https://doi.org/10.1038/s42256-022-00558-5

Developing moral AI to support decision-making about antimicrobial use

William J. Bolton, Cosmin Badea, Pantelis Georgiou, Alison Holmes & Timothy M. Rawson

The use of decision-support systems based on artificial intelligence approaches in antimicrobial prescribing raises important moral questions. Adopting ethical frameworks alongside such systems can aid the consideration of infection-specific complexities and support moral decision-making to tackle antimicrobial resistance.

Antimicrobials are drugs that kill or inhibit the growth of microorganisms. Their use or overuse is a major driver of antimicrobial resistance (AMR) in humans[1] – a leading global cause of mortality that killed more people than HIV/AIDS or malaria in 2019[2]. To address AMR, a multimodal approach is required that includes improving diagnosis, developing new antimicrobials and, critically, preserving the effectiveness of currently available agents. Recently, artificial intelligence (AI), machine learning (ML), and deep learning (DL) techniques, have shown great promise for providing decision support in healthcare[3,4]. Within the field of infection, models that support decision-making have been developed to select appropriate antimicrobials, predict COVID-19 outcomes[5–7] and anticipate AMR development[8,9]. However, the adoption and integration of such technology for antimicrobial prescribing and infection management remain challenging[10].

The development and adoption of AI-based decision-support tools to support improved antimicrobial use raises significant moral questions. Most notably, antimicrobial prescribing decision-support systems must, we argue, achieve a moral balance between the needs of an individual patient and those of the wider and future society. Currently, when making decisions about whether to prescribe antimicrobials to a patient, clinicians usually evaluate the potential risks and rewards[11,12]. Treating the patient may reduce suffering and disease progression, but risks driving the evolution of AMR as well as causing adverse effects[1,13]. Not prescribing antimicrobials has the opposite risk/reward profile. National and local guidelines are important in decision-making because they help clinicians follow evidence-based practice and understand when specific drugs should be administered in certain situations. However, in our experience the antimicrobial decision-making process is highly individualized and, unfortunately, is frequently undertaken with limited information, meaning it is unclear which antimicrobial the infectious pathogen is susceptible to, or even whether the patient could recover without treatment. This often results in difficult decisions that are not clear-cut, the consequences of which may or may not help the patient and may or may not harm future populations, and thus what decision is morally right is often unclear. Incorporating such concepts into AI systems is complex and if the 'interpretation problem'[14] is correct, then we may never be able to get perfect accuracy. Regardless, important progress can be made by working towards a consensus on the optimal approach to decision-making in general, and, especially in this context, a nascent field known as 'meta-decision-making for AI'[15]. In this Comment, we aim to explore potential ethical frameworks and nuances that may be applied to define what is ethical or not during the development of AI-based clinical decision-support systems (CDSSs) for antimicrobial optimization.

Moral paradigms in AI ethics

The question of how to build moral AI decision-making systems is very important. Building a decision system based on an underlying moral paradigm, called the top-down approach in AI, is one of the most significant ways of doing this[16]. The main moral paradigms employed in AI ethics are utilitarianism (a type of consequentialism), deontology and virtue ethics. Below, we compare the applicability of these different ethical theories to the moral use of antimicrobials to determine which approach is most appropriate to apply to AI decision-making systems in this scenario.

First, by taking the perspective of a utilitarian ethicist who believes that an action is only 'good' if it creates utility (frequently measured as happiness)[17], a moral balance may be achieved and applied to the development of AI-driven CDSSs. Utilitarianism in healthcare can be evaluated using several techniques. These include total, average, minimum and total-average utility[17]. Maximizing total-average utility is likely to be of greatest importance in the context of AMR and healthcare, given that it aims to optimise the average happiness for those people who are currently alive[17]. This aligns with the UK's General Medical Council (GMC) 'duties of a doctor' formulation whereby the objective is to maximise health and extend life for the patient being treated, while also considering wider society and providing equality[18]. Frameworks such as Bentham's felicific calculus are commonly used within utilitarianism and can be applied to complex healthcare questions to quantify the utility of an action[19]. Figure 1 provides an illustration of this calculus and its application to the decision to start antimicrobial treatment. When taken as a whole, this framework suggests that for prescribing antimicrobials to be justified, the intensity and duration of the utility gained for the individual patient must outweigh the negative effect on everyone else. The only utilitarian scenario in which this could occur is with what is known as a 'utility monster', who gains significantly greater utility from actions than others do, combined. This seems improbable to exist in practice, and thus, to maximize total average happiness, we must look to maintain, and promote, all life. Importantly, however, certainty and propinquity cannot be quantified without more information. AI models can therefore contribute toward utility evaluations by helping to estimate the effect of a particular agent on the development of AMR versus the likelihood of clinical efficacy. By combining Bentham's felicific calculus and AI-based CDSSs, we can quantify the

Variables	Description	Exemplar of starting antimicrobial treatment	Corresponding ad hoc utility value
Intensity	How strong is the pleasure?	Treating a relevant infection with antimicrobials has the potential to save that person's life	Highly positive utility
Duration	How long will the pleasure last?	Any extension of life is immeasurable, while it is reasonable AMR will continue in the near-term future	Positive utility
Certainty or uncertainty	How likely or unlikely is it that the pleasure will occur?	Limited information often means treatment may or may not be helpful and there is always an inherent risk of developing AMR	Neutral utility, without more information
Propinquity	How soon will the pleasure occur?	Treatment can be effective immediately; however, the same is true for the evolution of AMR	Neutral utility, without more information
Fecundity	The likelihood of further sensations of the same kind	–	Unable to assign
Purity	The likelihood of not being followed by opposite sensations	–	Unable to assign
Extent	How many people will be affected?	Prescribing antimicrobials affects the patient and those close to them, while the development of AMR is a certainty and may affect everyone, causing significant suffering and mortality	Immense negative utility

Fig. 1 | Overview of Bentham's felicific calculus variables and example application to starting antimicrobial treatment. The seven variables of Bentham's felicific calculus and their associated description are shown on the left. An example of using this algorithm to estimate the utility of the decision to initiate antimicrobial treatment is provided on the right.

utility of an antimicrobial prescription and understand the potential resulting individual and societal implications.

Deontological, or duty-based, ethics evaluates whether an action is morally good or bad based on whether one acts in accordance with one's duty[20]. Duty-based ethics is very amenable to rule-based decision-making; therefore, one could attempt to understand which perfect and imperfect duties could apply in this situation, based on, for instance, Kant's categorical imperative[21]. The UK's GMC 'duties of a doctor' and the Hippocratic oath can be considered relatively deontology-focused approaches, given that they take care of the patient as of primary concern[18], which is a deeply intentional and duty-focused value. However, these principles were designed to be universal and are not tailored for the significant ethical dilemma posed by antimicrobial prescribing.

Virtue ethics, in contrast, focuses on the moral character of the agent carrying out actions, an assessment of which can be made based on comparisons with a virtuous person who possesses and embodies the virtues[22]. In this context, virtues need to be defined, embedded in the moral agent, as previously described[14], and the question of 'What would a virtuous person do?' considered. One could argue that virtuous clinicians may act as moral exemplars for complex AI-supported decision-making[23], but common moral dilemmas arise in the context of decision-making to address AMR. For example, one might need to weigh the potential number of lives lost and the value of taking action to reduce the number of deaths versus taking personal responsibility for an individual's outcome, or consider whether acting in a high-pressure situation brings equal moral responsibility as not acting.

Finally, applied ethical theories can also be explored: for example, the four principles of medical ethics (autonomy, beneficence, non-maleficence and justice), which are commonly used as a platform upon which moral agents in healthcare should act[24]. When applying these principles at a societal level, one can consider that they should translate across time[25]. In this case 'justice', defined as the obligation of fairness in the distribution of benefits and risk, means that there is a responsibility to provide equal and fair care to everyone, no matter whether they are alive yet or not. Furthermore, as modern medicine only began in the nineteenth century with breakthrough discoveries from Louis Pasteur and Florence Nightingale, one can argue that the vast majority of individuals who will need care won't yet have been born. Therefore, for antimicrobials, which can be considered a finite and limited resource given the development of AMR, we must aim to optimise prescribing and reduce inappropriate use so that their associated benefits and risks are fairly distributed.

By comparing different ethical theories, we can infer that they may have contrasting viewpoints on what is considered morally right with regard to prescribing antimicrobials. We suggest that a utilitarian

approach is most appropriate for antimicrobial decision-making given the number of individuals potentially affected by AMR, and alignment between current best practice, Bentham's felicific calculus, and the principles of medical ethics indicating that antimicrobial resources should be fairly distributed.

Technological and clinical considerations for moral AI

Developing moral-AI-driven CDSSs that incorporate ethical frameworks may support the wider adoption of such systems. However, further technological and infection-specific clinical factors need to be considered[26,27]. Technical issues with AI such as transparency (explainability and interpretability)[23,28], bias[7], accountability[29] and adoption[10] have been extensively discussed in the literature, but what is morally acceptable from the perspective of an AI-assisted antimicrobial prescribing decision has yet to be fully defined. Model fairness is particularly important in the setting of infectious disease, where people of different ethnic backgrounds have different infection-related risks and outcomes, and research has shown a strong association between poor socioeconomic status, increased rates of infection and AMR[30,31], which has been emphasised during the COVID-19 pandemic[32].

Another critical consideration is, how should information be combined to reach a morally good decision? Data required for decision-making are unlikely to be processed through an individual model. For example, one system may output the anticipated antimicrobial risk or reward profile for the patient, while another produces a prediction for the likelihood of AMR development. As such, some sort of aggregation model, function[16] or an experienced clinician may be required, which gets particularly complex when patients' preferences are also taken into account, as is the case with shared decision-making[12]. These additional considerations must be accounted for by AI-driven CDSSs to provide prescribers with a high degree of confidence in their recommendations so that they can fulfil their 'duties of a doctor', while also incorporating, for example, a utilitarian societal benefit.

For infectious diseases specifically, additional factors must be investigated. Expert opinion on antimicrobial prescribing often deviates depending on the specific scenario, and relying on a single methodology such as an antibiogram can be unreliable or not correlate with expected response. This uncertainty hinders antimicrobial optimization and the development of AI-based CDSSs, but highlights why further research and holistic, well-balanced decisions are required. In addition, the evolutionary process of pathogenic microorganisms, and thus the development of AMR, occurs on a human life timescale. Hence, when tackling AMR, AI should be temporally dynamic so that it is sensitive to microbiological evolutionary changes. Furthermore, systems should be geographically revised given that infectious diseases, resistance rates and antimicrobial availability vary dramatically by region[1]. Moreover, heterogeneity is needed with regard to antimicrobial prescribing because uniform treatment drives AMR[1]. These factors increase the importance of algorithmic transparency and accessibility of live local medical data. Creating moral AI to support optimal antimicrobial prescribing is therefore very complex but remains a crucial endeavour to try and counteract AMR.

Towards moral AI decision systems

Moral frameworks have, by definition, been designed to help humans make ethical decisions. As our species advances into a new age with AI, we must consider how to ensure that such 'intelligent' decision systems promote moral decision-making. Regulators, users and developers should work through reasoning similar to that discussed above to determine what ethical theory is most applicable to AI decision-making in their specific scenario. Indeed, healthcare is particularly complex, and each speciality will have its own additional ethical and scientific factors that need to be investigated as part of moral AI. Similar to prescribing antimicrobials, other AI-supported decisions that potentially affect individuals beyond the immediate patient being treated are likely to entail comparable ethical dilemmas and arrive at allied conclusions. Organ donation is an example of this, in which medical resources are scarce and it is not possible to provide equal treatment to everyone[33,34]. As such, AI decision systems and policies must be carefully considered to ensure moral allocation. End-of-life care can also be considered here, given that death is strongly associated with a negative effect on others[19]. However, those dealing with choices that are almost entirely focused on the health of the patient, with limited external factors, such as decisions about knee replacement surgery or prescribing insulin for type 1 diabetes, are more likely to find that different moral paradigms, such as deontology, for example, may be appropriate. It is important to note, though, that understanding the potential effect of an AI-based decision on an individual patient, as well as on everyone else, is highly context and paradigm specific, and thus requires careful consideration.

Regarding antimicrobial decision-making, we believe a utilitarian approach is most suitable for developing AI-based CDSSs, and that focusing on the likelihood of drug effectiveness and that of resistance can have the biggest impact in supporting moral antimicrobial prescribing (Fig. 1). Furthermore, for antimicrobials, spatial and temporal considerations are critical to optimise treatment outcomes and minimise the development of side effects or AMR. Decision-making in antimicrobial prescribing is frequent, and both morally and technically complex; by applying ethical theories to specific scenarios and incorporating moral paradigms, we can ensure that AI-based CDSSs tackle global problems, such as the emerging AMR crisis, in a moral way.

William J. Bolton [1,2,3] ✉, Cosmin Badea[3], Pantelis Georgiou[1,4], Alison Holmes[1,5,6] & Timothy M. Rawson[1,5]

[1]Centre for Antimicrobial Optimisation, Imperial College London, London, UK. [2]AI4Health Centre for Doctoral Training, Imperial College London, London, UK. [3]Department of Computing, Imperial College London, London, UK. [4]Centre for Bio-inspired Technology, Department of Electrical and Electronic Engineering, Imperial College London, London, UK. [5]National Institute for Health Research, Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance, Imperial College London, London, UK. [6]Department of Infectious Diseases, Imperial College London, London, UK.
✉e-mail: william.bolton@imperial.ac.uk

Published online: 15 November 2022

References

1. Holmes, A. H. et al. *Lancet* **387**, 176–187 (2016).
2. Murray, C. J. L. *Lancet* **399**, 629–655 (2022).
3. Topol, E. J. *Nat. Med* **25**, 44–56 (2019).
4. Esteva, A. et al. *Nat. Med* **25**, 24–29 (2019).
5. Rawson, T. M., Ahmad, R., Toumazou, C., Georgiou, P. & Holmes, H. H. *Clin. Microbiol. Infect.* **23**, 524–532 (2017).
6. Peiffer-Smadja, N. et al. *Clin. Microbiol. Infect.* **26**, 584–595 (2020).
7. Wynants, L. et al. *BMJ* **369**, m1328 (2020).
8. Lv, J., Deng, S. & Zhang, L. *Biosaf. Health* **3**, 22–31 (2020).
9. Chindelevitch, L. et al. Preprint at http://arxiv.org/abs/2208.04683 (2022).
10. Rawson, T. M. et al. *Nat. Hum. Behav.* **3**, 543–545 (2019).
11. Brink, A. J. & Richards, G. *Curr. Opin. Crit. Care* **26**, 478–488 (Oct, 2020).
12. Butler, C. C., Kinnersley, P., Prout, H., Rollnick, S., Edwards, A. & Elwyn, G. *J. Antimicrob. Chemother.* **48**, 435–440 (2001).

13. Langford, B. J. & Morris, A. M. *Can. Pharm. J.* **150**, 349–350 (2017).
14. Badea, C. & Artus, G. Preprint at http://arxiv.org/abs/2103.02728 (2021).
15. Badea, C. & Gilpin, L. H. Preprint at https://doi.org/10.48550/arXiv.2210.00608 (2021).
16. Badea, C. Preprint at http://arxiv.org/abs/2109.03283 (2021).
17. Sinnott-Armstrong, W. in *The Stanford Encyclopedia of Philosophy* (ed. Zalta, E. N.) (Stanford Univ., 2021).
18. General Medical Council. https://www.gmc-uk.org/ethical-guidance/ethical-guidance-for-doctors/good-medical-practice/duties-of-a-doctor (April 2019).
19. Post, B., Badea, C., Faisal, A. & Brett, S. J. B. *AI Ethics* https://doi.org/10.1007/s43681-022-00230-z (2022).
20. Alexander, L. & Michael Moore, M. in *The Stanford Encyclopedia of Philosophy* (ed. Zalta, E. N.) (Stanford Univ., 2021).
21. Johnson. R. & Cureton, A. in *The Stanford Encyclopedia of Philosophy* (ed. Zalta, E. N.) (Stanford Univ., 2021).
22. Hursthouse, R. & Pettigrove, G. in *The Stanford Encyclopedia of Philosophy* (ed. Zalta, E. N.) (Stanford Univ., 2021).
23. Hindocha, S. & Badea, C. *AI Ethics* **2**, 167–175 (2022).
24. Beauchamp, T. L. & Childress, J. F. *Principles of Biomedical Ethics* (Oxford Univ. Press, 1979).
25. Meyer, L. in *The Stanford Encyclopedia of Philosophy* (ed. Zalta, E. N.) (Stanford Univ., 2020).
26. Lysaght, T., Lim, H. Y., Xafis, V. & Ngiam, K. Y. *Asian Bioeth. Rev.* **11**, 299–314 (2019).
27. Grote, T. & Berens, P. *J. Med. Eth.* **46**, 205–211 (2020).
28. Barredo Arrieta, A. et al. *Inf. Fus.* **58**, 82–115 (2020).
29. Price, W. N. II, Gerke, S. & Cohen, I. G. *JAMA* **322**, 1765–1766 (2019).
30. European Centre for Disease Prevention and Control. https://www.ecdc.europa.eu/en/publications-data/health-inequalities-financial-crisis-and-infectious-disease-europe (2013).
31. Alividza, V. et al. *Infect. Dis. Poverty* **7**, 76 (2018).
32. Williamson, E. J. et al. *Nature* **584**, 430–436 (2020).
33. Berrevoets, J., Jordon, J., Bica, I., Gimson, A. & van der Schaar, M. *Adv. Neural Inf. Process. Syst.* **33**, 20037–20050 (2020).
34. Persad, G., Wertheimer, A. & Emanuel, E. J. *Lancet* **373**, 423–431 (2009).

Acknowledgements
W.B. was supported by the UK Research and Innovation Centres for Doctoral Training in AI for Healthcare, http://ai4health.io (grant no. P/S023283/1).

Competing interests
T.M.R. was employed by Sandoz (2020), Roche Diagnostics Ltd (2021) and bioMerieux (2021–2022). These commercial entities were not involved in the design, collection, analysis or interpretation of data, the writing of this article or the decision to submit it for publication. All authors declare no other competing interests.

Additional information
Peer review information *Nature Machine Intelligence* thanks Zvonimir Koporc, Stuart McLennan and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.