Good decisions in an imperfect world

A human-focused approach to automated decision-making

Bettina Bacher[a]*

[a]*Institute of Law and Technology, Masaryk University, Brno, Czech Republic*

ORCID: 0000-0001-7067-4244

This work was supported by the project CyberSecurity, CyberCrime and Critical Information Infrastructures Center of Excellence (C4e; CZ.02.1.01/0.0./0.0./16_019/0000822)

Abstract: Legal rules are based on an imagined regulatory scene that contains presumptions about the reality a regulation addresses. Regarding automated decision-making (ADM), these include a belief in the "good human decision" that is mirrored in the cautious approach in the GDPR. Yet the "good human decision" defies psychological insight into human weaknesses in decision-making. Instead, it reflects a general unease about algorithmic decisions. Against this background I explore how algorithms become part of human relationships and whether the use of decision systems causes a conflict with human needs, values and the prevailing socio-legal framework. Inspired by the concept of Human-Centered AI, I then discuss how the law may address the apprehension towards decision systems. I outline a human-focused approach to regulating ADM that focuses on improving the practice of decision-making. The interaction between humans and machines is an essential part of the regulation. It must address socio-legal changes caused by decision systems both to integrate them into the existing value system and adapt the latter to changes brought forth by ADM. A human-focused approach thus connects the benefits of technology with human needs and societal values.

Keywords: automated decision-making; GDPR; heuristics and bias; Human-Centered AI; sociotechnical practices.

1 Introduction

In a 1979 essay titled 'To Err Is Human', scientist and poet Lewis Thomas discusses the infallibility of computers and the value of "generative mistakes"[1] only humans can make. "Mistakes are not believed to be part of the normal behavior of a good machine. If things go wrong, it must be a personal, human error [...]. The computer, at its normal best, is infallible."[2] Conversely, "[m]istakes are at the very base of human thought,

---

[1] Maria Popova, The Value of Being Wrong: Lewis Thomas on Generative Mistakes <https://www.themarginalian.org/2023/03/18/lewis-thomas-mistakes/>.

[2] Lewis Thomas, *The Medusa and the Snail, More Notes of a Biology Watcher* (Bantam 1979), 29.

embedded there, feeding the structure like root nodules. If we were not provided with the knack of being wrong, we could never get anything useful done."[3] Mistakes may lead to progress, and "[t]he capacity to [...] land lightly on the wrong side represents the highest of human endowments. [...] We are at our human finest, dancing with our minds, when there are more choices than two. [...] This process is called exploration and is based on human fallibility. If we had only a single center in our brains, capable of responding only when a correct decision was to be made, [...] we could only stay the way we are today, stuck fast."[4] Yet, our most challenging problems may be solved if computers learn to make mistakes: "Think of what we could gain from the near infinity of precise, machine-made miscomputation, which is now so easily within our grasp. [...] If it is a big enough mistake, we could find ourselves on a new level, stunned, out in the clear, ready to move again."[5]

Thomas' essay anticipated several aspects of the legal conversation about automated individual decision-making (ADM) long before it existed: the juxtaposition of human and machine-made decisions, the difference in their perception by humans, the generative potential of human fallibility, and the vague idea of salvation by infallible machines. Currently, suspicion seems to predominate the legal discourse on ADM. It is mirrored by Art. 22 GDPR[6], which gives the data subject the right to obtain

---

[3] Thomas (n 2), 30.

[4] Thomas (n 2), 31.

[5] Thomas (n 2), 32.

[6] Regulation (EU) 2016/679 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 [2016] OJ L119/1 (hereafter GDPR).

human intervention. According to a widespread narrative, humans decide more nuanced, accurately and humanely, while decision systems are prone to bias and absurd results due to a lack of human understanding. Consequently, humans seem necessary to protect the data subject from ADM by correcting algorithmic errors through human insight. Yet this division between "good" human and "bad" automated decisions contrasts with psychological research. It shows significant weaknesses in human decision-making. Humans are less rational and flawless than they wish to be. Cognitive biases can mislead decisions, and inconsistencies may occur not only between different people but within the same person whose decisions may be influenced by stress, mood, or hunger. Evidence shows that even simple algorithms outperform humans and make more consistent decisions.

Any regulation is connected to an "imagined regulatory scene" containing a "set of background understandings about the paradigmatic cases, practices, and areas of social life" to which the rules should apply.[7] "Phenomenological models" about reality are embedded in the law.[8] In this paper, I examine the imagined regulatory scene for ADM and possible consequences for its regulation. Based on interdisciplinary research, I first describe how the perception of ADM influences its use (3). Then, I analyse algorithms as a part of human relationships (4). I explore the distribution of power in the decision-making process, whether the use of decision systems changes it, and

---

[7] Jack M. Balkin & Reva B. Siegel, 'Principles, Practices, and Social Movements' (2006) 154 U Pa L Rev 927, 928.

[8] Linda Hamilton Krieger & Susan T. Fiske, 'Behavioral Realism in Employment Discrimination Law: Implicit Bias and Disparate Treatment' (2006) 94 Calif L Rev 997, 999.

whether ADM affects the socio-legal framework governing the relationship between the parties involved. Finally, I discuss how the law can embrace these insights to develop a more realistic regulatory scene and serve as an instrument to reduce the concerns about ADM (6). As a guiding principle, I propose a human-centered approach focusing primarily on making good decisions and increasing acceptance of ADM.

2 Definitions

I use the notion of ADM not in the narrow sense of art. 22 GDPR, but as an umbrella term for any decision not made solely by a human. It is irrelevant how sophisticated the decision system is. Equally irrelevant is the role allocation between the system and a possible human-in-the-loop (HITL). The system may decide fully automated or only make proposals to a HITL who ultimately decides.

I discuss only professional decisions taken on behalf of organisations that concern a third person, usually a human (e.g. hiring a new accountant).[9] Such decisions are most likely to be handled by ADM. Unlike Art. 22 GDPR, which applies to state actors and private parties, I focus on decisions made by the latter. Given the hierarchical relationship between the parties involved and the comprehensive procedural laws governing state decisions, discussing them as well would considerably increase the scope of this paper without substantially adding to its main purposes.[10]

---

[9] Notably, I will not discuss personal choices.

[10] On automation of administrative decisions see e.g. Cass R. Sunstein, 'Governing by Algorithm? No Noise and (Potentially) less Bias' (2022) 71 Duke LJ 1175; Aziz Z. Huq, 'Constitutional Rights in the Machine-Learning State' (2020) 105 Cornell L Rev 1875.

Decisions usually require some kind of judgment. It can be defined as "measurement in which the instrument is a human mind", measurement being "the act of using an instrument to assign a value on a scale to an object or event".[11] The scale used to make a judgment must not be numerical; any other hierarchy is possible. No matter the scale, making a judgment aims at a certain accuracy in the measurement[12], even though there is no certitude that a judgment is correct.[13] Matters of judgment allow for some level of disagreement but imply the belief that people "should agree".[14]

Kahneman et al. distinguish between predictive and evaluative judgments. A predictive judgment involves a prognosis, a presumption about something to happen in the future that can be verifiable or not.[15] If it is unverifiable, it is impossible to evaluate the quality of the decision based on the outcome. [16] Evaluative judgments are not based on a prognosis but are an assessment of the nature, character or quality of something according to a scale, often also by comparing several items that share some common properties.[17] However, the distinction gets easily blurred, notably if a decision involves

---

[11] Daniel Kahneman, Oliver Sibony & Cass R. Sunstein, *Noise, A Flaw in Human Judgment* (William Collins 2021), 39.

[12] Kahneman, Sibony & Sunstein (n 11), 39.

[13] Kahneman, Sibony & Sunstein (n 11), 43.

[14] Kahneman, Sibony & Sunstein (n 11), 43.

[15] Kahneman, Sibony & Sunstein (n 11), 47.

[16] Kahneman, Sibony & Sunstein (n 11), 47.

[17] Kahneman, Sibony & Sunstein (n 11), 51, do not define the term but give sentencing a crime, grading papers or awarding grants as examples for evaluative judgments.

the evaluation of multiple options and their various trade-offs that are themselves based on predictions.[18]

The decisions discussed here usually do not require a "discrete decision or judgment in the face of complexity", i.e. "when their constitutive considerations are numerous, contradictory, ambiguous, and incommensurate."[19] In the private field, arbitrators judging large cases may make such decisions. They can be influenced by the human weaknesses explained below[20], but at this point, it is unlikely that decision

---

[18] Kahneman, Sibony & Sunstein (n 11), 51.

[19] Dan Simon, 'A Third View of the Black Box: Cognitive Coherence in Legal Decision Making' (2004) 71 U Chi L Rev 511, 516.

[20] See e.g. Chris Guthrie, Jeffrey J. Rachlinski & Andrew J. Wistrich, 'Inside the Judicial Mind' (2001) 86 Cornell L Rev 777; Chris Guthrie, Jeffrey J. Rachlinski & Andrew J. Wistrich, 'Blinking on the Bench: How Judges Decide Cases' (2007) 93 Cornell L Rev 1; Doron Teichman & Eyal Zamir, Judicial Decision-Making: A Behavioral Perspective in Eyal Zamir & Doron Teichman (eds), *The Oxford Handbook of Behavioral Economics and the Law* (OUP 2014), 664; Nicole E. Negowetti, 'Judicial Decision-making, Empathy, and the Limits of Perception' (2014) 47 Akron L Rev 693. Regarding the heuristics and biases-concept presented below Christoph K. Winter, 'The Value of Behavioral Economics for EU Judicial Decision-Making' (2020) 21 German LJ 240. For complex decisions the concept of coherence-based reasoning offers psychological insight in human decision-making processes, see e.g. Simon (n 19); Keith J. Holyoak & Derek Powell, 'Deontological Coherence: A Framework for Commonsense Moral Reasoning' (2016) 142 Psychological Bulletin 1179; Dorothee Mischkowski, Andreas Glöckner & Peter Lewisch, 'Information search, coherence effects, and their interplay in legal decision making' (2021) 87 J of Economic Psychology 102445 (https://doi.org/10.1016/j.joep.2021.102445).

systems can make this kind of complex, non-routine judgment.[21] Therefore, they are not the focus of this paper. I will also not discuss the use of ADM in specific contexts like content moderation.

3 "Good" human and "bad" automated decisions

A recurrent topos in the discourse on ADM is the idea of the "good human decision" that considers human decision-making indiscriminately as superior. Proponents concede that humans may not be infallible but argue that they decide more adequately, nuanced, and individually than decision systems. Yet psychological insight substantially relativises this idea and suggests that its staying power might instead be connected to an unease about ADM.

---

[21] Digital technology may be used in the judiciary (and hence in arbitration) for other purposes. See e.g. Amnon Reichman, Yair Sagy & Shlomi Balaban, 'From a Panacea to a Panopticon: The Use and Misuse of Technology in the Regulation of Judges' (2020) 71 Hastings LJ 589; David Freeman Engstrom & Jonah B. Gelbach, 'Legal Tech, Civil Procedure, and the Future of Adversarialism' (2021) 169 U Pa L Rev 1001; David Freeman Engstrom, 'Digital Civil Procedure' (2021) 169 U Pa L Rev 2243; Lynn Winmill, 'Technology in the Judiciary: One Judge's Experience' (2020) 68 Drake L Rev 831; Daniel L. Chen, 'Judicial analytics and the great transformation of American Law' (2019) 27 Art Intell & Law 15. Regarding the lawyer's perspective see e.g. John Armour, Richard Parnham & Mari Sako, 'Augmented Lawyering' (2022) 2022 U Ill L Rev 71; Neel Guha, Peter Henderson & Diego A. Zambrano, 'Vulnerabilities in Discovery Tech' (2022) 35 Harv J L & Tech 581; Joshua P. Davis, 'Of Robolawyers and Robojudges' (2022) 73 Hastings LJ 1173.

3.1 Human weaknesses in decision-making

"Theory of mind" or "mentalising" describes the attribution of mental states to self and others.[22] There is a general, pervasive assumption that humans behave rationally in their individual and collective actions.[23] It allows us to regard the conduct of people as reasonable, coherent, and justifiable along a commonly accepted principle, not guided by unconscious, unacknowledged or dark motives.[24] In psychology, such normative theories about reasoning serve as standards to evaluate human behaviour to determine what counts as an error.[25] One concept is the "rational choice" model with its core principle of maximising subjective expected utility.[26] The "rational actor" evaluates the attractiveness of several options by assessing the probability and personal utility of an outcome, then chooses the best option based on the "optimal probability-weighted

---

[22] David Premack & Guy Woodruff, 'Does the chimpanzee have a theory of mind?' (1978) 4 Behav & Brain Sciences 515; Alvin I. Goldman, Theory of Mind in Eric Margolis and others (eds.) *The Oxford Handbook of Philosophy of Cognitive Science* (Oxford 2012), 402.

[23] Goldman (n 22), 408ff.

[24] Luc Boltanski & Laurent Thévenot, *On Justification, The economies of Worth* (Princeton UP, 2006), 43.

[25] Nick Chater & Mike Oaksford, Normative Systems: Logic, probability, and Rational Choice in Keith J. Holyoak & Robert G. Morrison (eds), *The Oxford Handbook of Thinking and Reasoning* (OUP 2012), 11.

[26] Dale W. Griffin and others, Judgmental Heuristics: A Historical Overview, in Keith J. Holyoak & Robert G. Morrison (eds), *The Oxford Handbook of Thinking and Reasoning* (OUP 2012), 323.

utility".[27] Rationality in one's acts is often presumed by law.[28]

Rational thinking and conduct are not factual prerogatives of decision-making, but the presumption of rationality is closely tied to the perception of a decision as valid. Implicitly, decision-making is about well-grounded, analytical reflections that produce comprehensible and accurate decisions.

The rational choice model is the theoretical point of reference for examining errors in human decision-making. Bias and noise are the leading causes of error.[29]

3.1.1 Bias

In general, bias describes "an inclination of temperament or outlook"[30] or "a proneness

---

[27] Griffin and others (n 26), 323. As a benchmark for coherence of the probability assessment serves Bayes' rule; see idem 323f. and Thomas L. Griffiths, Joshua b. Tenenbaum & Charles Kemp, Bayesian Inference in Keith J. Holyoak & Robert G. Morrison (eds), *The Oxford Handbook of Thinking and Reasoning* (OUP 2012), 22.

[28] The rational choice model is particularly important for the Law and Economics approach. See e.g. Russell B. Korobkin & Thomas S. Ulen, 'Law and Behavioral Science: Removing the Rationality Assumption from Law and Economics' (2000) 88 Calif L Rev 1051 and the famous discussion on the paper by Christine Jolls, Cass R. Sunstein & Richard Thaler, 'A Behavioral Approach to Law and Economics' (1998) 50 Stan L Rev 1471; reply by Richard A. Posner, 'Rational Choice, Behavioral Economics, and the Law' (1998) 50 Stan L Rev 1551 as a representant of the rational choice model; reply by Christine Jolls, Cass R. Sunstein & Richard Thaler, 'Theories and Tropes: A Reply to Posner and Kelman' (1998) 50 Stan L Rev 1593.

[29] Kahneman, Sibony & Sunstein (n 11), 63.

[30] Merriam Webster online, «bias».

to a particular kind of thought or action"[31]. In the context of human error in decision-making, the focus is on cognitive bias.[32] These are systematic and predictable errors of judgment.[33] Cognitive bias refers to "cases in which human cognition reliably produces representations that are systematically distorted compared to some aspect of objective reality."[34] In this instance, humans "draw inferences or adopt beliefs where the evidence for doing so in a logically sound manner is either insufficient or absent."[35] Individuals are often unaware of such distortions. In law, cognitive bias may affect the perception of facts, the interpretation of legal statutes and the decision-making process.[36]

---

[31] Merriam Webster online, «tendency».

[32] In English, bias refers to both cognitive biases and illicit discrimination. Sunstein (n 10), 1179, speaks of c-bias and d-bias to make a distinction. In statistics, bias describes a systematic tendency of a system.

[33] Kahneman, Sibony & Sunstein (n 11), 161.

[34] Martie G. Haselton, Daniel Nettle & Damian R. Murray, The Evolution of Cognitive Bias, in David M. Buss, *Handbook of Evolutionary Psychology* (2nd ed., Wiley 2016), 968.

[35] Martie G. Haselton, Daniel Nettle & Damian R. Murray, The Evolution of Cognitive Bias, in David M. Buss, *Handbook of Evolutionary Psychology* (Wiley 2005), 725.

[36] See e.g. Govind Persad, 'When, and How, Should Cognitive Bias Matter to Law' (2014) 32 Law & Ineq 31; Greg Pogarsky, Sean Patrick Roche & Justin T. Pickett, 'Heuristics and Biases, Rational Choice, and Sanction Perceptions' (2017) 55 Criminology 85; Joseph W. Rand, 'Understanding Why Good Lawyers Go Bad: Using Case Studies in Teaching Cognitive Bias in Legal Decision-Making' (2003) 9 Clinical L Rev 731; Jill C. Anderson, 'Misreading like a Lawyer: Cognitive Bias in Statutory Interpretation' (2014) 127 Harv L Rev 1521; Dan M. Kahan, 'Cognitive Bias and the Constitution' (2013) 88 Chi-Kent L Rev 367.

3.1.2 Noise

Noise is variability when consistency would be expected and desired.[37] Unlike bias, errors by noise are not consistent; they are the random, "unwanted divergence of judgments, the unreliability of the measuring instrument we apply to reality."[38] Due to its statistical nature, noise is visible only in recurrent decisions, i.e. decisions taken by interchangeable professionals in similar cases.[39] Yet singular decisions are unlikely to be less noisy[40], since "[w]herever there is judgment, there is noise."[41] Varied decisions may be connected to the ambiguity of the scale they are based on but also to the personality of the professional, which causes differences in ranking or categorisation.[42] Finally, a person may not always interact identically with a case or can be influenced by factors that are irrelevant to the case.[43]

3.2 Reasons for error in human decision-making

There is a large body of psychological research on error in human judgment. It cannot be discussed extensively in this paper. Instead, I only refer to key concepts relevant to the present context.

---

[37] Kahneman, Sibony & Sunstein (n 11), 4, 21.

[38] Kahneman, Sibony & Sunstein (n 11), 363.

[39] Kahneman, Sibony & Sunstein (n 11), 34.

[40] Kahneman, Sibony & Sunstein (n 11), 37f.

[41] Kahneman, Sibony & Sunstein (n 11), 33.

[42] Kahneman, Sibony & Sunstein (n 11), 75f, 366.

[43] Kahneman, Sibony & Sunstein (n 11), 79ff, 366f.

3.2.1 Decision-making and the two Systems of thinking

The heuristics-and-bias program by Kahneman and Tversky is the main psychological framework to demonstrate and explain errors in human decision-making.[44] It is based on the model of two-system thinking. The concept resulted from the observation that people approach cognitive tasks in two distinct manners[45] and evolved from the traditional separation of intuition and reason[46]. According to this model, there are two distinct systems concurrently at work in our thinking: System 1 and System 2. They describe cognitive processes that differ in speed, controllability and the content they operate on.[47] System 1 designates fast, intuitive thinking and operates without voluntary control, requiring little or no effort.[48] System 2 describes deliberate, but effortful thinking that is accessible to voluntary control (e.g. making choices).[49]

Making decisions may appear like a typical activity of System 2. Yet System 1 remains active throughout and provides intuitive suggestions while System 2 monitors

---

[44] An alternative I will not discuss is the fast-and-frugal heuristics concept by Gigerenzer and others; see Gerd Gigerenzer & Peter M. Todd, ABC Research Group, Simple Heuristics That Make Us Smart (OUP, 2000).

[45] On the evolution see Griffin and others (n 26), 339f.

[46] Daniel Kahneman & Shane Frederick, A Model of Heuristic Judgment in Keith J. Holyoak & Robert G. Morrison (eds), *The Cambridge Handbook of Thinking and Reasoning* (CUP 2005), 267.

[47] Kahneman & Frederick (n 46), 267.

[48] Daniel Kahneman, *Thinking, Fast and Slow* (Penguin 2011), 20ff.

[49] Kahneman (n 48), 21. All operations of System 2 require attention. Since there is only a limited capacity of attention available, activities of System 2 interfere with each other and it is difficult or impossible to perform them at the same time (idem, 23).

their quality.[50] These suggestions are called heuristics. They are simplifying operations people resort to when asked a difficult question.[51] If the target question is more difficult than another question (the heuristic question) that yields a credible answer, people provide the answer to the heuristic question as the correct answer to the target question.[52] Whenever the characteristics of the target question differ from the heuristic question, the substitution produces systematic bias.[53] There is, for example, a pronounced dominance of conclusions over arguments when emotions are involved.[54]

Heuristics can produce systematic errors. However, substituting the routine activity of System 1 by System 2 would be tedious and inefficient.[55] Kahneman concludes that "[t]he best we can do is a compromise: learn to recognise situations in

---

[50] Kahneman & Frederick (n 46), 267f.

[51] Kahneman, Sibony & Sunstein (n 11), 161.

[52] Kahneman (n 48), 97, see 98f. for a set of sample questions and their heuristic substitute; Kahneman & Frederick (n 46), 269.

[53] Kahneman & Frederick (n 46), 270.

[54] The so-called affect heuristic (Kahneman (n 48), 103; Kahneman & Frederick (n 46), 271). On the influence of anger and fear Jennifer S. Lerner & Dacher Keltner, 'Beyond valence: Toward a model of emotion-specific influences on judgement and choice' (2000) 14 Cognition and Emotion 473. See e.g. Daniel Kahneman, David Schkade & Cass R. Sunstein, 'Shared Outrage and Erratic Awards: The Psychology of Punitive Damages' (1998) J of Risk & Uncertainty 49, showing the influence of outrage on punitive damages. See Kahneman (n 48), 129ff, for other types of heuristics.

[55] Kahneman (n 48), 28.

which mistakes are likely and try harder to avoid significant mistakes when the stakes are high."[56]

3.2.2 The feeling of judgment completion

Making a judgment is also an internal emotional experience. When deliberating on a question that needs to be decided, people settle for an answer if they have the impression that they are not making progress anymore.[57] Kahneman et al. call this feeling "the internal signal of judgment completion".[58] It is unrelated to any external information about the validity of the judgment and only based on the internal sense of coherence.[59] The perception of coherence is closely connected with System 1. Its "measure of success" is "the coherence of the story it manages to create. [...] [It] is radically insensitive to both the quality and the quantity of the information that gives rise to impressions and intuition."[60] Therefore, it can result in erroneous decisions.

3.3 Do algorithms make better decisions?

As the previous section shows, the mere fact that a human decides is not a valid indicator of a good decision. This raises the question of whether algorithms[61] would be

---

[56] Kahneman (n 48), 28.

[57] Kahneman, Sibony & Sunstein (n 11), 48.

[58] Kahneman, Sibony & Sunstein (n 11), 48.

[59] Kahneman, Sibony & Sunstein (n 11), 49.

[60] Kahneman (n 48), 85f.

[61] Used here in the broadest sense, an algorithm is a set of specific steps taken to achieve something. It does not require a computer, although nowadays the notion is often used in this context; see Panos Louridas, *Algorithms* (MIT-Press 2019), 4f.

more adept at it, provided, of course, that the algorithm is built correctly, the dataset on the issue is correct, up to date and complete, and the question at hand is suitable to be decided by an algorithm.[62]

Neither humans nor algorithms always make ideal decisions. The comparison between humans and algorithms does not ask for perfection but, notably, for more accuracy and consistency. The result of the comparison depends on the type of decision at hand. For predictive decisions, there is ample and robust evidence that algorithms outperform humans.[63] In 1954, psychologist Paul Meehl compared clinical predictions based on

---

[62] "Algorithms do not build themselves. The Achilles' heel of all algorithms is the humans who build them and the choices they make about outcomes, candidate predictors for the algorithm to consider, and the training sample. A critical element of regulating algorithms is regulating humans." (Jon Kleinberg and others, 'Discrimination in the Age of Algorithms' (2019) 10 J of Legal Analysis, 113, 117). See David Lehr & Paul Ohm, 'Playing with the Data: What Legal Scholars Should Learn about Machine Learning' (2017) 51 UCD L Rev 653 regarding the basics of algorithm training.

Weaknesses of algorithms are not the focus of this paper; see e.g. the very nuanced analysis by Sandra G. Mayson, 'Bias in, Bias out' (2019) 128 Yale L J 2218; James A. Allen, 'The Color of Algorithms: An Analysis and Proposed Research Agenda for Deterring Algorithmic Redlining' (2019) 46 Fordham Urb LJ 219; Solon Barocas & Andrew D. Selbst, 'Big Data's Disparate Impact' (2016) 104 Cal L Rev 671; Crystal S. Yang & Will Dobbie, 'Equal Protection under Algorithms: A New Statistical and Legal Framework' (2020) 119 Mich L Rev 291; for examples Cathy O'Neil, *Weapons of Math Destruction, How Big Data Increases Inequality and Threatens Democracy* (Crown, 2016).

[63] Algorithms may generally outperform humans in a particular context, but the best humans in their field may outperform the algorithm (Cass R. Sunstein, 'The use of algorithms in

subjective assessment by professionals with statistical predictions based on scores or ratings, and the latter were consistently more accurate.[64] Thirty years later, Meehl concluded that "[t]here is no controversy in social science that shows such a large body of qualitatively diverse studies coming out so uniformly in the same direction as this one."[65] Since then, numerous studies confirmed[66] that an algorithm either matches or

---

society' (2023) Review of Austrian Economics <https://doi.org/10.1007/s11138-023-00625-z>, 9).

[64] Paul E. Meehl, *Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence* (Jason Aronson, Inc. 1996 [reprint]), notably 83ff.; see for a summary William M. Grove, 'Clinical Versus Statistical Prediction: The Contribution of Paul E. Meehl' (2005) 61 J of Clinical Psychology 1233; an overview of the controversy and the reasons for the negative reception by professionals gives e.g. Kurt Salzinger, 'Clinical, Statistical and Broken-Leg Predictions', (2005) 33 Behavior & Philosophy 91.

[65] Paul E. Meehl, 'Causes and Effects of My Disturbing Little Book', (1986) 50 J of Personality Assessment 370, 373f.

[66] See e.g. the overviews in Robyn M. Dawes, David Faust & Paul E. Meehl, Statistical Prediction versus Clinical Prediction: Improving What Works, in Gideon Keren, Charles Lewis (eds), *A Handbook for Data Analysis in the Behavioural Sciences* (Lawrence Erlbaum 1993), 352 with a list of 10 various fields where superiority was shown (academic success, business bankruptcy, longevity, military training success, myocardial infarction, neuropsychological diagnosis, parole violation, police termination, psychiatric diagnosis, violence); William M. Grove & Paul E. Meehl, 'Comparative Efficiency of Informal (Subjective, Impressionistic) and Formal (Mechanical, Algorithmic) Prediction Procedures: The Clinical-Statistical Controversy' (1996) Psychology, Public Policy, and Law, 293, and

even exceeds the accuracy of professionals, especially in low-validity environments[67]. Humans excel neither at assigning optimal weights to variables nor at applying them consistently.[68] Professionals also rarely get feedback on the accuracy of their decisions to learn from them.[69] So far, there are no clear patterns, neither for cases in which professional decisions are equally or more accurate than those by an algorithm nor for those rare instances when professionals were superior.[70]

Meehl found only a few circumstances when a human is preferable to a formula, later known as the "broken-leg rule".[71] These cases relate to facts that are too rare to be included in a prediction system but are so decisive they should be permitted to override the statistical prognosis if they occur.[72] Sunstein later summarised five prediction problems on which an algorithm will not do well: (1) foresee the effects of social interactions; (2) foresee the effects of context, timing, coincidence or mood; (3) make a

---

the meta-analysis by William M. Grove and others, 'Clinical Versus Mechanical Prediction: A Meta-Analysis' (2000) 12 Psychological Assessment, 19, analyzing a set of 136 studies.

67 Kahneman (n 48), 225. Low-validity environments entail a substantial degree of uncertainty and unpredictability (idem, 223).

68 Dawes & Meehl (n 66), 22.

69 Grove and others (n 66), 25.

70 Grove and others (n 66), 25.

71 See Meehl (n 65), 24f. In that example, the task is to predict whether a professor will go watch a movie on a specific night. The formula will make a prediction that under usual circumstances will be quite accurate. Yet the professional knows in addition that the professor recently broke his leg, reducing the probability of him going to the movies to almost zero.

72 Dawes & Meehl (n 66), 14.

prediction while missing local knowledge about particulars or what is happening on the ground; (4) identify people's preferences; (5) anticipate breakthroughs or shocks.[73] In category (3), humans might do better because of particular knowledge, whereas the phenomenon in category (1) is so complex that accurate predictions are impossible because there is no relevant data beforehand.[74]

Meehl made his observations with decisions relying on statistical methods, and indeed, simple models can perform as well as more complex ones[75]. Decision systems based on machine learning are trained with data from human decisions and are thus potentially incorporating the very flaws they are supposed to correct. Consequently, there is an extensive discourse on how to avoid such errors.[76] In general, though, Meehl's insights apply also in this case.[77]

3.4 Perception of Algorithmic Decisions

From a rational point of view, algorithms would replace human decision-making wherever they perform better. However, this is not the case, notably due to the

---

[73] Sunstein (n 63), 21f.

[74] Sunstein (n 63), 22.

[75] See e.g. Jongbin Jung and others, 'Simple rules to guide expert classifications', J Royal Statistical Society A (2020) 183, Part 3, 771.

[76] See e.g. Kleinberg (n 62), 138, regarding discrimination; Simon Caton & Christian Haas, 'Fairness in Machine Learning: A Survey', 2023 ACM Computing Surveys <https://doi.org/10.1145/3616865> on fairness; Mayson (n 62), 2262ff.

[77] See Jon Kleinberg and others, 'Human Decisions and Machine Predictions', (2018) Q J of Economics, 237; Sendhil Mullainathan & Ziad Obermeyer, 'Diagnosing Physician Error: A Machine Learning Approach to Low-Value Health Care', (2022) 137 Q J of Economics, 679.

perception of ADM. Research shows two opposing phenomena among professionals. On one hand, there is a profound distrust of humans towards ADM. It is termed algorithm aversion and results in the disuse of decision systems. On the other hand, professionals may be overconfident in their trust in decision systems, even when it is not warranted. This is called algorithm appreciation or automation bias and results in misuse of decision systems. Yet another perspective have recipients of an automated decision.

3.4.1 Algorithm aversion and disuse

Algorithm aversion describes a negative perception of ADM.[78] As a result, professionals disuse algorithms, i.e. they reject their abilities and fail to use them. [79] Burton et al. identify five different motives for algorithm aversion.[80] (1) An important role play false expectations about the algorithm that cause various responses, e.g. disuse because the professional feels confident in the domain or due to the belief that human mistakes are random whereas algorithms fail systematically. (2) Disuse can be a result of missing decision control because the system fails to address human needs such as agency, autonomy or control. (3) Organisational or social structures may lack incentives

---

[78] The term was first used by Berkeley J. Dietvorst and others, 'Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err' (2015) 144 J of Experimental Psychology: General 114, 114.

[79] John D. Lee & Katrina A. See, 'Trust in Automation: Designing for Appropriate Reliance', (2004) 46 Human Factors 50, 50.

[80] Jason W. Burton, Mari-Klara Stein & Tina Blegind Jensen, 'A systematic review of algorithm aversion in augmented decision making' (2020) 33 J Behav Dec Making 220, 223ff.

to use decision systems, notably regarding the expectations of what constitutes a good decision. (4) Decision systems may not be used because they are incompatible with the professional's decision style or way of thinking and fail to engage human intuition in decision-making. (5) Since the rationality of a decision is related to the environment in which it happens, disuse can occur when the professional cannot reconcile their view of what constitutes a good decision with the circumstances under which it needs to be taken.

The fourth motive is strongly connected to the feeling of judgment completion mentioned above. Using an algorithm deprives the professional of the internal signal giving them an intuitive sense that the decision is correct.[81] Professionals are reluctant to give up this emotional reward for an algorithm that does not produce perfect results and seemingly cannot replace the security provided by the internal signal.[82]

3.4.2 Algorithm appreciation, automation bias and misuse

Algorithm appreciation[83] or automation bias means that users overestimate the algorithms' consistency and accuracy[84]. It results in misuse because professionals, e.g.

---

[81] Kahneman, Sibony & Sunstein (n 11), 146.

[82] Kahneman, Sibony & Sunstein (n 11), 146.

[83] The term was first used by Jennifer M. Logg, Julia A. Minson & Don A. Moore, 'Algorithm appreciation: People prefer algorithmic to human judgment' (2019) 151 Organizational Behavior & Human Decision Processes 90, 90.

[84] S Mo Jones-Jang & Yong Jin Park, 'How do people react to AI failure? Automation bias, algorithmic aversion, and perceived controllability' (2022) 28 J of Computer-Mediated Communication 1, 2.

in healthcare or aviation[85], unduly rely on algorithms[86], preferring recommendations from decision systems and following their suggestions even if there is evidence that these are wrong or suboptimal[87]. Automation bias thus functions as a heuristic for decisions under uncertainty, similar to other heuristics.[88] An important role plays the "perfection scheme" that incorporates the idea of permanent flawless performance by decision systems.[89] It also seems that human preference for algorithms is highly influenced by context. Users preferred algorithmic recommendations for music or forecasts[90] but did not in scenarios that involved high-stakes issues like legal, medical or hiring decisions[91].

3.4.3 Perception by the recipient of the decision

Some of the findings outlined above may also influence the recipients of the decision, e.g., when they are asked for consent and have to decide whether or not they will agree to the use of ADM. In addition, there is specific research on the perception of ADM by the decision's recipient. Notably, recipients are often laypeople lacking knowledge about ADM. In one study, they underrated errors by algorithms and accepted fewer

---

85 See Raja Parasuraman, 'Humans and Automation: Use, Misuse, Disuse, Abuse' (1997) 39 Human Factors 230, 238ff.

86 Lee & See (n 79), 50.

87 S. Shyam Sundar, Jinyoung Kim, 'Machine Heuristic: When We Trust Computers More than Humans with Our Personal Information' (2019) CHI 2019 Paper 538, 3.

88 Parasuraman (n 85), 239.

89 Jones-Jang & Park (n 84), 1.

90 Logg and others (n 83), 92ff.

91 Jones-Jang & Park (n 84), 6.

errors than in humans.[92] Consumers in another study preferred ADM if it offered benefits compared to the human.[93] The perception of algorithmic decisions seems to depend on the task and the type of use. Mechanical tasks were not perceived differently, whereas in human tasks, ADM was perceived as less fair and trustworthy, evoking more negative emotions.[94] Also, algorithms are trusted less for tasks that seem subjective.[95] A study found that consumers react less positively to a favourable decision by ADM, whereas the effect is not as severe in negative decisions.[96] Some studies seem to indicate a preference for human decisions and broad concerns regarding fairness and

---

[92] Felix G. Rebitschek, Gerd Gigerenzer & Gert G. Wagner, 'People underestimate the errors made by algorithms for credit scoring and recidivism prediction but accept even fewer errors' (2021) nature: scientific reports 11:20171 (https://doi.org/10.1038/s41598-021-99802-y).

[93] Derek E. Bambauer & Michael Risch, 'Worse than Human?' (2021) 53 Ariz St LJ 1091.

[94] Min Kyung Lee, 'Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management' (2018) Big Data & Society 1. Mixed opinions were also found by Theo Araujo and others, 'In AI we trust? Perceptions about automated decision-making by artificial intelligence' (2020) 35 AI & Society 611. Yoan Hermstrüwer & Pascal Langenbach, 'Fair Governance with Humans and Machines' (2023) Psychology, Public Policy, and Law, advance online publication < https://psycnet.apa.org/doi/10.1037/law0000381> found that the fairness score rises with human involvement.

[95] Noah Castelo, Maarten W. Bos & Donald R. Lehman, 'Task-Dependent Algorithm Aversion' (2019) J of Marketing Research 56, 809.

[96] Gizem Yalcim and others, 'Thumbs Up or Down: Consumer Reactions to Decisions by Algorithms Versus Humans', (2022) 59 J of marketing Research 696.

effectiveness.[97] Legal and medical decisions by professionals were rated highest, and algorithms were rated lowest in terms of preference, accuracy, fairness, and ethicalness.[98] Patients are reluctant towards medical AI, particularly because it appears less sensitive to the uniqueness of each case.[99] Human discretion also plays an important role regarding the preference for professionals when it comes to making morally charged decisions.[100]

3.5 Conclusion: Perception shapes use

One might think that the application of technology largely depends on its utility. Yet the misuse and disuse of ADM discussed in the previous paragraphs show that human perception and human-machine interaction play a decisive role. The perception, however, is not uniform but often depends on the context of the decision or the type of question. This suggests that decision-making is often embedded in a social setting

---

[97] Pew Research Center, 'Public Attitudes Toward Computer Algorithms', 2016 https://www.pewresearch.org/internet/2018/11/16/public-attitudes-toward-computer-algorithms/.

[98] Joseph Eastwood, Brent Snook & Kirk Lutzer, 'What People Want From Their Professionals: Attitudes Toward Decision-making Strategies' (2012) 25 J Behav Dec Making 458.

[99] Chiara Longoni, Andrea Bonezzi, Carey K Morewedge, 'Resistance to Medical Artificial Intelligence' (2019) 46 J of Consumer Research 629.

[100] Johanna Jauernig, Mattias Uhl & Gari Walkowitz, 'People Prefer Moral Discretion to Algorithms: Algorithm Aversion Beyond Intransparency' (2022) 35 Philosophy & Technology 2 (https://doi.org/10.1007/s13347-021-00495-y); Yochanan E. Bigman & Kurt Gray, 'People are averse to machines making moral decisions' (2018) Cognition 181, 21.

connected to certain expectations. I will explore these socio-legal aspects in the next section.

4 Algorithms in human relationships

So far, I focused on the individual perception of ADM. Yet its appreciation also has a collective, societal dimension. Automated decisions are often made in the context of relationships based on a specific socio-legal setting that defines the parties' expectations. Decision systems become part of the interaction and affect its dynamic, which in turn may affect the socio-legal setting. As a result, ADM can appear as a challenge or even a contradiction to existing socio-legal values and concepts.

An example that is entirely different from this paper's topic describes this issue well. Based on artificial insemination, Bernstein discusses embedded values and the socio-legal acceptance process of technology.[101] She considers three layers of innovation: the technological structure or procedure, the social implications and the application of a technology.[102] The social implications lie at the core of the acceptance process and refer to "a bias that is inherent in the technology itself."[103] Bias, in this case, means the possibilities and limitations of a particular technology. These have repercussions on individuals and society that are independent of the specific application of a technology.[104] "The social implication at issue [...] is [...] the ability to create

---

[101] Gaia Bernstein, 'The Socio-Legal Acceptance of New Technologies: A Close Look at Artificial Insemination' (2002) 77 Wash L Rev 1035.

[102] Bernstein (n 101), 1041.

[103] Bernstein (n 101), 1041.

[104] Bernstein (n 101), 1041.

children without sexual intercourse between a man and a woman, traditionally consummated within the nuclear family. This ability placed the family value embedded in the technology at odds with the prevailing socio-legal conception of the family. The family value embedded in [artificial insemination] enabled family structures that were incompatible with the nuclear genetically related socio-legal conception of the family. It is this conflict between the alternative values of the family which runs through the history of [artificial insemination] and which society struggles to resolve."[105]

In this section, I explore whether the apprehension regarding ADM may be caused by a similar conflict with human needs, values and socio-legal notions as Bernstein describes. It may be exactly the properties of decision systems that let them outperform humans, notably their consistency and lack of "human factor", which have considerable effects on social interactions and contribute to the scepticism towards ADM.

4.1 Decision-making as sociotechnical practice

Decision-making is a social practice that is an inevitable part of everybody's life. Social practices are sets of behaviours occurring in a specific situation or context.[106] They frequently include common rules, purposes and perceptions.[107] Although practices often show as routine behaviour, this term contains no presumption about whether they are

---

[105] Bernstein (n 101), 1042.

[106] Andreas Reckwitz, 'Grundelemente einer Theorie sozialer Praktiken', (2003) 32 *ZfS* 282, 289

[107] Theodore R. Schatzki, Praxistheorie als flache Ontologie, in Hilmar Schäfer (ed), *Praxistheorie, Ein soziologisches Forschungsprogramm* (Transcript 2016), 29, 33.

yet emerging or already established.[108] The notion usually includes physical patterns that can also refer to the use of technology.[109] Nevertheless, practices encompassing the use of technology are often called sociotechnical practices, emphasising that technology is embedded in society and both evolve in a continuing process of mutual adaptation.[110]

With the use of decision systems, decision-making becomes a sociotechnical practice.[111] Technology is embedded in the decision-making process and the interaction of the parties involved.[112] It mediates social relations between humans, and in this case, "algorithms are the devices through which these social relations are produced."[113]

---

[108] Hilmar Schäfer, Praxis als Wiederholung, Das Denken der Iterabilität und seine Konsequenzen für die Methodologie praxeologischer Forschung in Hilmar Schäfer (ed), *Praxistheorie, Ein soziologisches Forschungsprogramm* (Transcript 2016), 142.

[109] Reckwitz (n 106), 289.

[110] Werner Rammert & Ingo Schulz-Schaeffer, Technik und Handeln: wenn soziales Handeln sich auf menschliches Verhalten und technische Artefakte verteilt in Werner Rammert & Ingo Schulz-Schaeffer (eds), *Können Maschinen handeln?: soziologische Beiträge zum Verhältnis von Mensch und Technik* (Campus 2002), 1, 21; see also Bruno Latour, 'On Technical Mediation – Philosophy, Sociology, Genealogy', Common Knowledge [1994] 3/2 29, 30ff; Günter Ropohl, *Allgemeine Technologie: Eine Systemtheorie der Technik* (3rd ed, Universitätsverlag Karlsruhe 2009), 29ff.

[111] Bettina Bacher, 'Decision-Making Practices in Transition: A Taxonomy of Automated Decision-Making' (2023) 9 *EDPL* 39, 41.

[112] Armin Nassehi, *Muster: Theorie der digitalen Gesellschaft* (CH Beck 2019) 200; Bruno Latour, *Reassembling the Social: An Introduction to Actor-Network-Theory* (OUP 2005) 71.

[113] Jack M. Balkin, '2016 Sidley Austin Distinguished Lecture on Big Data Law and Policy: The Three Laws of Robotics in the Age of Big Data' (2017) 78 Ohio St LJ 1217, 1223.

Technical properties, function patterns, and dysfunctionalities imprint themselves on the relationship, on the communication, and on the decision-making process itself. Technology thus changes the situation between humans and becomes itself a social actor.[114]

Schulz-Schaeffer developed a model that describes how a technical artefact becomes part of human or social interactions.[115] He refers to the concept of (social) actor by Coleman, who distinguishes two functional components of the self.[116] An actor has an object-self that experiences satisfaction (or lack thereof) and an acting-self that serves the object-self, trying to satisfy it.[117] The results created by the acting-self get their positive or negative significance from the perception by the object-self. The object-self represents the subjective, intentional side of the act while the acting-self represents its intentionless, impersonal side.[118] The role of the acting-self may be delegated to another human or a technical artefact that will provide standardised, typified acts.[119]

---

[114] Latour (n 112), 71.

[115] Ingo Schulz-Schaeffer, Technik als sozialer Akteur und als soziale Institution: Sozialität von Technik statt Postsozialität, in Karl-Siegbert Rehberg (ed), *Die Natur der Gesellschaft: Verhandlungen des 33. Kongresses der Deutschen Gesellschaft für Soziologie in Kassel 2006*, Teilbd. 1 u. 2 (Campus 2008), 710ff.

[116] James S. Coleman, *Foundations of Social Theory* (Harvard University Press 1990), 507.

[117] Coleman (n 116), 507.

[118] Schulz-Schaeffer (n 115), 712.

[119] Schulz-Schaeffer (n 115), 711.

In a purely human relationship, both sides have their object and acting-selves. The object-selves embody the respective interest in the collaboration, the intention of an action and the meaning of an interaction, but also the satisfaction in a completed transaction, while the acting-selves perform accordingly. There is unity between the two selves. This unity dissolves if a technical artefact becomes part of the relationship. It substitutes the acting-self on one side. The acting-self on the other side gets a different counterpart and interacts with a technical acting-self. The abilities and limitations of the technical artefact circumscribe, what actions and interactions it is capable of.[120]

The model shows that there are profound changes on the action level if technology becomes part of a relationship, whereas the perception of the interaction remains human. In the context of ADM, this perception is intimately linked to how humans appreciate the situation of decision-making in general, independently of technology.

4.2 Agency and vulnerability in decision-making

Decisions concerning another person epitomise the inevitable interdependence between humans, which is linked to what Bakan describes as the two fundamental modalities of human existence: agency and communion.[121] Agency is related to the interests of the self, whereas communion is connected to the interests of others.[122] Agency embodies

---

[120] Schulz-Schaeffer (n 115), 711.

[121] The distinction was introduced by David Bakan, *The Duality of Human Existence, Isolation and Communion in Western Man* (Beacon 1966), 14f.

[122] Andrea E. Abele & Bogdan Wojciszke, 'Agency and Communion From the Perspective of Self Versus others? ' (2007) 93 J of Personality & Social Psychology 751, 752.

striving to expand and protect the self and reach one's goals efficiently.[123] Communion manifests itself in "a sense of being at one with other organisms"[124] and "integrat[ing] the self in a larger social unit through caring for others"[125]. The need to dominate others is connected with agency, whereas communion can be seen in "noncontractual cooperation".[126] Bakan notes that "the very split of agency from communion, which is a separation, arises from the agency feature itself; and that it represses the communion from which it has separated itself."[127]

For both sides of the decision, its modalities are connected to the vulnerability of humans in their interactions with others. Inevitably, we must allow "other people to get into positions where they can, if they choose, injure what we care about since those are the same positions that they must be in to help us take care of what we care about."[128] We are vulnerable in our agency because someone may disregard our wishes or needs, limiting our autonomy. At the same time, we are vulnerable in communion because often our actions require "some completion by others if our intentions are to be fully carried out"[129], but these others may refuse to cooperate or act according to our

---

[123] Bakan (n 121), 15; Abele & Wojciszke (n 122), 752.

[124] Bakan (n 121), 15.

[125] Abele & Wojciszke (n 122), 752.

[126] Bakan (n 121), 15.

[127] Bakan (n 121), 15.

[128] Annette Baier, 'Trust and Antitrust', (1986) 96 Ethics 231, 236.

[129] Annette C. Baier, Doing Things with Others: The Mental Commons in Lilli Alanen, Sara Heinämaa & Thomas Wallgren (eds), *Commonality and Particularity in Ethics* (Macmillan 1997), 15.

requests. The perception of this vulnerability is influenced considerably by trust. It can be defined as accepting one's vulnerability to another's will.[130] To trust means letting others take care of something we care about, and this caring for "involves some exercise of discretionary powers."[131] A person who trusts forestalls the future and acts as if the future were safe[132], although entrusting discretionary powers means risking their abuse[133].

Decisions may for both sides take place at a crossroads that defines their respective future. Yet decision-making often conveys a more active, powerful role to the deciding entity and a more passive, weaker role to the recipient, frequently mirroring general societal power relationships. The recipient of the decision may feel particularly vulnerable regarding agency and communion. The vulnerability to the other's will can be felt acutely since being subject to a decision ultimately means giving up control.[134] This can affect the *sense* of agency. The psychological notion describes

---

[130] Baier (n 128), 235.

[131] Baier (n 128), 240.

[132] Niklas Luhman, Vertrauen (4th ed., Lucius & Lucius 2000), 9.

[133] Baier (n 128), 239.

[134] Not surprisingly, people were found to prefer deciding themselves to retain agency, even if giving up control meant maximizing their rewards; see Sebastian Bobadilla-Suarez, Cass Sunstein & Tali Sharot, 'The intrinsic value of choice: The propensity to under-delegate in the face of potential gains and losses' (2017) J of Risk & Uncertainty 54, 187. On the affective value of control see Lauren A. Leotti, Mauricio R. Delgado, 'The Inherent Reward of Choice' (2011) Psychological Science 22, 1310; Roy Shoval, Noam Karsh & Baruch Eitam, 'Choosing to choose or not', (2022) Judgment and Decision Making 17, 768.

the perception of agency, the feeling of being in charge when it comes to our actions[135] and not being helplessly exposed to our surroundings.

4.3 Effects of ADM on power relations and trust

As the previous section shows, decisions are situations with unequal distribution of power, agency and vulnerability. A trusting relationship between the parties can somewhat reduce at least the perception of this inequality, but not eliminate it. Decision systems become part of this sensitive interaction and may affect various aspects of the situation.

The use of ADM systems changes whom humans have to trust. In a purely human interaction, one side trusts another human. With ADM, the human counterpart must trust in technology and/or the deciding entity for using an accurate decision system. Trust in technology or, more precisely, in automation becomes a new, distinct dimension of the relationship. Automation means that a machine agent executes a function formerly carried out by a human.[136] Trust in automation describes "the attitude that an agent will help achieve an individual's goals in a situation characterised by uncertainty and vulnerability."[137] "Trust in automation guides reliance when the

---

[135] James W. Moore, 'What is the Sense of Agency and Why Does it Matter? ', (2016) Frontiers in Psychology 1271 (doi: 10.3389/fpsyg.2016.01272), 1. The sense of agency is a profound human need that is important for our health and well-being (idem, 4). See generally Patrick Haggard & Baruch Eitam (eds), *The Sense of Agency* (Oxford 2015).

[136] Parasuraman (n 85), 231.

[137] Lee & See (n 79), 54.

complexity of the automation makes a complete understanding impractical."[138] With increasing complexity, trust in automation is both harder to achieve and more relevant for people's ability to accommodate the move away from simple technology.[139] This applies to decision systems paradigmatically. Not only are they, or at least appear as, complex systems, making their application particularly challenging and dependent on trust, but they also operate in situations where trust plays a preeminent role.

In addition, it is highly relevant that artefacts lack human perception. Humans make sense of the world based on meaning. Deciding humans may have a completely different perspective than the recipient, but the latter can still rely on the fact that their counterpart's manner of perceiving the world is identical to their own. The recipient has a theory of mind regarding the deciding human and may hope they deliberate particularly careful because they fully understand the importance of the issue. This theory of mind breaks down when humans interact with intelligent systems because their manner of solving problems is "fundamentally alien to humans without training in computer science."[140] Notably, decision systems have no access to meaning. They may make sophisticated "decisions" based on probabilities and patterns they were programmed for, but they will never grasp the meaning of their actions or of the "decisions" they take. The more significant the decision appears, the harder it is to deal with a deciding entity unable to understand the implications of its actions.

---

[138] Lee & See (n 79), 52.

[139] Lee & See (n 79), 52.

[140] Mark O. Riedl, 'Human-centered artificial intelligence and machine learning' (2018) Hum Behav & Emerg Tech 33.

Decision systems also affect the interaction between the parties. In human dealings, both sides often perform intricate social rituals to establish mutual trust and stabilise the relationship. Although many human decisions happen without direct communication, the parties may, e.g. feel the need to communicate their respective views on the decision and its outcome if the issue at hand is particularly complex or significant. A decision system, however, acts in a standardised manner and is unable to participate meaningfully in such an exchange. Even if the artefact were to execute some of these rituals, the human counterpart would always know that it follows through with a script. In addition, decision systems subtly change the power[141] structure of the interaction. With a human counterpart, the recipient knows that the deciding human inevitably has been subject to decisions themselves and has the same needs for trust and agency. Implicitly, the recipient may hope that the deciding human uses their discretion adequately and behaves compassionately because they know that, in another instance, *they* will be at the receiving end of a decision. If a technical artefact becomes the deciding entity, the human recipient knows that the artefact will never be affected by a decision and does not have human needs. Artefacts neither trust nor require it, they neither have a sense of agency nor require it, and they are not torn between communion and agency as every human is.

---

[141] Power and control are important topoi in the sociological discourse on algorithms; see for an overview Jenna Burrell & Marion Foucarde, 'The Society of Algorithms' (2021) Annual Rev Sociology 47, 213. My focus lies on the power structure of the individual interaction, although its perception may be influenced by the general distribution of power in digital society.

Moreover, decision systems may affect an already unequal distribution of power. The inequality of agency and vulnerability is influenced by the means each party has at its disposal to achieve its respective goals in the decision. ADM can be used to enhance the possibilities of the party using the system to reach its goals or to increase accuracy, fairness, and consistency in decision-making, thus reducing vulnerability and increasing agency on both sides. But as long as only one side decides unilaterally on the application of ADM, the use of decision systems is connected to an inequality in agency, and there remains, if not a tendency, then at least a suspicion that the decision system serves primarily the ends of the party using it. Notably, decision systems create "relationships of informational power" because they increase the information one side has at its disposal, whereas the other side knows as little as before and has hardly any control over the information used for the decision.[142] An increase in information often results in increased agency and reduced vulnerability. The relationship of informational power exists even if the decision system is used for the mutual benefit of both parties. Still, its relevance increases if the decision system primarily serves the needs of the party with informational power. Ultimately, the "effects of Robotics [or decision systems] are always about the relationships of power between human beings or groups of human beings."[143]

4.4 Socio-legal shifts caused by ADM

The use of decision systems can thus have a profound influence on both trust and the distribution of power, which is itself a function of agency and vulnerability on both

[142] Belkin (n 113), 1226.

[143] Balkin (n 113), 1225.

sides. These elements also play an important role in the socio-legal framework at the base of decisions. I will explore next, how ADM affects this framework.

In private legal relations, the parties are generally presumed to be equal. It is a longstanding private law topic that this equality is purely formal in many regards because the factual leverage and bargaining power are often distributed unequally. Yet this formal equality is a fundamental socio-legal principle of private interactions. Since ADM can affect the power relationship between private parties, it adds to the already existing challenges to this formal equality.[144] Similar to the effect Bernstein described for artificial insemination, ADM may appear contradictory to this equality and exacerbate the visibility of inequality.

Additionally, the human counterparts of a decision system may feel less like the individual legal subjects they indeed are because they are not perceived as individuals but merely as part of some category in an algorithmic model.[145] The decision system does not interact with the actual human counterpart but with its digital identity as it was constructed based on (usually past) data associating positive and negative traits to them[146]. This is particularly delicate when the decision concerns an issue close to a

---

[144] The contradiction to equality by the use of decision systems per se needs to be distinguished from social inequality caused by the decisions taken by the system. On the latter see Frederic Gerdon et al., 'Social impacts of algorithmic decision-making: A research agenda for the social sciences, (2022) Big Data & Society <https://doi.org/10.1177/20539517221089305>

[145] Margot E. Kaminski, 'Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability' (2019) 92 S Cal L Rev 1529, 1542; Katrina Geddes, 'The Death of the Legal Subject' (2023) 25 Vand J Ent & Tech L 1.

[146] Belkin (n 113), 1236.

person's identity and their individual development in the future, e.g. getting hired. On a socio-legal level, the digital identity somewhat offsets the concept of the individual legal subject as a person with singular characteristics. The human counterpart is often unable to control their digital identity, which "interacts" on their behalf with the decision system and thus cannot know whether it "represents" them correctly. This is the most personal aspect of the relationship of informational power.

The law also contains many social practices that give a framework to decision-making, foster trust and lessen the feeling of defenselessness. Among private parties, the most important practice is the contract.[147] It allows one to enter cooperation without being overwhelmed by fear for individual needs. By conveying a sense of control over the interaction, contracts serve the sense of agency.[148] They "are designed for cooperation between mutually suspicious risk-averse strangers" and allow "trust with minimal vulnerability."[149] Yet a contract still requires trust "in the other party's good will and proper use of discretionary powers."[150] For its user, ADM may reduce the trust required in their counterpart because the use of decision systems often goes hand-in-

---

[147] Justin Hurwitz, 'Trust and Online Interaction' (2013) 161 U Pa L Rev 1579, 1598; Eli Bukspan, 'Trust and the Triangle Expectation Model in Twenty-First Century Contract Law' (2013) 11 DePaul Bus & Comm LJ 379, 382ff; see also Leon Yehuda Anidjar, 'Interpersonal Trust and Contract Theory Redux' (2020) 30 S Cal Interdisc LJ 1, 4ff. Relational contract theory describes the intricate interaction between parties particularly well, see i.a. Ian R. Macneil, 'Values in Contract: Internal and External ' (1983-1984) 78 Nw U L Rev 340.

[148] Robin Kar, 'Contract as Empowerment' (2016) 83 U Chi L Rev 759, 763, argues that contracts empower people to influence one another's action by the use of promises.

[149] Baier (n 128), 251.

[150] Baier (n 128), 251.

hand with an increase in information about the human counterpart based on data, notably profiling.[151] For the human counterpart, though, the use of decision systems in contractual relations may offset the feeling of control conveyed by the contract. The use of decision systems can move an issue "out of the space of immediate negotiability"[152] and contradict the longstanding socio-legal concept of contractual negotiations between autonomous, equal private parties. This concept may be pure fiction in many daily-life consumer contracts, no matter whether these take place in the analogue world or the digital space, but ADM reinforces this already existing contradiction between the law in the books and reality.

4.5 Conclusion: Technically mediated changes in relations

If technology becomes a part of social and individual relations, it mediates these interactions and may limit or expand the parties' range of action. Decisions are often delicate situations that let people acutely perceive their dependency on interaction with others. This exposure makes people particularly sensitive to potential or actual changes caused by technology. As the previous paragraphs show, ADM can cause shifts in trust, the perception of vulnerability and agency, and the distribution of power. These changes

---

[151] Bettina Bacher, 'Unterliegen automatisierte Entscheidungen im Privatbereich anderen Legitimitätsanforderungen als solche durch Menschen?' in Simon Schrör and others (eds) *Entscheidungsträger im Internet: Private Entscheidungsstrukturen und Plattformregulierung* (Baden-Baden 2022) 105, 117f.

[152] Shiv Issar & Aneesh Aneesh, 'What is algorithmic governance?', (2021) Sociology Compass <DOI: 10.1111/soc4.12955>, 5, observe this from a sociological perspective, that can be transferred to the law.

also concern socio-legal concepts that are foundational for private relations. This raises the question of whether regulation can address these socio-legal issues and encourage the use of ADM where it is beneficial.

5 Human-focused decisions

The previous sections explore the imagined regulatory scene for rules on ADM. A common thread in the ambivalent attitude towards ADM is that a non-human agent becomes part of a situation that is, per se, difficult for humans, and affects it. Even if this change is for the good, it points to the broader question of how to reap the full benefits of technology without sacrificing human and societal needs. This necessitates combining technology and its development with a normative discourse on values and socio-legal principles. Human-centered design is one path in this direction. It seeks to integrate multidisciplinary insight on different human factors into the design process to make technology work for people[153] and "considers the cognitive, physical, and organisational influences on human behaviour to improve human interaction with products and processes."[154] Together with human-computer interaction, human-centered design led the way to human-centered AI (HCAI). HCAI has yet to be adopted

[153] See John D. Lee and others, Designing for People: An Introduction to Human Factors Engineering (3rd ed, CreateSpace 2017), 2ff. In this perspective, human error in the interaction with technology appears as a symptom of poor design.

[154] Lee and others (n 153), 3.

widely[155], but the draft EU AI Act (AIA)[156], for example, claims to promote AI that serves humans. As a counterbalance to a technology-centered approach[157], HCAI puts humans at the centre of the AI lifecycle to improve its performance in a trustworthy, safe manner[158]. Garibay et al. describe HCAI as AI "that (1) is human well-being oriented, (2) is responsible, (3) respects privacy, (4) incorporates human-centered design and evaluation frameworks, (5) is governance and oversight enabled, and (6) respects human cognitive processes at the human-AI interaction frontier."[159] HCAI thus explicitly[160] connects technology with specific values[161], among them compliance with the law. The law is a building block of HCAI and is one of the predominant instruments

---

[155] Ozlem Ozmen Garibay and others, Six Human-Centered Artificial Intelligence Grand Challenges (2023) 39 Int J of Human-Computer Interaction 391 <https://doi.org/10.1080/10447318.2022.2153320>, Garibay and others (n 158), 393.

[156] Proposal for a Regulation laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts, COM(2021) 206 final, 21.4.2021, Explanatory Memorandum, 1. I will not discuss the AIA further since it is still subject to negotiations.

[157] Dominika Ewa Harasimiuk & Tomasz Braun, *Regulating Artificial Intelligence, Binary Ethics and the Law* (Routledge 2021), 49.

[158] Garibay and others (n 155), 393.

[159] Garibay and others (n 155), 393.

[160] Technology is never neutral.

[161] For an ethical exploration of AI see German Ethics Council, 'Humans and Machines – Challenges of Artificial Intelligence' (2023) < https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-mensch-und-maschine.pdf>.

to develop the values guiding HCAI, concretise them in various contexts, and anchor them in binding rules. Simultaneously, HCAI inspires technology-related law because it contains extra-legal normative concepts.

5.1 A human focus in digital law

Technology circumscribes what actions we *can* perform, whereas the law states what actions we *may* perform.[162] Both sides are connected by the imagined regulatory scene, the paradigmatic factual story at the base of a rule. Technological change can unhinge this connection between factual abilities and normative guidelines. For example, new possibilities for action may introduce new values or change existing social practices. A key issue is thus to realign the law with these new factual possibilities and, simultaneously, connect technology to extant values, often while it is still evolving. This requires a normative argument that is relevant for both the legal and the technical side, as is the case with HCAI. It can serve as an inspiration for a general human-focused approach to digital law because it provides a value-based foundation for technology and offers guidance at any stage of technical development. It requires but is also open to further concretisation, thus allowing for legal evolution.

A human focus on technology can serve as a distinct value and guiding principle in digital law. As a methodical consequence, the law must provide the conditions allowing the principle to unfold. A critical point is the general approach to technology. Too often, the law separates technology from humans or contrasts humans and automation. There is also an inherent danger of reducing a human focus to human

---

[162] Lyria Bennett Moses, 'Recurring Dilemmas: The Law's Race to Keep up with Technological Change' (2007) 2007 U Ill JL Tech & Pol'y 239, 245.

oversight or a HITL.[163] Ironically, these juxtapositions do not necessarily produce technology that serves humans[164] because they deepen the disconnection between humans, values and technology that a human-focused approach aims to bridge. Essential to this approach is, therefore, that the law systematically considers sociotechnical practices, i.e. how humans interact with technology.

A human-focused approach will rarely provide unequivocal guidance. Rather, it requires exploring human needs, how technology may be used to enhance individual lives and society, as well as whether technology affects existing social practices and causes conflicts with prevailing values or socio-legal concepts. Diverging interests and values must then be balanced, focusing on how to make the use of technology beneficial to humans by serving their needs or offering a remedy for their weaknesses. The overarching goal is not to protect humans from technology but to encourage sociotechnical practices that serve humans and prevent practices that harm them.

In the following sections, I outline how this approach may inspire a more nuanced regulation of ADM.

5.2 Better decision-making and decisions

The first step is always to identify relevant sociotechnical practices. In the case of ADM, it is decision-making (see above). A regulation of ADM must rethink the entire decision-making process and constructively incorporate technology to make optimal use

---

[163] See Lena Enqvist, ‘’Human oversight’ in the EU artificial intelligence act: what, when and by whom?’ (2023) 15 Law, Innovation and Technology 508.

[164] See Meg Leta Jones, 'The Ironies of Automation Law: Tying Policy Knots with Fair Automation Practices Principles' (2015) 18 Vand J Ent & Tech L 77.

of both human and technical abilities to reach good decisions.[165] In this process, digital technology may serve as a mirror because it uses aggregated data that shows hidden patterns and weaknesses in human decisions.[166] As a general rule, a function should be automated if the machine surpasses humans.[167] Technology may, for example, be used to improve consistency in decision-making, while humans may be necessary for interaction with the other side to correct factual mistakes. This concept may offer some guidance on when and what to automate, but it still incorporates the dichotomy between humans and machines that is meant to be overcome.[168] When humans collaborate with machines, they form a hybrid unit in which human-machine interaction plays an

---

[165] Online dispute resolution (ODR) is one of the best examples in this regard, see Pavel Loutocký, 'Possible approaches towards the architecture of online courts and their potential in the decision-making process' (2022) Jusletter IT 31.5.2022; Richard Susskind, *Online Courts and the Future of Justice* (Oxford 2019); Ethan Katsh & Orna Rabinovich-Einy, *Digital Justice, Technolgy and the Internet of Disputes* (Oxford 2017). Also see e.g. Hao-Fei Cheng and others, 'How Child Welfare Workers Reduce Racial Disparities in Algorithmic Decisions' (2022) CHI '22 No 162 <https://doi.org/10.1145/3491102.3501831>; Kate Donahue, Alexandra Chouldechova & Krishnaram Kenthapadi, 'Human-Algorithm Collaboration: Achieving Complementarity and Avoiding Unfairness' (2022) FAccT '22 < https://doi.org/10.1145/3531146.3533221>;.

[166] See Mayson (n 62), 2251ff., 2284.

[167] Joost C.F. de Winter & Dimitra Dodou, 'Why the Fitts list has persisted throughout the history of function allocation' (2014) 16 Cogn Tech Work 1, 4. The concept is the foundation of the Fitts list, also known as MABA-MABA (Men are better at-Machines are better at); see idem, Table 1.

[168] See de Winter & Dodou (n 167), 3ff. for an in depth discussion.

essential role (see below).[169] The principal question is, thus, how humans and machines collaborate optimally to reach a goal that serves humans.

While the respective parties' goals may differ considerably, an overarching aim of decision-making is to obtain a valid judgement in as many cases as possible. This lies usually in the interest of both sides. Abstract criteria may be hard to define, and what constitutes a valid decision often depends on the specific issue and context. Yet a central element of good decision-making is consistent, adequate and exact differentiation, i.e. treating equally what is equal and differently what is different. Also, a decision should not be a mere result of the distribution of power between the parties but be based on criteria connected to the issue at hand. Since it is often difficult or even impossible to evaluate a decision itself, it may be necessary to assess the decision-making process instead.[170] In that regard, procedural law can serve as an inspiration. It contains rules to, for example, prevent the decision from being made based on faulty facts or without the possibility of appeal.

As a consequence, rules on ADM need to be more discriminating.[171] The one-size-fits-all approach of Art. 22 GDPR is not adequate to the complexity of the question. Differentiation is necessary in several dimensions based on the social

---

[169] Rebecca Crootof, Margot E. Kaminski & W. Nicholson II Price, 'Humans in the Loop' (2023) 76 Vand L Rev 429 designate such hybrid concepts as HITL-systems; Latour (n 110), 33, describes human-machine composites as hybrid actors.

[170] Kahneman, Sibony & Sunstein (n 11), 51f.

[171] On criteria for differentiation see Bacher (n 111), 42ff; Tobias D. Krafft, Katharina A. Zweig & Pascal D. König, 'How to regulate algorithmic decision-making: A framework of regulatory requirements for different applications' (2022) Regulation & Governance 16, 119.

practices in question. It is relevant whether ADM is used for contractual proceedings (e.g., the conclusion of the contract) or to fulfil the contractual obligation (e.g., medical treatment[172]). In the second case, additional values inherent to the respective contract may be relevant. Predictive decisions will often benefit from the use of ADM, and the law should endorse it. On the other hand, it is unlikely that ADM is currently able to make complex decisions as those made by an arbitrator in large, non-routine cases.[173] Some matters are structurally inaccessible to an algorithm, as is often the case with free speech or artistic freedom on social media.[174] The law should, therefore, restrict the use of ADM where it is unlikely to produce valid results. If humans decide, the law may prescribe the use of decision-hygiene techniques.[175] Finally, good decision-making also includes rules to ensure that the data used allows correct decisions.[176]

5.3 Human-machine interaction as part of legal relations

With the use of decision systems, human-machine interaction becomes an inherent, specific dimension of the legal relationship and, thus, of regulation. Psychological

---

[172] In the case of medical treatments, Regulation (EU) 2017/745 of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC [2017] OJ L 117/1 applies.

[173] See Gizem Halis Kasap, 'Can Artificial Intelligence ("AI") Replace Human Arbitrators? Technological Concerns and Legal Implications' (2021) 2021 J Disp Resol 209.

[174] Bacher (n 111), 48.

[175] Kahneman, Sibony & Sunstein (n 11), 236ff.

[176] Daniel J. Solove, 'The Limitations of Privacy Rights' (2023) 98 Notre Dame L Rev 975, 1013f.

insight regarding human-computer interaction shows that humans behave towards computers in a manner that is "nonsensical when applied to computers but appropriate when directed at other humans."[177] Humans respond to a computer by unconsciously applying social scripts, i.e. patterns for human-human interaction.[178] Consequently, computers must be considered as social actors (CASA-paradigm) and the human-computer relationship as a fundamentally social one.[179] Many principles regarding human interaction known from psychology, communication sciences and sociology are therefore relevant to human-computer interaction and user interface design.[180]

The quality of the human-machine interaction is primarily a consequence of design choices.[181] These are accessible to legal regulation and must reflect the idea of technology serving humans. There is, for example, extensive research on how to increase trust and a sense of agency in humans when they interact with a machine.[182]

---

[177] Clifford Nass, Jonathan Steuer & Ellen R. Tauber, 'Computers are Social Actors' (1994) CHI 1994, 72, 72.

[178] Clifford Nass, Youngme Moon, 'Machines and Mindlessness: Social Responses to Computers' (2000) 56 Journal of Social Issues, 81.

[179] Nass, Steuer & Tauber (n 177), 77. Research on the CASA-paradigm now includes AI and robots, see e.g. Chad Edwards and others, 'Is that a bot running the social media feed? Testing the differences in perceptions of communication quality for a human agent and a bot agent on Twitter' (2014) 33 Computers in Human Behavior, 371.

[180] Nass, Steuer & Tauber (n 177), 77.

[181] The discourse on digitalization often omits that the majority of issues stem from design choices and are not a consequence of immutable natural laws or technological constraints.

[182] See e.g. Parasuraman (n 85), 237ff; Lingwei Cheng & Alexandra Chouldechova, 'Overcoming Algorithm Aversion: A Comparison between Process and Outcome Control'

Of course, the use of automated decision systems only makes sense if they produce valid decisions. The law must ensure developers adhere to certain standards, systems live up to their promises and are used correctly.[183] Certificates or quality labels can enhance the power of users or decision recipients.[184] Also, the upcoming AI Act classifies some decision systems as high-risk applications that must fulfil specific requirements.[185] In practice, there may be different criteria to define, for example, a non-discriminatory or fair system because they all come with their respective trade-

(2023) CHI '23 No. 756 <https://doi.org/10.1145/3544548.3581253>; Debora Zanatto, Mark Chattington & Jan Noyes, 'Human-machine sense of agency' (2021) Int. J Human-Computer Studies 156, 102716.

[183] This vast topic is beyond the scope of this paper and I must refer to other publications. Their diversity demonstrates the various instruments the law has at its disposal to increase a sense of security and trust towards technology. See e.g. Andrew D. Selbst, 'Negligence and AI's Human Users' (2020) 100 BU L Rev 1315; Karni A. Chagal-Feferkorn, 'How Can I Tell If My Algorithm Was Reasonable?' (2021) 27 Mich Tech L Rev 213; Bryan H. Choi, 'Software as a Profession' (2020) 33 Harv J L & Tech 557; Anat Lior, 'Insuring AI: The Role of Insurance in AI Regulation, 35 Harv J L & Tech 467.

[184] Garibay and others (n 155), 396, 419.

[185] Annex III of the Proposal for a regulation of the European Parliament and of the Council on harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts, Draft Compromise Amendments, 16.5.3023 and changes adopted by EU Parlament, 14.6.2023; see Andrej Krištofík & Pavel Loutocký, 'ODR and Online courts: What is their future after the AI Act;' (2024) Jusletter IT 28.3.204.

offs,[186] and the law should offer some guidance. Otherwise, companies or other entities without democratic legitimation will decide which criteria prevail.

Accessible to regulation are also user education, technical literacy of the public and, above all, technology rollout.[187] Frequently, it happens in a manner that makes systems available before they are full-fledged, putting early users in a situation akin to a medical trial, frequently without their knowledge. Such practices are inept at fostering trust in decision systems.[188]

5.4 Addressing socio-legal changes

The most complex issue of a human-focused regulation is addressing the socio-legal implications I discussed in section 4. At the same time, this is likely decisive for improving the acceptance of ADM. The analysis of changes caused by the use of technology is thus essential. Then, the law as a normative instrument allows the discourse about what *ought* to be, e.g. whether the use of decision systems affects existing social practices, to what extent this seems admissible or new sociotechnical practices desirable, and how the use of these systems can take place in sync with general social values.[189] Notably, changes should not merely reflect socio-economical

---

[186] Mayson (n 62), 2233ff.

[187] On the flaws of the often mandatory impact assessments see Andrew D. Selbst, 'An Institutional View of Algorithmic Impact Assessments' (2021) 35 Harv J L & Tech 117.

[188] It may be that such a trial period is inevitable from a development perspective. In that case, though, it should be transparent to all parties involved and come with adequate safety precautions.

[189] The emerging field of Socioinformatics can offer important insight in this regard; see e.g. the concept of Grounded Design that assesses the quality of information technology design

power relations. Hedging excessive power and balancing relevant interests is a familiar private law topos that extends to the use of technology in the context of ADM.

As a common thread, the law must negotiate the often unequal distribution of agency, vulnerability and power in decision-making and connect it with the use of technology to serve humans. If technology causes or exacerbates inequality, it must be justified by objective reasons related to its human-centered use.[190] While it can be difficult to address vulnerability since it is essentially a function of being at the receiving end of a decision, technology may considerably improve the sense of agency. It may enable the recipient to be more in control of the information used or allow to follow the decision-making process more closely, e.g. with automatic status updates. Decision systems can provide a different kind of accountability than the human black box.[191] The deciding entity is likely to use ADM also for economic reasons, e.g. to reduce the workforce or improve its marketing. Such one-sided economic advantages need to be balanced with the interests of the other side to ensure that the use of

---

based on the evaluation of emerging changes in social practices (Markus Rohde and others, 'Grounded Design – a praxeological IS research perspective', (2017) 32 J of Information Technology 163; Volker Wulf and others, 'The Praxeological Research Programme of Socio-Informatics – the Siegen School', (2021) 10 Acta Informatica Pragensia 333). On Socioinformatics in general Zdenek Smutny, 'Social informatics as a concept: Widening the discourse', (2016) J of Information Science 42, 681. From a socioinformatics perspective rather critical regarding ADM is Katharina A. Zweig, *Awkward Intelligence, Where AI Goes Wrong, Why It Matters, and What We Can Do about it* (MIT Press 2022), 193ff.

[190] In the medical field e.g., the use of technology may increase patients' feeling of helplessness, but its use can be justified if it improves treatment.

[191] Mayson (n 62), 2279.

technology is not a mere function of socioeconomic power. The idea of fairness in the use of decision systems may serve as an additional guideline.[192]

The ultimate question is whether legislation serving human needs must include the possibility for one of the involved parties to refuse the use of ADM. Again, this question is more complex than the approach in art. 22 GDPR makes it seem because it may require balancing the goal of a good decision better reached through ADM with other human needs like those for personal interaction. Also, the issue concerns both sides. Should a doctor be allowed not to use ADM even if it improves treatment? Does the deciding party always have to provide the option of a human decision at the request of the recipient, even in cases when systems outperform humans and a human decisionmaker often worsens the decision? Is a HITL a sufficient remedy for being faced with a non-human decision? Such questions must be distinguished according to different fields and types of decisions.

Finally, the law must ensure the prerequisite to trust and a sense of agency in the context of ADM: transparency, understood more comprehensively than as mere information[193]. Much of the mandatory information based on the GDPR remains buried in Terms of Service and privacy statements, which are the opposite of fostering trust

---

[192] Bacher (n 151), 122ff. Fair use of decision systems is not to be confused with fairness of ADM itself; see on the latter Deborah Hellman, 'Measuring Algorithmic Fairness' (2020) 106 Va L Rev 811.

[193] Heike Felzmann and others, 'Towards Transparency by Design for Artificial Intelligence', (2020) Science and Engineering Ethics 26, 3333, 3336ff.

and agency. Frequently, this information does not inform its recipients.[194] Notably, the data subjects are often unable to know whether they are subject to ADM at a specific moment. Yet this would be required to convey agency and build trust[195]. Beyond the informational aspect, transparency can serve as a guiding principle for the design of decision systems (transparency by design).[196]

6 Conclusion

Legal rules are based on an imagined regulatory scene that contains presumptions about the reality a regulation addresses. Regarding ADM, these include a belief in the "good human decision" that defies psychological insight into human weaknesses in decision-making and reflects a general unease about algorithmic decisions. Against this background I explore whether the use of decision systems causes a conflict with human needs, values and socio-legal notions. Frequently, decisions are situations with an unequal distribution of vulnerability and agency and, as a result, of power. Trust can considerably reduce the perception of this inequality. Decision systems become an integral part of this sensitive interaction and can affect these different aspects. In addition, their use may also appear contradictory to fundamental principles and values of the prevailing socio-legal framework, such as the formal equality of private parties.

---

[194] Solove (n 176), 995ff, distinguishes providing information and informing people so that they are able to make educated decisions.

[195] "Out of *past experience* develops a *present* orientation concerning the anticipation of *future behavior*." (Claus Offe, How can we trust our fellow citizens? in Mark E. Warren (ed), *Democracy and Trust* (CUP, 1999), 50).

[196] Felzmann (n 193), 3343ff.

Inspired by the idea of HCAI, I then outline a human-focused approach to regulating ADM. It focuses on improving the sociotechnical practice of decision-making to reach better decisions. Rather than juxtaposing humans and machines, the interaction between them becomes an essential part of the regulation. Also, it must address socio-legal changes caused by decision systems both to integrate them into the existing value system and adapt the latter to changes brought forth by ADM. A human-focused approach thus aims at reaping the benefits of technological change without sacrificing human needs and societal values.

Acknowledgements: I would like to thank Pavel Loutocký for his valuable feedback on an earlier version of this paper.