

AI ethics can't come from human medicine: Principles that guide doctors would 'make no sense' to a machine

The four core principles of medical ethics are respect for autonomy (patients should have control over how they are treated), beneficence (doctors should act in the best interest of patients), non-maleficence (doctors should avoid causing harm) and justice (healthcare resources should be distributed fairly).

The more than 80 AI ethics reports published are far from homogeneous, but similar themes of respect, autonomy, fairness, and prevention of harm run through most. And these seem like reasonable principles to apply to the [development of AI](#). The problem, says [Brent] Mittelstadt, is that while principles are an effective tool in the context of a discipline like medicine, they simply don't make sense for AI.

...

AI has no equivalent to a patient, and the goals and priorities of AI developers can be very different depending on how they are applying AI and whether they are working in the private or public sphere.

...

As a result, AI ethics has focused on high-level principles, but at this level of abstraction they are too vague to actually guide action. Ideas like fairness or dignity are not universally agreed on, and therefore each practitioner is left to decide how to implement them.

Read full, original post: [Where Should AI Ethics Come From? Not Medicine, New Study Says](#)