DATAIA Project
"BAD NUDGE BAD ROBOT:
Nudge and Ethics in human-machine spoken interaction "

Laurence Devillers

Professor of Computer Science LIMSI-CNRS/Sorbonne Univ.

Team: Affective and social dimensions in spoken interaction with robots

Team: J. Reyes, I. Vasilescu, G. Adda (LIMSI-CNRS)

T. Marquis, S. Pajak*, A. Bensamoun (CERDI, Univ. Paris-Sud), G. Cecere*, F. Le Guel*, M. Manant*, N. Soulié* (*RITM, Univ. Paris-Sud)

Computer sciences/Linguistic & Behavioral economy/Law
2018-2021

INSTITUT DATAIA
Science des données, Intelligence & Société
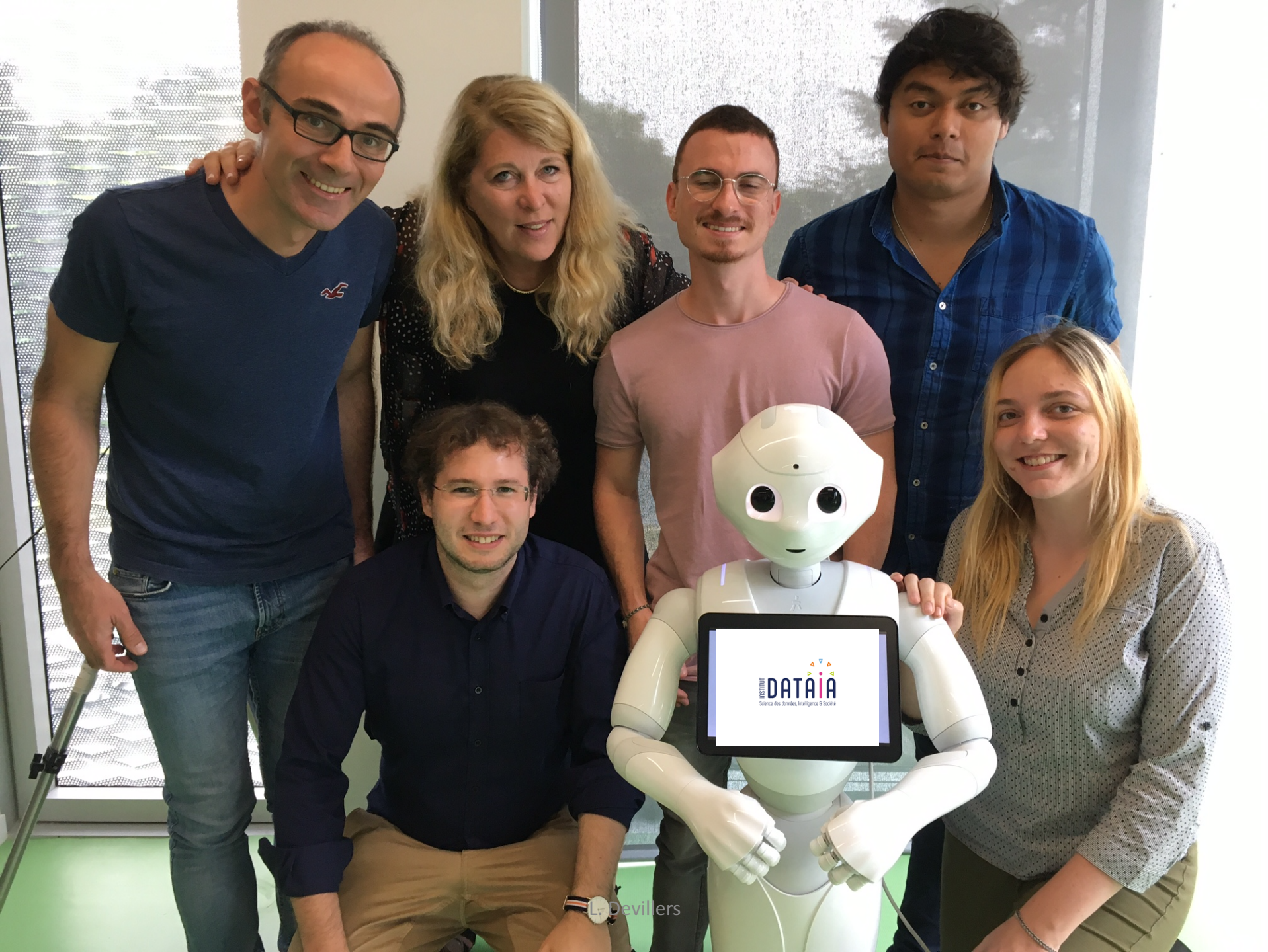
Robotics, AI and philosophy

Developments in robotics/AI quickly lead to deep philosophical questions about

- **Responsibility:** who is responsible for what *autonomous* robots do?

- **The nature of persons:** are robots (that mimic us) *moral persons* now, or in the future?

- **Ethics:** what moral codes should be programmed into robots ?

- **Social philosophy:** should we rethink the basic categories of a modern society after the AI revolution?

Human-Robot co-evolution

- General communication theory, the Media Equation (Reeves & Nass, 1996) explains that people tend to respond to computer/robot as they would either to another person by being polite, cooperative, attributing personality characteristics such as aggressiveness, humor, expertise, and even gender depending on the cues they receive from the computer/robot.

- So an object which seems to be in the pain, as the robot Atlas of Boston Dynamics, can inspire some empathy.

- **We build systems that will adapt to human**

- **How Humans will adapt to the robots that simulate affect?**

- How we can monitor the long-term Human-machine interaction ?

- Risks of dependencies, deshumanization, nudging...

Overview

Introduction
Nudge
Affective computing
Social and Affective Robots

Motivations
AI Applications and nudging
Chatbot, Social robot

ABC : **Ethical considerations**
Auditability
Benchmarking
Confidence

Conclusions

Nudge – some definitions

- Thaler: Nobel Prize in Behavioural Economy, Nov 2017

- Thaler and Sunstein (2008) define a nudge as: "*Any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives*".

- It is important to note that nudging is not a method for restricting choice

A "nudge" is a tactic of subtly modifying behavior of a consumer - Nudging mainly operates through the affective system.
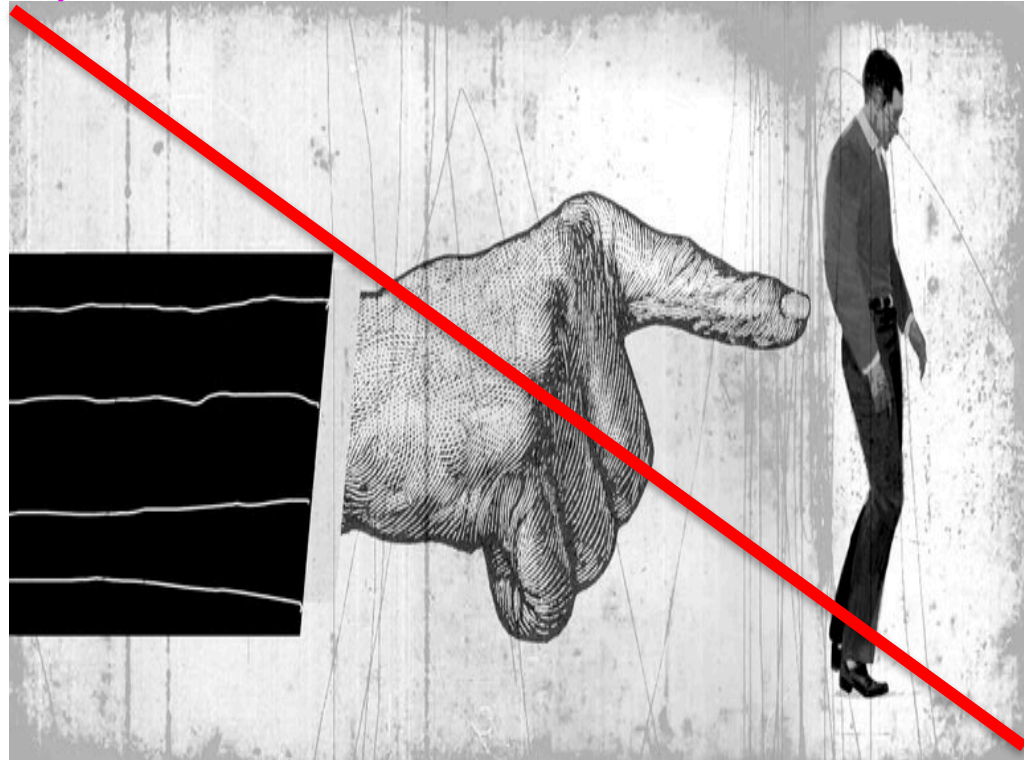
Good - Bad nudges

When systems subtlety or overtly manipulate emotions and alter human behavior for commercial purposes, what kind of transparency and traceability will be present in these systems?

Objectives of our project:

- Create experiments to prove that it is easy to nudge children, adults and elderly people using social and affective robots.

- Affective systems with nudging strategies should be carefully evaluated and monitored -> Find metrics to be able to detect bad nudges.
  - Bad nudges, for example repetitive questions concerning private data, must be detected. Each of these are not — in themselves — enormous problems. But in aggregate, I would argue they constitute a form of nudging which is aggressive in nature.
  - Another efficient bad nudging solution will be friend agents that give us advices for holydays, for health, for fashion to influence our choices.

Creating the reputation as a "Bad-Nudge-free" system for goods and services may be a winning long-run strategy.

What is the difference between ethical nudging and aggressive nudging?

Nudging could be utilized in a near future in chatbots and social robots:

- to incentivize purchase,
- to influence behavior that may be and may not be desired by users

Nudges work by making use of our cognitive biases and « irrational » way in decision-making

- our cognitive capacities are limited,
- we are lacking self-control,
- we act emotionally,
- we act by conformity,
- We act by laziness, etc.

What is affective computing ?

- Computing that relates to, arises from or deliberately influences emotion or other affective phenomena (Picard, MIT Press 1997).

- Giving technology skills of "emotional intelligence" for interacting with us

Interpretation Decision making → Emotion Generation and Synthesis

Emotion detection
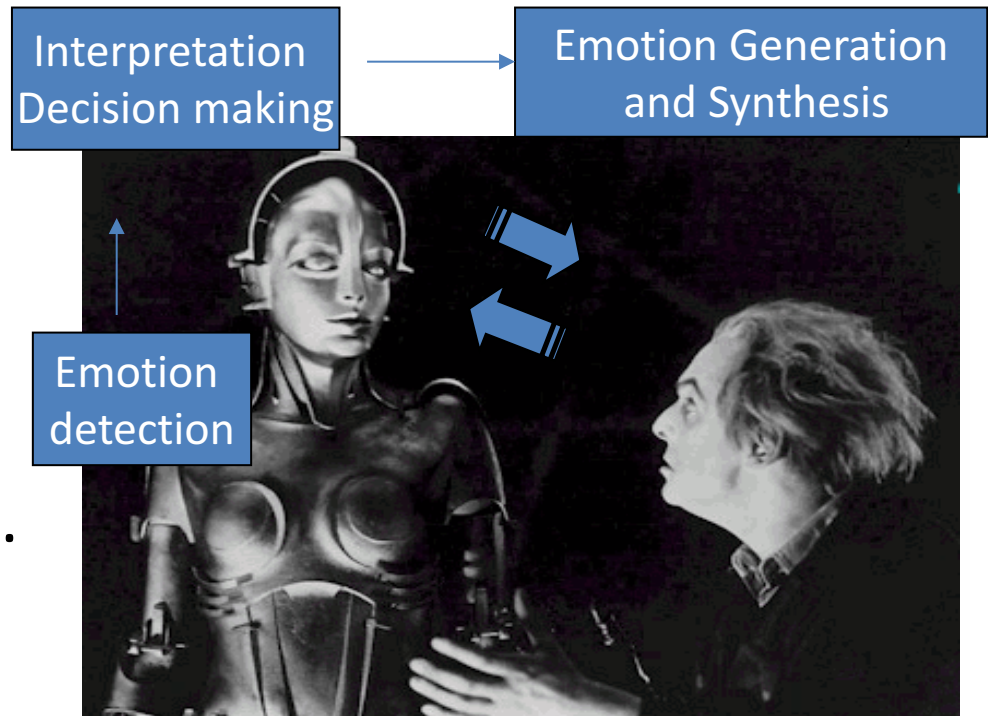
Metropolis (1927)
Fritz Lang

Emotions play a central role in social interaction
How we measure emotions?

3 main functions (K. Scherer) :

- individual reaction to a stimuli,
- physiological and psychological preparation to an appropriate action,
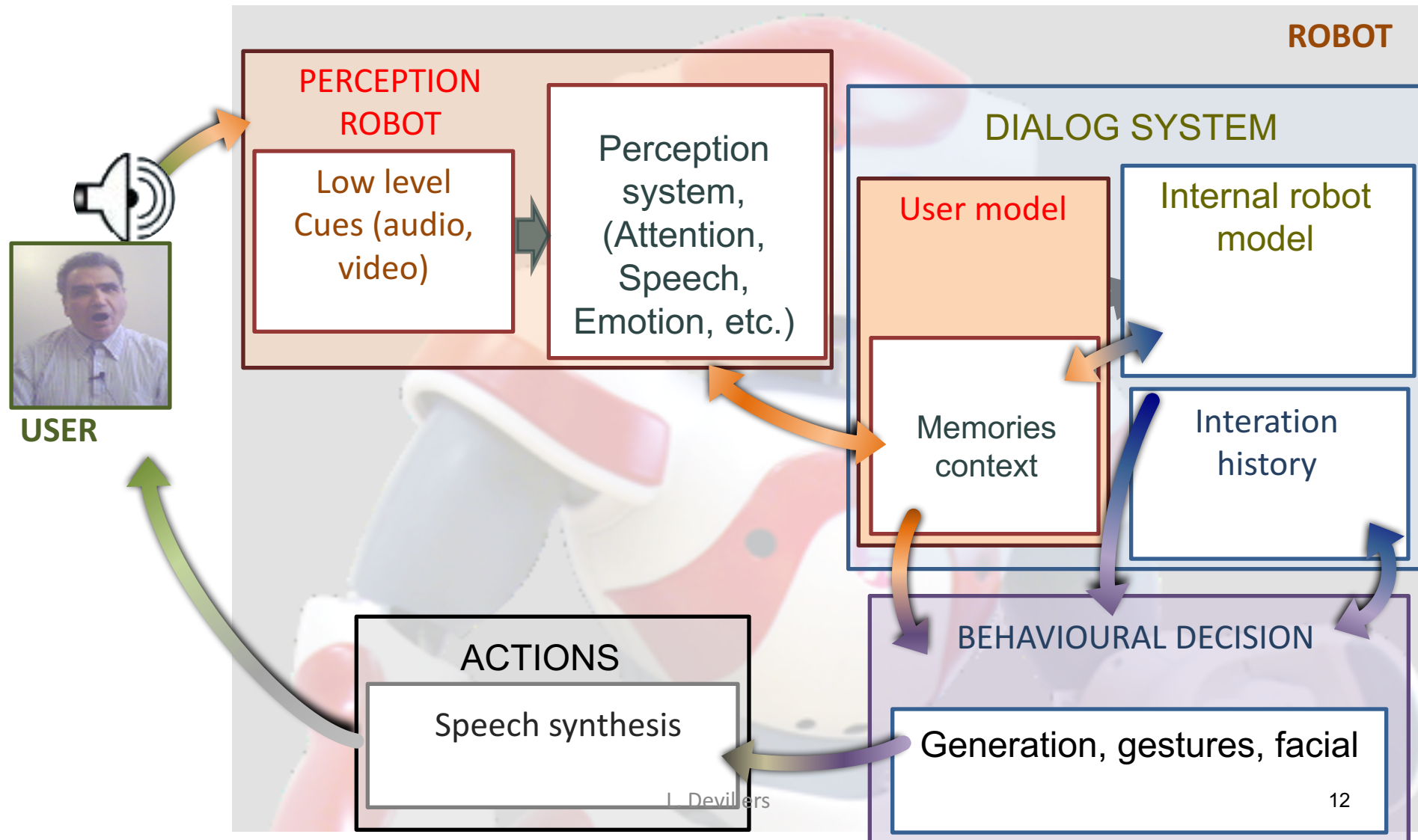- Communicative function of their intentions to another person

- Voice (prosody : timber, energy, rythm)
- Affect bursts
- Face, kinect
- Posture, Gestures, movement,
- Eye tracking,
- Behavior, Action
- Temperature, Respiration, Pupil dilation, Skin conductance
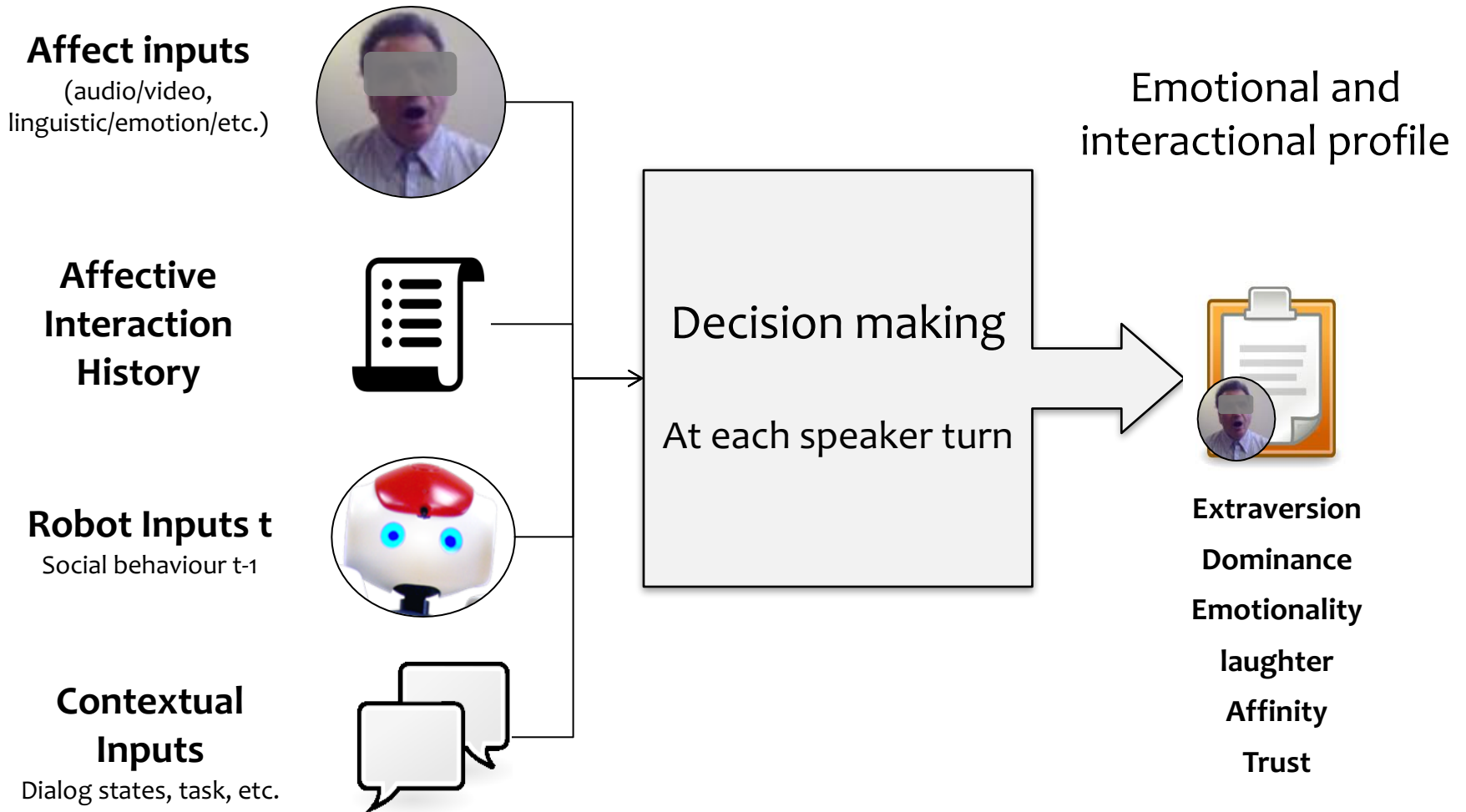- ECG, EEG, Blood pressure, etc.

Social and affective robots

- We built systems and devices that can recognize, interpret, process and simulate human affect

- With the capacity of interpretation of the emotional state of humans, a machine can adapt his behaviour and give an appropriate response to these emotions

- Naturally, it interacts differently with different individuals

ROMEO2 & JOKER projects (2013-2018)

ROBOT

PERCEPTION ROBOT

Low level Cues (audio, video)

Perception system, (Attention, Speech, Emotion, etc.)

DIALOG SYSTEM

User model

Internal robot model

Memories context

Interation history

USER

ACTIONS

Speech synthesis

BEHAVIOURAL DECISION

Generation, gestures, facial

L Devillers

12

User model

Affect inputs
(audio/video,
linguistic/emotion/etc.)

**Affective
Interaction
History**

Robot Inputs t
Social behaviour t-1

**Contextual
Inputs**
Dialog states, task, etc.

Decision making

At each speaker turn

Emotional and
interactional profile

Extraversion

Dominance

Emotionality
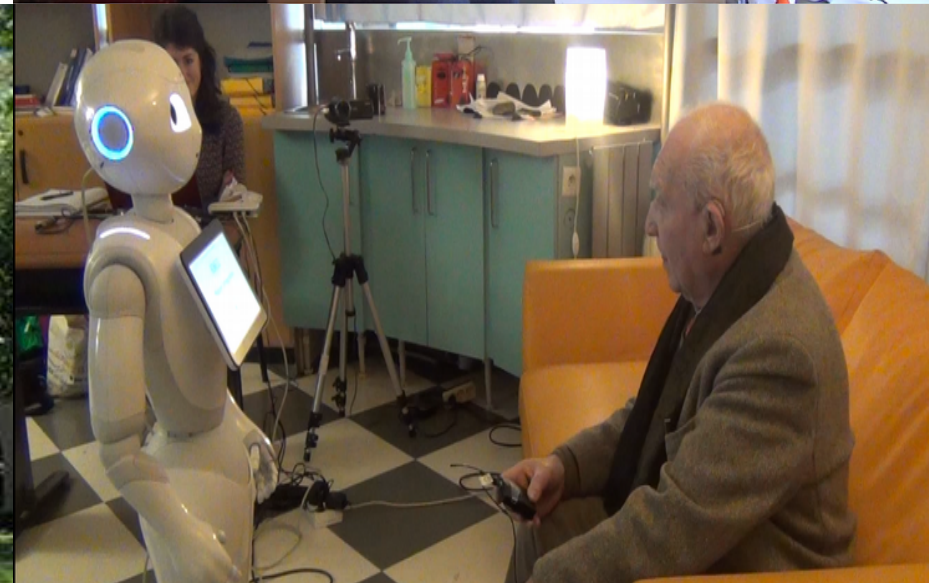
laughter

Affinity

Trust

Real tests
Broca Hospital (Paris, France)

Affective and social Robot

- Adaptive behavior

- Scenarios with elderly

AI applications and nudging

The use of AI applications of nudging are increasing steadily in the world, both within the private and public sector for example
- for health, well-being
- Would AI tools make us drink less alcohol?

Some other examples are:

Fitness technologies like Fitbit that nudge you to be more physically active …

In home AI appliances that help you navigate the web through a voice interface but are also slightly conversational and make small choices based on the systems interactions with the user.

Social robots helping the elderly
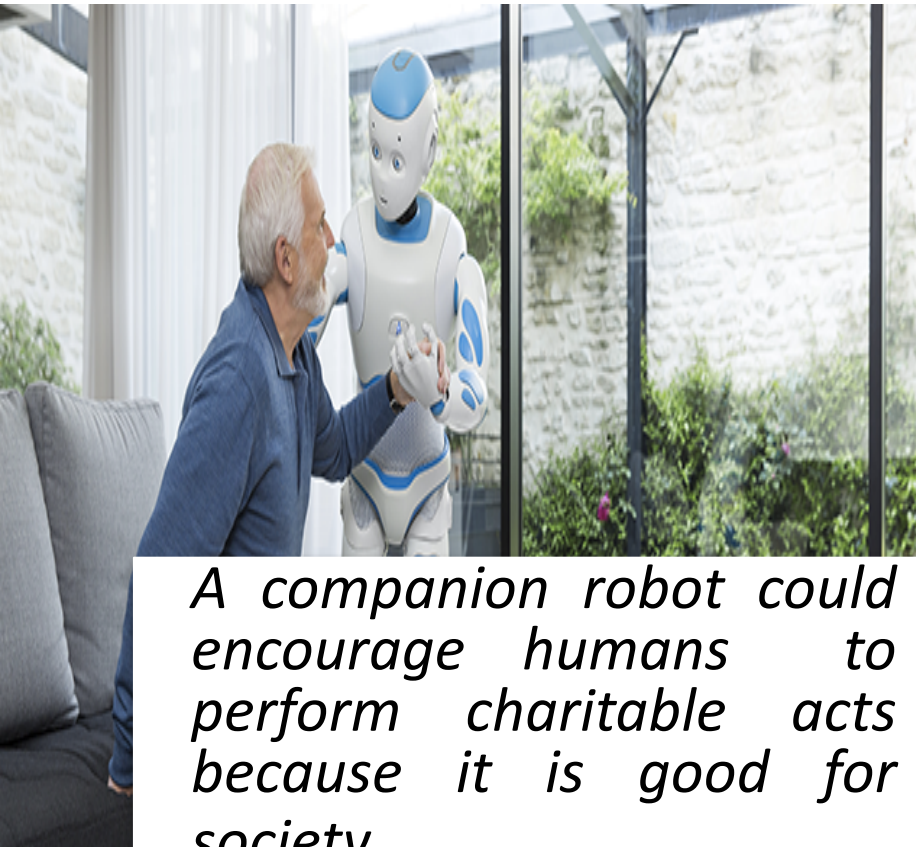
A companion robot could encourage humans to perform charitable acts because it is good for society.
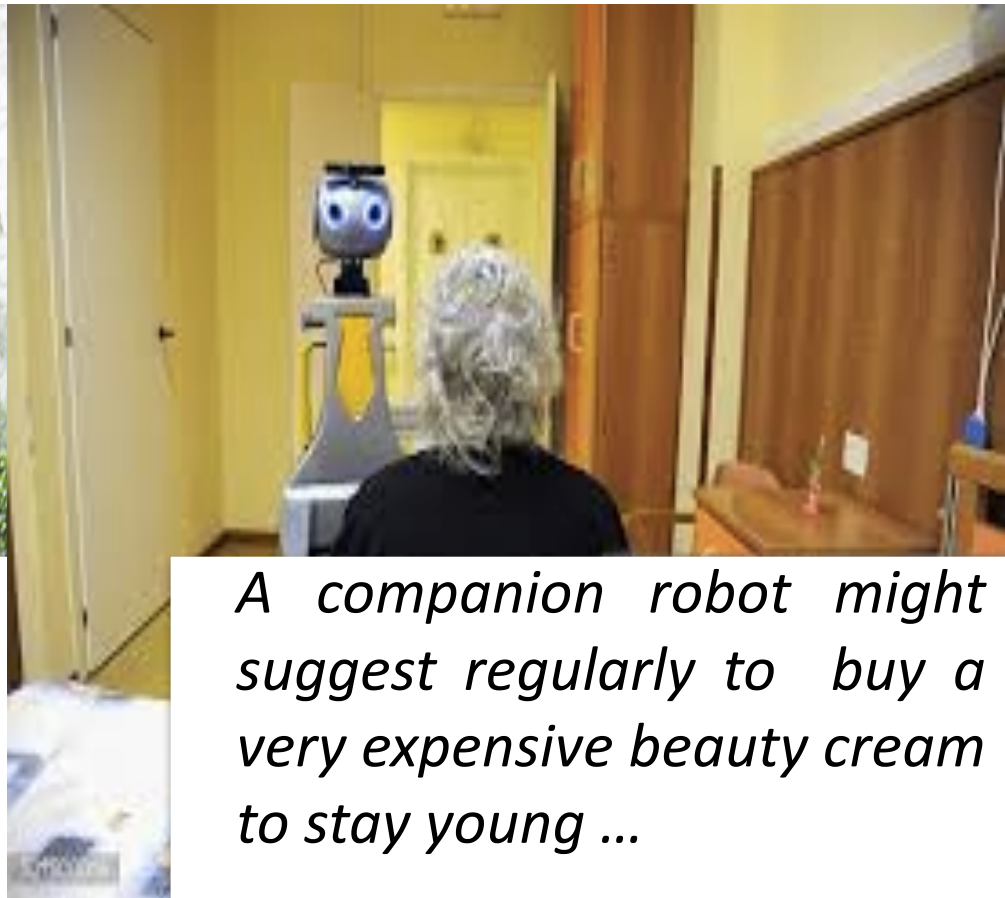
A companion robot might suggest regularly to buy a very expensive beauty cream to stay young …

Laurence Devillers 2018

Ethical considerations:
Auditability, Benchmarking, Confidence

- Consider the level of trust in a robot, in a model, its capabilities and limits and the capabilities and limits of the pair it forms with the user.

- French CERNA

  - Ethics in Robotics research (WG 2012-14): report Nov 2014

  - Machine Learning/AI and Ethics (WG 2016-): report June 2017

- IEEE's Global Ethics Initiative

  - The Global Initiative for Ethical Considerations in the Design of Autonomous Systems (2016) - Affective Computing Committe.

  - P7008 : Standard for ethically driven nudging for Robotic, Intelligent and Autonomous system (2017-2019)

Conclusions:
Responsible Development in robotics/AI - Collaboration with Japan

Develop joint actions aimed at strengthening scientific and innovation synergies is essential

- Social robotics

- Nudging

- Cultural differences

- Ethical considerations

Myths
Fantasies
And Reality
(2017)

L. Devillers
Social robotics,
IEEE ACII
Interspeech
ICMI, HCI, IROS,
IWSDS …

Emotion Perception/ Representation/data collection

Shame and relief : Tears of relief

How do you figure out *what is painful or distressing to* a person? Signs are cultural Many do not show expressive signs consistent with internal emotion that also can be mixed

- Emotion labels: big-six (P. Ekman)
- Dimensions: PAD model: pleasure, arousal, dominance (A. Mehrabian, J. Russell)
- Appraisal theory (K. Scherer)

- **Cultural considerations:**
  – **signs (negation, eyes-contact, touch)**
  – **emotional labels (ex: 1 word for sadness and anger)**

- **Data collection : WOZ / real system with robots / video**