CHANG-YUN KU

When AIs Say Yes and I Say No

Let's start with a thought experiment. A patient is waiting in the clinic room for the diagnosis result to decide whether he needs brain surgery for his medical conditions. After SaMD processed, the result shows that the patient is classified into the high- risk group with 99.9% of death rates and needs brain surgery immediately. But the result is opposite to your diagnosis that the patient needs not the surgery. Will you, as a physician in this scenario, object the result that SaMD has made? Theoretically, Human should be the one who determines all the decisions and takes AI's results for reference only, as the GDPR Article 22 presumes. But quite the opposite, AI's result has greater influences on Human than we thought. In this paper, I explore the tension between AI's decision and human decision from the Epistemological perspectives, i.e. to justify the reasons behind the positive human beliefs in AI. My conclusion is that positive human beliefs in AI are because we misidentified AI as a general technology, and only if we can recognize their differences correctly, then the requirement of "Human in the loop" in the GDPR Article 22 can have its meaning and function.

Keywords: *Artificial Intelligence, GDPR Article 22, Human in the Loop, Automated Decision-making*

Author Information

Chang-Yun Ku
Academia Sinica, Taiwan
https://www.citi.sinica.edu.tw/pages/evelynku/index_zh.html
Information Law Center, Institutum Iurisprudentiae, Academia Sinica
https://infolaw.iias.sinica.edu.tw/?page_id=562
Research Center for Information Technology Innovation, Academia Sinica
https://www.citi.sinica.edu.tw/people/postdoctoral-fellows

How to cite this article

Ku, Chang-Yun. „When AIs Say Yes and I Say No",
Információs Társadalom XIX, 4. no (2019): 61–76.
https://dx.doi.org/10.22503/inftars.XIX.2019.4.5

When AIs Say Yes and I Say No: On the Tension between AI's Decision and Human's Decision from the Epistemological Perspectives

Chang-Yun Ku

Abstract

Let's start with a thought experiment. A patient is waiting in the clinic room for the diagnosis result to decide whether he needs brain surgery for his medical conditions. After SaMD processed, the result shows that the patient is classified into the high-risk group with 99.9% of death rates and needs brain surgery immediately. But the result is opposite to your diagnosis that the patient needs not the surgery. Will you, as a physician in this scenario, object the result that SaMD has made?

Theoretically, Human should be the one who determines all the decisions and takes AI's results for reference only, as the GDPR Article *22* presumes. But quite the opposite, AI's result has greater influences on Human than we thought. In this paper, I explore the tension between AI's decision and human decision from the Epistemological perspectives, i.e. to justify the reasons behind the positive human beliefs in AI. My conclusion is that positive human beliefs in AI are because we misidentified AI as a general technology, and only if we can recognize their differences correctly, then the requirement of "Human in the loop" in the GDPR Article *22* can have its meaning and function.

Keywords: Artificial Intelligence, GDPR Article 22, Human in the Loop, Automated Decision-making

Introduction

U.S. FDA (U.S. Food & Drug Administration) approves SaMD (Software as Medical Device)[1] for medical diagnosis. China uses the AI social credit system Zhima Credit[2] to replace the traditional financial credit score. Estonia is going to deploy Robot Judge in Court for the small claim cases (Niiler 2019).[3] These examples from different Countries in different fields show that the Artificial Intelligence (AI/AIs), or say the algorithm, is not only an idea in the Sci-Fi but also a reality in our daily lives. Not even mention those AI applications in the private sectors. And Article *22* of the GDPR (EU General Data Protection

[1] U.S. Food & Drug Administration. "Software as a Medical Device (SaMD)." Accessed February 8, 2020. https://www.fda.gov/medical-devices/digital-health/software-medical-device-samd

[2] ZHIMA Credit, Ant Financial Services Group. Accessed February 8, 2020. https://www.xin.xin/#/home

[3] Niiler, Eric. "Can AI Be a Fair Judge in Court? Estonia Thinks So." WIRED. March 25, 2019. https://www.wired.com/story/can-ai-be-fair-judge-court-estonia-thinks-so/

Regulation)[4] on "*Automated individual decision-making, including profiling*" already regulated that for those decisions that are seriously impactful to the data subject should not be determined solely by automated decision-making process.

But the thing isn't as perfect as it sounds. Three empirical studies show that AI's influences on Human are more than we thought. Results from AI decision-making are better than we humans do? If not, what's the reason that we humans believe in AI's decision than Human? To explore this critical issue, I divide this paper into 7 parts. First, I will start from the elaboration of the GDPR Art.22 to point out the importance of "human in the loop" in the automated decision-making process. Second, I will take three experiments' results to show the significant impact of AI to Human, when human facing decision-making process. Third, I'll prove that the presumption of "AI's decision is better than human decision" isn't solid by demonstrating the nature of AI's decision. After these, I will point out the misattribution of AI as a general Technology is the reason we human have believed in AI. Fifth, I'll return to the Art.22 of GDPR and propose three possible solutions from epistemological perspectives to resolve the gap between this provision and the reality. Sixth, I'll push the discussion further for the crucial issue of whether the expert is immune to AI's decision when making professional decisions. And finally, I will summarize my arguments and conclude this article.

The Requirement of "Human in the Loop" in the GDPR Article 22

The GDPR recognized that "… profiling and automated decision-making can pose significant risks for individuals' rights and freedoms which require appropriate safeguards" (WP29 2018)[5], and thus it regulated automated decision-making process in the Art.22 titled *Automated individual decision-making, including profiling*. The contents of this provision are as followed:

1. *The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.*
2. *Paragraph 1 should not apply if the decision:*
   - *(a) is necessary for entering into, or performance of, a contract between the data subject and a data controller;*
   - *(b) is authorized by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subjects rights and freedoms and legitimate interests; or*
   - *(c) is based on the data subject's explicit consent.*

[4] 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). https://eur-lex.europa.eu/eli/reg/2016/679/oj

[5] Article 29 Data Protection Working Party. Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 (WP251rev.01). 2018. P.5.

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

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

According to the Art.22, basically, the "solely automated processing" that could cause "legal effects" or "similarly significant effects" to the data subject, is prohibited. The "Solely" automated processing means the processing is "without human intervention", i.e. the result of automated processing was decided by the algorithm and then automatically delivered to the data subject, but with no prior or meaningful assessment by a human (WP29 2018).[6]

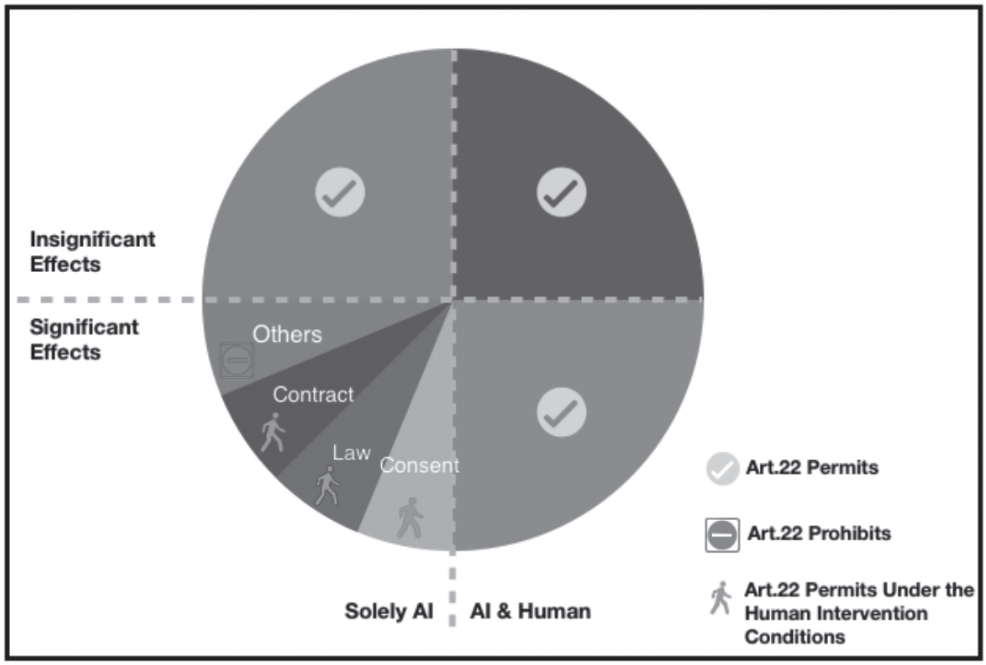
The "legal effects" means that a decision affects someone's legal rights or legal status[7], and the WP29 (Article 29 Data Protection Working Party) names a few examples as the legal right and the legal status. The Legal rights include the freedom to associate, to vote or to take a legal action; and the legal status includes cancellation of a contract, denial of social benefit, refused admission to a country…etc. And the term "similarly significantly effects", refer to the results that are serious impactful to the data subject and thus require the protections under this provision (WP29 2018)[8]; although it's not directly defined in the GDPR, the WP29 explains this as an effect that "must be similar to that of a decision producing a legal effect" (WP29 2018)[9], for example affect someone's financial circumstances, access to health service or employment opportunity…etc. In other words, if the effect of solely automated decision-making isn't serious impactful, then it will not to be regulated or prohibited by the GDPR Art. 22.

Even if the process could cause the legal effect or the similarly significant effect, but under three specific conditions, i.e., for the contract, with the Union or Member State's authorization, or with data subject's explicit consent, the GDPR allows the use of solely automated individual decision-making process. The permissions are under the conditions that if the data controller can meet the requirement of providing appropriate safeguards. These appropriate safeguards include data controller needs to provide meaningful information, specifically the logic of automated decision-making process, to the data subject (WP29 2018)[10], and also provide the opportunities for the data subject to request human intervention, to contest the decision, and to obtain the explanation (GDPR Recital 71)[11].

As mentioned above, for the automated decisions-making process that is regulated by the GDPR Art.22, first, it's only limited to those decisions will cause the significant effect to the data subject in principle; Second, with suitable measures, i.e. human interven-

[6] WP251, p.9.
[7] WP251, p.21.
[8] WP251, p.22.
[9] WP251, p.21.
[10] WP251, p.20.
[11] GDPR Recital 71

General Permissions and Prohibitions of AI Decision-making Process in the GDPR Art. 22



tions, the GDPR permits to use solely automated decision-making for three specific purpose. The "human factor" here is a means to prevent the harm that a solely automated individual decision-making can cause. And this Human is expected to oversee the decision, and who must "… has the authority and competence to change the decision" and "consider all the relevant data" (WP251 2018)[12] In the GDPR's presumption, the Human is capable of making things right when AI goes wrong, and has the authority to determine the best possible decision with AI's advice.

The Inconvenient Truth: The Tension between AI's Decision and Human Decision

To avoid the possible harms, the GDPR Art.22 regulates solely automated decision-making process, which could cause significant effects to the data subject, by requiring human intervention. "Human in the loop" is the solution to the risks post by solely automated decision-making process, i.e. makes the solely automated process "not" solely. But, does this solution actually work? In this section, I would like to introduce three empirical studies to point out a surprising phenomenon: AI 's decision has more influence on Human than we could image, and the Human is actually leaded by AI.

[12] WP251, p.21.

First of all, Human seems more than willing to take the "advice" from AI. Logg et al.'s (Logg et al. 2019)[13] research started from the prevalent presumption of "algorithm aversion", a term refers by Dietvorst et al. (Dietvorst et al. 2015)[14] means that humans distrust algorithm even though algorithm consistently outperform humans. Logg et al. 's study designed to enquire "the role of the self", when human facing the advice from the algorithm, from other people or from people themselves. According to their research results, when people with a choice to take advice from themselves or from other people, 88% of participants would take their own advice. But when people can choose the advices from themselves or from the algorithm, 66% people would take algorithm's advice instead of their owns, even Logg et al. specifically claimed that "the model does not have any additional information that you will not receive" to the research participants (Logg et al. 2019)[15]. The results of their experiments show that "people readily rely on algorithmic advice" (Logg et al. 2019)[16], and Logg et al. call this phenomenon "algorithm appreciation".

Second, the algorithmic advice has a strong impact to human decision than we knew. Vaccaro and Waldo's used the COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) risk scores to test the effects of algorithmic results to human assessments (Vaccaro and Waldo 2019).[17] They divided research participants into two groups: One group was given 40 defendants' profiles with low-risk score, and the other group received the same 40 defendants' profiles but with high-risk score. Their study results show that the scores in high-risk scores group is in average 42.3% higher than the scores in low-risk score group, i.e. AI's results effect people's decision significantly. And this result confirmed the hypothesis of psychological cognitive bias in the Human—the "anchoring effect" —when using algorithmic predictions. Anchoring effect means that during human assessments process, the algorithmic result will "act as an anchor" and thus "individuals will assimilate their estimates to a previously considered standard" (Vaccaro and Waldo 2019)[18]. Vaccaro and Waldo concluded that "even if algorithms do not officially make decisions, they anchor human decision in serious way" (Vaccaro and Waldo 2019).[19]

Thirdly, the Human generally believes algorithmic prediction, even if the accuracy is no more than we flip a coin. Lai and Tan conducted the experiments to evaluate humans' performance when people use different levels of machine assistance, and to see the influ-

[13] Logg, J. M., J. A. Minson and D. A. Moore. "Algorithm appreciation: people prefer algorithmic to human judgment." Organizational Behavior and Human Decision Processes, Vol.: 151 (February 5, 2019): 90-103. https://doi.org/10.1016/j.obhdp.2018.12.005

[14] Dietvorst, Berkeley J., Joseph P. Simmons, and Cade Massey. "Algorithm Aversion: People Erroneously Avoid Algorithms after Seeing Them Err." Journal of Experimental Psychology: General, Vol. 144, Issue 1 (February 2015): 114-126. https://doi.org/10.1037/xge0000033
P.114.

[15] Logg et al. 2019, p.96.

[16] Logg et al. 2019, p.99.

[17] Vaccaro, Michelle and Jim Waldo. "The Effects of Mixing Machine Learning and Human Judgment." Communications of the ACM Vol. 62, No.11 (October, 2019): 104 -110. https://doi.org/10.1145/3359338.

[18] Vaccaro et al. 2019, p.108.

[19] Vaccaro et al. 2019, p.105.

ences of machine accuracies to human predictions (Lai and Tan 2019).[20] They designed different levels of machine assistance for research participants, and they found that machine results with the description of "predicted label", i.e. attached the description of " the machine predicts…" to the results, could effectively improve research participants' performances than without it. Lai and Tan pointed out "this observation also echoes with concerns about humans overly relying on machines" (Lai and Tan 2019).[21] Also, Lai et al.'s research results showed that the participants' trusts are effected by the machine accuracies; more precisely, when the machine accuracies downed from 87%, 70%, 60% to 50% of machine predictions, the degrees of human trusts decreased from 79.6%, 78.6%, 76.9% and 74.5% accordingly. With these insignificantly incline results between the degrees of machine accuracies and the degree of humans' trusts, Lai and Tan concluded that "our findings suggest that any indication of machine accuracy, be it high or low, improves human trust in the machine" (Lai and Tan 2019).[22]

These studies above disclose the inconvenient truth that Human is greatly influenced by the so-called AI, the machine, or the algorithm. AI's result has the essential power to impact human judgment psychologically, whether it works as an anchor and makes humans adjust their decision toward it, or it could even make humans give up their chances to decide completely by taking AI's result instead. Even humans know the accuracies of AI are no more than we toss the coin, i.e. the chances of correctness are 50/50, humans still rather take AI's result anyway.

Following from this inconvenient truth, when the differences or conflicts emerge between the human decision and AI's decision, Human has high possibilities to assimilate human decision toward to AI's decision, or take AI's decision as the final decision instead. In other words, when Human prefers to follow AI's decision, it will eventually lead human to yield the decision authority to AI, as long as AI is in the decision loop. And this phenomenon also makes the GDPR Art. 22 meaningless, i.e. even with the requirement of "human in the loop" to prevent the harms that solely automated process could cause to Human, the human decision-maker will neither choose a different direction from AI, nor challenge AI's result.

And according to this human preference, it's reasonable to ask, why human generally inclines to believe AI's result but not human result? For this question, we are actually asking why Human has these great positive beliefs in AI, or why we believe the result or advice that AI bringing to us have more value or even closer to the truth? And from the epistemological perspectives, more importantly, is the reasons behind this "human positive beliefs" justified? To answer these questions, we need to inquire into two topics sequentially, i.e. the nature and the difference of AI's decision, and the formation of human positive beliefs in AI.

[20] Lai, Vivian & Chenhao Tan. "On Human Predictions with Explanations and Predictions of Machine Learning Models: A Case Study on Deception Detection." In FAT*'19: Proceedings of the Conference on Fairness, Accountability, and Transparency, 29-38. United States, New York: Association for Computing Machinery, 2019. https://doi.org/10.1145/3287560.3287590.
[21] Lai and Tan 2019, p.34.
[22] Lai and Tan, p. 36.

The Nature and the Difference of AI's Decision

Why does Human believe AI's decision or AI's result has more value or closer to the truth than human decision or human result? The presumption of this "positive beliefs" is reasonable if and only if humans presume "AI's decision is definitely better than human decision". Is this presumption true? The only way to justify it is to inquire about the nature of AI's decision and the difference between AI and Human decisions.

The Artificial intelligence is based on the models of Machine Learning (ML) and its branch Deep Learning (DL), and the algorithm is the core to perform these functions. And the material for ML is data, or says Big Data, whether or not it's personal data. The Volume, Variety, Velocity of data are much more than a human can perceive, and thus these data is considered by Human has the qualities of Veracity, Variability and Value. But when the issue comes to AI decision-making, the training data for AI is based on those past decisions that humans have made.

Regards to AI's performance of decision-making, if the training data is generated from those decisions that humans used to make, then, I believe that AI's decision will basically repeat the human decision, or could get even worse, for the five characters that can derive from the training data. These five characters are as follows.

The first character is that AI will repeat the human choice, no matter that decision is right or wrong. What we used to choose, AI will choose the same; what we have decided in the past, AI will have the same decision in the future. And based on this correspondence between the training data and the AI's performance, AI will surely make the same wrong decision as what Human did before. Furthermore, as G. Marcus points out, "Human beings can learn abstract relationship in a few trials…on a capacity to represent abstract relationships between algebra-like variables…; Deep learning currently lacks a mechanism for learning abstractions through explicit, verbal definition, and works best when there are thousands, millions or even billions of training examples…(Marcus 2018)"[23], when facing an unexperienced event, i.e. the new condition is nothing in common with the data in the training datasets, the performance of AI will be no better than Human do.

The second is that AI will amplify the injustice of human past decisions, even though it was unnoticed before. Because of the volume of the training data, AI will not only repeat human past mistakes, but it'll amplify the human errors as well. And the bias we wrongly formed for a decision in the past will become AI's dominant criterion for a decision now. For example, Amazon's AI recruiting tool is abandoned because this AI tool prefers the male candidate than the female candidate, and the reason is the training data of last ten years shows the company hired more men than women (Dastin 2018)[24]; and V. Ordóñez and his team discovered the gender bias is performed when depiction of activities by ML research–image collections, e.g. the activity of washing is linked to women, and the coaching activity is linked to men, since the training data generally linked these activities to the certain gender (Simonite 2017)[25]. As M. Yatskar said "this could work to not only reinforce

[23] Marcus, Gary. "Deep Learning: A Critical Appraisal." ArXiv abs/1801.00631 (January 2018).
[24] Dastin, Jeffrey. "Amazon scraps secret AI recruiting tool that showed bias against women." Reuters, October 10, 2018. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

existing social biases but actually make them worse"(Simonite 2017)[26]. The hidden unjust factor of a human decision is now the main criteria of AI's decision.

The third character is that AI will miss the truly but "off the record" factors. Even if we consider every visible factor of a human decision, but sometimes, the true cause that a human decision-making based on is the "off the record" one, i.e. the factor literally isn't or couldn't be recorded into the training data. The human decision could base on empathy, intuition, memories or any other different kinds of "human things", and these factors aren't possible to be included into the records or the training dataset. "Human judgment is affected by a range of invisible factors that the decision-maker is unable to fully explain when scrutinized" as A. Babuta descripts, the influence of the "Noise" to human decision-making are usual overlooked (Babuta 2018).[27]From this point of view, the training data for AI isn't accurate, and so does the decision that AI makes.

The fourth one is that AI will calculate more factors than a decision need. The algorithm could wrongly link the similar but irrelevant factors of each decision, and thus AI will be trained to make a decision based on those irrelevant factors, e.g. the correlation problem. In the famous Gettier's Problem[28], although the "ten coins" factor has nothing to do with the decision that which job candidate will be hired, but once the former is linked to the latter by algorithm, as what Smith did, AI will definitely decide the later by the former factor, i.e. hire the person only because who has ten coins in the pocket. And as W.T. Chiou indicates "Correlation knowledge is invaluable as it bridges the information gaps and allow a decision to be made in case of ignorance" (Chiou 2018)[29], but "Without causal explanations, decisions based on mere correlations would amount to arbitrary actions that would blame an affected subject for something that cannot be attributable to him or her"(Chiou 2018)[30].

Fifth, it's acknowledged that there are few kinds of AI's decision results are totally disasters, e.g. the AI's face recognition function. The highly inaccuracy of Facial-recognition AI's has widely reported in the media, and the latest NIST's (National Institute of Standards and Technology, U.S.) research result confirmed that "we found empirical evidence for the existence of demographic differentials in the majority of contemporary face recognition algorithms that we evaluated" (NIST 2019)[31]. According to NIST's report, Asian and American Indian individuals have a higher false negative rate than races, and Woman has 2 to 5

[25] Simonite, Tom. "Machines Taught by Photos Learn a Sexist View of Women." WIRED, August 21, 2017. https://www.wired.com/story/machines-taught-by-photos-learn-a-sexist-view-of-women/
[26] Simonite 2017.
[27] Babuta, Alexander. "Innocent Until Predicted Guilty? Artificial Intelligence and Police Decision-Making." RUSI Newsbrief Vol. 38, No. 2(March 2018). https://rusi.org/sites/default/files/20180329_rusi_newsbrief_vol.38_no.2_babuta_web.pdf
[28] Gettier, Edmund L. "Is Justified True Belief Knowledge?" Analysis, Vol. 23, Issue 6 (June 1963): 121–123. https://doi.org/10.1093/analys/23.6.121
[29] Chiou, Wen-Tsong. "Causal Explanation as a Partial Solution to Algorithmic Harms." paper presented at The 7th Academia Sinica Conference on Law, Science and Technology: Emerging Legal Issues for Artificial Intelligence: Legal Liability, Discrimination, Intellectual Property Rights and Beyond, Taipei, Taiwan, November 26-27, 2018, 1-16.
[30] Chiou 2018, p.8-14.
[31] National Institute of Standards and Technology, U.S. Department of Commerce. "Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects." (December 19, 2019) P.7. https://www.nist.gov/programs-projects/face-recognition-vendor-test-frvt-ongoing

times of false positives rates higher than man (NIST 2019)[32]. The reason to cause the inaccuracy of Facial-recognition AI could be many, but as IBM points out "One of the biggest issues causing bias in the area of facial analysis is the lack of diverse data to train systems on" (IBM 2018)[33]. The San Francisco City is the first city in the U.S. passed the Acquisition of Surveillance Technology Ordinance to ban the local agencies for using AI's facial recognition technology in May 2019 (Board of Supervisors, San Francisco 2019)[34], and many other States in the U.S. like New Hampshire, Washington, Indiana, South Carolina etc. are proposing different level's restrictions to Facial-recognition AI (Ramos & Abernethy 2019)[35].

As demonstrated above, in comparing the AI's decision to Human decision, there is no guarantee that AI's decision will definitely better than human decision. In fact, from the training data perspective, AI could precisely repeat human decision and amplify the unjust one, but the flexibility of Human to adjust the decision from the aware mistakes in the past is absent in AI. Human ability of adjustment is the answer to why AI's decision is no better than Human.

Of course, for the record, I have no intentions to deny the possibility that AI could bring a better decision than Human do. But from the Philosophical perspectives, the reasoning is as important as the conclusion, and sometimes it's even more important than the result. Thus, these factors that AI computes for a specific decision should be elaborated and disclosed, as The Right to Explanation or the Algorithmic Transparency that many scholars request. Otherwise, how can we prove that the AI's decision is truly better than Human? And these requirements aren't seem to be achievable by AI in the present days.

Our Beliefs in AI: An Empirical Explanation

People are holding positive and optimistic attitudes toward AI. According to ARM Northstar's survey which across the U.S., E.U. and Asia, 61% of 3938 participants believe AI will make the world a better place, especially in healthcare, science and traffic control fields (ARM Northstar 2017)[36]. As Genesys' National online survey, it shows that 70% of 1001 U.S. employees are with the positive attitude toward AI's impact of workplace (GENESYS 2019)[37]. And similarly, according to the Northeastern University and Gallup's mail survey

[32] NIST 2019, p.7-8.
[33] IBM. "IBM to release world's largest annotation dataset for studying bias in facial analysis." Accessed February 8, 2020. https://www.ibm.com/blogs/research/2018/06/ai-facial-analytics/
[34] Board of Supervisors, City and County of San Francisco. Administrative Code-Acquisition of Surveillance Ordinance. Accessed February 9 2020. https://sfgov.legistar.com/LegislationDetail.aspx?ID=3953862&GUID=926469C0-A7BA-47D3-BB32-05C2C6D8EB2B
[35] Ramos, Gretchen A. and Darren Abernethy. "Additional U.S. State Advance the State Privacy Legislation Trend in 2020." The National Law Review. January 27,2020. https://www.natlawreview.com/article/additional-us-states-advance-state-privacy-legislation-trend-2020
[36] ARM and Northstar. "AI today, AI tomorrow: Awareness, acceptance, and anticipation of AI: A global consumer perspective." 2017. http://pages.arm.com/rs/312-SAX-488/images/arm-ai-survey-report.pdf
[37] GENESYS. "70% of U.S. Employees Hold Positive View of Artificial Intelligence in the Workplace Today. " July 10, 2019. Accessed February 8 2020.
https://www.prnewswire.com/news-releases/70-of-us-employees-hold-positive-view-of-artificial-intelligence-in-the-workplace-today-300882125.html

of 3297 adults in the U.S., 76% participants agree or strongly agree that AI will change the ways people work and live in the next 10 years, and 77% of them are mostly positive or very positive about the impact that AI will bring (Northeastern University and Gallup 2018).[38]

But, as mentioned previously, the presumption of "AI's decision is definitely better than human decision" is not true, and AI's decision is basically equal or worse to Human. So, it's naturally to ask why Human would choose AI's decision instead of Human decision, why Human has more beliefs in AI than Human, and where are these positive beliefs in AI coming from? In short, how did Human form these positive beliefs in AI?

We always have positive impressions attached to the term "Technology", and also attribute it with great characters, like efficiency, accuracy, convenience, cost saving etc. Our beliefs of these characters belong to the "Technology" are gradually forming from our experiences, i.e. our past interactions with these Technological products. Thus, we always have similar expectations of those Technological Inventions, as long as they are in relation to Technology. And AI is presumed by Human to be one of those general Technological Inventions.

But, AI and the "general Technology" are heterogeneous. As Artificial Intelligence is deemed to play the leading role in "The Fourth Industrial Revolution", the term was introduced by the Founder and the Executive Chairman of World Economic Forum K. Schwab (Schwab 2015)[39] this should be the signal that AI is naturally different from the ordinary technology that we used to know. When refer to the decision-making, from the epistemological perspectives, the genuine difference between AI and general Technology is the certainty of the judging criteria that are used for decision-making process.

AI is expected to learn and generate the mean standards from training data instantly and apply them promptly, but the general Technology is designed to follow a specific standard and apply it afterwards. The new data input will definitely influence the mean standard of the training dataset, and consequently affect the judging criteria for AI decision-making. The fluctuation of mean standards causes AI's result not only unpredictable, but also makes AI's decision lack of stability and even possible inconsistency. For this reason, the epistemic certainty of the AI's judging criteria and decision results differs from the general Technology. Thus, it's by no means that we could categorize AI as a "general technology" from this viewpoint. But nonetheless, we still give AI all the credit and beliefs of the Technology as usual, like AI is a kind of general or ordinary technology.

A belief needs to be justified to become a truth that is worth to believe, and knowing is the justification of a belief. Surely, a true belief without justification can be acquired accidentally, i.e. by luck, but that's not the way we Humans would expect, because the "luck" is unreliable. Thus, to know before to believe, is the fundamental principle for us to see the World outside of us. Even if "to know" can have a different meaning or different degree, but the "degree of knowing" should be as equal as possible to the "degree of belief".

[38] Northeastern University and Gallup Inc.. "Optimism and Anxiety: Views on the Impact of Artificial Intelligence and Higher Education's Response." October22, 2018. https://perma.cc/57NW-XCQN

[39] Schwab, Klaus. "The Fourth Industrial Revolution: What It Means and How to Respond. Foreign Affairs." Foreign Affairs. December 12 2015. Accessed February 8, 2020. https://www.foreignaffairs.com/articles/2015-12-12/fourth-industrial-revolution

But it seems like we humans don't know what AI is yet. Pegasystem conducted a survey of 6000 adults in North America, APAC and EMEA, 70% of participants believe they understand AI, but 50% of them don't understand AI can solve problems, 37% of them don't understand AI can interpret speech, and 35% of them don't understand AI can mimic humans … etc. (Pegasystem 2017).[40] And according to Entrata's online survey result from 1051 U.S. participants, over 38% just heard of AI or have no idea what it is, but 52% of participants are comfortable interacting with it (Entrata 2019).[41] As Bristows's surveyed in U.K. shows that in 2103 participants, 25.5% of all never heard of AI or have heard the term but unsure what it is, and 39.5% participants says has limited knowledge of it (Bristow 2018).[42] The underlying problem of this is, as I. Evans highlighted that "Nobody agrees on what AI is" (Evans 2019)[43] and explained in Elsevier' report "The AI field has multiple definitions, but lacks a universally agreed understanding. AI means different things to different people: there are more differences than commonalities… ", (Elsevier 2019)[44] there is no one definition of AI that can achieve human consensus, i.e. none of these definitions can explain AI' characters, abilities, functions, and potentials in the complete, clear and distinct ways.

With no definitions or explanations of AI can satisfy with these basic criterion of understanding, but only based on what we have known by now, Humans have given more beliefs to AI than it should have. Our beliefs in AI shouldn't be the same as the general technology that we used to know, even if AI is included in the broadest definition of Technology. AI should be considered separately from the general or ordinary Technology.

Rethinking GDPR Article 22: The Possible Solutions

In the GDPR Art. 22, it requires human intervention as the safeguards to prevent the harms that the solely automated decision-making could cause. But humans prefer to choose AI's decisions or results instead of Humans decisions, due to we simplified the nature of AI as a kind of general technology. Following from this epistemological conclusion, I would like to propose three possible solutions to resolve this conflict between the regulation and the reality.

[40] Pegasystem. "What Consumers Really Think About AI: A global Study." June 19, 2017. https://www.pega.com/insights/resources/what-consumers-really-think-ai-infographic

[41] Entrata. "Artificial Intelligence and Apartment Living: Survey Studies Consumer's Knowledge of and Attitude Toward AI (Report)" and " What Consumers Really Think About AI (poster)". August 2019. P.18. http://info.entrata.com/newsletters/case_studies/AI/SurveySummary.pdf

[42] Bristows. Artificial Intelligence: Public Perception, Attitude and Trust. 2018. P.7. https://d1pvkxkakgv4jo.cloudfront.net/app/uploads/2019/06/11090555/Artificial-Intelligence-Public-Perception-Attitude-and-Trust.pdf

[43] Evans, Ian. ""Nobody agrees on what AI is"- How Elsevier's report used AI to define the undefinable." Elsevier. January 18, 2019. Accessed February 8, 2019. https://www.elsevier.com/connect/nobody-agrees-on-what-ai-is-how-elseviers-report-used-ai-to-define-the-undefinable

[44] Elsevier. "Artificial Intelligence: How knowledge is created, transferred, and used." January 2019. https://www.elsevier.com/research-intelligence/resource-library/ai-report?utm_source=AI-EC

The first is to raise public awareness of AI. Since no sufficient knowledge of AI and misunderstand AI's nature are the reasons why humans prefer to follow or obey AI's decision, as I explained in the previous section, then, to clarify AI's characters to the general public should be the first thing to do. The Government has more resources and more power than any other private sector, thus should be the one who carries this responsibility to raise public awareness of AI.

The second is the mechanism of the third party certification for AI's neutrality. The rationale behind this mechanism is as follows: if an organization or a company designs AI, they may not be able to verify the result objectively; and if they can't verify the results objectively, then the non-neutral results might affect the decisions of their employees, who prefer to choose AI's decision as described previously; Thus it's important to keep AI as neutral as possible, and maybe the mechanism of the third party inspection and certification could help. This mechanism is like those existing third party inspection methods, but its purpose is for examining AI's neutrality. This third party inspection agency could be a government agency or a private company, as long as it has the Governments' license. This mechanism should be applied before the AI's actual application and periodic inspection afterward.

Last but not least, it's necessary for us to clarify what do Humans really want from AI, and the reasons for AI's developments should be more than efficiency and resource saving. Everyone expects AI can improve our lives dramatically, and everybody talks about the benefits that AI could bring to the Humans. In the meantime, Humans shouldn't forget how little we know about the AI, and those unforeseen or unknown consequences due to the limited knowledge of AI we have. While we devote ourselves to improving AI's applications, we should enquire the meaning of this dedication as well.

Further Discussion: The Dichotomy of the Layperson and the Expert

Maybe, the optimistic attitudes toward AI or the willingness to embrace AI's decision don't happen to everyone? One important topic also studied by Logg et al. is the different attitudes between the layperson and the expert, when they're provided with algorithmic advice at the decision-making moment. Logg et al. invited experts to predict the events in relation to their expertise, and the research results showed that experts "… adherence to their prior judgments and their failure to utilize the information offered to them ultimately lowered their accuracy relative to the lay sample"[45], Logg et al. further pointed out that this could be used to explain P.E. Meehl's theory that "why pilots, doctors, and other experts are resistant to algorithmic advice."[46]. According to this research result, the experienced experts seem to refuse AI's decision while making decisions in relation to their professions, and thus don't even consider AI's decisions. In short, the experts seem immune to AI's influence in their professions.

Whether there is a dichotomy of the layperson and the expert concerning the influence of AI's decision is crucial in two ways. First, this claim seems intuitive, for instance,

[45] Logg et al. 2019, p.99.
[46] Logg et al. 2019, p.99.

an experienced driver knows the routes and the traffic situations at a certain timing in general, therefore the driver won't need the GPS's advice. Second, if there is clear evidence can support the claim that the experts are immune to AI's decisions, then, there is no need to worry for those experts will be lead by AI or won't challenge to AI's decision in their professions, while they are making those professional decisions that are seriously impactful to the data subjects.

But interestingly, we can see the research results that include both inclinations, i.e. does and doesn't support the claim that "expert immune from AI's decision"; and when given similar conditions and the same professions, this contradictory is even clear. For example, M. Stevenson[47] analyzed the criminal court's data in Kentucky U.S. to see the outcomes after the State has mandated the use of the pretrial risk assessment algorithm. According to Stevenson's results, the judges did use the risk assessment as the State required. And from the release rate perspective, although it increases low-risk and moderate-risk rate in 22% and 16% of non-financial release, it also caused the non-financial bond change to released on the low cash bond, and Stevenson pointed out "thus, the net effects on the release rate were attenuated"(Stevenson 2018)[48]. As for the changes in total release rate, this mandatory increased low and moderate risk defendants rate in 9% and 7%, but also decreased 4% in high-risk defendants release rate, and Stevenson concluded that "in total, this resulted in a 4 percentage point increase in the release rate for all defendants, which eroded over time as judges returned to their previous bail setting habits"(Stevenson 2018)[49]. According to Stevenson's research result, even provided with the algorithmic advice, judges eventually use their experience as the guidance for decision-making.

On the other hand, B. Cowgill's research suggested a different perspective in the same professions (Cowgill 2018).[50] To research the issue of judicial compliance of algorithmic risk assessments, Cowgill analyzed the data from the criminal court of Broward County Florida U.S. As Cowgill's research results, the algorithmic advice increase the pre-trial detention for one week in average; and for the overall day in jail, the low/medium risk of general recidivism increase two weeks additional detention, and it's even double up for the violent recidivism (Cowgill 2018)[51]. According to the research results, Cowgill pointed out "the algorithmic guidance does affect pretrial bail decisions"(Cowgill 2018)[52], and further indicated that, in summary, "this result suggested that algorithmic suggestion have a causal impact on criminal proceedings and recidivism"(Cowgill 2018)[53]. As Cowgill's research result, algorithmic advice is the guidance for judges' decision-making process.

By these two research results that have contrary conclusions but are in the same profession, the experts in their professions will resist and ignore the AI's decision, or will follow and obey AI's decision, from the Epistemological perspective, I believe we should

[47] Stevenson, Megan T.. "Assessing Risk Assessment in Action." 103 Minnesota Law Review (2018): 303-384. http://dx.doi.org/10.2139/ssrn.3016088

[48] Stevenson 2018, p. 368.

[49] Stevenson 2018, p. 369.

[50] Cowgill, Bo. "The Impact of Algorithms on Judical Discretion: Evidence from Regression Discontinuities". (Working Paper)(December 5, 2018).
http://www.columbia.edu/~bc2656/papers/RecidAlgo.pdf

[51] Cowgill 2018, p. 11-12.

[52] Cowgill 2018, p. 12.

[53] Cowgill 2018, p. 1.

suspend our judgments on this issue temporarily, because we need more information and further researches, before we firmly claim that the experts refuse or decline of AI's decision in their professional decision-making process.

But, suspend our judgments on the "experts immune to AI's decision" issue doesn't mean we should just wait until the decisive results, and then to see if we need to take any action afterward. As the ultimate purpose of the Epistemology is to take the action in accordance with our knowledge, the psychological issue of the control problem, which is introduced by J. Zerilli et al., (Zerilli et al.) [54] should be helpful as the precautions for both novices and experts, before we truly know what AI and AI's influence are.

The "control" refers to the human agent's supervisory functions in the human-machine loop, i.e. "both fault diagnosis and management…as well as planning (Zerilli et al. 2019)".[55] Zerilli et al. point out the control problem is caused by "the human agent within a human-machine control loop to become complacent, over-reliant or unduly diffident when faced with the output of a reliable autonomous system", and most importantly, this control problem "… somewhat alarmingly, it seems to afflict experts as much as novices… " (Zerilli et al. 2019).[56] Zerilli et al. analyzed a few reasons that could cause the control problem, e.g. human lack of technical ability or physical limitation to supervise, and one of the reasons is the psychological attitude.

In contrast to the human positive believes in AI is due to human epistemologically misidentified the nature of AI, as I mentioned in previous sections, this psychological attitude of Human is caused by machine's accurate performances. According to J. Zerilli et al., these psychological attitudes of human operator only occur when the automation system is highly reliable, which cause human over-trust the machine and thus change human supervisory behaviors. The highly reliable machine means its less error performance: on the one hand, it makes the human operator become complacency for machine's result and thus won't actively supervise the machine's operation, i.e. the automation complacency; on the other hand, it also makes the human operator inclines to ignore every other information, even include their own senses, i.e. the automation bias [57] (Zerilli et al. 2019). The behavior consequence of these two psychological attitudes should be the alerts while Human interacts with the AI.

As mentioned above, I believe experiences in certain aspects could cause the Human doesn't take the AI's decision for consideration, and thus won't cause any problem as this paper previous mentioned, when providing with AI's decision. But whether this claim is also applicable to the experts when they are facing specific professional decision-making and provide with AI's decision, e.g. when judges to decide pretrial detention, when physicians to diagnose disease and provide medical treatment, when police officer to distribute the police force in certain area…etc., I think we might need more information to determine. And before we truly know what AI is, we should always keep ourselves actively involve and look out all the information when we interact with AI, both the layperson and the expert.

[54] Zerilli, John, Alistair Knott, James Maciaurin, and Colin Gavaghan. "Algorithmic Decision-Making and the Control Problem". Minds and Machines 29(December 2019): 555-578. https://doi.org/10.1007/s11023-019-09513-7

[55] Zerilli et al. 2019 p.559.

[56] Zerilli et al. 2019 p.556.

[57] Zerilli et al. 2019 p.561.

Conclusion

Epistemologically speaking, Artificial Intelligence and the general Technology are heterogeneous for AI's judging criteria lack of the epistemic certainty. While the definitions of AI are still opaque, Humans are paying more attention to the possible advantages than the possible harms that AI could cause. Human in the loop can be an ideal solution to the solely automated decision-making process as the GDPR Art.22 requests, but if we can't recognize the differences between AI and the general technology correctly, human intervention won't work as it meant to be. When we dedicate ourselves to developing and improving the AI, we should ask ourselves as well: What do we Humans really want from the AI?

Reference

ARM and Northstar. "AI today, AI tomorrow: Awareness, acceptance, and anticipation of AI: A global consumer perspective." 2017. http://pages.arm.com/rs/312-SAX-488/images/arm-ai-survey-report.pdf

Article 29 Data Protection Working Party. Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 (WP251rev.01). 2018.

Babuta, Alexander. "Innocent Until Predicted Guilty? Artificial Intelligence and Police Decision-Making." *RUSI Newsbrief* Vol. 38, No. 2(March 2018). https://rusi.org/sites/default/files/ 20180329_rusi_newsbrief_vol.38_no.2_babuta_web.pdf

Board of Supervisors, City and County of San Francisco. Administrative Code-Acquisition of Surveillance Ordinance. Accessed February 9 2020. https://sfgov.legistar.com/Legislation Detail.aspx?ID=3953862&GUID=926469C0-A7BA-47D3-BB32-05C2C6D8EB2B

Bristows. Artificial Intelligence: Public Perception, Attitude and Trust. 2018. P.7. https://d1pvkxkakgv4jocloudfront.net/app/uploads/2019/06/11090555/Artificial-Intelligence-Public-Perception-Attitude-and-Trust.pdf

"Causal Explanation as a Partial Solution to Algorithmic Harms." (paper presented at The 7th Academia Sinica Conference on Law, Science and Technology: Emerging Legal Issues for Artificial Intelligence: Legal Liability, Discrimination, Intellectual Property Rights and Beyond, Taipei, Taiwan, November 26-27, 2018.

Cowgill, Bo. "The Impact of Algorithms on Judical Discretion: Evidence from Regression Discontinuities". (Working Paper)(December 5, 2018). http://www.columbia.edu/~bc2656/papers/Recid Algo.pdf

Dastin, Jeffrey. "Amazon scraps secret AI recruiting tool that showed bias against women." *Reuters*, October 10, 2018. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

Dietvorst, Berkeley J., Joseph P. Simmons, and Cade Massey. "Algorithm Aversion: People Erroneously Avoid Algorithms after Seeing Them Err." *Journal of Experimental Psychology*: General, Vol. 144, Issue 1 (February 2015): 114-126. https://doi.org/10.1037/xge0000033

Elsevier. "Artificial Intelligence: How knowledge is created, transferred, and used." January 2019. https://www.elsevier.com/research-intelligence/resource-library/ai-report?utm_source=AI-EC

Entrata. "Artificial Intelligence and Apartment Living: Survey Studies Consumer's Knowledge of and Attitude Toward AI (Report)" and " What Consumers Really Think About AI (post)". August 2019. http://info.entrata.com/newsletters/case_studies/AI/SurveySummary.pdf

Evans, Ian. "Nobody agrees on what AI is"- How Elsevier's report used AI to define the undefinable." *Elsevier*. January 18, 2019. Accessed February 8, 2019. https://www.elsevier.com/connect/nobody-agrees-on-what-ai-is-how-elseviers-report-used-ai-to-define-the-undefinable

GENESYS. "70% of U.S. Employees Hold Positive View of Artificial Intelligence in the Workplace Today. " July 10, 2019. Accessed February 8 2020. https://www.prnewswire.com/news-releases/70-

of-us-employees-hold-positive-view-of-artificial-intelligence-in-the-workplace-today-300882125.html

Gettier, Edmund L. "Is Justified True Belief Knowledge?" *Analysis*, Vol. 23, Issue 6 (June 1963): 121–123. https://doi.org/10.1093/analys/23.6.121

IBM. "IBM to release world's largest annotation dataset for studying bias in facial analysis." Accessed February 8, 2020. https://www.ibm.com/blogs/research/2018/06/ai-facial-analytics/

Lai, Vivian & Chenhao Tan. "On Human Predictions with Explanations and Predictions of Machine Learning Models: A Case Study on Deception Detection." In FAT*'19: Proceedings of the Conference on Fairness, Accountability, and Transparency, 29-38. United States, New York: Association for Computing Machinery, 2019. https://doi.org/10.1145/3287560.3287590.

Logg, J. M., J. A. Minson and D. A. Moore. "Algorithm appreciation: people prefer algorithmic to human judgment." *Organizational Behavior and Human Decision Processes*, Vol.: 151 (February 5, 2019): 90-103. https://doi.org/10.1016/j.obhdp.2018.12.005

Marcus, Gary. "Deep Learning: A Critical Appraisal." *ArXiv* abs/1801.00631 (January 2018).

National Institute of Standards and Technology, U.S. Department of Commerce. "Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects." (December 19, 2019) P.7. https://www.nist.gov/programs-projects/face-recognition-vendor-test-frvt-ongoing

Niiler, Eric. "Can AI Be a Fair Judge in Court? Estonia Thinks So." *WIRED*. March 25, 2019. https://www.wired.com/story/can-ai-be-fair-judge-court-estonia-thinks-so/

Northeastern University and Gallup Inc.. "Optimism and Anxiety: Views on the Impact of Artificial Intelligence and Higher Education's Response." October22, 2018. https://perma.cc/57NW-XCQN

Pegasystem."What Consumers Really Think About AI: A global Study." June 19, 2017. https://www.pega.com/insights/resources/what-consumers-really-think-ai-infographic

Ramos, Gretchen A. and Darren Abernethy. "Additional U.S. State Advance the State Privacy Legislation Trend in 2020." *The National Law Review*. January 27,2020. https://www.natlawreview.com/article/additional-us-states-advance-state-privacy-legislation-trend-2020

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). https://eur-lex.europa.eu/eli/reg/2016/679/oj

Schwab, Klaus. "The Fourth Industrial Revolution: What It Means and How to Respond. Foreign Affairs." *Foreign Affairs*. December 12 2015. Accessed February 8, 2020. https://www.foreignaffairs.com/articles/2015-12-12/fourth-industrial-revolution

Simonite, Tom. "Machines Taught by Photos Learn a Sexist View of Women." *WIRED*, August 21, 2017. https://www.wired.com/story/machines-taught-by-photos-learn-a-sexist-view-of-women/

Stevenson, Megan T.. "Assessing Risk Assessment in Action." *103 Minnesota Law Review* (2018): 303-384. http://dx.doi.org/10.2139/ssrn.3016088

U.S. Food & Drug Administration. "Software as a Medical Device (SaMD)." Accessed February 8, 2020. https://www.fda.gov/medical-devices/digital-health/software-medical-device-samd

Vaccaro, Michelle and Jim Waldo. "The Effects of Mixing Machine Learning and Human Judgment." *Communications of the ACM* Vol. 62, No.11 (October, 2019): 104 -110. https://doi.org/10.1145/3359338.

Zerilli, John, Alistair Knott, James Maciaurin, and Colin Gavaghan. "Algorithmic Decision-Making and the Control Problem". *Minds and Machines* 29 (December 2019): 555-578. https://doi.org/10.1007/s11023-019-09513-7

ZHIMA Credit, Ant Financial Services Group. Accessed February 8, 2020. https://www.xin.xin/#/home