Artificial Intelligence and Law
https://doi.org/10.1007/s10506-021-09302-7

ORIGINAL RESEARCH

Algorithmic disclosure rules

Fabiana Di Porto[1,2]

Accepted: 26 September 2021

Abstract

During the past decade, a small but rapidly growing number of Law&Tech scholars have been applying algorithmic methods in their legal research. This Article does it too, for the sake of saving disclosure regulation failure: a normative strategy that has long been considered dead by legal scholars, but conspicuously abused by rule-makers. Existing proposals to revive disclosure duties, however, either focus on the industry policies (e.g. seeking to reduce consumers' costs of reading) or on rule-making (e.g. by simplifying linguistic intricacies). But failure may well depend on both. Therefore, this Article develops a `comprehensive approach', suggesting to use computational tools to cope with linguistic and behavioral failures at both the enactment and implementation phases of disclosure duties, thus filling a void in the Law & Tech scholarship. Specifically, it outlines how algorithmic tools can be used in a holistic manner to address the many failures of disclosures from the rulemaking in parliament to consumer screens. It suggests a multi-layered design where lawmakers deploy three tools in order to produce optimal disclosure rules: machine learning, natural language processing, and behavioral experimentation through regulatory sandboxes. To clarify how and why these tasks should be performed, disclosures

Professor of Law and Tech at the University of Salento, Italy and Contract Professor of Innovation Law and Regulation at LUISS, Rome. Head of the Algorithmic Disclosure Regulation PRIN-funded Research Project 2017-22. As a Lady Davis Fellow for the 2019/20 academic year, I have been Forchheimer Visiting Professor at the Law Faculty and Research Associate at the Federmann Cyber Security Research Center of the Hebrew University of Jerusalem, Israel. This article has been written during the research period spent at and benefitted from the generous contribution of the Hebrew University, Israel. I am deeply grateful for insightful discussions to a number of people: Omri Abend, Ittai Bar Siman Tov, Netta Barak-Corren, Tamar Berenblum, Antonio Davola, Lex de Lange, Orr Dunkelman, Catalina Goanta, David Hay, Renana Keydar, Marco Lippi, Michael Livermore, Boaz Matan, Astorre Modena, Giorgio Monti, Monica Palmirani, Marilena Rizzo, Yuval Shany, Limor Shmerling Magazanik, Michal Shur-Ofri, Daniel D Sokol, Keren Weinshall, and Eyal Zamir; the participants to the Workshops 'Law, AI and Data Science: Challenges and Opportunities' held at Bar Ilan University, 18–19 December, 2019; the 'HUJI Federmann Cyber Security Summit Meeting' (13 May 2020) 'Private and Commercial Law Meeting' (22 June 2020), both held at the Hebrew University; the Conference 'Should Data Shape Private Law? Between Stereotypes and Personalization', organized jointly by the Universities of Tilburg, Maastricht and Osnabrück, 4–5 June 2020. I am thankful to Tatjana Grote for her wonderful research assistance. All mistakes remain my own.

Extended author information available on the last page of the article

Published online: 08 November 2021

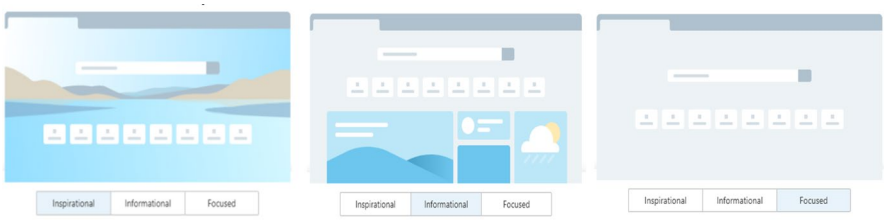
in the contexts of online contract terms and privacy online are taken as examples. Because algorithmic rulemaking is frequently met with well-justified skepticism, problems of its compatibility with legitimacy, efficacy and proportionality are also discussed.

Keywords Disclosure regulation · Failure · Consumers · Law and technology · Information duties · Machine learning · Algorithms · Natural language processing · Regulatory sandboxes · Knowledge graph · Due process

1 Introduction

Ms Schwarz had just finished editing her article, and was about to email it to a journal, when her laptop stuck: she was asked to choose among three different layouts of her browser's tab, so that next time she would open the app, the tab would look exactly like her preferred option, between an 'Inspirational', 'Informational' and 'Focused' appearance.

Here are the layouts:



The browser tab is an example of algorithmically 'targeted disclosure', through which a company[1] uses Information Technology[2] to convey information to consumers in a way that suits their preferences. It also provides a nice way to visualize the goal of this article: How can NLP and ML algorithms[3] be used to target disclosure rules at clusters (not individuals) of consumers to reduce the rules' failures? How could that be done in ways to ensure that disclosure rules be implemented automatically by the industry, thus significantly decreasing the cost of complying? And how could disclosure be differentiated to target the different informational preferences of consumers? Could that be done by rulemakers? Taking the economic ground for producing disclosure rules as a given, how should rulemakers proceed?

[1] The tab layouts are by Edge, the new Microsoft's web browser (2020).

[2] Here human–computer interaction.

[3] While the history of automation in legal science can be traced back to the early 1950s, the rise of highly performative NLP tools and ML algorithms has substantially widened the spectrum of possible applications: See Fagan (2016) and Medvedeva et al. (2019) for review.

We take online contract terms and online privacy disclosures as examples to illustrate our proposal. In fact, both are massively produced by providers of websites, are usually not subject to face-to-face negotiations with consumers but rather accepted on a take-it-or-leave-it basis. Far from reducing information asymmetry and increasing the bargaining power of consumers, the information conveyed through such duties not only increases obfuscation (Bar-Gill 2014), but is often 'weaponized' by the industry (Luguri and Strahilevitz 2021; Stigler Center 2019) to steer individuals towards behaviors that maximize the industry's profits (Thaler 2018). For instance, through lengthy and obscure contract terms, companies may increase the acceptance rate of online offers, or induce individuals to buy tied insurance services. Similarly, by framing the choice of lenient cookie policies in a more prominent way, users are induced to provide more personal data than they would according to their rational preferences.

That disclosure regulation online is prone to failure is nothing new (Ben-Shahar and Schneider 2014; Bakos et al. 2014; Marotta-Wurgler 2015). And its detractors are as numerous as are attempts to revitalize it. Among the latter, the most promising are the Law&Tech scholars (Ashley and Kevin 2017; Livermore and Rockmore 2019), some of whom suggest using NLP and ML tools to personalize, simplify and summarize disclosures. Some authors (Ayres and Schwartz 2014) propose to automatically detect unexpected or unfavorable terms in privacy disclosure policies, and presenting them in a separate warning box. Others suggest to frame terms of contracts differently to increase readability of online privacy policies, based on behavioral evidence (Plaut and Bartlett 2012); still others propose to base the decision of which parts consumers should pay special attention to on the requirements imposed by privacy law and consequently focus on choice provisions (Mysore Sathyendra et al. 2017). Moreover, others recommend to employ bots to highlight the content of platforms' privacy disclaimers or help educate consumers (Harkous et al. 2018). Finally, an important literature is emerging that identifies "probabilistic disclosures" as superior to discrete yes/no type disclosures (Levmore 2021).[4]

Although the proposed solutions are certainly promising, and come with the great benefit of only marginally intervening with the consumer's autonomy, they present some critical shortcomings.

First, they suggest intervening mostly on the implementation phase, namely to adjust fallacies at the firm-level disclosure policies, assuming that the sole reason disclosure regulations fail is the prohibitive cost of reading (Bartlett et al. 2019). For instance, a great work has been done using algorithms to show that online privacy policies are often incomplete (Contissa et al. 2018a, b; Lepina et al. 2019) or some of their clauses are linguistically imprecise (Liu et al. 2016). With regard to online contracts, an algorithm has been developed to automatically detect clauses that are potentially unfair under EU law (Lippi et al. 2018).

[4] Demonstrating that for some addressees, information provided in probabilistic terms may be more informative than generic one, and suggesting that law not only allows disclosure to be provided in diverse formats, but also safe harbors for probabilistic disclosures.

However, the source of failure may well depend *also* on how disclosure rules are formulated in the first place. For example, goals may be self-defeating: the GDPR, Article 12 requires data controllers to provide individuals with the information on their rights (stipulated in Articles 13 and 14) 'in a concise, transparent, intelligible and easily accessible form, using clear and plain language'. Being concise and intelligible at the same time can be two conflicting goals.

Second, the solutions proposed make assumptions about what consumers need to be warned of that are based on singular surveys. In this sense, they are static, instead of being continuously revived and dynamically updated with the support of live data. However, this would be necessary in order to react to shifts in both consumer preferences and business behavior. Designing legal solutions on evidence that is gathered through ad hoc experiments may be limiting, as new evidence may come about showing opposing results. For instance, in the US, the recently proposed 'Algorithmic Justice and Online Platform Transparency Act'[5] establishes new disclosure duties on online platforms[6] with regard to the algorithmic processes they utilize to 'promote content to a user' (i.e. personalized product/service). Although these data are made available to the regulator,[7] the FTC would only access past (not real time) information. In other words, this solution may not allow capturing in due time if consumers become reactive to some piece of information and not to other, or if companies adapt their disclosures to behavioral changes of consumers (as is the case with dark patterns)[8] or to counter changes in the law. Therefore, repeated experiments using real time data would be preferable in order to sustain regulatory intervention. Apart from that, no evidence is available of the number of consumers actually using, or the efficacy of the bots like CLAUDETTE (Lippi et al., at 136–137) and darkpatterns.com (Calo 2014; Stigler Center, at 28).

Before this background, this Article innovates in a relevant regard: it argues that algorithmic tools can and should be used that 'comprehensively' consider solutions at the drafting stage jointly with the implementation phase of disclosure regulation. This 'comprehensive approach' conceptualizes disclosure regulation as a process composed of rulemaking and implementation, and therefore suggests using algorithms to tackle the fallacies affecting each step singularly and all of them together (part one). As a unique contribution, this article elucidates on how to implement this 'comprehensive approach' in practice. From a technical point of view, lawmakers should deploy three tools in order to produce optimal disclosure rules: machine learning (ML), natural language processing (NLP), and behavioral experimentation through regulatory sandboxes (part two).

[5] H.R. 3611, 'Algorithmic Justice and Online Platform Transparency Act', 117th Cong. (2020–2021) of 28 May 2021.

[6] Section 4(1)(a) H.R. 3611, 117th Cong. requires that, for each process, users are informed of: (i) their personal data; (ii) in what way they are collected or created by the platform; (iii) how the latter uses them; and (iv) what methods are used to prioritize, assign weight, rank personal data to deny, amplify, recommend, or promote content to a user.

[7] Section 4(a)(2)(C) H.R. 3611, 117th Cong. (above, note 5).

[8] Section 2 (Findings) H.R. 3611, 117th Cong.

Lawmakers should begin (Phase One) by creating a dataset composed of disclosure rules, firm-level disclosure policies, and the case law pertaining to both (Sect. 3.1.1). Next, they should deploy NLP and other techniques to map out the causes of failure, rank the disclosures accordingly and find the best matches of law and implementation on the basis of such ranking (Sect. 3.1.2). The goal would be to develop an ontology of self-implementable rules that produces good outcomes in terms of readability, informativeness, and coherence, which could be dynamically updated. This ontology is called HOD or Hypothetically Optimal Disclosure (Sect. 3.1.3), which would only include disclosure rules that fail the least (according to our library of measurable failure indexes), and therefore give rise to the least disputed issues (out of the relevant case-law). HOD raise nonetheless questions of efficacy, legitimacy and proportionality that need to be addressed (Sect. 3.1.4).

For Phase Two suggests exploring the potential of regulatory experimentation in sandbox, as a viable solution. We propose testing HOD with regulatory experimental methods, as a unique solution. Regulatory sandboxes are thus presented as a means to pre-test of different layouts of HOD disclosures with stakeholders in a collaborative (co-regulatory) fashion; to ensure transparency and participation; target disclosures to increase efficacy; and cluster individuals (Sect. 3.2.1).

Section 3.2.2 explains how the sandbox is organized, both from a governance perspective and a technical one. The final outcome to sort with, once behavioral data from the sandbox are integrated, is the Best Ever Disclosures, or BED: an algorithm producing legal notices that would be targeted at clusters of consumers; updated continuously with rules, caselaw and behavioral data, and that would also be automatically implementable.

Section 3.2.3 explains how automatic implementation of BED on large scale works, both at the very first launch on the market, and successively, when amendments are needed.

Lastly, a discussion of possible drawbacks and wider effects of BED algorithmic disclosures on stakeholders is presented (Sect. 3.2.4) before concluding.

2 Part one: The case for a 'comprehensive approach'

2.1 Disclosure regulation in online markets: a failing strategy in need of a cure

Traditionally, Disclosure Regulation serves function of reducing information asymmetries that plague consumers (Akerlof 1970) and their unequal bargaining power (Coffee 1984; Grossman and Stiglitz 1980). In online consumer transactions we are

literally flooded with transparency and disclosure duties and policies. For instance, the EU Consumer Rights Directive (CRD)[9] contains several rules mandating the provision of information to the point of limiting the freedom to design e-commerce websites.[10] The CRD also relies on pre-contractual information requirements to protect consumers: online marketplaces must inform consumers about the characteristics of a third party offering goods, services, or digital content in the online marketplace[11]; state if the provider is a trader or not (in which case, other and less protective laws would apply) (Di Porto and Zuppetta 2020)[12]; or break down key information which involve costs.[13] Similarly, in the online privacy field, disclosure duties have flourished: 'cookie banners' (or more precisely 'consent management platforms', CMP) appear at any first website access requiring user consent for personal data processing, based on legal requirements in both the EU[14] and the US.[15]

However, the appropriateness of such duties to provide effective protection to consumers is knowingly poor. Online contract terms and online privacy policies are unilaterally designed by the platform, and are mass-marketed online, essentially on a take-it-or-leave-it basis (Bar-Gill 2014). In this scenario, the platform can determine the 'choice architecture' in which consumers act, thus deliberately exploiting irrational consumer behavior to increase its profits (sludging) (Thaler 2018). For instance, one may accept to provide more personal data than she would deem reasonable according to her preferences, or accept to buy more quantities of a given service.

Personalization has boosted the manipulative power of platforms, (Zuboff 2019) and digital firms have become skilled at developing 'dark patterns'. (Brignull 2013) through which the most vulnerable consumers are especially targeted (Stigler Center 2019).

Disclosure duties can do little to intercept or counter these practices or educate consumers. This because of the disproportionate informational disadvantage of which regulators suffer vis-à-vis the industry. Regulators do not possess granular and real-time data about users' behavior, nor can they observe changes in privacy policies made by the industry as a response. They can certainly run experiments and collect data, but do not have enough resources to do so on a regular basis, as can

[9] Directive 2011/83/EU, amended by Directive (EU) 2019/2161 of 27 November 2019 amending Council Directive 93/13/EEC and Directives 98/6/EC, 2005/29/EC and 2011/83/EU as regards the better enforcement and modernisation of Union consumer protection rules, *OJ L* 328, 18 December 2019, 7–28.

[10] CRD Art. 8(2) sets out clear requirements for the design of buttons in online consumer transactions: they may only state 'order with obligation to pay' or similarly unambiguous formulations.

[11] CRD Art. 6a(1)(a).

[12] If the user is not a consumer, then the relationship is one of one of Business-to-Business and hence covered by the (less protective) Regulation (EU) 2019/1150 of 20 June 2019 on promoting fairness and transparency for business users of online intermediation services, *OJ L* 186, 11 July 2019, 57–79.

[13] CRD Art. 6(1)(e).

[14] See the General Data Protection Regulation (GDPR) art. 7 and Alinea 32 (requiring 'informed' consent to data treatment).

[15] The California Consumer Privacy Act (CCPA) (requiring businesses to give consumers information about the data they collect and the way they use it, at the time or before they start collecting it, lit. in a 'notice at collection').

digital firms (e.g. by running A/B testing). In addition, educational campaigns for consumers do not seem a viable solution, not only because they suffer from a collective action problem (Bar Gill 2014), but also because they are costly for the industry.

Despite all this evidence, rule-makers continue to employ disclosure regulation massively in both online contract terms and the privacy contexts. To quote a few: traders of online marketplaces like Amazon shall inform consumers if their prices are personalized;[16] to ensure that reviews originate from real customers or are not manipulated, platforms must provide with 'clear and comprehensible information' about the 'main parameters determining the ranking' in research queries[17]; online general search engines (like Google, Edge and the like) must provide a 'description of the main ranking parameters and of the possibilities to influence such rankings against remuneration'.[18]

In the US, as seen, the information a consumer needs to be provided with must be given in a 'in conspicuous, accessible, and plain language that is not misleading'.[19]

All these new requirements do not innovate in terms of disclosure strategy, which remains based on long duties to provide information to some impersonal non-differentiated addressee (e.g. the average consumer). Rather, they rest on traditional and disproved assumptions: that individuals will read the disclosures by just using plain and intelligible language, or by putting the information in a given place of the platform's websites. For general terms and conditions this is requested by Articles 3 and 5, EU Regulation 2019/1150, and for rankings parameters by Article 6, CRD.

Based on previous massive evidence, however, we should expect that this avalanche of new information duties would not escape failure. For this paper argues that computer science solutions should be used 'comprehensively', that is: to tackle failures at the rule-making phase jointly with the implementation phase. Before illustrating how to implement this 'comprehensive approach' in practice (Phase Two), in the following we elucidate on our methodological approach.

2.2 Tackling failures at rulemaking and implementation stages

The idea behind the 'comprehensive approach' is to address failures at all levels through a two-step methodology. First we use text analysis to tackle failures of rules and policies; then we employ behavioral testing.

There is a reason if we separate this in two phases. Given the current state of art, algorithmic tools exist that allow intervening on texts and helping measuring failures pertaining specifically to the drafting of disclosures rules as well as firm-level

[16] CRD Alinea 45, and Art. 6(1)(ea) Lit. 'based on automated decision-making and profiling of consumer behavior.'.

[17] CRD Art. 6a(1)(a). The rationale is of course to ensure that reviews, on which rankings are based, originate from real customers, real purchasing experiences, no sponsorship nor contractual ties are supporting the reviews, and that no technical manipulation of the results occurred (Alinea 47). Such information should be 'made available in a specific section of the online interface that is directly and easily accessible from the page where the offers are presented'. (Ibid). The omission of such information may amount to an Unfair Commercial Practice (UCP) under Art. 7(4a) UCP Directive No. 2005/29/CE (Annex I, Nos. 23b, 23c).

[18] Art. 5, CRD.

[19] Section 4(a)(1)(A), H.R. 3611, 117th Cong.

disclosure policies in terms of readability, informativeness, and coherence (Phase One).

On the other hand, text analysis is not (yet) a good tool to identify and measure consumer behavior, nor the industry reaction to it. However, behavioral data is needed to assess how consumers and firms interact with the respective documents, and assess disclosures effectiveness with a view to overcome these types of failures. Hence, we will address the issue of how behavioral data can be generated and used in an inclusive and efficient way in a different part of the Article, by taking inspiration from the 'regulatory sandbox' model (Phase Two).

For completeness, one should mention a strain of literature that seems to consider both text and behavioral elements of disclosure, by addressing changes in the privacy disclosure based on rules from the GDPR as well as consumer expectations (behavioral element). More specifically, Schwartzhneider et al. (2018) claim that big disorders (like the Cambridge Analytica scandal) depend on the 'mis-alignment' between privacy notice and consumers' expectations (a behavioral element) regarding those notices and contend that such failure could be avoidable if both (i) a 'coherent flow' of information was identifiable between rules (principles level) and disclosure policies, and if (ii) the (average) consumer was not 'overwhelmed by the legal[istic] language.' Another noteworthy example is the work of Gluck et al. (2016) who link textual failures of overly lengthy privacy policies with behavioral elements (like the negative framing of disclosures offered to consumers).

In both cases, however, what lacks is a full picture, capable of capturing and measuring all failures of disclosures (not just length or legalistic language) at all different stages: the drafting of disclosure rules, their firm-level implementation, and behavioral failures when consumers are exposed (as well as their interactions).

3 Part two: Implementing the 'comprehensive approach'

3.1 Phase one: Getting to hypothetically optimal disclosures (HOD)

3.1.1 Mapping texts

Lawmakers should begin by creating three datasets composed of disclosure rules (Sect. 1), firm-level policies (Sect. 2), and the case-law pertaining to both (Sect. 3). Again, the domains of online privacy and online contract terms are taken as examples.

1. Disclosure rules as dataset: the De Iure disclosures

For the sake of simplicity, we term *De Iure disclosures* all rules where disclosure duties are set. In the privacy context, requirements to platforms to disclose information to individuals regarding their rights, how their data are collected and treated would fit this category. Just as examples we may quote: Sec. 1798.100(a) CCPA, which stipulates the duty of 'a business that collects a consumer's personal

information [to] disclose to that consumer the categories and specific pieces of personal information the business has collected'. GDPR Art. 12 requires data-controllers to provide similar information to data subjects.

Technically speaking, these rules can be understood as datasets (Livermore and Rockmore), that can be retrieved and analyzed through NLP techniques (Boella et al. 2013, 2015), easily searched (e.g. via the Eur-lex repository), modelled (e.g. using LegalRuleML) (Governatori et al. 2016; Palmirani and Governatori 2018), classified and annotated (e.g. through the ELI annotation tool)[20] For instance, the PrOnto ontology has been developed specifically to retrieve normative content from the GDPR (Palmirani et al. 2018).

While rules may be clear in stating the goals of required disclosure, it may well be that convoluted sentences or implied meaning appear that make the stated goal far from clear. Also, the same rule may sometime prescribe a conduct with a nice level of detail (if X, than Y), but it may include provisions that require, for instance, that information about privacy shall be given by platforms in 'conspicuous, accessible, and plain language'.[21] Even if governmental regulation is adopted specifying what these terms mean, they would not escape interpretation (Waddington 2020), and thus possible conflicting views by the courts.[22]

To help attenuate these problems, proposals have been made to use NLP tools to extract legal concepts and linking them to one another, e.g. through the combination of legislation database and legal ontology (or knowledge graph). Boella et al. (2015) suggest using the unsupervised TULE parser and a supervised SVM to automate the collection, classification of rules and extraction of legal concepts (in accordance with Eurovoc Thesaurus). This way, the meaning of legal texts will be easier to understand, making complex regulations and the relationships between rules simpler to catch, even if they change overtime. Similarly, LegalRuleML may be used to specify in different ways how legal documents evolve, and to keep track of these evolutions and connect them to each other.

2. Firm-level disclosure policies as dataset: the De Facto disclosures

The second dataset is that of firm-level disclosure policies, that we term *De Facto disclosures*. The latter include but are not identical to the notices elaborated by the industry to implement the law or regulations. We refer to the overly-famous online Terms of Services (commonly found online and seldom read). With regard to privacy policies, pioneering work in assembling and annotating them was undertaken by Wilson et al. (2016), resulting in the frequently used 'OPP-155' corpus. Indeed, ML is now standard method to annotate and analyze industry privacy policies (Sarne et al. 2019; Harkous et al. 2018).

[20] The European Legislation Identifier (ELI) is available online at: https://eur-lex.europa.eu/eli-register/resources.html.

[21] Section 4(1)(a) H.R. 3611, 117th Cong. (above, note 5).

[22] See *below* §3.

3. The linking role of case-law

The case-law would play an important role, serving as the missing link between legal provisions and their implementation. Indeed, courts' decisions help detect controversial text and provide clarification on the exact meaning to give both De Iure and De Facto disclosures. It follows that case outcomes and rule interpretation should be used to update the libraries with terms that can come out as disputed, and others that can become settled and undisputed.[23]

A good way to link the case law with rules is that proposed by Boella et al (2019) who present a 'database of prescriptions (duties and prohibitions), annotated with explanations in natural language, indexed according to the roles involved in the norm, and connected with relevant parts of legislation and case law'.

In the EU legal system, a question might arise if only interpretative decisions by the European Courts or also those of national jurisdictions should be included in the text analysis, given that the first would provide uniform elucidation that binds all national courts (having force of precedent), but most case-law on disclosures originates from national controversies and does not reach the EU courts. We know, for instance, that the EU jurisprudence saves to global platforms only a minor part of the costs they spend in controversies with consumers; the paramount ones are those platforms bear for litigations hold before national jurisdictions,[24] where there is no binding precedent, and the same clause can be qualified differently.

Moreover, differently from the US,[25] in Europe, only the decisions by the EU Courts are fully machine-readable and coded (Panagis et al. 2017),[26] while the process to make national courts' ones also so is still in the making (it is the European Case Law Identifier: ECLI),[27] although at a very advanced stage. Nonetheless, analytical tools are already available that allow to link the EU to national courts' cases. For instance, Agnoloni et al 2017 introduced the BO-ECLI Parser Engine, which is a Java-based system enabling to extract and link case law from different European countries. By offering pluggable, national extension, the system produces standard identifier (ECLI or CELEX) annotations to link case law from different countries.

[23] For instance, the 1985 US Supreme Court *Zauderer v. Office of Disciplinary Counsel* case established a rational basis review standard triggered by a provision requiring "factual and uncontroversial information" in the disclosure regulation. This is one example of how case law can link terms in the *de iure*-disclosures and the corresponding provisions in the de facto-disclosures. (Brannon 2019).

[24] One easy way to measure this is to estimate the costs platforms spend to insure themselves against the risk of lost controversies (and distinguish between EU and national ones).

[25] See the 'Caselaw Access Project', providing (free) access to the published decisions from nearly all US State and Federal Courts: https://case.law/.

[26] Texts of all judgments of the European courts can be downloaded for free from EUR-Lex (https://eur-lex.europa.eu/homepage.html).

[27] European Case Law Identifier (ECLI) is a computer readable and processable code that can be assigned to every judicial decision from every national or European court. Having an ECLI code assigned ensures that the database is indexed by the ECLI Search Engine, which is based on XLM, on an open source basis. ECLI 'facilitates automated linking of judgments to each other, to other legal sources or to academic writings'. See http://www.bo-ecli.eu/ecli/current-implementation.

Furthermore, the EU itself is increasingly conscious of the need to link European and national case law, resulting, for instance, in the EUCases project which developed a unique pan-European law and case law Linking Platform.[28]

As shown by Panagis (cit.), of algorithmic tools, citation network analysis in particular, can be extremely useful in addressing not only the question of which is the valid law but also which preceding cases are relevant as well as how to deal with conflicting interpretations by different courts. The latter is especially relevant in systems where there is no binding precedence (i.e. most national EU legal systems) and where, consequently, differing interpretations of certain ambiguous terms might arise. By combining network analysis and NLP to distinguish between different kind of references, it might be possible to assess which opinions are endorsed by the majority of courts and could thus be considered the 'majority opinion'. While other methods to analyze citations in case law might establish the overall relevance of certain cases in general, only the more granular methodology suggested by Panagis et al. seems well fit to assess which interpretations of certain ambiguous terms are "the truly important reference points in a court's repository". In this way, case law can be used to link the general *de iure* disclosures and the specific de facto disclosures while duly taking into account different interpretations of the former by different courts.

3.1.2 Mapping the causes of failure

To measure the causes of failure of both *de iure* and de facto disclosures is not an easy task (Costante et al. 2012). Nevertheless, quantitative indices are indispensable to conduct the following analysis, to make the information they store easily accessible and readable for machines and algorithms. Also, such indexes guarantee the repeatability and objectivity required for the sake of scientific validity.

In line with our 'comprehensive approach', for each stage, failures must be identified, mapped and linked with the failures at other stages, since these are inherently intimately related.

Therefore, we propose defining a standard made of three top-level categories of failure that can be used for both *de Iure* and de facto disclosures:

(i) Readability. *Length of text* can be excessive leading to *information overload*.
(ii) Informativeness. *Lack of clarity* and *simplicity* can lead to *information overload*. But also the *lack of information* can result in asymmetry.
(iii) Consistency. *Lack of same lexicon* and *cross reference* in the same document or across documents that may lead to incoherence.

Based on these three framework categories, we establish golden standard thresholds and rank clauses as optimal (O) or sub-optimal (S–O) (Contissa et al. 2018a). This way, we would for instance, rank as S–O a privacy policy clause under the 'length of text' index, if it fails to achieve the established threshold under the goal of

[28] EUCases LLOD, available at: http://www.eucases.eu/start.html.

Table 1 Indexes of failure of *de iure* and de facto disclosures—methodology and ranking

	Relevant failure index	Proxy	Methodology and Ranking (O/S–O)
Readability	Information overload	Length of text	No. of polysyllables on the basis of the length of the text *Rank*: e.g. if longer that X words (golden standard), then rank S–O *ALGO*: SMOG; Dale–Chall readability formula; Gunning Fog Index *Major Ref.* Bartlett et al. (2019)
Informativeness	Information overload	Complexity of text	*Syntactic*: No. of certain grammatical structures (nodes) containing complex text (e.g. conjunctive adverbs—however, thus, nevertheless—passives, modal verbs—could, should, might) *Rank*: E.g. if number of nodes containing complex tokens in clause is higher than X per sentence of a Y length (golden standard), than rank S–O *Major Ref.* Botel and Granowsky (1972) or Szmrecsanyi (2004) *Semantic*: use of complex, difficult, technical or unusual terms called 'outliers' (e.g. 'as necessary', 'generally') or of two or more semantically different CI parameters in information flows *Rank*: E.g. if clause contains more outliers than the number set in golden standard, then rank as S–O *ALGO*: LOF, CI in information flows *Major Ref.* Bartlett et al (2019) Shvartzshnaider et al. (2019)
	Information asymmetry	Lack of information	Presence of all information required by the law (e.g. identity of data controller, types of personal data collected; goals of treatment, etc.) *Rank* E.g. if clause omits more elements than all those necessary according to golden standard, then rank as S–O *Major Ref.* Liepina et al (2019) and Costante et al (2012) /or Contissa et al. (2018a, b)
Consistency	Internal and External	Interaction amongst clauses within the same text and across texts	Recurrence of same lexicon and cross reference between different clauses in the same document and across documents *Rank*: E.g. if a clause scores lower than the citation network gold standard for cross-reference links or evaluation of textual similarity, then rank as S–O *Major Ref.* Panagis et al. (2017) [citation net]; or Nanda et al. (2019) [similarity models]

Ranking: O = Optimal; S–O = sub-optimal

Ranking ought to be done per domain/sector

'clarity' as stated in the GDPR Article 12. At the same time, however, Article 12 or some of its provisions—as seen—may score S–O under other failure indexes, such as lack of clarity (vagueness). The case-law might help clarify whether this is the case.

In the following, we elaborate the methodology for designing a detailed system of indexes to capture the main causes of failure. Furthermore, we provide ideas on how to translate each indicator into quantitative, machine-readable indices. Table 1 summarizes our findings.

1. Readability. Information overload: length of text

The first quantitative index is readability. It is mainly understood as non-readership due to information overload, and measured in terms of '*length of text*'.

There is a large variety of readability scores (Shedlosky-Shoemaker et al. 2009), based on the length of text which are frequently highly correlated, thus 'easing future choice making processes and comparisons' between different readability measures (Fabian et al. 2017).

Among the many, we take Bartlett et al. (2019) proposing an updated version of the old (1969) SMOG. Accordingly, annotators establish a threshold of polysyllables (words with more than 3 syllables) a sentence may contain, in order to be tagged as unreadable by the machine,[29] and hence S–O. The authors suggest 'a domain specific validation to verify the validity of the SMOG Grade'.

This is especially relevant to make our proposal workable. Not all domains are the same and an assessment of firm-level privacy policies would clearly require to be made in each sector. For instance, the type of personal data a provider of health-related services collects would be treated differently from those of a manufacturer retailer dealing with non-sensitive data.

Under the Readability-Length of text index, sub-optimal disclosure clauses use more polysyllables than those established in the golden standard, set and measured using the revised version of SMOG proposed by Bartlett et al (2019)

2. Informativeness. Information overload: complexity of text

Lack of readability of disclosures may also depend on the complexity of text. The scholarship has suggested to measure it from both a semantic and syntactic points of view.

(a) Syntactic complexity

While most analyses of readability focus on the number of words in a specified unit (e.g. a sentence, paragraph, etc.) as a proxy for complexity, only few authors focus on analyzing the syntactic complexity of a text separately (Botel and Granowsky 1972). Although some scholars search for certain conditional or relational operators,

[29] A policymaker might decide to attach a legal effect, e.g. by establishing that the consumer would be bound only if the number of polysyllables is lower than a fixed threshold. Ibid. p. 9.

they usually do so with the aim of detecting sentences that are semantically vague or difficult to understand (e.g. see Liepina et al. (2019): see next para.).

Going back to Botel and Granowsky, they propose a count system which designates a certain amount of 'points' to certain grammatical structures, based on their complexity (the more complex, the more points). For instance, conjunctive adverbs ('however', 'thus', 'nevertheless', etc.), dependent clauses, noun modifiers, modal verbs ('should', 'could', etc.) and passives will be assigned one or two points respectively, whereas, for instance, simple subject-verb structures (e.g. 'she speaks') receive no points.[30] The final complexity score of a text is then calculated as the arithmetic average of the complexity counts of all sentences.[31]

An alternative approach is that of Szmrecsanyi (2004), who proposes an 'Index of Syntactic Complexity', which relies on the notion that 'syntactic complexity in language is related to the number, type, and depth of embedding in a text', meaning that the more number of nodes in a sentence (e.g. subject, object, pronouns), the higher the complexity of a text.[32] The proposed index thus combines counts of linguistic tokens like subordinating conjunctions (e.g. 'because', 'since', 'when', etc.), WH-pronouns (e.g. 'who', 'whose', 'which', etc.), verb forms (finite and non-finite) and noun phrases.[33]

Although this might be 'conceptually certainly the most direct and intuitively the most appropriate way to assess syntactic complexity', it is pointed out that this method usually requires manual coding.[34]

Since at least the last two measures seem to be highly correlated,[35] choosing among them might in the end be a question of the computational effort associated with calculating such scores.

Under the Informativeness-Syntactic complexity Index, S–O disclosure clauses (of a given length) use a number of complexity nodes that is higher than the standard, defined and measured using Botel and Granowsky (1972) or Szmrecsanyi (2004).

(b) Semantic complexity

Semantic complexity (or the use of complex, difficult, technical or unusual terms called 'outliers') is analyzed by Bartlett et al. (2019) who use the Local Outlier Factor (LOF) algorithm (based on the density of a term's nearest neighbors) to detect such terms.

Approaching the issue of semantic complexity from a slightly different angle, Liepina et al. (2019) evaluate the complexity of a text based on four criteria: (1)

[30] Ibid., p. 515.

[31] Ibid., p. 515.

[32] Ibid., p. 1034.

[33] These features are used to calculate the final complexity score as follows: $ISC(u) = 2 \times n(u, SUB) + 2 \times n(u, WH) + n(u, VF) + n(u, NP)$, which has been called a rather ad-Hoc-solution by the author himself. Ibid., p. 1035.

[34] Ibid., p. 1031. The article cited was published 16 years ago, therefore, some progress in the automatization of measuring syntactic complexity might have been made in the meanwhile.

[35] Ibid., p. 1037.

indeterminate conditioners (e.g. 'as necessary', 'from time to time', etc.), (2) expression generalizations (e.g. 'generally','normally', 'largely', etc.), (3) modality ('adverbs and non-specific adjectives, which create uncertainty with respect to the possibility of certain actions and events') and (4) non-specific numeric qualifiers (e.g. 'numerous', 'some', etc.). These indicators are then used to tag problematic sentences as 'vague'.

In a similar vein, Shvartzshnaider et al. (2019) base their assessment of complexity/clarity on tags, however, in a different manner. They analyze the phenomenon of 'parameter bloating', which can be explained as follows: building on the idea of 'Contextual Integrity' or CI and information flows,[36] the description of an information flow is deemed (too) complex (or bloated) when it 'contains two or more semantically different CI parameters (senders, recipients, subjects of information, information types, condition of transference or collection) of the same type (e.g., two senders or four attributes) without a clear indication of how these parameter instances are related to each other'[37] This results in a situation where the reader must infer the exact relationship between different actors and types of information, which significantly increases the complexity of the respective disclosure (at 164). Therefore, the number of possible information flows might be used as a quantitative index to measure the semantic complexity of a clause.

Under the Informativeness-Semantic complexity Index, a S–O disclosure clause contains more outliers or semantically different CI parameters in information flows than the number set in the golden standard, defined and measured using Bartlett et al (2019) or Shvartzshnaider et al. (2019) respectively.

3. Informativeness. Information asymmetry: lack of information

Another failure index of information asymmetry is the completeness of the information provided in a disclosure. Comprehensiveness has been investigated mainly at the firm-level disclosure policies, rather than the rulemaking (Costante et al. 2012). It must be noted that the requirement of completeness does not automatically counter readability. While an evaluation of the completeness of a disclosure clause is merely concerned with the question whether all essential information requested by the law is provided, readability problems mostly arise from the way this information is presented to the consumer by the industry. Therefore, a complete disclosure is not per se unreadable (just as an unreadable disclosure is not automatically complete) and the two concepts thus need to be separated.

[36] An information flow denotes the transmission of information from one actor to another. The concept of 'contextual integrity' (CI) is based on the notion that the assessment of an information flow's implications requires information on the full context of the flow, with the latter being operationalized by five CI parameters (senders, recipients, subjects of information, information types, condition of transference or collection).

[37] To illustrate this issue, consider the following fictive clause: 'Advertisers, app developers and specified partners <*three senders*> can request information on the content uploaded by you and your friends <*two subjects*> as well as your interactions with other pages <*two types of data*> '. While this clause might seem straightforward at first, a plethora of different information flows are conceivable based on the large number of different parameter values provided, thus keeping the consumer in the dark concerning the precise flow of her information. Shvartzshnaider et al. (2019), 164.

Several authors suggest tools to measure completeness, especially in the context of privacy disclosure. However, nothing impedes to transfer the approaches presented in this section to disclosures like terms and conditions of online contracts.

For instance, based on the above-outlined theory of CI, Shvartzshnaider et al. define completeness of privacy policies as the specification of all five CI parameters (senders, recipients, subjects of information, information types, condition of transference or collection). Similarly, Liepina et al. (2019) consider a clause complete if it contains information on 23 pre-defined categories (i.e. ' <id> identity of the data controller, <cat> categories of personal data concerned, and <ret> the period for which the personal data will be stored'). If information that is considered 'crucial' is missing, the respective clause is tagged as incomplete. Manually setting the threshold would then help define if a clause scores as optimal or not.

A similar, but slightly refined approach is presented by Costante et al. (2012, at 3): while they also define a number of 'privacy categories' (e.g. advertising, cookies, location, retention, etc.), their proposed completeness score is calculated as the weighted and normalized sum of the categories covered in a paragraph.120F.

For our purposes, a privacy or online contract disclosure clause could be ranked using the methodology suggested by Costante et al (2012) and Liepina et al., or alternatively, by Contissa et al. In both cases, however, corpus tagging would be necessary.

> Under the Informativeness-Lack of information Index, in sub-optimal disclosure clauses (of a given length) the number of omitted elements is higher than the pre-defined minimum necessary standard, defined and measured using Liepina et al (2019) and Costante et al (2012) or Contissa et al. (2018a, b).

4. Consistency of documents

One of the two root causes of failure identified above concerns the misalignment of the regulatory goals behind the duty to disclose certain information (as stated in the *de iure* disclosures) and the actual implementation thereof in the de facto disclosure. The general criterion that can be derived from this is that of *consistency*, which can be translated into two sub-criteria: internal and external. However, since their measure and computability are identical, they will be treated together.

Internal consistency denotes the recurrence of the same lexicon in different clauses of the same document as well as the verification of cross-references between different clauses within the same document. External coherence means the cross-references that refer to clauses contained in different legal documents. External coherence too can be understood both as the recurrence of the same lexicon across referred documents and the verification of the respective cross-references. For instance, one rule in the GDPR might refer to others both explicitly (e.g. Article 12 recalling Article 5) or implicitly (like the Guidelines on Transparency provided for

by the European Data Protection Board)[38]; or a privacy policy might refer to a rule without expressly quoting its article or alinea in the article.

Unfortunately, there is no common, explicit operationalization of internal and external coherence in the literature.

A first attempt to analyze cross-references in legal documents is made by Sannier et al. (2017), who develop a NLP-based algorithmic tool to automatically detect and resolve complex cross-references within legal texts. Testing their tool on Luxembourgian legislation as well as on regional Canadian legislation, they conclude that NLP can be used to accurately detect and verify cross-references (at 236). However, their tool would allow to construct a simple count measure of unresolved cross-referenced, which might serve as a basis for the operationalization of internal and external coherence, both in terms of the lexicon used (see above, 2nd cause of failure: complexity of text), and the correct referencing of different clauses (see above, 3rd cause of failure: lack of information). Nevertheless, this is far from the straightforward, comprehensive solution one might wish for.

A solution could be to rely on more complex NLP tools such as 'citation networks', as proposed by Panagis et al. (2017) or 'text similarity models', as suggested by Nanda et al. (2019).

The citation network analysis tool by Panagis et al. (2017), seems particularly straightforward, since it uses the Tversky index to measure text similarity. Therefore, using a tool such as theirs would automatically cover both the verification of cross-reference links as well as an evaluation of the textual similarity of the cited text.

Another promising option to capture text similarity is the model proposed by Nanda et al. (2019), who use a word and paragraph vector model to help measure the semantic similarity from combined corpuses.[39] After manually mapping the documents (rules provisions and respective policy disclosures), the corpuses are automatically annotated helping to establish the gold standard for coherence. Provisions and terms in the disclosure documents would then be represented as vectors in a common vector space (VSM) and later processed to measure the magnitude of similarity among texts.

This last two models especially come with the advantage of capturing the distance in implementation of rules-based disclosures by the industry policies. They seem therefore very promising in the aim of measuring both the distance in lexicon as well as the presence of cross-reference within the same disclosure rule or policy (and define the gold standard).

Under the Consistency Index, a sub-optimal disclosure clause scores lower than the gold standard for cross-reference links or lexicon similarity, measured using either the citation network tool by Panagis et al. (2017) or the similarity model by Nanda et al. (2019).

[38] See for instance the Edpb's Guidelines on Transparency under the GDPR of 11 April 2018, available at https://edpb.europa.eu/our-work-tools/general-guidance/guidelines-recommendations-best-practices_en.

[39] Although it is meant to measure legal transposition of EU Directives by national legislations (especially those of Italy, Ireland and Luxembourg), the model can be adapted to capture similarities between disclosure duties and their transpositions.

3.1.3 Getting to hypothetically optimal disclosures (HOD) through ontology

Preparing the texts in the *de iure* and de facto data sets means that we process the disclosures in each domain to rank them, thus collecting those that score optimal for each failure index. More specifically, per each clause or text partition of the disclosures in each (*de iure* and de facto) dataset, processing for the five analyzed indexes will provide a score, allowing to identify a set of optimal disclosure texts (see Table 2). So for instance, we should be able to select the optimal disclosure provision in the GDPR as far as its 'readability' index is concerned. The same should be for the clause of a privacy policy implementing that provision in a given sector (like e.g. the short term online home renting): imagine that is 'Clause X' of AirBnB disclosure policy. The two would form the 'optimal pair', under readability, of *de iure* and de facto privacy disclosures in the short-term online home renting sector. The same should be done for all clauses and each failure index.

The kind of coding (whether done manually or automatically) and training to employ, clearly depends on the methodology that will be chosen to perform for each of the failure indexes sketched above. In any event, labelling the disclosures might require some manual work by legal experts in the specific sector considered.

The next step is to link the two selected 'optimal pair' of *de iure* and de facto disclosures in the data sets, to reach a sole dataset of what we should term Hypothetically Optimal Disclosure, or HOD.

While, theoretically, a simple, manually organized, static database could be used to do so, the Law & Tech literature suggests a significantly more effective and flexible solution: the use of an ontology/knowledge graph (Shrader 2020; Sartor et al. 2011; Benjamins 2005)10 (Table 3).

As discussed above (I.A.1), legal ontologies are especially apt in this purpose, because they allow automating the extraction and linking of legal concepts, and to keep them up to date even if they change overtime (Boella et al 2015). Another reason is that some ontologies allow to link legal norms with their implementation practices, a feature that is relevant to us.

A good model for linking texts through ontology is provided for by the Lynx project[40] (Montiel-Ponsoda and Rodríguez-Doncel 2018). Lynx has developed a 'Legal Knowledge Graph Ontology', meaning an algorithmic technology that links and integrates heterogeneous legal data sources such as legislation, case law, standards, industry norms and best practices.[41] Lynx is especially interesting as it accommodates several ontologies able to provide the flexibility required to include additional nodes anytime rules or policies change.

To adapt the Lynx ontology to our needs, manual annotation to establish structural and semantical links of *de iure* and de facto disclosure datasets would

[40] http://lynx-project.eu/.

[41] http://lynx-project.eu/doc/lkg/ The Knowledge Legal Graph ontology reuses sources already available on an open access basis, as well as their metadata (such as the afore-mentioned ELI codes of EU case law) and other ontologies (a full list of which is available here: http://lynx-project.eu/data2/reference-ontologies).

Table 2 – Example of ranking of disclosure pair leading to HOD, based on failure index

Failure criteria → ↓ Disclosure pair			Readab (length)	Informativeness			Consistency	Ranking and HOD
				Syntactic compl	Semantic compl	Lack of info		
Rule	De iure disclosure	De facto disclosure						
CRD Art. 6	Partition X, Art. 6a(1)(ea)	AirBnB Portion Y ToS policy	Optimal (score 1)	Optimal (score 1)	Optimal (score 1)	Optimal (score 1)	Optimal (score 1)	HOD
	(info on personalized prices)	Expedia Port. W	Optimal (score 1)	Optimal (score 1)	Optimal (score 1)	S–O (score 0)	Optimal (score 1)	Not incl. in HOD
		Booking Port. Z	S–O (score 0)	Optimal (score 1)	S–O (score 0)	S–O (score 0)	S–O (score 0)	Not incl
		VRBO Port. XY	S–O (score 0)	S–O (score 0)	S–O (score 0)	S–O (score 0)	S–O (score 0)	Not incl

nonetheless be needed. That should be done taking into consideration the results of the ranking process, upon which optimal disclosure pairs are selected (Table 2, above). Hence, manual annotation in ontology would consist in functionally linking of only the latter texts, based on semantic relations between their contents.

In our model, nodes will be represented by the failure criteria sketched above. These nodes are already weighted as Optimal/Sub-Optimal and thus given a specific relevance, which allows an analytically targeted and granular nuancing of the ontology.

A further step consists in the assessment of the overall 'coherence' of HOD ontology. Coherence in this context is understood as a further failure index, consisting of *Lack of cross-reference between the Optimal principles-level rule* and the *corresponding Optimal implementing level policy* (Table 3).

Table 3 Using ontology to get to the hypothetically optimal disclosures (HOD)

Once De Iure and De Facto datasets prepared: Matching (linking) and Ranking through Knowledge Graph/Ontology, leading to 'HOD'		
Coherence/overall	Cross-validation amongst clauses across datasets	Verification of cross-referencing between principles-level (*de iure*) and application level (de facto) leading to incoherence *Rank* If 'clause-pair' scores lower than the gold standard for cross-reference links, then rank S–O *ALGO*: Lynx Legal Knowledge Graph Ontology + manual annotation *Major Ref.* Alschner and Skougarevskiy (2015)

After manual annotation, to cross-validate amongst clauses across datasets, this process would help to further verify if there is cross-reference between the optimal pairs, or between the principles-level of the de iure disclosure and the application level of the de facto disclosures, given that they might come from policies drafted by different firms.

A solution could be to rely on 'citation networks', as proposed by Alschner and Skougarevskiy (2015). Focusing on the lexical component of coherence, citation network would help to calculate the linguistic 'closeness' between different, cross-referenced documents[42] and to assess their coherence.[43]

This way, we will be able to give evidence to the overall optimal linked disclosures (i.e. showing the highest scores assigned to each and every pair per single sector domain) and hence to validate the overall coherence of HOD per given domain.

In conclusion, out of the linked data ontology HOD, we should be able to select the texts that fail the least, under a comprehensive approach. These are linked texts, made of the optimal rules (disclosure duties), linked to their optimal

[42] In Alschner and Skougarevskiy's work citation network allowed to compute the 'textual distance' between 1623 Bilateral Investment Treaties.

[43] Which the authors define as 'close mutual distances' between two treaties.

implementations (policies), whose terms are clarified through the case law and that score optimal for each and every failure index.

HOD are self-executing algorithmic disclosures, which specifications can be used by the industry to directly implement their content. This however opens a plethora of legal and economic questions regarding their efficacy, legitimacy and proportionality.

3.1.4 Limitations of HOD: legitimacy and efficacy

HOD are selected that are the optimal available algorithmic disclosures, but they are still prone to failure. We do not know how effective they might be in leading to behavioral change; how well they could inform real consumers and have them make a sensible choice (for a skeptical take: Zamir and Teichman 2018), given their diverse preferences (Fung et al. 2007). We do not have evidence if the optimal disclosure text regarding a given clause will perform well or not. For instance, imagine we are ranking disclosures in the short-term online renting sector, and that the HOD regarding information provision on the service ranking indicates that the optimal pair is "CRD Art. 5"—"AirBnB Terms, Clause X": what do we know about its efficacy? The HOD cannot tell.

Moreover, since the comprehensiveness of the proposed approach implies that HOD might complement or even partially substitute tasks that would normally be executed or at least supervised by democratically elected representatives, concerns of legitimacy arise. In the example done, once the optimal pairs identified through the HOD, the idea is that "CRD Art. 5"- "AirBnB Terms Clause X" would be automatically implementable. However, that would be problematic under legitimacy terms.

Lastly, HOD may lack proportionality, since they are addressed to undiversified, homogeneous consumers (the average ones), based on assumption of homogenous reading, understanding, evaluation, and acting capabilities (Di Porto and Maggiolino 2019; Casey and Niblett 2019). However, the same disclosure may well be excessively burdensome for less cultivated consumers, while being effective for well-informed, highly literate ones.

In the following, we explore these three issues separately.

1. Untested efficacy of HOD

Although they are hypothetically optimal inter-linked texts, constantly updated with new rules, industry policies, and case-law, easily accessible and simplified, not so costly to read and understand, the overall efficacy of HOD remains untested.

On this land stand the enthusiasts, like Bartlett et al. who purport that the use of text analysis algorithmic tools, which summarize terms of contracts and display them in graphic charts, 'greatly economize[s] on [consumers'] ability to parse contracts' (Bartlett et al. 2019). However, they do not provide proof that this is really so (if one excludes the empirical evidence supporting their paper). Paradoxically, the same holds for those who oppose the validity of simplification strategies and information behavioral nudging, like Ben-Shahar (2016). They consider that

'simplification techniques…have little or no effect on respondents' comprehension of the disclosure.' But again, this conclusion refers to the 'best-practice they surveyed'.

2. Legitimacy deficit of HOD

HOD suffer from a deficit of legitimacy. Because an algorithm is not democratically elected, nor is it a representative of the people, it cannot *sic et simpliciter* be delegated rulemaking power (Citron 2008, at 1297).

While in a not so far future it may well be that disclosure rules become fully algorithmic (produced through our HOD machine), a completely different question is whether disclosure we have selected as the hypothetically optimal might also become 'self-applicable', or, in other words, whether their adoption can become one step only, without any need for implementation. This is surely one of the objectives of HOD. By selecting the optimal rules together with the optimal implementation and linking them in an ontology, we aim at having self-implementing disclosure duties.

Hence, it is necessary to re-think of implementation as a technical process, strictly linked (not merged) with the disclosure enactment phase. But especially, we need to ensure some degree of transparency of the HOD algorithmic functioning and participation of the parties involved in the production of algorithmic disclosures.

Self-implementation of algorithmic rules is one of the least studied but probably the most relevant issues for the future. A lot has been written on the need to ensure accountability of AI-led decisions and due process of algorithmic rule-making and adjudication (Crawford and Schultz 2014; Citron 2008; Casey and Niblett 2019; Coglianese and Lehr 2016). However, whilesome literature exists on transparency and explicability of automated decision-making and profiling for the sake of compliance with privacy rules (Koene et al. 2019), the question of due process and disclosure algorithmic rule-making has been substantially neglected.

However, a problem might exist that the potential addressees of self-applicable algorithmic disclosure rules do not receive sufficient notice of the intended action. That might reduce their ability to become aware of the reasons for action (Crawford and Schultz at 23), respond and hence support their own rights.[44] Also comments and hearings are generally hardly compatible with an algorithmic production of disclosures; while they would be especially relevant, because they would provide all conflicting interests at issue to come about and leave a record for judicial review. The same goes for expert opinions, which are often essential parts of the hearings: technicians may discuss the code, how it works, what is the best algorithm to design, how to avoid errors, and suggest improvements.

[44] See Citron (*supra* note 152) p. 1284 (noting that the black box nature of algorithms can make their decisions non predictable, or non-fully compatible with the guarantees of due process. In Italy, an algorithm was used by the Ministry of Education to decide upon the allocation of high school chairs among teachers who had won a public selection in 2017. The decision being entirely delegated to an algorithm, it has been challenged before administrative courts and further annulled both on first instance and on appeal on discriminatory grounds (Tar Rome, Decision no. 9230/2018, on appeal Council of State, dec. 13 December 2019, no. 8472).

In the US system, it is believed that hearings would hardly be granted in the wake of automated decisions because they would involve straight access to 'a program's access code' or 'the logic of a computer program's decision', something that would be found far too expensive under the so-called *Mathews* balancing test (Crawford and Schultz at 123, Citron at 1284).

In Europe too, firms would most probably refuse to collaborate in a notice and comment rulemaking, if they were the sole owner of the algorithm used to produce disclosures, since that might imply to disclose their source codes, and codes are qualified as trade secrets (thus, exempt from disclosure).

Moreover, as (pessimistically) noted by Devins et al., the chances for an algorithm to produce rules are nullified, because 'Without human intervention, Big Data cannot update its "frame" to account for novelty, and thus cannot account for the creatively evolving nature of law.' (at 388).

Clearly, all the described obstacles and the few proposals thus far advanced are signs that a way to make due process compatible with an algorithmic production of disclosure rules is urgent and strongly advisable.

3. Lack of proportionality of HOD

Although it is undeniable that general undiversified disclosures may accommodate heterogeneous preferences of consumers (Sibony and Helleringer 2015), in practice, they may put too heavy a burden on the most vulnerable or less cultivated ones, while not generating outweighing benefits for other recipients or the society. In this sense, they may become disproportionate (Di Porto and Maggiolino 2019).

On the other side, also targeting disclosure rules at the individual level (or personalizing) (Casey and Niblett), as suggested by Busch (2019), may be equally disproportionate (Devins et al 2017) as can generate costs for the individuals and the society. For instance, if messages are personalized, the individual would not be able to compare information and therefore make meaningful choices on the market (Di Porto and Maggiolino 2019). That, in turn, would endanger policies aimed at fostering competition among products, which are based on consumers' ability to compare information about their qualities.[45] Also, targeting at the singular level requires necessarily to obtain individual consent to process personal data (for the sake of producing personalized messages) and also show one's 'own' fittest disclosure.[46]

[45] Ibid.

[46] Ibid, p. 23.

3.2 Phase Two: Integrating behavioral data into HOD: getting to the best ever disclosures (BED)

3.2.1 Experimental sandboxes to pre-test HOD

One way to possibly overcome the three claims (ensure transparency, participation, proportionality and efficacy of HOD disclosures) would be to integrate real-time behavioral data into the HOD algorithm and have it produce targeted, yet dynamic (i.e. fed by real-time data) self-implementable disclosures.

To achieve that, we suggest exploiting the potential of 'regulatory sandboxes'. In the following, we articulate how this tool could be used to conduct pretrial tests of HOD algorithmic disclosures. Such experiments serve the triple function of ensuring legitimacy of the algorithmic rulemaking by allowing participation and transparency; producing targeted disclosures to test their efficacy, and granting proportionality by clustering.

1. Regulatory sandboxes

Regulatory sandboxes are not new (Tsang 2019; Mattli 2018; Picht 2018). They exist in the Fintech industry, where new rules are experimented in controlled environments (thanks to simulations run over big data) before being implemented at large scale.[47] For instance, the UK's Financial Conduct Authority adopted a regulatory sandbox approach to allow firms 'to test innovative propositions in the market, with real consumers'.[48] Regulatory sandboxes can be conceptualized as venues for experimenting with co-regulation, in the sense that they foster collaboration between the regulator (which takes the lead) and the stakeholders to experiment with new avenues for rule production (Yang and Li 2018). Given their increasing relevance, they are being disciplined by the forthcoming EU Regulation on Artificial Intelligence (Article 45 ff.).

We argue for a regulatory sandbox model where, under the auspices of the regulator, stakeholders come together to pre-test the HOD algorithm to develop self-implementable targeted disclosure rules for consumers.

2. Pre-testing HOD to meet legitimacy claims

The main takeaway of the above discussion on having an algorithm legitimately producing disclosure rules, is that the human presence is irrepressible. That implies that a straight suppression of any transparency and participation guarantees (for the humans) in algorithmic rulemaking is not admissible.

[47] See the joint report by ESMA, EBA, and EIOPA, JC 2018 74, *FinTech: Regulatory Sandboxes and Innovation Hubs* (2019).

[48] Financial Conduct Authority, Regulatory sandbox, 10 February 2020, https://www.fca.org.uk/firms/innovation/ regulatory-sandbox (last accessed 16 June 2020).

Using regulatory sandboxes might remedy legitimacy concerns, as stakeholders will participate in real, and contribute to the regulatory process. Of this participation (i.e. of reactions, comments, etc.) data are tracked that feed the algorithm. Indeed, in the sandbox, the regulator sets up an agile group (of consumers, digital firms, legal experts, data scientists) for the ex-ante testing of HOD algorithmic disclosures in the course of a co-regulatory process. As real individuals interact with each other in the sandbox and their true responses to legal notices are registered and fed into the algorithm, they may constitute a good substitute for both notice and comment.

3. Pre-testing HOD for targeting and gather evidence of efficacy

Another reason why HOD disclosures need to be tested with real people in the sandbox is to check if they may actually change the behavior of addressees in the real world: e.g. if optimal acceptance of cookies by those adversely affected increases.

However, as said, to overcome what we consider the main limitation of the current scholarship on disclosure, we deem that experiments should not be occasional, but conducted on a 'real-time basis' and repeated. The sandbox mode is a proxy for real-time evidence of the recipients' actual reaction to the disclosures. The latter will be gathered later, when algorithmic disclosures will be implemented on the market (see below 1.1.3). Nonetheless, the sandbox mode would still greatly increase our understanding of what does not work, but most importantly, would provide behavioral data for reuse in the HOD algorithm to target the messages.

Elsewhere we purported that 'targeted disclosure' helps increase its effectiveness, as it allows to tackle the different groups of consumers showing homogeneous understanding capabilities and preferences with different messages. For instance, we might expect that consumers participating in the sandbox testing may react differently to the HOD-produced privacy disclosure and show different click-through attitudes. This might depend on their literacy, time availability, framing, and other bias. Exposing them to differentiated layouts instead of just one might increase their ability to overcome click-though.

But this needs to be tested. And the reactions of consumers traced by the algorithm. An example might clarify: only to the extent that targetization of privacy disclosure layouts also becomes optimal, meaning that it helps most consumers in a cluster overcome click-though, can HOD become really optimal, or Best.

4. Clustering to meet proportionality claims

Thus, targeting disclosures at 'clusters' is preferable. However, clustering is not an easy task, since clusters should be made of individuals showing similar preferences (e.g. all those who prefer detailed, long boilerplate of fine-print terms *vs* those who prefer synthetic warning messages). And humans are nuanced. A criterion should be set to form clusters, that can be either descriptive (what consumers in that group typically want to know) or normative (what they ought to know). Either way it should reflect sufficiently homogeneous cognition capabilities and preferences to reduce information overload and increase disclosure utility (Ben-Shahar and Porat 2021).

If data gathered in the sandbox show that a big group of consumers is especially exposed to the risk of overdue payment, then that could constitute a cluster (and a disclosure rule highlighting the consequences of payment delay, instead of a standardized all-inclusive warning list may be tested).

Only if testing sessions are repeated enough evidence is gathered of individual reaction that allows for clusterization. As known, the more data is gathered on the reaction and interaction of individuals, the easier is for the algorithm to identify clusters, based on its predictive capabilities.

Diversification of rules by clusters allows rulemakers to strike a balance between the use of predictive capabilities of algorithms, while at the same time conceiving of disclosure regulation that is compliant with the proportionality principle. In Ben-Shahar and Porat's words, a mandated disclosure regime that grants different people different warnings to account for the different risks they face gives all people better protection against uninformed and misguided choices than uniform disclosures do. (at 156).

Also, clusterization allows for targeted disclosure to be respectful of privacy and data protection rights, while preserving of innovation and the market dynamics (Di Porto and Maggiolino at 21). But even if personalized rules are permissible under a particular jurisdiction's privacy law, the state may economize by identifying clusters of people who share sufficiently similar characteristics and draft one disclosure rule for them instead of many disclosure rules for each of them.

3.2.2 Getting to best ever disclosures (BED) through regulatory sandboxes

1. Governance Design Issues

It is on the rulemaker to propitiate a regulatory sandbox, pooling together experimental groups, which would include, the final consumers, individuals representing digital firms (inclusive of platforms and SMEs,[49] which of course vary depending on the topic of algorithmic disclosures), and technical experts. Special attention should be paid to equal representation of stakeholders in each sector-specific sandbox.[50] As said, the goal of the group is to train the selected algorithm (the HOD) for designing different layouts of the best disclosures (Di Porto 2018).

Repeated sessions of tests and feedback would lead to elaborate, with the agreement of all participants, the final sets of targeted disclosures. The latter, by then would become, very emphatically, the Best Available Disclosures or BEDs, to be deployed at large scale (see below, Sect. 3.2.3).

Indeed, insofar as algorithmic HOD are fed-in with behavioral data on the reactions of real people (in an anonymized and clustered format), they could become differentiated, targeted, and timely, thus meeting the different informational needs

[49] It is especially important to select these stakeholders in a way that the interests of the business users are well represented before those of the platforms and enough receptive of those of final consumers.

[50] See Expert Group on Regulatory Obstacles to Financial Innovation (ROFIEG), *Thirty Recommendations on Regulation, Innovation and Finance*, 13 December 2019, at 70.

of recipients (Di Porto 2018 at 509; Busch 2019 at 312). So, for instance, to tackle the problem of online click-through contracts (people do not read standard form contracts before agreeing), one could provide different layouts to different groups of consumers, depending on their reading preferences. These layouts would target clusters of similar consumers and would be derived from behavioral data that was generated only in the sandbox as a separate, isolated environment, but not from data of every individual.

To be more concrete, each disclosure in the HOD algorithm would target each group showing similar characteristics. For instance, three groups, depending on their capabilities, may be detected:

i. the modest (to whom a super-simplified format may be preferable),
ii. the sophisticate (to be targeted through extensive disclosures), and
iii. the intermediate (a mix of the previous ones).

Testing may prove successful if exposure to the three layouts results in increased reading, understanding and, especially, meaningful choice (e.g. they start refusing third party tracking cookies).

Every choice the participants make will be tracked during the test, and this data will then feed the algorithm (on the technicalities of such feeding see below), providing it with information on how to produce the best disclosures, meaning those that fail the least to be read, understood, and give due course of action. At each session, new data will be recorded regarding how groups of individuals (firms, the regulator, and consumers) react to the provided information. Also, choices from firms regarding disclosure clauses should be tracked and feed in the algorithm.

So for instance, pre-contractual information regarding the right of withdrawal from distant contracts must be provided to consumers according to new Arts. 6 and 8 of the Consumers Rights Directive.[51] In particular, Art. 8 deals with the information provided on 'mobile devices', stating that notice on the right of withdrawal should be:

'provided by the trader to the consumer in an **appropriate way'**.

In a regulatory sandbox, various messages to provide such information would be tested before consumers and digital firms, meaning that both will respond to the different layouts. All such reactions would be coded, and feedback registered.

Testing is also relevant to implement rapid amendments to the algorithmic disclosures, both the texts and the graphic layouts (i.e. where the information is located in a mobile phone screen, or when is displayed on a mobile device) should reactions of consumers not occur.[52]

[51] See Arts. 6 and 8 of CRD, as amended by the New Deal for Consumers Directive 2019/2161.

[52] As noted by the WP29, algorithms are subject to bias and 'can result in assessments based on imprecise projections'. See WP29 (*supra*, note 138) p. 27. Therefore, it is crucial to 'carry out frequent assessments on the data sets…to check for any bias, and develop ways to address any prejudicial elements, including overreliance on correlations'. Those checks and audits require 'regular reviews of the accuracy

To help further this, the regulator should enjoy real-time monitoring powers. Indeed, the pre-testing phase also allows detecting with some precision what are the informational needs and understanding capabilities of the users. In this sense, algorithmic disclosures would produce useful information, by dynamically adapting their content and format to what the cluster recipients need at the time they need.

2. Technical issues: using knowledge graph/ontology

Diverse computational techniques could be used to develop algorithmic disclosures.

Like with HOD, also to get to the BED we suggest using a knowledge graph: this way, the textual libraries from the HOD can be enriched with behavioral data coming from the sandbox (hence, we start the process with three libraries).[53] In the knowledge graph, both the text and behavioral data will be integrated employing users' experience. To make a parallel, this operation resembles (but differs) the way Google search engine operates (through domains and supra-domains). When Google users are shown a picture and asked to 'confirm' that what they see is X and not Y, by clicking 'I confirm' they reinforce a node of the graph. Similarly, human stakeholders in the sandbox provide behavioral data that confirm the layout and text of a proposed clause, thus reinforcing nodes, and gradually strengthening the links in our BED knowledge graph.

For instance, per each group of consumers (modest, intermediate, sophisticate), the stakeholders will have to confirm the layout of a clause of privacy disclosure. The confirmation data of each group will feed the knowledge graph. If they see different layouts of cookie banners, confirmation will tell which one performed best in increasing the ability to avoid click-though (Table 4).

Behavioral data coming from the regulatory sandbox are also relevant to confirm or contradict the links and reality described by the graph. The human presence, as said, is essential to monitor if errors occur in the building of the knowledge graph: technicians supervising in the sandbox may intervene to eventually deactivate any error that may affect the algorithm (Yang and Li 2018, at 3267). That explains why we need technicians to participate in the sandbox, besides regulators, firms, and consumers.

Technically speaking, for the knowledge graph to be implemented, we need to connect all the data: the linked texts of the HOD and the behavioral ones coming from the sandbox. All of these data and information shall remain in the knowledge graph.

Footnote 52 (continued)
and relevance of automated decision-making, including profiling … not only at the design stage but also continuously' Ibid., p. 28.

[53] One key reason why knowledge graphs seem fit to do so 'is that they can provide a common (even if not neutral) language to express' the information of the different libraries. Also, they provide 'at the same time a tool for conceptual retrieval and a model of content which maintains strict references to the text', thus being very much in line with what is required in the context of this proposal. Benjamins (2005), at 116.

Table 4 Using ontology to test HOD in sandbox and get to the Best Ever Disclosures (BED)

Group A of Consumers vs HOD	→ Layout A	→ Confirmation	→ data	BED
	→ Layout B	→ Confirmation	→ data	
	→ Layout C	→ Confirmation	→ data	

To that end, we should use an ontology. The ontology serves to link all the pieces with concepts of the domain, supra-domain, and vertical domain. For instance, imagine we aim to link the term 'fintech' (domain) to the normative goal (supra-domain) to a sector-specific term, like 'transparency in financial fintech' (vertical domain).

Because most of the time, rules do not speak in such a detail, we need to use a meta-level to provide further instructions (Benjamins 2005, at 39). For instance, very often rules in the financial domain do not require retailers of financial products to disclaim full detailed composition of their products, but would instead require for general transparency. Therefore, we would need to provide a meta-level whereby to instruct the algorithm this way: 'When using the word "rules", link it to the concept "transparency", then link it to "disclaimer".

To sum up, the knowledge graph technology is used to refine the BED by performing the following tasks:

(i) memorizing the linked texts prepared in the HOD,
(ii) annotating them (through an ontology);
(iii) building a grid of theoretical-legal concepts, specific to a subject and goal, and to a sector, like in privacy.

In conceptualizing the sandbox, we should elaborate on the concepts typical of a specific sector (like privacy or online consumer contracts). To do so, we need to create relationships with a natural language sandbox, which serves to allow humans to participate in the sandbox, to either confirm or reject them. On this basis, we will provide them to the final consumers and the firms (i.e. the stakeholders). By saying that they are 'satisfied' (or 'confirm' the clauses/layouts), they will feed into the sandbox.

This should be repeated in several formats and for several times (sessions) until we get to the point where all participants are mostly satisfied and least dissatisfied. We should repeat this with the clauses of each disclosure per each of the 3 or *n* layouts we want to target the cluster consumers. In this way, we get to the BED we can implement at large scale.

3.2.3 Implementing BED at large scale

1. Automatic implementation of BED at large scale

After the sandbox testing, disclosure should be available, that are targeted at different groups and self-implementable at large scale: these are the BED. The expected output is that the BED algorithm can produce different rules with different messages to convey to each group of consumers (a); on the industry side, BED's specifications will be used for implementation (b); thus, firms' trade secrets will be safe (c).

(a) Allocation of consumers in the diverse clusters.

Once the BED algorithm producing automatic disclosure rules is launched on the market (implemented at large scale), users are first allocated a default intermediate group (b). However, they remain free to switch from one group to the other by choosing the preferred disclosure option.

Interactions with the algorithm will produce more data, that will be tracked and help further refining it. Choices made by the consumers between the three (or *n*) rule layouts and the switches among them, after due pseudonymization, may feedback into the BED algorithm and ameliorate it. On the contrary, individual choices made *due* to the BED (hence, their effects on a large scale) would not possibly be registered nor further analyzed due to privacy constraints,[54] unless a law expressly authorizes that.[55]

(b) BED's specifications in lieu of industry-led implementation

BED algorithmic disclosures are automatically implementable. However, for BED disclosure to be launched on the markets, the industry must make an effort to technically implement its specifications, which are made publicly available. Being the latter sector-specific, and thoroughly discussed among stakeholders in the sandbox, a lot of time and costs for producing disclosures will be saved to the industry.

Making specifications open to individuals and firms, is also a means to allow the regulator to monitor the efficacy of algorithmic disclosure. Furthermore, it allows for accountability of the disclosed information and the algorithmic decision.

(c) (continued) Without disclosing any trade secrets

Despite a broad consensus on an increased need for transparency when algorithmic decisions are involved, 'it is far from obvious what form such transparency should take' (Yang and Li at 3266). While the most straightforward response to this

[54] At the EU level, individual consent to data treatment would be required under the GDPR if the rule-maker wanted to test whether clustered disclosures were effective *after implementation* on a large scale (unlike the design phase). Even there, mass data treatment would possibly contrast with the *principle of minimization* of treatment (in this case, by public authorities).

[55] This is the case in the EU thus far: under EU law, consent is not the only legal ground legitimizing automated processing of personal data: Art. 22(2) *lit.* b) GDPR allows EU or Members states to adopt laws authorizing it, under the condition that same laws 'lay down suitable measures to safeguard the [individual]'s rights and freedoms and legitimate interests.' Hence, a statutory law may be adopted authorizing algorithmic production of disclosures.

heightened transparency requirement would probably be the disclosure of the source codes used by firms, this approach is not feasible as the latter are unintelligible to most lay persons and highly secretive.

When adopting a 'regulatory sandbox' solution, however, there would be no need for the platform to disclose any of its own algorithms (which might easily remain secret) to other stakeholders participating in the trials.[56] That is because the kinds of algorithms that are being used to get to the BED are publicly available.[57] The consumers, platforms and SMEs contribute with their behavioral data to feed the BED algorithm: for instance, in case of disclosures of standard form contracts, the experimental sandbox phase would consist of the stakeholders testing different formats of ToCs. Thus, they would be enabled to enhance their disclosures without having to publicize any of their algorithms or similarly sensitive information.

2. Post-implementation modification of the BED

As far as amendments to algorithmic disclosures are concerned, these could be done in the regulatory sandbox, and consequently implemented at large scale in an automated way. This is still another step, different from both the creation of the HOD algorithm, its testing with behavioral data to become BED and the latter implementation at large scale. Suppose we have already a BED algorithm working on the market that produces targeted disclosures for short-term rental service terms. Imagine that a new EU Regulation is adopted (e.g. Art 12(1) of the Digital Services Act)[58] amending the CRD and mandating digital providers to inform users about potential "restrictions"[59] to their services contained in the terms and conditions in an "*easily accessible format*" and written in "*clear and unambiguous*" language. Such new piece of law would require a refinement of BED, that we suggest doing in the sandbox, instead of starting the whole process from scratch.

This way, all modifications to BED algorithmic disclosures, that participants to the sandbox accept—and the regulator certifies—could become directly implementable by the digital firms on large scale, given that they have been 'pre-tested' in the sandbox. That would comply with the best practice identified by the already mentioned Guidelines of the WP29, and would allow such modifications to feedback into the BED algorithm to ameliorate it and, consequently, the disclosures.

Technically speaking, BED algorithmic disclosures update constantly depending on three factors: changes in the law/regulation or the jurisprudence (in which case

[56] Notoriously, algorithms are covered by IPRs (and are usually qualified as trade secrets). Legally speaking, under EU law, firms are not entitled to any general right to be informed about the overall system used to make automatic decisions, nor can they demand the full disclosure of the algorithm: see Recital 27 and Art 5(6) Regulation EU 1150/2019.

[57] See *supra* Sect. I.C (discussing how the HOD is built).

[58] European Commission, Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC (COM(2020)825), 15 December 2020.

[59] Meaning "any restrictions that they impose in relation to the use of their service in respect of information provided by the recipients of the service, in their terms and conditions": DSA, Art. 12(1).

the text libraries and nodes in HOD ontology update); change from the sandbox (i.e. update in the behavioral library, and consequently links to the texts through validation/confirmation in BED); change from real-world behavior after implementation on large scale.

To make a step further, one could also think of using sandboxes to "suggest" and "approve" rule modifications. This path would be another innovative, venue: all proposed modifications could be discussed, tested, and approved directly in the regulatory sandbox. That would clearly substitute the usual democratic process. So for instance, the regulator might propitiate and stakeholders in the sandbox agree to change the wording of a rule: they may agree to modify Art. 6a(1)(a), CRD and make information on the ranking parameters of search queries available.

> 'by means of *n* icons on the *x*-side of the presented offer',

instead of

> 'in a specific section of the online interface that is directly and easily accessible from the page where the offers are presented.'

as it currently is.

Once the change validated in the sandbox, the BED library is modified accordingly and can thus be directly implemented.

From a legal perspective, post-implementation modifications would not only be self-implementable, but could also be given a special effect: for instance, because they have been pre-tested and validated in the sandbox, they could produce a direct effect (or be enforceable) among the parties, or in some instances provide for safe harbors. For example, an amendment to the disclosure of a certain service's Terms of Contract in a given sector, which is agreed upon in the sandbox, and implemented in the algorithmic disclosure, could become immediately effective. Also, some of its clauses might escape liability.

3.2.4 Discussion of BED

1. Choice of algorithm provider

One possible limitation of our BED solution is selecting the algorithm provider. It seems problematic to have private parties providing the algorithm for rule-making purposes, since, as purported by Casey and Niblett, they would inevitably reflect their own interests in the definition of the objectives to pursue (Casey and Niblett, at 357). Also, there may be strong economic incentives for private parties for not disclosing information about how their rulemaking algorithm was created or why some results were generated. For instance, they might want to 'heighten barriers to competition, or favor one side because of repeat-player issues'..[60]

[60] Ibidem.

One possibility to overcome rent-seeking and riming rules by firm stakeholders could be for the state to open the provision of BED to competition, similarly to auctions hold for the provision of public goods (Levmore and Fagan 2021). Alternatively, the state could consider undergoing some type of approval process, similar to safety certification. However, that might prove costly.

2. Liability of digital operators

Why would the digital firms want to participate in the BED instead of producing their own disclosures? In the end, they have greater technical skills, knowledge and data about consumers to stay away of BED.

In addition, anything that happens in the sandbox implies some disclosure of trade strategies to the regulator, competitors, and consumers. Information is an asset, and even in the little margins left by the disclosure duties, firms might not want to share the way to convey it to their clients.

Also, the BED solution only holds for firms that operate through algorithms and big data technologies, while it leaves aside those not working in the digital sphere (think e.g. to SMEs who lack resources to invest in these technologies).

Moreover, there are industries (like the pharmaceutical) where the BED solution would not possibly be applicable, as full, lengthy, and complete disclosures are needed and not suppressible. Therefore, targeted and summarized information could not work.

While the last objection is insurmountable, one way to eventually commit digital operators to take part in the pre-testing and continue their support in the implementation of BED at large scale is that regulators establish a safe harbor.[61] The safe harbor would work for companies that commit in advance to the terms and clauses of the disclosures agreed upon by the participants in the sandbox (and of course in all subsequent periodical updates). Afterward, if a company fails to qualify for the safe harbor (because, for instance, it does not duly implement the technical specifications provided for in the BED), it may incur additional legal liability in case of litigation, provided that the plaintiff can prove that the disclosure fails the BED standard.

Eventually, one might consider the BED as a "minimum requirement for disclosure compliance" (e.g. at the Federal or EU levels), so that Member states would remain free to set stricter requirements and thus technical specifications to add to the BED. That way, national (member) states would be able to also take account of their own jurisprudence more widely and incorporate it into the algorithm to the level deemed appropriate.

On a more general reputational ground, engaging in the BED project might be convenient for the industry as firms might demonstrate to engage in pro-consumer

[61] This approach has recently been incorporated into the Australian 'Treasury Laws Amendment (2018 Measures No. 2) Bill 2019'. The bill (esp. Section 926B) facilitates the exemptions for firms participating in a regulatory sandbox to test financial and credit products from certain regulations for the time of testing and under certain conditions.

actions, while at the same time reducing their costs of compliance to disclosure regulation requirements.

3. Are recipients better off?

One possible drawback of algorithmic BED is that they may end up 'offering finite choices to users effectively forc[ing] them to guess the category under which their information falls.' (Citron 2008, at 1300) Also, it may well be that consumers are irresponsive, for reasons we are not able to assess, to the algorithmic targeted disclosures. For instance, as not all consumers are prone to intensive online marketing campaigns or dark patterns, it may well be that a noticeable portion of consumers is not becoming aware or that different pieces of information are needed for them in their decision-making process.

If we agree that this might be the case, we acknowledge that there is no evidence unless we try to seek some. And the BED project is especially aimed at providing the consumers with different types of information (instead of just one) to minimize their cognitive effort while maximizing her individual autonomy. As said, to prove the effectiveness of the provided information to also commit to a choice that maximizes her utility, consumers' online behavior ought to be tracked.

4. Which rulemaker?

On the rulemaker side, one limitation is about who—meant as which authority—should be given responsibility for designing and monitoring the applicability of BED disclosures. In our model, being disclosures sector- and topic-specific, the regulator participating in the sandbox would be, each time, the one responsible for the issue at stake. So for instance, if disclosures in the realm of distant contracts for energy provision are being discussed, then the energy regulator (together with the data protection agency) should take the lead of the testing. In a similar vein, the recently created Utah Fintech Sandbox will be administered by the Utah Department of Commerce.[62]

However, if that solution might accommodate national disclosures, where domestic rulemakers might be given legal responsibility for leading the project, one might wonder who should take the lead at the (US) Federal or EU levels. For instance, in Europe, one might wonder whether the Commission enjoys enough political support to do so, eventually with the support of the Jrc. That would also mean, because disclosures are usually written in different languages, that the translation language service of the EU should be included in the project.

[62] A similar program in Arizona, on the other hand, will be supervised by the Arizona Attorney General. KAYE AC (2019) Utah's new regulatory sandbox. Consumer Finance Monitor. Available at: https://www.consumerfinancemonitor.com/2019/06/11/utahs-new-regulatory-sandbox/.

4 Conclusion

Modern rulemaking has for centuries been a purely human activity. But algorithms are there to support in ways the legal scholarship has started exploring. This Article has drawn a roadmap to employ NLP and ML tools to help save disclosure regulation failure, its stated goal being to reframe how to create better disclosure rules. To do so, it has addressed three types of challenges: regulatory (why does disclosure regulation fail in the online privacy and consumer transaction contexts?); technical (what algorithms could best tackle both textual and behavioral failures of disclosures at the two, enactment and implementation, phases?); and legal (can algorithms legally produce self-implementing disclosure norms?).

To these, this article has provided solutions elucidating on how to build an algorithm for the linking of existing openly accessible datasets of *de iure* and de facto disclosures and then selecting those that fail the least. Further, it has addressed the question of how to attenuate legitimacy problems stemming from lack of democratic representativeness of the algorithm, by integrating elements of collaborative and procedural democracy (using a regulatory sandbox) into a knowledge graph. That, in turn, with the final goal of creating Algorithmic Disclosures, which are self-implementing rules.

In the future work, we intend to analyze how disclosure duties are created at the EU level, to check where possible source of failure might stand. To do so, we plan to use NLP and ML tools to analyze the feedback documents submitted by the stakeholders to the EU consultation process on new disclosure duties contained in the proposed Digital Services Act and Digital Markets Act 2020. This way we seek to identify possible semantic differences in the use and understanding of words that pertain to disclosure duties. If such differences exist, then they may provide fresh evidence of why disclosures fail.

Acknowledgements Funding was provided by Lady Davis Fellowship Trust, Hebrew University of Jerusalem.

Funding Open access funding provided by Università del Salento within the CRUI-CARE Agreement.

References

Agnoloni T, Bacci L, van Opijnen M (2017) BO-ECLI parser engine: the extensible european solution for the automatic extraction of legal links. In: Wyner A, Casini G (eds) Legal knowledge and information systems, proceedings of the 2nd workshop on automated detection, extraction and analysis of

semantic information in legal texts, June 16, 2017, London, UK, pp 113–118. https://ebooks.iospress.nl/publication/48052
Akerlof GA (1970) The market for "Lemons": quality uncertainty and the market mechanism. Q J Econ 84:488–500
Alschner W, Skougarevskiy D (2015) Consistency and legal innovation in the BIT Universe. Stanford Public Law Working Paper No. 2595288, p 2
Ashley KD, Kevin D (2017) Artificial intelligence and legal analytics: new tools for law practice in the digital age. Cambridge University Press, Cambridge
Ayres I, Schwartz A (2014) The no-reading problem in consumer contract law. Stan L Rev 66(3):545–610
Bakos Y et al (2014) Does anyone read the fine print? Consumer attention to standard-form contracts. Legal Stud 43(1):1–35
Bar-Gill O (2014) Consumer transactions. In: Zamir E, Teichman D (eds) The Oxford handbook of behavioral economics and the law. Oxford University Press, Oxford, pp 465–490
Bartlett R, Nyarko J, Plaut V (2019) Do you ever read the fine print? The potential and limitations of text analysis for consumer contracts. Unpublished https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=CELS2019&paper_id=271
Benjamins R (ed) (2005) Law and the semantic web: legal ontologies, methodologies, legal information retrieval, and applications. Lecture notes in artificial intelligence, 1st edn. Springer, Berlin
Ben-Shahar O, Chilton A (2016) Simplification of privacy disclosures: an experimental test. J Legal Stud 45(S2):S41–S67
Ben-Shahar O, Porat A (2021) Personalized law. Oxford University Press, Oxford
Ben-Shahar O, Schneider CE (2014) More than you wanted to know: the failure of mandated disclosure. Princeton University Press, Princeton
Boella G, Di Caro L, Leone V (2019) Semi-automatic knowledge population in a legal document management system. Art Intel L 27(2):228
Boella G et al (2013) Semantic relation extraction from legislative text using generalized syntactic dependencies and support vector machines. In: Morgenstern L et al (eds) Theory, practice, and applications of rules on the web, pp 218–225
Boella G et al (2015) Linking legal open data: breaking the accessibility and language barrier in European legislation and case law. In: Proceedings of the 15th international conference on artificial intelligence and law. Association for Computing Machinery, pp 171–175
Botel M, Granowsky A (1972) A formula for measuring syntactic complexity: a directional effort. Elementary Engl 49(4):513–516
Brannon VC (2019) Assessing commercial disclosure requirements under the first amendment. CRS Report No. R45700. Congressional Research Service, Washington, D.C. https://fas.org/sgp/crs/misc/R45700.pdf
Brignull H (2013) Dark patterns: inside the interfaces designed to trick you. The Verge
Busch C (2019) Implementing personalized law. Personalized disclosures in consumer law and data privacy law. U Chi LR 86:309–331
Calo R (2014) Digital market manipulation. Geo Wash LR 82(4):995
Casey AJ, Niblett A (2019) Framework for the new personalization of law. Univ Chicago Law Rev 86(2):359
Citron DK (2008) Technological due process. Washington Univ Law Rev 85(6):1249–1313
Coffee JC (1984) Market failure and the economic case for a mandatory disclosure system. Vi L R 70(4):717–753
Coglianese C, Lehr D (2016) Regulating by robot: administrative decision making in the machine-learning era. Geo L J 105(5):1147–1224
Contissa G et al (2018a) CLAUDETTE meets GDPR. Automating the evaluation of privacy policies using artificial intelligence. https://www.beuc.eu/publications/beuc-x-2018-066_claudette_meets_gdpr_report.pdf; http://uterms.software/documentation/
Contissa G, Docter K, Lagioia F, Lippi M, Micklitz H-W, Palka P, Sartor G, Torroni P (2018b) Automated processing of privacy policies under the EU General Data Protection Regulation. In: Palmirani M (ed) Legal knowledge and information systems. JURIX 2018: the thirty-first annual conference, pp 51–60
Costante E, Sun Y, Petkovi´c M, den Hartog J (2012) A machine learning solution to assess privacy policy completeness. In: ACM workshop on privacy in the electronic society, pp 91–96
Crawford K, Schultz J (2014) Big data and due process: toward a framework to redress predictive privacy harms. Boston Coll Law Rev 55(1):93–128

Devins C, Felin T, Kauffman S, Koppl R (2017) The law and big data. Cornell J L Pu Pol 27:357–413
Di Porto F (2018) In praise of an empowerment disclosure regulatory approach to algorithms. IIC Int Rev Intellect Property Compet Law 49(5):507–511
Di Porto F, Maggiolino M (2019) Algorithmic information disclosure by regulators and competition authorities. Glob Jurist. https://doi.org/10.1515/gj-2018-0048
Di Porto F, Zuppetta M (2020) Co-regulating algorithmic disclosure for digital platforms. Pol Soc. https://doi.org/10.1080/14494035.2020.1809052
Fabian B, Ermakova T, Lentz T (2017) Large-scale readability analysis of privacy policies. In: Proceedings of the international conference on web intelligence. Association for Computing Machinery, Leipzig, Germany, p 21
Fagan F (2016) Big data legal scholarship: toward a research program and practitioner's guide. Va J Law Technol 20(1):1–81
Fung A, Graham M, Weil D (2007) Full disclosure: the perils and promise of transparency. Cambridge University Press, Cambridge
Gluck J, Schaub F, Friedman A, Habib H, Sadeh N, Cranor LF, Agarwal Y (2016) How short is too short? Implications of length and framing on the effectiveness of privacy notices. Paper presented at the twelfth Symposium on Usable Privacy and Security (SOUPS 2016)
Governatori G, Hashmi M, Lam H-P, Villata S, Palmirani M (2016) Semantic business process regulatory compliance checking using LegalRuleML. In: Blomqvist E, Ciancarini P, Poggi F, Vitali F (eds) Knowledge engineering and knowledge management. Springer, Berlin, p 749
Grossman SJ, Stiglitz (1980) On the impossibility of informationally efficient markets. Am Econ R 70(3):393–408
Harkous H, Fawaz K, Lebret R, Schaub F, Shin KG, Aberer K (2018) Polisis: Automated analysis and presentation of privacy policies using deep learning In: Proceedings of the 27th USENIX Security Symposium, 15–17 August 2018, Baltimore, USA
Koene A et al (2019) A governance framework for algorithmic accountability and transparency. PE 624.262. European Parliamentary Research Service
Lepina R, Contissa G, Drazewski K, Lagioia F, Lippi M, Micklitz H-W, Pałka P, Sartor G, Torroni P (2019) GDPR privacy policies in CLAUDETTE: challenges of omission, context and Multilingualism. In: Proceedings of the third workshop on automated semantic analysis of information in legal text, ASAIL, pp 1–7
Levmore S (2021) Probabilistic disclosures for corporate and other law. Theor Inquiries Law 22(1):263–284
Levmore S, Fagan F (2021) Competing algorithms for law: sentencing, admissions, and employment. Univ Chicago Law Rev 88:367
Liepina R, Contissa G, Drazewski K, Lagioia F, Lippi M, Micklitz H-W, Pałka P, Sartor G, Torroni P (2019) GDPR privacy policies in CLAUDETTE: challenges of omission, context and Multilingualism. In: Proceedings of the third workshop on automated semantic analysis of information in legal text (ASAIL 2019)
Lippi M et al (2018) CLAUDETTE: an automated detector of potentially unfair clauses in online terms of service. arXiv preprint arXiv:1805.01217
Liu F et al (2016) Modeling language vagueness in privacy policies using deep neural networks. In: Association for the advancement of artificial intelligence fall symposium series
Livermore MA, Rockmore DN (eds) (2019) Law as data. SFI
Luguri J, Strahilevitz L (2021) Shining a light on dark patterns. J Legal Anal 13(1):67
Marotta-Wurgler F (2015) Even more than you wanted to know about the failures of disclosure. Jerusalem Rev Legal Stud 11(1):63–74
Mattli W (ed) (2018) Global algorithmic capital markets: high-frequency trading, dark pools, and regulatory challenges. Oxford University Press, Oxford
Medvedeva M, Vols M, Wieling M (2019) Using machine learning to predict decisions of the European Court of Human Rights. Artif Intell Law 28:237–266. https://doi.org/10.1007/s10506-019-09255-y
Montiel-Ponsoda, Rodríguez-Doncel EV (2018) Lynx: building the legal knowledge graph for smart compliance services in multilingual Europe. In: Rehm G, Rodríguez-Doncel V, Moreno-Schneider J (eds) Proceedings of the 1st workshop on LREC (Language Resources and Technologies for the Legal Knowledge Graph) workshop, 12 May 2018, pp 19–22. https://delicias.dia.fi.upm.es/members/vrodriguez/pdf/2018.legalkg.pdf

Mysore Sathyendra K et al (2017) Identifying the provision of choices in privacy policy text. In: Proceedings of the 2017 conference on empirical methods in natural language processing. Association for Computational Linguistics Copenhagen, Denmark, pp 2774–2779

Nanda R,· Siragusa G,·Di Caro L,· Boella G, Grossio L,·Gerbaudo M, Costamagna F (2019) Unsupervised and supervised text similarity systems for automated identification of national implementing measures of European directives. Artif Intell Law 27:199–225

Palmirani M, Martoni M, Rossi A, Bartolini C, Robaldo L (2018) PrOnto: privacy ontology for legal reasoning. In: EGOVIS2018, 7th International Conference, EGOVIS 2018, Regensburg, Germany, September 3–5, 2018, Proceedings. LNCS, vol 11032. Springer, pp 139–152

Palmirani M, Governatori G (2018) Modelling legal knowledge for GDPR compliance checking. In: Palmirani M (ed) Legal knowledge and information systems, p 101

Panagis Y, Sadl U, Tarissan F (2017) Giving every case its (legal) due. The contribution of citation networks and text similarity techniques to legal studies of European Union law. Paper presented at the 30th international conference on legal knowledge and information systems, JURIX 2017, Luxembourg, December 2017

Picht PG, Loderer GT (2018) Framing algorithms—competition law and (other) regulatory tools. MPI Research Paper No. 18-24.

Plaut VC, Bartlett RP (2012) Blind consent? A social psychological investigation of non-readership of click-through agreements. Law Human Behav 36(4):293–311

Sannier N, Adedjouma M, Sabetzadeh M, Briand L (2017) An automated framework for detection and resolution of cross references in legal texts. Requir Eng 22(2):215–237

Sarne D et al (2019) Unsupervised topic extraction from privacy policies. In: Companion Proceedings of the 2019 World Wide Web Conference on—WWW '19, vol 563. ACM Press, San Francisco, p 564

Sartor G, Casanovas P, Biasiotti M, Fernández-Barrera M (eds) (2011) Approaches to legal ontologies. Theories, domains, methodologies. Springer, Berlin

Shedlosky-Shoemaker R, Sturm AC, Saleem M, Kelly KM (2009) Tools for assessing readability and quality of health-related web sites. J Genet Couns 18(1):49–59

Shrader B (2020) What is the difference between an ontology and a knowledge graph? https://enterprise-knowledge.com/whats-the-difference-between-an-ontology-and-a-knowledge-graph/

Shvartzhneider Y, Apthorpe N, Feamster N, Nissenbaum H (2018) Analyzing privacy policies using contextual integrity annotations. arXiv preprint arXiv:1809.02236

Shvartzshnaider Y, Pavlinovic Z, Balashankar A, Wies T, Subramanian L, Nissenbaum H, Mittal P (2019) VACCINE: using contextual integrity for data leakage detection. In: The World Wide Web Conference on—WWW '19. ACM Press, San Francisco

Sibony A-L, Helleringer G (2015) EU consumer protection and behavioural sciences: revolution or reform? In: Alemanno, Sibony (eds) Nudge and the law. A European perspective. Hart Publ., pp 209–233

Stigler Center at Chicago Booth (2019) Report by the committee for the study of digital platforms—privacy and data protection subcommittee

Szmrecsanyi B (2004) On operationalizing syntactic complexity. In: JADT 2004: 7es Journées internationales d'Analyse statistique des Données Textuelles, pp 1031–1038

Thaler R (2018) Nudge, not sludge. Science 361(6401):1

Tsang C-Y (2019) From industry sandbox to supervisory control box: rethinking the role of regulators in the era of Fintech. In: Proceedings of the Comparative Corporate Governance Conference, Singapore, January 24, 2019, p 359

Waddington M (2020) Research note. Rules as code. Law Context 37(1):1–8. https://doi.org/10.26826/law-in-context.v37i1.134

Wilson S, Schaub F, Dara AA, Liu F, Cherivirala S, Giovanni Leon P, Schaarup Andersen M, Zimmeck S, Sathyendra KM, Russell NC, Norton T, Hovy E, Reidenberg J, Sadeh N (216) The creation and analysis of a website privacy policy corpus. In: Proceedings of the 54th annual meeting of the Association for Computational Linguistics (vol 1: long papers). Association for Computational Linguistics, Berlin, Germany

Yang D, Li M (2018) Evolutionary approaches and the construction of technology-driven regulations. Emerg Markets Finance Trade 54(14):3266

Zamir E, Teichman D (2018) Behavioral law and economics. Oxford University Press, Oxford

Zuboff S (2019) The age of surveillance capitalism: the fight for the future at the new frontier of power. Public Affairs

Authors and Affiliations

Fabiana Di Porto[1,2]

✉ Fabiana Di Porto
fabiana.diporto@unisalento.it; fdiporto@gmail.com

[1] Department of Economic Sciences, University of Salento, Lecce & Law Faculty, LUISS, Rome, Italy

[2] Former Forchheimer Visiting Professor, Law Faculty, Hebrew University, Jerusalem, Israel