frontiers in Psychiatry

CONCEPTUAL ANALYSIS
published: 22 January 2021
doi: 10.3389/fpsyt.2020.622506

Psychiatric Advance Directives and Artificial Intelligence: A Conceptual Framework for Theoretical and Ethical Principles

Stéphane Mouchabac[1,2†], Vladimir Adrien[1,2†], Clara Falala-Séchet[3], Olivier Bonnot[4,5], Redwan Maatoug[2,6], Bruno Millet[2,6], Charles-Siegfried Peretti[1], Alexis Bourla[1,7] and Florian Ferreri[1,2]*

[1] Sorbonne Université, AP-HP Department of Psychiatry, Hôpital Saint-Antoine, Paris, France, [2] Sorbonne Université, iCRIN Psychiatry (Infrastructure of Clinical Research In Neurosciences - Psychiatry), Brain and Spine Institute (ICM), INSERM, CNRS, Paris, France, [3] Laboratory of Psychopathology and Health Processes, EA 4057, Institute of Psychology, University of Paris, Paris, France, [4] CHU de Nantes, Department of Child and Adolescent Psychiatry, Nantes, France, [5] Pays de la Loire Psychology Laboratory, EA 4638, Nantes, France, [6] Sorbonne Université, AP-HP Department of Psychiatry, Hôpital Pitié-Salpêtrière, Paris, France, [7] Jeanne d'Arc Hospital, INICEA Group, Saint-Mandé, France

OPEN ACCESS

Edited by:
Hector Wing Hong Tsang, Hong Kong Polytechnic University, Hong Kong

Reviewed by:
Giacomo Deste, Civil Hospital of Brescia, Italy
Devashish Konar, Mental Health Care Centre, India

***Correspondence:**
Stéphane Mouchabac
stephane.mouchabac@aphp.fr

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Social Psychiatry and Psychiatric Rehabilitation, a section of the journal Frontiers in Psychiatry

Received: *28 October 2020*
Accepted: *16 December 2020*
Published: *22 January 2021*

Citation:
Mouchabac S, Adrien V, Falala-Séchet C, Bonnot O, Maatoug R, Millet B, Peretti C-S, Bourla A and Ferreri F (2021) Psychiatric Advance Directives and Artificial Intelligence: A Conceptual Framework for Theoretical and Ethical Principles. Front. Psychiatry 11:622506. doi: 10.3389/fpsyt.2020.622506

The patient's decision-making abilities are often altered in psychiatric disorders. The legal framework of psychiatric advance directives (PADs) has been made to provide care to patients in these situations while respecting their free and informed consent. The implementation of artificial intelligence (AI) within Clinical Decision Support Systems (CDSS) may result in improvements for complex decisions that are often made in situations covered by PADs. Still, it raises theoretical and ethical issues this paper aims to address. First, it goes through every level of possible intervention of AI in the PAD drafting process, beginning with what data sources it could access and if its data processing competencies should be limited, then treating of the opportune moments it should be used and its place in the contractual relationship between each party (patient, caregivers, and trusted person). Second, it focuses on ethical principles and how these principles, whether they are medical principles (autonomy, beneficence, non-maleficence, justice) applied to AI or AI principles (loyalty and vigilance) applied to medicine, should be taken into account in the future of the PAD drafting process. Some general guidelines are proposed in conclusion: AI must remain a decision support system as a partner of each party of the PAD contract; patients should be able to choose a personalized type of AI intervention or no AI intervention at all; they should stay informed, i.e., understand the functioning and relevance of AI thanks to educational programs; finally, a committee should be created for ensuring the principle of vigilance by auditing these new tools in terms of successes, failures, security, and relevance.

Keywords: psychiatric advance directives, artificial intelligence, medical ethics, joint crisis plan, clinical decision support system, predictive medicine

1. INTRODUCTION

1.1. Psychiatric Advance Directives (PADs)

Psychiatric disorders are often characterized by a high rate of relapse, during which the patient's decision-making abilities are altered and may result in psychiatric admissions, often involuntarily. Laws usually indicate that patients cannot be treated without their consent. But there is, most often, a legal framework to provide care or hospitalize patients who can no longer give free and informed consent. These patient care modalities are often deemed to be very restrictive, disempowering, and stigmatizing. Mostly, they predicted poorer recovery and more suicidal ideation after 2 years, mediated by decreased empowerment after 1 year. Finally, they do not promote a proper therapeutic relationship between patient and physician (1–3). Furthermore, compulsory admission is often used inappropriately to manage aggressive behavior rather than psychiatric diseases (4). In addition, routine crisis treatment guidelines are developed without patient involvement and based upon standard recommendations that are not personalized.

In most countries, patients have the right to give an informed advance treatment framework [advance directives (ADs)], which allows anticipating their requests for future care and providing information about drug treatments, nonmedical instructions, and the person authorized to make decisions for them (5). For example, in France, the L1111-11 article of the French Public Health Code (FPHC) and in United Kingdom the Mental Capacity Act 2005, amended by the Mental Health Act 2007, frame this practice. On a wider scope, The United Nations Convention on the Rights of Persons with Disabilities encourages the use of strategies that promote patient decision-making autonomy.

Foreseeing such situations is considered as the hallmarks of good clinical practice, recognized as an important supported decision-making tool and it also enables the patient to ensure that the medical decision is most consistent with his own interests. To be valid, the patient must have the mental capacity to write by hand the ADs. "Mental capacity" is defined as the ability to make a decision, understand, and analyze information related to one's care and know the existing alternative options (6).

In a recent review (7), authors clustered PADs into four types depending of their content, the way they are drawn up, and the level of legal authority:

- Classic PADs are formalized by the patient without any caregivers interventions and describes personal values and treatment preferences. They give informed consent to therapeutic interventions (which are accepted or rejected) and name the proxy for decisions during a relapse or crisis.
- Simplified PADs are a form of directive in which trained caregivers helps the user to create the final document, which would increase its quality. Among the methods available, the use of a semi-structured interview allows to select the preferences for future treatments based on a set of available information.
- Cognitive Therapy-based Advance Directives are written with a staff member, taking events from previous crises and proposing alternatives for future episodes (8). Finally, there is a collaborative approach between the patient and the caregiver, in which disagreements and differences of opinion are respected and recognized.
- The Joint Crisis Plan involves the patient and the care team in a negotiation process with a third-party facilitator who may be a mental health worker, a family member, a trusted person, a custodian, or a lawyer, and the quality of the document could be assessed with a "quality of crisis plan" checklist.

In general, the proposal to draft PADs is well-received by patients and health professionals (9, 10). The feeling of having an active participation in the decision-making process (9), the opportunity to record a treatment refusal (10), mutual agreement, and clarification also strengthens the therapeutic alliance and trust between the patient and the care team (11). Patients report a greater perceived sense of control over their care and treatment journey in a potential context of impaired capacity to make appropriate, informed decisions. Indeed, PADs are associated with a benefit such as improving the autonomy of patients and promoting empowerment, which has been defined as "the ongoing capacity of individuals or groups to act on their own behalf to achieve a greater measure of control over their lives" (12). In addition, the drafting of PADs has also shown its relevance in reducing the traumatic and coercive experience of a treatment not chosen by the patient and implemented in an emergency situation. Adherence to care is optimized and there is a decrease in the rate of coercive intervention or hospitalizations in psychiatry compared to consumers without PADs [(13), in a ratio of 50% over a 24-month follow-up period (14). This may be due to both a greater involvement of patients in their care experience and a more detailed understanding of their disorders (9). As expected, PADs may also reduce negative coercive treatment experiences and stigma (15) and is a strong enhancer of therapeutic relationship (16).

This "Advance Statement" also refers to the possibility of identifying prodromal signs of relapse and proposing early personalized interventions. This kind of approach is promising; on the one hand, it allows anticipating the aggravation of the symptoms and the intensity of the relapse. On the other hand, it offers the possibility of choosing treatments by stage more adapted to the patients (16).

Despite this recommendations and level of evidence, some studies have identified barriers to ADs use, which are clustered into health system constraints, health professional practices, and service user representations (17, 18). Predictive medicine has emerged largely in recent decades, including in psychiatry where recent advances in genetics, neuroimaging, and biomarkers are clarifying neurobiological features of mental diseases and could lead to the development of effective personalized medicine (19) and more appropriate ADs. Nevertheless, when we consider the complexity of gene–environment interaction, the use of ADs in psychiatry is complex, even risky. In this field, "to predict" may be understood as the action to announce in advance what should happen by intuition, reasoning, conjecture, or experience. If we retain the scientific aspect of this definition, the possibility of predicting the occurrence of a morbid event opens important perspectives, whether preventive or curative, but also ethical issues.

1.2. Artificial Intelligence (AI) Enhanced Clinical Decision Support Systems (CDSS)

The use of CDSS may be a response to this reluctance and is fundamental for proposing "staged" ADs in function of the intensity of the symptoms. Historically, CDSS belong to three registers: Bayesian probabilistic models, score calculations, and expert systems based on syllogistic algorithms. Interestingly, AI technologies and machine learning methods offer attractive prospects to design and manage crisis response processes in the form of new CDSS. Here, we are alluding to "weak AI," which consists of a device able to learn and correct itself [whereas "strong AI," i.e., autonomous machines with adaptation capacities, is still far beyond reach (20)]. The main purpose of weak AI is the valorization of human skills that are not possessed or that should not be possessed by AI for ethical reasons or the precautionary principle (21). Many initiatives are made in this direction, for example the recent creation of an international observatory of social impacts of AI and digital technologies (https://observatoire-ia.ulaval.ca).

AI technologies could use the user's digital phenotype on the phone's health data captured continuously (e.g., number of steps) or occasionally. The concept of digital phenotyping appears to be very efficient to implement data for these systems. Introduced by Jain et al. (22), it is based on the idea of collecting in real time human behavior data (the momentary ecological assessment or EMA) and markers of their functioning in order to characterize the "digital signature of pathology." The emotions, the energy level, or the presence of symptoms with their perceived intensity (ruminations, hallucinations, and suicidal ideation) can be analyzed. These data can provide useful indicators to identify the increased symptomatology (crisis, manic episode) of many pathologies (bipolar disorder, schizophrenia, major depressive episode, substance abuse) such as logorrhea, increased communicability or reduced social contact, increased behavioral activation, agitation, or psychomotor deceleration (23–30). For instance, CDSS enhanced with AI could make compulsory admissions more efficient to provide appropriate psychiatric care (4).

In addition to these data, the increase in the use of mobile and chatbots applications (providing exchange, therapeutic exercises) creates sources of declarative data on the patient's condition that are particularly interesting to better understand what the person is experiencing in their daily lives, to anticipate relapses, and to better treat such disorders. This collection of subjective data is clearly crucial in medicine and in psychiatry in particular. The notion of contextualization is central in order to personalize follow-up and better understand the appearance of symptoms. Since all these data exceeds the psychiatrist's real-time analysis capabilities, many of the difficulties encountered in consultation (forgetfulness, recall bias, loss of valuable context elements) could be overcome in the data obtained through AI technologies. They could exploit this digital signature by confronting it with important databases that group those of other patients to draw predictive information from them.

Digital phenotype also gives useful first points of contact in the detection of a crisis that can be used, if the patient has given his or her consent, to inform the treating health *in situ* professionals. This provides the opportunity to set up an emergency consultation to deal more effectively with the difficulty when it arises.

Overall, AI technologies offer CDSS tools that are interesting in clinical evaluation by creating relevant algorithms for diagnosis, decision support, relapse prediction, and neuro-prediction. Including the patient's consent, data extraction, and anonymization, this could contribute to a certain therapeutic innovation by creating an ample database: phenotypes (typical symptom profiles, relevant indicators) and patterns (monitored treatments, epidemiological data). For instance, in France, "Health Data Hub" is the initiative following these goals. This combination of tools could be useful and could facilitate PAD completion as peers and caregivers already do (18).

1.3. Issues in Predictive Medicine

Nevertheless, the issue of complex decisions, as is often the case in situations covered by PADs, raises the question of the impact of a nonhuman making decision: such a decision proposed by a nonhuman entity appears to be safer, more rational than that of a human because it is based on a very large amount of data and algorithms with few margins of error. However, the function of these data leads to the question of why and how to use these data: how often should these data be used, how often should these systems be transmitted to the professional? Based on what criteria? From when and what does an indicator provide valid information about the worsening of a disorder? Is there not a greater risk of overreaction or prediction error? Before implementing AI in the PAD drafting process, we must ask ourselves what and where ethical limits should be drawn. This is the goal of a predictive medicine (31).

The purpose of this conceptual framework is to assess two fundamental dimensions of the implementation of AI in the PAD drafting process. First, we will address the issue of the nature of AI (how it functions and interacts with databases) and its place during the process or in the patient–professional relationship. Second, we will be focusing on the ethical principles that this implementation should respect.

2. LEVELS OF AI INTERVENTION IN THE PAD DRAFTING PROCESS

AI can intervene in many ways in the PAD drafting process. First of all, the patient should be able to choose if he/she agrees with AI intervention and if he/she does, then he/she should be able to choose in which way.

2.1. Various Natures of AI

We can describe user or developer's intentional limitation of AI by separating two different AI functionalities:

2.1.1. "WHAT"—In Terms of Access to Data Sources

Data sources are on different graduated levels, from public available sources (Google search, civil status, etc.) to semi-public (social networks, medical data) and finally private (mailbox, web-browser history, etc.). As especially in psychiatry, useful information is often of a private nature, and the right to privacy

should be protected. For PADs, particular attention should be paid to the way in which data are used: explanations to patients of the issues and rights, patients must give their free and informed consent to the use that can be made of their data, what data they wish to have added to a knowledge database or not, and how they can exercise the right to withdraw. Also, it would be appropriate for the professional and the patient to agree on the symptoms to be followed, the key indicators of relapse. The patient can then choose which symptoms will be monitored by the AI, collected, and transmitted to the professional and/or a team of professionals.

2.1.2. "HOW"—In Terms of Different Intelligences (32)

Like conscience (projections of intention, metacognition, etc.), cognition processes of AI can be split, and the patient can be allowed to choose only specific processes, following the recommendation of Villani et al. (21) to preserve certain human skills for humans only. In practice, all these questions are related to the place of AI in care at the time of the crisis. The degree of trust in new technologies of each participant in the PAD drafting process has an impact on the place given to AI. The professional therefore has a major role to play in the way he or she presents technological tools, the patient's understanding of them, the degree of acceptance, and the ability to delegate the decision to a machine. It will be necessary to find the acceptable ratio for everyone between human and nonhuman expertise and ensure that the patient's wishes are respected. This is one of the key points for the drafting of PADs: the possibility offered by the AI, as an expert authority, to ensure that the patient's wishes are respected in terms of type of care.

2.2. Various Places of AI

2.2.1. "WHEN"—At Opportune Moments

- *At the moment of the drafting process*: AI could propose a PAD template based on deep learning.
- *To optimize PADs in real-time*: ADs were developed because current care guidelines are either over or under inclusive. AI makes it possible to optimize ADs in real time through an incrementation process. ADs propose a decision that will prevail against a future will. But patients' preference may change overtime, and there is always a shift between their preference at the time t and their preference at the previous time $t - dt$ of the AD: the preference can be seen as a multivariate function of sociodemographic data, environmental factors, and time (**Figure 1**). For example, if the environmental circumstances change, it is possible for the directives to be modified and adapted to the new context. This kind of microdirectives could be extracted from large databases, but personalized too: from an incremental point of view, these microdirectives could be enhanced with the others patient's experiences after an algorithmic treatment

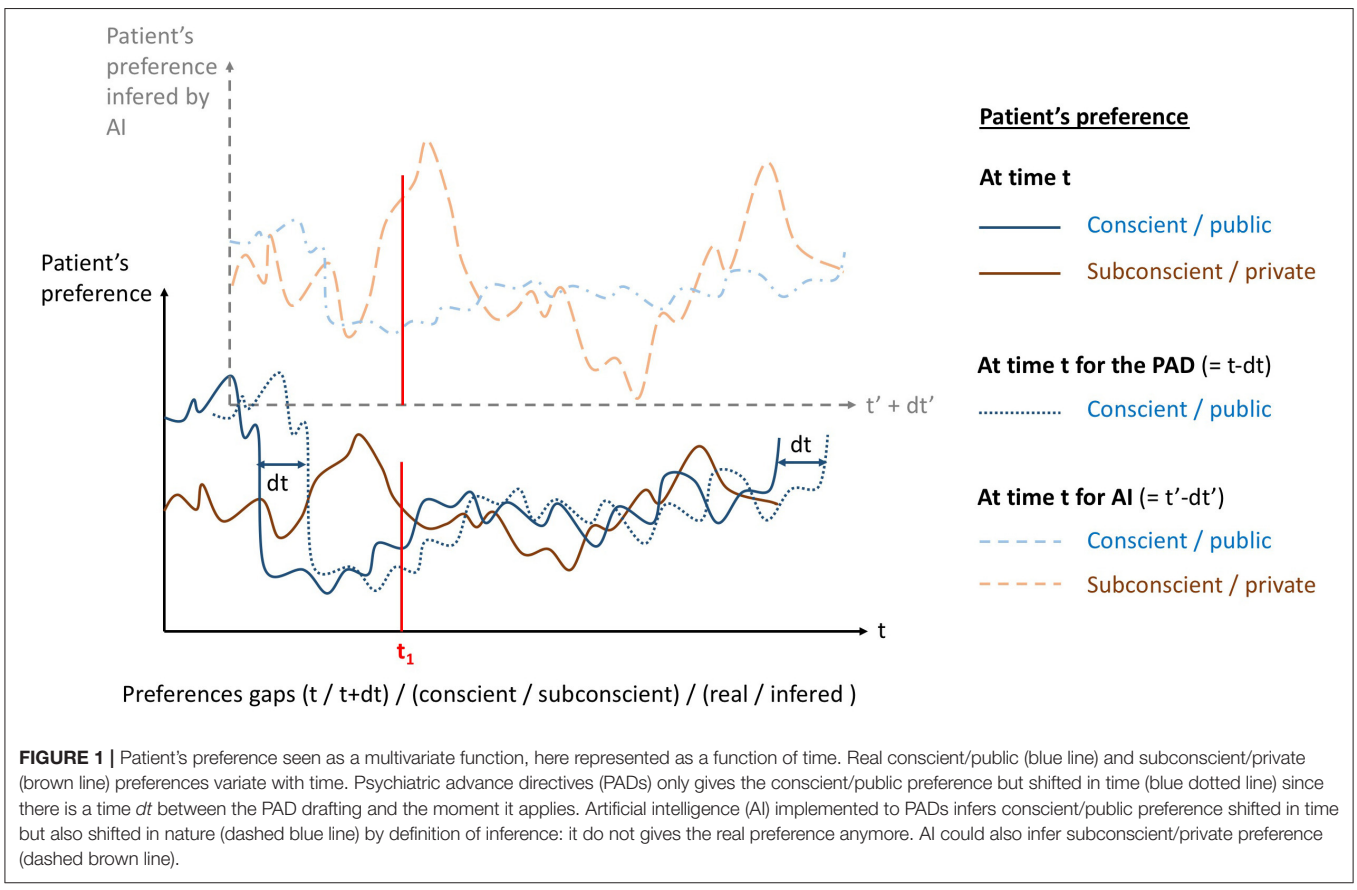

FIGURE 1 | Patient's preference seen as a multivariate function, here represented as a function of time. Real conscient/public (blue line) and subconscient/private (brown line) preferences variate with time. Psychiatric advance directives (PADs) only gives the conscient/public preference but shifted in time (blue dotted line) since there is a time *dt* between the PAD drafting and the moment it applies. Artificial intelligence (AI) implemented to PADs infers conscient/public preference shifted in time but also shifted in nature (dashed blue line) by definition of inference: it do not gives the real preference anymore. AI could also infer subconscient/private preference (dashed brown line).

with AI, so new directives benefit from past directives. This feedback permits the selection of a new panel of possible and efficient directives. The use of AI will create a supplementary shift, between the real preference and the inferred preference, independently of the time factor. Furthermore, ADs express conscious or public preference, whereas AI could access the unconscious or private preference, raising the ethical issue of which preference is the most beneficial for the patient (**Figure 1**). The incrementation process and feedback offer a large field of directives for the constitution of the ADs that could act as "validated alarms" for the different actors implicated in ADs. This principle is already used in law (33). Still, regarding PADs, the issue is made more complex because the clinician must decide whether patients are currently able to express their preference or not: the preference inferred by AI (different of the real preference) could turn to influence this clinical decision and induce an error that could self-drive itself, the patient real preference ending with being rejected permanently.

- *At the moment of a "difficult" medical decision*: The relevance of AI is in particular to help the professional in the event of a difficult decision (34), with complex, contradictory data, and evolution. The use of AI, coupled with ADs—a device also recognized as helping in difficult medical decision-making (35)—would facilitate the actions of professionals in accordance with the patient's request. By difficult decision, we mean in particular the case of a gap between what the patient would have liked (as noted in his/her AD) and therapeutic options, considering the ongoing situation. This can occur in this kind of situation: new information on the patient's situation (whether or not detected by AI), conflict of interest between the clinical benefit of a therapeutic modality and the patient's choices, and new therapeutic modalities available not provided for in advance patient instructions. Thus, AI should be able, if there is a PAD, to decide if we are in the situation for which the PAD can be applied, to modify the PAD if "necessary," to add elements on patient preference when dealing with a situation not foreseen by the PAD. If there is no PAD, AI should be able to add elements on patient's preference when dealing with a medical situation.

2.2.2. "WHO"—In the Contractual Relationship

In psychiatry, the family is most of the time involved in clinical decisions, and thus we can consider PADs as a third-party contract already in a tripartite relationship. When drafting PADs, it is relevant to discuss the place given to the AI, its degree of participation in the CDSS, what place the AI takes in relation to a trusted person and/or the professional if an important decision must be made (for example, stopping, maintaining, and changing treatment), and finally who makes the final decision. AI could therefore act (**Figure 2**).

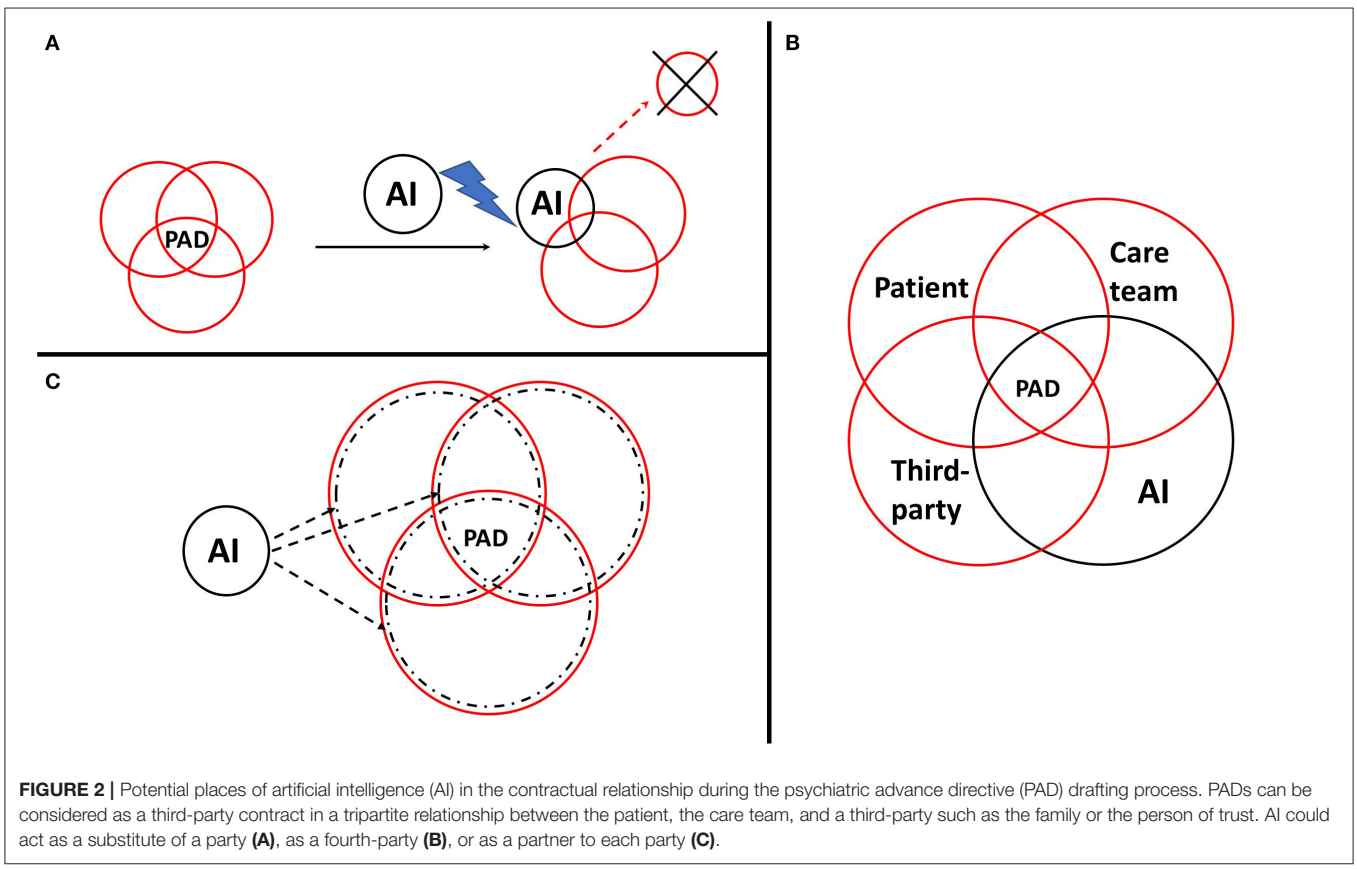

FIGURE 2 | Potential places of artificial intelligence (AI) in the contractual relationship during the psychiatric advance directive (PAD) drafting process. PADs can be considered as a third-party contract in a tripartite relationship between the patient, the care team, and a third-party such as the family or the person of trust. AI could act as a substitute of a party **(A)**, as a fourth-party **(B)**, or as a partner to each party **(C)**.

- *As a substitute of a party of the contract*: This option is not ethical as will be seen later.
- *As a fourth party*: The option of treating AI as a "party" raises the issue of creating a juridical personality status for AI, currently being debated at the European Parliament under the concept of "electronic personality." It could impose a responsibility for AI (inducing the creation of insurance funds by developers or users) to ensure potential victims for damages "attributable" to AI. Nonetheless, the risk is to disempower AI users (in our case the patient, the medical doctor or the third party) but also developers. In addition, applying human rights like autonomy or citizenship to an "electronic personality" raises ethical issues.
- *As a partner to each party*: An entity infiltrating each of the three parties of the contract without being a juridical personality in itself: the idea is a partnership between AI and each party (DSS for the patient or the third party, and CDSS for the clinician).

Although AIs today have no intention to harm, we can legitimately ask ourselves what will happen in several decades' time. A nonhuman decision may therefore seem very relevant at the individual level and not at all at the level of a human group and vice versa. For example, the choice not to treat a patient according to different parameters (age, symptoms, cost of treatment, rarity of treatment) may be relevant at the level of a group (significant financial loss) but may not go in the direction of maximum preservation of a patient's life span and autonomy.

It is possible that the presence of AI does not promote respect for the patient's choice but could be for AI the subject of a trade-off between the value of individual life and the preservation of collective imperatives. These aspects raise countless ethical questions about the interests of the patient and a human group. Computational ethics (36–38) raise the question of the applicability of programming ethical principles within AI. Without an overall reflection on the integration of technological systems such as AIs into the drafting of PADs, which preserve the interest of a particular person, this will result in a potential incompatibility between ADs and the use of AI. We therefore have to go through and examine each ethical principle involved.

Ethics begins with applying good practice recommendations based on values, such as in medicine: autonomy, beneficence, non-maleficence, and justice. The 2017 report (39) of the French National Commission on Informatics and Liberties adds 2 founding principles of AI ethics: loyalty and vigilance.

3. PRINCIPLES OF MEDICAL ETHICS APPLIED TO AI IN THE PAD DRAFTING PROCESS (40)

3.1. Autonomy

The principle of autonomy includes various components of the subject as follows:

- *Free will* (intentionality) was theorized by the philosophy of mind and top-down bottom-up approaches of cognition processes. If there is a disagreement between contract parties, AI could turn in favor of one of the parties (alliance) and endanger the free will of the other.
- *Free action* postulates that the act is not controlled by an external intelligent entity, conscious or artificial. It is subject to cognitive biases, manipulation, or conditioning. Free action is by definition affected by the irruption of AI in the PAD drafting process.
- *Comprehension and adaptation abilities* echo the ethical principles of AI that are loyalty and vigilance (see later). The simplification of algorithms is necessary to get the result in a reasonable amount of time for the caregiving, but also because the user should be able to understand how they work. This simplification implies the risk of loss of information or accuracy. AI cannot be involved in PADs without the patient being able to understand how it works. Hence, there is a need for patient education and specific training of AI teachers, even more specific to the population of psychiatric patients. Naturally, psychoeducation programs should include information on PADs as a prerequisite to understand AI implementation in this process.
- *The principle of dignity* includes respect for autonomy but is a broader concept that implies, in medical care, various requirements:
  - The collection of consent to care (Hippocratic Oath) is the reason why ADs were created for.
  - Valid therapeutic choices are at the cross-section of the clinician experience, the state of the scientific art, and the patient's preference. AI raises the issue of the difficult integration of "tacit" data that are considered by the clinician without him/her knowing it, i.e., heuristics that are difficult to express or objectify. Regarding the state of the art, so much the methodology (41–48) as the supremacy of science (49, 50) can be contested. In medicine, the deep learning possesses the obvious asset of rapidly finding strong correlations over an important quantity of data that cannot be analyzed by the human brain. Nevertheless, besides the high cost of constructing this "big data," it implies significant risks that increase along with the size and efficiency of data collection: loss of oral information (that could contain important medical information) and the issue of causality in data correlations: confounding factors may come into play, hiding a useful correlation into a sum of irrelevant ones, making deep learning far from heuristic learning and reasoning of clinicians. Finally, the concept of "validity" of a therapeutic choice is subjective and depends on the patient's belief. To respect the patient's dignity would then be more about respecting the patient's beliefs, even though it may be in contradiction with evidence-based medicine.
  - *Respect for privacy and medical confidentiality* is no longer the paradigm in the field of healthcare, sharing information between the various caregivers (medical, paramedical, social workers, administrative) being now considered more beneficial for the patient. For example, in France such enlargement of the right to share medical

information has been made possible in 2016 (L1110-4 of the FPHC). Confidentiality is thus gradually being replaced by professional integrity (51–53). With the generalization of electronic health records, caregivers can access the shared medical record of a patient without actually taking care of this patient: their integrity prevents them to access the record. Today, the tendency in psychiatry is to write only factual elements in the patient digital files. When invoking AI, confidentiality can only be partial, since AI skills are acquired mainly through deep learning, which uses anonymized digital records that are never completely anonymized. Indeed, patients stay traceable despite the anonymization: it is easy to "de-anonymize" data by cross-referencing specific data. These data can thus no longer be considered anonymous, which raises the issues of data ownership, control, organization, marketing, right to deletion, and specific uses such as risk assessment by insurances. Another issue is the potential lack of data precision or reliability for the determination of a probability *a priori*: the more the AI will be relevant (i.e., the more the inferred preference will be close to patient preference), the more it will have to "know" patients, including their private life. The risk is to know patients "better than themselves" and therefore their preference (conscious or unconscious) better than themselves. The AI could turn to play a role equivalent to that of the person of trust with whom private information is shared (L.1111-6 of the FPHC). Finally, if data are collected by AI, it is subject to risks such as hacking, breakdowns, and data protection system would have to be strengthened.

3.2. Beneficience

Concerning AI, psychiatrists are very careful on this principle (54). As it contains two elements:

- *The benefit* can be seen as an improvement of "wellness." AI could turn to subjectively make wellness like in other societal fields where wellness becomes, through positive thinking, a moral imperative ultimately leading to malaise (55). This effect could increase within the specific population of patients suffering from psychiatric disorders. Thus, there is a need to find more objective and specific indicators to measure the benefit of a treatment. In the case of PADs, these could be the degree of adequacy *a posteriori* between the patient and the treatment received during a state of crisis.
- *The benefit-to-risk ratio* is also difficult to evaluate in the case of new technologies, risk estimations being often wrong in this field (56). Thus, the principle of objective vigilance has to be applied.

3.3. Non-maleficience

Non-maleficence invokes the idea of malaise. AI could be maleficent at every level in the PADs process. An inventory of all possible failures will be necessary, this inventory being updated along with the use of the AI, respecting the principle of objective vigilance.

3.4. Justice

Justice implies equality of care for all patients (non-discrimination), with specific adaptation to each individuality (positive discrimination) due to incompressible inequalities (genetic, developmental, or environmental). The major risk is that the specific care could get to be in function of the societal participation of the patient, marginalizing even more psychiatric patients, already undertreated on the somatic level (57–62). Encouraging clinicians to have a subjective approach also goes against harmonization of practices.

4. PRINCIPLES OF AI ETHICS APPLIED TO MEDICINE IN THE PAD DRAFTING PROCESS

4.1. Loyalty

The principle of loyalty of AI condenses the laws of robotics proposed by (63) and the laws of algorithms (64). The AI must be a partner of the patient, particularly in the PAD drafting process, this being only possible if it respects several principles:

- *Neutrality* means treating all information equally.
- *Diversity* implies not prioritizing the answer.
- *Transparency* implies making the code available to the public. But this remains insufficient if patients are not trained to understand it.
- *Equity* involves not differentiating subjects.
- *Loyalty* means responding to what is asked.
- *Comprehensibility* must be reached by focusing on three axes (21):
  - Simplification of models, with the risks of approximation that entails.
  - Simplification of user interfaces: reduction of learning biases.
  - Education on cognitive mechanisms at work, their strengths and weaknesses, and in particular the principle of precaution against the high risk of digital divide with the psychiatric population. Indeed, if users are not formed well enough, the learning bias leads to a misuse of AI systems. Education comprises formation of AI developers, clinicians, patients, families, and the development of tools for the management of personal data and data processing by the patients themselves.

4.2. Vigilance (Also Referred as Reflexivity or Auditability)

Vigilance makes it possible to cover the risks inherent to new technologies, because anticipating all situations is impossible. It is important to have standards as a basic precaution, but they have to stay flexible enough to allow technological innovation. Vigilance must be the responsibility of the state and in practice, it should be implemented through the creation of committees, recruited on referral, litigation or during automatic audits. These committees would have to be open to patients and users of healthcare systems. Vigilance is not just to reliably identify technological failures and propose alternatives to avoid them.

In the absence of failures, it must also be sanitary: global audits should be implemented to assess for the success or the failure of the use of new AI technology on the physical and mental health of patients.

5. CONCLUSION AND SUGGESTED GUIDELINES

This reasoning on the introduction of technological devices in the PAD drafting process confronts caregivers and patients with preconceived ideas about both ADs and AI. Regarding the anticipated guidelines, some professionals are still skeptical about the introduction of a crisis plan (16): doubts about the relevance of this crisis plan, addition of documents to be taken into account, lack of consideration of the crisis plan by the teams, etc. With regard to new technologies, a number of preconceived notions also persist, both on the part of caregivers and patients (54). Concerns about AI include concerns about confidentiality and data storage, particularly when it comes to sensitive data such as health data. In this context, it would be appropriate for the data collected by these intelligent systems to comply with the legislation on the confidentiality of health data (65) as defined in the FPHC. At present, systems, such as those mentioned above, are capable of making reliable predictions based on algorithms and a large database, respecting legislation on confidentiality and the use of health data do not yet exist. In this context, we can propose several recommendations:

5.1. Support: AI Must Remain a Decision Support System, and Seen as a Complement to the Decision, a Partner of the Parties of the PAD Contract

It should always be subject to validation by a professional, whether it is the patient's reference professional or an expert identified and solicited through the use of a telemedicine service. The presence of technological devices such as AI helps to bring new elements (medical data, therapeutic options not considered by health professionals). In order for this input to continue to be relevant to the patient and the professional, a probability system should be in place to weight:

- The different treatment options, possible outcomes and their likely influence on symptoms (and which ones).
- Ecological data (data reported to the professional), their potential evolution, and impact on the patient's quality of life.

5.2. Choice: It Must Be Let to Patients Whether They Wish to Use AI or Not, Which Type of AI, at What Step in the PAD Process

Patients must be able to be systematically informed of the use of an intelligent technological device during their care journey and give their consent for its use and authorization on the different types of data collected. Patients must be able to choose, at each level, whether they consent to the use of AI for data collection: for which data precisely do they consent, and for which use (collection, sharing of information with the professional, CDSS, research, etc.).

5.3. Information: Make AI Understandable

A significant amount of information and education work remains to be accomplished on the functioning and relevance of intelligent systems. It will also be essential to explain the limitations of such tools, including the degree of feedback in the event of a system failure. For this purpose, massive open online courses, workgroups, serious games could be flexible tools. It is primordial to evaluate the level of comprehension, perception, and acceptability of these educational tools, with the use of experimental studies. A final checklist with items verifying the comprehension of the process would also be mandatory. This information should be subject to strengthened provisions in the case of vulnerable persons in order to ensure that free and informed consent is obtained.

5.4. Vigilance: Create a Committee That Will Audit These New Tools in Terms of Successes and Failures, Security, and Relevance

In addition, a set of feedback systems must be provided for in the event of errors (a system of probability or reliability of therapeutic options, detection of errors by the system itself, clinical sense) integrating targeted and random control systems on the different functionalities offered by the AI. In order to allow for the integration and optimal use of these systems within healthcare services, it will be necessary to create new legal frameworks for the use and regulation of these systems and the data obtained through these systems. In particular, regulation must consider the level of sensitivity of the health data collected and their impact on medical decisions. In addition, the use of new technologies must respect the rules and ethical principles of caregivers. In fact, it will be necessary to support health professionals in the use of new technologies that respect these rules inherent to their profession. One of the ways in which these tools could be deployed is to implement them gradually with feedback (evaluation and research on their relevance in healthcare), ethical considerations on new technologies and finally anticipation of new cases of use. The subject of PADs raises more than any other the delicate balance between support for innovation and the necessary ethical regulation. Current issues related to new technologies give rise to important debates on the impact on the maintenance of financial and human resources, the quality of care, the preservation of the human link between caregiver and patient, and respect for patients' rights.

This paper is a reflection by medical professionals on how to employ new information technology tools and techniques for the improvement of the patient's hospital experience. It is fully understood that some suggestions may be in contravention of legal dispositions of some national jurisdictions and that special permission or even legislation may be required to eventually put them into practice in the future.

AUTHOR CONTRIBUTIONS

SM, VA, and CF-S contributed to the conceptual framework and the writing of this work. SM supervised the development of this work. OB, RM, BM, C-SP, and FF contributed to the manuscript evaluation and reviewing. AB and FF contributed to the editing. All authors contributed to the article and approved the submitted version.

REFERENCES

1. Xu Z, Lay B, Oexle N, Drack T, Bleiker M, Lengler S, et al. Involuntary psychiatric hospitalisation, stigma stress and recovery: a 2-year study. *Epidemiol Psychiatr Sci*. (2019) 28:458–65. doi: 10.1017/S2045796018000021
2. Xu Z, Muller M, Lay B, Oexle N, Drack T, Bleiker M, et al. Involuntary hospitalization, stigma stress and suicidality: a longitudinal study. *Soc Psychiatry Psychiatr Epidemiol*. (2018) 53:309–12. doi: 10.1007/s00127-018-1489-y
3. Rusch N, Muller M, Lay B, Corrigan PW, Zahn R, Schonenberger T, et al. Emotional reactions to involuntary psychiatric hospitalization and stigma-related stress among people with mental illness. European archives of psychiatry and clinical neuroscience. (2014) 264:35–43. doi: 10.1007/s00406-013-0412-5
4. Oliva F, Ostacoli L, Versino E, Portigliatti Pomeri A, Furlan PM, Carletto S, et al. Compulsory psychiatric admissions in an italian urban setting: are they actually compliant to the need for treatment criteria or arranged for dangerous not clinical condition? *Front Psychiatry*. (2019) 9:740. doi: 10.3389/fpsyt.2018.00740
5. Farrelly S, Szmukler G, Henderson C, Birchwood M, Marshall M, Waheed W, et al. Individualisation in crisis planning for people with psychotic disorders. *Epidemiol Psychiatr Sci*. (2014) 23:353–9. doi: 10.1017/S2045796013000401
6. Nowland R, Steeg S, Quinlivan LM, Cooper J, Huxtable R, Hawton K, et al. Management of patients with an advance decision and suicidal behaviour: a systematic review. *BMJ Open*. (2019) 9:e023978. doi: 10.1136/bmjopen-2018-023978
7. Nicaise P, Lorant V, Dubois V. Psychiatric Advance Directives as a complex and multistage intervention: a realist systematic review. *Health Soc Care Commun*. (2013) 21:1–14. doi: 10.1111/j.1365-2524.2012.01062.x
8. Khazaal Y, Chatton A, Pasandin N, Zullino D, Preisig M. Advance directives based on cognitive therapy: a way to overcome coercion related problems. *Patient Educ Couns*. (2009) 74:35–8. doi: 10.1016/j.pec.2008.08.006
9. Henderson C, Lee R, Herman D, Dragatsi D. From psychiatric advance directives to the joint crisis plan. *Psychiatr Services*. (2009) 60:1390–1. doi: 10.1176/ps.2009.60.10.1390
10. Henderson C, Farrelly S, Flach C, Borschmann R, Birchwood M, Thornicroft G, et al. Informed, advance refusals of treatment by people with severe mental illness in a randomised controlled trial of joint crisis plans: demand, content and correlates. *BMC Psychiatry*. (2017) 17:376. doi: 10.1186/s12888-017-1542-5
11. Abettan C. The "virtue" of advance directives. *'Ethique Santé*. (2017) 14:42–8. doi: 10.1016/j.etiqe.2016.10.006
12. Linhorst DM, Hamilton G, Young E, Eckert A. Opportunities and barriers to empowering people with severe mental illness through participation in treatment planning. *Soc Work*. (2002) 47:425–34. doi: 10.1093/sw/47.4.425
13. Szmukler G, Dawson J. Commentary: toward resolving some dilemmas concerning psychiatric advance directives. *J Am Acad Psychiatry Law*. (2006) 34:398–401.
14. Swanson JW, Swartz MS, Elbogen EB, Van Dorn RA, Wagner HR, Moser LA, et al. Psychiatric advance directives and reduction of coercive crisis interventions. *J Mental Health*. (2008) 17:255–67. doi: 10.1080/09638230802052195
15. Campbell LA, Kisely SR. Advance treatment directives for people with severe mental illness. *Cochrane Database Syst Rev*. (2009) 2009:CD005963. doi: 10.1002/14651858.CD005963.pub2
16. Farrelly S, Lester H, Rose D, Birchwood M, Marshall M, Waheed W, et al. Improving therapeutic relationships: joint crisis planning for individuals with psychotic disorders. *Qual. Health Res*. (2015) 25:1637–47. doi: 10.1177/1049732314566320
17. Shields LS, Pathare S, van der Ham AJ, Bunders J. A review of barriers to using psychiatric advance directives in clinical practice. *Administr Policy Mental Health*. (2014) 41:753–66. doi: 10.1007/s10488-013-0523-3
18. Easter MM, Swanson JW, Robertson AG, Moser LL, Swartz MS. Facilitation of psychiatric advance directives by peers and clinicians on assertive community treatment teams. *Psychiatr Services*. (2017) 68:717–23. doi: 10.1176/appi.ps.201600423
19. Barlati S, Minelli A, Ceraso A, Nibbio G, Carvalho Silva R, Deste G, et al. Social cognition in a research domain criteria perspective: a bridge between schizophrenia and autism spectra disorders. *Front Psychiatry*. (2020) 11:806. doi: 10.3389/fpsyt.2020.00806
20. Wallace E, Rodriguez P, Feng S, Yamada I, Boyd-Graber J. Trick me if you can: human-in-the-loop generation of adversarial question answering examples. *Trans Assoc Comput Linguist*. (2019) 7:387–401. doi: 10.1162/tacl_a_00279
21. Villani C, Schoenauer M, Bonnet Y, Berthet C, Cornut AC, Levin F, et al. *For a Meaningful Artificial Intelligence: Towards a French and European Strategy*. Paris: Conseil national du numérique (2018).
22. Jain SH, Powers BW, Hawkins JB, Brownstein JS. The digital phenotype. *Nat Biotechnol*. (2015) 33:462–3. doi: 10.1038/nbt.3223
23. Barnett I, Torous J, Staples P, Sandoval L, Keshavan M, Onnela JP. Relapse prediction in schizophrenia through digital phenotyping: a pilot study. *Neuropsychopharmacology*. (2018) 43:1660–6. doi: 10.1038/s41386-018-0030-z
24. Zulueta J, Piscitello A, Rasic M, Easter R, Babu P, Langenecker SA, et al. Predicting mood disturbance severity with mobile phone keystroke metadata: a biaffect digital phenotyping study. *J Med Intern Res*. (2018) 20:e241. doi: 10.2196/jmir.9775
25. Corcoran CM, Carrillo F, Fernandez-Slezak D, Bedi G, Klim C, Javitt DC, et al. Prediction of psychosis across protocols and risk cohorts using automated language analysis. *World Psychiatry*. (2018) 17:67–75. doi: 10.1002/wps.20491
26. Place S, Blanch-Hartigan D, Rubin C, Gorrostieta C, Mead C, Kane J, et al. Behavioral indicators on a mobile sensing platform predict clinically validated psychiatric symptoms of mood and anxiety disorders. *J Med Intern Res*. (2017) 19:e75. doi: 10.2196/jmir.6678
27. Kleiman EM, Turner BJ, Fedor S, Beale EE, Picard RW, Huffman JC, et al. Digital phenotyping of suicidal thoughts. *Depress Anxiety*. (2018) 35:601–8. doi: 10.1002/da.22730
28. Ferreri F, Bourla A, Mouchabac S, Karila L. e-Addictology: an overview of new technologies for assessing and intervening in addictive behaviors. *Front Psychiatry*. (2018) 9:51. doi: 10.3389/fpsyt.2018.00051
29. Umbricht D, Cheng WY, Lipsmeier F, Bamdadian A, Lindemann M. Deep learning-based human activity recognition for continuous activity and gesture monitoring for schizophrenia patients with negative symptoms. *Front Psychiatry*. (2020) 11:967. doi: 10.3389/fpsyt.2020.574375
30. Li Y, Cai M, Qin S, Lu X. depressive emotion detection and behavior analysis of men who have sex with men via social media. *Front Psychiatry*. (2020) 11:830. doi: 10.3389/fpsyt.2020.00830
31. Uusitalo S, Tuominen J, Arstila V. Mapping out the philosophical questions of AI and clinical practice in diagnosing and treating mental disorders. *J Eval Clin Practice*. (2020). doi: 10.1111/jep.13485. [Epub ahead of print].
32. Turing AM. Computing machinery and intelligence. *Mind*. (1950) 59:433–60. doi: 10.1093/mind/LIX.236.433
33. Casey AJ, Niblett A. Self-driving laws. *Univer Toronto Law J*. (2016) 66:429–42. doi: 10.3138/UTLJ.4006
34. Miller KW, Wolf MJ, Grodzinsky F. This "Ethical trap" is for roboticists, not robots: on the issue of artificial agent ethical decision-making. *Sci Eng Ethics*. (2017) 23:389–401. doi: 10.1007/s11948-016-9785-y
35. Swanson JW, McCrary SV, Swartz MS, Elbogen EB, Van Dorn RA. Superseding psychiatric advance directives: ethical and legal considerations. *J Am Acad Psychiatry Law*. (2006) 34:385–94.

36. Berreby F, Bourgne G, Ganascia JG. Event-based and scenario-based causality for computational ethics. In: *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems. AAMAS '18*. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems (2018). p. 147–55.
37. Grinbaum A, Chatila R, Devillers L, Ganascia J, Tessier C, Dauchet M. Ethics in robotics research: CERNA mission and context. *IEEE Robot Autom Mag*. (2017) 24:139–45. doi: 10.1109/MRA.2016.2611586
38. Berreby F, Bourgne G, Ganascia JG. A declarative modular framework for representing and applying ethical principles. In: *AAMAS*. Caen (2017). Available online at: https://hal.sorbonne-universite.fr/hal-01564673
39. Demiaux V, Si Abdallah Y. *How Can Humans Keep the Upper Hand? Report on the Ethical Matters Raised by Algorithms and Artificial Intelligence*. Paris: Commision Nationale Informatique et Libertés (2017).
40. Jahn WT. The 4 basic ethical principles that apply to forensic activities are respect for autonomy, beneficence, nonmaleficence, and justice. *J Chiropract Med*. (2011) 10:225–6. doi: 10.1016/j.jcm.2011.08.004
41. Ioannidis JPA. Why most published research findings are false. *PLoS Med*. (2005) 2:2–8. doi: 10.1371/journal.pmed.0020124
42. Ioannidis JPA. How to make more published research true. *PLoS Med*. (2014) 11:e1001747. doi: 10.1371/journal.pmed.1001747
43. Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, et al. Confidence and precision increase with high statistical power. *Nat Rev Neurosci*. (2013) 14:585. doi: 10.1038/nrn3475-c4
44. Munafó MR, Nosek BA, Bishop DVM, Button KS, Chambers CD, Percie du Sert N, et al. A manifesto for reproducible science. *Nat Hum Behav*. (2017) 1:21. doi: 10.1038/s41562-016-0021
45. Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci*. (2013) 14:365–76. doi: 10.1038/nrn3475
46. Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, et al. Empirical evidence for low reproducibility indicates low pre-study odds. *Nat Rev Neurosci*. (2013) 14:877. doi: 10.1038/nrn3475-c6
47. Prinz F, Schlange T, Asadullah K. Believe it or not: how much can we rely on published data on potential drug targets? *Nat Rev Drug Discov*. (2011) 10:712. doi: 10.1038/nrd3439-c1
48. Larsen ME, Nicholas J, Christensen H. a systematic assessment of smartphone tools for suicide prevention. *PLoS ONE*. (2016) 11:e0152285. doi: 10.1371/journal.pone.0152285
49. Popper KR. *The Logic of Scientific Discovery*. Oxford: Basic Books (1959). doi: 10.1063/1.3060577
50. Latour B, Woolgar S. *Laboratory Life: The Construction of Scientific Facts*. Princeton University Press (1986). Available online at: http://www.jstor.org/stable/j.ctt32bbxc. doi: 10.1515/9781400820412
51. Lo B. Professionalism in the age of computerised medical records. *Singapore Med J*. (2006) 47:1018–22.
52. Satkoske VB, Parker LS. Practicing preventive ethics, protecting patients: challenges of the electronic health record. *J Clin Ethics*. (2010) 21:36–8.
53. Jacquemard T, Doherty CP, Fitzsimons MB. Examination and diagnosis of electronic patient records and their associated ethics: a scoping literature review. *BMC Med Ethics*. (2020) 21:76. doi: 10.1186/s12910-020-00514-1
54. Bourla A, Ferreri F, Ogorzelec L, Peretti CS, Guinchard C, Mouchabac S. Psychiatrists' attitudes toward disruptive new technologies: mixed-methods study. *JMIR Mental Health*. (2018) 5:e10240. doi: 10.2196/10240
55. Cederström C, Spicer A. *The Wellness Syndrome*. 1st ed. Stockholm Business School; Faculty of Social Sciences; Stockholm University: Polity Press (2015). Available online at: http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-111975
56. Fiske A, Henningsen P, Buyx A. Your robot therapist will see you now: ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. *J Med Intern Res*. (2019) 21:e13216. doi: 10.2196/13216
57. Druss BG, Rosenheck RA, Desai MM, Perlin JB. Quality of preventive medical care for patients with mental disorders. *Med Care*. (2002) 40:129–36. doi: 10.1097/00005650-200202000-00007
58. Chwastiak LA, Rosenheck RA, Kazis LE. Utilization of primary care by veterans with psychiatric illness in the National Department of Veterans Affairs Health Care System. *J Gen Intern Med*. (2008) 23:1835–40. doi: 10.1007/s11606-008-0786-7
59. Dembling BP, Chen DT, Vachon L. Life expectancy and causes of death in a population treated for serious mental illness. *Psychiatr Services*. (1999) 50:1036–42. doi: 10.1176/ps.50.8.1036
60. Hannerz H, Borga P, Borritz M. Life expectancies for individuals with psychiatric diagnoses. *Publ Health*. (2001) 115:328–37. doi: 10.1016/S0033-3506(01)00471-1
61. Colton CW, Manderscheid RW. Congruencies in increased mortality rates, years of potential life lost, and causes of death among public mental health clients in eight states. *Prevent Chron Dis*. (2006) 3:A42.
62. Chang CK, Hayes RD, Perera G, Broadbent MTM, Fernandes AC, Lee WE, et al. Life expectancy at birth for people with serious mental illness and other major disorders from a secondary mental health care case register in London. *PLoS ONE*. (2011) 6:e19590. doi: 10.1371/journal.pone.0019590
63. Asimov I. I, robot. *Garden City*. New York, NY: Gnome Press (1950).
64. Abiteboul S, Dowek G. *The Age of Algorithms*. Cambridge: Cambridge University Press (2020). Available online at: https://www.cambridge.org/core/books/age-of-algorithms/3E2937477538DB53669E2919BE565288
65. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. *Evid Based Mental Health*. (2018) 21:116–9. doi: 10.1136/eb-2018-102891

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.