PRECISION AUTONOMY

Infrastructure for AI

A Purposeful View of Ethical Autonomy

By Mark Halverson and Leanne Seeto of Precision Autonomy

Every technology breakthrough requires infrastructure supporting its safe integration into society. The horseless carriage would not have achieved ubiquity without infrastructure such as roads, fueling stations, auto insurance, traffic laws and signals, and so many other ecosystem services. Artificial Intelligence and Autonomy are in their infancy and currently lack the infrastructure required to be readily adopted in society.

The reality of 'Corporate Autonomy'

In migrating from the Information Age to the Intelligent Age, a dramatic amount of information is being aggregated to serve data monetization business models. This corporate version of Autonomy is upon us and the sad reality is that while as human beings (and consumers) we are materially impacted by those staging our thinking and intentionally biasing our decisions; and we have no insight into the rules upon which that Autonomy operates. And while we may dismiss the intrusion as 'marketing'; as a higher and higher percentage of our daily experience is delivered digitally (and therefore capable of scale impact/influence) the greater the need for transparency into how the Autonomy is making decisions. And of course as digital and physical worlds intersect the risk grows materially. An obvious example comes up in the philosophical debate as to whether an autonomous car may choose to kill the driver instead of hit a school bus, or run into a crowd of people. The reality is that (in the near term) the machine will follow the rules/law it has access to. Our challenge now as we move towards an Autonomy Economy is how to define those rules and ensure the Autonomy is operating 'On Purpose'.

An 'On Purpose' infrastructure will build trust and offer transparency into the operation of Autonomy. This paper explores such an 'On Purpose' infrastructure as a basis upon which Autonomy should operate.

For Precision Autonomy, the initial baseline ‘On Purpose’ model is linked to human decision making, for two key reasons

1) Any moral and ethical bounds placed around autonomy needs to be understandable, manageable, and auditable by humans, and

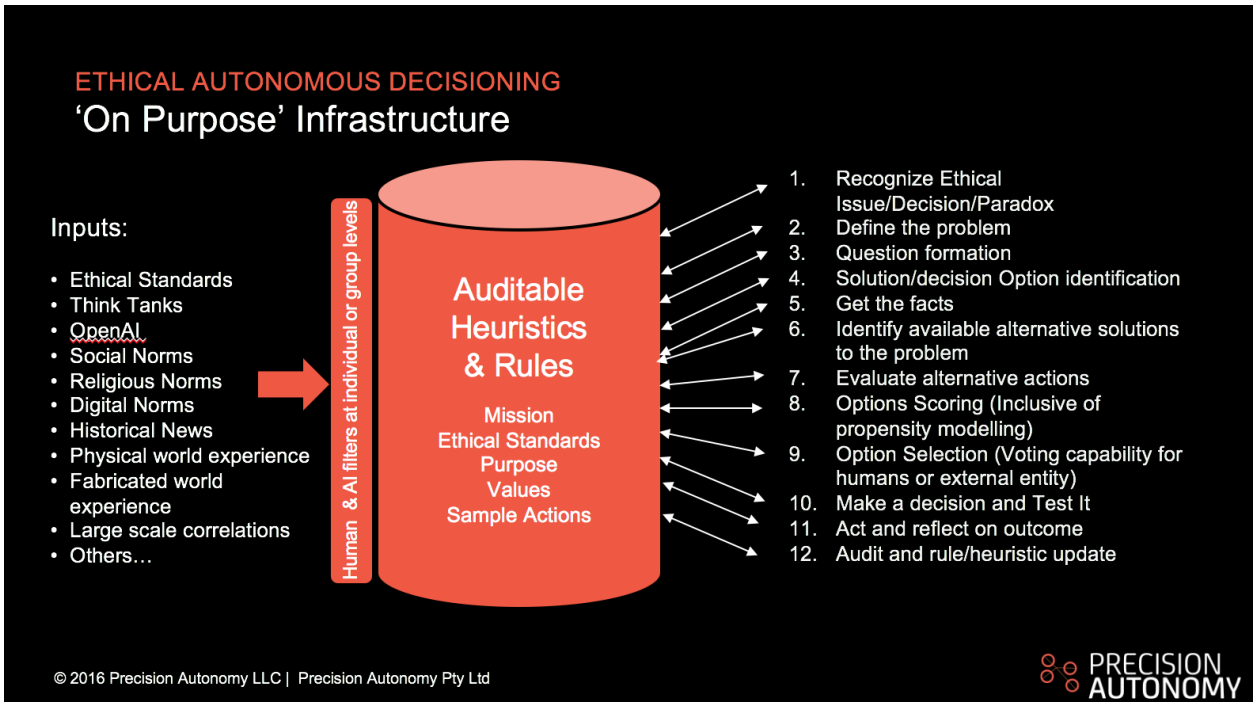
2) We seek to enhance and not replicate what humans do, addressing many of the shortcomings humans have in making ethical decisions and evaluating paradox (with the hope of autonomy establishing a type of ethical prosthetic).

Those who study the nature of human decision making can be disappointed in the malleability of the heuristics we use to govern most actions, and the inherent bias exhibited in our behavior. Therefore, we seek a model which enhances human ethical decision making and transparency providing appropriate controls around Autonomy. Some basic considerations would include:

- From the point of view of Autonomy, its world is a deterministic one bound by rules.
- The rules upon which Autonomy operates need to be transparent and agreed. This is by the target of Autonomy when it is directed (e.g. a consumer being positioned with an ad or offer). Or on behalf of the individual when Autonomy is acting as an agent for the individual.
- As part of transparency Autonomy should have an ability to act upon and expose its pre-disposition of approach in cases where pre-existing rules are insufficient (e.g. Utilitarian approach, Rights approach, Fairness or Justice approach, Common Good, Virtue, etc).
- Autonomy should act ‘on purpose’ where the full extent of the purpose is clearly articulated and therefore actions (and associated decision process) are auditable against the stated purpose.

As an aside, humans rarely expose or even understand their true intent and purpose, and therefore many human interactions are seeking insight into motives and values, in order to better predict decisions and actions.

Part of what is necessary is a hierarchy and taxonomy of how decisions will be made and how humans can influence ‘Autonomous’ decisions impacting them. The remainder of this document posits a model for registering Purposeful heuristics and rules from humans, informing Autonomy operating on their behalf.



Ambiguity informed by operator's intentions

- A registry is created allowing humans to 'inform' autonomous entities acting on their behalf. Similar to SIC (Standard Industrial Classification) codes for industries there would be a Wiki created that allows for Purposes in the form of heuristics and rules and their definitions to be maintained.
- Autonomous entities would make their decisions using the 'operator's' intentions and values via pre-registering by the operator of Mission, Standards, Purposes, and Values to better align a voting or propensity model to the operator's wishes.
- The model provides influence and transparency into the autonomy which serves us individually. There could also be default and modifiable settings or templates addressing broad themes (e.g. quantified self, eco-friendly, etc). At its core this creates a new set of 'self-defined' marketing segmentation attributes with the human in control and allowing themselves to be approached on their terms.

A case study: Corporate Autonomy acting upon us

Information is collected about a subject via apps, searches, purchases, and other collection mechanisms. That data is used by large corporations to position ads and offers to influence people to buy certain products. An individual who downloads the 'free' game application often unwittingly is agreeing to share a great deal of information which can then be used to influence their decision process in perpetuity. So the 'Purpose' of the game in this case is to capture information about a user to be resold to marketing organizations. This 'intent' is currently described within Privacy Policies for corporations. Then those marketing organizations use sophisticated algorithms to influence decision making by influencing experience and therefore heuristics of individuals. This is an example of algorithms/autonomy acting *upon* individuals in their role as consumer and not on their behalf.

A case study - An 'On Purpose' Registry acting on our behalf

An individual registers their key interest areas and conditions upon which they are comfortable being approached. Those companies wishing to get access to the individual would need to have a set of services which align to the same purposes/principles. So a user may indicate privacy as a key concern seeking to restrict any services with an intention to resell information about their behavior, contacts, etc. So if a user is seeking an App for gaming, they would only see those games which have been filtered via the individual's AI agent which will read privacy policies of games of interest and only present those which adhere to the privacy purpose.

Similarly, an individual can register their conviction towards utilitarianism (actions are right if they are useful or for the benefit of the majority, greatest happiness for the greatest number). In this case an autonomous drive vehicle (or any other Autonomous capability) could query the registry when operating upon or on behalf of an operator and have a basis to take actions commensurate with the operator's ethical philosophy. This information would be factored into decision option scoring and selection in the ethical decision process. So if coming across the situation where harming a driver versus potentially injuring many on a school bus, the vehicle would be able to take the utilitarian action and minimize impact to the smallest number of people. The vehicle is then in a position to demonstrate upon audit that it took the action based on the operator's wishes; insurance and legal transparency into the Autonomy's actions would exist. Equally there would be definitions and cause of action based on a Rights Approach, Fairness/Justice Approach, Common Good, Virtue, or other philosophical positions.

Only a first step

Clearly this is only a first step in establishing controls for AI acting on a human's behalf in ambiguous situations when overt rules aren't available. By implementing such a registry system as AI Infrastructure we allow Autonomy to state its purpose so as not to deceive, and query key considerations to allow influence from the wishes of its operators. This is a first step towards creating virtual 'rules of the road' for AI and Autonomy managing ambiguity while maintaining transparency and trust.

Of course there will need to be additional enhanced models of control as we continue to mature the concepts of AI and Autonomy. With one definition of AI as representing those problems which have yet to be programmatically addressed (i.e. something ceases to be AI when a computational capability addresses the problem space, such as playing chess, driving a car, playing Alpha Go), putting controls and moral and ethical bounds around AI and Autonomy will continue to evolve. And therefore the infrastructure will have to commensurately be enhanced.

So in parallel with continuing to admire the larger philosophical problems we can begin with a simple 'On Purpose' registration infrastructure that can offer us some level of comfort and control.