

# PROMETHEUS AI

Phase 2

Sean Stappas  
260639512

ECSE-499: Honours Thesis II

Supervised by: Prof. Joseph Vybihal

December 7<sup>th</sup>, 2017

## Abstract

Prometheus AI is a model of the human brain with the goal of controlling and coordinating multiple robots in different environments. This could be useful in environments hazardous for humans, such as in the aftermath of a nuclear power disaster, or in outer space. The model consists of four layers: the Neural Network (NN), the Knowledge Node Network (KNN), the Expert System (ES), and the Meta Reasoner (META). The NN classifies the signals coming from the robots' sensors and sends formatted tags to the KNN. The KNN represents memory and can initiate cascaded activation of memories in the form of tags, which are passed on to the ES. The ES is a simple logic reasoner and provides recommendations for actions to the META. The META represents high-level thinking and makes an intelligent decision for the robots. The tasks for this thesis were to implement the KNN and ES layers in Java and to supervise work done by undergraduate students on Prometheus. This was achieved using specific design criteria and extensive feedback from the project supervisor, Prof. Vybihal. Personal design criteria included using object-oriented programming principles, optimizing the system for speed and space efficiency, and maximizing code readability and testability. Continuous integration tests were created with TestNG and TravisCI and extensive documentation was written in Javadoc.

## Acknowledgments

Under my supervision, several undergraduate students worked on the Prometheus layers in Prof. Vybihal's lab. Over the summer, Isaac Sultan and Si Yi Li made contributions to the ES and KNN layers, respectively. This semester, Mohammad Owais Kerney and Michael Ding showed interest in the NN and META layers, respectively.

# Contents

<b>List of Figures</b>	<b>2</b>
<b>List of Tables</b>	<b>3</b>
<b>1 Introduction</b>	<b>4</b>
<b>2 Background</b>	<b>4</b>
2.1 Neural Network . . . . .	4
2.2 Knowledge Node Network . . . . .	4
2.3 Expert System . . . . .	7
2.4 Meta Reasoner . . . . .	8
2.5 Summary . . . . .	8
<b>3 Problem</b>	<b>8</b>
<b>4 Design</b>	<b>8</b>
4.1 Efficiency . . . . .	8
4.2 Object Oriented Programming . . . . .	8
4.3 Readability & Documentation . . . . .	8
4.4 Testing . . . . .	8
4.5 Quality Assurance . . . . .	9
<b>5 Implementation</b>	<b>9</b>
5.1 Tags . . . . .	9
5.2 Knowledge Node Network . . . . .	10
5.3 Expert System . . . . .	11
5.4 Testing . . . . .	11
5.5 Documentation . . . . .	13
<b>6 Results &amp; Tests</b>	<b>13</b>
<b>7 Impact on Society and the Environment</b>	<b>13</b>
7.1 Use of Non-renewable Resources . . . . .	13
7.2 Environmental Benefits . . . . .	13
7.3 Safety and Risk . . . . .	13
7.4 Benefits to Society . . . . .	13
<b>8 Conclusion</b>	<b>14</b>
<b>A UML Diagrams</b>	<b>15</b>
<b>References</b>	<b>16</b>

## List of Figures

1	Prometheus AI model. . . . .	4
2	High-level model of the Knowledge Node of the KNN. . . . .	5
3	Thinking forwards in the KNN. . . . .	6

4	Thinking backwards in the KNN. . . . .	6
5	Lambda thinking in the KNN. . . . .	7
6	Prometheus AI model with labeled input and output. . . . .	9
7	Setup for the <code>testKNN()</code> method. . . . .	12

## List of Tables

1	Abbreviations in Prometheus. . . . .	4
2	Test setup for <code>testES()</code> . . . . .	12
3	Test setup and activation for the ES portion of <code>testKNNandES()</code> . . . . .	12

## Listings

# Abbreviations

The following abbreviations will be used throughout the report:

Table 1: Abbreviations in Prometheus.

NN	Neural Network
KNN	Knowledge Node Network
KN	Knowledge Node
ES	Expert System
META	Meta Reasoner
OOP	Object-Oriented Programming

# 1 Introduction

The goal of the Prometheus project is to create an artificial intelligence system to control and coordinate multiple robots. All the thinking and decision-making is done by Prometheus, which collects data from all the robots' sensors.

Applications for this type of system include robots in hazardous environments, such as in outer space (Mars, Moon, etc.), in nuclear plants after a disaster, and in military zones. The system could ideally learn from this data in a given environment and apply its learning to new environments.

The system is loosely inspired from the structure of the human brain, and is composed of the following four layers [1]: the Neural Network (NN), the Knowledge Node Network (KNN), the Expert System (ES), and the Meta Reasoner (META), which can be seen in Figure 1.

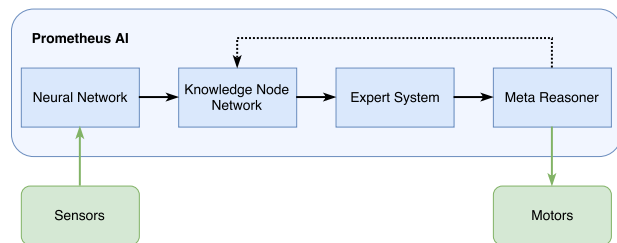


Figure 1: Prometheus AI model. Blue represents Prometheus and green represents the robots.

The theory needed to understand these four layers will be given in Section 2. The assigned task will then be described in Section 3. Design criteria will be discussed in Section 4. The description of the work done this semester can be found in Section 5, and the results and tests will be discussed in Section 6. Finally, the possible future impact of this project on society and the environment will be explored in Section 7.

# 2 Background

## 2.1 Neural Network

The NN layer consists of a network of neurons with a structure similar to neurons in the human brain. It is the interface between the robots' sensors and the rest of the AI system.

The robots in Prof. Vybihal's lab are equipped with two types of sensors: camera and ultrasonic. The ultrasonic sensor can measure distance between the robot and nearby objects, and the camera can take images of what the robot is facing.

The NN gathers raw sensor data and will build an abstract view of the robot's surrounding environment. For each camera image, it will achieve two main goals:

1. Classify objects observed in the image.
2. Localize objects in the image.

Ultimately, the classification and localization of objects will produce abstract informational tags, which will be passed on to the KNN. These tags can therefore be characteristics of objects in the world, such as distance. The tags can also have associated confidence values produced by the NN, which can be converted to belief within the KNN.

## 2.2 Knowledge Node Network

The KNN layer represents memory in the human brain. It takes in the tags provided by the NN and outputs tags based on its knowledge of

Mention scope of thesis (where my part fits in to bigger whole)

Mention that this section draws heavily from previous report.

the environment. The KNN is based around interconnected Knowledge Nodes (KNs), which are abstract structures representing memories and their connections to other memories. A simple model of a Knowledge Node (KN) can be seen in Figure 2.

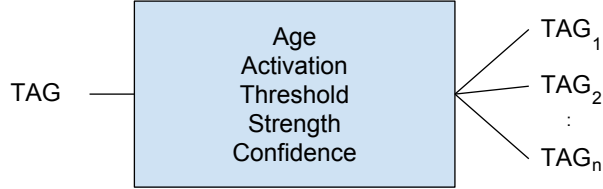


Figure 2: High-level model of the Knowledge Node (KN) of the KNN [2].

KNs have an input tag representing some information, and output tags representing information related to the input tag. When a KN is stimulated it becomes “excited”. Tags in the KNN can become “active”, meaning the associated KN is stimulated beyond a threshold. We say that the associated KN “fires” in this case. For example, if the NN determines that it observes a ball and passes that information on to the KNN, then a “ball” tag may become active in the KNN. This would excite the KN with input tag “ball”. If that KN fires, other tags related to that observation can become active. For example, tags representing ball characteristics such as “round object” may become active. This tag in turn may be connected to another KN, potentially causing more activation.

There are different ways that KN excitation could be implemented. With simple linear activation, excitation would increment the node’s activation parameter, which initially starts at 0. If  $activation \geq threshold$ , the KN fires, causing the activation of the output tags. This description corresponds to forwards thinking. The activation can also be achieved using a sigmoid function, which is more representative of neurons in the brain [3].

There is a strength value associated with every KN which represents how much weight an activation has and therefore how quickly that node will fire. A simple implementation of strength would be as a constant coefficient multiplying

the activation parameter. So, instead of checking when the activation is greater than the threshold, one would check if  $activation * strength \geq threshold$ , where the strength is positive. The strength value could be 0 however, which effectively shuts off the KN. Another way of implementing strength is to check  $activation + strength \geq threshold$ . In this case, strength can be negative and make the KN fire later than normal.

Strength can be seen as the firing predisposition of neurons in the brain as a result of learning [2]. Indeed, learning can increase the synaptic strength between neurons and cause early firing of those neurons [4]. For example, a person who has had a bad experience with spiders would fire their fear response upon seeing a spider more quickly than one without that fear.

Every KN can have a belief value associated with it, representing how certain the KNN is that the input tag is true. In this case, the belief value would be stored within the KN itself. These belief values can come from the NN, since a neural network can associate confidence values with its classifications of objects [5]. When initiating a think cycle, the belief values at each stage can be multiplied with each other to produce a new confidence value, making the KNN less certain of a memory as it searches through its tree of KNs. This represents how belief changes when thinking. Indeed, some memories in the brain require a great deal of thinking to reach and, as such, can be less certain than other memories. This can lead to the recollection of false memories [6].

Note that the activation of a KN and the belief value associated with it are two different concepts. Indeed, a KN may be fired even if the KNN knows that it is false. This can represent an agent “lying” to itself or repeating a thought process it knows to be false. This is known as “self-deception” in humans and it has been argued that this can be used for positive mental health [7].

Age represents how long it has been since a KN has been excited. The idea is that, after a KN has aged a certain amount of time, that node will be discarded, similarly to how old memories are discarded in the brain [8]. The intuitive im-

plementation of aging is to constantly increment some age value associated with every KN, and discard nodes whose age is greater than some threshold. The age would be reset to 0 when the node is excited. This would however be unnecessarily computationally intensive, requiring some kind of constant updating of every single node. A more efficient way of doing it would be to save a timestamp for every KN when it is excited, and if the KNN attempts to excite a KN whose previous timestamp value is too far into the past, that node is instead deleted.

Thinking in the KNN represents an activation routine, which activates tags and potentially fires KNs. The KNN has four main ways of thinking: direct, forwards, backwards, and lambda. The version of thinking to be done by the KNN is chosen by the META.

Direct thinking is a simple lookup of the desired Tag and excitation of the associated KN. This represents the simplest form of thinking in the KNN.

Forwards thinking is depicted in Figure 3. Firing a KN can cause forward activation of more KNs, hence the “forwards” naming.

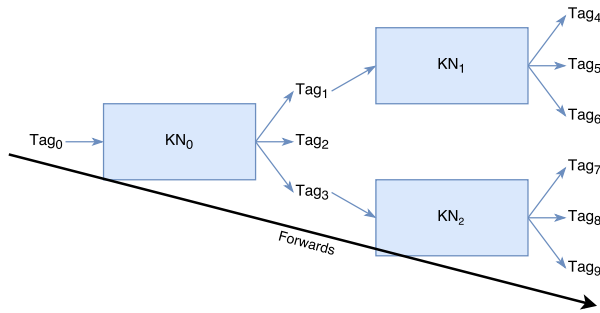


Figure 3: Thinking forwards in the KNN.

Thinking backwards starts at the output tags of KNs, and works backwards, as can be seen in Figure 4. In its simplest form, it checks output tags of KNs and, if all of them are active, the input tag must be active as well, so that KN is fired.

This type of thinking can be more difficult to implement than forwards thinking, since one must decide what percentage of active output tags of a KN corresponds to an active input tag. For instance, should a KN fire when 100% of its

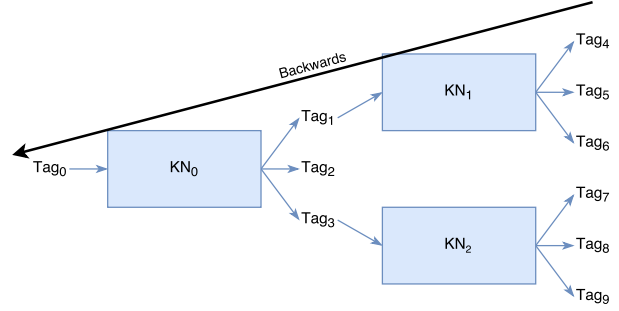


Figure 4: Thinking backwards in the KNN.

output nodes are active, or when 75% are active? This relates to the belief with which the KNN believes the tag associated with that node to be true. As a concrete example, if you observe an object that has four wheels, seats, and a steering wheel, how confident are you that that object is a car? Realistically, humans will often classify what they observe with some uncertainty [9] and this can be represented with backwards thinking. This type of thinking occurs constantly in the background in humans [2], and this is something to keep in mind during implementation.

Lambda thinking uses a combination of forwards and backwards, and can be seen in Figure 5. It is called “lambda” thinking because the shape of the thinking trajectory matches a Greek uppercase lambda ( $\Lambda$ ). It first looks at output tags of KNs and propagates activation backwards, similar to backwards thinking. After a certain amount of nodes are fired, it will then start activating forwards, similar to thinking forwards. An interesting question is how far backwards should lambda thinking go before starting to cascade forwards? This relates to how general one wants to explore before searching for a more specific value.

This type of thinking occurs in humans when using analogical reasoning to find a memory [10]. In essence, when a person wants to locate a memory that is not directly accessible, they will explore related memories, move their way “backwards” to concepts related to the desired memory, and think “forwards” to focus in on the desired memory. For example, if one is asked where they were in 2002 on a specific date, they most

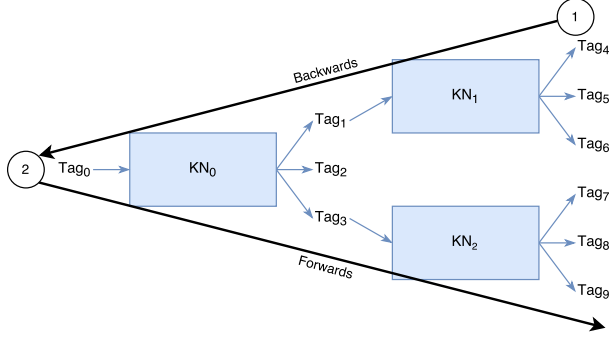


Figure 5: *Lambda thinking in the KNN.*

likely would not remember. If they explore related dates and events in their life, however, they may be able to narrow down and extract that memory. In the context of the Prometheus system, lambda thinking will be attempted if all other forms of thinking fail (forwards and backwards).

All forms of thinking can continue until there are no more KNs to activate, which corresponds to natural quiescence. There can also be a fixed number of thinking cycles, which represents how much effort is being put into thinking. Indeed, in humans, thinking is done with varying degrees of effort [11], and it would be interesting to model this.

The result of all forms of thinking is a collection of activated tags, which are passed on to the ES.

## 2.3 Expert System

The ES layer is a basic logic reasoner. It is not aware of its current reality or any context. It takes in the tags provided by the KNN and interprets them as either facts, recommendations or rules.

Facts are simple calculus predicates showing that something is true. Here are some examples of fact tags and their meanings:

- (A) A itself is true or active.
- ( $A = 1$ ) A is equal to 1.
- ( $A > 1$ ) A is greater than 1.
- (A ?) A can take any value.

As a more concrete example, a fact can represent a certain measurement, like (*distance = 5*) representing the robot's distance from a wall as measured by one of its sensors.

Recommendations represent suggestions for actions to be taken by a robot. For example, (*@turn\_left*) is a recommendation for a robot to turn left, if it sees a wall directly in front of it and must avoid it, for example. These are recommendations and not commands because the META can decide whether or not to actually take that action.

Rules are many-to-many structures with facts as inputs and tags as outputs. This can be seen in Equation (1), where  $m \geq 1$  and  $n \geq 1$ , i.e., there must be at least one input fact and one output tag. Each output tag can either be a fact or a recommendation. When all the input facts become active, the output tags become active and the rule itself is said to be active. In this way, a rule can represent a logical AND of all its input facts.

$$Fact_1 \cdots Fact_m \rightarrow Tag_1 \cdots Tag_n \quad (1)$$

The runtime of the ES consists of the following general steps [12]:

1. Reset.
2. Add facts and rules.
3. Think.
4. Send recommendations to META.

The most important part of the previous process is the thinking stage, which represents the activation routine of the ES. This consists of first iterating through all the rules in the ES and checking if they are active by inspecting the lists of facts and recommendations. Rules may then become active and cause cascading activation of more rules. This can continue until there are no more rules to activate, which corresponds to natural quiescence. There can also be a fixed number of thinking cycles, which represents the effort put into thinking, similarly to the KNN. The recommendations activated as a result of thinking are passed on to the final layer, the META.



## 2.4 Meta Reasoner

The META layer represents high-level reasoning in the human brain. It is aware of its environment and context, and makes decisions based on what it believes to be right. It is paranoid, and constantly checks whether the tags reported by the rest of the AI system make sense based on its expected view of the world. If it decides to make a decision, it sends a command to the actuators of the robots to decide how to move. If it is not happy with the recommendations from the ES, it may send a query back to the KNN to initiate another think cycle and generate new recommendations.

## 2.5 Summary

With this full description of the Prometheus AI model, the system with labeled input and output can be seen in Figure 6.

# 3 Problem

The main task for this thesis is to construct the KNN and ES in Java. The requirements are that these layers should contain basic versions of the functionality described in Section 2. Another important task is to supervise the work of other undergraduate students working on Prometheus and to put systems in place to ensure a certain standard for the code. The design criteria for both these tasks will be discussed in Section 4, and their implementation will be discussed in Section 5.

# 4 Design

## 4.1 Efficiency

A very important consideration when designing the system is speed. Since the robots may have to react very quickly to stimuli in the environment, the reasoning in the AI must be as fast as possible. This is especially true in the hazardous environments for which this system could be useful for, as specified in Section 1.

## 4.2 Object Oriented Programming

Another important design choice is to leverage object-oriented programming (OOP) as much as possible. OOP allows extensive planning before even beginning to write code, which can identify any flaws in the initial design. It also allows the code to be very clean and reusable [13]. Since Java is the programming language chosen for the project, OOP is also the natural way to proceed. OOP principles such as polymorphism and encapsulation will be followed closely. One important OOP principle to use is abstraction. Indeed, the system should be as abstract as possible, while still performing its desired task. For instance, the system should be general enough to perform under simulations, as well as in real-life environments. It should also ideally be able to perform in vastly different environments, with varying tasks. Encapsulation can also be very useful, since each layer of the system has unique, localized functionality that does not need to be visible from the rest of the system.

## 4.3 Readability & Documentation

One final design criterion is that the code written should be very easy to understand. This means implementing each method and class in the most intuitive way possible and providing good documentation to support the code. This is to ensure that anyone wanting to work with the code or looking to understand how the system works has an easy time doing so.

## 4.4 Testing

An important criterion is that the code be heavily tested. This includes integration tests, covering end-to-end functionality, and unit tests. The unit tests should be behavior-driven, so that a developer reading the tests can easily understand the purpose of the element being tested. The aim is that at least 95 percent of the code be covered by tests.

Talk about more complex principles (Liskov, dependency inversion, injection, SOLID)

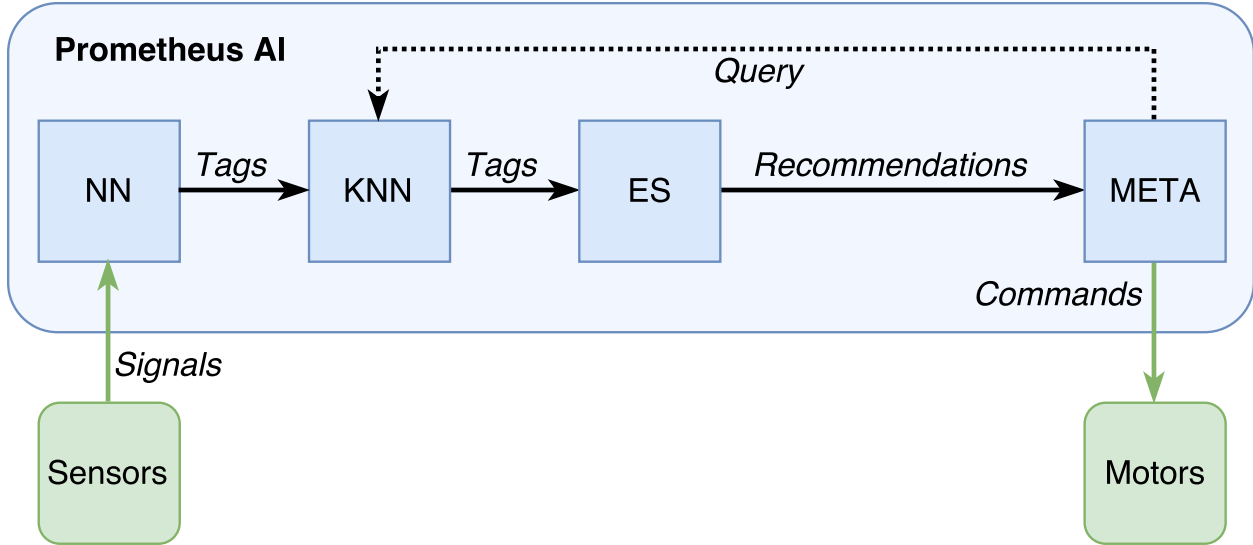


Figure 6: Prometheus AI model with labeled input and output.

## 4.5 Quality Assurance

With the numerous undergraduate students showing interest in working on Prometheus, there must be a system in place to ensure the quality of the code is upheld. This may include automatic testing of the code, as well as human review of code before being committed.

## 5 Implementation

### 5.1 Tags

As described in Section 2, the entire system revolves around tags passed from layer to layer. For this reason, a lot of thought was put into the proper design of these tags.

The tags need to be as general as possible. A natural choice for this structure would be a Java `String`, which would be relatively simple to pass around the system. However, these tags represent various concepts; each tag can either be a fact, a recommendation or a rule. If implemented as `Strings`, the tags would have to be encoded on creation to represent each concept and decoded on use to retrieve the important information. This seems like a bad use of the OOP principles of Java. Furthermore, if specific functionality is needed in the future for each tag type, that can easily be implemented with a Java class.

For these reasons, the tags are implemented using a `Tag` Java class, with `Recommendation`, `Fact`, and `Rule` subclasses. This is a good example of using the inheritance principle of OOP. This should also make manipulating the Tags faster, while incurring a slight memory overhead. To store these Tags in a database, they can be converted to JSON format. On read from the database, they can be easily decoded.

The `Tag` class has an associated `Type`, which can take the following `enum` values: `FACT`, `RECOMMENDATION`, and `RULE`. This is used to distinguish the three types of Tags.

The `Fact` and `Recommendation` classes at this point are little more than wrapper classes around a `String` value. The `Rule` class has the following important fields:

`inputFacts` List of input Facts.

`outputTags` List of output Tags.

The `Tag` class itself is abstract. This means that an object may not be directly instantiated as a `Tag`, but must be instantiated as one of its subclasses. This makes sense, since a tag *must* be one of the three types: rule, recommendation or fact. This is an example of using the abstraction principle of OOP.

All of these classes were placed inside the `tags` package. A UML diagram of the `tags` package can be seen in ?? of the Appendix.

## 5.2 Knowledge Node Network

All code relating directly to the KNN was placed in the `knn` package of the project. A UML diagram of the `knn` package can be seen in ?? of the Appendix.

The KNN layer is based around the `KnowledgeNodeNetwork` Java class. This class has the following fields:

`mapKN` One-to-one `HashMap` of input `Tags` to associated `KnowledgeNodes`.

`activeTags` `HashSet` of active `Tags`, corresponding to input `Tags` of fired `KnowledgeNodes`.

The choice of `HashSet` for the `activeTags` field is not arbitrary. Indeed, `HashSets` were used for most of the collections with variable size in the ES and KNN. The original specifications mentioned using `ArrayLists`, but, since there is no specific iteration order necessary for most operations in the ES and KNN, these collections were changed to `HashSets`. `HashSets` are also faster because they have  $O(1)$  access time, whereas `ArrayLists` have  $O(N)$ , where  $N$  is the number of elements in the collection. Access time is important because the tags in the KNN and ES are accessed often. `HashSets` also have the advantage of only permitting unique elements. This is useful because there should never be two copies of the same tag or the same KN in the system.

The `KnowledgeNode` class implements the functionality of a KN, which has the following important fields to implement the functionality described in Section 2:

`inputTag` Input Tag.

`outputTags` Array of output `Tags`.

`activation` int starting at 0, incrementing when the `KnowledgeNode` is excited.

`threshold` int threshold that causes firing of the `KnowledgeNode`.

`strength` int that biases the activation of a `KnowledgeNode`, causing early firing. The `KnowledgeNode` is fired if  $(\text{activation} \cdot \text{strength}) \geq \text{threshold}$ .

`confidence` int representing the belief that the `inputTag` is true (0 to 100).

`age` int representing the age of the `KnowledgeNode`.

The most important method in the `KnowledgeNodeNetwork` is `think()`, which chooses either `thinkForwards()`, `thinkBackwards()`, or `thinkLambda()`. These methods implement the functionality described in Section 2 and return the `Tags` activated as a result of thinking. At this point, `think()` chooses `thinkForwards()` by default. In the future, the choice will be determined by META (see ??).

Without a parameter, the `think()` method runs to natural quiescence. There is also an overloaded version of `think()` that takes an `int numberOfCycles` as a parameter. This parameter represents the thinking effort described earlier. Similarly, `thinkForwards()`, `thinkBackwards()`, and `thinkLambda()` all have overloaded versions with `numberOfCycles` as a parameter. This is a good example of the polymorphism principle of OOP.

Currently, only the `thinkForwards()` method is fully completed. The version that runs until natural quiescence can be seen in ??. The version with a fixed number of thinking cycles can be seen in ??. One can see that both `thinkForwards()` methods call `forwardThinkCycle()`, which can be seen in ??. This is where `Tags` can become active (line 9), being added into `activeTags`.

One can see that the `forwardThinkCycle()` method calls `excite()` (line 5) on a `KnowledgeNode`, which returns the `Tags` activated as a result of excitation. The method can be seen in ??. This, in turn, may call

`fire()` on line 5 to fire the `KnowledgeNode`, activate its output tags, and return the `Tags` activated as a result of firing. The method can be seen in ???. All these methods were made `private`, since they should never be accessed outside the KNN. This is a good example of the encapsulation principle of OOP.

A simple version of `thinkBackwards()` was also completed, activating a KN only if all its output `Tags` are active.

### 5.3 Expert System

All code relating directly to the ES was placed in the `es` package of the project. A UML diagram of the `es` package can be seen in ??? of the Appendix.

The ES layer is based around the `ExpertSystem` Java class. This class has the following fields:

<code>readyRules</code>	<code>HashSet</code> of <code>Rules</code> that have not been activated yet.
<code>activeRules</code>	<code>HashSet</code> of active <code>Rules</code> .
<code>facts</code>	<code>HashSet</code> of active <code>Facts</code> .
<code>recommendations</code>	<code>HashSet</code> of active <code>Recommendations</code> .

The most important method in the `ExpertSystem` is `think()`, which implements the functionality described in Section 2. Without a parameter, the `think()` method runs to natural quiescence. This can be seen in ???.

There is also an overloaded version of `think()` that functions similarly to that of the `KnowledgeNodeNetwork`. This can be seen in ???. Both variants of `think()` return the `Set` of `Recommendations` that become active as a result of thinking, which is passed on to the `META`. Both versions also call a method named `thinkCycle()` which represents a single cycle of thinking. The code for this method can be seen in ???.

Finally, we see in the `thinkCycle()` method a call to the `addTag()` method on line 21, which adds a `Tag` to the ES depending on its type. The

code can be seen in ???. The method is `public` because it could be used by a user of the ES to set the initial data structures. It returns `true` if the `Tag` is added successfully. The `addRule()`, `addFact()`, and `addRecommendation()` methods simply add a `Tag` to the `readyRules`, `facts`, or `recommendations` `Sets`, respectively, and return `true` if the ES did not already contain the `Tag` to be added.

### 5.4 Testing

All tests on the system were conducted using the TestNG framework in Java, which provides a simple and intuitive way to create assertions in tests. These tests were placed in the `test` package of the project. A UML diagram of the `test` package can be seen in ??? of the Appendix. The primary methods of interest in the test package are `testKNN()`, `testES()`, and `testKNNandES()`.

The test setup for the `testKNN()` method can be seen in Figure 7. It creates a KNN, with every node having a `threshold` value of 1 to simplify the activation process. The system starts out with `A` as the only active `Tag`, and `testKNN()` will make the KNN `think()` until natural quiescence. It is therefore expected that all the `Tags` shown in Figure 7 will become active by the end of the thinking process. The initial and final states are both asserted with TestNG in `testKNN()`, with positive results.

The test setup for the `testES()` method can be seen in Table 2. The columns from left to right (ignoring State) correspond to the elements in `readyRules`, `activeRules`, `facts`, and `recommendations`, respectively. The first row corresponds to the initial setup of the ES, and the last row corresponds to the expected final state. Both the initial and final states are asserted with TestNG in `testES()`, with positive results.

The `testKNNandES()` method has the same setup as `testKNN()` in Figure 7, except the output active `Tags` from the KNN are passed on to the ES. The resulting setup and activation in the ES can be seen in Table 3. The initial and final states were also tested with TestNG with positive results.

Remake these old tests. Name them 'simple'.

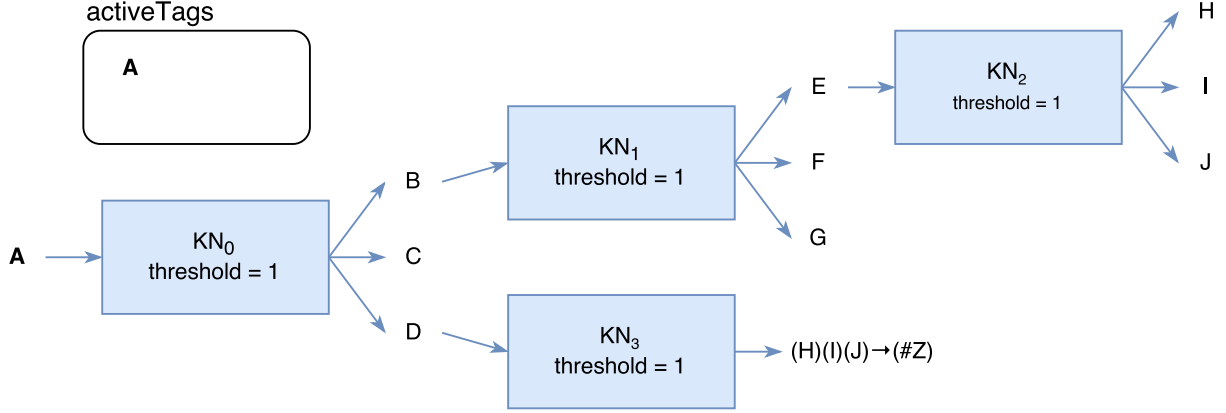


Figure 7: Initial setup for the *testKNN()* method.

Table 2: Test setup for *testES()*. Middle activation steps omitted.

State	Ready Rules	Active Rules	Active Facts	Active Recommendations
Initial	$(A)(B) \rightarrow (D)$ $(D)(B) \rightarrow (E)$ $(D)(E) \rightarrow (F)$ $(G)(A) \rightarrow (H)$ $(E)(F) \rightarrow (\#Z)$		$(A), (B)$	$(\#X), (\#Y)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Final	$(G)(A) \rightarrow (H)$	$(A)(B) \rightarrow (D)$ $(D)(B) \rightarrow (E)$ $(D)(E) \rightarrow (F)$ $(E)(F) \rightarrow (\#Z)$	$(A), (B),$ $(D), (E)$ $(F)$	$(\#X), (\#Y), (\#Z)$

Table 3: Test setup and activation for the ES portion of *testKNNandES()*.

State	Ready Rules	Active Rules	Active Facts	Active Recommendations
Initial	$(H)(I)(J) \rightarrow (\#Z)$		$(B), (C), (D)$ $(E), (F), (G)$ $(H), (I), (J)$	
Final		$(H)(I)(J) \rightarrow (\#Z)$	$(B), (C), (D)$ $(E), (F), (G)$ $(H), (I), (J)$	$(\#Z)$

## 5.5 Documentation

Extensive documentation was created in Javadoc<sup>1</sup>. UML diagrams were also created for each package in the source code to accompany the documentation (see Appendix A).

## 6 Results & Tests

## 7 Impact on Society and the Environment

### 7.1 Use of Non-renewable Resources

As purely a software project, there are no physical materials needed to construct this system. For this reason, less emphasis will be put on this section. The only related physical resources potentially needed are the materials needed to construct the robots and the computers to house the software. One area of concern could be the energy source of the robots themselves, which would probably vary depending on context. Ideally, the source should be a renewable one. For example, solar panels could be placed on the robots to provide energy.

### 7.2 Environmental Benefits

Prometheus could be used in a context beneficial to the environment. For example, the system could be used as a tool to control robots after oil spills, where the robots could theoretically contain the problem faster than humans and thus limit the risk on the environment. There is already research being done on employing robots in this context. Indeed, MIT's Senseable City Lab has been working on Seaswarm, a system composed of a fleet of vehicles to help clean future oil spills [14]. An AI like Prometheus to control a swarm of robots like this could be very valuable.

<sup>1</sup>The Javadoc can be found here: <http://cs.mcgill.ca/~sstapp/prometheus/index.html>

## 7.3 Safety and Risk

It is critical that, once this system is completed, it is used in an ethical way, and for the right purposes. One example of use that may cause ethical concern is in a military setting, where an AI system like the one described here could be used in a battlefield in place of soldiers. Indeed, the US Department of Defense has plans to employ AI for autonomous weapons to attack targets without human intervention in the future [15]. This would have the advantage of potentially saving human soldiers' lives [16]. However, there are also concerns because this would make it much easier to start a battle. More than 3000 AI and robotics researchers have signed an open letter arguing against a military AI arms race [17]. They argue that autonomous weapons would not be beneficial to society, since they would be ideal for assassinations, destabilizing countries or selectively subduing or killing a population.

Another possible issue is the future loss of jobs to be done by humans, with automated systems like Prometheus replacing physical labour. This problem can be solved as a society, perhaps by having more support for universal basic income (UBI). Telsa CEO Elon Musk is a proponent of the idea, saying that UBI will be necessary in the future [18]. Microsoft co-founder Bill Gates suggests possibly having a tax on robots to help pay for this income [19].

In the very long term, there are concerns with the possibility that an AI system might achieve intelligence and awareness close to a human. If such an AI were to obtain "consciousness" in its own way, should that entity be entitled to its own rights, like humans or animals are? The issue has been mentioned by the Institute for the Future, calling for the possibility of "robo-rights" in the future [20].

### 7.4 Benefits to Society

This type of system could be extremely useful in many contexts in society. For instance, the system could be used to send and control robots in an area that would otherwise be very danger-

ous for humans. For example, robots are often used in the aftermath of nuclear disasters to prevent the unnecessary loss of human life. Indeed, even in the Chernobyl nuclear disaster, the Soviet authorities employed the use of robots to avoid losses of human life [21]. With an AI like Prometheus, the robots could be controlled in an intelligent manner.

The system could also be used to further space exploration, with an AI controlling multiple robots exploring the surface of Mars, for instance. The value of AI in this context has been clearly shown, with the Mars Curiosity rover recently being upgraded to have its own AI system using computer vision to identify rocks [22]. This system can therefore help further important cutting-edge research in space exploration.

## 8 Conclusion

This semester, prototypes of the Expert System (ES) and Knowledge Node Network (KNN) layers of the Prometheus AI model were completed in Java based on personal design criteria and feedback from the supervisor. The design of the prototypes has required a great deal of thought and planning to ensure efficiency and proper behaviour. These prototypes were tested using the TestNG Java framework with positive results, and extensive Javadoc was produced. Possible impacts on the environment and society were also discussed.

The main goal for next semester is to finalize the entire system, implementing more complex features that were omitted for this prototype stage. This will also require proper integration between the work done in this report (KNN and ES), and outside this report (NN and META). Finally, the entire system will have to be tested, in simulation and with physical robots.

## A UML Diagrams



UML Diagrams of ES and KNN packages



Guice dependency graph diagram



## References

- [1] J. Vybihal, “Full AI model,” 2016.
- [2] ———, “Knowledge nodes,” 2017.
- [3] J. Palka and C. Stevens, “Neurophysiology — a primer,” 1967.
- [4] D. O. Hebb, *The organization of behavior: A neuropsychological theory*. Psychology Press, 2005.
- [5] T. Mitchell, *Machine Learning*, ser. McGraw-Hill international editions - computer science series. McGraw-Hill Education, 1997. [Online]. Available: <https://books.google.ca/books?id=xOGAngEACAAJ>
- [6] H. L. Roediger and K. B. McDermott, “Creating false memories: Remembering words not presented in lists,” *Journal of experimental psychology: Learning, Memory, and Cognition*, vol. 21, no. 4, p. 803, 1995.
- [7] S. E. Taylor, *Positive illusions: Creative self-deception and the healthy mind*. Basic Books, 1989.
- [8] M. P. Mattson and T. Magnus, “Ageing and neuronal vulnerability,” *Nature Reviews Neuroscience*, vol. 7, no. 4, pp. 278–294, 2006.
- [9] D. C. Knill and A. Pouget, “The Bayesian brain: the role of uncertainty in neural coding and computation,” *Trends in neurosciences*, vol. 27, no. 12, pp. 712–719, 2004.
- [10] J. Vybihal and T. R. Shultz, “Search in analogical reasoning,” 1990.
- [11] D. Kahneman, *Thinking, fast and slow*. Macmillan, 2011.
- [12] J. Vybihal, “Expert systems,” 2016.
- [13] J. L. Popyack, “Object Oriented Programming,” [https://www.cs.drexel.edu/~introcs/Fa15/notes/06.1\\_OOP/Advantages.html?CurrentSlide=3](https://www.cs.drexel.edu/~introcs/Fa15/notes/06.1_OOP/Advantages.html?CurrentSlide=3), 2015.
- [14] “A swarm of robots to clean up oil spills,” <https://sap.mit.edu/article/standard/swarm-robots-clean-oil-spills>, December 2010.
- [15] D. Gershgorn, “The US army wants to weaponize artificial intelligence — Quartz,” <https://qz.com/767648/weaponized-artificial-intelligence-us-military/>, August 2016.
- [16] B. Allenby, “The difficulty of defining military artificial intelligence,” [http://www.slate.com/articles/technology/future\\_tense/2016/12/the\\_difficulty\\_of\\_defining\\_military\\_artificial\\_intelligence.html](http://www.slate.com/articles/technology/future_tense/2016/12/the_difficulty_of_defining_military_artificial_intelligence.html), December 2016.
- [17] “Open letter on autonomous weapons - future of life institute,” <https://futureoflife.org/open-letter-autonomous-weapons/>.
- [18] A. Kharpal, “Tech CEOs back basic income as AI job losses threaten industry backlash,” <http://www.cnn.com/2017/02/21/technology-ceos-back-basic-income-as-ai-job-losses-threaten-industry-backlash.html>, February 2017.

- [19] “Bill Gates: Job-stealing robots should pay income taxes,” <http://www.cnbc.com/2017/02/17/bill-gates-job-stealing-robots-should-pay-income-taxes.html>, February 2017.
- [20] “Robots could demand legal rights,” <http://news.bbc.co.uk/2/hi/technology/6200005.stm>, December 2006.
- [21] G. Dvorsky, “A museum of robotic equipment used during the Chernobyl disaster,” <http://io9.gizmodo.com/a-museum-of-robotic-equipment-used-during-the-chernobyl-512831778>, 2013.
- [22] “How does mars rover Curiosity’s new AI system work? — astronomy.com,” <http://www.astronomy.com/news/2016/08/how-does-mars-rover-curiositys-new-ai-system-work>, August 2016.