AI and Ethics
https://doi.org/10.1007/s43681-022-00246-5

ORIGINAL RESEARCH

Artificial moral experts: asking for ethical advice to artificial intelligent assistants

Blanca Rodríguez-López[1] · **Jon Rueda**[2]

Received: 8 August 2022 / Accepted: 29 November 2022

Abstract
In most domains of human life, we are willing to accept that there are experts with greater knowledge and competencies that distinguish them from non-experts or laypeople. Despite this fact, the very recognition of expertise curiously becomes more controversial in the case of "moral experts". Do moral experts exist? And, if they indeed do, are there ethical reasons for us to follow their advice? Likewise, can emerging technological developments broaden our very concept of moral expertise? In this article, we begin by arguing that the objections that have tried to deny the existence (and convenience) of moral expertise are unsatisfactory. After that, we show that people have ethical reasons to ask for a piece of moral advice in daily life situations. Then, we argue that some Artificial Intelligence (AI) systems can play an increasing role in human morality by becoming moral experts. Some AI-based moral assistants can qualify as artificial moral experts and we would have good ethical reasons to use them.

Keywords Artificial intelligence · Dialogical assistant · Moral advice · Moral expertise

1 Introduction

Morality is so-considered a hallmark of humanity [1]. Since humans are social and moral animals, we tend to consider others when modulating our behavior. When we are not entirely clear about how to act, we sometimes ask others for advice. But seeking advice is not as trivial as it may seem. It is not only the content of the advice that matters, but also who the adviser is. Although all typical human adults routinely make morally relevant decisions, we do not all have the same knowledge and skills. Thus, who are the best people to give a piece of *moral* advice? Are there individuals better qualified to advise others on moral matters?

In most domains of human life, we are willing to accept that there are experts with greater knowledge and competencies that distinguish them from non-experts or laypeople. Interestingly, this tendency creates more misgivings in the case of "moral experts". The philosophical dispute over the existence or not of moral experts has been protracted for almost half a century. Since the 1970s, the very concept of 'moral expertise' has increasingly been the subject of analysis by a variety of authors from different ethical traditions [2–5]. Today this continues to be a lively debate. Are there any moral experts? If indeed there are, do we have ethical reasons to follow their advice? Likewise, can emerging technological developments broaden our very concept of moral expertise?

This article aims to provide a philosophical answer to the above questions. First, we show that the arguments that have tried to deny the existence of moral experts are unsatisfactory. As most of the counter-arguments are refutable, we have no reason to believe that the existence of moral experts is any more controversial than in other areas such as economics, healthcare, architecture or fitness. Second, we argue that seeking moral advice is a common human experience and that we have ethical reasons to ask for a piece of advice from the so-considered moral experts. Third, we think that the debate about moral expertise should be updated in light of recent developments in Artificial Intelligence (AI). AI commonly refers to a set of computational technologies that are capable of performing tasks that would normally require human intelligence—such as object detection, complex problem solving, language translation, or predictive judgements. We advance the thesis that some AI systems

✉ Jon Rueda
ruetxe@gmail.com; ruetxe@ugr.es

[1] University Complutense of Madrid, Madrid, Spain

[2] University of Granada, Granada, Spain

Published online: 09 January 2023

can play an increasing role in human morality by becoming moral experts.

To meet those purposes, the structure of our article is the following. In the first section, we start by clarifying the concepts of 'expertise' and 'moral expertise', and some recurring distinctions that are often made about the latter. Moreover, we present the most recurrent objections against moral experts and show how these can be refuted. Then, in the second section, we analyze the ethical issues raised by the phenomenon of seeking moral advice and relate it to the controversy that surrounds moral expertise. After that, in the third section, we defend that some AI systems could be in fact considered moral experts. To narrow our argument and to illustrate the previous claim with a particular example, we focus on the AI-based Socratic Assistant (or also called SocrAI) devised by Francisco Lara and Jan Deckers [6, 7]. To conclude, we argue that SocrAI can be considered an artificial moral expert to whom people would have cause to ask for advice.

2 Moral expertise

The concept of expertise is exclusionary and restricted as long as it is based on a set of characteristics not possessed by others or possessed to a much lesser extent [8]. That is to say, expertise is generally acknowledged relative to a comparison or contrast class [4, 9]. Consider, for instance, the following generic definition of expertise given by Jonathan Matheson and colleagues:

> Someone S is an expert in domain D at time T with respect to population P just in case S possesses an unusually extensive body of knowledge in D at T and S has unusually extensive skills to apply that knowledge at T to new questions and problems compared to others in P [10].

Although that broad definition seems mainly uncontroversial, expertise becomes a more contentious issue when it is applied to the moral domain. Do *moral* experts exist? And if they exist, who are they? How do we come to recognize them? Similarly, what arguments support its existence and what arguments deny it? In this section, we address these questions and attempt to briefly answer them by offering a plausible characterization of moral expertise.

To begin with, the delimitation of the salient features of moral expertise is challenging. It has been argued that moral experts are characterized by the possession of some skills, knowledge and values [11, 12]. These competencies can be moral or non-moral [3]. For instance, Peter Singer famously claimed that moral experts are not only acquainted with moral theories and moral concepts, but they also need to have abilities in logic, argumentation, in getting information about concrete facts or even having more time to think about moral issues [2]. Singer argued that moral philosophers, insofar as they generally fulfil such competencies, could be qualified as moral experts. Whether philosophers are more likely to possess moral expertise is something that has been widely debated and that can be accepted from different theoretical approaches.

Following Karen Jones and François Schroeter's view, the previous Singer's conception of moral expertise can be understood as an intellectualist model [13]. This model is based on the acquisition of knowledge of ethical theories and on the abilities to subsume the relevant facts to a particular controversy under those theories, especially through reasoning capabilities that avoid fallacies. Conversely, Jones and Schroeter stated that another predominant conception of moral expertise comes from particularist approaches related to virtue theory, which are based on the importance of moral perception and practical wisdom through habituation. As they put, according to this model, "the core capacity grounding moral expertise is a capacity to discern the moral salience of considerations in particular contexts." [13, 14] Whereas both models are quite dissimilar, they do share two characteristics. First, both can endorse the claim that philosophers are more likely to be moral experts than non-philosophers. Second, both conceptions relate to the idea that moral experts provide action-guiding judgments about particular circumstances or problems. In what follows we will not specifically deal with the controversy about the differences between philosophers and non-philosophers, but we will rather focus on the issue of action-guiding verdicts.

A clarification is needed first with regard to the action-guiding role of moral experts. The way we conceive of the function of many experts is not restricted to giving us advice on what concrete course of action we should take. Many experts do not tell us what we should do, but rather indicate important factors (which we may not have paid attention to) so that we can form a personal and autonomous decision on the basis of our preferences. This support in the processes of decision-making, as we will see in the third section, is central to our discussion below, since we will show that dialogical assistance would be an attractive function that artificial moral experts could fulfil from a procedural ethics perspective.

Having said that, there are other distinctions regarding moral expertise that are worth clarifying. An interesting distinction was addressed by Arthur Caplan and by François Bayles, who reflected on the differences between *moral expertise* and *moral experts*. [12, 15] Philosophers may have moral expertise as long as they have been trained to be proficient in moral theories and analytical skills, but this does not

necessarily make them moral experts. Whereas moral expertise relates to 'know that', moral experts also need to 'know how' [12].[1] That distinction is akin to another one that differentiates between *propositional expertise* from *performative expertise*. While propositional or descriptive expertise "it is constituted by knowledge *that* something is the case", performative expertise is constituted "by knowledge of *how* to do something, that is by the superior, accomplished or definitive performance of some act or class of actions." [8] [2] In the same vein, Julia Driver claimed that there are different forms of moral expertise:

> there are at least three distinct forms moral expertise can take: there is the expert judger, who does a better job of arriving at true moral judgments, the expert practitioner, who acts morally well more than others, and the expert in moral analysis who has greater than normal insight into the nature of morality (in some respect) [9].

Holding these distinctions in mind, moral expertise does not seem to be an all or nothing category. It rather seems that it can be subjected to different degrees and sub-specifications. With this in view, we can define a moral expert as "someone who knows what people ought to do or is at least capable of helping people see more clearly (perhaps through questioning) what they have good moral reasons for doing" [17].

Thus, when we wonder if moral experts exist—it would be peculiar if they did not—, we also recognize experts in a wide variety of domains, from engineering to cooking, from medieval history to medicine. In fact, the arguments in favor of their existence found in the literature are little more than replies to the objections posed against it. Nevertheless, there have been various objections to the existence of moral experts. Henceforth, we highlight some of the most recurrent ones and show how they can be refuted.

The first objection is the *Argument from the absence of objectivity*. The apparently most destructive argument is the alleged lack of objectivity in the moral realm. Expertise is based on objective knowledge of a particular matter. Thus, for there to be moral experts, there should be objective moral knowledge [8, 17].[3] Since there is no objectivity in ethics (i.e. no access to independent moral truths or objective moral facts), there cannot be moral experts. (This argument can also be sustained from emotivist positions in metaethics that claim that morality is ultimately a matter of subjective feelings) [3, 9, 18].[4]

This argument has three problems. First, it begs the question of the lack of objectivity on ethical matters. If one does not accept the premise that there is no objectivity in ethics (as is the case with moral realists), this argument does not hold. Second, even if we grant subjectivism, it is still possible to have standards for ranking subjective judgements [8]. This can happen in disciplines such as literature, music or film criticism. Third, and more importantly, the initial premise that expertise is only limited to objective knowledge is highly doubtful. In other disciplines such as economics or genetic counselling, for instance, expertise is recognized not only from previous knowledge (that is not always necessarily related to objective truths), but also from particular skills [15]. Expertise in those areas depends on value judgements (and also on probabilistic information), and thus we do not expect 'truth' in their insights, but knowledgeable answers supported by good reasoning. Similarly, some have argued that the existence of moral experts is not related to their judgments being true or referring to objective moral facts, but rather to their ability to coherently justify their judgments [19].

The second objection is the *Argument from disagreement*. There is deep disagreement within moral philosophy about moral matters. This disagreement about a wide range of moral problems suggests that non-experts should not defer to the normative recommendations of alleged moral experts [20].

This objection can be counter-argued, though. The existence of expert disagreement among moral philosophers is highly overrated, as in many cases the same moral conclusion is reached from different moral theories [8]. Also, disagreements are sometimes more factual than evaluative. [9, p. 291]. There is no more profound disagreement in ethics than in other areas of knowledge and there are also substantive agreements about great deals [8] (e.g. the impermissibility of causing gratuitous suffering to beings with moral status). Finally, acknowledging the fact that experts may disagree on particular issues does not necessarily imply denying the existence of expert judgements, though it makes it advisable to consult more than one, as we often do in medical or legal matters [11].

The third objection is the *Argument from "we are all moral experts"*. As each person has the capacities and the moral sense about what is right, anyone can claim to be a moral expert [12] pp. 73–5. Likewise, from Kantian philosophy, all people have the component of morality in

[1] [15] bring the distinction between 'know that' and 'know how' from [16]

[2] [8], p. 127. However, Bruce Weinstein distinguished between descriptive expertise and performative expertise in more different sense: "The descriptive expert, according to Weinstein, is able to make expert moral judgements about what, morally, ought to be done in this situation, and has the capacity to justify such judgements. The performative expert is able to ''get it right'' without necessarily having the capacity to justify the judgement or explain how he got it right." [23, cited in 24].

[3] Cited in [8].

[4] Cited in [3, 9]

themselves and therefore everyone—not only those who have special capacities—can be moral by following the goodwill to respect the moral law [3]. From common-sense morality, moreover, it is also argued that ordinary people have the moral resources (the set of basic moral maxims and capacities to elementary judgments about right and wrong), which they use in their daily lives [8]. Consequently, to some extent "*we are all ethical experts*, and so effectively none of us are." [21, italics in the original sentence].

The third objection can also be replied. To begin with, the arguments of this category have also been heavily contested, sometimes by pointing out the problematic relation between moral philosophy and common sense morality [22] and other times by acknowledging the role of the philosopher even when claiming that moral philosophy is somewhat based in common sense morality [8]. Also, even if we assume that we are all moral agents and cannot avoid making moral judgements and decisions, we can still claim that some people are better than others in this regard [23].

The fourth and last objection is the *Argument from autonomy and democratic liberalism*. The view that every person has the necessary capacities to autonomously lead a moral life is supported by democratic liberalism [12]. The liberal value of autonomy entails that people must make their own moral judgements of how they should live their life. Hence, individuals should not solely follow the judgments of others (alleged moral experts included), but they would also need to autonomously decide what ought to be done [8].

The counterargument to this objection goes as follows. Autonomy is no more incompatible with the existence of experts in morals as it is in other areas (e.g. medical experts and patient's autonomy). Autonomy also implies that choosing an expert and following specific advice depends on each one of us [5]. Indeed, in the third section, we will show that moral experts (human or artificial) who have roles closer to procedural ethics can be especially respectful of autonomy.

As none of the objections seem to be categorical, we will accept that moral experts exist. As Matheson and colleagues stated, there are people who

> "possess a deep understanding of the relevant facts, issues, and arguments—indeed the entire body of major scholarly literature surrounding a topic—and are able to use that understanding to engage new problems and questions about the topic. Further, some of these people also have the personal and communication skills to competently serve as advisors to families in need of navigating an ethical challenge in an informed way." [10]

We want to underline the importance of 'understanding'. Though in ordinary language 'knowledge' and 'understanding' are used interchangeably, literature on epistemology has established useful differences between them. While knowledge is classically defined as justified true beliefs, understanding—which is not univocal—requires "seeing the way things fit together" [24, p. 218], the "grasping of explanatory and other coherence-making relationships in a large and comprehensive body of information" [25], or to "'grasp' or 'see' how the various parts of the model relate to one another." [26] As is commonly recognized, expertise in general [27] and moral expertise in particular [5, 10, 13, 28–30], requires understanding in the above-mentioned sense.

We must next answer what reasons might prompt us to seek ethical advice, especially from moral experts.

3 The ethics of asking for moral advice

If moral experts who can give moral advice indeed exist, the next question is when and who is in need to ask for such advice. We understand the idea of 'moral advice' in a broad way. Moral advice is not just reduced to the kind of statements such as "you should do *X*", but can also include different kinds of interactions that help us in moral decision-making processes, such as "if you hesitate between *X*, *Y* or *Z*, you should take into account *A*, *B*, or *C*, in order to make your decision". People have reasons to follow the advice of moral experts in many situations [22]. To be in need of moral advice and to ask for it is a common human experience.

We all follow moral rules that serve us well in normal, daily life circumstances: do not lie, do not harm, keep your promises and so on. But there are some circumstances in which these moral rules are not enough. Sometimes the rules conflict in a particular case or one can face a situation so new that they are not sure which rule, if any, applies. In these situations, we talk about moral dilemmas or perplexities. Dilemmatic or perplexing situations are especially frequent when the topic is controversial and new situations arise. Moreover, there are situations in which we are too personally involved, too passionate or biased and, therefore, we have good reasons not to trust our own judgement. When one is in these situations, the right thing to do is to look for moral advice—and we could probably consider that not doing so is somehow morally faulty.

Once we know the "when", it is not difficult to point to the "who": all of us can be in these situations, including those who can be considered moral experts. This is not exclusive to the moral realm. Think, for instance, in medicine or law. Both are areas in which dilemmas or personal involvement are as common as in morality, and where it is not unusual for an expert to ask for the advice of another expert.

More peculiar to the moral realm seems to be the question about how to identify a moral expert. After all, there are credentials for physicians or experts in law and there

seems to be not such a thing for moral experts [8][5]. For some, moral philosophers have a good claim to be moral experts [2, 5], even if their expertise does not include (*qua* philosophers) the necessary knowledge of factual matters. In this case, their advice should be supplemented [17], as they have the relevant skills and knowledge, both moral and non-moral (see above). In fact, moral philosophers are often hired to teach courses in medical or other professional ethics or are asked to join ethical committees. For others, moral philosophers are better considered moral cartographers [11] or cannot claim exclusivity as moral experts [12], as there are others with similar claims [4]. Leaving the question of formal qualifications aside, there are people with better claims to have it right in moral judgments and decisions, at least regarding particular areas—e.g. environmental ethics.

Some wonder how a non-expert can identify a moral expert in the absence of formal qualifications [10, 29], especially if some of the candidates disagree. But we do not think this is an insurmountable problem. If the purported experts, let us say moral philosophers, disagree, this is a reason to consult with more than one expert, as we do in medical or legal matters. And even if we do not think that being a moral philosopher is at least a partial qualification—though we do—we should not despair. Being a non-expert does not mean that you know nothing about morality and that you are at a loss when judging moral matters. To begin with, we all have a good amount of moral practice so even if we are not moral experts, we are not absolute laypeople either and the fact that we are not all experts in reasoning does not mean that we are unable to acknowledge a good argument. According to our definition, a moral expert is not someone who simply tells us what we should do, but someone who has the capacity to explain to us the reasons behind different courses of action. Indeed, moral experts can be identified by considering how often they answer moral questions correctly [29, 30][6] or by looking how well other people who followed the advice of a particular expert are doing, including whether their moral lives have improved [10, 31].

The last question we want to address in this section is if we should follow the advice of a moral expert. For some authors, especially the moral philosophers, there are reasons "to believe that others should not lead their lives by the lights of moral judgments they do not themselves make" [8]—reasons sustained by the value of autonomy in moral life. The general view is that for an action to have moral value it is not only necessary that the action itself is morally correct, but also that the agent can judge with their own reason that the action is indeed morally correct and that they understand the reason why it is correct. As we have already pointed out, moral autonomy is compatible with asking for moral advice, and this can even be the morally correct thing to do if one has some reason to trust somebody else in a given circumstance more than to trust themselves. The decision to ask for moral advice is, or can be, an autonomous one.

That being said, we can wonder if following the advice given by an expert is compatible with the value of autonomy. Following advice does not mean to follow it blindly. It is usually understood that to act morally you should not only know that action *A* is correct, but it is also necessary to know the reason why, as well as the relation between this set of reasons and the rightness of the action *A*. [13, 28] Our definition of a moral expert in the previous section includes their role in explaining the reasons for their advice. Many times, we are able to understand these reasons and, in these cases, following the advice is totally autonomous. Even so, some object that it is risky to ask and follow moral advice as we can get too used to it and that can make us "moral cripples". This objection is easy to dismiss. First, it is an empirical claim with no empirical data to support it. Second, it could be the case that we, in fact, gain some moral knowledge by consulting a moral expert and following their advice [17].

In the literature, following and advising once you understand the reasons is different from pure deference [29] or moral testimony [28]. Some deem this as not defensible from a moral perspective, since one acts on someone else's judgement without understanding the reasons, and this is only correct for small children as they still lack the capacity of reason, moral or otherwise. Besides the value of moral understanding and the moral unworthiness of actions without previously mentioned reasons we do not understand, some point out the need to justify our actions to others and the goodness of developing a virtuous character as reasons to object to moral deference [28]. But even this can be contested [13, 17, 30]. Of course, moral deference does not deliver moral understanding, but even if we agree that acting right is not the only important thing in morality, neither is moral understanding. It is difficult to deny that doing the right thing also counts for something. Even if one does not understand the reasons behind a piece of moral advice or when the matter is so urgent that the person simply has no time to consider the reasons, it could be morally wise to follow the expert's advice as long as they have good reasons to trust them and the decision to defer to them can be perfectly autonomous. Trying to increase one's moral understanding is undoubtedly a morally good and wise thing to do, but when this is one's only moral worry, it is as morally faulty as not to worry about it at all.

[5] Though some authors claim that this is not completely true. See [4].

[6] Provides a vivid example of someone *A* who tends to react on the spur of the moment in a way that, once he has cooled down, he himself considers morally wrong. This person has a friend *B* who reacts in a way that *A* endorses after a couple of days. We can undoubtedly say that *A* considers that *B* provides correct moral answers.

4 AI-based ethical advisors as moral experts

So far, we have argued that moral experts exist and that people have ethical reasons to ask for moral assistance in daily life situations. In this section, we advance two novel arguments. We shall first defend that the concept of moral expertise is not necessarily restricted to human beings, but it could also soon be applied to AI systems that meet the characteristics presented in the first section. Then, following the argumentation of the second section, we claim that it may be ethically instructive to seek moral advice from such AI experts, although not in an equal way.

Firstly, in the near future, there may be AI systems that can be considered moral experts. AI could interact with humans about moral issues taking the form of conversational bots, voice assistants or robotic agents. In fact, the use of AI as an ethical advisor or moral enhancer has increasingly been discussed in recent times [32–34]. For instance, Alberto Giubilini and Julian Savulescu tentatively suggested—though they did not thoroughly argue—that their proposal of the Artificial Moral Advisor could play the role of a moral expert, but being "an expert more informed and more capable of information processing than any other human moral expert we trust." [34, p. 177]. The preceding claim is interesting because it not only hints that there may be AI systems that are moral experts, but also that these systems may surpass human moral experts in some competencies. The latter question is controversial (as it suggests a kind of ontological superiority of AI for certain functions) and we will not resolve it here, so we will leave it open. Rather, we are more interested in convincing, below, that there are AI systems that could be moral experts.

To support our first assertion that there could be AI systems with moral expertise, we will focus on a more promising AI moral assistant that has recently been devised by Francisco Lara and Jan Deckers. This theoretical (i.e., not already used) model has been called the Socratic Assistant [6] or, in a catchier way, SocraAI [7]. SocrAI is a conversational bot (in principle without a robotic body) that aims to enhance human moral decision-making processes.[7] This deliberative system was conceived in the wake of the moral enhancement debate to overcome the shortcomings of *bio*-enhancement proposals through an AI model that would develop the dialogical Socratic method. Although the debate of moral enhancement through AI is highly stimulating, we cannot fairly address it here.

We will focus on SocrAI because we believe it is the most complete and attractive AI moral assistant proposal to be considered a moral expert. Among the features of SocrAI, Lara and Deckers include providing empirical support, improving conceptual clarity, understanding argumentative logic, testing whether one's judgement may possess ethical plausibility, raising awareness of personal limitations, and advising on how to execute one's decisions [6]. SocrAI's axiological neutrality[8] would also help those assisted to make reflective decisions for themselves, therefore avoiding the risk of becoming a moral cripple as mentioned. This AI voice assistant would thus guarantee the full participation of the user [7, p. 10]. Overall, SocrAI fulfils the moral and non-moral capabilities normally attributed to moral experts, which we have extensively discussed in the first section.[9]

However, Lara and Deckers have never gone so far as to characterize this system model as a moral expert. We believe that such a conception is possible—and even necessary to encourage future use of these AI systems to support moral decision-making. We believe that one possible reason why Lara and Deckers have not considered SocrAI as a moral expert is because of their intention to separate their proposal from other artificial moral counsellors.[10] These authors understand the Socratic assistant more as an instructor (in moral education) rather than as a counsellor. Lara and Deckers' justification is that rival models of moral assistants did not sufficiently protect users' autonomy through their counselling processes. Although this reason is understandable, we think that, if we broaden the notion of asking for advice

[7] The discussion of the technical plausibility of developing this particular AI system from an engineering or computer science perspective is relevant, but beyond the scope of this article. However, the proponents give some general clues in this regard. According to Lara (2021), the algorithms of SocrAI could be developed through a hybrid design strategy that combines top-down ethical principles with bottom-up learning. The Socratic assistant should thus be opened to learn from users' reactions to update the language processing, data mining, and functional skills of the programme. IBM's Project Debater (based on supervised learning algorithms and which has adopted a neural network for processing natural language created by Google called BERT) would be a reference example for the algorithmic development of this dialogic assistant. The discussion on the technical issues of this system should be further elaborated in the future. The issue is important, for example, because if failures in the system were to occur, this could degrade the quality of the moral expertise. We thank an anonymous reviewer for this comment.

[8] To avoid misunderstanding, by 'axiological neutrality' we do not mean that AI systems in general are "neutral" technologies that are not imbued with values. Rather, Lara and Deckers use this expression to refer to one of the classic requirements of procedural ethics, namely, the absence of substantive position-taking at the normative level.

[9] The definition of moral expert adopted in this paper is partial in terms of knowledge. Thus, expressions such as "deep understanding" used in the first part can reasonably be considered inappropriate when applied to machines, especially for philosophers impressed by the "Chinese Room Argument". See [35, 36]. Although this debate is unfortunately beyond the scope of this paper, we should remark that we do not claim that a machine can act morally but only that it can provide moral advice regardless of whether it has a deep understanding in the sense claimed by Searle or not.

[10] Francisco Lara has (in a personal communication) confirmed this fact to us.

(as an autonomous activity that is not strictly action-guiding but decision-supporting), this problem would be overcome. Therefore, SocrAI could be considered an artificial moral expert which, moreover, would be less problematic in ethical terms than other AI virtual assistants as we shall see below.

Though the two definitions of 'moral expert' in the first section are intended to apply to humans, we believe that they can be used to define an AI device without necessarily claiming that it would be considered a person. McConnell's definition tells us what we should expect from a moral expert, namely that "(…) is at least capable of helping people see more clearly (perhaps through questioning) what they have good moral reasons for doing". This is precisely what SocrAI offers. More complex, and interesting, was Matheson's definition, which tells us what characteristics an expert should possess in order to help us in McConnell's sense. We have claimed that SocrAI has access "to the relevant facts, issues, and arguments" and is able to use those to "engage new problems and questions", but it is debatable that an IA system can be said to have understanding. Terms used in the definitions of "understanding ("grasp" or "see") seem to make reference to things that only persons can do. We do not need to take sides in this debate. The important thing is not what makes an AI system different from persons, but what is similar. The similarity resides in the AI system's ability to point to "coherence-making relationships in a large and comprehensive body of information" or to the reasons why an action is morally wrong. When SocrAI points to some ambiguity in the way we are using concepts, or to faults in our reasoning, it does not matter if terms such as "grasp" or "see" are properly used or not. In any case, it helps us to understand these things.

One last consideration is related to the relationship between trusting, expertise and autonomy—which we have addressed in the second section. Trust is fundamental for deference, and we have defended that in some cases moral deference can be justified and compatible with autonomy. But this cannot be applied to an AI expert such as SocrAI, not because trust is a difficult topic in AI, but for a simpler reason. As long as SocrAI's purpose is to provide the user with relevant facts and help with conceptual clarification and moral reasoning, it does not give pieces of advice of the form "you should do *X*" that the user could be tempted to defer to. Rather, SocrAI provides the user with the means to understand what could be considered a correct moral judgement in a particular instance.

Secondly, our further claim is that people have good reasons to ask for ethical assistance from AI-based moral advisory systems. Indeed, AI has significant potential for the domain of counselling [37]. Moral counselling is not an exception. If seeking advice is part and parcel of our socio-moral lives, as we have argued in the previous section, AI expert systems may play a role in our demand for moral counsel. Moreover, if AI ethical advisors become widespread in the future—for example, through their inclusion in smartphone apps—, it may even become an everyday activity to seek advice from such artificial moral experts. However, we believe that this possibility is not unproblematic. Not every type of system would be equally desirable in ethical terms.

The AI ethical advisors proposed by Savulescu & Maslen [32] and Giubilini & Savulescu [34] can be objected to on the grounds of the passivity to which they relegate users. In their proposals, users were rather passively receiving the advice (or even the verdict on the course of action) from the AI—even if it originated from the users' own values. In contrast, SocrAI allows the advice process to be a two-way street. The Socratic assistant is based on the active flow of arguments between the AI and the user. In our opinion, these characteristics would make this system ethically more desirable than the other proposed models for at least two reasons. Firstly, the fact that the very act of asking for counselling takes place through a dialogical process not only makes it ethically more instructive, but it also attributes a leading role to the autonomy of each user. The user is the one who will make their own decisions (hopefully) based on the reasons that they have found to be the best in their dialectical exchange with the intelligent assistant. Secondly, this autonomy is reinforced by the axiological neutrality of the system. SocrAI enhances the procedural aspects of decision-making, without privileging any substantive ethical position. In consequence, two of the most prominent objections to moral expertise—the argument from disagreement and the argument from autonomy—are again surmounted. The Socratic assistant is an artificial moral expert that preserves the core value of autonomy and that respects the fact that users may hold different normative values.

All in all, thinking about the possible characteristics and functioning of these AI advisory models is important. This task should be carried out now, given that they are proposals whose future developments we still have time to modify. In any case, uncritical deference to these systems (even to the most optimal) would be undesirable. In our view, we must promote expert systems that respect the autonomy of users and manipulate as little as possible. If we were to achieve this, artificial moral experts would be beneficial assistants to form our own moral judgements in the most reflective and unbiased way. Moreover, in the case of SocrAI, Francisco Lara points out that the use of this system could have a motivating effect in translating our moral judgements into practice: "the motivational force of the decisions will increase even more as the individual considers that such decisions are the result of a demanding learning process in which it was constantly necessary to debate with an expert." [7, p. 21].

To summarize, some AI moral assistants could be considered moral experts. As long as AI-based advisory systems

may become a salient phenomenon in the future, we should be concerned about the characteristics of these expert systems. SocrAI would be one of those to whom we would have good reason to turn to for moral advice.

5 Conclusion

We can all find ourselves in circumstances where we do not know what we should do from a moral point of view. Thus, we can all benefit from receiving a good piece of moral advice. In the first part of this paper, we have concluded in favor of the existence of the so-called moral experts and adopted a definition that we find compelling. In the second part, we have defended that when we are in morally troubling circumstances, looking for (and often following) moral advice, far from being antithetical to morality, is the morally right thing to do. Finally, in the third part, we have claimed that some AI systems, especially SocrAI, could fill the role of a moral expert.

Being at a loss to know what to do from a moral point of view may not be so common as needing assistance to arrive at our favorite vacation destination (or is it?). However, ethical disorientation is much more worrisome. The consequences of getting it wrong are far more serious. If we are willing to use a GPS, we have all the reason to welcome the possibility of SocrAI, a moral compass that can dialogically guide us on the winding routes of morality.

Acknowledgements This article is part of the research project EthAI+3 (Digital Ethics. Moral Enhancement through an Interactive Use of Artificial Intelligence), funded by the State Research Agency of the Spanish Government (PID2019-104943RB-I00) and the project SOCRAI3 (Moral Enhancement and Artificial Intelligence. Ethical aspects of a virtual Socratic assistant), funded by FEDER Junta de Andalucía (B-HUM-64-UGR20). Jon Rueda thanks the funding of an INPhINIT Retaining Fellowship of the La Caixa Foundation (Grant number LCF/BQ/DR20/11790005). We also thank the audience of the Universidad de La Laguna of Tenerife for their insightful questions, as well as two anonymous reviewers of this journal, and the comments of Pedro Garrido that help us to improve our writing.

Funding Funding for open access publishing: Universidad de Granada/CBUA.

Data availability statement Not applicable.

Declarations

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.

References

1. Ayala, F.J.: The difference of being human: morality. Proc. Natl. Acad. Sci. USA **107**(SUPPL. 2), 9015–9022 (2010). https://doi.org/10.1073/pnas.0914616107
2. Singer, P.: Moral experts. Analysis **32**(4), 115–117 (1972)
3. Burch, R.W.: Are there moral experts? Monist **58**(4), 646–658 (1974)
4. Miller, P.: Who are the moral experts? J. Moral Educ. **5**(1), 3–12 (1975). https://doi.org/10.1080/0305724750050101
5. Szabados, B.: On 'moral expertise.' Can. J. Philos. **8**(1), 117–129 (1978)
6. Lara, F., Deckers, J.: Artificial intelligence as a socratic assistant for moral enhancement. Neuroethics **13**(3), 275–287 (2020). https://doi.org/10.1007/s12152-019-09401-y
7. Lara, F.: Why a virtual assistant for moral enhancement when we could have a socrates? Sci. Eng. Ethics (2021). https://doi.org/10.1007/s11948-021-00318-5
8. Archard, D.: Why moral philosophers are not and should not be moral experts. Bioethics **25**(3), 119–127 (2011). https://doi.org/10.1111/j.1467-8519.2009.01748.x
9. Driver, J.: Moral expertise: judgment, practice, and analysis. Soc. Philos. Policy **30**(1–2), 280–296 (2013)
10. Matheson, J., McElreath, S., Nobis, N.: Moral experts, deference & disagreement. In: Watson, J., Guidry-Grimes, L. (eds.) Moral Expertise. Philosophy and Medicine, vol. 129, pp. 87–105. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-92759-6_5
11. Crosthwaite, J.: Moral expertise: a problem in the professional ethics of professional ethicists. Bioethics **9**(4), 361–379 (1995). https://doi.org/10.1111/j.1467-8519.1995.tb00312.x
12. Caplan, A.: Moral experts and moral expertise: do either exist? In: Hoffmaster, B., Freedman, B., Fraser, G. (eds.) Clinical Ethics: Theory and Practice, pp. 59–87. The Humana Press, Totowa (1989)
13. Jones, K., Schroeter, F.: Moral expertise. Anal. Kritik **34**(2), 217–230 (2012). https://doi.org/10.1515/auk-2012-0204
14. McDowell, J.: Virtue and reason. Monist **62**, 331–350 (1979)
15. Baylis, F.: Persons with moral expertise and moral experts: wherein lies the difference? In: Hoffmaster, B., Freedman, B., Fraser, G. (eds.) Clinical Ethics: Theory and Practice, pp. 89–99. The Humana Press, Totowa (1989)
16. Weinstein, B.: The possibility of ethical expertise. Theoret. Med. **15**, 61–75 (1994)
17. McConnell, T.C.: Objectivity and moral expertise. Can. J. Philos. **14**, 193–216 (1984)
18. Ryle, G.: On forgetting the difference between right and wrong. In: Melden, A.I. (ed.) Essays in Moral Philosophy. University of Washington Press, Seattle (1957)
19. Yoder, S.D.: The nature of ethical expertise. Hastings Cent. Rep. **28**(6), 11–19 (1998)
20. Cross, B.: Moral philosophy, moral expertise, and the argument from disagreement. Bioethics **30**(3), 188–194 (2016). https://doi.org/10.1111/bioe.12173

21. Cowley, C.: A new rejection of moral expertise. Med. Health Care Philos. **8**(3), 274 (2005). https://doi.org/10.1007/s11019-005-1588-x
22. Gordon, J.S.: Moral philosophers are moral experts! A reply to David Archard. Bioethics **28**(4), 203–206 (2014). https://doi.org/10.1111/j.1467-8519.2012.02004.x
23. McConnell, T.C.: Objectivity and moral expertise. Can. J. Philos. **14**(2), 193–216 (1984)
24. Riggs, W.D.: Understanding 'virtue' and the virtue of understanding". In: DePaul, M., Zagzebski, L. (eds.) Intellectual Virtue, pp. 203–226. Clarendon Press, Oxford (2003)
25. Kvanvig: The Value of Knowledge and the Pursuit of Understanding, p. 192. Cambridge University Press, New York (2003)
26. Grimm, S.R., Baumberger, C., Ammon, S. (eds.): Explaining Understanding: New Perspectives from Epistemology and Philosophy of Science, p. 88. Routledge, New York (2017)
27. Scholz, O. R.: Symptoms of Expertise: Knowledge, Understanding and Other Cognitive Goods (2018).
28. Hills, A.: Moral testimony and moral epistemology. Ethics **120**(1), 94–127 (2009). https://doi.org/10.1086/648610
29. McGrath, S.: Skepticism about moral expertise as a puzzle for moral realism. J. Philos. **108**(3), 111–137 (2011)
30. Enoch, D.: A defense of moral deference. J. Philos. **111**(5), 229–258 (2014)
31. Riaz, A.: How to identify moral experts. J. Ethics **25**(1), 123–136 (2021). https://doi.org/10.1007/s10892-020-09338-y
32. Savulescu, J., Maslen, H.: Moral enhancement and artificial intelligence: Moral AI? In: Romportl, J., Zackova, E., Kelemen, J. (eds.) Beyond Artificial Intelligence. The Disappearing Human-Machine Divide, pp. 79–95. Springer, New York (2015)
33. Klincewicz, M.: Artificial intelligence as a means to moral enhancement. Stud. Logic Grammar Rhetoric **48**(1), 171–187 (2016). https://doi.org/10.1515/slgr-2016-0061
34. Giubilini, A., Savulescu, J.: The artificial moral advisor. The "ideal observer" meets artificial intelligence. Philos. Technol. **31**(2), 169–188 (2018). https://doi.org/10.1007/s13347-017-0285-z
35. Searle, J.: Minds, brains and programs. Behav. Brain Sci. **3**, 417–457 (1980)
36. Searle, J.: Minds, Brains and Science. Harvard University Press, Cambridge (1984)
37. Fulmer, R.: Artificial intelligence and counseling: four levels of implementation. Theory Psychol. **29**(6), 807–819 (2019). https://doi.org/10.1177/0959354319853045