# Are Autonomous Vehicles The Sentient Robots We Were Promised?

**Michael Braun**
BMW Group Research, New Technologies, Innovations
85748 Garching, Germany
michael.bf.braun@bmw.de

Position paper submitted to CHI 2018 workshop "Interacting with Autonomous Vehicles: Learning from other Domains"


## Abstract
Digital assistants are gaining acceptance in modern life and it is likely that autonomous vehicles in the future will also be controlled with their help. The advantages are plentiful: natural language interaction makes the system more human, automation gives us more time for the important things. But how must these agents and their personalities evolve to give us the perfect user experience? We take a look at current developments in personality-based human-robot interaction and try to widen our focus by looking at how humanoid agents are portrayed in science fiction. These concepts show potentials but also hazards which could come with humanized computers. In our opinion, autonomous vehicles are the predestined breeding grounds for sentient assistants. We however need to ask ourselves what dangers we conjure with them.

## Author Keywords
HCI, Automotive UI, Automated Vehicles, Virtual Assistants, Smart Agents, Sentient Computing, Ubiquitous Computing

## Introduction
We have witnessed a great deal of digitalization in the recent past, yet we are still waiting for the smart robots we were promised by industry visions and science fiction. We are willingly opening our homes to digital assistants, we are staring at our phones 10 hours a day, is there not enough

data by now to adapt interaction to exactly what we want? There are two reasons why autonomous vehicles might be the missing stone in the path to truly smart robots which we can interact with completely naturally. On one hand, the devices we are currently using cannot entirely understand what is happening around them. Mobile phones and smart home assistants are just not receptive enough. And even if they had more environment sensors, there is more to understanding context than watching the surroundings [10]. Modern cars however can sense the world around them, get additional information from other cars and infrastructure, and most importantly: monitor their occupants wholly without being in the way. On the other hand, there are reasons to believe that the providers of modern day software do not actually intend to make interaction less intruding, as they partly rely on revenue from advertising which can only be guaranteed by constant interplay [2]. Automotive manufacturers however are currently concerned to minimize driver distraction and deliver a different kind of user experience to the passengers than producers of smart gadgets.

With the hypothesis in mind that autonomous cars will provide the perfect breeding grounds for truly natural HMI, we share a glimpse of our vision for future interaction inspired by the current developments of digital agents and by the ideas we stumble upon in popular media and fiction.

## Personality in Human-Robot Interaction

Smart agents like Siri and Alexa have introduced natural language interaction into everyday life. They come with limited functionality: in the end they can do nothing more but answer questions if the user knows how to ask them right. What makes these agents special however is the human touch they express through the usage of voice output and conversational behavior. Users get the feeling they are talking to a humanoid helper instead of a machine, establishing some kind of emotional bond. Siri's developers also gave her a somewhat feisty personality, compared to the rather neutral conspecifics Alexa and Cortana. One behavior all of them show in one way or another is trying to engage the user with funny novelty features like jokes and fun facts, which can be interpreted as an approach to build a social relationship. [7]

User interaction in the field of robotics has developed even more personal in recent years, as service robots with narrow application areas and thus well defined user groups can actively address their users' abilities and behavior. The Robot *Murphy* assists children in doctors' waiting rooms and as it assumes the child to be intimidated and sick, it acts as if itself needs company in trying times. The developers justify the behavior with the statement *misery loves miserable company*, so the children's sorrows can be mitigated by helping the robot [12]. This use case goes well with the model of De Graaf and Allouch who investigate how humans get to be acquainted with each other and how friendships are formed. They identify *similarity*, *reciprocal sympathy*, and *intimacy* as key driving factors for the start of a friendly relationship [5]. We can see these factors being addressed in Cortana's personality being adjusted to its surrounding cultural habits [1] or in the behavior of other social robots like *Robin* [4]. This robot was envisioned to help children learn to deal with diabetes. It actively demands help by the user, e.g. when it cannot get up, and it can lose interest in the user to show independence.

When virtual assistants in cars learn to express their own personalities, a lot of these factors will play a role in how well the agent is going to be perceived. Forming a friendship with the car's agent can be influenced through similarity – Are they sporty or rather comfortable drivers? Do they talk a lot? – through reciprocal sympathy – Drivers and

assistant spend a lot of time together to learn about each other – and through intimacy – Cars are vulnerable towards their drivers as they depend on their benevolent maintenance. This way, autonomous cars can benefit from the advances in human-robot interaction by building upon established means of initial bonding and further improve the relationship between agent and user when they learn about preferences, tastes, and boundaries.

## Humanoid Agents in Science Fiction

Potentially good visions on how artificial intelligence can shape the future of the world can also be found in science fiction. Dystopian novels tell us how their authors, who have intensively thought about possible futures, envision technology and society to change. In *QualityLand* by Marc-Uwe Kling for example, smart devices like self-driving cars know everything about their users and display themselves with humanoid personalities [8]. The protagonist encounters a robotaxi which initially adapts to his preferences but later takes offense in the guest's poor taste in music. It boasts about how much better of a driver it is compared to humans and talks about the joy it gets from denouncing traffic violators. Delivery drones in this world are constantly anxious to receive good ratings. Digital assistants by different manufacturers despise each other openly. While this sounds overdone for comedic effect, the idea to make devices as human as possible – including their bad character traits – could lead to a way of interacting with technology which is accessible to anyone and by the way incorporates entertainment for the user. In *QualityLand*, the simplification of user interaction is taken to the extreme as in many situations the only possible input by the user is to say *Okay*, which is justified with the fact that the corporations behind those devices know all about the user anyway and can therefore predict their desires.

Another example for the personification of technology is given in the movie *Blade Runner 2049* [6]. The personal assistant *Joi* accompanies the protagonist throughout the movie as his female sidekick. Only on second thought one realizes its manipulative nature: Joi incorporates a highly sexualized approach to lure its user into loving her, gaining his trust. This can be seen as an effective way of building customer loyalty but also as an open invitation to manipulate users through emotional leverage.

What we can learn from these examples is that certain humanizing techniques to improve interaction, e.g. reciprocal sympathy as stipulated in the previous chapter, can also be abused to manipulate users. We are used to interact with real humans who are driven by complex actuators, are capable of altruism, and are usually not purely motivated by profit. When big corporations can mould their agendas into human form and coat them with sex-appeal and niceties, consumers have to get used to be on the watch for potential exploitation when interacting with technology.

One last example from science fiction could solve trust issues concerning an assistant's genuineness: as depicted in the episode *White Christmas* in the *Black Mirror* series, the assistant can be an identical, digital clone of the user's consciousness with knowledge of all their preferences, traits, and feelings [3]. The catch in this approach (apart from the technological hurdles) lies in ethical considerations and whether virtual clones deserve a right to self-determination.

There is no way of saying where we are headed. As the general tendency in these examples however goes towards a very emotional way of interaction with humanoid agents, or the complete abandonment of interaction due to technology's omniscience, we want to stimulate the debate about these ideas and what they can mean for future interaction with autonomous vehicles.

## Sentient Assistants and Autonomous Vehicles

From a technical point of view, modern cars are the perfect environment to finally turn the visions of ubiquitous computing into reality. In the past, the realization of initial concepts of computers which are small, everywhere, and merge invisibly into the environmment [9, 13] has been thwarted by the development of powerful mobile phones which focus all interaction in one handheld device [2]. Modern cars however already possess manifold sensors to monitor what is happening around the vehicle, they can interact with infrastructure to get information about the rest of the world, and they will soon be able to measure their passengers' states with cameras and through physiological sensing. All this while the user sits inside the device with possible actuators for interaction all around them.

In the light of this infrastructural development, sentient computers, meaning systems able to silently predict our wishes, and corresponding novel interaction paradigms may find a platform in the automotive sector. As an example, Schmidt et al. introduce the idea of *Intervention User Interfaces*, inspired by the opinion that "there is no sense in staying in the interaction loop while autonomous operations run according to the user's expectations" [11]. They prophesy UIs which unobtrusively inform the user of automation processes and intervention options, and allow for an immediate change in system execution when desired. Combined with a humanoid personality which can befriend the user through aforementioned behavioral patterns we are very close to a sentient personal assistant – like Joi.

We as designers of these systems have to consider what it means to build sentience, not just from a technical perspective but also ethically. How should we handle conflicts based on emotional matters, can AIs and humans be at strife? Are user commands to be preferred, even if the assistant knows better? How can we prevent humanoid assistants from emotionally manipulating humans? And do we need to teach users what they can trust an assistant with?

We hope to inspire a vivid discussion on the prospect of sentient assistants in autonomous cars with this submittal.

## REFERENCES


1. Marcus Ash. 2015. Cortana brings Cultural Savviness to New Markets. https://blogs.windows.com/windowsexperience/2015/07/20/cortana-brings-cultural-savviness-to-new-markets/ (2015).
2. Matthew P. Aylett and Aaron J. Quigley. 2015. The Broken Dream of Pervasive Sentient Ambient Calm Invisible Ubiquitous Computing. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '15)*. ACM, New York, NY, USA, 425–435. DOI: http://dx.doi.org/10.1145/2702613.2732508
3. Charlie Brooker and Carl Tibbetts. 2016. "White Christmas", Black Mirror Season 2, Episode 4. *Channel 4* (2016). http://www.imdb.com/title/tt3973198
4. Lola Canamero and Matthew Lewis. 2016. Making New "New AI" Friends: Designing a Social Robot for Diabetic Children from an Embodied AI Perspective. *Int J of Soc Robotics* 8 (2016), 523–537. DOI: http://dx.doi.org/10.1007/s12369-016-0364-9
5. Maartje Margaretha Allegonda de Graaf and Somaya Ben Allouch. 2017. The Influence of Prior Expectations of a RobotâĂŹs Lifelikeness on UsersâĂŹ Intentions to Treat a Zoomorphic Robot as a Companion. *Int J of Soc Robotics* 9 (2017), 17–32. DOI:http://dx.doi.org/10.1007/s12369-016-0340-4

6. Hampton Fancher, Michael Green, Philip K. Dick, and Denis Villeneuve. 2017. "Blade Runner 2049". *Columbia Pictures* (2017). http://www.imdb.com/title/tt1856101

7. Gary Alan Fine. 1983. Sociological Approaches to the Study of Humor. *Handbook of Humor Research* (1983), 159–181. DOI: http://dx.doi.org/10.1007/978-1-4612-5572-7_8

8. Marc-Uwe Kling. 2017. *QualityLand*. Ullstein, Berlin, DE.

9. Garret Romaine. 1999. The Invisible Computer: Why Good Products Can Fail, the Personal Computer is so Complex, and Information Appliances are the Solution. *Technical Communication* 46, 3 (1999), 395–395.

10. Albrecht Schmidt, Michael Beigl, and Hans-W Gellersen. 1999. There is more to Context than Location. *Computers & Graphics* 23, 6 (1999), 893 – 901. DOI:http://dx.doi.org/https://doi.org/10.1016/S0097-8493(99)00120-X

11. Albrecht Schmidt and Thomas Herrmann. 2017. Intervention User Interfaces: A New Interaction Paradigm for Automated Systems. *interactions* 24, 5 (aug 2017), 40–45. DOI: http://dx.doi.org/10.1145/3121357

12. Daniel Ullrich, Sarah Diefenbach, and Andreas Butz. 2016. Murphy Miserable Robot - A Companion to Support Children's Well-being in Emotionally Difficult Situations. *CHI'16 Extended Abstracts, ACM* (2016), 3234–3240. DOI: http://dx.doi.org/10.1145/2851581.2892409

13. Mark Weiser. 1999. The Computer for the 21st Century. *Mobile Computing and Communications Review* 3, 3 (1999), 3–11.