Autonomous Systems Laboratory
Universidad Politécnica de Madrid

About the MetaTool HE Project

Ricardo Sanz

ASLab-TN-2022-002 v 1.1

August 3, 2022

This is a brief note describing the recently signed **Horizon Europe MetaTool Project**. This project addresses the use of *advanced AI architectures* inspired by human neuroscience to improve the capability for bodily and situational understanding of *autonomous robots*. It is one of the projects approved in the Horizon Europe call HORIZON-EIC-2021-PATHFINDERCHALLENGES-01-01 of 2021 on the topic of *Awareness Inside*.

While being an independent, European multipartner project, this project continues the long-term research line on *Autonomous Systems* (ASys) of the *Autonomous Systems Laboratory* of the UPM. The UPM MetaTool team includes also researchers from other UPM research groups on robotics.

The project will start on October 1, 2022 and will be alive for four years.

1 MetaTool Project Content

1.1 About the project

Title: MetaTool: A metapredictive model of synthetic awareness for enabling tool invention

Abstract: Around 3.3 million years ago our ancestors made the first tool. They imagined a new utensil and then knapped a stone until it became an efficient tool for cutting. Tool creation was an outstanding technological milestone for humanity providing us with unprecedented control over our environment. This ability required cognitive capabilities, such as prediction, metacognition, abstraction, and creativity—all of which are associated in humans with awareness. Current artificial intelligence systems and robots largely lack these capabilities and cannot even monitor and evaluate the consequence of their actions let alone develop new tools to address environmental challenges.

METATOOL aims to provide a computational model of synthetic awareness to enhance adaptation and achieve tool invention. This will enable a robot to monitor and self-evaluate its performance, ground and reuse this information for adapting to new circumstances, and finally unlock the possibility of creating new tools. Under the predictive account of awareness, and based on both neuroscientific and archeological evidence, we will: 1) develop a novel computational model of metacognition based on predictive processing (metaprediction) and 2) validate its utility in real robots in two use case scenarios: conditional sequential tasks and tool creation.

METATOOL will provide a blueprint for the next generation of artificial systems and robots that can perform adaptive, and anticipative, control with and without tools (improved technology), self-evaluation (novel explainable AI), and invent new tools (disruptive innovation). Tool-making and tool-invention are outstanding technological milestones in human history. A similar breakthrough can now be envisioned in engineering. We already have algorithms to enable machines to use tools and now it is time to develop robots that create tools.

Partners: The consortium is composed of seven partners from three EU countries (ES, NL, DE) and UK:

1 UNIVERSIDAD POLITECNICA DE MADRID ES - Coordinator
2 STICHTING RADBOUD UNIVERSITEIT - NL
3 TECHNISCHE UNIVERSITEIT DELFT - NL
4 HUMBOLDT-UNIVERSITAET ZU BERLIN - DE
5 SENTA BW - NL
6 PAL ROBOTICS SL - ES 7 THE UNIVERSITY OF SUSSEX - UK

Timespan: 2022-2026.

From UPM: There are two UPM research groups involved in this project:

- **Autonomous Systems Laboratory** (ASLab): Ricardo Sanz, Esther Aguado, Virgilio Gómez
- **Robotics and Cybernetics** (R&C): Claudio Rossi, Ernesto Gambao, Miguel Hernando

Project Manager: Ricardo Sanz, Departamento de Automática, Ingeniería Eléctrica y Electrónica e Informática Industrial. ETS Ingenieros Industriales; coordinator of UPM Autonomous Systems Laboratory.

1.2 Scientific and technical objectives

This project addresses synthetic awareness as a metacognitive process built over the *predictive brain hypothesis*. We particularly focus on the development of a computational model that enables an artificial agent to go from tool use to tool invention as an extension of its own body.

This general aim will be achieved by addressing three specific objectives:

- To study the consistency of metacognitive capabilities and awareness as a facilitator for the emergence of tool making from the perspectives of archaeology and neuroscience. The working hypothesis is that monitoring the uncertainty in a predictive system facilitates cognitive offloading, i.e., to transform complicated outcomes into external novel tools.
- To develop a computational model of synthetic awareness based on metacognition and the predictive processing account of perception and action to unlock the mechanisms of tool invention. We generalize our current models of deep active inference to hierarchical neural predictive filters and extend them with two monitoring metalayers and suitable inductive learning and processing biases to foster the development of the representation needed for achieving tool invention.
- To validate our synthetic awareness technology on embodied artificial agents. We evaluate the model in AI-based simulated and robotic systems in two use cases: conditional task problem solving under uncertainty and tool invention. The goal is to achieve Technology Readiness Level (TRL)-4 and TRL-3 respectively.

In UPM we will establish a testbed using PAL Robotics Thiago robots to explore tool use and tool creation and establish a benchmark where to measure the potential advances of the ”awareness inside” approach.

1.3 Project Organization

The project is organised into six work packages and will have a duration of four years. Figure 1 shows the overall organization of the work. The six workpackages are the following:

WP1 Foundations leaded by TU Delft develops the anthropological tool use fundamentals.

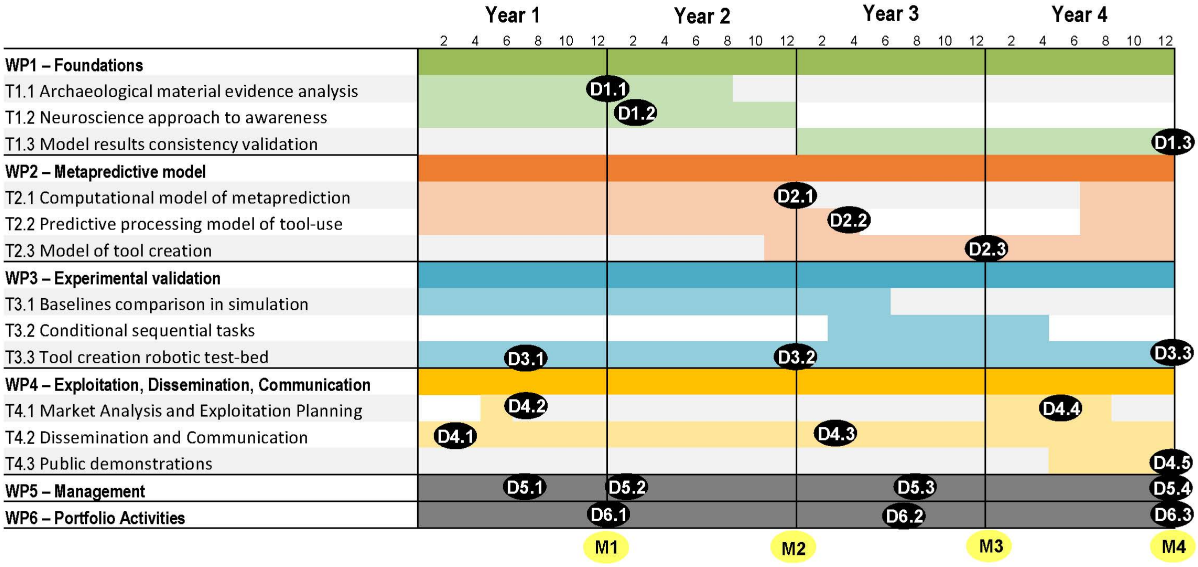

Figure 1: Overal organization of project activities.

WP2 Metapredictive Module leaded by SRU develops the systems neuroscience architecture for tool-extensible self-awareness.

WP3 Experimental Validation leaded by UPM addresses the use of the architecture in tool-oriented improvement of behaviour of robots.

WP4 Communication, Dissemination and Exploitation leaded by SENTA coordinates project communication activities.

WP5 Project Management is leaded by UPM.

WP6 Porfolio Activities is leaded by UPM to address collaboration with other projects in the call.

2 EC Programme Topic

This section describes the Horizon Europe workprogramme topic that this project is addressing. This information comes from the Funding & Tenders portal of the EC.

Topic: EIC Pathfinder challenge: Awareness inside

Topic ID: HORIZON-EIC-2021-PATHFINDERCHALLENGES-01-01

Programme: Horizon Europe Framework Programme (HORIZON)

Call: EIC Pathfinder challenges

Type of action: HORIZON-RIA — HORIZON Research and Innovation Actions

2.1 Topic description

Awareness and consciousness have been on the Artificial Intelligence agenda for decades. Progress has been difficult because it has been hard to agree on exactly what it means to be aware. Most researches would agree though that we do not have any truly aware artificial system yet, that awareness is much more than a sensorial sophistication and that it is much more than any Artificial Intelligence as we know it. But, what is it then that a user would expect from a service or device that has 'awareness inside'?

Most scientific and philosophical accounts of awareness are based on a human subject perspective and at an individual level. They address the question of what it means for an individual human subject to be aware of, e.g., the environment, time or oneself and how one can assess awareness in this context. The problem is relevant, certainly, since many clinical and cognitive conditions can be linked to awareness issues. The concept is also relevant to emerging technologies as it has been argued, for instance, that humans will not accept robots (or chatbots, or decision support systems) as trustable partners if they cannot ascribe some form of awareness and true understanding to them.

The individual human-centric concept of consciousness hinders the application of awareness as a measurable feature of any sufficiently complex system. The study of awareness in other species and artefacts, or even more elusive concepts such as social awareness require a new perspective applicable to many systems. It can then also serve to attack the inter-subjective state and experience of awareness (i.e., what is it like to interact with an aware robot that, most probably, does not have the same kind of awareness than the human?), or to include non-conscious objects into the sphere of awareness (e.g., to become aware of the time without looking at the watch).

For technologies, awareness principles would allow a step-up in engineering complex systems, making them more resilient, self-developing and human-centric. Awareness is a prerequisite for a real and contextualised understanding of a problem or situation and to adapt ones actions (and their consequences) to the specific circumstances. Ultimately, awareness serves the coherent and purposeful behaviour, learning, adaptation and self-development of intelligent systems over longer periods of time.

Specific conditions for this challenge Proposals are expected to address each of the following three expected outcomes:

- New concepts of awareness that are applicable to systems other than human, including technological ones, with implications of how it can be recognised or measured. It will require to elucidate the relationship between, among others, complexity and awareness, information structure and representation, the environment and its perception, distributed versus centralized awareness, and time awareness. This will lead to better approaches for defining aspects of awareness over different temporal, spatial, biological, technological and social scales.
- Demonstrate and validate the role and added-value of such an awareness in an aware technology, class of artefacts or services for which the awareness features lead to a truly different quality in terms of, e.g., performance, flexibility, reliability or user-experience. The specific expected outcome is a proof of principle of technologies far beyond the current state of the art or a laboratory-validated prototype enabling evaluation of the proposed technology's awareness features, relying where relevant on neuroscientific and psychological methods, and possibly in a range of application areas. As examples, projects could investigate the implications of 'awareness inside' for safer robots or self-driving cars, for better resilience of critical infrastructure, in artefacts that compensate for consciousness disorders, in decision support (e.g. for surgery, economics or epidemiology), or for chatbot-based conversation, language learning or translation.
- Define an integrative approach for awareness engineering, its technological toolbox, the needs and implications and its limits, including ethical and regulatory requirements. On this aspect specifically, the projects that will be funded under this challenge are expected to collaborate and contribute to the wider ethical, societal and regulatory debate since, ultimately, new awareness concepts may lead to a redefinition of how we look at the relation between humans, other species and smart technologies. The gender dimension in research content should be taken into account, where relevant, to maximise user experience.

Specific Topic Conditions: Proposals are required to comply with the Trustworthy Artificial Intelligence principles.

Title:	About the MetaTool HE Project
Author:	Ricardo Sanz
Date:	August 3, 2022
Reference:	ASLab-TN-2022-002 v 1.1 Final
URL:	http://www.aslab.upm.es/doc/controlled/ASLab-TN-2022-002-v1.1.pdf

UPM ASLab MetaTool