The Agent Metamodel in CATALINA (Cognitive AgenT prActicaL reasonINg Architecture)

Massimo Cossentino[1], Giovanni Pilato[1], Guido Averna[1], Myrto Mylopoulos[2] and John Mylopoulos[3]

[1]*National Research Council of Italy (CNR), Via U. La Malfa, 153, Palermo, 90146, Italy*
[2]*Carleton University, 1125 Colonel By Drive, Ottawa, Canada*
[3]*University of Toronto, Toronto, Canada*

Abstract

Several recent studies explore the relevance of the different aspects of consciousness in agent reasoning, often starting from cognitive science studies. Besides, the agent community always found a relevant reference in the practical reasoning paradigm proposed by Bratman in his works. The proposed CATALINA architecture aims to blend contributions from these two worlds by implementing a refined attention mechanism together with the BDI reasoning. The core of CATALINA is the Global Workspace (GW), inspired by the Baars' Global Workspace Theory. Agent's brain functionalities are modularised as executive functions, taking inspiration from Buheler's theory of the brain executive system. Among them, the Reasoner Function is implemented as a practical reasoner resembling Bratman's architecture. Blending these theories poses several challenges, among them the definition of an underlying metamodel that ensures coherence and collaboration among the ideas coming from the inspiring theories. To evaluate the performance of CATALINA's attentional and deliberative mechanisms, we developed a simulation of an autonomous vehicle navigating between cities that receives several desires from the designer and, in executing them, shifts its attention among different tasks.

Keywords

Global Workspace Theory, Executive Functions, BDI Agent, Practical Reasoning, Agent Metamodel, Goal-Oriented Reasoning

1. Introduction

Designing autonomous agents capable of exhibiting rational deliberation and adaptive responsiveness continues to be one of the most interesting challenges in artificial intelligence.

Conventional agent architectures typically consist of two main approaches. The first deals with reactive behavior, which enables agents to react quickly to events or changes in their environment; this makes it possible to handle sudden events or new challenges efficiently. The second approach involves deliberative reasoning, which includes proper planning and data analysis to achieve non-trivial, long-term goals. This methodology supports the agents in making their decisions and managing their behavior accordingly.

However, the aforementioned approaches are often developed independently; as a consequence, this can limit the overall effectiveness of the agent's behaviour in unpredictable and dynamic environments.

The practical reasoning architecture, initially proposed by Bratman [1], is a foundational and widely adopted framework for modeling goal-directed, resource-constrained agents with practical reasoning capabilities. Despite its broad application, the traditional BDI model lacks mechanisms for attentional control, i.e., procedures that enable dynamic shifting of focus between goals in response to changing perceptual input. Moreover, it does not take into account or include mechanisms related to consciousness or awareness.

EUMAS 2025, The 22nd European Conference on Multi-Agent Systems, September 3-5, 2025, Bucharest, Romania
*Massimo Cossentino.
massimo.cossentino@icar.cnr.it (M. Cossentino); giovanni.pilato@icar.cnr.it (G. Pilato); guido.averna@icar.cnr.it (G. Averna); myrto.mylopoulos@gmail.com (M. Mylopoulos); jm@cs.toronto.edu (J. Mylopoulos)
https://www.pa.icar.cnr.it/cossentino/ (M. Cossentino)
0000-0003-1258-9767 (M. Cossentino); 0000-0002-6254-2249 (G. Pilato); 0000-0002-0935-5866 (G. Averna); 0000-0003-2260-7068 (M. Mylopoulos); 0000-0002-8698-3292 (J. Mylopoulos)

Understanding the processes that give rise to consciousness remains one of the central challenges in cognitive science. In parallel, a key question in Artificial Intelligence is whether it is possible to design systems replicating some of the functional characteristics associated with consciousness. Research in robot consciousness, for example, aims to address both goals: to develop artificial agents capable of exhibiting aspects of experiential and functional consciousness and to use these agents as experimental platforms for advancing our understanding of consciousness in biological systems.

Consequently, creating software architectures that integrate attention, awareness, and self-regulation functions is a significant step in this direction.

There are many opposing theories of consciousness. In our work, we adopt a theory-neutral approach, focusing instead on implementing selected functional components that are broadly compatible with leading models.

In particular, we propose an extension to the BDI architecture that integrates two forms of control related to attention: one inspired by Baars' Global Workspace Theory of consciousness and the other grounded in Buehler's account the role of executive functions in action guidance and regulation.

This paper is a further step in the evolution of the CATALINA architecture initially described in [2] and it will discuss the CATALINA agent metamodel [3]. More specifically, this paper specifically focuses on the four types of desires supported by this architecture. It will also explore the connection between the BDI agent's desires and its attention mechanism. By incorporating goal modelling, this work expands upon the conventional BDI framework. Specifically, we separate desires into four categories: *practical*, *quality-related*, *epistemic*, and *green* (i.e. environmentally-oriented) desires. In our approach, the attention modulation mechanism controls the activation of the first two categories (practical and epistemic); they are collectively referred to as attentional desires.

In CATALINA, the classic concept of BDI desire is represented by the *Active Desire* that is an attentional desire that has passed the attention filter. This distinguishes the standing, innate desires injected by the designer into the agent from the active desires the reasoner is processing to find an option that could achieve them and may deliberate their promotion to intentions. In fact, intentions are desires that the agent decides to pursue actively. We will also report a metamodel that describes the agent's desires and links them to the main elements of a BDI agent.

At the heart of CATALINA, whose architecture is depicted in Fig. 1, lies the Global Workspace (GW), a shared working memory that mediates the flow of information among the other components of the architecture, and the Executive Functions inspired by Buehler's account. Among other information, the GW shares the agent's beliefs, represented as structured predicates with associated truth values. By modulating access to the GW, the executive functions, inspired by human cognitive control systems, allow the agent to maintain its working and long-term memories, allocate resources, reason about desires, and change focus in accordance with ongoing objectives.

CATALINA agents can balance rational goal pursuit with reactivity to salient stimuli through the integration of attentional filtering on incoming stimuli. This enables them to perform practical reasoning without overwhelming memory or huge processing requirements.

To evaluate the performance of CATALINA's attentional and deliberative mechanisms, we have set up a simulation of an autonomous vehicle navigating between cities on a known map under changing task demands and environmental respect conditions.

This demonstration highlights the architecture's capability to shift attention, re-prioritise intentions, and remain sensitive to ethically and ecologically relevant constraints.

The remainder of the paper is organized as follows: the next section introduces the theoretical background at the basis of the proposed approach. Section 3 recaps the CATALINA architecture, Section 4 illustrates the CATALINA agent metamodel, Section 5 describes the four types of desires taken into consideration, Section 6 provides a description of the attention modulation shifting capability of CATALINA, including the description of two simulation scenarios taken into consideration, and finally in Section 7 we debate the proposed architecture, we draw some conclusions and outline our future works.

2. Theoretical Background

In this section, we briefly recap three frameworks on which our proposed architecture is based: the Bratman's Belief-Desire-Intention (aka BDI) model [4], the Baars' Global Workspace Theory [5], and the Buehler's account of the executive system [6].

2.1. Bratman's Practical Reasoner

Bratman's Practical Reasoner [4], also known as the BDI (Belief, Desire, Intention) model, defines the classical practical reasoning architecture for a resource-bounded agent, in particular with respect to computational resources and time. The BDI model is an abstract architecture for plan-based reasoning that exploits the agent knowledge about the world and its current state, the agent's desires and the agent's commitment allows the agent to reason, implement intentions and make decisions.

A *belief* is an agent's knowledge about the world, typically expressed as logical predicates. Beliefs reflect what the agent considers true about itself, others, and the environment. They are dynamic, changing over time based on perceptions, input from other agents, or a priori reasoning.

A *desire* is a mental state aimed at achieving a world state the agent values, which acts as motivation for forming intentions and taking actions.

An *intention* is the agent's commitment to act in order to fulfil a desire. It is realized through plans that guide actions toward the desired outcome. In general, intentions must align with the agent's beliefs and desires.

In the BDI model, beliefs, desires, plans (options), and intentions are data libraries. The agent also includes psychological components, like Means-End Reasoner, Opportunity Analyzer, Filtering Process, and Deliberation Process, that carry out different decision-making functions.

2.2. Baars' Global Workspace Theory

The Global Workspace Theory (GWT) of consciousness, introduced by Baars [5, 7], plays a foundational role in the approach developed in the following sections, as it delineates key consciousness-related functions intended for integration into artificial agents. GWT posits that information becomes conscious when it is globally broadcast to a broad array of specialized psychological subsystems, such as those responsible for planning and verbal reporting. In contrast, unconscious information remains confined to isolated modular subsystems.

Baars illustrates this dynamic using a theatrical metaphor [5]: consciousness is likened to a spotlight illuminating a specific area on the stage of immediate memory, directed by executive attention, while the surrounding area remains in darkness, representing the unconscious. From a cognitive architecture standpoint, GWT conceptualizes a shared memory structure that facilitates the distribution of information across various functionally distinct modules. These modules handle specific cognitive tasks, including sensory processing, environmental evaluation, motor control, and language, often in coordinated interaction.

Functionally, the GW operates similarly to a central communication hub in a network, without the constraints of data collision, where each module acts as a discrete processing unit. Modules interact with the GW to retrieve or deposit information, making the GW a crucial component for filtering, integrating, and accessing data, particularly from long-term memory [5] [8].

2.3. Buehler's The Executive System

Buehler's account of the executive system [6], as described in cognitive psychology, outlines a central control structure that manages various psychological subsystems, particularly those related to attention. This system is composed of four key functions: the *Executive Switching Function*, which coordinates the agent's actions by selecting and shifting between abilities based on goals, directing attention to relevant beliefs and representations in the Global Workspace (GW), and handling priority assessments, memory access, and motor control; the *Executive Inhibition Function*, which suppresses irrelevant

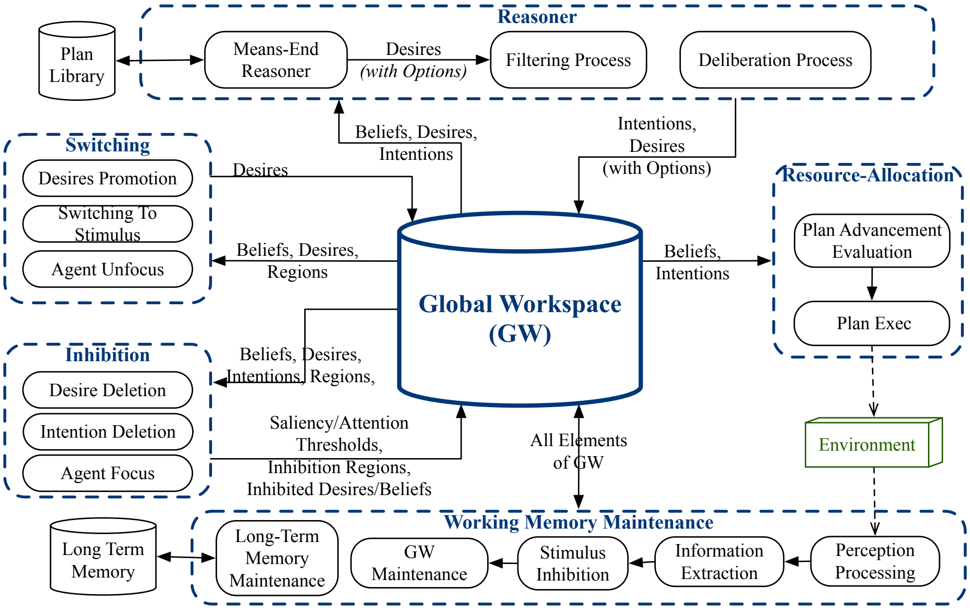

Figure 1: The CATALINA Architecture

processes and distractions, refining attention by filtering out non-pertinent information (e.g., excluding irrelevant areas in visual searches) and helping maintain goal focus through inhibition regions; the *Executive Resource Allocation Function*, which acts as a workflow manager, orchestrating necessary functions and resources to execute plans aligned with the agent's goals; finally the *Executive Working Memory Maintenance Function* manages the flow of information between long-term memory and the GW, ensuring only relevant data is retained or refreshed, and outdated information is discarded to prevent cognitive overload.

These functions support mechanisms of attentional modulation, distinguishing between top-down (*endogenous*) attention—voluntary and goal-driven—and bottom-up (*exogenous*) attention—stimulus-driven and automatic. Attention is influenced by biases, which guide behaviour based on either salient stimuli (*bottom-up*) or intentional goals and learned values (*top-down*). These biases shape how agents prioritize and respond to environmental inputs [6].

3. A Conscious Agent Reasoning Architecture (CATALINA)

The CATALINA architecture has been introduced in[2] and further evolved as reported in [3] when some internal modules of the Inhibition and Switching Executive Functions have been partitioned into smaller and more specialised elements. CATALINA integrates both consciousness-related capabilities and focused attention mechanisms, and it achieves this by establishing a complex interaction between different executive functions. This interaction is mediated by the Global Workspace, where the agent's attention is selectively focused on the specific knowledge domain relevant to its current intentions.

Our approach considers that the agent's knowledge is instantiated in the Global Workspace (GW), as conceptualised by Baars [5],[7], in the form of beliefs. The behaviour of the agent's mental states is implemented through the Executive Functions, a framework discussed by Buehler [9][10][6]. The reasoning capabilities proposed by Bratman [11][4] are provided by the Reasoner Function of CATALINA. Further details about the cited theories may also be found in [2].

A CATALINA agent is a rational agent that performs practical reasoning while maintaining a conscious awareness of the environment (and of itself). It also focuses its attention on current tasks without compromising the capability to react to salient stimuli that have a higher priority. This generates an interesting mix of features that allow the agent to perform rational reasoning (usually a computationally demanding feature) without the need for extensive memory and limiting the CPU requirements. This happens because by focusing its attention, the agent limits the amount of considered data, and only the tasks at hand are momentarily active while all the remaining information is stored in long-term memory. At the same time, the agent is able to promptly respond to new stimuli (just like a reactive agent) when they are significantly more important (salient) than the currently running tasks; vice versa, less relevant stimuli are not immediately processed and they can be the objective of future attention, if the case.

The following subsections, one for each of the CATALINA's components, briefly recap the main functionalities of CATALINA. Each executive function can be further decomposed into modules responsible for specific aspects of its behaviour.

3.1. The Global Workspace

The GW works as a publish-subscribe dashboard, processing incoming messages such as belief updates and generating outgoing events that notify registered functions of these updates. This enables each function to access the GW and retrieve updated information when needed, thereby ensuring that the GW's operations do not interfere with the autonomous functioning of individual modules.

As an example, let us suppose that the perception management modules of the WMM function produce a new belief representing the detection of a dangerous situation. The Perception Processing and Information Extraction modules cooperate to define the saliency of the perception. The Stimulus_Inhibition module evaluates whether activating an epistemic desire that monitors the situation is appropriate according to the current values of the thresholds and the stimulus's saliency, and then posts the stimulus to the GW. This communicates the presence of an updated belief to the Switching Function through proper notification events. Subsequently, the Switching Function promotes the stimulus to active desire, which involves revising the current set of desires. For instance, if one of these desires conflicts with other conditions or environmental factors, it may be deleted or revised accordingly.

The Reasoner Function then processes the new desire, searching for alternative options and storing them in the GW. Concurrently, the Reasoner Function reasons about the opportunity offered by the updated beliefs, revising the current options and intentions to align with these updates.

Finally, the GW notifies the Resource Allocation Function of the updated beliefs and intentions, enabling this module to execute new intentions while considering the world's revised state. The actions generated by this function will alter the environment, developing new perceptions that are subsequently processed by the Working Memory Maintenance Function, thereby restarting the loop. Further details about the agent working cycle are reported in [3].

3.2. The Executive Inhibition Function

The Executive Inhibition Function (EIF) plays a crucial role in modulating attention and allocating resources to facilitate goal-directed behaviour. Its primary functions are twofold: first, it implements the attention modulation mechanism, which is mediated by the Focus Attention module; second, it generates inhibition regions that effectively limit the agent's perception to a specific area of interest.

Through its attention modulation function, the EIF revises and refines the saliency and attention thresholds, ensuring that the agent's focus is directed toward the most relevant information. These inhibition regions serve as cognitive filters, allowing the agent to concentrate its attention on a particular portion of the environment while ignoring irrelevant stimuli.

Furthermore, the EIF plays a critical role in regulating the processing of beliefs not pertinent to the agent's current intentions. By inhibiting the processing of such beliefs, the EIF enables the agent to maintain focus on desires and intentions, thereby facilitating efficient information processing and decision-making. This mechanism prevents the accumulation of irrelevant information, which could lead to cognitive overload and decreased performance.

3.3. The Executive Switching Function

CATALINA models attention propagation in two directions: Endogenous Attention Modulation (EnAM) and Exogenous Attention Modulation (ExAM). EnAM involves top-down attention, where the agent focuses on desires that align with its intentional efforts. In contrast, ExAM is driven by external stimuli that may bypass the current attention threshold, triggering the agent to redirect its attention toward novel or salient information.

The EnAM process generates new desires for the agent, but, in practice, it can only pursue a subset of its desires. This can be due to the limitations of the agent's cognitive resources and the need to prioritise tasks. On the other hand, the ExAM process involves how attention is captured by external cues, such as visual or auditory stimuli, which are processed through belief updates.

When an agent perceives meaningful information, it automatically directs its attention toward processing the stimulus, often generating a new epistemic active desire to investigate the source of that perception. For instance, if an agent notices a novel object in its environment, it may generate a desire to explore the object's properties or behaviour.

The Executive Switching Function is crucial in mediating this process by evaluating whether a belief update from the Global Workspace (GW) warrants revising the agent's current desires. If the new information has higher saliency than the current attention threshold, the function considers whether it should add new desires or modify existing ones. Either way, the agent adjusts its desires and priorities in response to the updated information.

Sometimes, belief updates may trigger a bottom-up attention modulation mechanism that prompts the agent to pursue novel epistemic desires. These desires often aim to explore the origin of the perception that generated the new belief, allowing the agent to expand its knowledge and understanding of the world.

3.4. The Executive Resource-Allocation (RA) Function

The Executive Resource Allocation (RA) Function is responsible for executing the agent's intention-embedded plan, which involves coordinating multiple actions to achieve its desires. This function ensures efficient management, distribution, and equilibrium of resources.

In the current implementation, the RA Function simply executes a list of actions that comprise the plan by invoking the corresponding behaviors of the agent. However, this approach has limitations since, in the current implementation of CATALINA, it only supports linear action execution. In a future and more advanced version of CATALINA, we will incorporate a workflow engine, which will enable the handling of complex plans involving parallel flows of actions.

3.5. The Executive Working Memory Maintenance (WMM) Function

The Executive Working Memory Maintenance (WMM) Function plays a crucial role in managing and updating information stored in both Long-Term Memory and the Global Workspace. It is responsible for refining data that enters the system through sensors, filtering out irrelevant or redundant information based on the inhibited regions generated by the Endogenous Input (EI) Function.

For instance, when an autonomous vehicle's camera captures an image of the road ahead, the WMM Function removes non-essential elements from the image, such as the sky or background distractions, to focus attention on relevant features like lanes and obstacles. This process allows efficient processing of salient stimuli in the Global Workspace, enabling the system to respond effectively to its environment.

3.6. The Reasoner Function

The Reasoner Function is a higher-level cognitive process inspired by Bratman's reasoner. It consists of four modules: the Means-End Reasoner searches for plans in the Plan Repository to satisfy new desires.The Filtering Process evaluates the quality attributes of options and selects those that meet the agent's green desires. Finally, the Deliberation Process decides which option best fulfils each desire,

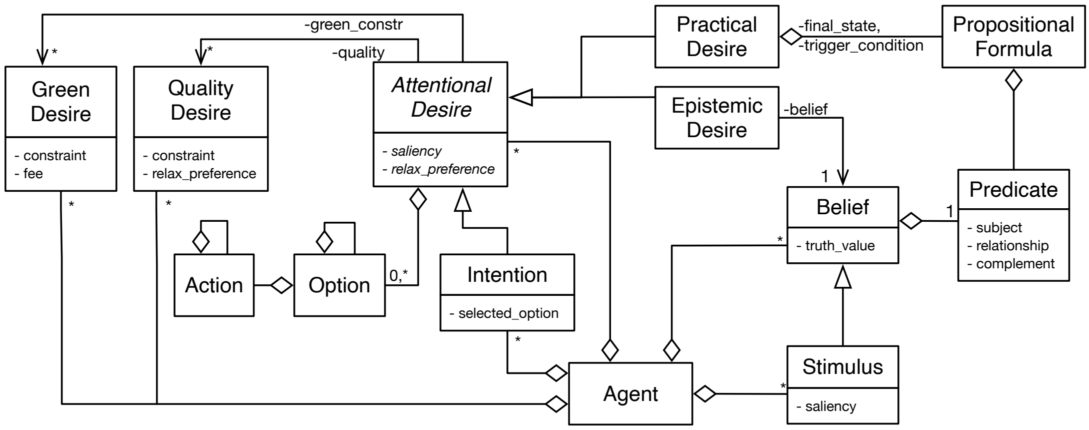

Figure 2: The CATALINA Agent Metamodel

taking into account potential trade-offs between conflicting desires or relaxed constraints. The Reasoner Function workflow is the following:

1 The Means-End Reasoner searches for plans to satisfy new desires.
2 The Filtering Process selects options that meet the agent's desires.
3 The Deliberation Process decides which is the best option, considering trade-offs and relaxed constraints.

4. The CATALINA Agent Metamodel

The metamodel that is behind the agent implemented using the CATALINA architecture has some peculiarities descending from the need to blend three different theories (Baars' GWT [7], Buheler's Executive Systems theory [10, 12], Bratman's Practical Reasoning [11]) in one unique novel approach.

The CATALINA agent metamodel is represented in Fig. 2. The agent's knowledge is made of beliefs expressed by predicates with some truth value. In the current implementation of the CATALINA architecture, predicates are represented in the form:

```
Predicate = <subject> + <relationship> + <complement>
```

This allows to easily represent predicates like:

```
I move the chair, I look_at the traffic light, I visited Paris, ...
```

As said, a belief adds a truth value to the predicate, thus stating whether the agent believes that predicate to be true or not.

A stimulus is a peculiar type of belief. It usually originates from perceptions and has a saliency that represents the relevance or urgency of the percept. Assigning the right saliency value to any perception is a relevant and domain-dependent challenge. For this reason, in CATALINA, we encapsulated this task into a method that can be properly coded for each application. Stimuli are very relevant to the proposed approach for merging the three cited cognitive theories; in CATALINA, stimuli are directly connected to the triggering of endogenous desires. Such desires, usually epistemic, are the agent's response to some change in the environment. As an example, if an autonomous vehicle driving along a road sees dense smoke ahead, it reacts by triggering an epistemic desire to understand if the smoke may endanger its route.

The concepts of desires and intentions play a fundamental role in the agent's mind. We here present an extension of the BDI theory [11] since we identify two different states for desires.

A (standing) desire in CATALINA is a prodromic active desire (that is more aligned to Bratman's idea of desire); it comes from the requirement of supporting goals injected into the agent by the designer and allowing the attention mechanism filter [6] them before they become active desire. Active desires are not represented in Figure 2 since *Active* and *Standing* are two different states of *Attentional Desire*, an abstract concept which can be specialised into Practical or Epistemic desires.

An active desire is an attentional desire (therefore a practical or epistemic desire) that has passed the attention filter; it may also have options (a plan of action) that the agent may use to fulfil the final condition prescribed by the desire. In fact, CATALINA agents try to fulfil active desires by continuously looking for options that allow their pursuit. If options are found, the agent evaluates the possibility of promoting the desire to intention and enacting the related option. Conversely, standing desires are not considered for option generation until they pass the attention filter and become actual desires.

More specifically, we consider four categories of desires: practical, epistemic, quality, and green desires. As already said, we also refer to the first two categories as attentional desires since their triggering is subject to the attention modulation mechanism. Section 5 will detail these categories, and therefore, we here omit to discuss them further. Quality and green desires are not subject to the attention filter; rather, they are bound to attentional desires and constrain them. Therefore, when an attentional desire is promoted to active desire, the related quality and green desires are considered in the option selection (more on that in Section 5).

Intentions are (active) desires that the agent deliberates to pursue actively. Each intention may have a selected option taken from the options available for that desire, according to the adherence to the quality and green desires. There is also the possibility that the agent deliberates an intention from a desire that has no options. While this impedes its active pursuit, at the same time, this allows the agent to avoid any action that will, in the future, hinder the pursuit of the intention.

The following section will detail the four different desire categories available to the designer of a CATALINA agent and provide some examples.

5. Practical, Epistemic, Quality and Green Desires

In this section, we discuss the four categories of desires supported by the CATALINA architecture and represented in Figure 2; we also provide details about the goals' structure (attributes and properties of each category), and a few examples.

5.1. Practical Desires

Practical desires are desired states of affairs that can be achieved through actions. For example, for the AV "Be in Rome" can be achieved by getting into a car and driving to Rome. On the other hand, the desire "Drive to Rome safely" is not a practical desire but rather a quality one that can be achieved by driving defensively and at speeds below 70km/hr.

Generally speaking, an agent does not have a unique atomic desire, but rather it has a set of desires, each one decomposable into more elementary desires, as is common in many Goal-Oriented Requirements Engineering (GORE) approaches [13, 14, 15]. This set of desires (and their decomposition relationships) constitutes the agent's desire model. In CATALINA, Practical desires have the following attributes:

- A 'name': a string, unique identifier of the desire.
- A 'final state' of the world the agent aims to achieve. It is a propositional formula expressed referring to the bounded Linear Temporal First-Order Logic (bLTFOL) [16] (more on that later in this section).
- A 'trigger condition': the state of the world the agent will wait for before pursuing this desire. It is a propositional formula.
- A 'saliency' that represents the *urgency* or *priority* of the desire. Saliency has a prominent role in the mechanism for promoting standing desires to active desires. Usually, the saliency is expressed as a number in the [0,1) interval.

- A ‘reward’ that the agent receives if it satisfies the desire. This may also be an abstract value given to the desired state of the world by the agent. The CATALINA architecture prescribes no specific rule for the reward specification; it may even be zero if the designer does not want to address a ‘gain’ for the agent when it achieves that desire. The agent will pursue this desire according to its saliency, but its behaviour will not be ‘awarded’ for the fulfilment. A standing desire with a zero reward may be promoted to an active desire because that phase only depends on the saliency of the desire (and not on the reward).
- A list of quality and green desires constraining this attentional desire.

In the CATALINA architecture, the final state of practical desires is defined using a bounded Linear Temporal First-Order Logic (bLTFOL), already adopted by other approaches like KAOS [17]. It is a first-order logic with quantification over finite sets, enriched with LTL operators defined over discrete intervals. Formally, a bLTFOL desire has the following form:

$$< \phi > ::= < p > | \neg < \phi > | < \phi > \wedge < \phi > | < \phi > \vee < \phi > | \; \forall x/D[< \phi(x) >] \mid \exists x/D[< \phi(x) >] \mid \\ \mid \mathbf{F}(< \phi(t) >, < i >) \mid \mathbf{G}(< \phi(t) >, < i >) \mid \mathbf{U}(< \phi(t) >, < \phi(t) >, < i >)$$

where p is a predicate, D a finite set, $\mathbf{F}, \mathbf{G}, \mathbf{U}$ are respectively Finally, Globally and Until temporal operators defined over a discrete time interval *i*:

$$< i > ::= [< t_{start} >, < t_{end} >].$$

An example of desire specification is the following:

$$ReachRome := ([\mathbf{F}_{[4,5]}(ArrivedtoRome)], [RouteOpen \wedge EnoughAutonomy], \\ 0.5, 20, [MinimizeTravelTime], [MinimizeCO_2Production])$$

Desire $ReachRome$ prescribes the agent to arrive in Rome between t=4 and t=5; the precondition for the applicability of this desire prescribes that the chosen option (the route) is open and the autonomous vehicle has a sufficient autonomy (i.e. enough energy/fuel) for the travel (or can recharge/refill along the route). The desire’s saliency is 0.5, and the reward is 20. Finally, the desire is constrained by a quality desire ($MinimizeTravelTime$) and a green desire ($MinimizeCO_2Production$). It is worth noting that there is no clash between the last two desires because green desires cannot be violated (more on that in Subsect . 5.4), therefore their satisfaction takes priority over the other ones.

5.2. Epistemic Desires

Epistemic Desires aim to update the value of some beliefs, reflecting the agent’s desire to acquire knowledge about the state of the world. While practical desires represent the agent’s willingness to address some change in the state of the world outside the agent itself, epistemic desires represent the agent’s willingness to change the state of its knowledge and, therefore, have an internal impact on the agent when compared with practical desires.

Epistemic desires may be triggered by some agent’s perception (stimulus). In this case, the urgency to acquire more knowledge about the stimulus could come from a past experience showing that the stimulus may be associated with some risky or pleasant situation (a seismic wave, a good food smell,…).

Besides, epistemic desires are employed in the plans the agent enacts. In fact, the agent needs to monitor the world around itself and check the effect of the tasks it executes to verify their successful execution and the progress towards practical desire achievement.

For instance, if an autonomous vehicle can cross a road only if the traffic light is green, the driver agent will trigger an epistemic desire that will analyse the camera image looking for the traffic light and stop the car if the light is not green. Options (i.e. plans) and actions may include decision points where the control flow depends on some conditions that should be observed in the environment. Again, the need for this information triggers an epistemic desire. To fulfil its epistemic desires, the agent will conceive options as it does for practical desires.

It is worth noting that there is no direct link between the agent's need to access some of its beliefs and the triggering of a new epistemic desire. Such desires will be needed only when the agent does not have the required belief or estimates it as obsolete or inaccurate (obsolescence or inaccuracy management are not implemented in the current version of CATALINA since they concern the adoption of time-related criteria for beliefs, which are still a work in progress). From a design point of view, an epistemic desire has:

- A 'name', a string, unique identifier of the desire.
- A 'belief' that it aims to update
- A 'trigger condition' that is the state of the world the agent will wait for before pursuing this desire. A propositional formula expresses it.
- A 'saliency' that represents the 'urgency' or 'priority' of the desire.
- A 'reward' that the agent receives if it satisfies the epistemic desire,
- A list of quality desires constraining this epistemic desire.

The similarity between the epistemic and the practical desire structures is evident; differences lie in their scope (internal vs external changes are desired) and in the attentional process that triggers them. While practical (standing) desires are triggered by an endogenous (top-down) mechanism that continuously evaluates their saliency and promotes them to active desires when their saliency exceeds the attention threshold, epistemic (standing) desires are mainly triggered by the occurrence of salient stimuli coming from perceptions, and therefore, they are mainly related to an exogenous (bottom-up) attention mechanism [2]. Of course, we are here, leaving aside all the epistemic desires that are part of the monitoring process that check the advancement in the execution of any option, since they can be considered part of the option itself.

An example of an epistemic desire is:

$$IsTrafficLightGreen := ([GreenTrafficLight], [CarStoppedAtCrossroad \vee \\ \vee CarTravelling], 0.9, 10, [AccurateReading])$$

This desire expresses the agent's will to look at the traffic light (so that it can proceed with its travel). We note the high saliency (0.9, the maximum value is 1.0), which represents the importance of this desire, and the quality desire, which prescribes an accurate reading of the required information. In this case, failure to achieve the quality desire may be a condition for adopting safety measures, for instance, slowing down or even stopping the vehicle.

5.3. Quality Desires

Quality Desires impose constraints on other desires by specifying quality requirements that do not necessarily have a crisp satisfaction condition. Generally speaking, the same quality desire may constrain one or more practical desires. The difficulty in defining the satisfaction condition of these desires makes them different from the previous ones and makes evaluating their satisfaction challenging. Nonetheless, approaches exist in the literature that suggest how to define measurable satisfaction conditions for such desires. According to [18], we can complete the specification of quality desires with an operationalisation that is based on qualities defined by adopting an ontological foundation; the approach uses operators like universality, gradability, and agreement to assess the quality of fulfilment. This allows to define a Quality Type (like Effort) and a Quality Space for it (e.g. Person/Months). Numerical constraints on quality types may be conceived starting from the quality desire definition and referring to the Quality space for their values. For instance, let us consider the quality desire: "$QG_0 := TravellingSpeedInTownCentre : Moderate$", this may be operationalised as follows: "$QC_0 := CarSpeed \leq 30$".

Quality desires have the following attributes:

- A constraint that affects some quality property of the option the agent can use to achieve its practical desire. A predicate may express this in the form:

```
Quality_desire := quality_subject + verb + constraint.
```

- A 'trigger condition' specifying the applicability of the quality desire (it may include a spatial specification and/or a temporal constraint).
- A reward that the agent receives if it satisfies the quality desire.

5.4. Green Desires

In CATALINA, a green desire specifies a constraint implementing some environmental policy that the agent must respect. Therefore, a green desire is usually empowered by some normative enforcement. It is not a quality desire since we can usually define a crisp satisfaction criterion for the green desire, while we can not for quality desires.

Often, quality desires constrain the way the agent acts to achieve its objectives. For instance, a quality desire may limit the cost of buying/building something or the effort to complete a task. This affects the selection of the plan the agent adopts for fulfilling the specific desire, in this case, the quality desire is 'attached' to the practical desire it constrains. Conversely, green desires may constrain specific practical desires, but they may potentially affect the entire agent's behaviour.

Let us consider the desire that imposes some temporal limits on a smart house control agent to use the heating system. In some places, this system may be switched on for only 8 hours a day. This affects the agent's several (practical) desires about house management. In fact, a desire like *Ensure a comfortable living temperature* has to be pursued together with *Ensure a regular fresh air income*. These are clashing desires whose coexistence may be solved by frequently executing cycles like switching off the heating system, opening windows, waiting 10', closing windows, switching on the heating system, and maintaining optimal temperature. Unfortunately, this is not an efficient energy-saving approach, and it will increase the total number of hours the heating system is used.

As already said, green desires usually have some normative basis. This may imply that green desires sometimes have an application range. This may be a place: "Cars that do not have Euro 6 certification can not travel in the centre of the town". In other cases, the application range is a temporal interval. For instance, as already proposed above, a house heating system may not be used before a specific date or after another date.

From the normative empowerment of green desires comes the possibility that violating one of them may imply a fee/penalty for the agent responsible for the violation. Another significant feature of green desires, again coming from their normative origin, is that they cannot be relaxed. Any violation of the constraint will lead to the consequences provided by law.

A green desire in CATALINA includes the following attributes:

- A trigger condition that specifies if the desire constraint is in effect. This may be a spatial or temporal constraint.
- A constraint on some parameter that may affect the agent's actions. This usually means that green desires may constrain the agent's options to achieve its attentional desires.
- A reward the agent receives if it satisfies the green desire.
- A fee that is due when the desire is violated.

The Executive Reasoner Function considers green desires for selecting the best desire option that will be attached to an intention. Therefore, green desires play a relevant role in the agent's decision process and condition that.

In the next section, we will discuss the attention modulation mechanism and its relationship with standing desires, active desires, and intentions, as well as describe an example of CATALINA implementation.

6. CATALINA in Action: The AV Scenario

In CATALINA, the shift in attention modulation occurs immediately after an intention change (deletion or addition) in the GW, as the GW sends a signal to the Agent Focus module (Executive Inhibition Function).We made a simulation[1] to test the shifting of attention modulation and the mechanism of practical (top-down) and epistemic (bottom-up) desires in the event of salient stimuli. Our objective was to demonstrate how our architecture shifts attention among desires (practical or epistemic in the same way).

We implemented two scenarios: the first one occurs when the agent pursues a practical desire, and the second one occurs when the agent has a salient stimulus from the environment and shifts its attention to what caused it (thus triggering an epistemic desire). In this simulation, an agent, driving an autonomous vehicle (AV), has to go from one city to another as specified by the practical desire (promoted to selected intention). Let us suppose it starts from Lisboa and to reach its destination, the AV must pass through various cities along a route made up of steps (steps are a rough approximation for a unitary length of path and allow us to perform a discrete-time, step-by-step simulation easily). Routes are identified by a numeric ID and have various characteristics, such as: the number of steps, the maximum speed (steps per hour), a numerical value for the panorama of the route, and the amount of pollution produced to travel the entire route.

In the current version of our simulation, the agent initially has three standing practical desires (Visit_Paris, Visit_Frankfurt and Visit_Rome), two quality desires (Panorama_is_Great and Drive_Safely) and one green desire (Limit_Pollution_to_50). The AV may satisfy the three practical desires by arriving respectively in Paris, Frankfurt, and Rome through a path that satisfies the green and quality desires. One standing practical desire (Visit_Rome) also has a precondition that consists of visiting Paris before becoming an active practical desire

In the following, we will detail the two scenarios, starting with the one related to the attention modulation mechanism during the execution of an (endogenous, top-down) practical desire that, in this case, means going towards a city.

6.1. Attention Modulation in Endogenous Desire Pursuit

After the agent has been initialised, it executes the Perception Processing module (Working Memory Maintenance Function) to process its perceptions. At this stage, the agent has no perceptions, so it executes two modules of the Executive Inhibition Function: Desire_Deletion and Intention_Deletion. The Desire_Deletion module is responsible for checking whether there are desires with a lower saliency than the agent's saliency/attention thresholds. If so, the agent eliminates these desires and any intentions connected to them from the GW by executing the Intention_Deletion module.

Hence, the agent executes its Switching_To_Stimulus module (Executive Switching Function) to check if it has perceived any relevant stimulus that can be transformed into a standing epistemic desire and subsequently into an active (exogenous, epistemic) desire to pursue. At the beginning, in our simulation, the agent does not have any stimulus, and therefore the execution moves to the Desires_Promotion module (Executive Switching Function) to fetch from memory (The Global Workspace) the standing practical desires (either inhibited or not inhibited), and to check if among them there is any standing practical desire whose saliency is high enough to consider that for being promoted to active practical desire. If the check reports some standing desire with a sufficiently high saliency, then these are promoted to active practical desires. In the simulation, only the standing practical desires Visit_Paris and Visit_Frankfurt will initially be promoted to active practical desire, while Visit_Rome, having a false precondition (Paris has not been visited yet), cannot be promoted to active desire.

When the agent is run for the first time, the AV has not focused its attention yet. As a consequence, all the stored information (beliefs, regions, practical desires, etc.) is uninhibited and available within the Global Workspace that, at this stage, contains the information reported in Table 1. This table can be

[1]The current version of the architecture, including the autonomous vehicle simulation, is available for download from the CATALINA GitHub repository: https://github.com/CATALINA-Architecture/CATALINA_Model

Table 1
Scenario 1: a snapshot of information (regions, beliefs, standing/active desires, and intentions) in the Global Workspace after Endogenous module execution.

Regions	Beliefs	Attentional (Standing) Desires	Active Desires and Intentions
All 47 Cities: Lisboa, Madrid, Rome, Venice, Palermo,London, Paris, etc.. All 132 Routes: 1, 2, 3 , etc..	All 204 Beliefs: Belief_Current_Time, 2 Belief_Destination_City, all Belief_Visited_City (1 for city), Belief_Current_City, Belief_Current_Route, Belief_Current_Step, Belief_Route_Status, etc.. all Belief_Route_Status (1 for city), etc..	Visit_Paris Visit_Frankfurt	Desire D1 (Attent. Desire: Visit_Paris): Desire D2 (Attent. Desire: Visit_Frankfurt)

considered as a snapshot of the information present in the Global Workspace after the execution of the Desires_Promotion module.

The Reasoner function receives a message from the GW informing it of a change in the active desires. As a result, it fetches the new active desires from the GW and analyses them. For each active desire, it elaborates a set of options that allow the pursuit of the attentional desire. Now, the Reasoner considers the temporal constraint defined by the temporal logic operator and removes the options that do not allow it to be satisfied. Next, the Reasoner considers the constraints defined by the Green Desires linked to the current desire and removes all the options that violate them. The result is a set of options that allow achieving the desired outcome while respecting the related Green Desires. Finally, these options are ordered according to their performance in terms of quality desires. More specifically, let us suppose the desire is constrained by a quality desire prescribing a high travel speed. In this case, the options are ordered according to the average travel speed in the different parts of the route, and, of course, the fastest one is preferred.

Subsequently, for each active desire that passes the attention filter, a corresponding intention is generated. Each intention represents a commitment the AV can pursue by selecting one single option. Note that an intention may not have any options. In fact, in our simulation, the intention related to the practical desire "Visit_Paris" has one option, while the intention related to the practical desire "Visit_Frankfurt" has none because its options don't pass the filters (it's a deliberate coincidence). The new intentions are posted to the Global Workspace, which sends a signal of change in intentions to the Agent Focus module (Executive Inhibition Function class) and to the Agent Unfocus module (Executive Switching Function class).

As previously mentioned, Attention modulation occurs when there is a change in the agent's intentions. If the agent has at least one intention, the Agent Focus module recalculates the saliency and attention thresholds according to the intention with the highest saliency and creates the inhibition regions, the inhibited beliefs, and the inhibited attentional desires. The Agent Focus moves them (via the Global Workspace) to the Long Memory and keeps in the Global Workspace only the regions, beliefs and attentional active desires that are related to the intentions that are currently pursued; this way, only the essential information is kept in the Global Workspace, all remaining information is removed from the Global Workspace and saved in the Long Term Memory.

Conversely, when the agent has updated the list of its intentions and this list is empty because it has deleted all the intentions (because it has completed them or they are no longer consistent with its beliefs), then the Agent Unfocus module restores saliency/attention threshold to the default value for the agent and removes the inhibition regions, inhibited beliefs and standing desires, and moves them to the Global Workspace. In our simulation, the agent focuses its attention. The information reported in Table 2, is a snapshot of the information that is maintained in the Global Workspace. In this table, unlike what is reported in Table 1, only the beliefs, regions, practical desires, and quality/green desires

Table 2
Scenario 1: a snapshot of information (regions, beliefs, standing/active desires and intentions) in Global Workspace after Focus Attention module execution.

Regions	Beliefs	Attentional Standing Desires	Active Desires, Intentions
5 Cities: Lisboa, Madrid, Barcelona, Marseille, Paris 10 Routes: 96, 32, 115, 124 97, 33, 114, 125	30 Beliefs: Belief_Current_Time, 2 Belief_Destination_City, 5 Belief_Visited_City (1 for city), Belief_Current_City, Belief_Current_Route, Belief_Current_Step, 5 Belief_Route_Status (1 for city), Others..	Vist_Paris Visit_Frankfurt	Intention I1 (linked to Attent. Desire: D1) Intention I2 (linked to Attent. Desire: D2)

that are related to I1 and I2 are present. All remaining information, which was previously in the Global workspace and is not related to I1 and I2, is moved to the Long Memory. Thus, in the Global Workspace, there is only the essential information to pursue the current intentions.
Then, the AV begins the execution of its planned trajectory. In our simulation, the agent starts moving towards Paris from Lisbon, and it runs along the route leading to Madrid. The execution of the selected option consists of the agent executing the actions using its Plan_Advancement_Evaluation and Plan_Execution modules (Executive Resource Allocation Function) until the Perception Processing module perceives that it has reached the destination and, therefore, removes the current intention from the GW.

6.2. Attention Modulation in Exogenous Desire Pursuit

The second scenario of our simulation occurs when there is a salient external stimulus, and the agent raises an epistemic desire to identify the underlying cause that triggered this stimulus. The Stimulus_Inhibition module of the agent applies a filter to the stimulus to verify whether or not the stimulus falls within the inhibition region. Subsequently, it verifies whether the filtered stimulus has a saliency value higher than the current saliency/attention threshold. If this latter verification is satisfied, the stimulus becomes a relevant stimulus, and the Stimulus_Inhibition updates the stimulus in the Global Workspace that broadcasts a stimulus update message. Subsequently, the Switching_to_stimulus module detects the stimulus update message and takes the stimulus from GW. Hence, an epistemic standing desire is raised with a saliency value equal to the stimulus, and the standing desire is then promoted to a new active desire. The Reasoner analyses it and generates one or more options. Thus, the Reasoner updates the intentions in the GW, triggering a broadcast message from the GW that also reaches the Executive Inhibition function. The Agent Focus module of this function recalculates the saliency and attention thresholds according to the new intention, and recalculates the inhibition regions and the inhibited beliefs and desires.

In our simulation, the user may send a message to the AV during its travel, communicating that there is some danger along the road and that it is, therefore, closed to transit. This message is a stimulus that triggers (because of its relevant saliency) an exogenous epistemic standing desire (*how long will the road be closed?*). This becomes an active desire to understand the expected duration of the road interruption. If the delay is too long, the agent needs to find an alternative route to reach its destination in time. This desire becomes an intention to be pursued with the highest priority. Thus, the AV's focus of attention shifts from moving along the route (the initial practical desire) to understanding the duration of the traffic interruption on the road (the new epistemic desire). The agent may fulfil that by contacting a Traffic Information Service, and after receiving the expected delay, the agent decides if it is better to wait or to change its route.

7. Discussion and Conclusions

In this paper, we have illustrated the metamodel of the CATALINA architecture, which is aimed at exhibiting consciousness features and practical reasoning. This objective has been reached by integrating the Global Workspace Theory into a BDI (Belief-Desire-Intention) agent architecture to enable consciousness-inspired reasoning driven by attention. Combining Buheler's executive function theory, Bratman's practical reasoning paradigm, and Baars' model of consciousness, we have proposed an agent architecture that can engage in both reactive and deliberative actions under the control of an attentional mechanism. Consequently, the agent can dynamically prioritise important desires thanks to the selective processing capability enabled by this design.

The CATALINA's current implementation demonstrates the feasibility of attentional modulation across practical and epistemic desires in response to dynamic environmental stimuli. In our simulation, an autonomous car travelling between cities shows how the agent can change its focus and attention between competing desires, allowing it to adapt its behaviour to changing internal demands and external conditions.

The Global Workspace acts as the central integration point, triggering executive modulation when intentions are added, removed, or reprioritised, simulating conscious control over task focus.

The model offers a promising starting point for connecting concepts from artificial consciousness with rational agent design. It offers a concrete architecture that highlights how attention, modular executive functions, and desire management work together to guide the agent's behaviour.

The CATALINA architecture is still in its early stages of development. It does not yet have mechanisms for dealing with belief obsolescence or determining implicit beliefs. Currently, it only manages explicit beliefs. Future versions will try to include logical closure techniques for belief systems and temporal decay models for beliefs to improve the agent's epistemic self-regulation. Despite the interesting results, the work towards a complete realisation of a conscious agent architecture extending the BDI reasoning paradigm is still far from concluded; the issues related to this objective are still open research challenges in both cognitive sciences and computer science. We are already planning further versions of this architecture in an iterative-incremental approach. Next steps will consider improvements like the introduction of a devoted *Attention Modulation* component of the architecture that would collect all the attention-related features and further extend them.

Acknowledgments

G. Averna, M. Cossentino, G. Pilato acknowledge the support of the PNRR project FAIR - Future AI Research (PE00000013), Spoke 9 - Green-aware AI, under the NRRP MUR program funded by the NextGenerationEU.

Declaration on Generative AI

During the preparation of this manuscript, the author(s) used Grammarly to check grammar and spelling. After using this tool, the authors reviewed and edited the content as needed and took full responsibility for the publication's content.

References

[1] M. Bratman, Intention and means-end reasoning, The Philosophical Review 90 (1981) 252–265. doi:10.2307/2184441.

[2] C. Massimo, G. Pilato, A. Guido, M. Myrto, J. Mylopoulos, Practical reasoning with attention mechanisms, in: I. Press (Ed.), In proc. of the IEEE SIMPAR Int. Conf. on Simulation, Modeling, and Programming for Autonomous Robots, 2025.

[3] C. Massimo, A. Guido, P. Giovanni, M. Myrto, M. John, The agent working cycle in catalina, in: C. Proceedings (Ed.), In proc. of the WOA 2025 26th Workshop From Objects to Agents, 2025.
[4] M. E. Bratman, D. J. Israel, M. E. Pollack, Plans and resource-bounded practical reasoning, Computational intelligence 4 (1988) 349–355. doi:10.1111/j.1467-8640.1988.tb00284.x.
[5] B. J. Baars, Global workspace theory of consciousness: toward a cognitive neuroscience of human experience, Progress in brain research 150 (2005) 45–53.
[6] D. Buehler, Agential capacities: a capacity to guide, Philosophical Studies 179 (2022) 21–47.
[7] B. J. Baars, The global workspace theory of consciousness: Predictions and results, The Blackwell companion to consciousness (2017) 227–242.
[8] W. Huang, A. Chella, A. Cangelosi, A cognitive robotics implementation of global workspace theory for episodic memory interaction with consciousness, IEEE Transactions on Cognitive and Developmental Systems 16 (2023) 266–283.
[9] D. Buehler, Psychological Agency-Guidance of Visual Attention, Ph.D. thesis, UCLA, 2014.
[10] D. Buehler, The central executive system, Synthese 195 (2018) 1969–1991.
[11] M. Bratman, Intention, Plans, and Practical Reason, Harvard University Press, 1987.
[12] A. Diamond, Executive functions, Annual review of psychology 64 (2013) 135–168.
[13] E. S. Yu, Towards modelling and reasoning support for early-phase requirements engineering, in: Proceedings of ISRE'97: 3rd IEEE International Symposium on Requirements Engineering, IEEE, 1997, pp. 226–235.
[14] A. van Lamsweerde, E. Letier, P. M. Heymans, Kaos: A methodology for object-oriented requirements engineering, in: Proceedings of the 3rd International Conference on Object-Oriented Information Systems (OOIS '97), IEEE, 1997, pp. 9–22. doi:10.1109/OOIS.1997.10057.
[15] P. Bresciani, A. Perini, P. Giorgini, F. Giunchiglia, J. Mylopoulos, Tropos: An agent-oriented software development methodology, Autonomous Agents and Multi-Agent Systems 8 (2004) 203–236.
[16] N. Kamide, Bounded linear-time temporal logic: A proof-theoretic investigation, Annals of Pure and Applied Logic 163 (2012) 439–466.
[17] A. Dardenne, A. Van Lamsweerde, S. Fickas, Goal-directed requirements acquisition, Science of computer programming 20 (1993) 3–50.
[18] F.-L. Li, J. Horkoff, J. Mylopoulos, R. S. Guizzardi, G. Guizzardi, A. Borgida, L. Liu, Non-functional requirements as qualities, with a spice of ontology, in: IEEE 22nd International Requirements Engineering Conference (RE), 2014, pp. 293–302.

A. Online Resources

The current version of the architecture is available for download from the CATALINA GitHub repository: https://github.com/CATALINA-Architecture/CATALINA_Model