

# Artificial Moral Advisors: A New Perspective from Moral Psychology

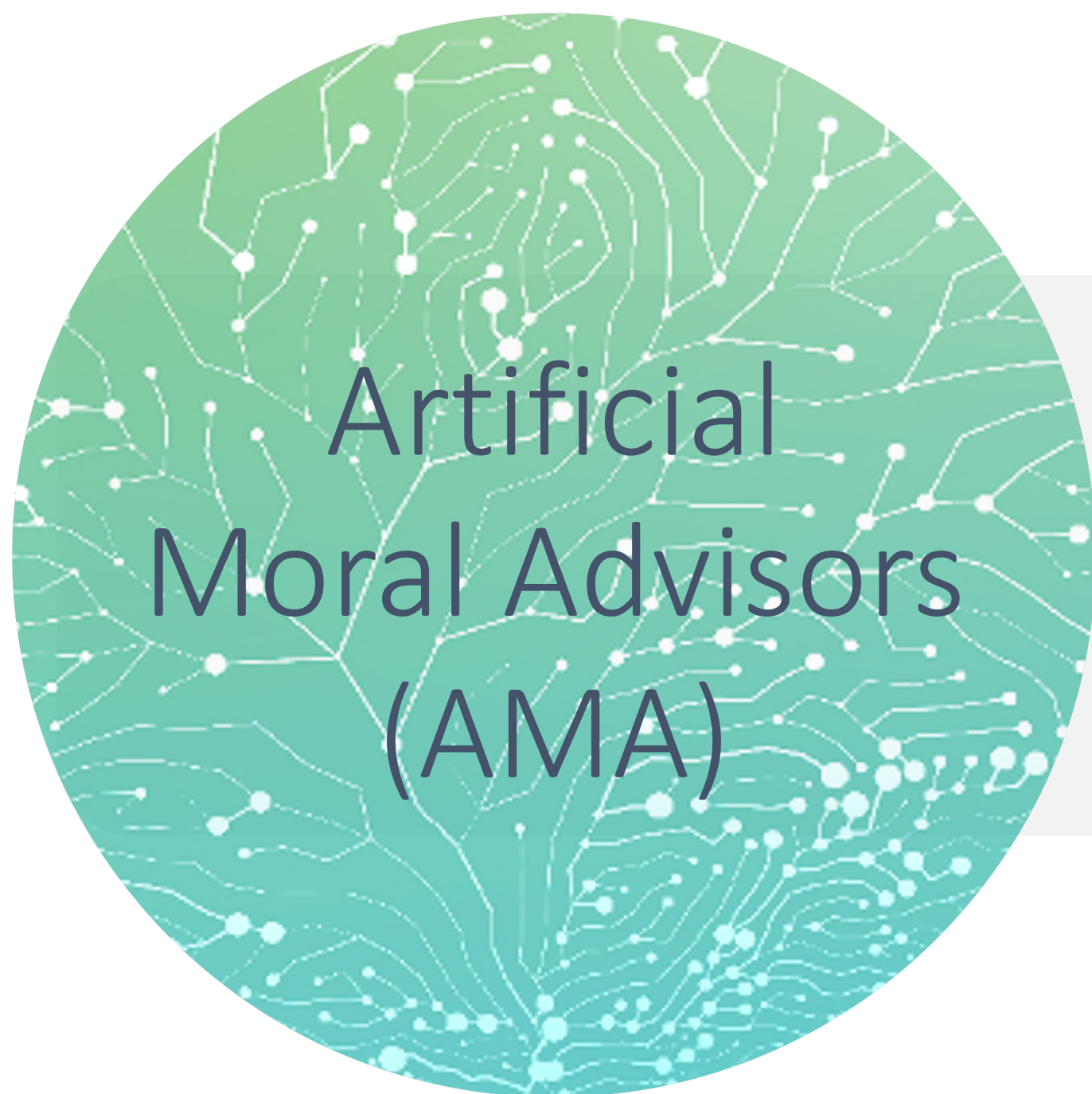
Yuxin Liu<sup>1,3</sup>, Adam Moore<sup>1</sup>, Jamie Webb<sup>3,4</sup>, and Shannon Vallor<sup>2,3</sup>

<sup>1</sup>Department of Psychology, School of Philosophy, Psychology and Language Sciences

<sup>2</sup>Department of Philosophy, School of Philosophy, Psychology and Language Sciences

<sup>3</sup>Centre for Technomoral Futures, Edinburgh Futures Institute

<sup>4</sup>Centre for Biomedicine, Self and Society, Usher Institute



Artificial  
Moral Advisors  
(AMA)

*"... a type of software that would give us moral advice more quickly and more efficiently than our brain could ever do, on the basis of moral criteria we input."*

Giubilini & Savulescu (2018, p. 171)

## AMA Internal Configuration

### AMA Moral Value Input

Lack of universal consensus:  
what morality to input?

Conflicts between principles &  
the lack of exception-less universals

Updating or shifting of moral values:  
how to keep track and update in time?

### AMA Information Processing

Dual-process of moral judgement:  
multiple framework of processing

Varying cognitive representations  
underpinning judgement & decision-making:  
no reference of true moral values

Lay theories of meta-level moral preferences

## Responding to AMA Moral Advice

### Acting More Morally?

Passive acceptance?  
Little role of the human

Ignoring the AMA:  
No motivation to act

Irreducibility &  
inescapability of moral  
decisions

Accepting/rejecting the AMA:  
responses to AMA are themselves by nature  
human moral judgments

Motivated cognition &  
selective info processing

Moral degradation

Exploitation of a polarising AMA

### True Moral Enhancement?

Utilitarian approach:  
most compatible with the AMA

Kantian perspective:  
AMA facilitating moral autonomy?

Existentialist account:  
AMA encouraging 'inauthentic' behaviour?

Virtue ethics tradition:  
AMA as a full moral exemplar?

### AMA in a more restricted context

Clinicians/physicians regularly face moral dilemmas  
Large pool of experts and data  
Principle-based approach: principlism

Poor record of performance  
Human judgements after all  
High-stakes risk of moral deskilling

Positive  
Use Case:  
A healthcare  
AMA?