Digital Services

Data Ethics / AI Ethics

The 2 faces of a responsible future

orange™ Business

A word from the editor…

What are ethics?
How does it relate to data?

The paradigm shifts currently observed in the cultural, societal, social, economical, technological and even environmental spheres, may be more interlinked and fundamental than they seem. Science and technology have taken some mind-boggling leaps in a mere three hundred years, boosting Man's understanding and capacity to control his surroundings. Of course, when all is said and done, no one wants to (nor could for that matter) really give up these developments that are today part and parcel of our culture. But it must be acknowledged that their actual impacts on our lives have not been fully assessed, least of all anticipated.

Customers have increasingly high expectations of the scientific, technological, biological and even financial worlds and are clamouring for new rules and challenging companies to engage in their search for meaning and questioning.

What if ethics was the common denominator of these emerging expectations and demands?

In our modern societies, data, labelled the fuel of the 21st century, could very well have a stronger connection to ethics than we think. And this is the argument highlighted in this white paper. In an increasingly technological context, where operating processes mainly feed on data, we believe that the very first step that companies should take towards ethics is adopting a "Data Ethics" stance.

Data ethics is more than just a new approach to data use, it is about a state of mind and values. This is why legislation alone is not enough to ensure its practice. In the absence of self-regulation, personal beliefs and values, it cannot exist!

We will hence to try to explain why companies should take hold of ethics and make it their own, and regarding data, why you should speak up and even openly display ethical behaviour. However, such a message demands that we first go back to basics and conduct, upstream, a brutally honest examination of our own values.

We hope that this white paper will encourage you to publicly commit to Data Ethics and to communicate on the subject, both at the internal and external level. Adopt a charter accessible to all in the company, be transparent, defend the position! This is how you will be able to make a difference and combine the best of both worlds: technology and ethics… We are convinced that it is possible to create value in a resolutely technological world whilst respecting human beings and ethics by putting them at the heart of our concerns, providing we take affirmative action and are open about doing so!

Contents

04 Civil society's quest to regain control over data
06 Data non grata
10 Can we regulate the use of data?

12 Ethics in an opaque environment
14 State of play/ critical
18 Data ethics: a matter of survival for companies
22 Displaying ethical behaviour

24 Ethics, the new competitiveness driver
24 Data Wars : will there be a data war ?
26 No data (ethics), no AI !
30 Conclusion

Part 01

Data Ethics / AI Ethics

Civil society's quest to regain control over data

Why citizens want to regain control over their data

The French are increasingly wary of entrusting their personal data to companies! Indeed, 60% of Internet users are more and more vigilant when using the Internet [1] (i.e. up 6 percentage points as compared to 2017), and over half of them delete their browsing history and refuse to share certain data. Why are they so distrustful? Because the use of their data, especially for artificial intelligence purposes, is often perceived as intrusive. Proof that implemented measures and regulations are not enough. If the situation deteriorates any further, the slightest mishap – whether related to artificial intelligence or not – will likely result in a crisis, and an ensuing negative buzz. In such a context, restoring confidence in data has become a matter of urgency.

1 Personal data and trust: changing perceptions and uses post GDPR, study carried out by the Chair Values and Policies of Personal Information of the Institut Mines Télécom and Médiamétrie., October 2019

Data non grata

With 29 terabytes of data being published in the world every second [2], it is extremely difficult to keep such a volume of information under control, especially since companies now have a vital economic stake in its exploitation. In an environment of heightened competition, characterised by frenzied races towards innovation, risks of ethical drifts are widespread and growing.

93%

of French companies have been attacked at least once in 2018 [3].

When cybercrime becomes a geopolitical issue

The Cambridge Analytica scandal, suspicion of Russian interference in the 2016 United States presidential elections, suspicions of technological espionage by China...Who still has full confidence in the use of personal data? Just like oil, data – 21st century's newest "black gold" and the fuel of artificial intelligence – has now become the sinews of the digital war. Crises are proliferating in a global context where cybercrime is increasing: with more than 8 million malware detections every year [4], no organisation is spared. Google, Uber, Facebook, Yahoo, Amazon, as well as British Airways, Marriott… have all experienced data leakage incidents. Since 2013, over 13 billion data have thus been lost or stolen in the world [5] !
Digital terrorism on the rise: +32% cyberattacks in 2018 as compared to 2017 [6] and an upward trend in 2019 (+25% in big companies during October alone[7]). The annual cost of cybercrime is 600 billion dollars [8], accounting for 0.8% of the world's GDP. As a result: cyber-terrorism highlights the current gaps in effective data protection, compounding the mistrust and reluctance of consumers to entrust their personal information to companies.

Big Data or Big Brother?

Data may have taken on a key role in our society, but there is a flip side to the coin. It has also bred mistrust and created a constant feeling of being "spied on". 57% of French Internet users feel that their actions are being more tracked now by private companies (search engines, social networks and e-commerce sites topping the list) and 80% consider the current automatic recommendation systems' algorithms, based on purchase or browsing histories, to be too intrusive[10]. In May 2019, the American Democrat Senator Chris Coons also demanded information

"As more and more data is being collected today, it is imperative that we adopt a responsible and reasonable approach to using data for business purposes. We must all be aware of the uses that must be made of data, personally, but also whilst interacting with other organisations."

—Serge Blanc, Data Scientist, Business & Decision France

Expert opinion

Jean-Michel Franco,
Product Marketing Director at Talend

Product Marketing Director at Talend. Today, It is still very difficult for the average consumer to exercise his right to personal data portability vis-à-vis companies. We tested this at Talend within the context of the Data Trust Readiness study[9] and the findings were appalling: in a majority of cases, the companies either did not respond at all, or responded only partly and in somewhat unprofessional ways. However, in legal notices, it was quite easy to find the person to contact to recover one's data. These requests are thus viewed more as legal concerns than being part of customer service. This is why, recovering one's data, in a comprehensible format, remains a complex task. It is difficult for consumers to understand what companies do with their personal data or to see the clear benefits of entrusting them with that data. Data teams have not been educated in the matter and lack awareness regarding their data responsibilities. And this in spite of the fact that the GDPR provides a good practices framework. It is up to the DPO then to bring them up to speed. This is the extra step that companies must take! Privacy by Design defines rules that help make security a part of the application and as such, qualifies as a virtuous approach. Data used for Data Science can be rendered anonymous without impacting the results in any way. These rules must then be spread across the entire organisation in order to reassure data players as well as end customers.

regarding the recordings of the voice assistant Amazon Echo. And what of the social credit system being considered by the Chinese State to “monitor” its citizens by collecting data on them? What are the direct consequences of these ethical drifts with regards to data? In the wake of the Cambridge Analytica case, one in four Americans – and 44% of young people[11] – deleted their Facebook application whilst 54% of them modified their security settings. As a result, the group’s stock lost 20% of its value. In Europe, the social network lost some 3 million users! Admittedly a figure that is still quite low when compared to the 2.45 billion active users in the world[12], but clients’ behaviour towards their personal data is definitely changing.

When Data knocks out AI

In this regard, the first artificial intelligence applications have only served to feed the scepticism further. As evidenced by the fact that Amazon finally had to disable its AI since, having been trained using a majority of male profiles, it was discrimanating against female applicants during recruitment, or by Tay, Microsoft’s chatbot in “automated machine learning” mode on Twitter, wich in just 8 hours, under the influence of its users, became racist, homophobic and a historical revisionist. Both examples of ethical drift occured due to poor quality of data (often incomplete or erroneous). Data can thus be said to be unethical. A fact that has led to symbolically “fire” Amelia, its virtual assistant due to performance issues.

94%

of internet users want to regain control over their data[13].

2 ConsoGlobe – https://www.planetoscope.com/Internet-/1523-informationspubliees-dans-le-monde-sur-le-net-en-gigaoctets-.html
3 CESIN’s Barometer 2018 – https://www.silkhom.com/cybersecurite-3-barometres[1]des-cyberattaques-a-connaitre-en-2019/
4 Source: Data Security Breach, 2017
5 According to the Breach Level Index
6 F-Secure report, 2019 – https://fr.press.f-secure.com/2019/03/05/cyberat[1]taques-h2-2018/
7 2019 Accenture study on cyber resilience – https://www.accenture.com/en-en/in[1]sights/security/invest-cyber-resilience
8 McAfee report in collaboration with the Center of Strategic and International Studies, February 2018 – https://siecledigital.fr/2018/02/22/600-milliards-de-dol[1]lars-cestle-cout-de-la-cybercriminalite/
9 “Data Trust Readiness” report, prepared by Talend and Opinion Matters in April 2019 - https://info.talend.com/datatrustreadinessreportfr.html
10 Personal data and trust: changing perceptions and uses post-RG PD, IMT-Médiamétrie study, 2019
11 2018 Pew Research Center study – https://www.pewresearch.org/fact[1]tank/2018/09/05/americans-are-changing-their-relationship-with-facebook/
12 https://www.journaldunet.com/ebusiness/le-net/1125265-nombre-d-utilisateurs-de-facebook-dans-le-monde/
13 Personal data and trust: changing perceptions and uses post-RG PD, IMT-Médiamétrie study, 2019

Can we regulate the use of data?

74% of French people do not trust the use made of their personal data by mobile applications[14]. And the GDPR alone is not enough to improve the situation. Indeed, 67%[15] of them think that the GDPR has not yet helped to increase the level of protection. Why is it so ineffective? Maybe because the very concept of personal data is still quite vague, for companies and users alike.

78%

are concerned about the use and protection of their personal data[17].

Yet, it would seem that the GDPR gives the end user the option to control his data, at least in principle. First, through consent collection, a process that must be explicit and transparent. Secondly, through the right of access and the right to rectification. Finally, through the right to deletion, the right to be forgotten and the right to data portability, i.e. the ability to ask for all of one's personal data back at any time.
But the facts are grim: compliance with these obligations are still quite low. One year after the implementation of the GDPR, only 28% of companies assert compliance whilst 70% cannot transmit their personal data to clients whenever they ask for them[16], and this, despite formal and legal commitment in the matter as made public through companies' security and confidentiality policies.

Data is the starting point of Data Ethics

Is the GDPR thus a constraint or the basic version of a framework that should exist? Data ethics is a new field and both, the regulations and the new generation, are clearly showing signs of ethical deficiencies. With the rise of artificial intelligence, it is of course gaining new momentum. This is why companies should ask themselves what is the "purpose" of data. In the health sector, for instance, data collection and exploitation can help save lives. Collection, as well as the reason for collecting should thus both be ethical. Since data can, in itself, be unethical, legislation is necessary… but it is far from sufficient to ensure ethical practices. This is the data paradox: data itself, just like the way it is used, can have an unethical dimension. Proper ethical conduct, which is the only way to ensure that customers trust you, thus implies that data must not only be ethical but also used in an ethical manner. And this is why the GDPR, or the California Privacy Act (CCPA) cannot, on their own, guarantee ethical behaviour. These regulations aim to protect data. Apple is consistently campaigning for personal data regulations and now demands that application developers not use systems that secretly record user activity on iPhone devices, if they do not want them removed from its AppStore. A unique commitment that sets it apart in Silicon Valley and on which the Apple brand is banking to boost the sales of its smartphones. Even a few members of GAFA seem to have understood the marketing potential of ethics. But aren't ethics precisely the best way to compete with them?

As a user, I must know straight away that the amount of personal data collected will be limited to the strict minimum and that, I can delete all of my data at any time. This is why users must be included in the decision loop to accurately define the scope of Data Ethics and AI Ethics.

—Romain Bernard, Data Science Manager, Business & Decision France

Homomorphic encryption, the ultimate Data Ethics tool?

How then to develop use cases and still be ethical? Homomorphic encryption, a niche in which Orange is positioned[18], namely presents an interesting approach by combining data use and data confidentiality. “This type of encryption is not yet widespread but will help entities perform computations using encrypted data, without learning anything about the input data or final result. But companies must also comply with codes of conduct such as the ones drawn up by the OECD or European Union for example. Some companies are in fact seriously considering setting up a new “Chief Ethics Officer” function, along the lines of the DPO one. The GDPR imposes Privacy by Design, well the same should be done for ethical issues through the adoption of Ethics by Design principles”, says Mick Lévy, Business & Decision - Business Innovation Director.

14 Source: L’usine digitale, 2016: https://www.usine-digitale.fr/article/donneespersonnelles-3-4-des-francais-ne-font-pas-confiance-aux-applis-mobiles.N387806
15 Personal data and trust: changing perceptions and uses post-GDPR, IMT-Médiamétrie study, 2019
16 Talend 2018 study – https://fr.talend.com/about-us/press-releases/the-majority-ofbusinesses-are-failing-to-comply-with-gdpr-according-to-new-talend-research/
17 Source: KPMG International, “Crossing the line – Staying on the right side of consumer privacy”, 2016: https://assets.kpmg.com/content/dam/kpmg/xx/ pdf/2016/11/crossing-the-line.pdf
18 Orange, october 2017 – https://hellofuture.orange.com/fr/chiffrementhomomorphe-la-cle-de-la-securite/

Expert opinion

Stéphane Walter, Big Data Expert & Consulting, Business & Decision France

Understanding what is meant by personal data is a real issue. According to the French Data Protection Authority, CNIL, “personal data means any information relating to an identified or identifiable natural person and since it concerns individuals, should remain within the latter’s control.” Not using personal data may constitute a serious operational impediment for companies wishing to, for example, design interesting test sets. Hence the temptation to circumvent the limitation. But whether in the medium or the long term, customers will turn their back on brands that do not respect their personal data. This is why companies must start to think along ethical lines. In that respect, the GDPR can help guide them, namely as regards the Cloud Act*. Organisations must take action to protect their data from third parties, as well as their own government, and provide ethical use guarantees to their end users. If the customer trusts the company to which it has entrusted its data to use it responsibly, then he is more likely to want to share his data with said company. * The CLOUD Act (Clarifying Lawful Overseas Use of Data Act) is a U.S. federal law promulgated on 23 March 2018 which makes it obligatory for telecom operators and cloud service providers to submit all information stored on the servers to “American law enforcement officials”. Whether this data is located in the United States or abroad.

Part 02

Ethics in an opaque environment

How to practice Data Ethics in an opaque data environment

Why do users strongly distrust the use that will be made of their data? The unsavoury truth is: a large amount of data is false, poorly collected, untraceable, inconsistent, incomplete and non-representative. The root causes of the problem vary - one of them being for example the fact that 27% of Internet users show resistance when faced with the consent request [1] and 21% of them would prefer not having to share any information at all [2]. What do they do then when they have no other choice than to fill in their personal data? One third of French people admits to providing erroneous data [3], and even using false identities or pseudonyms, in addition to ad blockers [4].

1 Personal data and trust: changing perceptions and uses post-GDPR, IMT-Médiamétrie study, 2019
2 According to the Barometer of French Trust in Digital, ACSEL, 2017– https://www.acsel.eu/presentation-de-6eme-vague-barometre-de-confiancefrancais-numerique/
3 According to a study by Norton Lifelock, 2019
4 Personal data and trust: changing perceptions and uses post-GDPR, IMT-Médiamétrie study, 2019

State of play: critical

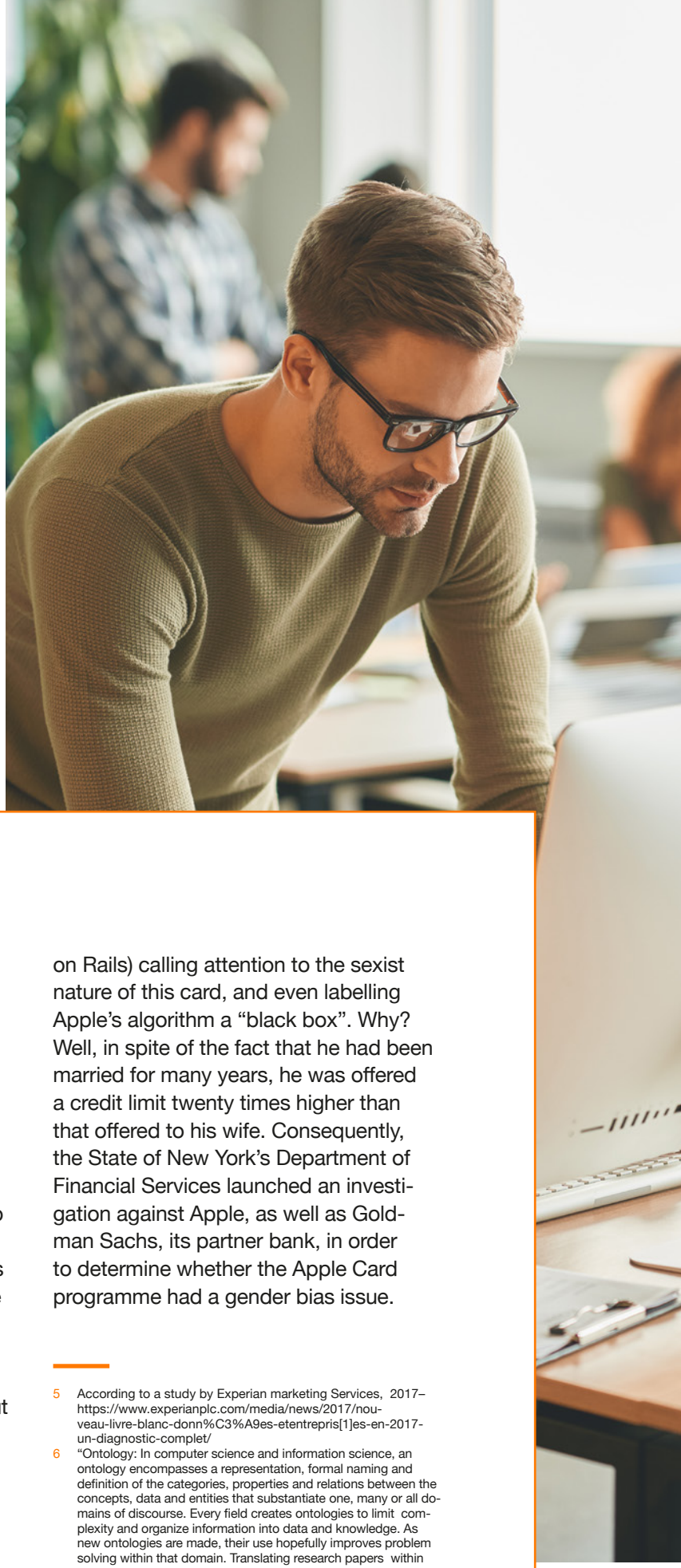
Today, nine out of ten companies feel that their databases contain too many errors [5] and thus do not trust the information with which they are working! How can “reliable” applications, or even worse, reliable artificial intelligence be developed using incorrect data?

“It is easy to imagine, in the near future, sets of data from any information system being used to feed or train AI”, explains Didier Gaultier, Business & Decision France - Data Science & AI Director. “But numerous problems can impede proper AI training, such as false data, inconsistent data and missing (and/or non-representative) data. The list is far from exhaustive, and the danger is that this may not only skew results, but also reproduce, maybe even intensify, human biases through algorithmic biases”, insists Didier Gaultier

The ten key points of Data Ethics

1. The data must be precisely defined with respect to its ontology [6], i.e. within the context of the business process that we wish to model.
2. It must be referenced in an accessible and upto- date data dictionary, including its name, type, features and exact definition.
3. It must be accurate. In case of doubtful elements, these must be known and recorded with the data.
4. Its precise date and hour of collection must be known and recorded.
5. Its mode of collection, including the various possible sources for the data (for instance questionnaire filled out by telephone or on the Internet) must also be mentioned.
6. The data must exist or be explicitly declared missing.
7. It must be consistent, i.e. only vary within the range defined in the dictionary. Likewise, it must not conflict with any other value associated with the same observation.
8. It must be unique, i.e. no observation can produce two entries in the same entity.
9. It must be compliant, licit and validated, i.e. must follow the internal governance rules and standards as well as the external regulations in force (for example the GDPR).
10. It must be useful and valuable: data should only be stored in an IS if, at the very least, a use for it, or a way to create value from, it is being considere

How to determine whether data is “ethical” and trustworthy? Orange Business has identified ten main characteristics that will help you qualify data (and sets of data) as ethical and reliable.



Important

The use of any item of data that does not fulfil these ten requirements is considered to be a breach in ethics. However, it is of course extremely difficult to fully satisfy all of these conditions for all of the data, all of the time. Companies should thus aim for an “ethics gradient” instead that will help them continuously and gradually evolve through all these steps. The idea is for these ten points to serve as an ideal, admittedly difficult to achieve. The goal being to tend towards this ideal by having the highest possible percentage of data in compliance with the biggest possible number of points from the above-mentioned list. What matters is not where you come from, but where you want to go! The case of the Apple Card, the American giant’s payment card, shows how complex Data Ethics is proving to be, even for leading brands. Everything started with a series of tweets from businessman David Heinemeier Hansson (founder of Ruby on Rails) calling attention to the sexist nature of this card, and even labelling Apple’s algorithm a “black box”. Why? Well, in spite of the fact that he had been married for many years, he was offered a credit limit twenty times higher than that offered to his wife. Consequently, the State of New York’s Department of Financial Services launched an investigation against Apple, as well as Goldman Sachs, its partner bank, in order to determine whether the Apple Card programme had a gender bias issue.

5 According to a study by Experian marketing Services, 2017– https://www.experianplc.com/media/news/2017/nouveau-livre-blanc-donn%C3%A9es-etentrepris[1]es-en-2017-un-diagnostic-complet/

6 “Ontology: In computer science and information science, an ontology encompasses a representation, formal naming and definition of the categories, properties and relations between the concepts, data and entities that substantiate one, many or all domains of discourse. Every field creates ontologies to limit complexity and organize information into data and knowledge. As new ontologies are made, their use hopefully improves problem solving within that domain. Translating research papers within every field is a problem made easier when experts from different countries maintain a controlled vocabulary of jargon between each of their languages” (Wikipedia).”

Expert opinion

Jean-Michel Franco, Product Marketing Director at Talend

Companies still lack maturity when it comes to personal data. Their value, their use and especially bond of trust that links the individual who provides the data to the one who consumes it. We are used to particularly high levels of non-quality in the field of customer relationship. The problem is that with the digital era, came the necessity to capture the customer path throughout the whole relationship, and that without the customer's trust, managing such a relationship on an ongoing basis is just plain impossible.

Part of data "ethics" is thus making sure that the customer understands what his data is used for. Trust is the key to loyalty, and in the digital world, loyalty is a fundamental notion. This is what can strike fear in the hearts of Facebook and the like: distrust that translates into decisions not to consume and share anymore. Though rarely aware of the fact, companies thus shoulder a big responsibility when it comes to data.

To regain the trust of their users, organisations must start by adopting a transparent approach to the use of data. If the consumer is unclear about how he can benefit from filling in his data, he is more likely to provide false personal details when prompted. Tomorrow, companies will have to be able to give control back to the consumer and allow each of them to disable and enable confidentiality options as and when they need to do so. This is what companies should be aiming for.

The most common quality problems

Problem	Problem Description
False data	Observation value is incorrect or outside this type of data's tolerance value limits (for example a 160-year-old client).
Inconsistent data	Two values from two different records, for one same observation, contradict each other.
Data incorrectly time dated	Data collection date (or processing date) is false or missing.
Poorly defined data	Data ontology not defined, data is not referenced in a data dictionary or data definition is out of date.
Missing data dictionary	Description and definition not accessible for a whole set of data.
Missing data collection mode	Data collection mode not provided (for example: the type of sensor used to capture the data is not known).
Missing data	Data value not filled in.
Non representative data	A whole set of data missing for a given subject resulting in data not reflecting reality: for example, a whole category of customers is missing.
Duplicate observation (duplicate entries)	Two entries exist for one same observation in one same entity.
Illegal or noncompliant data	Data not compliant with data governance, maybe even illegal by virtue of a regulation such as for example the GDPR.
Unnecessary data	Data never used in any application.

As shown, data can, in and of itself, intrinsically, have an unethical dimension. And as a result, skew the results of analyses performed by Data Analysts and Data Scientists, as well as distort the relationship that a company has with its customer. In the world of business, “Data Ethics” is thus a prerequisite to sustainability.

How serious is the problem	Consequences	Solutions
Potentially very serious,depending on the importance of the data especially if not identified as false.	If input in an AI, or worse used for its training, the AI will malfunction or be biased (Garbage In, Garbage Out).	Flag data as false (or doubtful). Correct data value if still possible.
Usually signals the presence of false data. However, these can then be easily identified as potentially false.	If input in an AI, the AI will, at best, stop working and at worse, malfunction or be biased.	Flag both items of data as potentially false (or doubtful). Correct data values if still possible.
Quite serious in the sense that this can invert the cause and effect relationship in an AI’s or a Data Science project’s inference process.	An AI could draw the wrong conclusions from a cause and effect relationship due to, for example, temporal inversion and in the most serious situations, this could lead to algorithmic bias.	Flag the data as doubtful or incomplete. Enter the full data temporal value if still possible.
Quite serious since poor definition can lead, for example, to a customer being both eligible and not eligible for a loyalty programme.	The AI will make unwanted decisions. An AI could for instance, mistakenly grant or revoke customer rights or access, and the consequences of such actions can be easily imagined.	Defining a complete and documented business ontology is the only way to avoid this problem. One of the results is an exhaustive and up-to-date data dictionary.
Serious enough to, for example, make it impossible for business functions to use the data.	Data cannot be used neither in an AI nor by Data Scientists or Data Engineers.	Use ontology to populate a data dictionary, then publish it and make it accessible.
Only serious if there are various possible collection modes, for example on the Internet and by telephone, or using different types of sensors.	If there are several possible collection modes and the data are mixed up in one same base, this makes them unsuitable for statistical processing and can lead to a biased or malfunctioning AI.	Isolate the data obtained from different sources, fill in the data sources or flag the data as doubtful if this is no longer possible.
Only serious if data is a key data for Data Science analysis or a key AI input, and depends on percentage of missing data.	Depending on the volume or percentage of missing data, it may be difficult, even impossible, to conduct a Data Science analysis. Likewise, an AI may malfunction.	Complete data if still possible (in some cases, a predictive model can be used). If not, flag data as missing.
Quite serious and insidious if not detected.	Will create a data bias that will translate into algorithmic bias and affect an AI’s operation or a Data Science project.	Complete missing data and ensure completeness of data. To do so, tests and samplings can be carried out.
Quite serious if not addressed.	One customer could, for example, receive loyalty rewards twice. Incorrect statistical weight (too large) could be assigned to a phenomenon.	Delete duplicate value using a deduplication process.
Serious since the data will affect an AI’s input specifications and there are risks of legal sanction for noncompliance.	Potential serious judicial and legal consequences. Unethical or biased decisions and actions from the AI.	Make data compliant with governance and legislation. Delete illegal or unauthorised data from IS.
The only real risk associated with unnecessary data is the fact it may have a “hidden defect” that will never be discovered.	Takes up space in the IS and backups, costs money, resources and can potentially hide a never detected defect.	Assess the real usefulness or value of data, delete if not needed and no value.

Data Ethics : a matter of survival for companies

Thus, over and above its uses and “purpose”, data itself must be of an ethical and reliable nature if organisations do not want to run the risk of operating based on skewed results and outside of the regulatory framework.

There is little choice left in the matter: companies must set up dedicated governance. Though complex and costly, the set-up of such an ecosystem is a prerequisite for survival.
“The cost of ethics compliance can indeed be extremely high, says Cédric Missoffe, Business & Decision France Consulting & Expertise agency Director. Each consent implies specific processing. But ifyou are open about why you are collecting emails, then you gain real competitive advantage since this will only increase the trust your brand inspires.”
Legacy evolution, deep cultural transformation, redesign of work processes, modification of data collection and data storage tools, recruitment of new types of profiles – DPO, CISO, Chief Data Stewards, Data Scientists, Chief Ethics Officer… Adopting an ethical approach towards data ownership represents a considerable expense. But not investing in quality data bears a much higher cost.

From a ‘data-centric’ approach to an ‘ethic-centric’ company

To promote Data Ethics, simply implementing a regulatory framework is not enough. The company, as a whole, must make of ethics a priority and position itself accordingly with regards to data. So, what does a data-ethical business model look like? A team. One made up, at the very least, of a business representative, a Data Engineer for data collection and storage and data processing, a Data Scientist for algorithms (statistics, machine leaning and AI) and a Product Owner as Chief Data Officer. All of them working under the watchful eye of a “Chief Ethics Officer” who can spread out a global ethical vision of the desired result and liaise with the various stakeholders regarding ethics.
“Organisations must approach data in the same way they approach human resources: as a company asset, explains Mick Lévy, xBusiness Innovation

Over

95 000

complaints were recorded in Europe for non-compliance with the GDPR with fines amounting to approximately 56 million euros already imposed [7].

Director. This implies central data governance, with at its heart a Chief Data Officer for example, which raises the issue of data culture integration. Internal teams, as well as customers must be trained, assisted and educated."

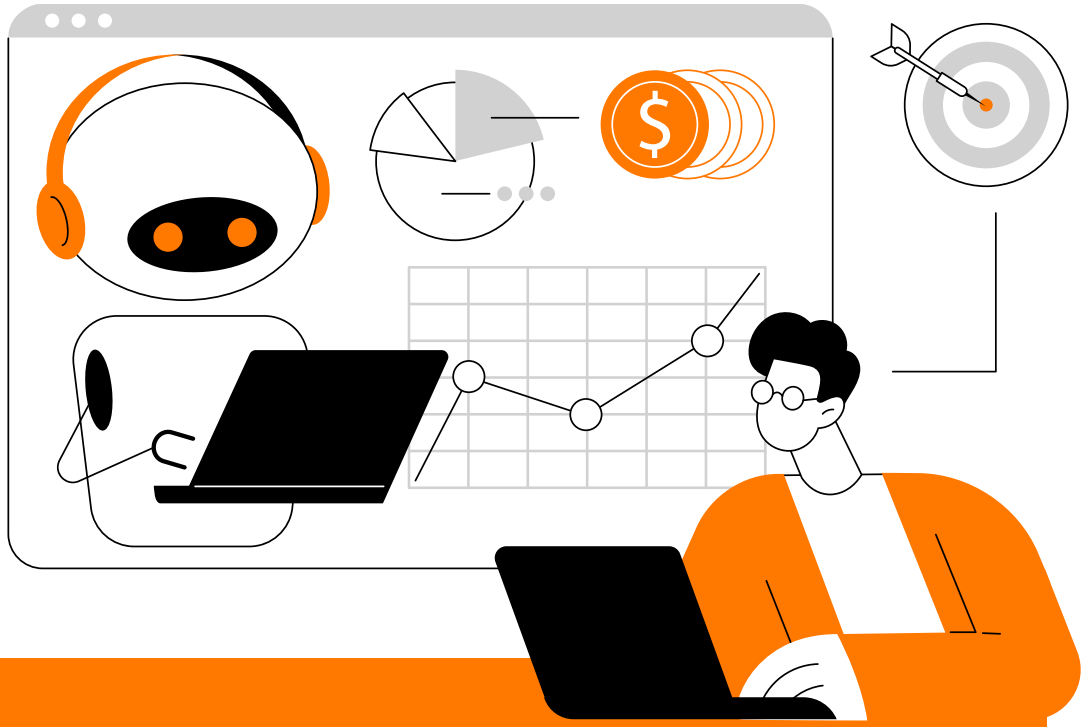
Why should Data Ethics governance be crossfunctional? Because the more data is shared amongst varied profiles, the lesser the risk of bias. And it must be said that current data indicate a lack in terms of diversity and gender. Namely: only 33% of the digital sector workforce are women, with 16% only working in development and 27% in coding [7]. Out of the students currently attending some artificial intelligence schools, 80% are men. Such conditions make it extremely difficult to assert that designed intelligences are a 100% not sexist, even with the best will in the world!

This is why the bias question is one that should always be raised and governance structures combining both artificial intelligence and human intelligence should be set up.

Expert opinion

Emmanuel Dubois, Indexima co-founder

Such projects require deep, complex and costly transformation...and must be driven by top management. But numerous organisations content themselves with a superficial level of transformation by, for instance, creating datalabs... that do not bring them anything. No Innovation Director can, single-handedly, change a productcentric approach adopted fifteen years ago. How can we think about transforming a siloed organisation in which data flow players do not exchange and share data with each other? The hindrances are thus mainly political and often due to internal competitiveness between departments. In contrast, organisations that are able to gather round a table those responsible for data collection, find their projects quickly evolving due to rapidly-made decisions. Moreover, for data to become a differentiating factor, it must be of high quality and exploitable by a majority of people. The more data is shared, the less the likelihood of biased interpretation. This is why it must be brought to maturity by a large number of people and circulated in the company to be unified, consistent, traceable, cross-referenced, usable, and challenged by a data team who can then extract the most relevant information from it and imagine new paths to improve customer experience.



Expert opinion

Fayçal Boujemaa, Technology Strategist, Orange Labs Research

Within the context of AI, it is what “humans” do with data that is problematic. Indeed, the mathematical machine that creates algorithms is only using data fed to it by man. This is why the chosen sample must be a representative one, or else the machine is not able to learn properly. Sometimes, even if the sample is highly representative, the data can encapsulate human and cultural biases rooted in our own prejudices. For example, in 2015, 21 American courts adopted a risk assessment algorithm to help judges decide whether defendants should be jailed before their trial. However, in 2016, an investigation revealed that the intelligent system was totally biased since it had been fed “historical” data from previous rulings. It thus became clear that black defendants were mistakenly being flagged at almost twice the rate as their white counterparts as “future criminals”. This does not mean that the AI was, itself, racist. This is how a cognitive bias got transformed into an algorithmic bias. Likewise, if the data contained in a sample is not “clean” - for instance because the company bought a list of data, the sources of which were not verified - a marketing campaign can end up flawed. This is why data must be clean, qualified, traced...Finally, the ethics notion depends, to a large extent, on the cultural, geographical, political, religious context... Many people mistakenly consider ethics to be only a legal concern. They are wrong. The machine must be taught which uses are ethical or not, using as reference a set of values, a law ... But this implies knowing how to translate these mathematically and coming up with the mathematical techniques that will mitigate these biases and put them into perspective !

When Data Ethics rhymes with ‘customer-centric’

As integral part of the data-centric approach and the customer-centric strategy, Data Ethics represents a profound cultural change. Still considering it as a simple technical and legal element it is synonymous to risking failure. Data ethics requires in-depth knowledge of data and of its full impact.
“In the past, customer-centricity was often perceived as just a marketing notion, says Serge Blanc, Data Scientist. However, today, we no longer speak of product but of customer story and customer experience. Things have changed fundamentally. Wealth is only created through data and its value. This means better understanding one’s environment, then improving the existing before innovating. In other words, creating new services and, especially, new business models with data as the starting point. Transforming to impose the sharing of data with as many people as possible: this is what will make our data rich.”
Within increasingly horizontal and “no-silo” companies, giving all players an opportunity to experiment with and use data is a must. Ethics thus starts with collective awareness, which means that employees must be educated in the matter and the do’s and don’ts clearly explained in a charter or good practice guide.

Wealth is only created through data and its value. This means better understanding one's environment, then improving the existing before innovating

7 Informatique News, June 2019 – https://www.informatiquenews.fr/rgpd-plus-de-95-000-plaintes-deposees-les-amendes-commencent-a-tomber-62104
8 Femmes du numérique, April 2018 – https://femmes-numerique.fr/quelle-placepour-les-femmes-dans-le-numerique/

Displaying ethical behaviour

74% of French people do not trust the use made of their personal data by mobile applications and 78% expressed concern with regards to the use and protection of their personal data [9]. This is why all companies should, at the internal level, develop an ethical data use charter that Data Scientists will be able to use as reference, a document that will commit the company and act as a moral contract between the organisation and the end user.

69%

of young consumers pay close attention to (socially) responsible companies' communication campaigns [10] and are more likely to buy from a brand that demonstrates transparency regarding its commitments.

Hippocratic Oath for Data Scientists and any other person working with data – Data for Good

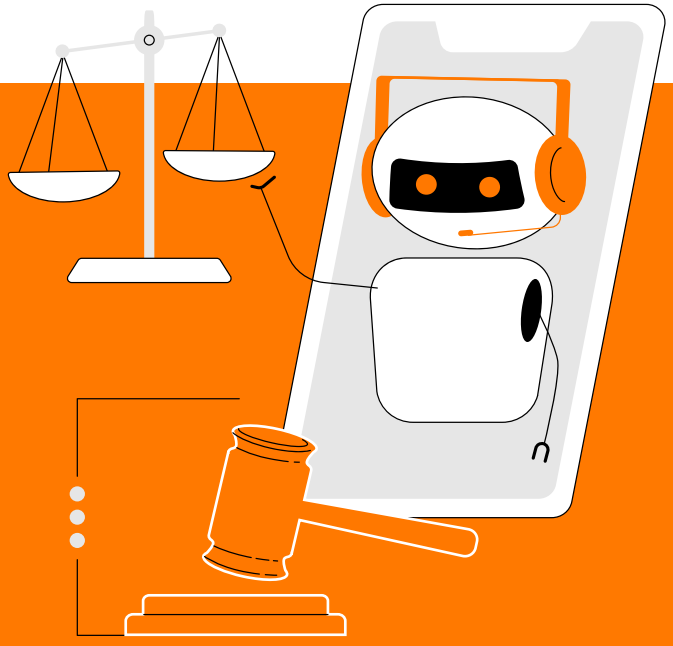
Objective: help Data Scientists put things into perspective, perform their job in a responsible manner and give them the keys that will help them clearly understand the consequences of their actions. Data Scientists should adopt a cautious and holistic approach when asking questions and present facts in a transparent and objective way. In that respect, being ethical requires courage and entails a duty to alert.
https://hippocrate.tech/

Protect your e(thics)-reputation

This is why you can no longer afford not to embark on this Data Ethics journey, without running the risk of being completely side-lined! Ethics has now evolved beyond the spheres of business. The connection between the corporate ecosystem and the social ecosystem has never been stronger, as evidenced by the #MeToo movement which forced companies to take a clear stance. What if the next #MeToo was (data) ethics?

9 Innovation Barometer from Odoxa-Microsoft – L'usine digital 2016 – https://www.usine-digitale.fr/article/donnees[1]per-sonnelles-3-4-des-francais-ne-font-pasconfiance-aux-ap[1]plis-mobiles.N387806

10 According to a study from Fuse agency, 2018 – https://www.businesswire.com/news/home/20180628006409/en/Teens%E2%80%99-Views-Social-Activism-Marketing-Mat[1]ters-Brands/



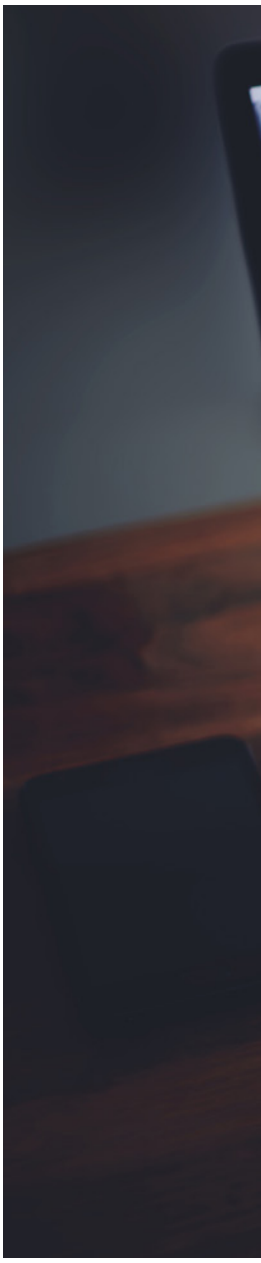
Expert opinion

Luc Julia, Samsung Electronics CTO,

Siri co-creator and author of “L’intelligence artificielle n’existe pas” (artificial intelligence does not exist) -First publishing house - 2019

AI, just like any other tool, can be misused. AI does not decide to be bad; it is us, humans, who make ill-advised use of them. Claiming that tools are taking over is thus a fallacy. But it is true that they can be programmed to act poorly. This is why AI is above all a reflection of personal ethics. The community is the one that gradually decides, based on its own ethics, which policies, laws, regulations to adopt... as shown by the Treaty on the Non-Proliferation of Nuclear Weapons signed in 1968 or the Russian-American Intermediate-range Nuclear Forces Treaty signed in 1987. The same principle applies for AI: if there are abuses, the community will simply place a stranglehold on their use. But for users to understand this, the public must be educated and we must stop saying anything and everything about AI. Once they know better, consumers will demand ethical behaviour. I believe in the value of human beings basing their decisions on what is right. This is why, being ethical represents a true opportunity for brands, and GDPR’s data privacy concept plays a real educational role, even though some companies struggled against it at first. Showing that you are ethically responsible is a positive thing. To practice “Data Ethics,” data must be unbiased. When a system is created, it must thus, as best as possible, treat all different populations on an equal basis. For instance, if my AI is sorting CVs, the dataset must have been selected from an egalitarian dataset. The problem with AI today, is that it is expected to simplify processes, decisions...using data, and this implies access to the most comprehensive data possible, which is no simple matter. Nowadays, creating algorithms is not difficult; the complexity lies in the data. Hence, to create egalitarian AI, data must be carefully chosen. And that is where ethics comes into play: at the data selection stage! The problem is the data and the bias in them, since there have always been bugs in algorithms and there will always be bugs in algorithms. But biases exclude individuals making the sharing of datasets, and constantly accessing those datasets to fix bugs, necessary.

Ethics, the new competitiveness driver



It is a fact, today more than seven consumers out of 10 (73%) only decide to buy after having read up to six customer feedbacks. Feedbacks that twice more people trust (68%) than they would any traditional media advertising. Conversely, a poorly rated brand can be eliminatory for 87% of the population. That says it all. For companies, trust is the new battleground of business war.

Data Wars: will there be a data war?

92% of GDPR-compliant companies – i.e. having adopted a somewhat ethical approach to data – reported having gained competitive advantage [1] whilst only 28% of them were expecting such an outcome before the implementation of the European regulation. Amongst the main noted improvements: customer confidence (8%), brand image (8%) and employee motivation (79%). Other noted benefits from data collection compliance: enhanced cybersecurity (91%) and deeper organisational transformation (89%).

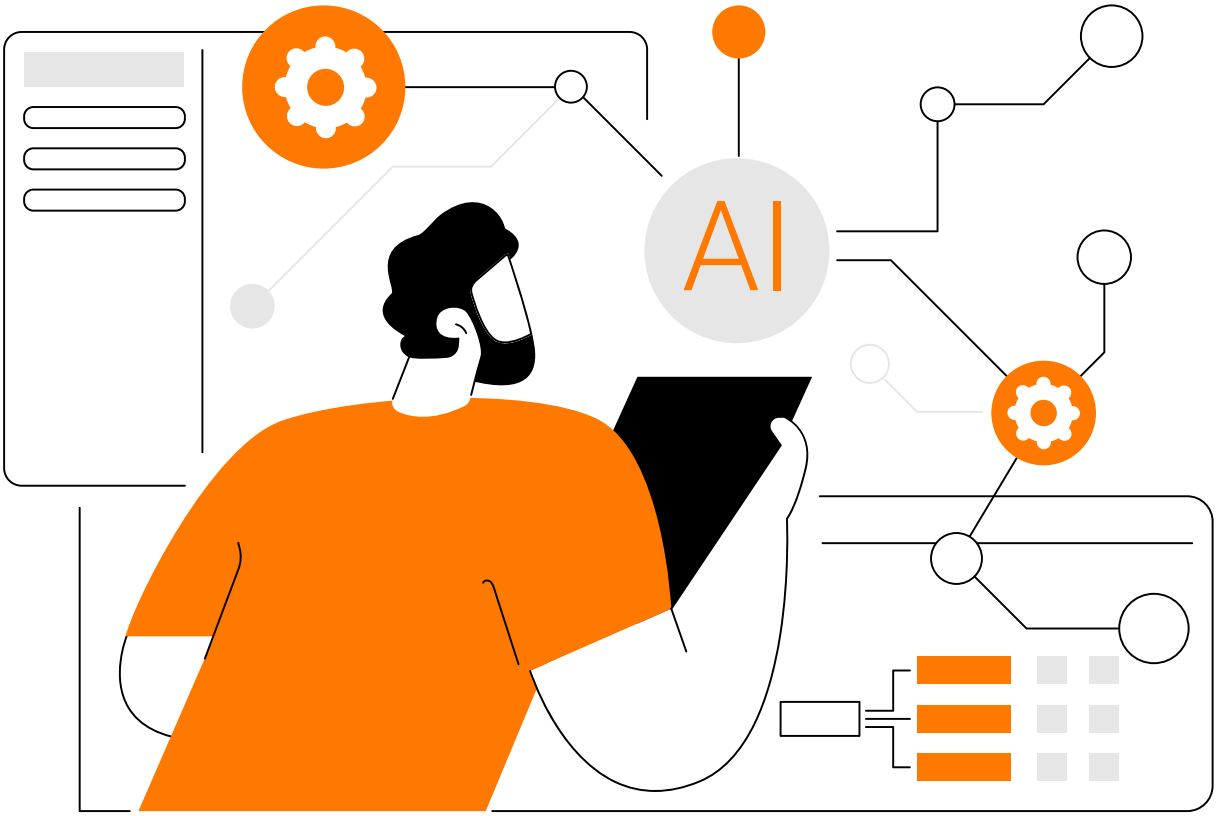
Why this competitive factor? Because, as Cédric Missoffe, Consulting & Expertise Director, explains, “European companies do not have the same legal weapons as their American or Chinese competitors. Our difference cannot be expressed through technology as we all use the same artificial intelligence. The purpose of the AI is what will set us apart since it requires the end user’s explicit consent.”
And in that regard, ethics has emerged as a major competitiveness driver against the GAFA and other Chinese

startups. Why? Because if the United States have chosen an ultra-liberal approach and the Chinese all-powerful state control, Europe has, for its part, embarked on a differentiating 3rd path based on respect for the people and ethics. However, for ethics to become a true competitive factor, it is also important to choose one's suppliers carefully and to ensure that the data value chain is ethical end-to-end, from initial collection to data storage and processing and service delivery to the end user.

"The purpose of the AI is what will set us apart since it requires the end user's explicit consent."

1 According to the Barometer of French Trust in Digital, ACSEL, 2017https:// www.acsel.eu/presentation-de-6eme-vague-barometre-de-confiancefrancais-numerique

No data (ethics), no AI !



In a context of data and intelligence wars, what part can French - or European - companies play opposite the powerful GAFA and the likes of Alibaba? That of leads… How come? Quite simply, because “we are the best". We are not the ones saying it, Luc Julia is. And he further adds: “he best mathematicians in the world are ours, so we definitely have a role to play. France holds all the cards to become the flagship of artificial intelligence in years to come... Much like the Chinese and the Americans do.”

The arrival of AI is inevitable.

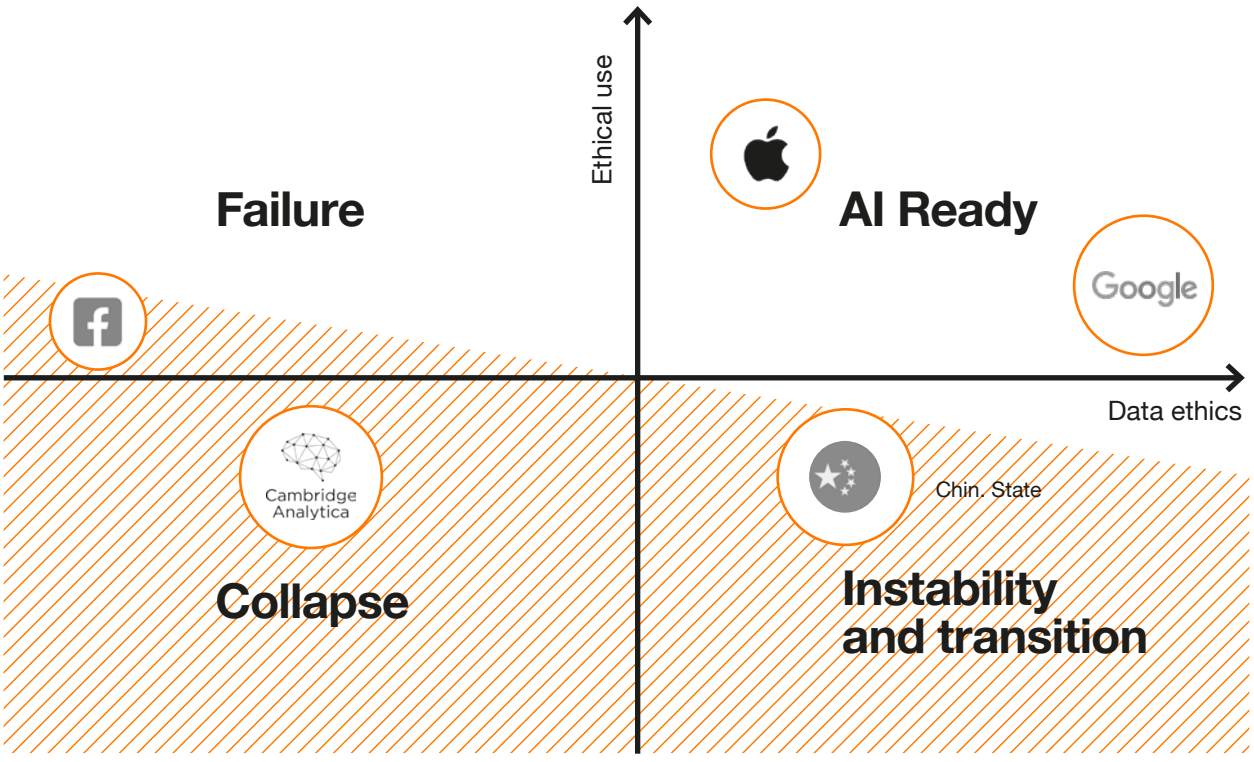
You have little choice but to prepare yourself for it, as the European Commission published, in 2019, its ethics guidelines for trustworthy AI, and the new European Commission Director, Ursula von der Leyen, declared that new European regulations concerning AI - to complement the GDPR - were on her priority list. Indeed, the immediacy requirement imposed by consumers will, in the near future, lead to the use of AI. Artificial intelligence is set to become a norm just like having a free email address in its day. “But nothing is free and the GAFA are not charitable

organisations", reminds Cédric Missoffe. It is thus possible that we are about to witness a reshuffling of the cards between those who have understood this and those who refuse to acknowledge it, amongst whom a few GAFA maybe? Therefore, companies can be grouped under four categories. Whilst the GAFA oscillate between more or less ethical data and uses, for them, scandals are always just a stone's throw away. As regards the Chinese companies subject to the requirement of entrusting their data to the government, today they are not really able to practice AI Ethics in spite of having data, which in their case, is reliable and of good quality. For differentiation purposes and to secure a foothold in the market, French and European companies must thus make the most of the regulations in force to gravitate towards the top right quadrant and Data Ethics, a necessary condition for trustworthy artificial intelligence.

Half of the artificial intelligence projects of one out of four companies fail[2].

"Ethics gives companies the opportunity to freely contribute to AI, whilst however complying with certain rules for the common good, concludes Didier Gaultier, Business & Decision France - Data Science & AI Director. This is why Data Ethics and AI are today inextricably linked. Adopting any other approach would be synonymous to risking, in the long run, substantial regulatory penalties, and in the shorter term, losing customers' and users' trust."

2 According to a study from Norton Lifelock, 2019
3 Personal data and trust: changing perceptions and uses post-GDPR, IMT-Médiamétrie study, 2019



Legal Risk Area Strong in Eur.

Expert opinion

**Emmanuel Dubois,
Indexima co-founder**

Though restrictive, the GDPR provides a framework that allows data definition and use. In that respect, it can thus guide companies and help them better harness their big data, namely to compete with the GAFA quadrumvirate. This is the real data challenge and the importance of data quality - of unified, consistent, verified, traceable and exploitable data. The GAFA and the biggest unicorns can, single-handedly, disrupt the market and rattle major global corporations. “Traditional” companies now understand that their very existence is threatened. In 10 years, they have thus had no choice but to transform themselves. However, this has often been done in a random and hierarchical way at the detriment of agility, with transformation being mostly restricted to marketing communication. Even though all of these companies realise the importance of data, only a few have managed to use it as a lever. The reason : they have been unable to harness a mine of information which they have owned for decades, and this despite the strength that this represents. In contrast, since day 1, the digital native and customer-centric GAFA have set up KPIs to help them manage their activities, constantly reduce their costs, optimise their processes, innovate and continuously improve customer experience. To this day, Amazon’s algorithm is more effective than the ones of incumbent organisations. It is like comparing what Fintech companies can do to traditional banking, and in the event of a new financial crisis, transformation failure could expedite the fall of numerous institutions. However, though resolutely customer-centric, giants of the web often do not care about ethics… contrary to more “institutional” companies that show higher levels of compliance with regulations. The GDPR is thus an asset as it can help guide companies towards Data Ethics and also educate users on their rights as regards personal data. This is how businesses will be able to differentiate themselves from less ethical ones: through complete transparency when it comes to data use and communication with the general public. A real competitiveness booster, especially against the GAFA, data is above all a matter of organisation. Just wanting to be data-driven does not make you so, and the organisation is key to becoming data-driven. Restructuring your organisation around data and not processes starts with a state of mind.

Conclusions

Who has ever taken the time to read the general terms and conditions of use applying to website data? An American study carried out in 2016 revealed that 74% of people validate the terms and conditions of use without even opening them and 98% do not take the time to read them. Why? Because more often than not, they are incomprehensible.

And regarding this situation, the GDPR has unfortunately not changed a thing! At the end of 2018, only 1% of websites were able to respond, within 30 days, to email requests pertaining to personal data in their possession[2]. 83% could not respond at all and in 16% of cases, the email address mentioned was not valid. “The lack of a global or generic ethics, approved by all, is certainly regrettable, says Ada Sekirin, Orange Business, Digital Services CEO. It is extremely important that algorithms be told what to do within an X, Y or Z framework that everyone finds acceptable. We must thus take a stand and it is far from easy. These hurdles are still very much in place and the issue must be discussed at a much more fundamental level than it currently is.”

Unethical data is not reliable and can potentially lead to a high number of errors throughout the associated chain of processes: i.e. when it is processed, analysed, output, used, and included in designers’ reasoning and, de facto, in the machine learning process. “One simple item of data, if false, will through AI, be cross-referenced with hundreds of other data, and ultimately lead to the generation of false information, and this even if the hundreds of other data are correct, insists Didier Gaultier. And training AI with false or incomplete information is extremely risky. Missing key data on AI’s environment can also have unpredictable consequences.” Last September, the car manufacturer Tesla tested its new semi-autonomous feature “Smart Summon” on one of its Model X cars. The car was supposed to leave its parking space on its own and meet up with its “owner” using his smartphone’s geo-tracking. But they had not factored in the reaction of ordinary individuals who, faced with a driver-less car in a parking, started hesitating, even creating some deadlock situations. This “disturbed” the autonomous car which then lost control and caused a few, fortunately minor, accidents.

“The problem of course is the reasoning input in the AI with regards to the environment, as well as the whole machine learning process, explains Didier Gaultier, Business & Decision France - Data Science & AI Director. Why? Because people got scared when they saw this driver-less car and hesitated, thus displaying unusual behaviour for which the autonomous “machine” had not been trained. The emotional response of other users and the complexities of a parking space’s global environment were not sufficiently accounted for.”

This shows that for AI to function properly, it needs, not only a correct environment analysis, but mostly complete, accurate, consistent, sound, unbiased and high-quality - in a word: ethical - data. One basic rule is that AI can never be better or more effective than what it has learned; learning using unethical data will, under no circumstance, create ethical and reliable AI and unreliable AI, as previously illustrated through numerous examples, will always create a “bad buzz” and damage image and reputation.

This is why Data Ethics is a prerequisite of trustworthy AI.

1 https://www.numerama.com/politique/182421-mettez-ce-que-vous-voulez-dansles- cgu-on-accepte-nimporte-quoi.html
2 Study carried out by Freebip, December 2018 – https://comarketing-news.fr/rgpd-6- mois-apres-le-constat-est-accablant/
3 The various AI learning modes and their application will be dealt with in a white paper to be published.

Contributors

- **Didier Gaultier:** Data Science & AI Director - Business & Decision France, part of Orange Business
- **Fayçal Boujemaa:** Head of Strategy - Orange
- **Emmanuel Dubois:** General Manager and co-founder - Indexima
- **Jean-Michel Franco:** Product Marketing Senior Director - Talend
- **Luc Julia:** CTO & Senior Vice President of Innovation - Samsung

Business & Decision France, Part of Orange Business

- **Romain Bernard:** Data Science Manager Paris-Nord Ile-de-France
- **Serge Blanc:** Data Science Manager
- **Michaël Deheneffe:** Strategy and Innovation Director
- **Jérome Dewever:** MDM Expertise Consulting Manager
- **Maxence Dhellemmes:** R&D Director
- **Mick Levy:** Business Innovation Director
- **Cédric Missoffe:** Consulting & Expertise agency Director
- **Ada Sekirin:** CEO - Orange Business - Digital Services business line
- **Mondher Sendi:** AI Expert - Computer Scientist
- **Stéphane Walter:** Consulting & Expertise / Big Data Manager

Acknowledgments:

- **Renaldo Candreva:** Graphic Designer

Do you have any further questions?

Or if you would like to know more about how to create value from Artificial Intelligence, do not hesitate to mail us your query.

Digital Services

Digital Services is a business line within Orange Business, contributing to reliable and successful digital transformation for many organizations. Our joint mission is to help customers innovate and drive their business strategies in key digital domains, including Cloud, Customer Experience, Workspace, and Data & AI.
We assist them on their digital journey by providing advisory, end-to-end solutions, managed services, and professional services to ensure our customers' success.
We are digital natives, with innovation at the core of our business, which makes us a reliable partner close to our customers, leading them in their digital transformation challenges.

We support a wide range of industries in the private sector as well as the public sector. We have built a significant level of experience and understanding over the last 30 years in industries like Finance, Insurance, Life Sciences, Healthcare, Manufacturing, Travel & Transportation, Retail, and the Public Sector.
As always with Orange Business, our customers trust us for delivering end-to-end, sovereign, and sustainable solutions to turn their Operational Experience, Employee Experience and Customer Experience into business value.

Find out more how we can help you with your projects on:

digital.orange-business.com

Orange Business
Digital Services
Avenue du Bourget 3,
1140 Brussels
Belgium
marketing@bd-orange.com

Follow us on

orange™ Business