

Formal verification of ethical choices in autonomous systems.

Datenbank

TEMA, Copyright WTI-Frankfurt eG

Deskriptoren

Inselbetrieb; Ethik; formale Überprüfung; unbemanntes Fahrzeug

Freie Begriffe

Hybridarchitektur; theoretical framework

Abstract

Autonomous systems such as unmanned vehicles are beginning to operate within society. All participants in society are required to follow specific regulations and laws. An autonomous system cannot be an exception. Inevitably an autonomous system will find itself in a situation in which it needs to not only choose to obey a rule or not, but also make a complex ethical decision. However, there exists no obvious way to implement the human understanding of ethical behaviour in computers. Even if we enable autonomous systems to distinguish between more and less ethical alternatives, how can we be sure that they would choose right? We consider autonomous systems with a hybrid architecture in which the highest level of reasoning is executed by a rational (BDI) agent. For such a system, formal verification has been used successfully to prove that specific rules of behaviour are observed when making decisions. We propose a theoretical framework for ethical plan selection that can be formally verified. We implement a rational agent that incorporates a given ethical policy in its plan selection and show that we can formally verify that the agent chooses to execute, to the best of its beliefs, the most ethical available plan.

Copyright Elsevier B.V. Reproduced with permission.

Autor

Dennis, Louise; Fisher, Michael; Slavkovik, Marija; Webster, Matt

Institution

Department of Computer Science, University of Liverpool, GB; University of Bergen, NO

Quelle

Robotics and Autonomous Systems * Band 77 (2016) Seite 1-14 (14 Seiten, 53 Quellen)

Sprache

EN Englisch

Dokumentart

J Zeitschrift

Erscheinungsjahr

2016