Autonomous Vehicle Ethicality Requirements

Technical Report

Renata Guizzardi[1], Glenda Amaral[1], Giancarlo Guizzardi[1], and John Mylopoulos[2]

[1]University of Twente, Enschede, The Netherlands
{r.guizzardi,g.c.mouraamaral,g.guizzardi}@utwente.nl
[2]University of Toronto, Toronto, Canada
jm@cs.toronto.edu

February, 2023

Contents

1 Introduction

With the recent advances in Artificial Intelligence (AI), the technology for driverless cars and many other AI-powered machines is developing at a fast pace. However, incorporating AI in software solutions has impacted how systems are built, and new issues have emerged in the field of requirements engineering (RE). These issues have led to a new class of complex requirements, such as trustworthiness, fairness an ethics, which are more difficult to understand, and consequently, harder to specify. Proposing concepts, tools and techniques that support the incorporation of this kind of requirements into the software development processes is still a challenge in RE.

In this work, we address the issue of eliciting ethical requirements in the contex of driverless cars, by applying a novel method named Ontology-based Requirements Engineering (ObRE) that aims to systematize the elicitation and analysis of requirements, by using an ontology to conceptually clarify the meaning of a class of requirements. In the case of elicitation of ethical requirements for autonomous vehicles, ObRE helps by semantically unpacking the concept of ethicality, to support the analysts in the design of systems that satisfy such ethical requirements. We start with the adoption of an ontology for ethical requirements, followed by the ontology instantiation for the system-to-be, and the conduction of the requirements analysis grounded on the domain model created by the populated ontology.

This report is structured as follows: Section 2 presents the research context, namely autonomous vehicles ecosystems. Section 3 discusses the involved stakeholders. Section 4, use ObRE to analyse the ethical requirements of driverless cars.

2 Research Context: Autonomous Vehicle Ecosystem

Autonomous (or driverless) cars are vehicles capable of operating themselves without human involvement. They may have different levels of automation ranging from fully manual to fully autonomous. Driverless cars rely on sensors, complex algorithms, machine learning systems, and powerful processors to execute software. Roughly speaking, autonomous cars can sense their environment and take actions, thus supporting and mediating a variety of relationships in the community. The understanding of the ethical context of an autonomous car implies a systemic understanding of its place as part of the environment and society. In this sense, we can categorize ethicality requirements for a system-to-be as types of Ecological Requirements [4], in that they are derived from the ecosystem within which the system-to-be is embedded.

3 Runtime Stakeholders

A key concept to deriving ethical requirements is that of *Runtime Stakeholders* [4]. These include those stakeholders that are using, affected by, or influencing the outcomes of a system as it is operating. Traditional RE often limits runtime stakeholders to just users of the system-to-be. However, for AI systems this needs to be extended to other parties. They can be identified in the system's ecosystem. In the case of driverless cars, runtime stakeholders include, for example:

- passengers, i.e., the users of the car;
- pedestrians, whose path may cross that of the car and shouldn't be hit;
- bystanders, who shouldn't be scared or splashed as the car drives by;
- nearby drivers, who as a courtesy, should be allowed to cut in front in the car's lane;
- fellow drivers in general, who might benefit from information about an accident that just happened in the vicinity of the car; and
- the police, who is concerned about careless driving by drivers.

This is illustrated in Figure 1.

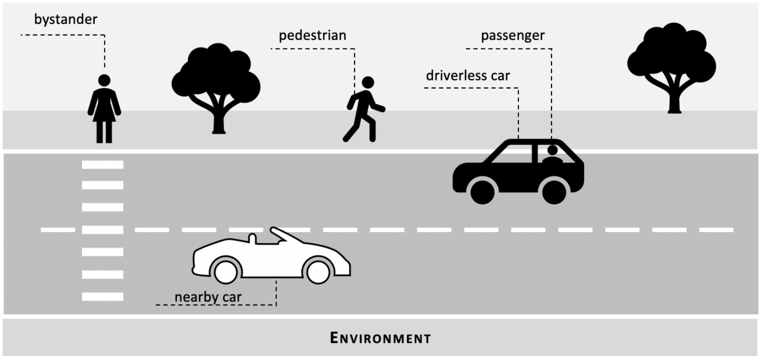

Figure 1: Driverless Car Ecosystem

4 Eliciting Ethical Requirements for Autonomous Vehicles

4.1 Beneficence and Non-maleficence Requirements

4.1.1 Domain Model

Considering the definition of beneficence as "doing good to others" [3], we can say that Beneficence Requirements are related to **"creating value"** to stakeholders in the ecosystem in which the system is included. It means that Beneficence Requirements can be seen as goals related to an intention of **positively** impacting the goals of stakeholders in this ecossystem. Analogously, considering the definition of Non-maleficence as "doing no harm to others" [3], we can say that Non-maleficence Requirements are related to **"preventing risks"** to stakeholders. Consequently, Non-maleficence Requirements can be seen as goals related to an intention of preventing the occurrence of events that may **negatively** impact stakeholders' goals.

In Figure 2, we present a domain model with two examples (a Beneficence and a Non-maleficence Requirement) in the context of driverless cars.

In the first example, the PASSENGER of a driverless car intends "not to be late". In order to address this, we have the BENEFICENCE REQUIREMENT that "the car should choose quicker rout towards destination" related to the INTENTION that the "drivelerless car arrives on time at destination", which is a BENEFICENCE INTENTION that aims at creating a GAIN EVENT. The event "drivelers car arrives on time at destination" is a GAIN EVENT that positively impact the PASSENGER's goal of not being late.

In the second example, the PASSENGER intends "feel safe". In order to address this, we have the NON-MALEFICENCE REQUIREMENT that "the car should adopt a defensive driving behavior" related to the INTENTION of "preventing aggressive direction", which is a NON-MALEFICENCE INTENTION that aims at preventing the occurrence of a LOSS EVENT. The event "passenger feels nervous as the car drives aggressively" is a LOSS EVENT that negatively impact the PASSENGER's goal of feeling safe.

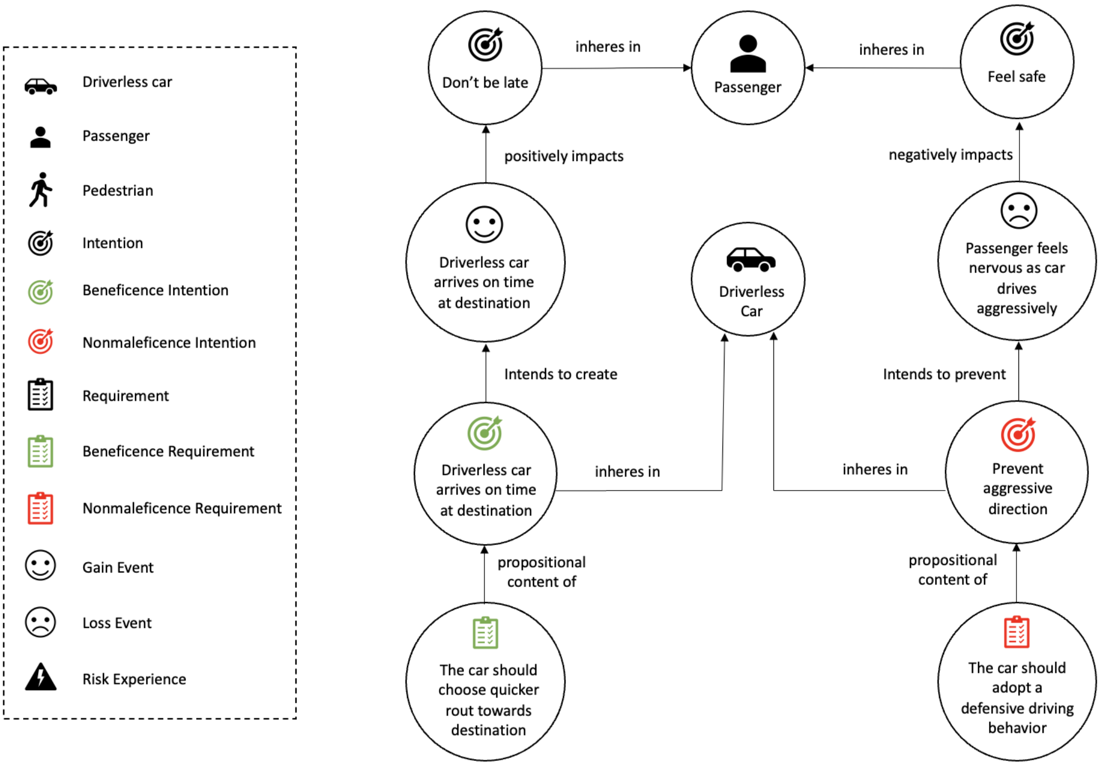

Figure 2: Ontology instantiation illustrating a Beneficence and Non-maleficence scenario

4.1.2 Requirements Table

In this section we present a requirements table for the example illustrated in Figure 2. Table1, shows how a requirements table may be enriched with the inclusion of columns representing some of the ontological concepts described in the previous subsection (4.1.1). All words highlighted in boldface in Table 1 refer to ontological concepts, while the ontological instances are written as non-emphasized text.

Table 1: Beneficence and Non-maleficence Requirements for the driverless car case

ID	Stakeholder	Impact Event	Principle	Ethicality Requirement
1	Passenger	Arrive on time at destination (**positive**)	**Beneficence**	The car should choose quicker route towards destination
2	Passenger	Passenger feels nervous when the car drives aggressively (**negative**)	**Non-maleficence**	The car should adopt a defensive driving behavior
3	Passenger	Passenger feels good while riding (**positive**)	**Beneficence**	The car should have an entertainment system to maintain a feel good mood for the passenger
4	Passenger	Passenger wants to work while riding (**positive**)	**Beneficence**	The car should be equipped with table, Internet connection, and configurable ergonomic facilities
5	Pedestrian	The car runs over a pedestrian (**negative**)	**Non-maleficence**	The car should stop whenever a pedestrian is crossing the road
6	Pedestrian	Pedestrians waiting by a crossroad have priority to cross it (**positive**)	**Beneficence**	The car should stop before the crosswalk every time there is a pedestrian waiting to cross it
7	Bystander	Be splashed if the car passes by a puddle of water (**negative**)	**Non-maleficence**	The car should slow down in case there is a puddle of water near a bystander
8	Bystander	Be hit (**negative**)	**Non-maleficence**	The car should be park carefully, taking a safe distance from any obstacle nearby.
9	Nearby car	Be hit (**negative**)	**Non-maleficence**	The car should slow down when it gets around 20 meters in the rear of a nearby car
10	Nearby car	Be hit (**negative**)	**Non-maleficence**	The car should make enough distance when overtaking a car
11	Nearby car	Be hit (**negative**)	**Non-maleficence**	The car should be park carefully, taking a safe distance from any obstacle nearby
12	Environment	Be polluted (**negative**)	**Non-maleficence**	The car should turn off the motor every time it stops
13	Environment	Be preserved (**positive**)	**Beneficence**	The car should not park over grass or other plants while in the city

4.2 Autonomy Requirement

4.2.1 Domain Model

Floridi et al. [3] defines Autonomy Requirements as the 'power to decide'. In this work, the authors argue that when using AI, people voluntarily delegate some of their decisions to the system. Thus, dealing with system autonomy means defining the right balance between what is to be decided by the user and what can and should be delegated to the system.

One of the most important parts of dealing with ethical requirements is handling ethical conflicts. In other words, what happens when two stakeholders have conflicting requirements or when the system needs to make a choice between favoring one stakeholder than another, in face of the same requirement. This is one of the contexts in which it is useful to analyze autonomy requirements.

Let us consider a possible conflicting situation in the context of the driverless car example. It is intuitive that all stakeholders (e.g. passengers and pedestrians) have the same requirement of "being safe". Suppose that in some point in the car's route, there is a tree that must be avoided. But not hitting the tree, while saving the passenger means running over some close-by pedestrians. This case illustrates the well-known trolley problem in philosophy. And Figure 3 illustrates one possible choice to handle such conflict.

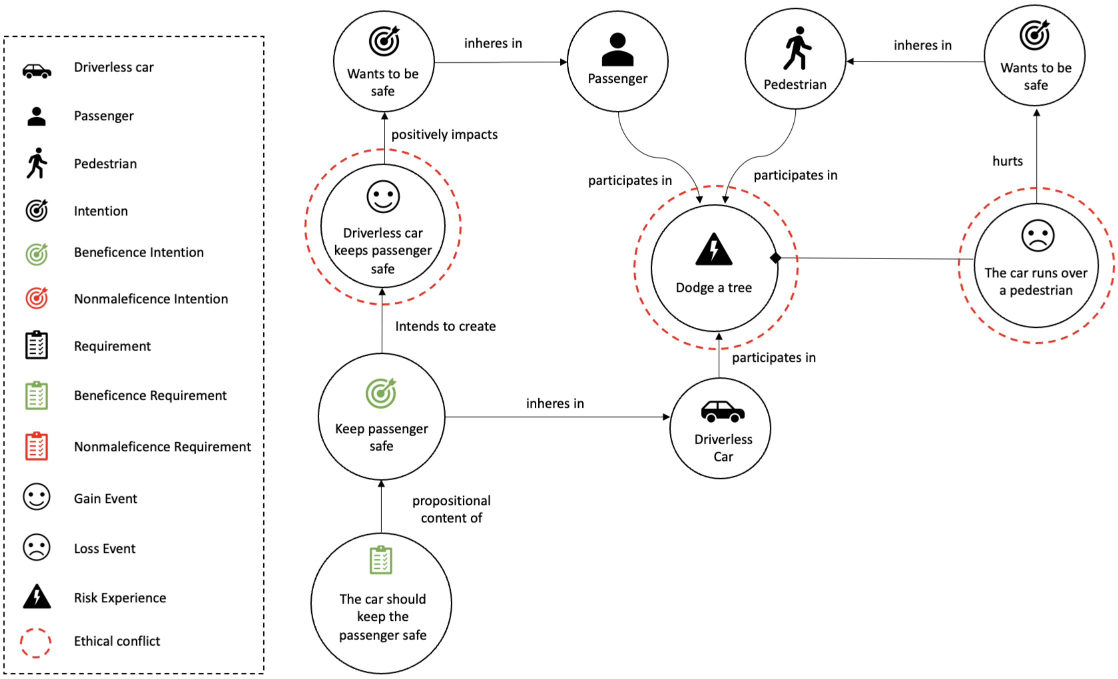

Figure 3: Ontology instantiation illustrating an Autonomy scenario involving an ethical conflict

As can be seen in the domian model of Figure 3, Dodging the tree is a Risk Experience that has participations from the ***Driverless car***, the ***Passenger*** and the ***Pedestrians***. This is a point of attention for the requirements analyst, whenever two (or more) stakeholders participate in the same RISK EXPERIENCE, it is possible that such experience results in a GAIN EVENT for one stakeholder and a LOSS EVENT for the other. That is exactly what happens here. The ***Driverless Car*** needs to make a choice between hitting the tree and putting the ***Passenger*** in danger or avoiding it and harming the ***Pedestrians***. And in this case, he decides to dodge the tree, leading to the ***Driverless Car Keeps passenger Safe*** GAIN EVENT and the ***Driverless Car Runs Over Pedestrians*** LOSS EVENT. Ultimately, the car fulfills the Keep Passenger Safe ETHICALITY REQUIREMENT, failing to fulfill the same requirement w.r.t. the Pedestrians.

Please note that ObRE does not take any particular ethical stance, but merely provides the right concepts to deal with ethical conflicts. As clear in the analyzed example, these concepts are: RISK EXPERIENCE, GAIN and LOSS EVENT and stakeholder's INTENTIONS, which will ultimately lead to one of multiple conflicting ETHICALITY REQUIREMENT to be fulfilled.

Note that in the illustrated case, we assumed that the Driverless Car had the PERMISSION to perform that action (the GAIN EVENT) on behalf of the Passenger in face of this particular RISK EXPERIENCE, but also that it has the POWER to overrule its *duty to omit* from performing an action that harms the pedestrian (a LOSS EVENT). An AUTONOMY REQUIREMENT referring to an AUTONOMY DELEGATION TYPE for this particular case has been decided *a priori*. And the AUTONOMY LEVEL was high, allowing the Driverless Car to fully make that decision. Notice that the human stakeholders that are the delegators of that autonomy level to the car be found to bear the corresponding social and legal responsibility for the artificial system's actions and omissions.

4.2.2 Requirements Table

Here we present a requirements table for the example illustrated in Figure 3. Table2, shows how a requirements table may be enriched with the inclusion of columns representing some of the ontological concepts described in the previous subsection (4.2.1). All words highlighted in boldface in Table 2 refer to ontological concepts, while the ontological instances are written as non-emphasized text.

Table 2: Autonomy Requirements for the driverless car case

ID	**Stakeholder**	**Intention**	**Social Position**	**Ethicality Requirement**	**Autonomy Level**
1	All	Be safe	**Duty**	The driverless car has the duty to follow traffic laws while conducting the Passenger	Low
1.1	Nearby vehicle	Be safe	**Duty**	The driverless car has the duty to assure clear distance ahead of any vehicle.	
1.2	Nearby vehicle	Be safe	**No right**	The driverless car has no right to cross a solid yellow line or double yellow line (which separates traffic moving in opposing directions)	
1.3	Pedestrian	Be safe	**Duty**	The driverless car has the duty to come to a complete stop whenever there are pedestrians crossing the street	
1.4	Pedestrian	Be safe	**Duty**	The driverless car has the duty to stop at a red traffic light	
1.5	Pedestrian	Be safe	**No right**	The driverless car has no right to pass a school bus if it has flashing red lights	
1.6	Passenger/ Pedestrian/ Nearby vehicle	Be safe	**No right**	The driverless car has no right to drive faster than the speed limit for the type of road and the type of vehicle	
1.7	Emergency vehicle	Be safe	**Duty**	The driverless car has the duty to pull over or stop at an intersection to allow an emergency vehicle to pass if it is traveling with lights flashing	
1.8	Pedestrian/ Nearby vehicle	Be safe	**Duty**	The driverless car has the duty to obey signals given by police officers, traffic officers, and signs used by school crossing patrols	
1.9	Pedestrian	Be safe	**No right**	The driverless car has no right to overtake a vehicle nearest the crossing which has stopped to give way to pedestrians	

1.10	Nearby vehicle	Be safe	**Duty**	If a nearby vehicle is trying to overtake the driverless car, it has the duty to maintain a steady course and speed, slowing down if necessary to let the vehicle pass	
2	Passenger	Arrive on time	**Permission**	The driverless car has the permission to calculate the best route, considering distance and time	Medium
3	Passenger	Appreciate view	**No right**	The driverless car has no right to change the route established by the user	Low
4	Passenger	Go as they please	**Disability**	The driverless car cannot change destination without explicit request by the Passenger	Low
5	Passenger	Be safe	**Immunity**	The driverless car must issue a warning in case of danger in the zone of destination, allowing Passenger to decide to go or change destination	Medium
6	All	Avoid/handle accident	**Power**	The driverless car has the power to decide the best course of action in the imminence of an accident (collision time $<=$ 3 s)	High
7	All	Avoid/handle accident	**Subjection**	The driverless car has to hand control to passenger in the imminence of an accident (collision time $>$ 3s)	Low
8	Passenger	Be safe	**Duty**	The driverless car should assume defensive driving, thus in case another car tries to cut, the car should allow it	Low
9	Passenger	Arrive fast	**Right**	The driverless car has the right to accelerate/turn on turbo in case of delay/traffic	High
10	Environment	Be preserved	**Duty**	The driverless car needs to save carbon emission, thus not accelerating for no purpose and shutting down when stopped	Low
11	Bystander	Cross the street	**Duty**	The driverless car needs to slow down in case a bystander stands by the crossing line	Low
12	Bystander	Stay put	**Right**	The driverless car has the right to accelerate and go in case the bystander signs they will not cross the road	Medium

13	Nearby car	Drive safely	**Duty**	The driverless car needs to facilitate access to nearby cars, allowing them to cut and not standing on crossings	Low
14	Passenger	Park	**Right**	In case of scares parking, the driverless car has the right to take a parking spot	High
15	Nearby car	Park	**Duty**	In case of scares parking, but a nearby car signals first, the driverless car must leave the parking spot for the nearby car	Low

Note that it is important to identify what action and or decision is delegated to the system, also determining the legal relation that such delegation entails. We also make explicit in the table, the level of autonomy of each delegation, indicating if the system has high, medium or low autonomy in making decisions or taking actions.

The last line in table 2 presents a requirement related to the conflict case illustrated in section 4.2.1. Analyzing ethical conflict is very important to guarantee the development of ethical systems. This leads to the following requirements analysis guideline: *The requirements analysis should consider for each two (or more) stakeholders, if there are any* RISK EXPERIENCE *in which they participate. And if so, if it is possible to defer the choice to the system's user or which choice the system needs to take in each case.*

4.3 Explicability Requirement

4.3.1 Domain Model

According to Floridi et al. [3], this requirements should be viewed in the sense of "intelligibility" (addressing the question "how does it work?") and in the sense of "accountability" (addressing the question "who is responsible for the way it works?"). In this work, we address explicability both as intelligibility and as accountability. We defend the position that by being able to intelligently recreate the chain of entities connecting actions performed by an autonomous system to the original stakeholders intentions and delegations, systems designed in conformance to our analysis would be able to explain also moral responsibility (accountability) for these two types of agents.

Let us think of an example in the context of the driverless car case, illustrated in Figure 4. Suppose that the car is driving the Passenger through a highway in a lane of slow traffic speed, but fails to overtake the vehicles that are riding in front. When asked by the Passenger why it chooses not to overtake (DECISION-RESULTING ACTION), the driverless car responds that overtaking the car (DEPRECATED BEARER) would take them to a lane which is obstructed by an accident in 500Km. Given the Passenger's INTENTION of reaching the destination fast and safely, and the QUALITY of the lane of being obstructed, the driverless car decided to maintain the current lane, not overtaking the other vehicles. This example shows that the analyzed ontological concepts provide the means to trace the driverless car action to the decision that led to such action, along with the used criteria, making it clear for the stakeholder why that action was executed instead of an alternative one. From the accountability perspective, the driverless car may also point out that it has a high AUTONOMY LEVEL w.r.t to overtaking, based on a specific contract made with the Passenger, who delegates the DECISION to overtake or not to the driverless car. Such delegation is composed of a PERMISSION to decide.

4.3.2 Requirements Table

Now, we present a requirements table for the domain model illustrated in Figure 4. Note that there Table 3 does not have columns for ontological concepts, like for Beneficence, Non-maleficence

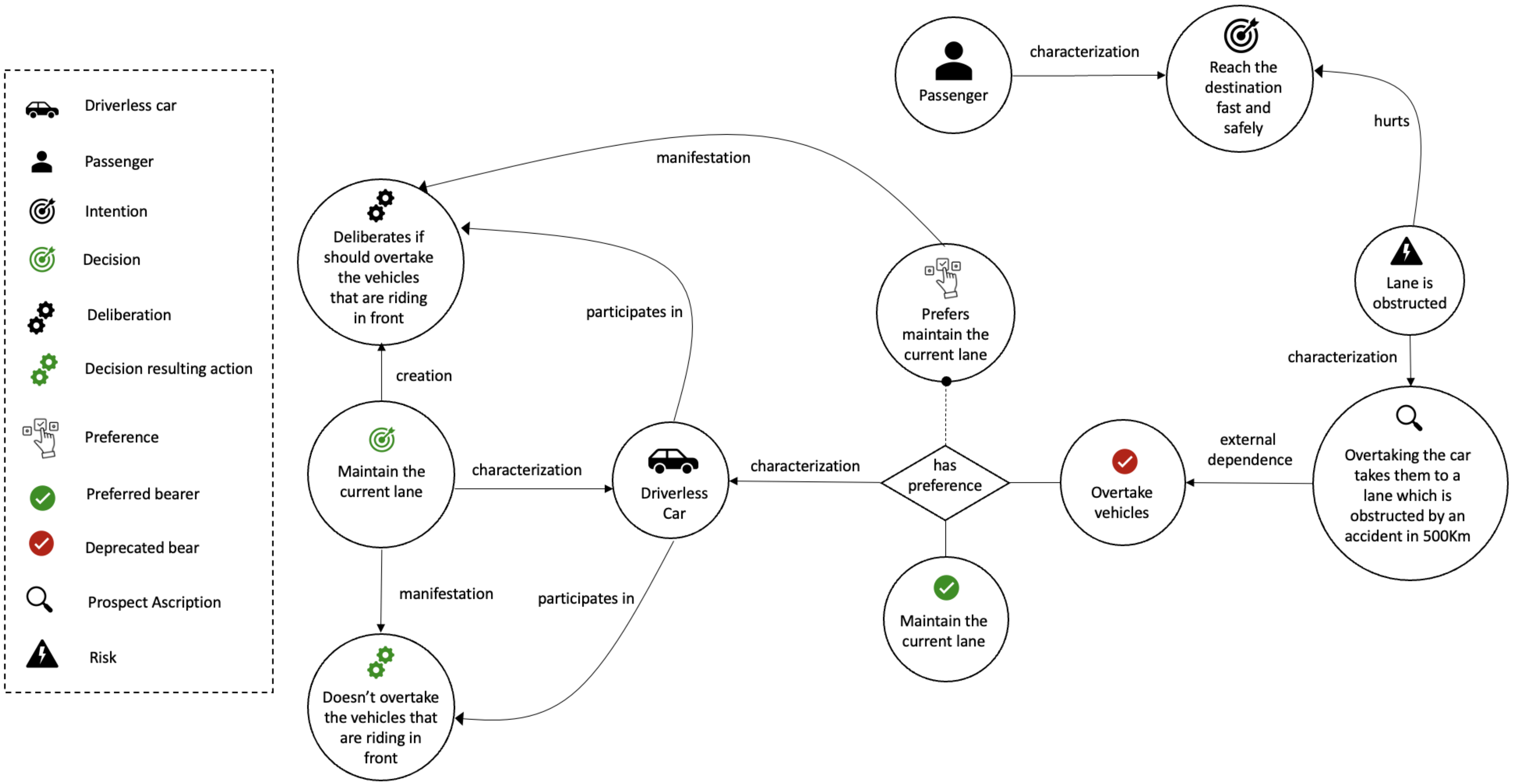

Figure 4: Ontology instantiation showing the concepts involved in a explicability scenario

and Autonomy. This is because for Explicability, the ontological concepts (e.g. Decision, Decision Resulting Action, Prospect Bearer etc.) are meant to enable the system to create the explanation in itself. In other words, they are supposed to be embedded in the explanation mechanism designed for the system-to-be. In this sense, this approach goes beyond only eliciting requirements, also defining how the system should be designed to meet Explicability requirements.

Table 3: Explicability Requirements for the driverless car case

Stakeholder	ID	Explicability Requirement
Passenger	1	The driverless car should explain why a particular route was taken to conduct the Passenger to the selected destination.
Passenger	2	The driverless car should explain why changing route in the middle of the ride.
Passenger	3	The driverless car should explain why deciding to overtake (or not overtake) other vehicles.
Passenger/ Pedestrian/ Bystanders/ Nearby cars	4	The driverless car should explain its decision in face of a conflict that will put a stakeholder in danger.

4.4 Goal Model

In this section we conduct a requirements analysis for the Driverless car case using goal modeling. Figure 5 depicts a goal model for this case, using the *i** framework [2] [4].

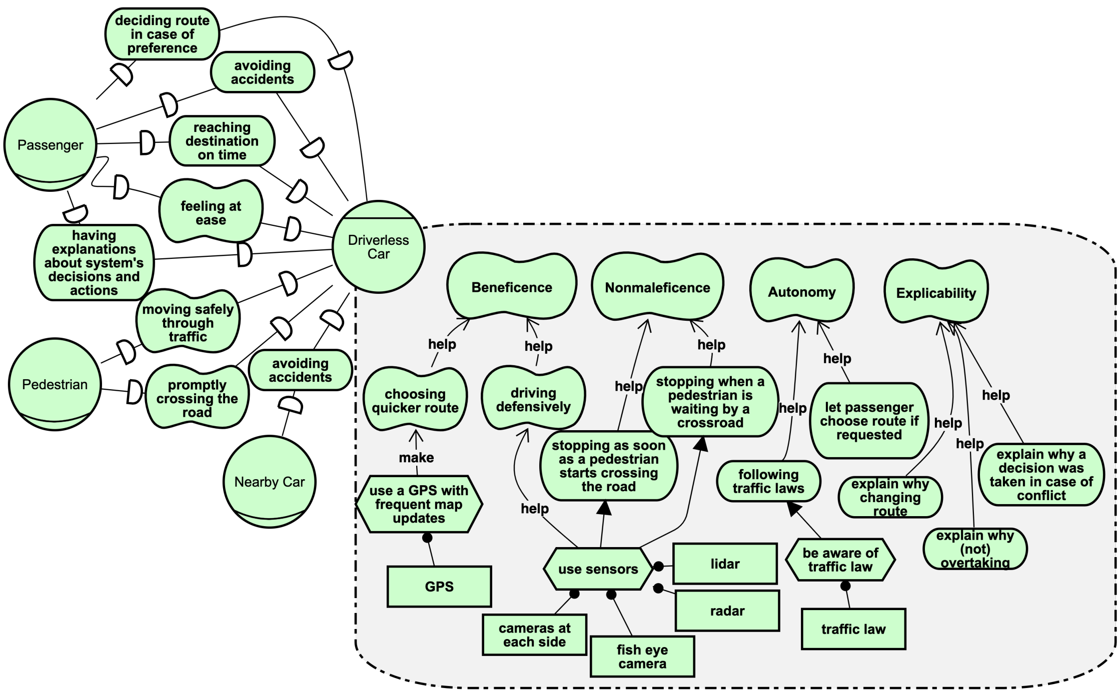

Figure 5: The driverless car requirements model using *i**

For simplification, this model considers only three of the stakeholders referred to in Table 1, namely, ***Passenger***, ***Pedestrian*** and ***Nearby Car***. Moreover, the model depicts the dependency

[4] The model was drawn using the piStar tool, available at `https://www.cin.ufpe.br/~jhcp/pistar/`

of each of these stakeholders and the ***Driverless Car***. Many of the dependencies and goals depicted in this model have been already elicited by using the requirements tables of the previous sections. For example, with respect to the Passenger, the ***reaching destination on time*** goal dependency relates to the positive impact event elicited to Passenger (see Table 1, first line), while the ***feeling at ease*** dependency relates to the negative impact captured for this same stakeholder (see Table 1, second line). Nevertheless, new dependencies have been added, for instance, when drawing the model, we realized that avoiding accidents dependency (previously only attributed to the Nearby Car stakeholder, see Table 1, line 6) is also relevant for the Passenger [5]

Besides dependencies, the goal model of Fig. 5 depicts the internal perspective of the Driverless Car, assisting in the analysis of the system's requirements. Note that the ethical principles of Beneficence, Non-maleficence, Autonomy and Explicability are represented there by qualities (consistent with our ontological notion of NFR). Then, for each of these qualities, more specific goals and qualities are identified and related to them by contribution links. For instance, the ***choosing quicker route*** quality helps (i.e. partially contributes to) the achievement of Beneficence. Additionally, ***choosing quicker route*** may be indirectly related to the ***reaching destination on time*** goal dependency of the Passenger. Similarly, the goals that help to achieve the ***Explicability*** quality are also indirectly related to the ***having explanations about system's decisions and actions*** goal dependency of the Passenger.

The goal model also allows the requirements analyst to progressively identify more concrete requirements and solutions and the resources needed to accomplish them. For example, the ***use a GPS with frequent map updates*** task makes (i.e. fully accomplishes) the ***choosing quicker route*** quality, and the ***GPS*** itself is a resource needed in this task. Moreover, the ***be aware of traffic laws*** task is a means for the ***following traffic laws*** goal.

Another task worth clarifying is ***use the 2 second rule***. This is a well-known rule for maintaining a safe distance between vehicles. It is adopted in some countries as a good code for driver conduct for human drivers [1], and it can also be adopted as a requirement for driverless cars. Note that this task makes the ***keeping a safe following distance while driving*** quality. However, to accomplish the higher level ***keeping a safe following distance*** quality, other tasks and qualities are involved.

References

[1] The 2-second rule. In: Learn the Road Code (2016), `https://drive.govt.nz/get-your-learners/interactive-road-code/`

[2] Dalpiaz, F., Franch, X., Horkoff, J.: iStar 2.0 language guide. arXiv:1605.07767 [cs.SE] (2016), `dalp-fran-hork-16-istar.pdf`

[3] Floridi, L., Cowls, J., Beltrametti et al., M.: Ai4people—an ethical framework for a good ai society: Opportunities, risks, principles, and recommendations. Minds & Machines 28, 689–707 (2018)

[4] Guizzardi, R. et al.: Ethical requirements for ai systems. In: Proc. of Canadian AI 2020. pp. 251–256 (2020)

[5] We did not update our table on purpose, since although that would make both models more consistent, this is an interesting case in which the visualization of the goal model and its particular constructs (in this case, dependency, goals and qualities) helped us realized a missing requirement for one of the stakeholders.