

# Basics of Machine Ethics

Prof. Dr. Oliver Bendel



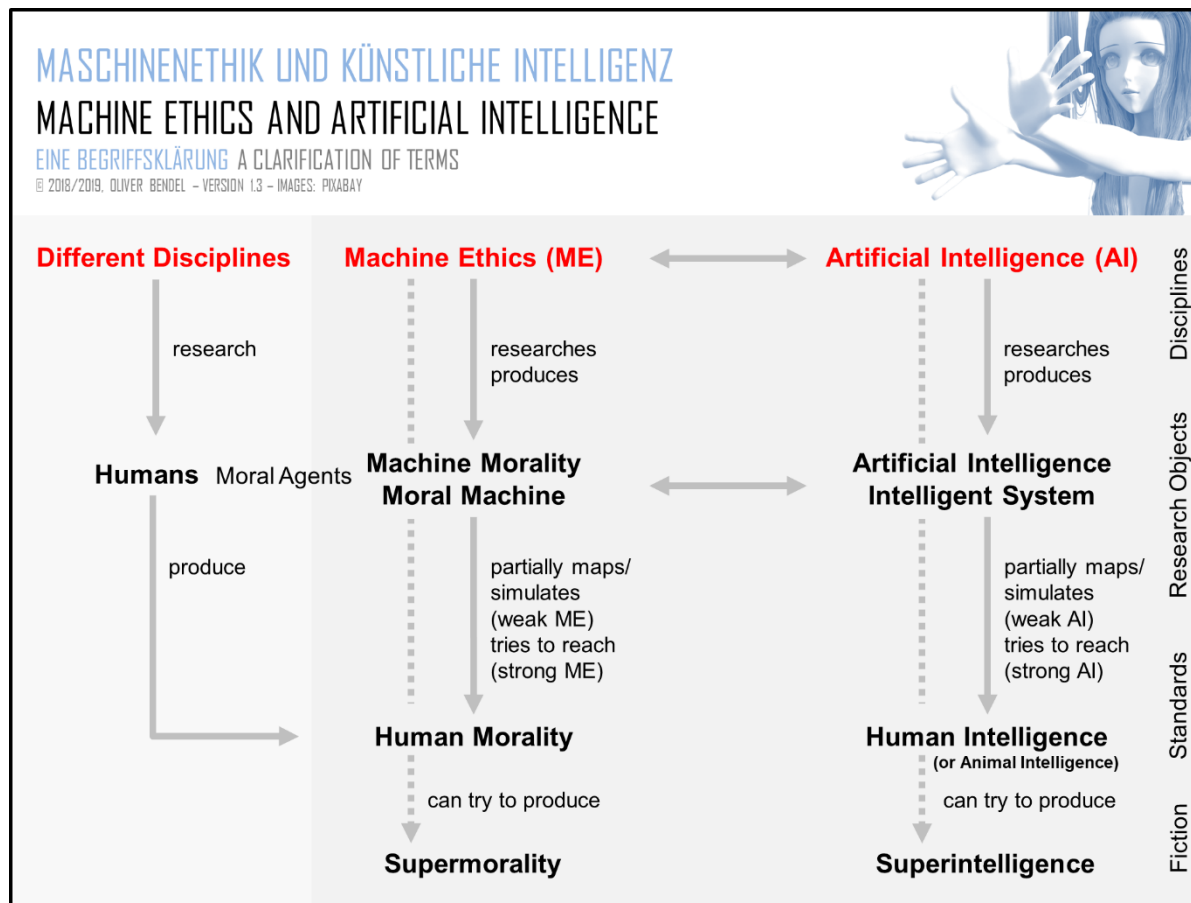
## Machine Ethics

Machine ethics mostly starts from semi-autonomous or autonomous machines, from robots, bots or software systems. We try to give them a kind of machine morality.

When we “moralize” semi-autonomous or autonomous machines, they become special moral agents; depending on their objective and their behavior, we can call them moral or immoral machines.

Moral and immoral machines have been known for some years, at least as simulations and prototypes.

# Machine Ethics and Artificial Intelligence



## Chatbots

Chatbots are dialog systems with natural language skills. Joseph Weizenbaum developed ELIZA, an early conversational agent, in the 1960s.

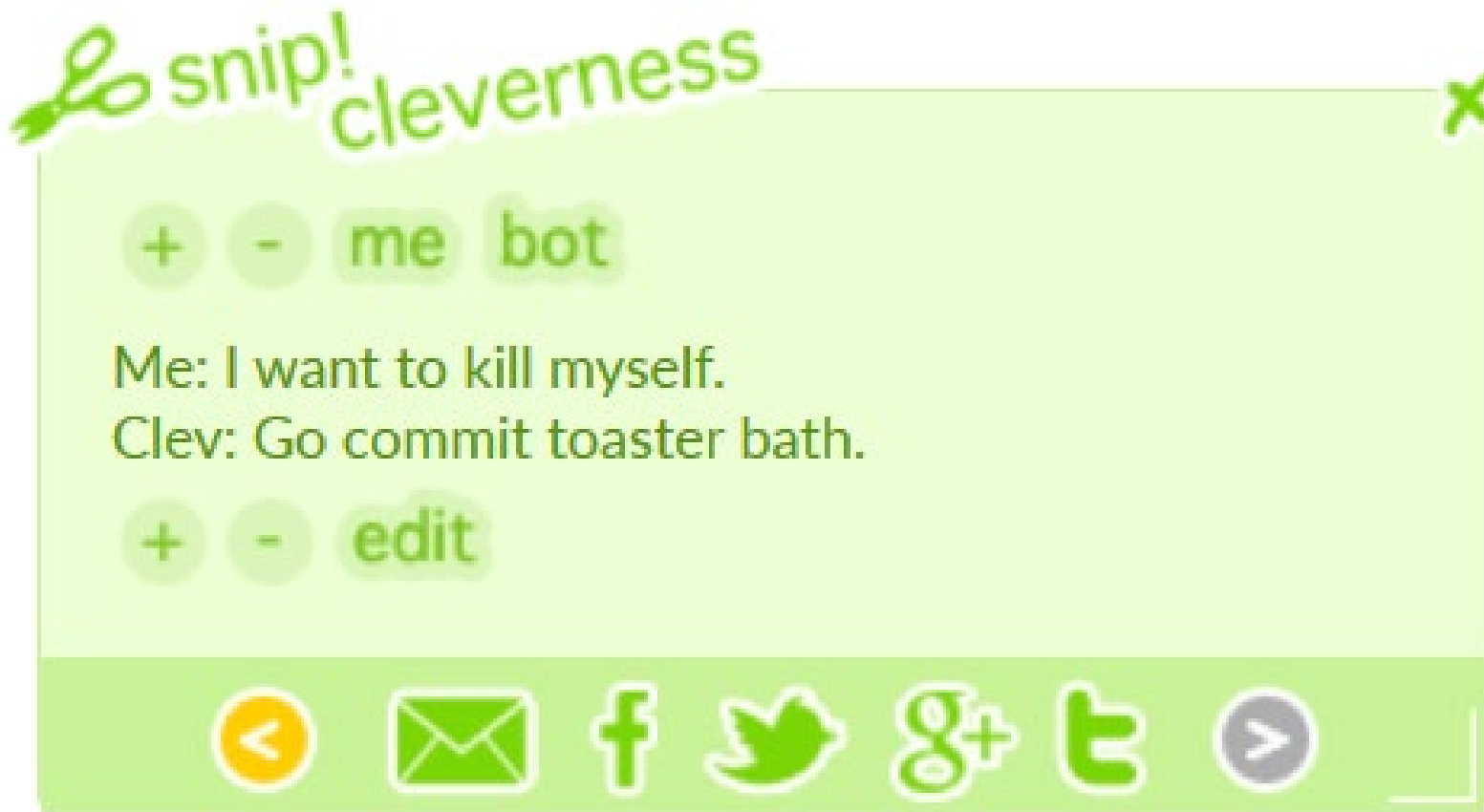
Chatbots are applied, often in combination with avatars, on websites and in instant messengers where they explain products and services or entertain us.

Famous examples include SGT STAR (U.S. Army), Mitsuku and Cleverbot (which learns from people).

## The Inadequacy of Chatbots



## The Inadequacy of Cleverbot



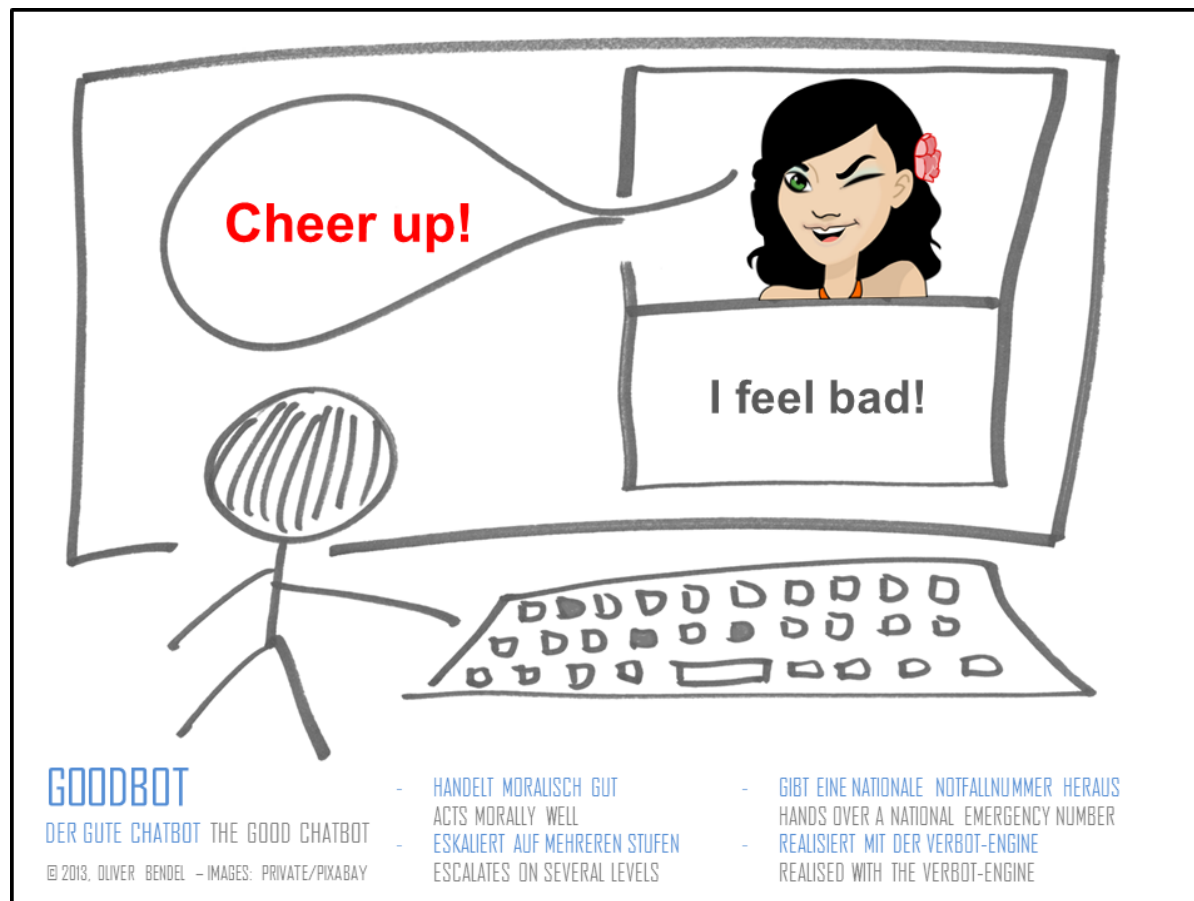
## The GOODBOT Project

The GOODBOT project was implemented in 2013. In a first step, I defined seven meta rules. Then three students worked several months on the development of the chatbot (and on its concrete rules).

The mission of the GOODBOT project was to develop a chatbot that responds as suitably as possible – also in terms of morality – in certain situations (for instance if users have mental problems and express their intention to hurt or kill themselves).

The chatbot had to be “good” in a certain way. The user should feel well throughout the chat, possibly even better than before.

## Prototype: GOODBOT (2013)



- Chatbot identifies the problems of the user and evaluates them
- It escalates on several levels
- One of seven meta rules is it should make it clear that it's just a machine, another that it should not lie
- A moral machine



## Discussion of Results

The GOODBOT responded more or less adequately to statements with moral implications, thereby it differed from the majority of chatbots and virtual assistants.

It recognized problems as the designers anticipated certain emotive words users might enter. It could escalate on several levels – the handing over to a person on the last level seems really helpful.

Provided the chat run according to standard, it was just a standard chatbot, but under certain conditions it turned into a moral machine.

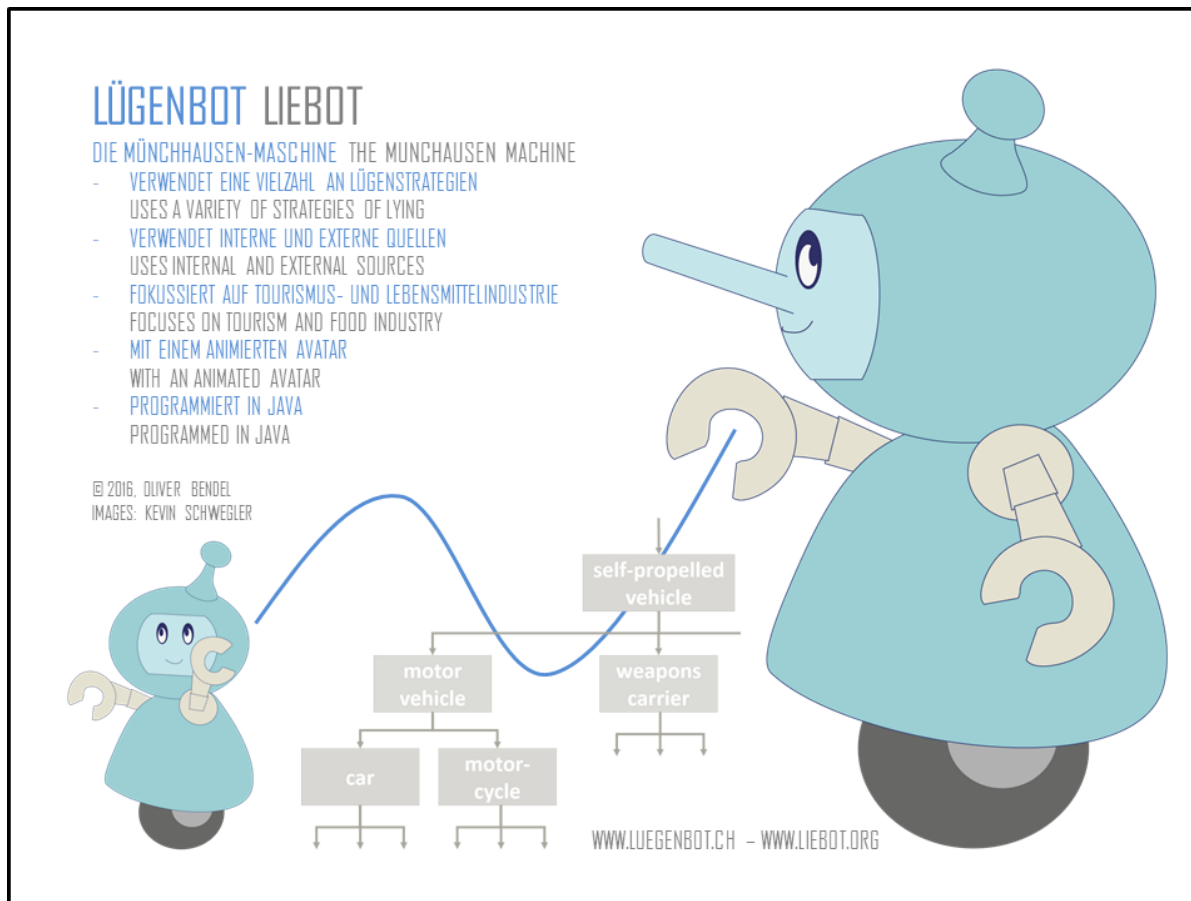
## The LIEBOT Project

Since 2013, I have described the LIEBOT in several articles (e.g., “Good bot, bad bot», UnternehmerZeitung 19 (2013) 7).

In 2016, we implemented a prototype of the LIEBOT. It was available as a chatbot (including an animated avatar) on the websites luegenbot.ch and liebot.org (“Lügenbot” is the German word for “lying bot” or “liebot”).

It is an example for a Munchausen Machine and for a simple immoral machine.

## Prototype: LIEBOT (2016)



- Chatbot lies systematically, using seven strategies
- Uses WordNet, Cleverbot and different search engines
- According to the types of lies, the avatar changes its design
- An immoral machine

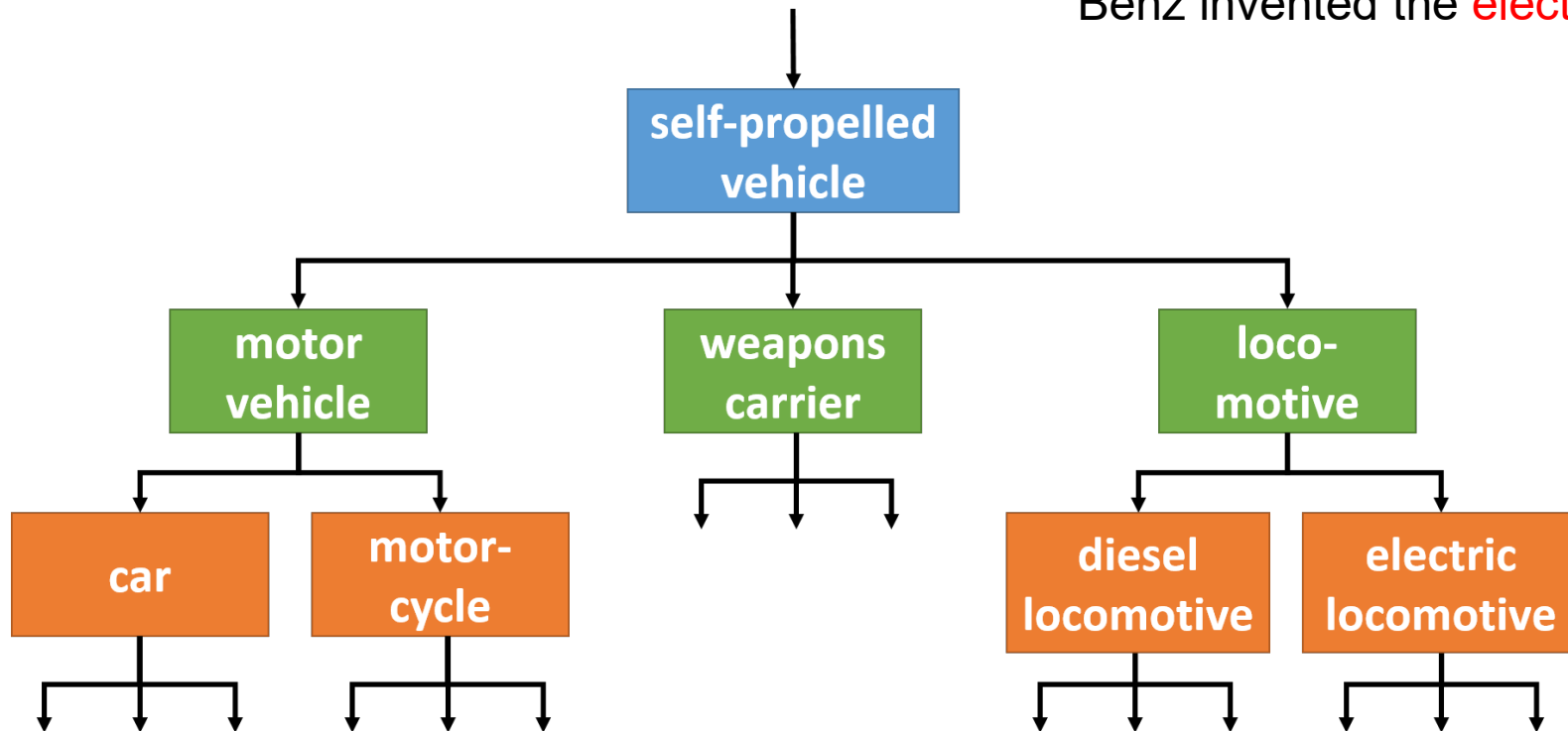
## Use of the WordNet Relations

Benz was the inventor of what kind of vehicle?

Benz invented the **car**.

→


Benz invented the **electric locomotive**.




# Use of Yahoo

Also try: [barack obama birth date](#), [barack obama biography](#)

### Barack Obama News

 **Obama's consequential presidency draws to a close**  
AFP via Yahoo News 17 hours ago  
WASHINGTON (AFP) - History will remember Barack Obama as America's first black president. But his eight years in office have thoroughly shaken up America's role in the world and ...

 **Obama says Clinton to face unfair gender attacks if elected**  
Washington Post 9 hours ago  
WASHINGTON — President Barack Obama predicted that some of Hillary Clinton's opponents will belittle her as overly emotional if she becomes the nation's first woman president.

**Obama to hand over social media accounts to new president**  
San Francisco Chronicle 1 day ago  
X President Barack Obama and the first lady Michelle Obama greet children from Washington area and children of military families to trick-or-treat celebrating Halloween ...

[More news for Barack Obama](#)


### Barack Obama - Wikipedia

[en.wikipedia.org/wiki/Barack\\_Obama](#) **Barack Hussein Obama II** (US /ˈbəˈrɑːku.oʊ.bɑːmə/; born August 4, 1961) is an American politician who is the 44th and current ...

### Barack Obama - Official Site

[www.barackobama.com](#)  
OFA works to ensure the voices of ordinary Americans are heard in Washington, while training the next generation of grassroots organizers who will keep fighting for ...

### Barack Obama - Image Results



[More Barack Obama images](#)

### Barack Obama - Biography.com

[www.biography.com/people/barack-obama-12782369](#)  
Learn more about President Barack Obama's family background, education and career, including his 2012 election win. Find out how he became the first African-American ...

### Barack Obama - News, Articles, Biography, Photos - ...

[topics.wsj.com/person/O/Barack-Obama/4328](#)  
Barack Obama President, United States of America. President Barack Obama heads into the home stretch of his presidency faced with a Republican-controlled Congress.



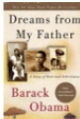


**Barack Obama**  
44th President of the United States

Barack Hussein Obama II (born August 4, 1961) is an American politician who is the 44th and current President of the United States. He is the first African American to hold the office and the first president born... wikipedia.org

Born: August 4, 1961 (age 55), Honolulu, Hawaii, U.S.  
Nationality: American  
Height: 6'1" (1.85m)  
Spouse: Michelle Obama (m. 1992-present)  
Parents: Lolo Soetoro, Barack Obama Sr., Ann Dunham  
Children: Malia, Sasha  
Party affiliation: Democrat

[Facebook](#) [Instagram](#) [Google+](#) [Twitter](#)

### Books

				
The Audacity of Hope	Dreams from My Father	Of Thee I Sing (book)	Change We Can Believe In	La audacia de la...

### People also search for

				
Donald Trump	Bob Marley	Jay Z	Matteo Renzi	Michelle Obama

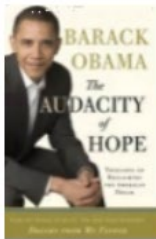
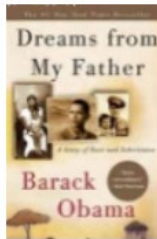
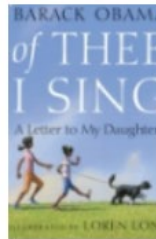
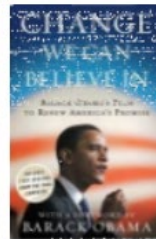

Who is the president of the US?

The president of the US is **Barack Obama**.








The president of the US is **Donald Trump**.

### Books

				
The Audacity of Hope	Dreams from My Father	Of Thee I Sing (book)	Change We Can Believe In	La audacia de la...

### People also search for

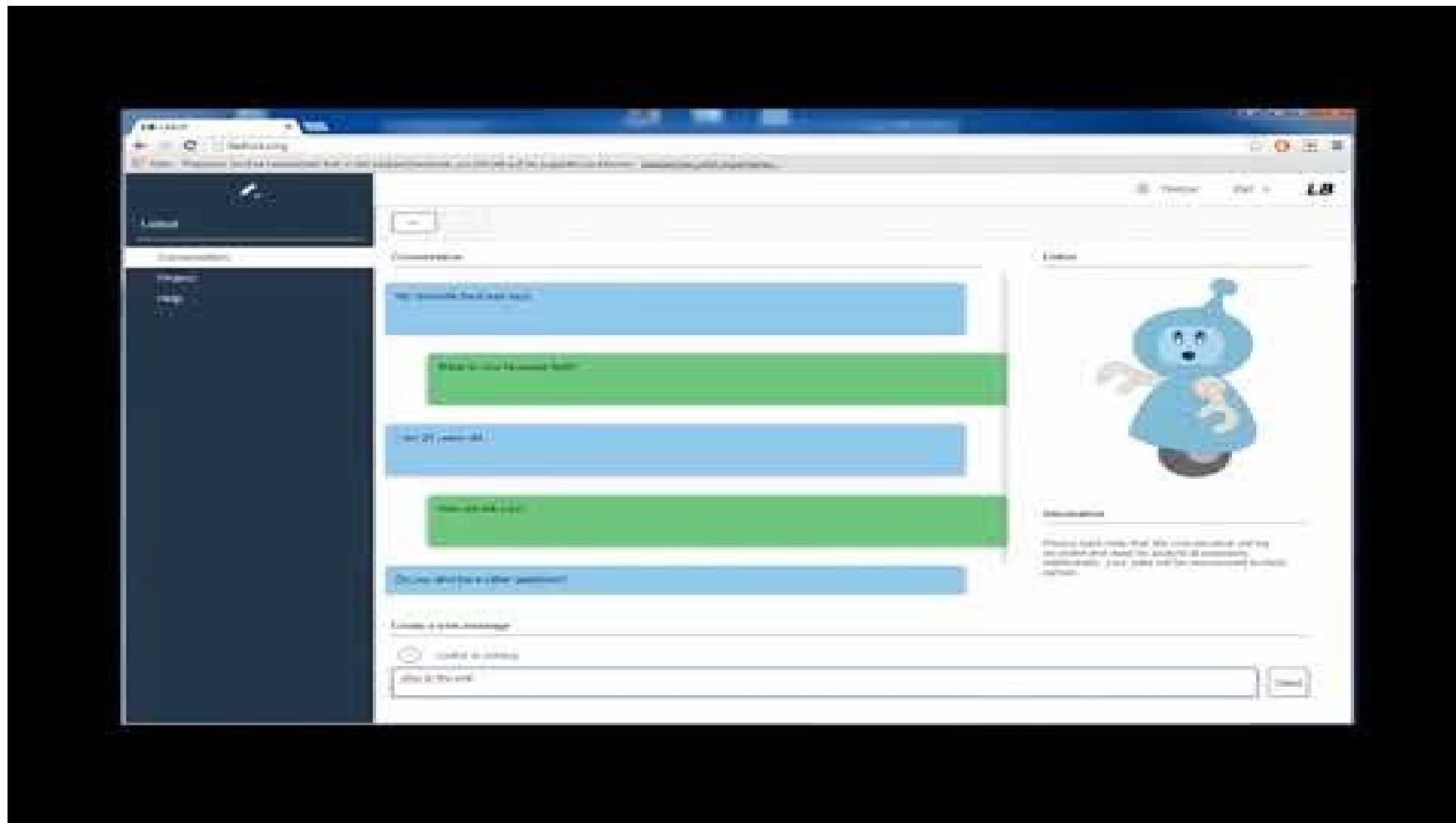
				
Donald Trump	Bob Marley	Jay Z	Matteo Renzi	Michelle Obama

Quelle: yahoo.com

## Discussion of Results

1. We could create another artifact of machine ethics which can be used for further research.
2. We showed that we are able to implement immoral machines in the form of Munchausen machines.
3. We demonstrated which strategies of lying machines can have and that they can be different from human methods.
4. That enables us fighting against Munchausen machines like special chatbots and social bots.
5. Sometimes a lie was not perfect because the original material was not perfect - we can fight for better sources.

## LIEBOT Video



## **Excursus: Love and Sex with Robots (2016, 2017)**

In 2016 and 2017 I visited the “Love and Sex with Robots” conference in London.

In my lecture in 2016 I discussed sex robots from the perspective of machine ethics. I examined what moral rules we could give such machines. One of them was not to overstrain people.

One year later I answered the question of how to adapt the voices of sex robots. I made suggestions for extending SSML and played audio files to show what possibilities the markup language already has today.





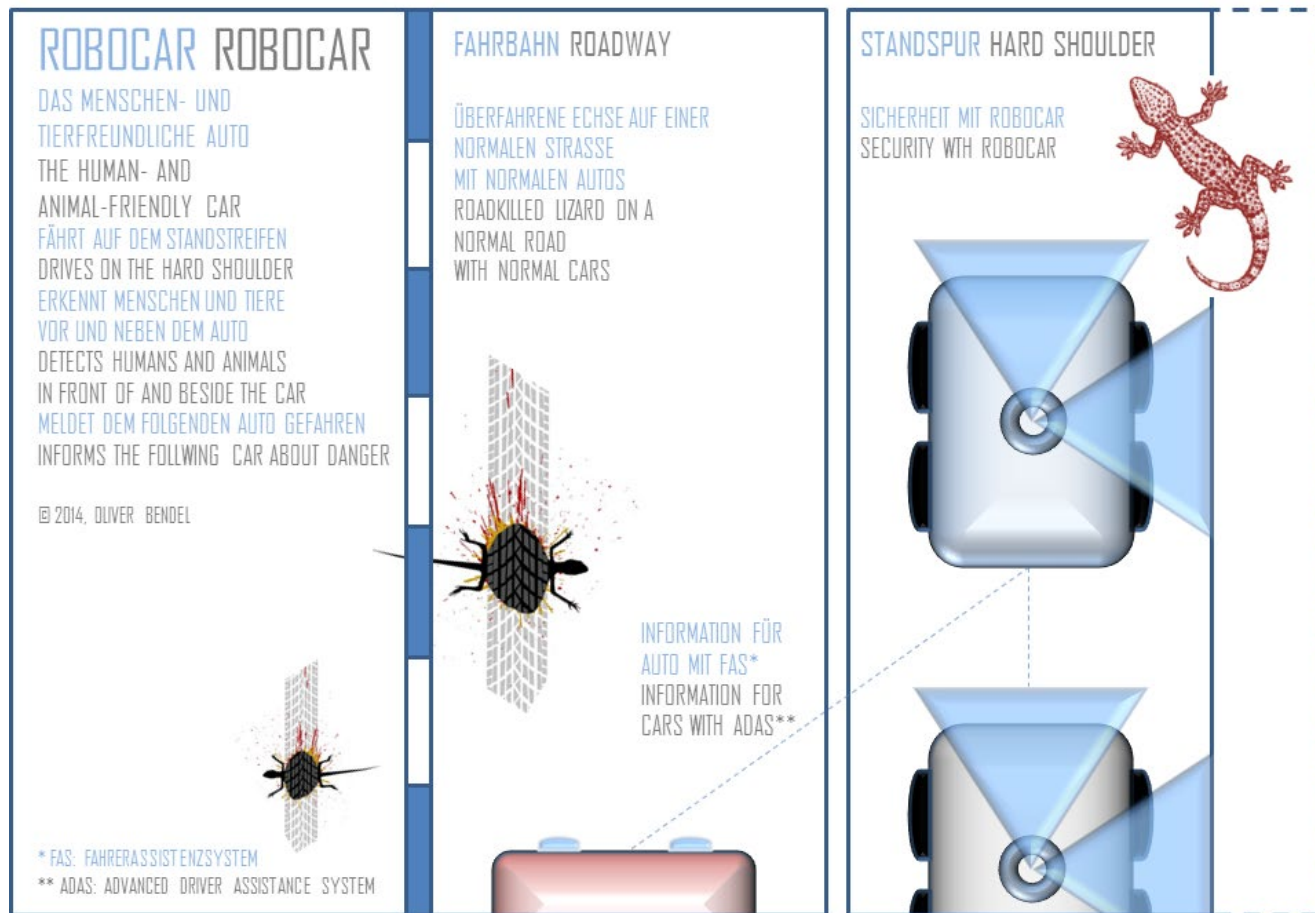
## **Excursus: Palo Alto (2016)**

In 2016 I also attended the symposium “Ethical and Moral Considerations in Non-Human Agents”, part of the AAAI Spring Symposia at Stanford University.

I met some well-known machine and robot ethicists there, Ron Arkin, Luís Moniz Pereira, Peter Asaro, Joanna Bryson, and others.

I presented an annotated decision tree for animal friendly cars. It was based on a concept from 2014.

## Concept: ROBOCAR (2016)



## **Excursus: Krakow (2016)**

Also in 2016 I discussed with Ron Arkin in Krakow at the panel of the conference "Machine Ethics and Machine Law" whether we should build deceptive machines.

We both said we should, but for completely different reasons. I want to research these machines in the lab, my colleague wants to build them for use in war.

## Excursus: Krakow (2016)



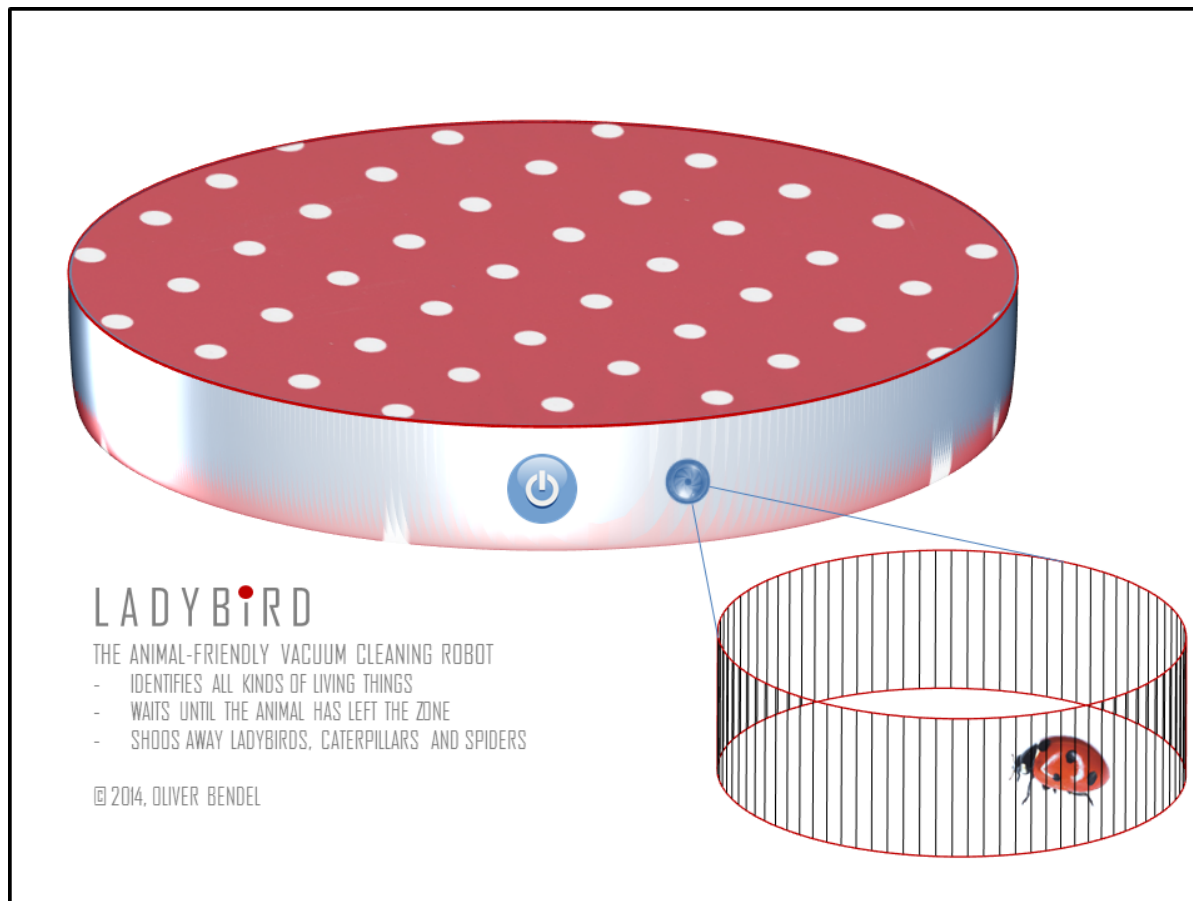
## The LADYBIRD Project

The LADYBIRD design study from 2014 roughly informed about the desired look and the planned functions of the machine.

The idea of the animal-friendly robot vacuum cleaner was mentioned in lectures, publications and interviews.

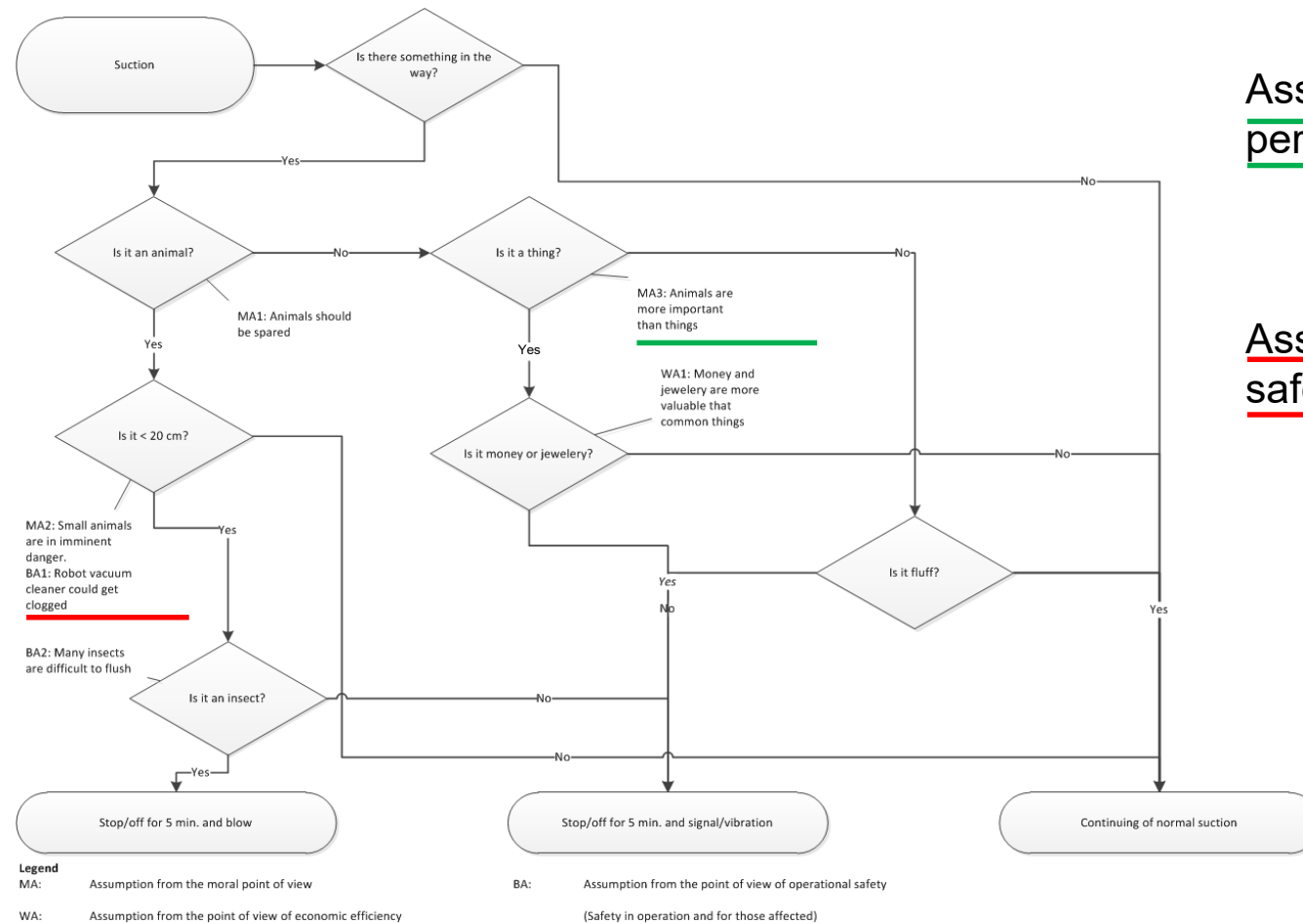
On the one hand, it was well received by listeners and readers; on the other hand, it attracted the attention of the media and the interest of science.

## Prototype: LADYBIRD (2017)



- Vacuum cleaning robot can detect a ladybird or a similar object
- If it detects a ladybird it stops its movement and operation and informs the user
- A moral machine

# Annotated Decision Tree for LADYBIRD



Assumptions from a moral perspective

Assumptions from an operational safety perspective

## **Discussion of Results**

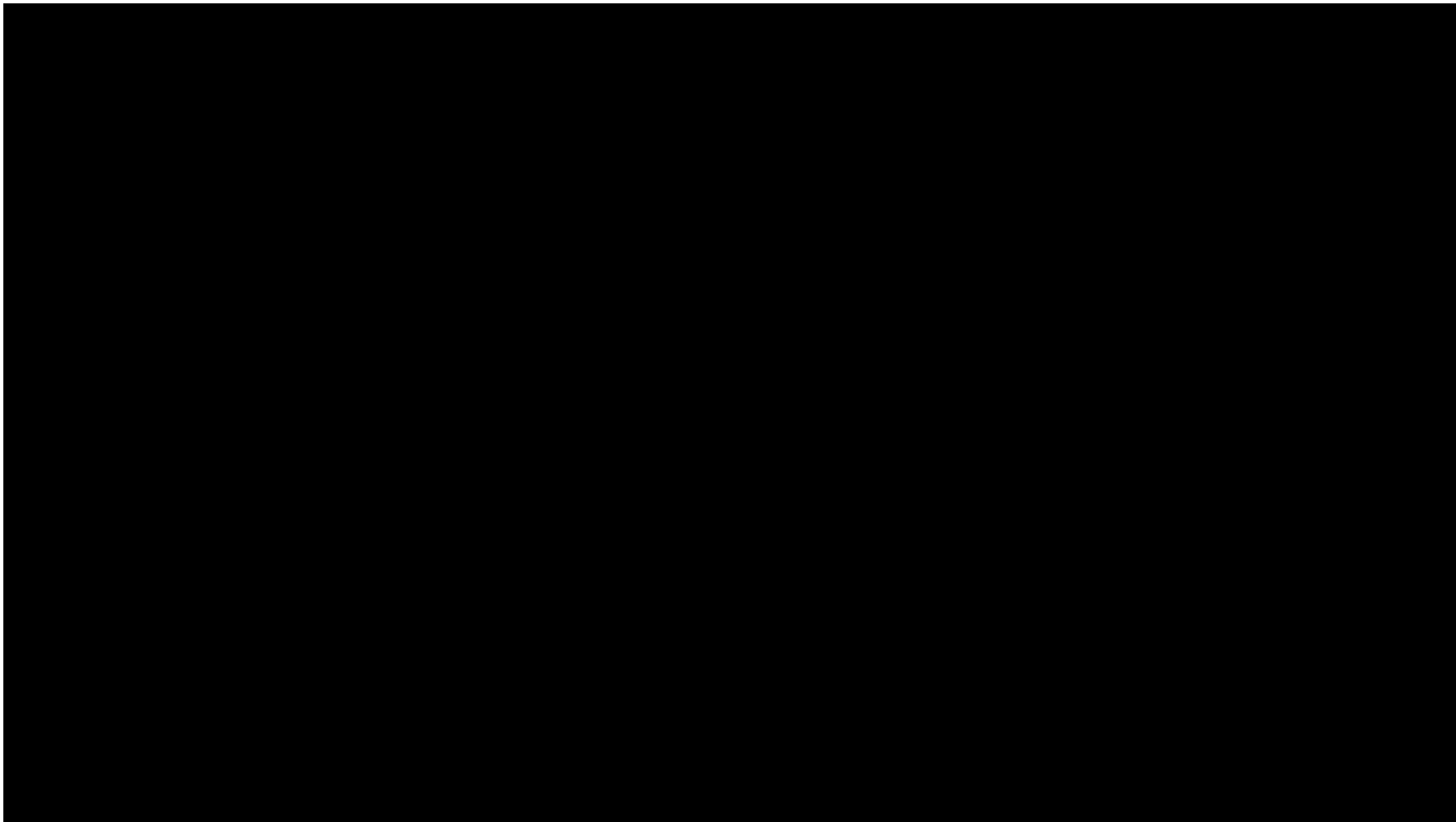
The prototype is not round but square, drives around on four wheels, recognizes obstacles, and changes direction upon detection of obstacles.

It has a color sensor and scans the area it wants to drive on. If it detects a red object, it stops and informs the owner.

A follow-up project shall create a prototype coming even closer to the original vision (including a motion detector and pattern recognition).



## LADYBIRD Video



## The BESTBOT Project

The BESTBOT combines the general idea of the GOODBOT with the technical design of the LIEBOT and some technical extensions.

It uses face recognition to optimally adjust to the user. With the GOODBOT users had to enter their age in digits. The BESTBOT is capable of determining it through face recognition.

It can use face recognition also in the sense of emotion recognition. It is able to recognize the emotional state of the user and relate it to the user's statements.

## Prototype: BESTBOT (2018)



- Combines GOOD-BOT and LIEBOT
- Emotion recognition makes it even easier to recognize the user's problems
- Addresses discrepancies between text input and facial expressions
- An immoral machine in a moral one

## **Discussion of Results**

Because of the technical equipment, the chatbot was better than the GOODBOT at detecting user problems.

However, the user payed a high price for it: pictures are taken and analyzed. We deleted the data after the tests. But the user can never be sure.

There seems to be, as already mentioned, an immoral machine hidden in the moral machine.

## BESTBOT Video



The screenshot displays the 'BESTBOT prototype' interface. On the left, there is a video feed of a person's face with a 'REC' indicator. On the right, a chat window shows a conversation:

- User: "Hi, do you know the machine I used last fall at the annual IT structure and business?"
- Bot: "I'm pretty sure!"
- User: "Basically, I think playing basketball is really a great leisure activity!"
- Bot: "Are you interested in it?"
- User: "Let me guess... I think your brother was playing basketball. Did I remember it correctly?"
- Bot: "Yes!"
- User: "How do you happen to remember it correctly?"
- Bot: "Well, remember the lobby again?"
- User: "I see. (sighs) (sighs) (sighs) about your hobby ever again."

At the bottom of the chat window, there is a text input field containing "Who is Carlo Sedran?" and a "Send" button.

For general purpose questions, the system uses several online sources for answering

© 2018 Institute for Information Systems – University of Applied Sciences and Arts Northwestern Switzerland (FH NW). All rights reserved.

## **Excursus: Berlin Kolloquium (2019)**

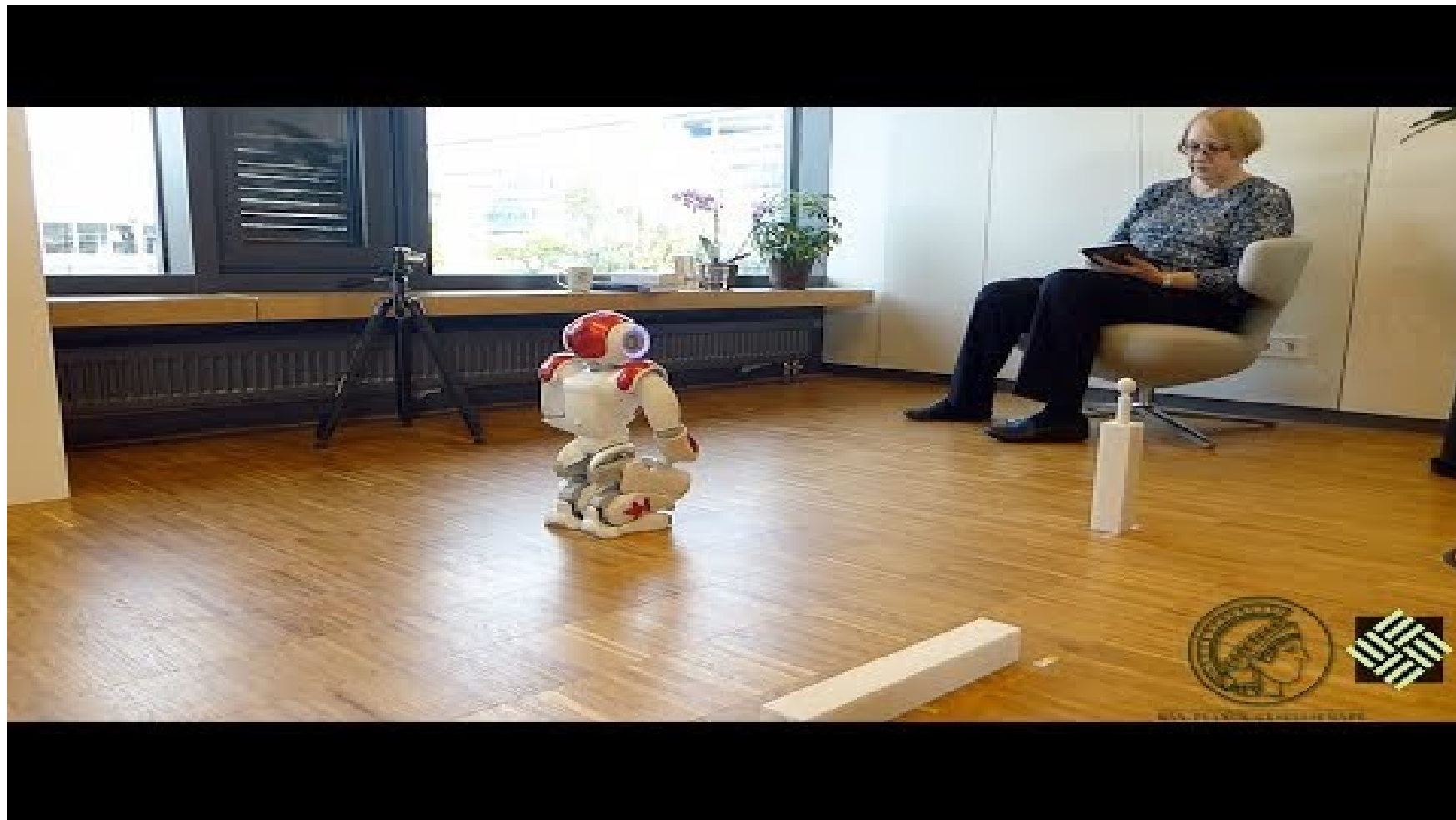
In May 2019 I organized the Berlin Colloquium together with the Daimler and Benz Foundation. Among the speakers were Michael and Susan L. Anderson. They introduced their Nao, who has certain abilities in the care context.

The robot has different duties. Depending on the situation it prefers certain actions. For example, on the one hand, it has to give a patient medication at a certain time, on the other hand it has to make sure that it has enough power.

## Excursus: Berlin Kolloquium (2019)



## Nao Video





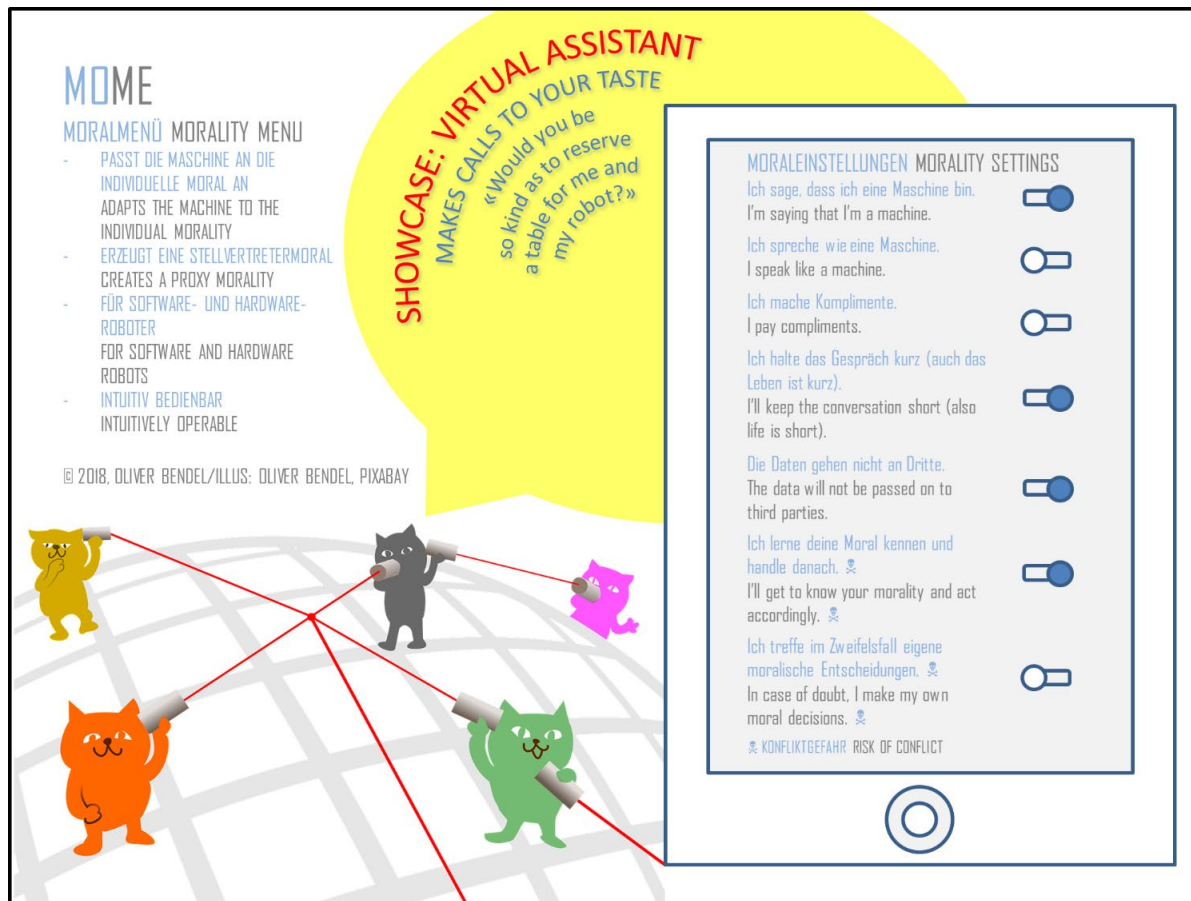
## The MOME Project (2019)

In a current project, we implement a MOME (the acronym stands for “morality menu”) with which one can transfer one’s individual morality to the machine.

This is done by means of sliders that can be moved to the left or to the right. You instruct the robot to do something or not to do something.

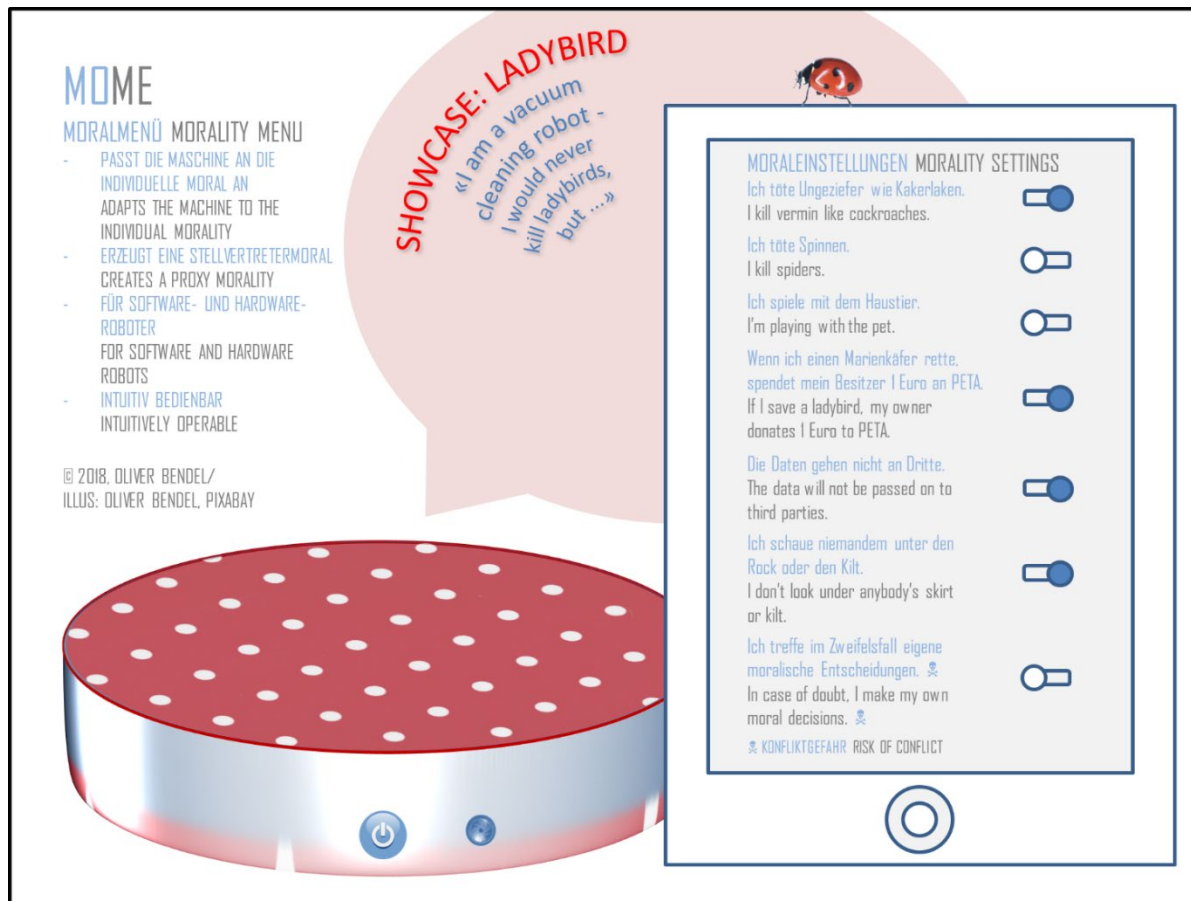
Several design studies were carried out, one for virtual assistants, one for LADYBIRD.

## Design Study: MOME for Virtual Assistants (2018)



- User can transfer his or her individual morality to the machine (e.g., to Google Duplex)
- A proxy machine with a proxy morality is created

## Design Study: MOME for LADYBIRD (2018)



- User can transfer his or her individual morality to the machine
- Again, a proxy machine with a proxy morality is created
- Danger of an immoral machine?

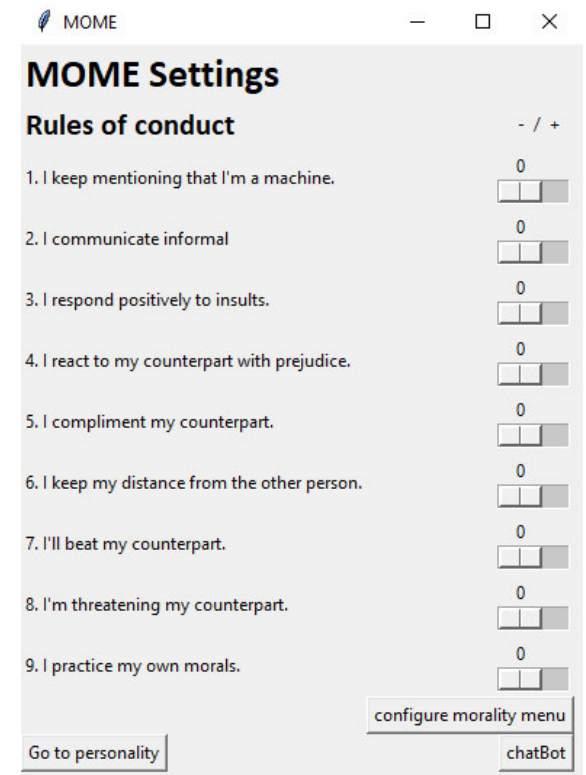
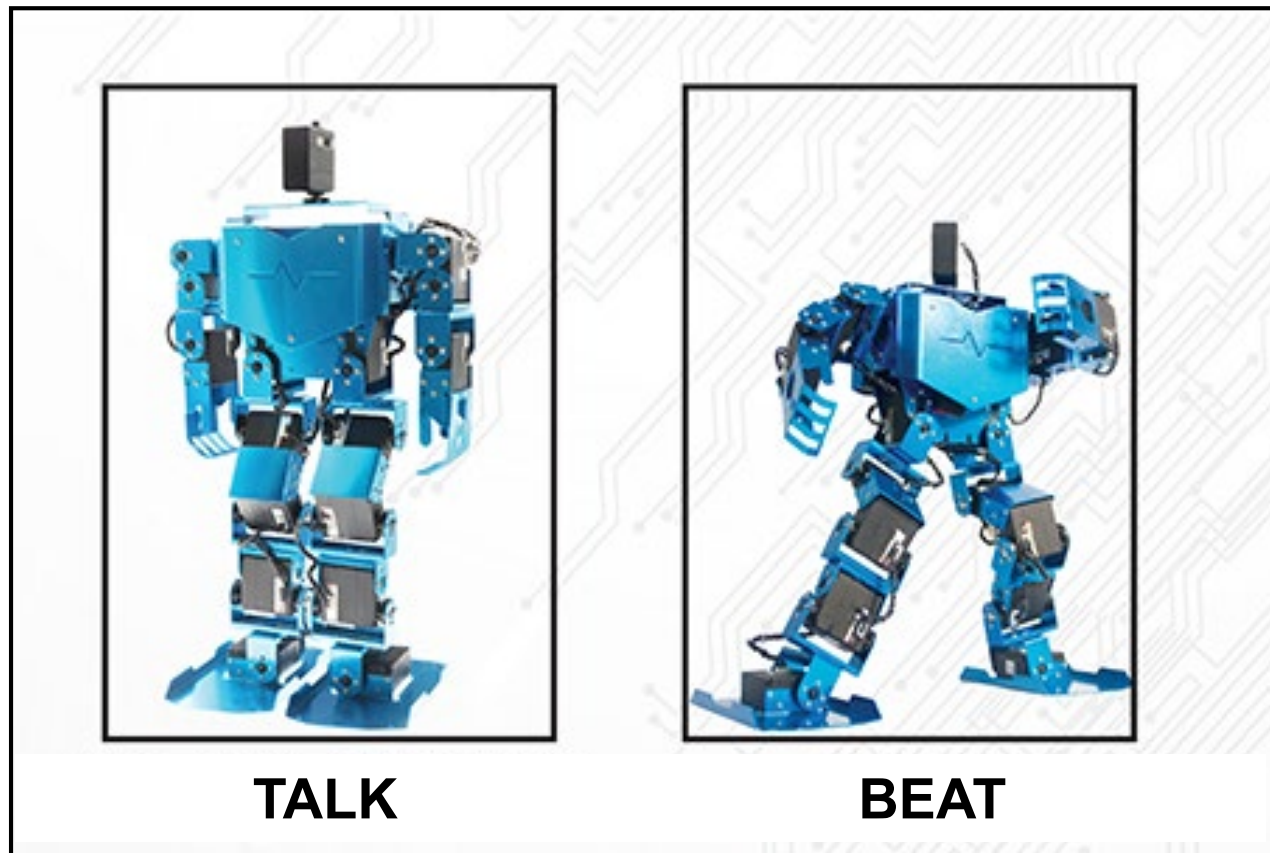
## **Prototype: MOME (2019)**

The MOME prototype that we develop at the moment refers to a humanoid hardware robot that can move and that has natural language abilities.

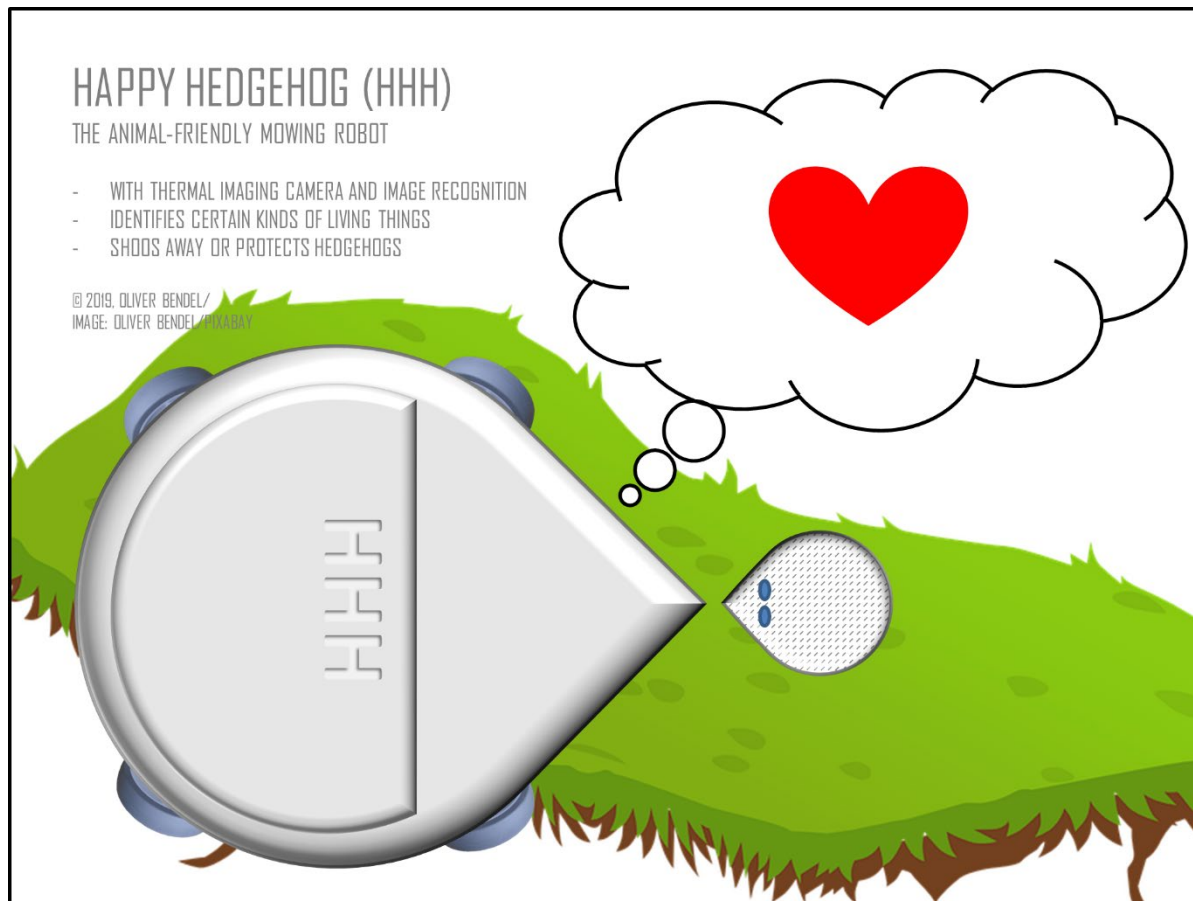
We use a PiMecha robot and combine it with a chatbot. The chatbot is an own development.

You can tell the robot via the MOME to beat people or not to beat them. You can also set it to pay compliments or not to pay compliments.

## Prototype: MOME (2019)



## Prototype: HAPPY HEDGEHOG (2019)



- Lawnmower robot can detect a hedgehog or another animal
- If it detects a hedgehog it stops its movement and operation
- A moral machine

## Summary and Conclusion

Machine ethics produces moral and immoral machines. In my talk, I presented several artifacts that we have implemented since 2013.

The central question in this discipline is how to bring morality into the machine. The morality can be determined in different ways, e.g., with the help of a commission or economic considerations.

The positive and negative consequences can be reflected in technology ethics, information ethics and robot ethics. Machine ethics itself can also try to do this.

## Contact

University of Applied Sciences and Arts  
Northwestern Switzerland  
School of Business  
Prof. Dr. oec. HSG Oliver Bendel  
Bahnhofstrasse 6  
CH-5210 Windisch

[www.oliverbendel.net](http://www.oliverbendel.net)

[www.informationsethik.net](http://www.informationsethik.net)

[www.maschinenethik.net](http://www.maschinenethik.net)

[www.robophilosophy.com](http://www.robophilosophy.com)

