

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/213892>

Please be advised that this information was generated on 2021-04-18 and may be subject to change.

10. 'Legal by Design' or 'Legal Protection by Design'?

Mireille Hildebrandt

Published on: Jun 02, 2019

Updated on: Sep 07, 2019



Yuko Nasaka 1938

Untitled 1964

© reserved

Collection of TATE, TATE Modern, ref. T14778

Image released under [Creative Commons CC-BY-NC-ND \(3.0 Unported\)](https://creativecommons.org/licenses/by-nc-nd/3.0/).

[Yuko Nasaka](#) was part of the Japanese avant-garde that called for novelty in both the process of creation and its result. Stepping outside the boundaries of one's tradition to reinvent its core tenets requires a combination of loyalty and daring. Legal protection by design invites lawyers to rethink their relationship with text, and the need to amplify practical and effective legal protection by way of design - without, however, turning law into techno-regulation.

Policymakers, lawyers and other folk often speak of 'regulating technologies'.

This is an interesting phrase, because it can mean many things, depending on how you 'read' it. In the old days, most lawyers and policymakers would understand it in the sense of **technologies being the object of legal regulation**. The law can, for instance, impose requirements on the fabrication, design, sale and use of cars, knives, guns, housing, office space, washing machines, toys, or medical instruments. These requirements may concern safety, privacy, or a technology's potential to violate copyright, to disseminate child pornography or to generate pollution of the environment. They may be aimed at protecting weaker parties, critical infrastructure, national or public security or the environment. The default response that technologies are the object of regulation may, however, be changing.

The same phrase ('regulating technologies') can also refer to **technology as a 'subject' that is regulating human behaviour**, e.g. by way of speed bumps, digital rights management (DRM) technologies, news feed algorithms that determine what news we perceive and other default settings that determine our 'choice architecture'. Here the object of regulation is not a technology but human behaviour. So, technology can be either the object or the subject of regulation (and maybe both), whereas law is usually only seen as a subject of regulation (that which regulates).

This may be about to change due to the pervasive effects of two types of technologies that impact the environment of the law: machine learning (ML) applications that e.g. decide a person's credit worthiness or employability, and distributed ledger technologies (DLTs) that allegedly self-execute transactions and agreements without and beyond the law.

In this chapter the focus will be on how ML and DLTs change the environment of the law, the substance of legal goods (such as legal certainty, equality before the law, inalienability of personality rights, fairness and human dignity) and the extent to which this affects legal protection. One of the main challenges here concerns the regulatory effects of these novel technologies and the potential incompatibility of legal protection

and **'techno-regulation'** (defined as the regulatory effects of a technology, whether or not intended).

10.1 Machine learning (ML)

To understand the relevance of ML for legal protection it may help to look at a very simply example, such as AB testing. Imagine that the provider of a website or an IoT display or portal (webshop, platform, news agency, fanpage, smart energy meter display or portal) wants to 'optimise' its site to achieve higher performance in terms of influencing its visitors' purchasing behaviours, their reading habits, political opinions or energy usage behaviours. To do so, they employ software that enables the following process:

- The current webpage is called version A,
- its design is changed in a minimal way, e.g. the colour or place of a button, the position of a text block, the type and number of clicks required to access other webpages within the site.
- The slightly transformed page is called version B,
- 50% of visitors are directed to version A, the other 50% to version B.
- The software conducts automated measurement of their clickstream behaviours, possibly including those captured over the next day (possibly across various other websites based on tracking cookies).
- The software calculates which version generated the more desirable behaviour.
- The version that is more effective is now the default page.
- The whole process is repeated with another slight change.
- AB testing can be targeted at specific types of people or even be personalised.

Let's see if this qualifies as an example of ML. In his handbook on Machine Learning, Tom Mitchell recounts that:

A computer program is said to learn

- from experience E
- with respect to some class of tasks T
- performance measure P

if

- its performance at tasks in T,

- as measured by P,
- improves with experience E.

As to **type of task T**: This clearly sets out that machines do not learn anything if no task is defined. In this case the task will be defined by the website 'owner', together with the software provider, because the definition of what counts as desirable behaviour needs to be translated into machine-readable language. A webshop may find increased purchasing behaviour desirable, though they may also formulate more complex tasks, based on a segmentation of the visitors: they may prefer to increase the purchasing behaviour of people who buy expensive products, or of people who are likely to buy more than one product over the course of a specified period of time.

As to **experience E**: Note that the experience of this software is limited to clickstream behaviours of visitors of the page, even if they can be followed on other sites. It may be that their behaviours on other sites are not within the tracking-scope of the software provider (e.g. in offline shops or via another browser), whereas those unknown behaviours are actually more relevant for an inference about their preferences. The software's experience, however, is necessarily limited to the available training data.

As to **performance metric P**: It may be that a simple performance metric, such 'clicks on one product', or 'buys at least two products', does not really say much about the preferences of the visitors, because these behaviours are instances of situated behaviour that depends on many other factors. These other factors may be more indicative of their preferences. To test both versions against each other, one may need to test 6 or 7 different performance metrics to obtain a better picture of what qualifies as an accurate measure of achieving desirable behaviour.

10.1.1 Exploratory and confirmatory ML research design

AB testing can be done by way of an exploratory research design, meant to generate hypotheses about what kind of behaviour is more lucrative for the webshop. This implies recognition that such AB testing is a matter of real-time experimentation. As Hofman, Sharma and Watts write:

In exploratory analyses, researchers are free to study different tasks, fit multiple models, try various exclusion rules, and test on multiple performance metrics. When reporting their findings, however, they should transparently declare their

full sequence of design choices to avoid creating a false impression of having confirmed a hypothesis rather than simply having generated one (3). Relatedly, they should report performance in terms of multiple metrics to avoid creating a false appearance of accuracy.

Claiming success based on such AB testing is a very bad idea, and usually amounts to what statisticians call p-hacking. For a reliable prediction one needs a confirmatory research design, that provides tested and testable hypotheses about the preferences of visitors. As Hofman, Sharma and Watts write:

To qualify research as confirmatory, however, researchers should be required to preregister their research designs, including data preprocessing choices, model specifications, evaluation metrics, and out-of-sample predictions, in a public forum such as the Open Science Framework (<https://osf.io>).

As one can understand providers of marketing software that enables micro-targeting or underlies behavioural advertising will not be inclined to deposit their research design, including pre-processing choices, at the OSF. We may conclude from all this that:

1. ML is used to influence or nudge people into behaviours that are desirable from the perspective of whoever pays for the software, and,
2. such software may not be as effective as some may either hope or fear.

10.1.2 Implications of micro-targeting

Instead, the result of micro-targeting based on flawed research design may be that visitors of websites and users of IoT interfaces are confronted with a personalised choice architecture that is meant to lure them into desirable behaviour, but has two unintended consequences:

1. a fragmented public space that e.g. algorithmically favours extreme content to hold onto people's attention, and,
2. undesirable discrimination based on data points that systematically disadvantage certain categories of people.

These consequences are not necessarily envisaged by developers or users of the software; they are brought about by mistaking – potentially crappy – exploratory research design for robust confirmatory research design.

This raises two issues for legal protection. First, the mining and inferencing of behavioural data may violate **specified fundamental rights, such as privacy, data protection, non-discrimination and freedom of expression**. Behavioural data are often personal data and the mining of such data may infringe the privacy of those unaware of the rich profiles that can be built from such data, often combined with features that are inferred from such data. This may be in direct violation of the fundamental right to data protection, depending on how the data is mined and shared, on what ground, and with what purpose. Based on micro-targeting, the mining and inferencing of behavioural data may also violate the freedom of expression, since this right also includes the freedom to receive information free of censure. Micro-targeting based on AB testing could shield information from certain people, because there is no added value for the website owner in providing them with such information. We have entered the era of ad-driven-content, where the algorithms that infer what content is most conducive to attracting visitors may be prioritised in order to increase ad revenue. The use of 'low hanging fruit' to train ML algorithms will easily result in all kinds of unwarranted bias, due to the bias that is inherent in the so-called training data. Even if the right kind of data is available, the choice of the feature space, the hypothesis space, the task that is formulated and the performance metric that is chosen may result in a biased outcome that systematically discriminates people based on their race, ethnicity, religion, political preferences, gender or sexual orientation.

An example of such bias is the proprietary COMPAS software, sold by Equivant (formerly Northpointe), where COMPAS stands for Correctional Offender Management Profiling for Alternative Sanctions. COMPAS is used by courts in the US to assess the risk that an offender will recidivate (i.e. commit another offence after being released). This risk co-determines the parole or sentencing decisions. The risk score is based on a limited number of data points that have been found to correlate with re-offending. COMPAS is the result of an ML research design that tested 137 features to infer which 6 features were actually predictive. After Julie Angwin conducted own research on similar training data, she claimed that COMPAS discriminates people based on their race. More precisely, she found that

1. within the set of offenders that did not recidivate, the error rate for black persons may have been as high as that for white persons, but the error for black persons meant they were wrongly given a higher risk score,
2. whereas the error for white persons meant they were wrongly given a lower risk score.

According to Equivant this was the result of the fact that black persons on average had a higher risk of reoffending, Equivant accused Angwin of methodologically flawed methodologies, saying the laws of statistics were responsible for the disparate outcome of the risk score. As a use case the accusation of racial discrimination has generated a flood of scientific literature on fairness in ML, underpinned by requests for transparency and accountability, basically demanding that business and government employs FAT ML (fair, accountable, transparent machine learning applications). The literature demonstrates, that many different definitions of what qualifies as fair ML are possible, leading to different research designs. For instance, in the case of COMPAS, one could argue that fairness requires that the 'learner' is trained to come up with a risk score that does not result in disparate errors for black and white persons who do not recidivise. The COMPAS case returns in more detail in chapter 11, section 11.3.2.1.

10.1.3 Implications of micro-targeting for the Rule of Law

The second issue for legal protection concerns the extent to which decisions based on ML-inferences violate **core principles of the Rule of Law**, such as transparency and accountability, or more precisely **(1) the explainability of the decision-making process, (2) the justification of the decision, and, (3) the contestability of the decision**. The second and third requirements concern the decision. In public administration, decisions must be taken in accordance with the legality principle, meaning that the justification must be based on law and citizens have a right to contest the decision in a court of law. In the private sector, however, the freedom of contract and the freedom to dispose of one's property may provide the justification. These freedoms, however, are restricted, for instance due to the prohibition to discriminate in the context of employment, or to discriminate based on gender or race. Both in public administration and commercial enterprise, ML-based decision-making may incur invisible discrimination that is actually prohibited, for instance based on race. Such discrimination will often be unintended and invisible because it is based on a concerted set of features that correlate with race and therefore act as proxies for race. This means that such discrimination need not be based on a deliberate attempt to use race as a relevant feature; even if one removes race as a feature altogether, the proxies will probably sustain the discrimination. Legally speaking this may be qualified as indirect discrimination, which is often explicitly defined and prohibited (unless

justified). What matters here is that without explainability of the ML application, it may be very difficult to check the extent to which discrimination occurs.

Apart from prohibited discrimination, decision-making based on applied ML may have other repercussions. Imagine that the risk profile that is applied to a person is based on

1. the average risk in a specified class of people,
2. whereas that average risk does not apply to each member of that class.

In that case individuals are basically treated on the basis of a score that probably does not apply to them. Even if such classification of individuals does not involve prohibited discrimination, it may be seen as unfair. For instance, on average women may have a risk of 1 out of 8 to suffer from breast cancer. Depending on a woman's age, the occurrence of breast cancer in her ancestry and family, her lifestyle and other factors, her risk will stray from '1 out of 8', to potentially much higher or lower risk. Treating each and every woman as if her risk is 1 out of 8 would therefore be unwise, and in the case of e.g. a health insurance premium one might argue this is unfair. This explains why the explainability of decisions based on the application of ML has become a serious issue of legal protection.

In terms of the GDPR, personalised targeting based on machine learning would most often fall within the scope of art. 4(4):

'profiling' means any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements;

Art. 21 GDPR stipulates that data subjects have a '**right to object**' to profiling based on the grounds of art. 6(e) and (f), i.e. based on a public task or public authority, or on the legitimate interest of the data controller, and to profiling 'to the extent that it is related to direct marketing'. Next to this, data subjects have a '**right not to be subject**' to profiling when this is a form of automated decision-making that significantly affects data subjects (art. 22). In section 10.3.3.3 we will explain to what extent the right not to be subject to automated decisions provides 'legal protection by design'. Note that this right is not only applicable to profiling but also to other types of automated decisions, such as those involving self-executing code.

10.2 Distributed Ledger Technologies (DLTs), smart contracts and smart regulation

As the development and usage of DLTs and/or Blockchains are in full flux, so is the terminology. Sidestepping discussions on the correctness of either term we will use the term DLT to cover the whole range of technologies framed as

- **distributed databases (ledgers)**
- that store transactions based on
- **decentralised infrastructures (the core code)**
- that enable self-executing code, based on
- a specific combination of **security technologies** (notably hashing and encryption)
- that incentivise 'miners' or 'validators' to partake in a reasonably trustworthy **consensus mechanism**
- that supposedly ensures the **integrity** of the data stored in the ledger, and of the sequence of such storage.

DLTs are often promoted as providing a form of **trustless computing** that enables immutable, transparent and secure storage of transactions, with a guarantee against ex-post manipulation of previous transactions, thus ensuring the integrity of both the sequence and the content of the transactions (which e.g. protects against 'double spending'). Often DLTs are 'sold' as enabling **disintermediation**, meaning that users need not connect with a traditional institution to engage in trustworthy transactions with parties they do not know or do not trust. The idea is that the ledger allows them to interact with others in a fully transparent way, with certainty that neither the other party nor any third party can manipulate stored transactions. In a sense the promise is that the technology can take over the role of a trusted intermediary by way of a fully predictable sequence of events that self-executes tamper-free transactions.

Before unpacking these claims it is crucial to distinguish between **public and private** and between **non-permissioned and permissioned DLTs**, as well as their combinations. The difference between public and private DLTs can be defined in terms of who can '**read**' the content, and the difference between permissioned and non-permissioned in terms of who can **add or 'write'** new content. Bitcoin builds on public non-permissioned DLTs, meaning that anybody can check the content and submit new content. By now, commercial enterprises, financial institutions as well as government

agencies probe business cases for DLTs, often resorting to private permissioned versions that lack part of the lure of a decentralised system, because with private permissioned DLTs only a specified set of players is allowed to read and write on the ledger. Let us note up front that this means that private non-permissioned DLTs basically require users to trust:

1. the traditional intermediary that employs the DLT, and
2. those who write the code for a particular type of transaction, and
3. those who write the protocols that constitute the infrastructure ('core developers').

Taking into account that most users do not understand computer code, such DLTs basically reinforce the role of the institutions that employ them; they require more trust, not less and they certainly do not achieve disintermediation.

10.2.1 Smart contracts and smart regulation

For this chapter, the relevance of DLTs concerns so-called **smart contracts** and **smart regulation**, i.e. the use of DLT to self-execute either an agreed contract or specified policy decisions based on regulatory competence. As to the first we can think of a contract of sale that self-executes once triggered (when the system detects payment it transfers the object, or the other way round). Note that this may work perfectly if both the payment (e.g. cryptocurrency) and the assets (e.g. an electronic proof of ownership) are within the system (often referred to as on-chain). Off-chain payments or off-chain transfer of assets, however, will require the use of 'oracles', i.e. software applications that interface between the ledger and the real world, or other systems.

As to the self-execution of regulatory constraints this assumes that a competent authority **translates** its policy into machine readable code (an act of interpretation) and **defines** what kind of data-input triggers the execution of the code (another act of interpretation). Some have observed that this conflates legislation with its execution and even with adjudication (in case of disagreement about the content of the contract). This would mean that the checks and balances of the Rule of Law, notably the separation of the powers of legislation, administration and adjudication, are disrupted. This, in turn, would require new types of safeguards (legal remedies) to enable the contestation of the ensuing decisions - thus making sure smart regulation and smart contracts remain 'under the Rule of Law'.

A quick round-up of critique regarding some of the claims made about DLTs, notably with regard to smart contracting and smart regulation:

Immutability:

- if parties agree on the code (and if the code is not corrupted or otherwise disabled), their agreement will be executed without recourse to remedies or reinterpretation. One could argue that the immutability of the self-executing code entails legal certainty, though changing circumstances may result in the opposite, precisely because the code is not adaptive;
- if a party does not understand code and agrees to oral or written communication that differs from the code, the immutability becomes a problem and will certainly not deliver legal certainty;
- in the case of a permissioned DLT the immutability may be overruled, depending on the governance structure (which is not decentralized in that case).

Trustless computing:

- if parties do not know each other but wish to engage in transactions, a smart contract **is often said to enable trustless computing**, to the extent that the protocols of the platform and the code of the contract are trustworthy and do what parties legitimately expect;
- parties are basically asked to trust that the protocols of the underlying infrastructure are trustworthy, and the program language aligns with the intent expressed;
- in the case of permissioned private DLTs, users are required to trust those who control the DLT, the protocols that form its 'constitution', and the code that is run on their behalf or on behalf of the other party.

Transparent transactions:

- to the extent that parties have access to the source code of the infrastructure (public DLTs) and to the programming code (the smart contract itself), and to the extent they can understand it, there is **transparency**;
- if parties have no access to the code (private DLTs), or do not understand code, transparency cannot be assumed.

Security:

- if all works as hoped for, the execution of the contract is **secured**;

- if the protocols and/or code are sloppy or if new bugs appear, in the case of a so-called 51% attack, or if the miners/validators stop maintaining the system, the contract and/or the whole system may be hacked and/or dissolve.

Anonymity:

- depending on how parties access the smart contract ecosystem, they may remain anonymous or at least pseudonymous;
- transparency in public DLTs may imply that anonymity is an illusion, also considering the use of e.g. behavioural analytics to reidentify users.

Safety:

- to the extent that the underlying system and the smart contract itself operate as agreed, the transaction could be called safe;
- if circumstances change, requiring adaptation of the contract or decision, the self-executing nature of the code may create unsafe outcomes for users, especially if they cannot identify or sue whoever is liable (the provider of the DLT, the contracting party, or the government agency may e.g. be in another jurisdiction);
- if either the underlying system or the smart contract code is hacked, if off-chain input is incorrect, or if the provider cannot be held liable, one or more parties to the contract may lose their input.

Correctness:

- to the extent that off-chain input is correct, the on-chain execution of the contract will be executed correctly (as long as the code does what parties agreed to);
- to the extent that off-chain input is incorrect, the error or false input is automated (and due to the immutability this may be hard to correct).

From the perspective of law, the employment of DLTs raises many questions. In the context of this chapter I focus on whether operating self-executing code via a DLT must be seen as 'legal by design' or as 'legal protection by design' (preparing the ground for the topic of 10.3). Do smart contracts or smart regulation guarantee that the behaviour of parties of the contract or the addressees of regulation is 'legal by design' or 'legally compliant by design'? To prepare the ground I will first discuss the question whether smart contracts are **contracts in the legal sense** (10.2.2), and whether smart regulation is **law in the legal sense**, to be discussed in (10.2.3).

10.2.2 The legal status of 'smart contracts' under private law

As to contracts in the legal sense, we need to investigate what legal conditions must be fulfilled for 'something' to qualify as a legally binding contract. These legal conditions can be found in private law, which is mostly national law (there is e.g. no binding European private law). I refer to section 3.2.2, where some of the basics of a valid contract were discussed, based on Dutch private law. Though other jurisdictions may have different legal conditions some of the underlying assumptions remain the same.

First of all, an obligatory agreement is a more-sided act where parties aim to establish specified legal effects, such as a legal obligation to pay a price in exchange for the transfer of the property of a good or the provision of a service. In the common law a contract requires 'consideration' (tit for tat) to be valid. The intent to be bound by the contract can be inferred from the declarations of the parties, though sometimes it can also be inferred from their actions – if such actions have generated the legitimate expectation that one has consented to the contract. In most – if not all – jurisdictions, a valid contract requires a sufficiently specified offer by one party that is accepted by the other party. If the acceptance was mistakenly inferred from certain behaviours, whereas in fact there was no acceptance, the contract would be considered **void** (as one of the constitutive conditions does not apply). If the offering party, however, legitimately inferred acceptance from the other party's behaviour, the contract may nevertheless be valid. Also, most jurisdictions have safeguards in place in the case that acceptance is based on duress or undue influence, mutual mistake or fraud. If this can be proven, the contract becomes **voidable**, at the request of the party that wishes to 'undo' the contract. Again, in most jurisdictions, there are no formal requirements for contracts in general, which means they can be concluded in whatever way (speech, writing, shaking hands, real-time exchange of a good and the payment). Specific contracts, such as the sale of real-estate, do have formal requirements (e.g. of a deed) which usually involves a trusted third party (e.g. a notary public).

Does a smart contract qualify as a legal contract? Based on the above, there are at least three issues:

1. Can we assume that sending a message to a smart contract (code on the ledger) implies the will to be bound (and thus acceptance of an offer)?
2. Does computer code count as an expression of the content of a contract (and thus as a sufficiently specified offer)?

3. Can a party invoke voidability because they cannot read the code?

The fact that most contracts have no formal requirements could be used as an argument that sending a specific message to the code on the DLT may count as an expression of one's intent to enter into the contract as defined in the code. However, the jury is still out on whether computer code counts as an expression of the content of a contract just like a written contract supposedly does. To count as such an expression, the code must be sufficiently determinate for both parties to understand the legal effect of the contract (i.e. the legal obligations it generates). If the accepting party does not read code, they can either

- argue that they did not accept the content of the code because their legitimate expectations about that content - as inferred from negotiations, advertising or other expressions by the offering party - do not match the code, which means the contract is void; or they can
- argue that the contract is voidable because of e.g. mistake or fraud.

If we assume that the contract is valid, we need to still look into the legal effect of a valid contract, because in most jurisdictions such legal effect is not limited to the literal wording of the contract, but extends to

1. what both parties should reasonably expect, considering the circumstances,
2. while a number of legal constraints may apply that co-determine the content of the contract.

The latter constraints may derive from either private or public mandatory law (see 3.1.2 and 8.1.1), which cannot be overruled by contractual stipulations (whether in speech, writing or code). To build flexibility into a contract, or a policy, they often contain concepts with an open texture that leave parties or competent authorities some room to adapt the contract to the concrete circumstances that cannot all be foreseen. Think of terms such as reasonably, timely, state of the art or trustworthy, which can only be interpreted in the light of the circumstances that parties confront when performing the contract. Unforeseen changes in circumstances will also have an impact on the content of the ensuing legal obligations, as when one party can claim force majeure. Whereas the 'smart contract' will self-execute, force majeure may overrule the obligation to perform the contract, meaning the execution may have to be undone (which may be impossible and/or the party that benefits may not be identifiable, or in a far-out jurisdiction, meaning they cannot be sued).

All this also happens to 'normal' contracts, and with 'normal' decision making in public administration, but it is crucial to highlight that smart contracts and algorithmic decision-making in the sense of smart regulation do not necessarily solve these problems and may indeed create extra problems, precisely due to the non-adaptive nature of self-executing code. Those who wish to remedy these new problems by creating adaptive code must realise that this implies foreseeing all possible future scenarios, which is by definition not possible. Though it may prevent some problems, it still implies that legislation (a contract can be seen as legislating how parties should act), execution (a contract should clarify what counts a performance) and interpretation (the meaning of a contract depends on the circumstances) are all predetermined upfront by whoever writes the code.

Legal scholar Allen argues that smart contracts will be part of what he has called the 'contract-stack', which involves speech acts, behaviour, written documents, deeds, electronically signed documents and – potentially – also self-executing code. This implies that contract law will be transformed to accommodate the use of self-executing code, e.g. by way of legislation, case law and doctrinal innovation. Similar arguments can be made for smart regulation, which could similarly be seen as a 'regulatory-stack', involving legislative Acts that grant regulatory competences, policy documents, government agency's behaviour patterns, decision-making processes and procedures, and – potentially – also self-executing code.

10.2.3 The legal status of 'smart regulation' under public law

With the term 'regulation' I refer to rules promulgated by public administration or independent supervisors instituted by an Act of the legislature (usually called 'regulators' in the US and the UK, e.g. the Federal Trade Commission; in the EU we can think of the EDPS or the national DPAs). Such rules are either

1. part of an explicitly attributed competence to create and impose rules, or
2. a way to provide transparency about how a regulator will make use of its discretionary competence (in that case those rules form a policy).

Many government decisions affect individual citizens, such as the granting of a permit, social security, or a decision on taxation. Many of the arguments provided in the previous section can be repeated here, and do not merely apply to implementation via DLTs but also to other forms of algorithmic (automated) decision-making. It simply

means that the relevant rules are interpreted and translated into non-ambiguous code, to enable their self-execution.

As with private law contracts, smart regulation will necessarily be overinclusive and underinclusive (or both), due to its lack of adaptive flexibility. The need to formalise will – in a sense – freeze future responses into a template that necessarily overlooks changing circumstances and may not reflect developments in case law, which may result in the code violating rights instead of enforcing compliance. In that respect it is crucial to remember that these rules and policies, as well as their machinic automation, fall under the Rule of Law. Instead of understanding 'smart regulation' as a kind of law, it is therefore better understood as public administration. This means these rules and policies, as well as their machinic translations, must at some point be contestable in a court of law. Those subject to decisions based on smart regulation should be capable of requesting a justification of the decision in accordance with the legality principle. Note, however, that a **justification** is not equivalent with an **explanation**, which rather serves as a means to make the decision contestable as to its justification.

10.3 'Legal by Design' or 'Legal Protection by Design'?

Some authors claim that self-executing code could be used to ensure that the conduct of legal subjects will be 'legal by design' (LbD). What they mean to say is that one can interpret the content of a contract, the content of policy guidelines, or even the content of legislation such that it becomes amenable to a translation into computer code. So-called Turing complete languages have been developed in the realm of DLTs, to write 'smart contracts' that – as we have seen in section 10.2 - supposedly self-execute whatever has been agreed by the parties. One can imagine similar attempts to ensure compliance at the level of regulatory rules.

10.3.1 Legal by design (LbD)

LbD is a subset of what other authors have termed '**techno-regulation**'. This refers to the fact that technologies often induce or inhibit and enforce or preclude certain types of behaviours, which has a de facto regulatory effect. As mentioned in the introduction to this chapter, such regulatory effects can be

1. the result of deliberate design of a technology (requirements that specify which functions must be engineered) or
2. the unintended result of design choices made with other intentions, or of unforeseen usage of the technology.

In the latter cases we speak of side-effects, though we should take note that such side-effects may be more prominent or influential than the intended effects. LbD is a specific subset of techno-regulation that is (1) the result of deliberate design choices, where (2) those choices aim to ensure compliance with legal obligations by way of technical enforcement.

LbD involves two steps. First, it involves a specific (non-ambiguous) **interpretation** of the relevant legal norm, and, second, it involves the **translation** of that interpretation into a programming language. Note that these steps can be analytically distinguished, but may be conflated in practice (thus hiding the act of interpretation). Due to the need to select an interpretation that can be translated into unambiguous machine language, such interpretations may be overinclusive or underinclusive compared to the relevant legal norm.

For example, a legal obligation for an employee to drive a truck from A to B within a reasonable time scale could be part of a smart contract between an employer and an employee. As the performance of the contract takes place off-chain, an oracle must be put in place to provide clear signals about whether or not this legal obligation has been fulfilled. To define what performance counts as 'reasonable', taking into account various types of circumstances, the contract must be interpreted beforehand and translated into a set of input variables for the oracle. As discussed in section 10.2.2 'reasonableness' is not a subjective concept under contract law as it will have to be interpreted in line with relevant case law, while taking account of the unique circumstances of the case at hand. This makes it highly unlikely that a smart contract can be equated with 'legal compliance by design', due to the rigidity of the code.

Another example could be that the legally allowed level of pollution caused by a car is integrated into smart regulation that rules out delivery of non-compliant cars by the car manufacturer. To enable this, however, the cars must be tested before leaving the factory, which necessarily disregards the actual pollution caused on the motorway. This, again, implies that there is no absolute guarantee that the car manufacturer is 'legally compliant by design'.

In both examples LbD seems to be an inept term for what is actually achieved. As long as this is kept in mind, incorporating checks and balances (including legal remedies if the lawfulness is contested), smart contracts and smart regulation may nevertheless contribute to (though not guarantee) compliance.

10.3.2 Legal protection by design (LPbD)

Legal protection by design (LPbD) is another matter. It does not aim to guarantee enforcement of whatever legal norm, but rather aims to ensure that legal protection is not ruled out by the affordances of the technological environment that determines whether or not we enjoy the substance of fundamental rights. The term 'legal' here, involves two important requirements of law in the context of a constitutional democracy:

1. The scope of LPbD should be determined by way of democratic participation, for instance in the context of participatory technology assessment and involvement of the democratic legislature;
2. Those subject to such LPbD should be able to contest its application in a court of law.

Techno-regulation in general does not include these requirements and neither does LbD, which is often focused on excluding the involvement of trusted third parties. These two requirements thus distinguish LPbD from other types of 'by design' solutions, for instance 'value sensitive design' or 'privacy by design'. The latter are often proposed as ethical requirements. However, such ethical design is problematic for two reasons. First, as it will not level the playing field, companies that apply such ethical design may be pushed out of the market. Second, it makes protection dependent on the ethical inclinations of those who develop and market the choice architecture of citizens, instead of demanding that such choice architecture must meet minimum standards that provide effective and practical protection. For readers interested in confrontation of law and ethics, see chapter 11.

10.3.3 LPbD in the GDPR

10.3.3.1 Data Protection Impact Assessment

Three interesting examples of LPbD can be found in the GDPR. First, the legal obligation to conduct a data protection impact assessment (DPIA) in art. 35, which is

compulsory if the introduction of a new technology is likely to present a high risk to the rights and freedoms of data subject:

1. Where a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data. A single assessment may address a set of similar processing operations that present similar high risks.

(...)

3. A data protection impact assessment referred to in paragraph 1 shall in particular be required in the case of:

- a) a systematic and extensive evaluation of personal aspects relating to natural persons which is based on automated processing, including profiling, and on which decisions are based that produce legal effects concerning the natural person or similarly significantly affect the natural person;
- b) processing on a large scale of special categories of data referred to in Article 9(1), or of personal data relating to criminal convictions and offences referred to in Article 10; or
- c) a systematic monitoring of a publicly accessible area on a large scale.

(...)

7. The assessment shall contain at least:

- a) a systematic description of the envisaged processing operations and the purposes of the processing, including, where applicable, the legitimate interest pursued by the controller;
- b) an assessment of the necessity and proportionality of the processing operations in relation to the purposes;
- c) an assessment of the risks to the rights and freedoms of data subjects referred to in paragraph 1; and

d) the measures envisaged to address the risks, including safeguards, security measures and mechanisms to ensure the protection of personal data and to demonstrate compliance with this Regulation taking into account the rights and legitimate interests of data subjects and other persons concerned.

(...)

11. Where necessary, the controller shall carry out a review to assess if processing is performed in accordance with the data protection impact assessment at least when there is a change of the risk represented by processing operations.

Recital (75) adds some considerations concerning the question what constitutes the likelihood of a high risk to the rights and freedoms of natural persons.

The risk to the rights and freedoms of natural persons, of varying likelihood and severity, may result from personal data processing which could lead to physical, material or non-material damage, in particular:

- where the processing may give rise to discrimination, identity theft or fraud, financial loss, damage to the reputation, loss of confidentiality of personal data protected by professional secrecy, unauthorised reversal of pseudonymisation, or any other significant economic or social disadvantage;
- where data subjects might be deprived of their rights and freedoms or prevented from exercising control over their personal data;
- where personal data are processed which reveal racial or ethnic origin, political opinions, religion or philosophical beliefs, trade union membership, and the processing of genetic data, data concerning health or data concerning sex life or criminal convictions and offences or related security measures;
- where personal aspects are evaluated, in particular analysing or predicting aspects concerning performance at work, economic situation, health, personal preferences or interests, reliability or behaviour, location or movements, in order to create or use personal profiles;
- where personal data of vulnerable natural persons, in particular of children, are processed; or
- where processing involves a large amount of personal data and affects a large number of data subjects.

Art. 35 basically requires controllers to err on the side of caution by foreseeing risks to the rights and freedoms of natural persons. One could qualify this as the introduction

of the principle of precaution in data protection law. Note that the assessment does not merely regard potential violations of the rights and obligations stipulated in the GDPR but focuses on 'rights and freedoms' in a more general sense, which links up with the goal of the GDPR as formulated in art. 2.2: '[t]his Regulation protects fundamental rights and freedoms of natural persons and in particular their right to the protection of personal data'. Moreover, the assessment of such a risk is not limited to data subjects but refers to 'natural persons', which includes e.g. individuals that run a risk to be discriminated even though their personal data are not (yet) being processed.

10.3.3.2 Data protection by default and design (DPbD)

Art. 35.7(d) clearly indicates that a DPIA incorporates an assessment of the need for data protection by default and by design (DPbD), as it requires an inventory of 'the measures envisaged to address the risks, including safeguards, security measures and mechanisms to ensure the protection of personal data and to demonstrate compliance with this Regulation taking into account the rights and legitimate interests of data subjects and other persons concerned'. This brings us to art. 25, which requires to design systems that process personal data in a way that ensures data minimisation by default, while incorporating all other GDPR obligations into the design of the system:

1. Taking into account the state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for rights and freedoms of natural persons posed by the processing, the controller shall,
 - both at the time of the determination of the means for processing and at the time of the processing itself,
 - implement appropriate technical and organisational measures, such as pseudonymisation, which are designed to implement data-protection principles, such as data minimisation,
 - in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and protect the rights of data subjects.
2. The controller shall implement appropriate technical and organisational measures for ensuring that,
 - by default,

- only personal data which are necessary for each specific purpose of the processing are processed.

That obligation applies to

- the amount of personal data collected,
- the extent of their processing,
- the period of their storage and
- their accessibility.

In particular, such measures shall ensure that by default personal data are not made accessible without the individual's intervention to an indefinite number of natural persons.

Here again, we can observe a requirement to err on the side of caution, basically echoing longstanding security principles, such as 'select before you collect'. In paragraph 2, for instance, we read that technical and organisational measures must be in place to ensure that only data that are necessary for each specific processing purpose are being processed (data minimisation and purpose limitation). Though 'privacy by design' has deep roots in privacy engineering communities, the big difference here is that this is no longer a matter of choice, or of 'being ethical' about one's processing operations.

DPbD is not to be taken lightly, but also does not require what is not feasible. The requirement takes into account 'the state of the art, the cost of implementation and the nature, scope, context and purposes of processing' (first paragraph), meaning that measures must be doable, also in light of the business model. However, this does not mean that anything goes if the business model does not fly without taking disproportionate risks with the rights and freedoms of natural persons. Here again, as with the DPIA, those risks must be taken into account when designing (engineering) the processing operations. The proportionality depends on 'the risks of varying likelihood and severity', meaning that the higher the risks the more protection must be implemented 'by design'.

Clearly, both the DPIA and DPbD take a so-called 'risk approach' to the protection of personal data. Though some have interpreted this as a sign that the EU legislature favours a cybernetic understanding of risk and regulation to a rights-based approach, it seems more likely that the risk approach aims to introduce some lawfully required

precaution on the side of data controllers, to sustain and enable an effective and practical protection of the rights and freedoms of natural persons.

When reading the carefully crafted, balanced and reasonably complex requirements to embed relevant legal norms in the architecture of personal data processing, it is evident that neither the DPIA nor DPbD aim to produce processing systems that are 'legal by design'. Instead, they warrant and introduce legal obligations to embody 'legal protection by design' in technical systems that would otherwise render the protection of an individual's rights and freedom illusory.

10.3.3.3 Automated decisions

This brings us to a **third example of LPbD** in the context of the GDPR that is highly relevant for both ML applications and DLTs, as it targets the implications of automated decisions. Art. 22 GDPR reads:

The data subject shall have the right not to be subject to a decision

- based solely
- on automated processing, including profiling,
- which produces legal effects concerning him or her or
- similarly significantly affects him or her.

The legal effect of the four legal conditions (two of which are alternative), is a prohibition. Even though this prohibition is formulated in a rather complicated way, the EDPB (formerly Art. 29 Working Party) has clarified that this 'right not to be subject to' must be understood as a prohibition.¹ Note that each term in this set of legal conditions requires interpretation that is not obvious in the light of technologies such as ML and DLT. For instance, which of the decisions taken by machines in the course of a machine learning operation qualifies as a decision in the sense of art. 22.1: the decision of an algorithm to adept weights within a neural net, where such a decision will result in a refusal to provide credit? or, the decision to select 4 of the 19 features that have some impact on a specified health risk, where such a decision results e.g. in a person being advised to undergo an operation or in a person being charged with tax fraud? Does 'solely' refer to machine decisions that directly affect a data subject (e.g. online acceptance of a health insurance), or also to decisions that have been prepared by a software program but are 'stamped' by a human person who, however, does not understand how the system came to its conclusion and cannot explain to the data subject why she was not e.g. selected for a job interview? The

EDPB finds that '[t]he controller cannot avoid the Article 22 provisions by fabricating human involvement'.² Does the fact that automated processing is qualified as 'including profiling' imply that 'smart contracts' that do not involve profiling in the sense of art. 4(4) do not fall within the scope of art. 22? Note that English grammar answers that question, due to the fact that a comma is inserted after processing (check the rules for restrictive and non-restrictive modifiers).

When does a decision produce legal effect? The EDPB clarifies that this is the case if the decision 'affects someone's legal rights, such as the freedom to associate with others, vote in an election, or take legal action. (...) affects a person's legal status or their rights under a contract'.³ Any other 'similarly significant effect' also results in a prohibition, e.g. as the EDPB writes:⁴

For data processing to significantly affect someone the effects of the processing must be sufficiently great or important to be worthy of attention. In other words, the decision must have the potential to:

- significantly affect the circumstances, behaviour or choices of the individuals concerned;
- have a prolonged or permanent impact on the data subject; or
- at its most extreme, lead to the exclusion or discrimination of individuals.

It is difficult to be precise about what would be considered sufficiently significant to meet the threshold, although the following decisions could fall into this category:

- decisions that affect someone's financial circumstances, such as their eligibility to credit;
- decisions that affect someone's access to health services;
- decisions that deny someone an employment opportunity or put them at a serious disadvantage;
- decisions that affect someone's access to education, for example university admissions.

Having laid out the scope of the prohibition, art. 22 continues with 3 exceptions:

2. Paragraph 1 shall not apply if the decision:

- a) is necessary for entering into, or performance of, a contract between the data subject and a data controller;

b) is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or

c) is based on the data subject's explicit consent.

Here again, a number of questions can be raised. The reader is advised to carefully study the EDPD Guidelines on Automated Individual Decision Making and Profiling, to gain a proper understanding of how these exceptions must be interpreted.

1. In the cases referred to in points (a) and (c) of paragraph 2, the data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision.

So, in the case of a decision based on automated processing that is necessary for a contract or a decision based on consent, access to human intervention is required, both to express one's point of view and to contest the decision. This is related to recital (71) that adds another requirement:

In any case, such processing should be subject to suitable safeguards, which should include specific information to the data subject and the right to obtain human intervention, to express his or her point of view, to obtain an explanation of the decision reached after such assessment and to challenge the decision.

Here we find the right to obtain an explanation of the decision, which many authors interpret as being a precondition to be able to contest the decision (as required in art. 22.3). By now, a number of scientific papers have been published on the 'the right to an explanation' and 'explainable AI', which is deemed highly relevant also due to potential unwarranted bias. This 'right to an explanation' can also be read into the transparency requirements in art. 13.2(f), 14.2(g) and 15.1(h) which all three require that the following information will be provided:

- the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases,
- meaningful information about the logic involved, as well as
- the significance and the envisaged consequences of such processing for the data subject.

This basically means that data controllers have a **legal obligation** provide such information, both when the data has been provided by the data subject (art. 13), and when data has not been obtained from the data subject (art. 14), and that data subjects have a **right** to obtain such information (art. 15). Note that the obligation to provide these 3 types of information does not depend on a request by the data subject but must be provided anyway. Just imagine what this could mean for an IoT system that runs on real-time ML applications, or for online credit applications based on ML inferences of credit worthiness.

4. Decisions referred to in paragraph 2 shall not be based on special categories of personal data referred to in Article 9(1), unless point (a) or (g) of Article 9(2) applies and suitable measures to safeguard the data subject's rights and freedoms and legitimate interests are in place.

The exceptions generally do not apply to automated decisions that are based on art. 9 data. Now think of unintended machine bias based on proxies that result in indirect racial discrimination as described above in 10.1. There is no case law yet on how this prohibition must be interpreted, but we can imagine that art. 22.4 may provide far reaching protection if properly interpreted in a balanced way.

Art. 22 repeatedly speaks of 'suitable measures to safeguard the data subject's rights and freedoms and legitimate interests'. The EDPD clarifies that this includes technical measures. They write:

Errors or bias in collected or shared data or an error or bias in the automated decision-making process can result in:

- incorrect classifications; and
- assessments based on imprecise projections; that
- impact negatively on individuals.

Controllers should carry out frequent assessments on the data sets they process to check for any bias, and develop ways to address any prejudicial elements, including any over-reliance on correlations.

Systems that audit algorithms and regular reviews of the accuracy and relevance of automated decision-making including profiling are other useful measures.

Controllers should introduce appropriate procedures and measures to prevent errors, inaccuracies or discrimination on the basis of special category data. These

measures should be used on a cyclical basis; not only at the design stage, but also continuously, as the profiling is applied to individuals. The outcome of such testing should feed back into the system design.

These types of 'safeguards' exemplify how LPbD can be turned into an operational requirement that guides the design of personal data processing systems, ruling out unwarranted violations of data protection law, while providing practical and effective protection at the level of the technical and organisational infrastructure.

References

On machine learning:

Mitchell, Thomas. 1997. *Machine Learning*. 1 edition. New York: McGraw-Hill Education.

Mitchell, Tom M. 2017. 'Key Ideas in Machine Learning'. In *Machine Learning*, draft for the second edition, 1-11.

On p-hacking and other risks in ML:

Berman, Ron, Leonid Pekelis, Aisling Scott, and Christophe Van den Bulte. 2018. 'P-Hacking and False Discovery in A/B Testing'. SSRN Scholarly Paper ID 3204791. Rochester, NY: Social Science Research Network.

<https://papers.ssrn.com/abstract=3204791>.

Hofman, Jake M., Amit Sharma, and Duncan J. Watts. 2017. 'Prediction and Explanation in Social Systems'. *Science* 355 (6324): 486-88.

<https://doi.org/10.1126/science.aal3856> (quotation p. 487).

Hildebrandt, Mireille. 2018. 'Preregistration of Machine Learning Research Design. Against P-Hacking'. In *BEING PROFILED: COGITAS ERGO SUM*. Amsterdam University Press.

On bias in ML applications:

Angwin, Julia, Jeff Larson, Surya Mattu, and Kirchner. 2016. 'Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks.' ProPublica. 23 May 2016. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

Barocas, Solon, and Andrew D. Selbst. 2016. 'Big Data's Disparate Impact'. *California Law Review* 104: 671-732.

Chouldechova, Alexandra. 2017. 'Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments'. *Big Data* 5 (2): 153-63.
<https://doi.org/10.1089/big.2016.0047>.

Yong, Ed. 2018. 'A Popular Algorithm Is No Better at Predicting Crimes Than Random People'. *The Atlantic*, January.
<https://www.theatlantic.com/technology/archive/2018/01/equivant-compass-algorithm/550646/>.

On the potential and real effects of ML on public space, democracy, and freedom of expression:

Pariser, Eli. 2011. *The Filter Bubble: What the Internet Is Hiding for Your*. Penguin.

Sunstein, Cass R. 2016. *The Ethics of Influence: Government in the Age of Behavioral Science*. Cambridge University Press.

Tufekci, Zeynep. 2018. 'How Social Media Took Us from Tahrir Square to Donald Trump'. *MIT Technology Review*, no. September/October.
<https://www.technologyreview.com/s/611806/how-social-media-took-us-from-tahrir-square-to-donald-trump/>.

Re the fundamentals of 'smart contracts':

Buterin, Vitalik. 2014. 'A Next-Generation Smart Contract and Decentralized Application Platform. White Paper.' Ethereum Platform.

Nakamoto, Satoshi. 2008. 'Bitcoin: A Peer-to-Peer Electronic Cash System'.
<http://www.bitcoin.org/bitcoin.pdf>.

Szabo, Nick. 1997. 'Formalizing and Securing Relationships on Public Networks'. *First Monday* 2 (9). <http://firstmonday.org/ojs/index.php/fm/article/view/548>

Re the writing of 'smart contracts':

Seijas, Pablo Lamela, Simon J. Thompson, and Darryl McAdams. 2016. 'Scripting Smart Contracts for Distributed Ledger Technology'. IACR Cryptology EPrint Archive 2016: 1156.

Re 'blockchain' and the GDPR:

Finck, Michèle. 2018. 'Blockchains and Data Protection in the European Union'. *European Data Protection Law Review* 4 (1): 17-35.
<https://doi.org/10.21552/edpl/2018/1/6>.

Re compatibility of 'smart contracts' with art. 22 GDPR:

Art. 29 Working Party WP251rev.01, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679:
https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053

CNIL, September 2018, 'Solutions for a responsible use of the blockchain in the context of personal data':
<https://www.cnil.fr/sites/default/files/atoms/files/blockchain.pdf>

Re legal contracts and 'smart contracts':

Allen, J.G. 2018. 'Wrapped and Stacked: "Smart Contracts" and the Interaction of Natural and Formal Language'. *European Review of Contract Law* 14 (4): 307-343.
<https://doi.org/10.1515/ercl-2018-1023>.

Cornell, N. and K. Werbach. 2017. 'Contracts Ex Machina'. *Duke Law Journal* 67 (2): 313-82.

Raskin, M. 2017. 'The Law and Legality of Smart Contracts'. *Georgetown Law and Technology Review* 1 (2): 304-41.

Verstraete, Mark. 2018. 'The Stakes of Smart Contracts'. SSRN Scholarly Paper ID 3178393. Rochester, NY: Social Science Research Network.
<https://papers.ssrn.com/abstract=3178393>.

Re Legal by Design and Legal Protection by Design

Filippi, Primavera De, and Samer Hassan. 2016. 'Blockchain Technology as a Regulatory Technology: From Code Is Law to Law Is Code'. *First Monday* 21 (12).
<http://firstmonday.org/ojs/index.php/fm/article/view/7113>.

Hildebrandt, Mireille. 2017. 'Saved by Design? The Case of Legal Protection by Design'. *NanoEthics*, August, 1-5. <https://doi.org/10.1007/s11569-017-0299-0>.

Lippe, Paul, Daniel Martin Katz, and Dan Jackson. 2015. 'Legal by Design: A New Paradigm for Handling Complexity in Banking Regulation and Elsewhere in Law'. *Oregon Law Review* 93 (4). <http://papers.ssrn.com/abstract=2539315>.

Van den Berg, Bibi, and Ronald E. Leenes. 2013. 'Abort, Retry, Fail: Scoping Techno-Regulation and Other Techno-Effects'. In *Human Law and Computer Law: Comparative Perspectives*, edited by Mireille Hildebrandt and Jeanne Gaakeer, 67–87. *Ius Gentium: Comparative Perspectives on Law and Justice* 25. Springer Netherlands. http://link.springer.com/chapter/10.1007/978-94-007-6314-2_4.

Footnotes

1. Art. 29 Working Party WP251rev.01, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679, at 19. [↵](#)
2. *Ibid*, at 21. [↵](#)
3. *Ibid*, at 21. [↵](#)
4. *Ibid*, at 21-22. [↵](#)