Charting Algorithmic Accountability: A Comparison of the Notions of Explanation and Time in Legal Rights and Pre-Emptive Machine Learning Algorithms[1]

Abstract

The paper compares the notions of time and explanation in the framework of legal rights with that in pre-emptive/predictive machine learning. The aim of such comparison is to understand how accountability is conceived in the framework of legal rights, and how it is different from what accountability may mean in the context of ML algorithms. This comparative discussion becomes especially relevant with the possibility of a 'right to explanation' under the European Union's General Data Protection Regulation. In this context, the paper outlines the differences in understandings of time, uncertainty and explanability in each legal rights and predictive ML frameworks and the corresponding implications for power relationships between those impacted by such algorithmic decisionmaking, those who design or operate them, and the algorithms themselves.

I. Introduction: ML Causation in Context of Right to Explanation?

One relevant aspect to machine learning algorithms of the EU's General Regulation on Data Protection, coming into force in May 2018, is the Right to Explanation.[2] According to Article 22(1) of the GDPR, the data subject has a "*right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.*" Even in cases where such decision is made with the explicit consent of the data subject, the data controller is obliged to safeguard the subject's rights, freedoms and legitimate interests – "a*t least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision*."[3] Such a provision can only be executed when the data subject has some grounds- viz. an explanation- of the machine learning decision making process in order to challenge it. This is further supported by Recital 71 of the Regulation, according to which data processing and profiling should be subject to safeguards including "*specific information to the data subject and the right to obtain human intervention, to express his or her point of view, to obtain an explanation of the decision reached after assessment and to challenge the decision.*" All this leads to the obvious question of what constitutes 'explanation' in law, specifically 'a right' to explanation, and how do these conceptions of explanation interact with machine learning?

---

1 Submitted to ICML Causal Machine Learning Workshop, Stockholm (15 July 2018) as early ideation paper for a work in progress. Author: Smarika Kumar, Ph.D. Candidate, European Law School, Humboldt Universität zu Berlin. Contact: smarika.kumar@gmail.com for clarifications and citation requests.

2 Bryce Goodman & Seth Flaxman, EU Regulations on Algorithmic Decisionmaking and a "Right to Explanation," Cornell University Library, available at <https://arxiv.org/abs/1606.08813>, (last visited, 15 May 2018.)

3 Article 22, of Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), available at <http://eur-lex.europa.eu/eli/reg/2016/679/oj> (last visited, 23 May 2018)

The present ideation paper is a preliminary approach to this question through a comparison of explanation making and accountability generation techniques in the language of legal rights and in machine learning pre-emptive algorithms.[4] To outline these differences, I focus on the role time plays in constructing (or not, or at least, differing) notions of causation and explanation in both the case of legal rights and pre-emptive machine learning algorithms. Through such an analysis, my objective is to illustrate that how time is constructed has implications for the attribution of responsibility for human actions, and therefore, for power relationships.

II. How Time Constructs Agency in Systems: Understanding Time as 'Time Binding'

What is time and how should we understand it in the context of legal rights and ML pre-emptive algorithms? The understanding of time in its functional component of 'time binding' in the work of Niklaas Luhmann and Elena Esposito is critical in this regard. Luhmann understands time as a character of systems, which is produced by systems to make sense of occurrences in the world. This function of making sense of the world is what Luhmann calls 'time binding,' which indicates "*the generation of structures in the autopoetic process of continuous self-renewal of the system, and not simply the coming into being of factual states (atoms, suns, ozone hole, etc.) of some duration*."[5] This occurs because "*although time itself cannot be bound, it can bind by giving events structural value.*"[6] Accordingly, for Luhmann, time is used when something is observed with the aid of the distinction drawn between 'before' and 'after.'[7] This understanding of time might be clarified with the thought experiment that everything which exists, exists simultaneously. When an observer discerns similarities or reiterations or distinguishes between the before and after, including to attribute effects to causes, he does this only with the aid of distinctions between the 'past' and 'future' which he employs, and under the assumption that his own observation-operations are coincidental with what he observes.[8]

In this framework, operatively closed systems are guided by their immediate past. However, since the aid of memory is employed to calculate the behavourial consistency, and thus project similar outcomes for similar events, inconsistencies can occur in the absence of the praxis of time: "*Just as binocular vision produces spatial depth for the purpose of resolving self-produced visual inconsistencies, the ever more complex*

4 The term "pre-emptive algorithm(s)" here refers to algorithms that mine personal data gathered from an individual's interactions in her '*onlife* world' to make guesses or predictions using machine learning techniques, about the likely actions and risks which the individual will take in the future. The term is used in the sense of Yeung's understanding of the same in her taxonomy of big data driven, machine learning algorithms. Karen Yeung, Algorithmic Regulation: A Critical Interrogation, in Regulation & Governance 3 (2017), p.7.

5 Niklas Luhmann, Risk: A Sociological Theory (trans. R. Barrett), Berlin/New York (1993),p. 53.

6 Supra n. 5, p. 54

7 Supra n. 5, p. 51

8 Supra n. 5, p. 34

memory generates temporal depth in the form of the twin horizons of the past and the future."[9] In this sense, memory-aided operations cannot take everything that they observe as occurring simultaneously, because this would lead to inconsistency and disorientation. And because of this reason, Luhmann asserts, the concept of time with its temporal distinctions of the 'past', 'present' and the 'future' manifests itself: "*The 'before' and 'after' of an event become discrete, and highly complex systems finally become capable of seeing the future in the mirror of the past and of orienting themselves by the difference between the past and the future.*"[10]

The point of interest for Luhmann is not so much what the 'true' nature of time is, and how far it aligns with 'our' or a 'legal' or a 'technological' notion of time, but rather what does a particular form of time allow the system to do? In particular, what does the specific direction of time, the temporal priority where the present succeeds the past and is succeeded by the future makes possible? Sociologist Elena Esposito's work in this regard, which builds on Luhmann's analysis of time and future uncertainty can serve as a useful interjection here. For Esposito, the temporal order and its related possibility of time binding allows us to exploit the uncertainty of the future without being paralysed by it.[11]

Esposito outlines that mere actuality without the notion of time is the space where the possible and necessary coincide, since it is the space where what is, could not be different.[12] This encompasses the 'immediate presence of the present' which leaves no room for freedom, or the possibility of exercise of human agency, since for that we require a certain notion of contingency which allows us to separate actual data and compare them to other possibilities and choose between them. Such contingency is developed through reference to other observers, or through reference to other presents, which are then projected as the 'past' or 'future' of a particular moment, viz. of the actual present. For Esposito as for Luhmann, this generation of time through the formulation of 'past,' 'present' and 'future' in modern societies allows us to vary, design, correct, and learn from experiences of the 'past' for decisions in the 'present' to carve a better 'future': "T*ime allows the system to separate itself from its own operations and its situation, linking it with other (past and future) situations in a complex framework of connections where uniformities, influences and corrections can be found. Time is, therefore, seen as a structure of systems, a structure that gives an order to operations and connects them to one another*."[13] Such time binding allows for human decision-making and the exercise of human agency even in environments with insufficient information.

III. Time Binding in Legal Rights Framework: Affording Causal Explanations

9 Supra n. 5, p. 35

10 Supra n. 5, p.36

11 Elena Esposito, The Future of Futures, Edward Elgar (2011), p.19.

12 *Supra* n. 11, p.21

13 *Ibid.*

How is the notion of time binding relevant for the legal rights framework? Legal rights are often encoded in social norms: To claim a legal right is to assert the legitimacy of a social norm. This is clearly seen in cases where the right to privacy is claimed in new technological contexts, for example: Helen Nissenbaum has extensively outlined how claims for right to privacy largely succeed/are legitimised in those contexts where there already exists a social expectation/norm embodying respect for privacy.[14]

What happens when such social norms are violated? In these cases, law needs to respond by using its force to 'correct' the norm deviant behavior. However, the use of such force is not arbitrary and constructs accountability through its decisionmaking process through a detailed foray into discovering the agent responsible for such deviation. In other words, 'who' violated the norm-based right? Who is responsible? These questions require an 'explanation' of the events surrounding violation of said right. Additionally, these are not just questions of fact, but of policy, with implications for power relationships and require legal analysis. Hart and Honore's theory of legal causation has been influential in order to make this legal analysis to explain the events surrounding a legal rights violation.[15] Such definition of legal causation assumes a certain direction of events: A *starts* a process which *leads to* e, etc.[16] This clues us in to the idea that the construction of causation in rights-violation cases appears upon a specific construction of time. In this sense, legal rights operate with a certain form of time binding, viz. they construct moments in a particular order of past, present and future.

What exactly is this form? Time binding in the legal rights framework is *linear*, since the social norm embedded in said legal rights sets certain social expectations, which are consistent in the entire duration in which those rights are valid. Because norms codified as legal rights lay down expectations for human behaviour, they introduce a time binding which offers consistency over different moments. For example, if a new social norm specifies that the collection and sharing of data through home assistant devices is illegitimate[17], then in the legal

---

14 Helen Nissenbaum, Privacy in Context: Technology, Policy, and the Integrity of Social Life, Stanford (2010), p. 162. However, it should be noted in these contexts that social norms are never a given, but always contested, and therefore have power dynamics to them, a point on which Nissenbaum has been critiqued in her failure to address, See, for example, Simon Dawes, Privacy and the Public/Private Dichotomy, 104(1) Thesis Eleven (2011), pp.115-124.

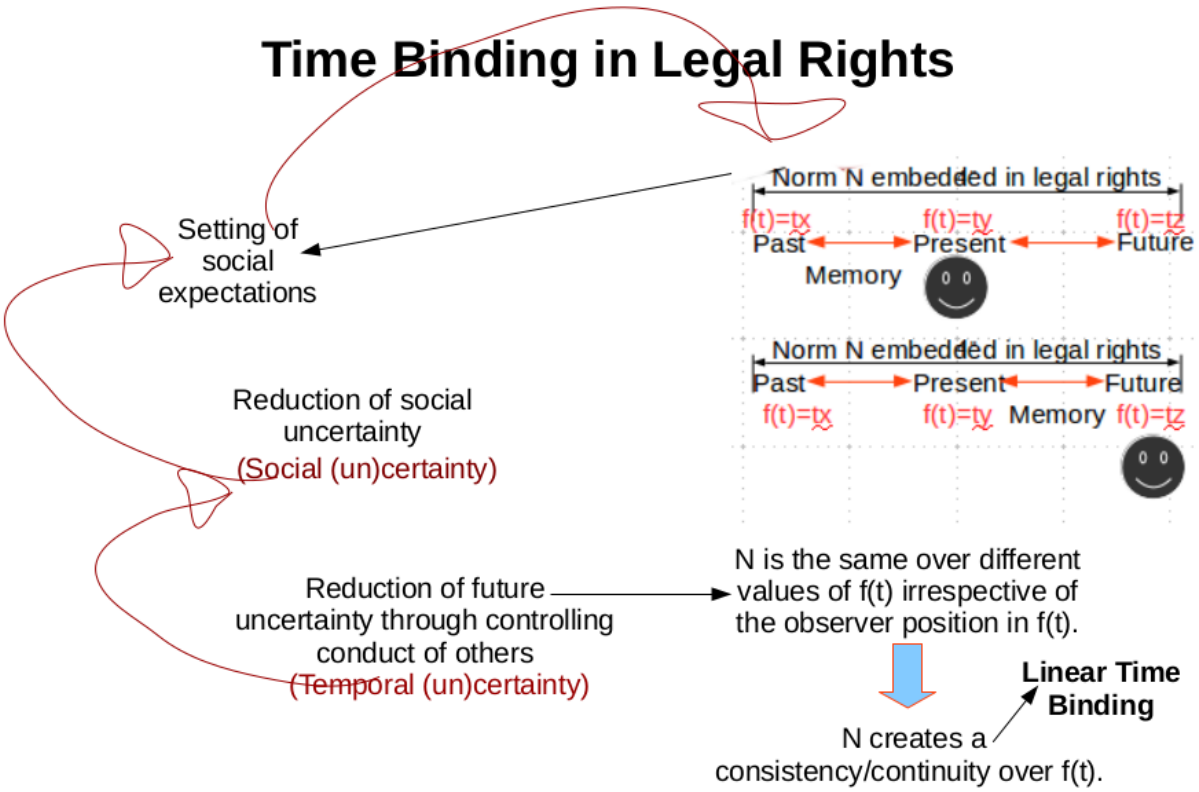
15 To be precise, Hart and Honore propose the following definition of legal causation: „Agent A causes an event e, that might involve agent B, if either of the following holds: 1.A starts some physical process that leads to e, or 2.A provides reasons or draws attention to reasons which influence the conduct of B, who causes e, or 3. A provides B with opportunities to cause e., and 4. All the important negative variants of clauses 1,2,3," see HLA Hart and Tony Honore, Causation in the Law (1959).

16 Jos Lehmann, Joost Breuker, & Bob Brouwer, Causation in AI and Law, 12(4) Artificial Intelligence and Law (2014), p. 289.

17 The notion that a social norm can be determinedly ‚specified' is of course, highly reductive and has been used here merely for purposes of illustration and should not be used to allude to real world scenarios. In reality, ‚specified' social norms (including those embedded in legal rights) have underlying power relations and are always contested, see *supra* n. 14. To treat any position as a ‚specified' social norm in legal discourse then, is to reify and reproduce the power relationships existing in society into law.

rights framework, a person X acting in 2016 should operate with the expectation that this norm would also hold in 2017, 2018 or beyond. And in case X violates the norm in 2017 through certain actions in 2016, legal causation will construct a chain of events, viz. an explanation, such that the norm is consistent both at the time when X acted in 2016 and when it was actually violated. This holds true for any point in time after 2016. Such normative assumptions in the legal rights framework afford *a consistent expectation* of social behaviour 'over time' without which it would be impossible to construct the causal chain which allows for a legal explanation in case of rights violations.

The possibility of causal explanations is also an indicator of how uncertainty is managed in a system, For instance, through the linear time binding described above, temporal uncertainty is translated to social uncertainty. As Esposito remarks, "T*he more one is free to construct one's own temporality, the more one must take the equal and yet opposite freedoms of others into account. The uncertainty of the future is multiplied by the uncertainty of the behavior of all other operators who are oriented to the same future, further increasing the complexity, but also offering the possibility of structuring it. Time bindings have social costs, in that they constrain others and become intertwined with their constraints. However, they also have social support, in so far as the uncertainty appears less uncertain if others handle it in the same way.*"[18]



18 *Supra* n. 11, p.28.

The conversion of temporal uncertainty to social uncertainty such time binding form has the advantage that it allows norm-based rights to address social uncertainty by prescribing normative behaviour. Even in the case of violation of such rights, the legal language of causation allows for the norm to be reaffirmed by offering explanation for the events which led to the violation of the norm. Such explanation, of course, is directed to pinning accountability for the rights violation. But through its workings, it allows for future (temporal) uncertainty to be addressed, since it reinforces social support for the norm by directing the behaviour of others in a certain way (social certainty).

IV. Time Binding in Pre-Emptive ML Algorithms: Affording Comparisons and Correlations

Having outlined the role of causation in constructing explanation in the legal rights framework, I'll now examine the role causation does or does not play in the case of ML based algorithmic predictions. While big data and ML are a more contemporary development, their genealogy can be traced to practices of risk regulation that have been around for atleast the last 200 years.[19] Unlike the rights-based regulation system, where the focus is on preventing or addressing harm, the focus is on controlling the greatest potential threats to achieving regulatory objectives, as determined through ex-ante assessments of their probability and consequences.[20] Accordingly *the use of pre-emptive algorithms is understood as a form of "risk-based prioritisation" which employs statistical prediction techniques to assist in the allocation of enforcement resources.*[21]

Given this background, how may one understand time binding in pre-emptive ML?: As previously discussed, the temporal order in the rights discourse manifests as linear time where determinations the 'future' can be shaped in the 'present' and explanations for the 'present' in the 'past.' This configuration allows the system to have continuity and to retain its identity (in the sense of unity) from one moment to another, from one present to another. Risk-based time binding, on the other hand, appears as radical discontinuity. Unlike in rights-based analysis, in risk analysis, one sees the future from the present future knowing that everything one knows about the future in the 'present future' is possible to change in the 'future present.'[22] Such an orientation lets time appear as a recursive order, since one looks also to the possibilities of the future self to make analyses in the present. In the case of pre-emptive policing

19 Karen Yeung, Algorithmic Regulation: A Critical Interrogation, in Regulation & Governance 3 (2017).

20 A-L. Beaussier, D. Demeritt, A. Griffiths, and H. Rothstein, Accounting for Failure: Risk-Based Regulation and the Problem of Ensuring Healthcare Quality in the NHS, 18 Health, Risk and Society, 205-224.

21 A Griffiths, A-L Beaussier, D Demeritt, and H Rothstein, Intelligent Monitoring? Assessing the Ability of the Care Quality Commission's Statistical Surveillance Tool to Predict Quality and Prioritise NHS Hospital Inspections, 26 BMI Quality and Safety, 120-130.

22 The terms 'present future' and 'future present' are used in the sense Luhmann uses them, so that 'present future' refers to how one sees the future in the moment of now, and 'future present' refers to the now that will pass in the future. "One could also assert that it is necessary to draw a distinction between the present future (when the prognosis is made) and the future present (the point in time when the event occurs), see supra n. 35, p. 73.

algorithms for example, this is evident when predictions about the future possibilities of crime are used to define risks in the now.[23]

Risk-based analysis then requires a new form of temporal organization - "*one that supports the disappearance of the minimal constraints produced by the familiar mode of time binding, the minimal coherence that makes us think that in the future we shall still recognize the norm.*"[24] And because risk analysis does not constrain possibilities of a future present through a norm, the order of time it constructs is recursive, not linear: "*What we believe now, or the information we have now, may not hold, so we have to place ourselves in a future present to make sense or make decisions about the present future. This still produces a time binding form, in so far as it allows us to order different moments, their likeness or differences in a particular way, however circular or recursive that way may be. This is the time binding of risk".*[25]

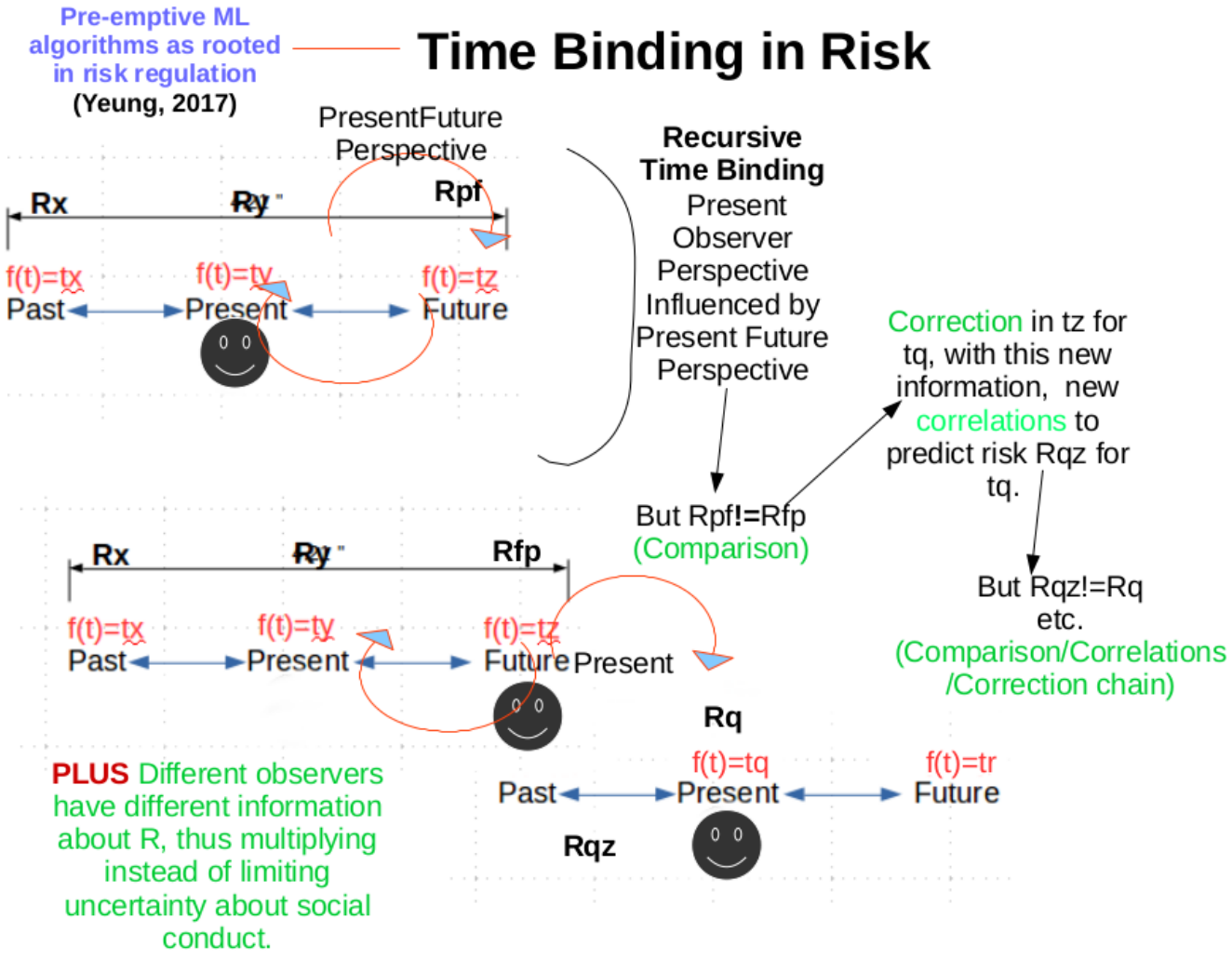
Unlike legal rights framework, from the perspective of risk analysis, temporal uncertainty and social uncertainty thus stand in tension with each other because there is no overarching temporal perspective (like a norm) which can coordinate individual behaviour in the present and the future. Instead, the recursive experience of time creates multiple temporal perspectives for the 'future' not just for the individual but also multiple social perspectives for different individuals. In other words, the uncertainty of one individual's 'future' is multiplied by, rather than offset through or translated into, the uncertainty of others' behaviour: "*What multiplies is the diversity of perspectives that come into play and are accepted (the difference between present present and future present, the difference between the perspective of the one who decides and that of those affected by the decision)*."[26]

23 Lyria Bennett Moses & Janet Chan, Algorithmic Prediction in Policing: Assumptions, Evaluation, and Accountability, in Policing and Society, 8 November 2016. p. 5

24 *Supra* n.11, p.33

25 *Ibid.*

26 *Ibid.*



Such multiplicity of perspectives can have the effect of paralysing decision-making for an uncertain future (made even more uncertain by risk's radical discontinuity,)[27] or make such decision-making seem all the more random. Additionally, such multiplicity does not offer the possibility of explanation in the way linear time binding affords in the legal rights framework. How is then uncertainty addressed in the 'risk society' through applications like pre-emptive ML algorithms?

In contrast to the time binding of rights to establish temporal certainty through social certainty and thus causal explanations, the key feature of this risk regulatory form of time binding is its ability to offer possibilities for comparison and correction of decisions in the future presents by focusing on consequences of decisions. According to Esposito, "*The present creates the future, not as an identity, but as a difference, and this even with regard to itself. One only discovers afterwards the meaning of what one has done, on the basis of the consequences and of the reactions it provoked (that is, enabling the different perspectives of the future and of other observers.) These remain different, but on the basis of the decision, this difference can be used to compare and to correct, thereby building an identity that would not be there if one had not made*

27 *Ibid.*

the decision."[28] This seems to imply that accountability in risk regulation systems like pre-emptive ML algorithms is produced not via causal explanations, but rather by comparing the forecast of the 'present future' with a 'future present.' In case of dissimilarity between the two scenarios, the accountability question asked is not "who" is responsible? (as in rights based frameworks), but rather, how can this be "corrected?"

V. Conclusion: Political Implications and ML Accountability Beyond 'Explanation'?

The present ideation paper has thus argued that risk regulation as embedded in pre-emptive ML algorithms produces its own form of recursive time binding which stays distinct from rights based linear time binding. Additionally, like discussed earlier rights based time binding relies on causation to produce explanations, and accountability for deviation from norm and production of uncertainty. However in risk regulation and in pre-emptive ML, uncertainty is regulated not through an explanatory but rather a comparative mechanism which allows for corrections into decisions to be taken in the future present. What becomes important then in this scenario is to find patterns in data from the world or correlation, which can then be used to make comparison between two or more different kinds of decisions or individual actions and their consequences.

The following table summarises the comparison this paper has undertaken in this regard.

Legal Rights	**Pre-Emptive Machine Learning Algorithms**
Social Norm based time binding	Risk based time binding
Linear time binding	Recursive time binding
Temporal uncertainty consolidated through social uncertainty	Temporal uncertainty multiplied through social uncertainty
Creation of continuity over time function	Creation of 'radical discontinuity' over time function
Construction of causal chains, and therefore, explanation	Construction of chains of comparison, finding correlations and corrections.

28 *Ibid.*

Temporal uncertainty addressed through creating accountability in human agents of causation	Temporal uncertainty addressed through changing future consequences which is implied in both human and non-human factors.

One remarks in the risk-based pre-emptive ML scenario that uncertainty is not sought to be avoided, but rather features of the present, decisions of the present sought to be made in reference to a present future, which would likely be different from the present- both in terms of changed circumstances, and expectations and behaviours of other individuals. And when that present future does not turn out as expected then in a future present, or when others do not behave in the future present as imagined, the individual revises her 'risk taking' behaviour. Pre-emptive ML, like risk regulation, thus defer the problem of explanation and accountability for behaviour to a future present, which of course, never arrives, because of the recursive time in which risk operates. In this sense, risk regulating operations like pre-emptive algorithms, unlike norm-based legal rights, do not operate upon the premise of explanable behaviour or circumstances as a mode to address the question of accountability.

When linear time-bound explanation is not the basis for assigning accountability or for ensuring responsible behaviour, it creates new configurations of power relationships between actors impacted by algorithmic decision-making and those designing or using them as well as the algorithms themselves, in the recursive time bound context of pre-emptive ML. What new modes of power does recursive time binding afford, and how can they be addressed? Interpretations of a 'right to explanation' and corresponding ML algorithmic accountability under the GDPR will largely depend on the (mis-)understanding of such new configurations of politics.