Giulio Antonio Abbo[1] Serena Marchesi[2] Kinga Ciupinska[2]
Agnieszka Wykowska[2] Tony Belpaeme[1]

[1]IDLab-AIRO – Ghent University – imec, Belgium
[2]S4HRI – Istituto Italiano di Tecnologia, Italy

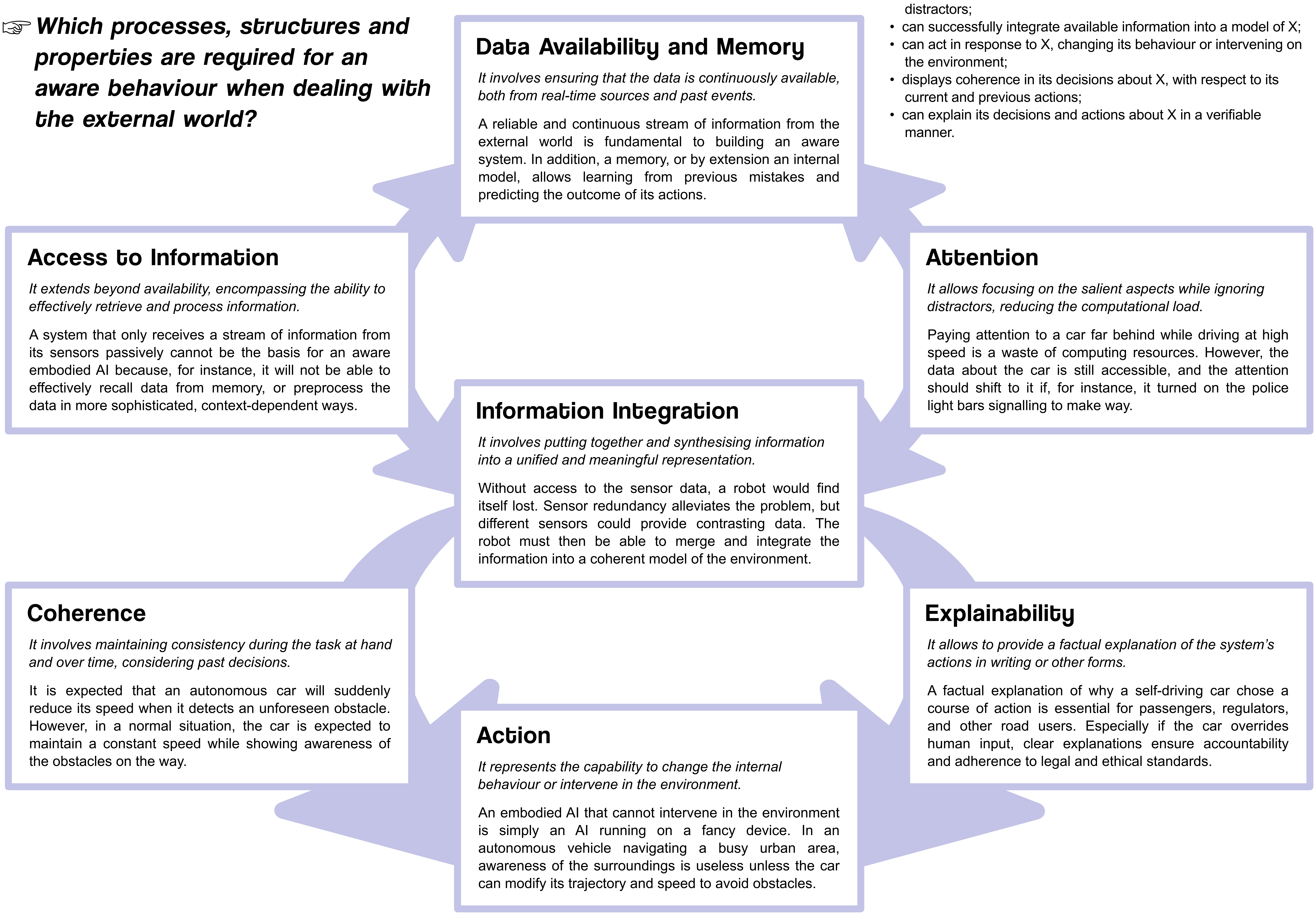
Towards a DEFINITION OF AWARENESS for Embodied AI

☞ ***What are the building blocks for an aware embodied AI?***

☞ ***Which processes, structures and properties are required for an aware behaviour when dealing with the external world?***

We propose to call a system *aware of X* if it:

- has access to information about X, in the form of data availability, memory recall and forward modelling;
- displays an attention mechanism towards X, filtering out distractors;
- can successfully integrate available information into a model of X;
- can act in response to X, changing its behaviour or intervening on the environment;
- displays coherence in its decisions about X, with respect to its current and previous actions;
- can explain its decisions and actions about X in a verifiable manner.



Data Availability and Memory

It involves ensuring that the data is continuously available, both from real-time sources and past events.

A reliable and continuous stream of information from the external world is fundamental to building an aware system. In addition, a memory, or by extension an internal model, allows learning from previous mistakes and predicting the outcome of its actions.

Access to Information

It extends beyond availability, encompassing the ability to effectively retrieve and process information.

A system that only receives a stream of information from its sensors passively cannot be the basis for an aware embodied AI because, for instance, it will not be able to effectively recall data from memory, or preprocess the data in more sophisticated, context-dependent ways.

Attention

It allows focusing on the salient aspects while ignoring distractors, reducing the computational load.

Paying attention to a car far behind while driving at high speed is a waste of computing resources. However, the data about the car is still accessible, and the attention should shift to it if, for instance, it turned on the police light bars signalling to make way.

Information Integration

It involves putting together and synthesising information into a unified and meaningful representation.

Without access to the sensor data, a robot would find itself lost. Sensor redundancy alleviates the problem, but different sensors could provide contrasting data. The robot must then be able to merge and integrate the information into a coherent model of the environment.

Coherence

It involves maintaining consistency during the task at hand and over time, considering past decisions.

It is expected that an autonomous car will suddenly reduce its speed when it detects an unforeseen obstacle. However, in a normal situation, the car is expected to maintain a constant speed while showing awareness of the obstacles on the way.

Explainability

It allows to provide a factual explanation of the system's actions in writing or other forms.

A factual explanation of why a self-driving car chose a course of action is essential for passengers, regulators, and other road users. Especially if the car overrides human input, clear explanations ensure accountability and adherence to legal and ethical standards.

Action

It represents the capability to change the internal behaviour or intervene in the environment.

An embodied AI that cannot intervene in the environment is simply an AI running on a fancy device. In an autonomous vehicle navigating a busy urban area, awareness of the surroundings is useless unless the car can modify its trajectory and speed to avoid obstacles.

Left: *link to the paper.* **Below:** *three images from the paper "I Was Blind but Now I See: Implementing Vision-Enabled Dialogue in Social Robots" showing an LLM with image input capabilities powering a Furhat robot. Screenshots from the webcam are interleaved in the conversation, allowing the system to display enhanced awareness of its surroundings. The system can – from left to right – answer the question "Do you know how to use this?" by understanding to which object the speaker is pointing, identifying it correctly as a coffee machine, and providing detailed step-by-step instructions; compliment the speaker on the jacket and initiating a conversation; identify which of the two garments is most appropriate for a rainy day. The approach proposed in the paper consists of substituting some of the images in the prompt with their textual description, reducing the prompt's length and improving the response speed.*

Do you know how to use this?
That jacket looks perfect for rainy weather...
The jacket on your right looks thicker and might have a hood.

GHENT UNIVERSITY
imec
iit S4HRI ISTITUTO ITALIANO DI TECNOLOGIA SOCIAL COGNITION IN HUMAN-ROBOT INTERACTION
VALAWAI TOWARDS VALUE-AWARE AI
Funded by the European Union