrain from any attempts to extract de-anonymised information about the company's customers from the data).

TRANSPARENCY

Transparency requires ensuring that informed consent can be achieved in the simplest, clearest, and most accessible way possible.

Recommendations

– Find simple ways to explain what data you collect, your privacy policy, and how you will use customer data (i.e., for what purpose and what results you expect).
– Use summaries and other ways to explain the key points of your Terms and Conditions.
– Work with other companies to create new standards for more readable and understandable Terms and Conditions.

EXAMPLE – PHOTOGRAPHER

A photographer offers special event and personal photo shoots along with various options for purchasing the resulting images. His general terms and conditions (GTC) stipulate that the photos can be made available to customers digitally via an external online service provider so that customers can download them. The photographer can keep the photos for one year after the order has been processed so that customers may reorder prints, and also has the right to use the photos for advertising purposes (on a specific website, on posters, etc.). Copyright remains with the photographer. A customer would like to receive information about the general terms and conditions. The photographer can respond to this request as follows:

– The photographer delivers the GTC to the customer before the contract is concluded. He or she summarises the purposes for which the data will be used and promises not to use the data for other purposes.
– The photographer informs the customer that he or she can delete the photographs after the order has been completed, but that the customer will have to re-upload them if he or she wishes to place another order for prints.
– He explicitly asks the customer whether he or she may use the photographs for his own advertising purposes.
– Finally, the photographer ensures that the external online service provider adheres to the same rules regarding the handling of data (in particular, not using the data for any other purpose) and informs the customer accordingly.

ACCOUNTABILITY

The accountability recommendations are designed to ensure that you can use data collected by third parties in a responsible manner.

Recommendations

- If your product depends on data from other companies, make sure that this data has been collected in an ethical manner.
- Do not use data from untrusted sources.
- Only acquire data from partners who are transparent about both their data collection practices and any restrictions associated with the use of the data.

EXAMPLE – MONETISATION OF SOCIAL NETWORK DATA

A pharmaceutical company considers buying data from a web platform providing social networking and self-tracking opportunities to patient groups. The pharmaceutical company obtains information about the data collection practices of the web platform. Are the data obtained in such a way that the purpose of the data collection and the link with the industry, including the monetisation of data, is transparently and openly communicated to the user? If not, the pharmaceutical company does not acquire data from this platform and does not use its services.

STEP 2: DATA STORAGE AND MANAGEMENT

BASIC ETHICAL ORIENTATIONS

HARM AVOIDANCE

The recommendations for harm avoidance at this step are intended to help ensure that no damage can be done with stored data.

Recommendations

- Implement appropriate cyber security measures and apply them consistently across all forms of data storage (e.g. cloud-based).
- Develop contingency plans to minimize the extent of damage in the event of a data breach.
- Define rules for the deletion of data with the aim of reducing the risk of damage from unauthorized access.

EXAMPLE – HEALTH DATA IN THE CLOUD

A health data analytics company develops customised data collection and analytics tools for large hospitals. The company provides to hospitals a cloud service to host all data, allowing hospitals to focus on patient care and outcome improvement. In order to guarantee the hospital the necessary cyber security, the data analytics company implements the highest standards of cyber security and commits to strict compliance with these standards (e.g., the hospitals' data are accessible only to select employees in the software company using two-factor authentication, and these employees are trained to recognise social engineering attacks).

JUSTICE

Justice requires enabling customers to use their data for the benefit of third parties and the common good.

Recommendations

– Support your customers in effectively using the right to data portability.
– Enable your customers to easily receive and use their data (provided it has not been anonymised) for third-party applications.

EXAMPLE – DATA DONATION

A company that collects and processes customers' personal data in order to provide services wants to make it easier for its customers to donate their own data (if they wish) for valuable purposes (for example, non-profit research projects or public administration projects). By reusing data, the company also wants to contribute to digital sustainability. The following measures could be useful here:

– Setting up an information management system on the company's platform that allows the company's customers to manage their personal data themselves. This includes, in particular, a user-friendly option to download all or selected data in a machine-readable form and export them into other systems.
– Establishing a technical interface with other information management systems and data escrow platforms, so that the company's clients can make their personal data directly available to other service providers.

AUTONOMY

The recommendations related to autonomy are intended to ensure that data storage and management do not pose any risks to the privacy of customers.

Recommendations

- Examine whether your technical measures for protecting the privacy of your customers are appropriate and proportionate.
- Keep abreast of developments in relevant security technologies.
- Ensure that stored anonymised data cannot be traced back to the individuals to whom they refer.
- Offer your customers uncomplicated options for restricting the storage of their data.

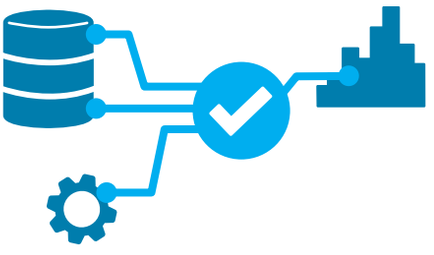
EXAMPLE – VISUALISATION OF DIGITAL SELF-DETERMINATION

A company has put the digital self-determination of its users at the centre of its business activities. It provides a personal information management system and a market-place for personal data. The platform enables its users to make their data profile or selected excerpts thereof available to individual companies, public administrations, or non-profit organisations at a certain price or free of charge; the company obtains a share of the price paid. The users have an overview at all times of which data goes to which recipients at what price and can also cancel these offers. The added value of the platform for the user consists in the strengthening of their digital self-determination through the following elements:

- The users' data, such as purchasing data, mobility data, or health data, are aggregated and visualised. For example, users can see how healthy or sustainable their purchases have been or what their mobility behaviour (e.g., number of steps) has been like in recent months.
- The data is stored in encrypted form and cannot be viewed by the company itself.
- If a customer wants to pass on parts of his dataset to a third party, an automatic check is carried out to determine whether the dataset defined in this way can be traced back to the customer. If the risk of this is high, the customer is warned against disclosure and must explicitly consent to it.

PROCEDURAL VALUES

CONTROL



The recommendations regarding control are intended to ensure that no information relevant for assessing the ethical use of data is lost in the management of data (e.g. data cleaning, deletion of data, etc.).

Recommendations

- Document any changes and adjustments to data management requirements as you move data to different locations, especially when services are outsourced to the cloud.
- Implement technical requirements to ensure traceability of changes to stored data.
- Consider in advance what to do with the data of inactive accounts or former customers.

EXAMPLE – TRACEABILITY OF DATA USE

A company uses socio-demographic data about its customers as well as data from customer interaction (ordering behaviour, correspondence, etc.) to classify its customers into different customer groups by means of data-analytical models, and to forecast customers' future behaviour. These models are used in customer relationship management. In addition, some results of the model-based forecasts are shared with other companies in the same group. This data is subject to constant change. In order to be able to trace at any time which data has been used for further processing (data analyses, forecast models), each dataset is enriched with metadata describing where the data comes from, when the data was passed on to the analysts or model builders, what restrictions on use are linked to the data, etc. This makes it possible to trace every single dataset that has been used for analysis or modelling, while ensuring that analysts and modellers use the data in an appropriate manner.

TRANSPARENCY

Transparency requires ensuring that customers are informed about all aspects of data handling – storage, management, and protection – thus creating (well-placed) trust.

Recommendations

- Make it clear to your customers how their data is protected and how access is logged.
- If it is necessary to store data for a very long time (e.g. for legal reasons), explain this to your customers at the time of data collection.

EXAMPLE – INFORMATION ON DATA RETENTION

The Federal Act on the Surveillance of Post and Telecommunications obliges telecom operators in Switzerland to store and retain the metadata relating to their customers' telecommunications for six months. This is intended to facilitate the clarification of criminal data and the search for missing persons. All Swiss citizens also have the right to access their retained data. A telecom provider draws attention to this fact in an appropriate and detailed manner by indicating which data is retained, including:

- Who was on which website and when, and which apps were used;
- With whom, when, and from where communication took place;
- Who has logged on to the internet, when, and for how long;
- Who sent whom an e-mail or other text message, and when;
- Where a mobile phone user is currently located and the phone's location history for the last six months.

ACCOUNTABILITY

The recommendations regarding accountability are intended to ensure that the roles of all persons involved in data management have been sufficiently clarified.

Recommendations

– Define at the organization level who is involved in the assessment, reporting, and implementation of ethical standards in data management.
– Define clear rules regarding access to stored data (who may have access, when, under what conditions, and how access is logged or tracked).

EXAMPLE – THE ROLE OF THE CHIEF DATA OFFICER

In many organisations the role of the Chief Data Officer (CDO) has become established. The CDO oversees the use of data to ensure that an organisation gets the maximum benefit from its data while maintaining its own standards. These standards typically set forth the entire data management process and thus ensure data quality. Both data management and data quality are addressed in this Recommendation. Although the office of Chief Data Officer is usually established for business rather than for ethical reasons, it is nevertheless one of the duties of the Chief Data Officer to establish and monitor data ethics standards.

BASIC ETHICAL ORIENTATIONS

HARM AVOIDANCE

The recommendations on harm avoidance for this step are intended to ensure that the risk potential of model generation and other data-driven analytics are identified early enough in the process of data analysis and model generation.

Recommendations

– Assess the foreseeable possibilities of misuse of your models.
– In the course of model development, avoid in particular any measures that could lead to the de-anonymisation of the data used.
– Integrate control mechanisms into your models to monitor and limit the potential damage that could result from their practical use; this applies in particular to technologies that may be used to violate privacy or individually harmful consequences of incorrect decisions due to inaccuracies in the model.
– Consider imposing limits on the distribution of the model where appropriate.

EXAMPLE – AI SECURITY ASSESSMENT

One of the key players in the AI field (Deep Mind) has developed a theoretical framework to guide the assessment of the security of its data-based products. The frame-work has three elements: (1) Specification (see the example of Step 3, Control, below); (2) Robustness, which requires identifying environments not represented in the dataset that may lead the AI to misbehave (e.g., a household cleaning robot tested in a pet-free home may perform differently in an environment that includes pet hair) and testing the possibility of building adversarial inputs; (3) Assurance, which consists of two aspects: inspecting systems (summary statistics, automated inspections, and tools designed to make AI decisions explainable or interpretable after use) and enforcement mechanisms (e.g., interruptibility, that is, a reliable off-switch). Source: https://deepmind.com/blog/article/specifying-ai-safety-problems

JUSTICE

The recommendations on justice address the problem of indirect discrimination as it relates to data-based models and predictions. Indirect discrimination occurs when a decision rule is applied to everyone in the same way but has an unjustifiably different effect on different groups of people (e.g., women vs. men). This is particularly relevant for models and forecasts on the basis of which decisions are automatically made, or which support human decision making (e.g., recommendations).

Recommendations

- Ensure that any data set used for machine learning is diverse enough to avoid bias due to poor training data.
- Explain what type(s) of discrimination are intended and built into your model; what type(s) of unwanted, indirect discrimination have been removed from your model, if any; and what form(s) of indirect discrimination are still present, if any.
- Check for the presence of indirect discrimination with test data.
- Develop in-house expertise on how to deal with the indirect discrimination that can result from machine learning; this should also include technical (using statistics and computer science) methods to eliminate such injustices.
- Where appropriate, introduce relevant technologies to prevent indirect discrimination.

EXAMPLE – BIASED DIAGNOSTIC TOOL

An interdisciplinary team of medical scientists produces a predictive model of patients' responses to treatment based on the data from genomics, epigenomics, and proteomics. The model can provide a prediction, for any given patient and proposed treatment, of whether that treatment will be successful for that patient, based on the patient's genomic (etc.) data. However, due to resource constraints, a decision is made to prioritise the patients most likely to benefit. It turns out that 90% of men of European descent, but only 10% of men of African descent are predicted to benefit from the treatment, so there is a disproportion between the people recommended to receive the treatment. This raises the following question: do black people have a reasonable complaint against this inequality? Initially it is argued that the unequal chance of benefiting from the treatment justifies this inequality. Yet, a data scientist points out that it is possible to increase the sensitivity of the prediction for blacks (i.e. a larger proportion of black patients who will be healed will be identified and given the cure), with the collateral effect that this reduces its precision (i.e. a larger proportion of black patients who will not be healed will be wrongly predicted to be healed and given a cure). To evaluate this trade-off, a stakeholder panel involving patient and equality organizations is organized. The panel concludes that the tweaked algorithm is in the interest of the black patients, for whom (unlike white patients) no other effective treatment is available.

AUTONOMY

An orientation toward autonomy requires ensuring that affected parties outside the company are able to make appropriate assessments of the ethical aspects of model-building if these models have an influence on the autonomy of these parties.

Recommendations

– Together with the stakeholders concerned, identify ethical issues relevant to the use of your data-based models or the knowledge gained from the data.
– Enable stakeholders or other interested parties to provide comments on the model development, especially regarding the objectives embedded in the model (e.g., the classification goal or the loss function).

EXAMPLE – WORLD CAFÉ WITH STAKEHOLDERS

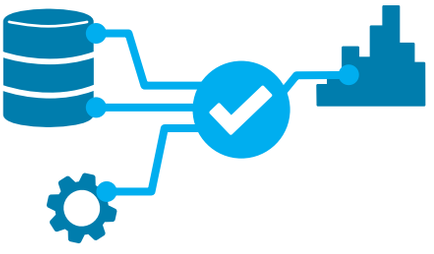
A Swiss telecommunications company maintains a dialogue with its most important stakeholder groups regarding its sustainability strategy. Representatives of external stakeholders discussed key sustainability issues with the company at a World Café (a format for holding small-group discussions). Three major topics were identified:

– Information on the risks of the G5 mobile phone standard,
– Training of teachers and SMEs on digitisation issues by creating digitisation labs for SMEs and schools,
– Creation of increased incentives for suppliers to reduce CO2 emissions.

The participants in the World Café appreciated the opportunity to exchange directly with the company. Based on the positive feedback received, the company intends to organise further discussion events involving its stakeholders to complement the prior stakeholder survey. This type of stakeholder dialogue can also be used to specifically discuss data-based models, e.g., the risk that these models generate discrimination.

PROCEDURAL VALUES

CONTROL



The recommendations on control are intended to ensure that the ethical aspects of data analysis are systematically documented.

Recommendations

– Justify the appropriateness of the algorithms used in relation to the given data set and the specific analysis task; the justification should include documentation relevant to the assessment of the risks and the measures taken to mitigate the risks in the light of the principles of harm prevention, equity, and autonomy referred to above.
– Ensure that any knowledge gained or model generated can be traced back to the original data and the algorithm used to create it.
– If your analysis is based on data and/or models provided by others, ensure that this input was produced in a way that meets your own ethical standards.

EXAMPLE – CONTROLLING AI MODELS

One of the key players in the AI field (Deep Mind) has developed a theoretical framework to guide the assessment of the security of its products. The framework has three elements. The first element, Specification, implements the value of Control. The designer distinguishes the 'ideal specification' (the wishes of the designer), the 'design specification' (e.g., the reward function used in a reinforcement learning system), and the 'revealed specification' (the actual behaviour of the system). The first element of Control is to clearly document the ideal specification (the purpose you want to achieve) and the design specification (how this is translated into code). It uses a data-driven method (e.g., inverse reinforcement learning or simulated scenarios) to identify how the revealed specification departs from the ideal specification (i.e., when the AI departs from the wishes of the creator). Source: https://deepmind.com/blog/article/specifying-ai-safety-problems. A similar approach is defined in ABOUT ML, the methodology developed by Partnership for AI involving transparency by documentation. See: https://www.partnershiponai.org/wp-content/uploads/2019/07/ABOUT-ML-v0-Draft-Chapter-2.pdf

TRANSPARENCY

The recommendations on transparency for this step are intended to ensure that both users and expert auditors of the analyses and models are able to assess them in terms of their ethically relevant characteristics. This applies in particular to the limits of the models, whereby the communication of the limits of the model should be expressed with concepts tailored to the target group who needs to use the model.

Recommendations

- Explain the logic of your models and the decisions that shaped them in a way the users of the models can understand.
- Make the limitations of your models (e.g., confidence intervals) clear to the users of your models.
- Explain to users the possible consequences of such limitations (including indirect discrimination).
- Document the measures taken to meet your ethical standards with respect to data-based modelling: for example, measures taken to eliminate or reduce indirect discrimination in your model.
- Inform users about these measures in an appropriate, targeted manner; inform the logic and limitations of the model to the public if the models are used in decision-making contexts that may affect human rights.

EXAMPLE – INFORMATION ABOUT AN AI MODEL ADAPTED TO TARGET AUDIENCE

Suppose a data scientist trains a machine learning model with information about applicants for a line of credit to predict whether they will make timely payments over a two-year period. This machine learning prediction is used to decide whether an applicant qualifies for a line of credit. Three different groups may wish to receive an explanation of the model:

1) Data scientists, who would like to understand the behaviour of the model as a whole, not just in specific instances;
2) Loan officers, who need to understand why a given application was accepted or rejected when comparing it with other applications;
3) Applicants who did not qualify for a line of credit and who would like to understand why, and what they could do to reverse this judgement.

For each stakeholder, a different type of explanation is provided. The AI Explainability 360 suite offers:

1) Software that produces an approximation of the product of the algorithm in the form of directly interpretable, rule-based models that provide global understanding of the algorithm's behaviour;
2) Software that generates prototypical profiles of successful applicants;
3) Software able to identify and communicate to an applicant what individual change in their profile would have changed the decision of the AI model, everything else being equal.

Source: Tutorials page of the AI Explainability 360 Toolkit, retrievable at https://aix360.mybluemix.net

ACCOUNTABILITY

Accountability means ensuring that in the complex process of data analysis and model generation the responsibilities for adherence to ethical recommendations are respected.

Recommendations

- Assign well-defined responsibilities to the people involved in model development to ensure that ethical standards for data analysis and knowledge generation are met.
- At the organisational level, ensure that ethical standards for data analysis and knowledge generation are met and that emerging ethical problems can be identified.

EXAMPLE – DATA GOVERNANCE BOARD

A media company interested in developing new business models with data has set up a Data Governance Board. This board meets at least quarterly (with ad hoc meetings as needed) and involves representatives of all business areas, especially the teams active in product development and the legal department. This board discusses the ethical questions of new business ideas openly and under the guidance of a person who has received training in ethical issues. The discussions are recorded in the form of written minutes. The joint learning experience raises awareness of ethical issues in everyday professional life. In critical cases, these questions also reach the higher management and are discussed there. The minutes of the group discussions help to present the problem as precisely as possible, which the management appreciates as an efficient mode of communication.

BASIC ETHICAL ORIENTATIONS

HARM AVOIDANCE

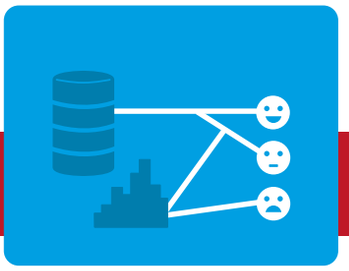
The recommendations for harm avoidance at this step aim to ensure that the potential for abuse of data-based products and services is recognised.

Recommendations

- Do not distribute data-based products and services until you have investigated the extent to which unskilled users can damage the products and/or skilled users can abuse them.
- To the extent possible, ascertain whether your data-based products and services are in fact being misused by third parties (e.g., if software is being used for purposes other than those for which it was designed and tested).
- Determine whether your product could be used to de-anonymise anonymous data or to collect sensitive data from originally unproblematic personal data.

EXAMPLE – ABUSIVE BOT INTERACTION

A company uses artificial intelligence to develop an information bot designed to answer general questions (similar to the telephone-accessible information service once widely available). The software bot learns from user interactions how satisfied users are with the answers it provides; it was trained with extensive, publicly accessible information such as Wikipedia. In the case of controversial topics, the bot provides answers containing all aspects of the controversy in order to maintain neutrality. It soon became clear that the software bot increasingly provides 'neutral' answers even when the information is factually secure: for example, it leaves open the possibility that the earth could be flat. The reason for this was that supporters of conspiracy theories expressed dissatisfaction with the answers of the bot in large numbers and thus influenced the bot's learning. The developers had neglected to plan for systematic abusive interactions with the bot. Once it was clear that the abuse had taken place, countermeasures could be taken.



JUSTICE

Justice requires ensuring that the use of data-based products and services does not lead to undesirable social effects such as indirect discrimination or the stigmatisation or exclusion of groups of people. The ethical duty to assess and address indirect social impacts is proportional to the financial and technological capabilities of a company and the extent to which its data-based products and services affect society.

Recommendations

- Avoid using data-based products and services whose indirect discriminatory effects have not been studied and quantified, particularly where there is a risk of significant indirect discrimination.
- Avoid using data-based products and services that may inadvertently stigmatise whole groups of people or threaten their reputation.
- Assess whether your data-based products and services exclude vulnerable groups and avoid such exclusion if possible.
- Monitor automated decision systems or recommendation systems for unintentional discrimination even if the data product has been tested for indirect discrimination; the test data may have different statistical properties than the real data, and new forms of discrimination may arise.

EXAMPLE – VALUE TRADE-OFFS IN COVID APPS

Many European digital contact tracing apps, such as Swiss Covid (used in Switzerland), Immuni (Italy), and Corona Warn App (Germany), have chosen a decentralised architecture implemented by a Google/Apple API that significantly limits the data that can be collected for further analysis. However, the decentralised architecture makes it very difficult to ascertain whether certain social groups are unequally impacted by notifications and false positives. If the app were made compulsory, or even linked to significant social advantages (such as the right to use public transportation), not knowing what groups are notified and inequalities in the rate of false positives would imply ignoring important elements about justice in deployment. Thus, the app as designed protects privacy, but prevents the collection on information relevant to the justice implications of the app's use and functionality. There is thus a tension between justice and privacy in the deployment of the app. So, if the app were to be made compulsory, Google/Apple should modify the protocol to enable the systematic collection of anonymised data, allowing assessment of correlations between the rate of notification (and of false positive notifications) and various demographic features (sex, ethnicity, income, etc.) – this in order to accurately identify populations disproportionately at risk.

AUTONOMY

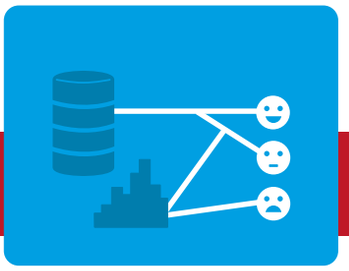
The recommendations on autonomy are intended to ensure that feedback reporting negative effects of your data-based products and services is taken seriously.

Recommendations

– Implement and follow up on processes that allow affected parties to report negative results of your data products.
– Follow expert debates in the various application contexts of your data products that provide evidence of inacceptable systematic deviations in automated decision-making processes.

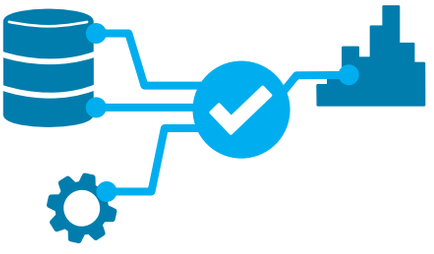
EXAMPLE – CITIZEN JURY

Deep Mind teamed with the (UK) Royal Society for the encouragement of Arts, Manufactures and Commerce (RSA) on a project to encourage meaningful public engagement on the real-world impact of AI. The RSA conducted a citizens' jury to explore the use of AI for decision-making. This was used to discuss general questions about AI, for instance, what sort of engagement the public needs to feel comfortable with AI, and what are citizens' 'red lines' (i.e., questions relevant prior to deployment and at the design stage; see Step 3 – Autonomy). To implement the principle of autonomy as described here, the same means (a citizen forum) could be used to facilitate discussion of negative aspects of the deployment of a technology after deployment. Source: https://www.thersa.org/action-and-research/rsa-projects/economy-enterprise-manufacturing-folder/tech-and-society/forum-for-ethical-ai



PROCEDURAL VALUES

CONTROL



The recommendations on control are intended to ensure, at a general level, that a company's values are appropriately reflected in the data-based products and services that the company intends to offer on the market.

Recommendations

– Determine the data ethics priorities for your company – these should be determined by your ethical duties and your mission as a company – and derive corresponding requirements for your data-based products and services.
– Assess whether ethical standards have been followed at each step of the process that led to your product or service (whether carried out within your company or by others) and whether they are consistent with the mission, values, and public image of your company.

EXAMPLE – ENFORCEMENT PROBLEMS OF AI GUIDELINES

The NGO AlgorithmWatch has started to compile guidelines for 'ethical AI'. However, of the more than 160 guidelines now included in the AI Ethics Guidelines Global Inventory (2020), only a small fraction (10) have adequate oversight and enforcement mechanisms in place. Regular reviews of such guidelines, such as those conducted by AlgorithmWatch, and recurring critical discussions could create public pressure to make guidelines resulting from data ethics standards controllable.

TRANSPARENCY

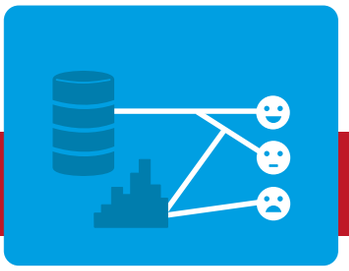
Transparency requires providing adequate information on your compliance with the recommendations set out in this Code and contributing to the maintenance of ethical data practices.

Recommendations

- Communicate your ethical standards and practices to anyone who may be affected by the results of your data product.
- If your company produces data, models, or data products for third parties, provide all information necessary for the ethical use of these data/models/products; for example, if a third party wishes to build a model based on your data, provide enough information about the data set to enable the application designer to avoid unintentionally introducing biases into the model.

EXAMPLE – RESALE OF MODEL-GENERATED KNOWLEDGE

A company uses data to generate new information through data analysis, which in turn is made available to others for further processing. In the data the company uses, it has specified the purposes for which it may use the data and has stored this in metadata. The company's own data ethics guidelines define fundamental values, such as non-discrimination, respect for persons, self-determination, transparency, accountability, etc. In particular, the company has committed itself to ensuring the traceability of the data, meaning that the origin and purpose of the data remain known all along the value chain. These principles are attached to the information sold. Through sporadic enquiries, the company ensures that the third parties basing their products on the company's results and models are informed about the purpose of the data as it was originally used by the company.



ACCOUNTABILITY

The recommendations on accountability aim to ensure that your company's top management can appropriately assume responsibility for its products.

Recommendations

- Prevent responsibility dilution in the complex process of developing data-based products and services.
- Implement processes to ensure that ethical standards are maintained in the delivery of the data product.
- In the event of an ethical problem with your data products, accept responsibility and do not name scapegoats.
- Be aware that the ethical standards of your data products may be strongly influenced by the ethical standards of other companies with whom you interact (e.g., because they have collected the data you use).
- Take these influences into account when evaluating the ethical standards of data products and take responsibility for the final product.

EXAMPLE – CHIEF INFORMATION SECURITY OFFICER

Part of the management's responsibility is to ensure data security (this is an element of step 2). The role of the Chief Information Security Officer (CISO), which has now been introduced in many companies, can serve as an example. When setting up such a function, the following should be avoided:

- Lack of real authority vested in the position, so that the CISO can only formally, but not effectively, carry out his or her task;
- Installation of the CISO as a scapegoat to be sacrificed when a mistake is made.

Conversely, the following characteristics of a CISO can contribute to success:

- Focus on change management and transformation, if a company wants to develop substantially;
- Focus on operational and technical issues in order to develop workable solutions for ongoing operations;
- Focus on regulatory issues and compliance where these are of major importance to the company;

These differentiations are helpful not only for the sub-area of information security, but also for the other areas mentioned under steps 1, 3, and 4.