From decision support systems to autonomous agents: how can we ensure ethical practice?

John Fox[1,2] **and Omar Khan**[2]
[1]Dept. of Engineering Science, Oxford University; [2] The OpenClinical project

Abstract[1]**.** AI techniques have long been a focus for research in developing new tools for clinical decision-making. In most the computer is in a purely supportive role with clinicians and their patients who take the decisions. However it can be surprisingly simple to turn decision support tools into autonomous agents that make the decisions themselves. This paper briefly discusses some of the ethical questions raised by this development using the CREDO decision support platform and the OpenClinical.net knowledge repository to provide examples for discussion.

1. INTRODUCTION

As with stocks, shares and the gold market the perceived value of AI waxes and wanes. Stanley Kubrick's *2001 a space odyssey* captured everyone's imagination in 1968; the Japanese 5th Generation created an enthusiastic bubble in the early 70s, to be punctured in the 80s; the defeat of world chess champion Garry Kasparov by the supercomputer *Deep Thought* and the extended operation of NASA's *Deep Space 1* probe by an "autonomous agent" regenerated funders' interest, and Spielberg's *A.I. Artificial Intelligence* renewed public appetite, and expectations.

As with all markets however AI bulls and AI bears tell the most convincing stories by turns. The bears who say we are far from constructing a general intelligence seem now to be in the ascendant, pointing out that few automata are truly autonomous and most "robots" are largely pre-programmed or remotely controlled. *Avatar*, the film that most recently captured the public imagination as *A.I.* did a decade before, focused on devices that project and amplify human capabilities rather than act autonomously.

Yet, as with markets, the swing of short-term sentiment between bulls and bears isn't the whole story. Despite constant turbulence the long-term trends in the gold price and in AI are solidly "up". In the world of stocks and shares we are seeing the emergence of autonomous systems in high frequency (and in some cases predatory) trading which are already reputed to be making serious money in global markets.

Healthcare is a rather different domain for applying AI but we are beginning to see the same capabilities emerging. In the *Safe and Sound* project we and our colleagues showed how AI systems may populate the digital future in healthcare, demonstrated in a video that shows a benign future of human and artificial agents cooperating for the benefit of patients, clinicians and medical research[2]. The narrative of the video follows a fictitious patient in her "journey" through the diagnosis and treatment of breast cancer, showing how many different tasks and medical services can be automated and orchestrated using AI and other techniques.

A key station on the cancer journey is the *multi-disciplinary meeting*, where all the members of the clinical team discuss each patient to decide what clinical actions and treatments to recommend. The photograph below shows a multidisciplinary meeting at the Royal Free Hospital in London where the team reviews each patient's history, imaging and lab results, current research and so on[3].

Various kinds of imaging are projected at the front of the meeting room to support the clinical discussion. Our decision support system is a novel addition, projected on the left. It is designed to support real-time recording and interpretation of available clinical data and it can assist in most decisions taken by the team (e.g. risk assessment and prognosis, staging and eligibility for participation in trials, test selection and treatment recommendations). The system has been in routine use for about 4 years by the team at the Royal Free and the results, in terms of improved compliance with evidence-based clinical guidelines, are very promising (Patkar et al, 2012).

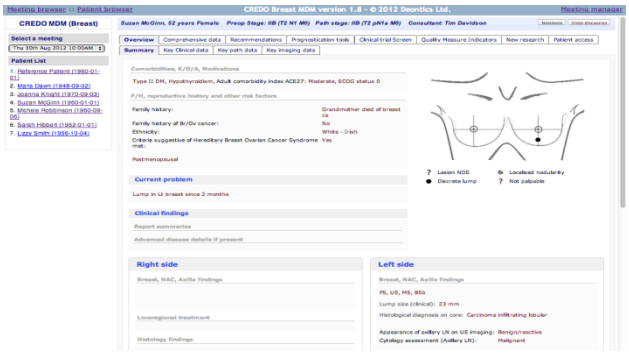
This application builds on a cognitive model of decision-making based on our work in argumentation theory and formalized in the PRO*forma* agent specification and knowledge representation language (Fox, Glasspool and Sutton, 2003; Sutton and Fox, 2003). The data interpretation and decision support capabilities are now implemented in a generic technology for clinical decision support and multidisciplinary patient care called CREDO (Fox, 2014). The argumentation approach is pivotal because it facilitates evidence-based decisions and engages all stakeholders

[1] **Acknowledgement**: We wish to thank Tom Sorrell, Professor of Politics and Philosophy and Head of the Interdisciplinary Ethics Research Group at Warwick University, for his helpful comments on a previous draft of the paper (http://www2.warwick.ac.uk/fac/soc/pais/people/sorell/). Any residual errors or misunderstandings of ethical matters are of course the responsibility of the authors. Thanks also to Mor Peleg, Haifa University, for some suggestions for improving the presentation.

[2] Safe and Sound www.clinicalfutures.org.uk, a grand challenge project funded by the UK Digital Economy programme, was a collaboration between Oxford University, Edinburgh University, Imperial College/St. Mary's Healthcare and UCL/Royal Free Hospital; video on YouTube (http://www.youtube.com/watch?v=zoxpuzH4B_0)

[3] Published with kind permission of Mr. Mo Keshtgar, who leads the breast team at the Royal Free Hospital.

in the decision-making by providing access to the rationale for all recommendations while allowing users to challenge or override inappropriate options. The screen shot below shows an entry screen of the application but a better idea of the functionality and use of argumentation can be seen in a video demonstration at Royal Free Charity Web site. CREDO is now being used to develop a range of applications for other cancers and other clinical settings and to explore ways of supporting patients in their personal healthcare decisions.



Entry screen of web version of the CREDO breast cancer support system

Most CREDO applications to date have been developed to provide decision support, not to take decisions autonomously. For obvious professional and ethical reasons all decisions remain the responsibility of the clinical team. However, the PRO*forma* agent language is designed to permit autonomous operation (Das et al, 1997; Fox and Das, 2000; Fox et al, 2003). With little more than a "flick of a switch" any decision, action or plan described in a PRO*forma* knowledge base can be enacted without human supervision.

The simplicity of this change deserves attention. Looking after cancer patients demands a large slice of every healthcare budget, and multi-disciplinary meetings (which are the standard of care in the UK and many other countries) require the time and expertise of a lot of highly paid professionals, as is evident from the photograph above. If evidence emerges that decision-making and care planning could be carried out without human supervision, yet be effective and medically safe, the potential financial benefits are substantial and the pressure to deploy systems like CREDO in this way would surely grow.[4] The ethical questions arising from this are obvious and we investigate some of these in section 2.

2. SAFE, SOUND AND ETHICAL

Medical decision-making has been an important setting for the discussion of ethical questions in professional practice. The following ethical principles (Beauchamp and Childress, 2008) have been very influential in the English speaking world.

- *Beneficence*: do good.
- *Non-maleficence*: do no harm.
- *Distributive* justice: be fair.
- *Patient autonomy*: respect patient self-direction.

As AI researchers develop the autonomous clinical systems of the future we will be expected to respect principles. There seems no reason to think that systems like CREDO should not support the first three principles but clearly the technology has the potential to violate the last.

If systems like CREDO could be rolled out in an autonomous mode should they? In order to address the question of patient autonomy it seems likely that new principles over and above the traditional ones are likely to be needed. For example:

- *Intelligibility*: a healthcare agent should be able to engage in natural and cooperative interaction with its users (clinicians, patients)
- *Personalisation*: an agent should accommodate the users' personal goals and preferences so long as they don't conflict with the above principles.
- *Justifiability* an understandable rationale must be available for all recommendations, and particularly for automated actions, at whatever level of detail the user may reasonably require.
- *Controllability*: the user must be able to modify the system's assumptions and goals, and the system must be able to adapt appropriately and safely to such changes.

One of the reasons for developing a decision model based on naturalistic argumentation was to ensure that the rationale for clinical decisions would be intelligible rather than implicit in a black box decision-procedure. Argumentation also facilitates expression of the users goals and preferences into the decision support system to ensure they retain control of the criteria used during decision-making.

It is still possible, however, to deploy clinical decision-support applications without giving the user access to the facilities offered by an explicit decision model. There is already an example of this in the case of the **NHS Choices** symptom checker, which is designed to support members of the public when making first line decisions about whether to seek emergency or other help for common clinical problems (http://www.nhs.uk/NHSDirect/Pages/Symptoms.aspx). Like CREDO the knowledge base is written in PRO*forma* so the argumentation for the different options could be provided in a patient-friendly format. However, the designers appear to have decided that the users of the service need not have access to the reasons for the recommendations made by the software.

We can see some practical reasons for this, though the decision is open to debate. The NHS Choices advice system is of course a much simpler, consumer oriented, application than the services offered to healthcare professionals by CREDO, and it is a reasonable supposition that many users will not wish to bother with a detailed rationale for the recommendation. Providing this will simply add complexity and cost without significant public benefit. Nevertheless if healthcare agents are rolled out routinely without considering issues of autonomy then, as with automated trading systems, those without understanding and skills will be excluded from many benefits and the potential for failures and abuse seem likely to grow.

[4] The idea that responsibility for landing a plane full of people could be delegated to an automated system was once outrageous; today we take autolanders for granted, not least because they can get aircraft down safely in conditions that human pilots might not.

3. OPEN ACCESS, OPEN SOURCE AND OPENCLINICAL

In this section we continue the theme of what the responsibilities of healthcare systems implementers should be but we move the focus from the design and delivery of individual decision agents to considering the supply chain that will be needed to support the development and deployment of healthcare agents on an industrial scale.

Why is talk of "industrial scale" relevant here? The starting position is that good decision-making and intelligent service delivery are pivotal to delivery of high quality patient care in every medical sector (primary, secondary and tertiary) and every specialty (oncology, cardiology, endocrinology, accident and emergency and so on). We also know that throughout medicine healthcare professionals have too much to do, too much to know and too little time -- with the effect that errors occur at a worrying level, reducing clinical quality and eroding patient experience, and leading to avoidable patient harm and waste of resources.

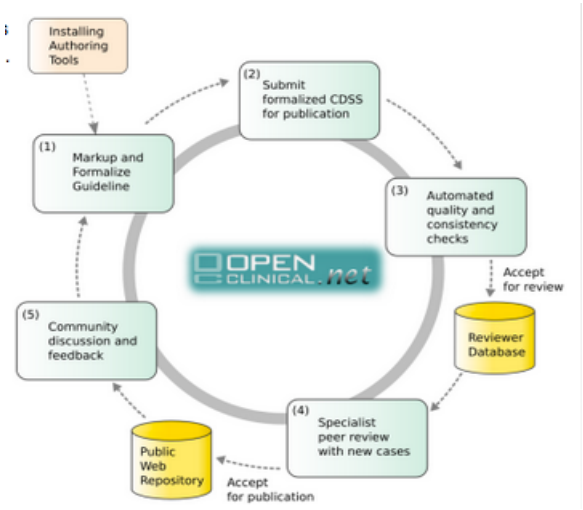
Clinical decision support systems can improve quality of decision-making; the evidence for this is now strong (e.g. Garg et al, 2005; Kawamoto et al 2005; Fox et al, 2006) and the last 10 years has seen a rapid acceleration of their commercial adoption and clinical penetration. As the variety and number of operational clinical decision systems continues to grow the opportunity to use them in autonomous mode will be increasingly recognized and is likely to be appealing because of the potential financial benefits as well as improvements in service quality and patient safety.

We believe that the ethical issues we have discussed in the context of the cancer multidisciplinary team are general, and are going to come upon us rapidly in coming years. Unfortunately every clinical setting is different, every application has different challenges, and there are no established methods or standards for developing and deploying decision *support* systems let alone for autonomous health agents. Furthermore the differences from one clinical situation to the next obscure the generic requirements and issues. The result is that most products and applications are "one-offs" and the academic research world and the commercial marketplace are fragmented.

The OpenClinical project was established in 2001[5] to begin to address this, first, by providing an information service and web portal for raising awareness and promoting good engineering practice and open standards for designing clinical decision support systems and medical knowledge management services (http://www.openclinical.org). It was intended later to develop a standardized and open access, open source repository of medical knowledge that could provide a scalable platform for creating, sharing and evaluating knowledge based services for supporting patient care (http://www.openclinical.net). The project has now reached the point where good governance has become a significant concern, and it is hoped that this paper will attract discussion of ethical issues in this context.

From OpenClinical.org to OpenClinical.net

OpenClinical offers an opportunity to support the standardization, formalisation and distribution of machine interpretable medical knowledge (see for example Peleg et al, 2005 and Peleg et al, 2010). OpenClinical.**net** was established in 2005 and has been developing steadily in a collaboration between Oxford University, University College London and the Royal Free Hospital[6].



OpenClinical.net knowledge authoring and publishing lifecycle

OpenClinical.net is presently based on two foundations: the open standard PRO*forma* agent modelling language, and the *Tallis* software platform, an "end to end" lifecycle for authoring and publishing open access and open source knowledge bases, as illustrated in the figure above.[7]

Repertoire: an open repository of applications

The OpenClinical.net knowledge base is called *Repertoire*. This is an open access and open source repository of chunks of knowledge that we call "publets" covering many medical specialties and types of clinical service. (The term is a contraction of "publication" and "applet".)

A design framework, development tools, editorial and knowledge management services, are now operational and we aim to make these free for use as soon as is practical. *Repertoire* presently

[5] The OpenClinical.org web site was under continuous development until 2011. Thanks to the editor, Richard Thomson, it had about 600 pages of material covering technical, clinical, and policy aspects of medical knowledge systems with many links to other sources, projects and demonstrations. The site was widely used for reference and teaching and at peak was receiving more than 350,000 visitors per annum (about 15% UK; 50% USA). The growing adoption of clinical decision support and other knowledge-based services and dramatic growth of eHealth over the last decade, however, has meant that it has become difficult for a tiny organisation to keep it up to date. Nevertheless we believe that it made a valuable contribution and we are now reconfiguring the project as OpenClinical.net and in a more sustainable form.

[6] With the cooperation of Deontics Ltd, a commercial spinout from Oxford, UCL and Cancer Research UK.

[7] Though we wish to emphasise that we aim to support other formats as well when this becomes practical see: http://www.openclinical.net/knowledge-publishing/content.html

includes about 50 example publets for different clinical tasks and in different specialties. Our vision is that anyone, working as an individual or under the auspices of professional bodies, healthcare providers and other organisations will be able to prepare knowledge of best practice in machine interpretable formats like PRO*forma*, and make them available as open access and open source publets on *Repertoire*.

Once the project is launched members of OpenClinical.net will be able to create and publish applications on *Repertoire* as illustrated by the lifecycle diagram. Existing applications can also be downloaded into the *Tallis* authoring system in order to adapt or re-purpose them for different healthcare settings, organisations or even countries. We hope to launch OpenClinical.net as a free service before the end of 2014.

4. FROM REPERTOIRE TO CHOREOGRAPHY

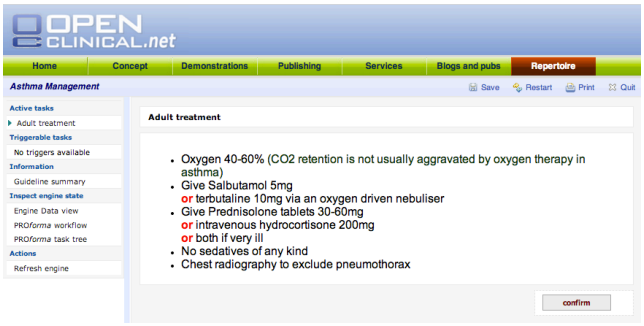
It is also very possible that *Repertoire* could offer a resource for research of various kinds, such as research on healthcare agents.

Scenario: Annie Boddie, an elderly lady with known diabetes and high blood pressure, complains to her doctor about some difficulty in breathing. Breathing difficulty is not caused by diabetes or hypertension or, to her doctor's knowledge, by the treatments that she is currently receiving for either condition. Dr. Smith considers starting Annie on a standard pathway for investigation, diagnosis and treatment of her complaint, but decides against this because it is not clear whether the standard pathway would be effective or safe in a patient with multiple conditions. Dr. Smith decides to use a care planning service *MyCareflow***,** available from *Agents Inc*. to provide guidance on the appropriate way forward. *MyCareflow* is a workflow planning system that searches for specialised care pathways and other resources on the web and uses them to assemble a patient-specific care plan. An important resource that *MyCareflow* frequently uses is the repository of evidence-based decision making and workflow models on the OpenClinical repository.[8]

Even at its present very limited size *Repertoire* could be viewed as a useful resource for *Agents Inc*. Annie's doctor wants to know whether the standard asthma care pathway would be appropriate. Given such a query the *Agents Inc.* choreography software comes into play. The *Agents Inc Choreographer* (*AIC*) queries *Repertoire* for available publets for diagnosis and treatment of possible asthma, to which it responds with a link to an *OpenClinical* publet that is a knowledge engineered version of a text guideline originally published by the British Thoracic Society and Scottish Guidelines Network. *AIC* then checks publet metadata and meta-knowledge for the application, which indicates that the provenance of the BTS/SIGN guideline is good, the OpenClinical.net application is current and it is appropriate for a patient with Annie's presenting symptoms.

AIC then spawns a manager agent for the publet, which uploads the knowledge base and initiates the appropriate workflow and decision support services for Annie's doctor (see screen shot below which is taken during a run off the application that is currently on *Repertoire)*.

Since *Repertoire* publets are declarative *AIC* can also search the knowledge and check whether there are any potential drug contra-indications or other issues of using this pathway for a patient who is also being treated for diabetes and hypertension. In this scenario there are no problems identified and the manager agent proceeds to assist Dr. Smith in making recommendations to Annie for her care.



Screen shot of decision support for medication decision part way through the OpenClinical asthma pathway

5. WHAT META-DATA AND META-KNOWLEDGE SHOULD *REPERTOIRE* EXPOSE?

The scenario is at this moment a fiction, but we believe that the capability it illustrates is technically within reach. However, in the context of this paper its purpose is only to reveal some of the challenges of having a knowledge base (or any number of knowledge bases from different providers) which might be accessed and used by autonomous agents, in order to safely plan and manage services for patient care.

The concept seems very timely since a growing public concern is with the growing problems of *co-morbidities* (providing high quality care for patients with multiple conditions) associated with the ageing population and *poly-pharmacy* (safe prescribing of multiple medications whose potential interactions may not be known to individual prescribers).

To date we have viewed the role of OpenClinical.net as providing a means by which individual clinicians and professional organisations can author and publish their specialist knowledge of best clinical practice in a machine readable form. We have also come to recognize, however, that *OpenClinical.net* could offer to our fellow researchers a platform for studying and experimenting with multi-agent systems in healthcare. I the present context, it also offers a setting in which to explore the kinds of ethical obligations that we might consider knowledge publishers of the future should be under.

The stakeholders to whom a knowledge publisher like OpenClinical.net would likely owe obligations to include at least the following:

[8] All the names in this scenario are fictional

- The recipients of care (the patients)
- Those who are professionally and legally responsible for efficacy and safety of care (the clinicians)
- The technical service providers (e.g. *Agents Inc*.)

One way of exploring what the publisher's ethical obligations are is to ask what metadata and meta-knowledge should be available alongside the primary knowledge content in the OpenClinical.net publet repository and that should be exposed to the stakeholders.

The metadata that it would seem appropriate to include is of two kinds. First, *static metadata*, including

1. *Documentation* comparable to the conventional Dublin core metadata (including citation information like Title, Creator, Author, Subject, Date, type etc)
2. *Assessments* that may relate to the *quality* of the publets (including scope, provenance, evidence grade etc)
3. A *safety case* might be required for every publet at an appropriate level of rigour (see discussion paper OpenClinical.org, 2002).

We might also propose to include *dynamic metadata* that are updated when a publet or other knowledge resource is accessed and used. Among the dynamic data that might be included are:

4. *Use data*: how many times has the publet been used, when, by whom etc?
5. *Outcome data*: patient information captured by a publet and clinical outcomes feedback provided by human users or software agents that track the patients' electronic records.

A rather different category of information that we may wish to expose for external review and make accessible to software agents that use *Repertoire* content is *meta-knowledge* or knowledge that an autonomous software agent can use to reason about the medical content and clinical use of a publet. Again we see opportunities for exposing two different forms of information.

1. *Descriptive meta-knowledge* is information that will permit an agent choreographer to reason about the relevance, appropriateness and safety of a plan or care pathway for a particular patient, by analyzing the representation of the plan itself. For example, in the asthma application discussed earlier the *AIC* should be able to understand the purpose or *goal* of the pathway that was envisaged by its authors; the *preconditions* that must be satisfied for a patient to be eligible or for it to be used safely, and *post-conditions* of using the application that may have implications for subsequent decision-making and care, or the diagnosis and treatment of concurrent problems.
2. *Context-updated meta-knowledge* is information that can be instantiated with patient specific data and other context information while the pathway is in use (e.g. beliefs, goals, rules, decisions and plans). Even though a pathway may have been selected appropriately to achieve a particular goal it may still be uncertain whether this was the best plan and the agent should be able to update its assessments of relevance, appropriateness and safety of the publet as the patient's circumstances change.

6. SUMMARY AND CONCLUSIONS

As this AISB50 symposium suggests I believe that it will soon be possible, and may be economically irresistible, to deploy healthcare agents that look after significant parts of our care without the oversight or supervision of human specialists. Achieving the kinds of scenario presented here will require new research and good engineering, but we believe that these kinds of capability are within reach. Our discussion of the ethical questions and possible approaches to addressing them is tentative as we cannot imagine all the opportunities or the risks that the deployment of autonomous healthcare agents will create. However, we believe that OpenClinical.net offers many ways of investigating these questions and look forward to it being used in this way.

7. REFERENCES

Beauchamp T L and Childress J L Principles of Biomedical Ethics Oxford University Press, USA 2008 (6th Edition)

Das S, Fox J, Elsdon D, Hammond P. "A flexible architecture for a general intelligent agent*" Journal of Experimental and Theoretical Artificial Intelligence*, 9, 407-440, 1997.

Fox J and Das S. *Safe and Sound: Artificial Intelligence in Hazardous Applications,* MIT Press and American Association of Artificial Intelligence, 2000.

Fox J, Beveridge M and Glasspool D "Understanding intelligent agents: analysis and synthesis", *AI Communications* 2003.

Fox J, Patkar, V, Thomson R; "Decision support for healthcare: the PRO*forma* evidence base" *Informatics in Primary Care*, 2006, 49-54(6).

Fox, J., Glasspool, D. W. and Modgil, S. (2006). "A canonical agent model for healthcare applications". *IEEE Intelligent Systems,* 21(6), 21-28.

Fox J, Glasspool D, Grecu D, Modgil S, South M, Patkar V, "Argumentation-based Inference and Decision-Making – A Medical Perspective" *IEEE intelligent systems*, Nov/Dec, 2007, 34-41.

Fox J, "Credo for decision science and engineering" (to appear) 2014

Garg A, Neill K, Adhikari J, McDonald H, Rosas-Arellano P, Devereaux P J,Beyene J, Sam J, Haynes B "Effects of Computerized Clinical Decision Support Systems on Practitioner Performance and Patient Outcomes: A Systematic Review" *JAMA*. 2005;293 (10):1223-1238.

Kawamoto K, Houlihan C A, Balas A, Lobach D L, Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success, *BMJ* 2005;330:765

OpenClinical, 2002 "Quality, safety and ethics in the use of computers to advise on patient care" Discussion Paper drafted for OpenClinical.org and available at http://www.openclinical.org/qseOCGreenPaper.html

OpenClinical, 2013 A platform and infrastructure for creating and sharing knowledge of best clinical practice and evidence---based decision making OpenClinical.net: White Paper 1, 2013

Patkar V, Acosta D, Davidson T, Jones A, Fox J, Keshtgar M "Using computerised decision support to improve compliance of cancer multidisciplinary meetings with evidence-based guidance" BMJ Open June 2012

Peleg M, Steele R, Thomson R, Patkar V, Rose T, Fox J "Open-Source Publishing of Medical Knowledge for Creation of Computer-Interpretable Guidelines" *Proc. AI in Medicine* Europe, 2005, Springer.

Peleg M, Fox J, Greenes R and Rafaeli S "Sharing Clinical Guideline Knowledge: can the dream come true?" Panel at Medinfo 2011, http://www.openclinical.org/sharingGLknowledge.html

Sutton D, and Fox J "The Syntax and Semantics of the PRO*forma* guideline modelling language". *J. Am. Med. Inform. Assoc.* 1264-1279, 2003.