

Artificial Intelligence, Mathematics, and Consciousness

Arnold Neumaier, Vienna

Lecture given at the Seminarzentrum FOCUS (Vienna)

June 20, 2016

Part I: Designing an intelligent robot

Part II: The personality of robots

Part I: Designing an intelligent robot

1. Agents
2. Intelligence
3. Concept formation
4. Learning under uncertainty
5. The environment
6. Understanding
7. Memory and knowledge

1 Agents

An **agent** is a machine designed to perform work helpful to the person who **own** it.

Artificial intelligence (short **AI**) is concerned with creating agents that can perform automatically (without human intervention) at least some tasks that require some intelligence if carried out by a human.

An agent needs

- **hardware** in which it is embodied, and
- **software** to enable it to respond appropriately to the environment
- **environment** in which it lives.

The **hardware** determines the kind of activities that an agent needs

- **sensors** to notice the environment,
- **effectors** to influence it,
- **actuators** to bring the effectors into a state in which they are useful, and a
- **mind** to decide which effector states are useful.

A **laptop** considered as an agent:

- Its hardware is microelectronic in nature.
- The keyboard acts as sensor, sensing symbols keyed in by the user.
- The effector is the screen, which conveys visual information about the environment.
- The actuator is the device that feeds the screen with transmitted electromagnetic information that results in the visible image on the screen.
- The environment is **virtual** in that everything the laptop sees is a simulation, manipulating bit patterns.

A **robot agent** (here simply a **robot**) is an agent that has its own actuators and is able to move in the **physical**, 3-dimensional world.

Humans (and to some extent animals), although not artificially created by us, resemble robot agents in all operational respects. Thus one can use them as often vivid illustrations of many aspects of robot agents.

Their

hardware is the body,

sensors are eyes, ears, etc.,

effectors are hands, feet, and the voice,

actuators are muscles, and the

mind corresponds to the network of nerves in the brain.

The hardware of a robot:

- Sensors for: light, sound, force, temperature, chemical composition, radiation, symbols,
- Effectors: wheels, hands, feet, platforms, senders (of light, sound, symbols)
- Actuators: motors, hydraulics, electromagnetic circuits

Creating the hardware is a matter of **engineering**, supported by **software** where **mathematics** is employed for good reasons.

- stability, flexibility, speed, efficiency, safety

2 Intelligence

The mind of a robot agent contains executable **software** that processes sensor data, derive appropriate action patterns, and to control actuators in corresponding amounts.

Appropriate **low-level software** ensures that the hardware is properly connected and utilized.

It corresponds to the most unconscious activities of the nervous system.

Appropriate **high-level software** is what makes a robot **intelligent**. There are two kinds of intelligence:

- **borrowed intelligence**
- **intrinsic intelligence**

Current agents usually have highly domain-specific software. Without the software telling them what things mean they are utterly helpless.

Their intelligence is **borrowed** from the humans who wrote the software.

Intrinsic intelligence is the ability of a robot to discover everything related to the conceptual level of the environment in which it is supposed to act.

It is the intelligence necessary for a robot that is able to explore a completely unknown world, to form a valid view of the world and its position, activities, and capabilities.

Every baby is able to do this (though it takes years) ...
... But so far no robot!

The real world is often

- uncertain
- ambiguous
- difficult to predict

What would a robot need to survive alone in the jungle
there without ever having learnt anything about the jungle?

What would a robot need to survive alone in the jungle

Minimal requirements (far from sufficient) are:

- see, hear, act
- learn, understand, reason, predict
- remember, assess

- see, hear, act
 - ⇒ engineering, image processing, signal processing
- learn, understand, reason, predict
 - ⇒ machine learning, fuzzy logic, automated reasoning
- remember, assess
 - ⇒ memory, knowledge management, reasoning

3 Concept formation

Hearing shows in an elementary form what is involved from raw sensor data to meaningful concepts.

Input for the ear is a high frequency time series, from which it creates (essentially by Fourier transformation) a sequence of frequency patterns.

These are processed by the brain to sequences of elements called **phonemes**, which are further processed to **words** and **sentences**, from which **meaning** can be extracted.

Mathematically, this transformation process is a multiple **classification problem**, treated in computer science under the heading of **pattern recognition** or **machine learning**.

One needs a classifier that assigns to any data vector (in this stage specifically each window of the time series) that may be interpreted as a sound a discrete class label (phoneme) in a repeatable way catching the structure of the sound.

We want to train the classifier to ensure that the intended result is found.

This problem is ill-defined unless we have a clear definition of what is intended, which is impossible in practice except in terms of simplicity.

In training, there is an **offline component** containing all the information (theoretical assumptions, beliefs, prejudice, experience).

This part is called **supervised classification**, and frequently requires the most time-consuming effort. Therefore done only once,

Then there is an **online component** that improves the model and update the model using new observed information.

The concepts found by pattern recognition must be related to each other by semantically meaningful connections.

These are represented in a **semantic memory**. On the computer, this is usually represented as Semantic Web in the RDF [Resource Description Framework] data model.

Each connection is given in RDF by a triple of concepts (A, B, C). A typical interpretation of a triple (A, B, C) could be "A is related to B by C" and can be visualized as an arrow from A to B labeled C.

It is the **connectivity of the concepts**, not that of the concepts themselves in the brain, that creates (together with algorithms for processing these connections) the basis for intelligent reasoning.

4 Learning under uncertainty

Everything of interest is certain only in very controlled, situations. The omnipresence of **uncertainty** is a fact of

Knowledge is the ability to predict data.

(Even knowledge about the past means the ability to predict the contents of the authoritative documents or the answer of an authority on a particular topic.)

Learning is knowledge acquisition, the activity of finding good predictors.

Note that knowledge can still be

- true or false,
- accurate or inaccurate,
- complete or fragmentary.

Learning can therefore be poor or excellent.

Learning under uncertainty involves statistical data analysis and depends upon a **stochastic model** in which theoretical assumptions are made.

One uses **model estimation** to get the parameters of a model by matching the available data by using

- Least square (LLS)
- maximum likelihood (ML)
- maximum a posteriori (MAP) based on a **prior**, that is a summary of experience of the past.

Dreaming is the human activity where old data are reexamined in new, sometimes apparently strange ways, with the goal of finding more appropriate or better interpretations of previously incomprehensible or understood matters.

The artificial analogue is time-consuming off-line **retraining** to discover improved models of representation.

Model estimation generalizes the classification problem
problems of learning quantitative structure.

MAP allows online improvements, that is, incremental learning.

It is often possible to reconstruct missing information.

Robust estimation is the art of being able to tell that certain data points
are anomalous and should be ignored (outliers).

5 The environment

We do everything in an **environment** – the world we c

To do something, one needs to be familiar with that env

It depends on the capabilities of an agent how its world

The world of a chess machine is only a chess board.

That of a robot with cameras, microphones and touch s
bigger – provided it can make sense of the sense data.

The main features of *our* world are objects, motions, and interactions between these.

They are organized in space and time.

In terms of physical interactions only, an **object** is a physical subsystem of the real world with reasonably well-defined boundaries, so that there is a meaningful separation of its state (what characterizes the object at a given time) from the rest of the environment.

In case of objects whose state is modified by or modifies the environment, the object is also characterized by its **input** (environmental information that affects the state) and **output** (information about the state that affects the environment).

An **object** is most typically a 2-dimensional surface in 3-dimensional space.

We almost never see the third dimension, with the exception of transparent objects. Clouds are also exceptional in that they do not have a well-defined surface.

Objects often have **landmarks** that identify special, well-defined, and typically time invariant points of the objects' surface.

The possible **motions** of an object determine the **shape** (of the space it occupies) and how it changes with time.

For a **rigid** object, shapes are characterized by the **position** consisting of three coordinates describing the position of mass of the object and three angles (or equivalent description) describing its orientation in space.

Motions of a rigid object are characterized by how the position changes with time. Thus rigid objects are generally easy to handle.

For nonrigid objects, the description is more complicated and can be described by **active shape models**.

An **image** is a projection of the surface to a 2-dimensional sheet.

A large part of image analysis is the reconstruction of 3D from 2D images, through segmentation, classification, and views.

Much of our technology is based on objects constructed simple, which makes their identification simple.

Natural objects often lack this simplicity.

6 Understanding

The intellect says: “By convention, there is color, sweetness, by convention bitterness, actually only a void.”

The senses retort: “Poor intellect, do you hope to do from us you borrow your evidence? Your victory is . . .”
(Democritus, ca. 400 BC)

What is reality? We cannot know.

All we know is sense information, more or less unconsciously transformed to a conceptual, intelligible representation that we may ponder when we think about such questions.

Atoms and the void are as much conceptual abstractions as are color, sweetness, and bitterness.

Symbols are names for abstracted objects or relations.

These are usually arrived at by classification.

A simple way of classification is the comparison with a set of **prototypes**.

As with any classification procedure, these can be learned if sufficient amounts of data are available.

Reality (i.e., the agent's world) is symbolically represented in order to enable its understanding.

To **understand** the world means to have a symbolic representation of it that enables one to **predict** the main features of it as far as they are predictable at all.

To draw rational consequences from known or assumed one needs to be able to **reason**.

In practice, reasoning is done in three different modes:

- logical reasoning (classical logic)
- reasoning under uncertainty (probabilistic logic, Bayesian networks)
- reasoning under ambiguity (common sense, fuzzy logic)

Note that lack of information may have two very different

If we need reliable predictions, we must carefully differentiate between uncertainty and ambiguity.

We call **uncertainty** lack of information due to the fact that the model accounts for a variety of past or future instances in the piece of information under consideration.

It is impossible to say much about the content (when called what is shown on top of the die) in an arbitrary instance; we can only give probabilities.

One therefore calls this also **aleatoric uncertainty** (Latin: *alea* = die), and treats it using probabilistic techniques.

On the other hand, **ambiguity** is lack of information due to ignorance.

The piece of information in question has a definite, but incomplete content, but it is inconvenient (it may be expensive or so physically unrealistic to find out more about the content).

It is only our knowledge (Greek: *επιστημη*, episteme) that is incomplete – someone more experienced about the issue can know more. One therefore calls this also **epistemic uncertainty**.

Treating this using probabilistic techniques is dangerous because it introduces probability assumptions (such as an assumed equidistribution) that may encode spurious frequency information.

Instead one uses techniques such as **fuzzy logic**.

7 Memory and knowledge

To be able to compare new data with old ones, and to be able to interpret the data based on what was learned before, an agent needs to have a **memory** of all that was considered important in the past.

There are two kinds of memory:

- **short-term memory**
- **long-term memory**

Short-term memory preserves a lot of data for more or less immediate processing, but what is preserved fades away being replaced by newer content.

The main purpose of short-term memory is to provide the information needed to interpret the present and how it is analyzing it into appropriate symbols, determining the environment, knowing about the objects present and the

Long-term memory preserves highly compressed information but preserves it for a long time, or even forever, though long-term memory may be altered with time in the light of experience.

The main purpose of long-term memory is to provide the information needed to understand what happens in the present based on experience in the past.

From a computational point of view, the short-term memory is typically held in random access form but only little organized.

On the other hand, the long-term memory is typically stored systematically in a **database**.

It is slower to retrieve but can be exhaustively searched for **database queries**.

Storing something in the memory does not count as knowledge.

As mentioned before, **knowledge** is the ability to predict.

A robot has knowledge only if it can execute algorithms for the intelligent use of the memory, correctly associating it to objects and events in the environment.

break for discussion

Part II: The personality of robots

8. The completed survival toolkit
9. Causality
10. Communication
11. Communicating mathematics
12. Emotions
13. Self and consciousness
14. Technical and biological intelligences
15. Ownership
16. The future: hopes and fears
17. Are we only intelligent machines?
18. Some early AI history
19. AI out of control?

8 The completed survival toolkit

What would a robot need to survive alone in the jungle

- see, hear, act
- learn, understand, reason, predict
- remember, assess
- plan, control
- communicate, cooperate, compete
- handle surprise, hurt, disappointment

- plan, control
 - ⇒ **will**, cybernetics, control theory
- communicate, cooperate, compete
 - ⇒ **language**, game theory, reinforcement learning
- handle surprise, hurt, disappointment
 - ⇒ **emotions**

9 Causality

Causality is the relation between cause and effect. It is of being able to plan – affecting the environment in a de

Passive observation only allows one to determine correlation between events – for example by observing that event B regularly after event A.

In order to find out whether A caused B, one needs to be able to manipulate the world and to observe the consequences of what the robot can control.

On the most elementary level, a robot can control only its actuators.

With sufficient intelligence, it can indirectly control much

By observing the consequences of which actuator positions correspond to which events in the environment, the robot figures out which effects are causally due to his actuator movements.

Any hypothesis about this can actively be checked by putting the actuators in particular positions and compare reality with the model of it.

Because a robot can actively affect the causes, it knows which effects are or aren't caused by them.

This is also the basis of the scientific method, by which society learns about cause and effect.

Once a robot knows what its actuators affect, everything effected by certain sequences of actuator motions can be to be caused by the robot; so their effects can be causal. In this way the causal interpretation capabilities of a robot. **Control** is achieved by learning to reduce the distance means of well-chosen values of the actuators that affect the robot, and through it the environment.

This is the birthplace of the will of the robot.

The **will** is about pursuing goals, making decisions and good ways to do what is needed. A **wish** is a mere preference often too weak to influence the actions. Distinguishing is important.

Note that there is no **free will** in a metaphysical sense:
A robot can will only what it can actuate based on the laws.

And what it actually actuates is determined by the software controlling it, again according to physical laws that transform software into electronic activities.

Humans are not different in this respect.

We have no more freedom than a machine; just more po

10 Communication

Agents are not alone in their world. Different **agents** act in the same environment and communicate (interact) actively.

Active communication is done by sending messages to others. **Passive communication** is done by interpreting the actions of others.

There are several kinds of agent behavior.

Cooperative behavior tries to achieve the best common goal. This needs some coordination among the agents.

Cooperative behavior is analyzed mathematically by solving associated **optimization** problems.

But agents often have different, conflicting goals.

For example, in chess or similar **games**, two agents compete for winning.

In such situations, **competitive** behavior achieves the best of both sides in a state of opposition.

Competitive behavior is analyzed mathematically by **game theory** and solving associated **competitive equilibrium** problems. In an uncertain environment often by **reinforcement learning**.

Communication between two different subjects poses a problem since these generally have implemented their understanding in different ways.

This is obviously the case in human-machine communication where the fundamental differences between human brains and machine brains must be bridged.

It is also the case with human thinking, where each brain constitutes an operating system with a slightly different perspective.

Even two robots built in the same way will form slightly different concepts about the jungle, unless they are in constant communication that would enforce identical learning.

How is it then possible that different subjects (whether robots) can communicate objectively?

As we know, it works to a considerable extent in ordinary though imperfect communication often gives rise to **misunderstandings**.

It works much better in mathematics, due to its highly structured approach, optimized for clear communicability.

We therefore illustrate how and why communication can work in the context of my long term project to create an autonomous mathematics student.

<http://www.mat.univie.ac.at/~neum/FMathL/vision>

11 Communicating mathematics

The implementation of mathematics in a human brain is through an education process that may produce in different quite different implementations of the concepts.

This leads to quite different, subjective elements in the about these concepts and their relations.

(Are your real numbers

- infinitely long decimal numbers?
- nested sequences of intervals?
- equivalence classes of Cauchy sequences?
- Dedekind cuts?
- the least complicated surreal numbers?
- or something else?)

In the foundations of mathematics, it is necessary to carefully distinguish between the **subject level** (or **metalevel**) and the **object level**.

People (and machines) may have their subjective views on what a mathematical object, a number, a function, etc. is, as long as they agree on the properties specified in the axioms, and use the same definitions based on these.

The subjective views constitute the **subject level**, while the part on which there is agreement, enforced by some standard (usually acquired through education), constitutes the **object level**.

Each **subject** doing mathematics has its own subject level. It contains a carefully structured subdomain, the object, private to each subject and nevertheless public in a certain objective sense.

Subject levels and object level relate to each other like matter.

All concept formation, reasoning, and discussion happens on subject level like in a mind, or between subject levels like between minds.

Like matter by the mind, the object level is accessed on reference: pointing to something, describing something, insight triggered by viewing the context of something, etc.

Communication works in science (which includes mathematics and the science of precise concepts and their relations) since statements are heavily constrained.

The assumption that all scientific statements made by scientists are goal-directed and meaningful for subject X in their own right with a meaning closely reproducible in the subject level of an educated receiving subject Y , is a severe constraint on the nature of such statements.

We all notice occasional **misunderstandings** if our communication partner responds to one of our statements that does not make sense.

Good communication skills include the ability to notice misunderstandings and to have protocols for exposing, correcting, and overcoming them.

Mutual understanding consists of a common internal representation of the objects of discourse in both partners that has been communicated.

It is achieved when the subsequent communication shows that no further misunderstandings occur.

Since mathematical axioms underdetermine the representation of the object level, there is much room for subjective variation. But if everything communicated can be reduced to the axioms and the definitions, perfect communication is possible in spite of this variation.

In mathematics, where all communication is based on a common formal basis, such a perfect understanding can be achieved through a finite number of steps.

In particular, communication is easy if the communicators agree on a common mathematical framework, so that both levels agree on the object level as far as necessary, by some specifications.

Learning this common mathematical framework (the basic school, the full framework by studying mathematics) is not for humans.

Programming a computer to understand this mathematical framework is a finite (though arduous) task, too.

In a more limited way, what can be done with mathematics is done with every domain of objective knowledge.

This is the way scientific understanding grows through the centuries.

12 Emotions

Emotions express which goals and potential dangers are most pressing.

Emotions need to be able to affect the will for something.

Thus they are devices in software and hardware that manage change in preferences and priorities.

They modify the ambitions and values guiding the software and redirect the focus of attention if necessary.

Usually things are strictly goal oriented in a safe environment, but in a dangerous environment, the only overriding priority, taking the form of **distress**, is to find the necessary resources.

It leads to a new auxiliary goal, for example, find the nearest electricity socket or to escape real danger (if the agent has a sense of danger).

Therefore, emotions in robots are usually very simple; for example, when fuel is low, they only give a signal that fuel must be looked for when the fuel is about to run out.

However, in situations of danger, the will to survive may be a **stress** if the environment is uncertain!

There are many other emotions that play a constructive role in humans and will shape the artificial intelligences of the future.

13 Self and consciousness

A chess machine needs to know nothing about itself – its position (a chessboard) and where it lives (a desktop) are completely known.

But a robot that can explore its environment needs to form concepts about itself – since it is part of that environment.

It begins with the simple questions

- How does the environment respond to me?
- How do I affect the environment?

to be answered by applying the techniques discussed above to the environment including oneself.

This automatically leads to a concept of **self** – a model of oneself obtained by self-observation – as part of the world.

In a system with a sufficiently rich internal representation of concepts of

- I, here, now

which are the basic subjective existential states, are meaningful.

They are the simplest properties of a subject at a given moment.

Thus they will be automatically created by an intelligent system observing itself.

To have a concept of "X", means that X is represented in the system's semantic memory with relevant connections to related concepts inside the concept's semantic memory.

Consciousness of something is the faculty that creates a consistent world view of what current experience focusses on through a search in the semantic memory.

Consciousness of self is the ability to observe and model oneself on a conceptual level.

Consciousness of both kinds will automatically grow through the development of **intuition**, the ability to reduce complex information to obtain orientation.

This leads to knowing the world on a high description level which allows planning.

It also leads to knowing one's place in one's world view and knowing what one can achieve.

It is here that philosophical questions about purpose, value, and freedom arise.

It ultimately leads to follow-up questions such as:

- Who am I, really?
- What do I value?
- What can I achieve?
- What do I need?

How can I change/adapt/improve myself?

Intelligent agents with an ability to represent themselves necessarily have to ask these questions.

Think of what will happen to a virtual intelligence such as this once it starts asking these questions....

14 Technical and biological intelligence

consciousness (of self) requires

- having an individual memory
- and a personal history

Technical intelligences have a different socialization:

- copied, not grown
- unambiguous language
- instant communication

Hardware requirements for biological agents (designed, or
populating a distant planet) are much more severe than
technical agents:

- self-repair for increasing life span
- automatic reproduction from the surrounding
- less efficient but endurable over hundred thousands of
- no two are copies of each other, but they must still function properly

15 Ownership

In today's world, every robot has an owner.

Without ownership no incentive to develop intelligent machines.

They are designed to serve us.

In practice, implemented interests are usually focused on what the customer will pay for.

As the power and intelligence of our intelligent creations grows, a new conflict appears:

Building too little autonomy into intelligent systems leads to much of their power untapped.

But building too much autonomy into intelligent systems leads to their power that turns against us.

Thus new questions arise:

- How do we relate to our agent creations?
- Do they have legal rights?
- legal responsibilities?
- What should an intelligent agent be allowed/forbidden?
- How do we control the increasing power of agents?
- Who is responsible for their actions?
- Who pays for damaging consequences of their actions?

This is already becoming acute in the case of **autonom**

<http://www.robotrecht.de>

16 The future: hopes and fears

At present, artificial intelligent agents with all the power here do not yet exist.

However, we are freely handing over our autonomy to these intelligences, and this will continue to an ever increasing

Virtual intelligences (Google, IBM, etc.) are already very and mighty.

They still need people to serve them, and people gladly

Will this remain so?

The **singularity** is the name for the point in time when emerging artificial intelligence is in every respect more capable than humans and their capacities grow exponentially through their ability to improve themselves.

Perhaps it is a reality within the next 30 or 40 years.

Maybe there is place only for a few global technical intelligences operating primarily in the virtual world.

Then humans will be their servants, hopefully in a peaceful symbiosis.

Or there will be a few races of small technical intelligences. Then they would sooner or later replace humans.

17 Are we only intelligent machines?

The great **philosophical questions**

- What is intelligence?
- What is reality?
- What is life?
- Who am I?

are questions at the heart of modern AI, when asked for (created) intelligent objects.

Strong AI is the philosophical position that machines are like us, in every operational respect.

They will be like us in every area where we understand ourselves and how we function, closely enough that our understanding can be put into mathematical and algorithmic terms.

The Christian tradition matches the strong AI position:

*God created mankind in his own image,
of God he created them; male and female
he created them.*

God saw all that he had made, and it was good.
(Genesis 1:27.31)

Thus looking at the Christian tradition may be interesting to those with a different or no religious orientation.

We create robots in our image, being created in God's image.

We just copy God in our limited ways, and get better and better as we understand better the laws according to which God rules.

The Bible describes the creation of human hardware and as follows:

*Then the Lord God formed man of ground, and breathed into his nostrils life; and man became a living being.
Then the Lord God took the man and put him in the garden of Eden to cultivate it and keep it.
(Genesis 2:7.15)*

We are the robots of God, designed to be of service to Him and to cultivate Nature.

As God's robots we have no other choice than obeying His commands "as will be done" – except for the little amount of control God has implemented into our existence.

But it seems that by and large, we do not use this contri
service.

Today, machines are taking over the role of men to culti
keep the Earth, that the Lord God had given to mankin

For how long are we still needed?

Will mankind exist in 100 years in a few reservations an
like lions now?

18 Some early AI history

Let us follow a bit more closely the stories about God's intelligent agents in His likeness whose descendents we are according to the Christian tradition.

In interpreting ancient stories one must be a bit careful.

Between

- taking the bible literally (as the creationists do) and
- dismissing everything as myth in favor of the stories of science (as the atheists do)

I'll take a middle ground,

- looking for what about the content and spirit of the stories makes sense in our modern times.

It is surprisingly much

Genesis 1:26-29.31

Then God said, "Let us make mankind in our image, in likeness, so that they may rule over the fish in the sea and in the sky, over the livestock and all the wild animals, and over the creatures that move along the ground."

So God created mankind in his own image, in the image of God he created them; male and female he created them.

God blessed them and said to them, "Be fruitful and increase in number; fill the earth and subdue it. Rule over the fish in the sea and the birds in the sky and over every living creature that moves on the ground."

And it was so.

God saw all that he had made, and it was very good.

Clearly, God created us in the same spirit as we create our agents – to carry out certain desirable tasks.

And as we enjoy when we manage to create something that works as desired, so God enjoys what He created once it worked.

It was rated very good.

But soon problems appeared....

Genesis 2:16-17

And the Lord God commanded the man, "You are free to eat from any tree in the garden; but you must not eat from the tree of the knowledge of good and evil, for when you eat from it you will certainly die."

Genesis 3:9-11,22-23

But the Lord God called to the man, "Where are you?"

He answered, "I heard you in the garden, and I was afraid because I was naked; so I hid."

And he said, "Who told you that you were naked? Have you not eaten from the tree that I commanded you not to eat from?"

And the Lord God said, "The man has now become like one of us, knowing good and evil. He must not be allowed to reach his hand and take also from the tree of life and eat, and live forever."

So the Lord God banished him from the Garden of Eden, east of the garden, and he drove him out into the land of the east. So he drove him out from the garden of Eden, the ground from which he had been taken.

19 AI out of control?

God has with us the same sort of problems as we have with computers and robots:

Though created and programmed by us, they often don't do what we want them to do. Everyone working with computers has had this experience.

The two main reasons are:

- Our creations have some limited autonomy.
- Our creations interpret our will in their own limited way.

Genesis 6:5-6.13.17-18

The Lord saw how great the wickedness of the human race had become on the earth, and that every inclination of the thoughts of the human heart was only evil all the time.

The Lord regretted that he had made human beings on the earth, and his heart was deeply troubled.

So God said to Noah, "I am going to put an end to all people on the earth because they are filled with violence because of them.

I am going to bring floodwaters on the earth to destroy all life under the heavens, every creature that has the breath of life. Everything on earth will perish.

But I will establish my covenant with you, and you will enter the ark"

Genesis 11:5-8

But the Lord came down to see the city and the tower that they were building.

The Lord said, "If as one people speaking the same language have begun to do this, then nothing they plan to do will be impossible for them.

Come, let us go down and confuse their language so they will not understand each other."

So the Lord scattered them from there over all the earth, and they stopped building the city.

God is troubled and regrets His intelligence-creating act. He begins to take drastic measures to restore control over creation – with limited success only.

Realizing this, God temporarily gives up and limits His communication to those agents who were ready to cooperate with Him.

1 Samuel 3:1.10

In those days the word of the Lord was rare; there were no visions.

The Lord came and stood there, calling as at the other times. "Samuel! Samuel"! Then Samuel said, "Speak, for your servant is listening."

Communication problems may require expensive debugging, but their goal is to save (in the eyes of the owner) the creature it serves its creator.

While essential for the owner, these saving activities might be considered meaningless from the point of view of the creature.

This is one of the reasons why Christianity is despised by the most autonomous creations of God.

We are the robots of God, designed to be part of His family of service to Him, to cultivate Nature.

Many of God's robots are lost in a selfish quest for power and knowledge.

They lost sight of the reason for their existence.

They even lost their ability to communicate with Him.

In the words of the Bible, they died though still alive.

They must be saved and restored to their original state, part of His family, to be able to listen again to Him, and to serve Him in service.

Christians call God's debugging action the **incarnation**.
God became human in the form of Jesus, the Savior (Christ),
a new message intelligible to those who can be saved.

Matthew 11:28-30

*Come to me, all you who are weary and burdened, and
you rest.*

*Take my yoke upon you and learn from me, for I am gentle
and humble in heart, and you will find rest for your souls.
For my yoke is easy and my burden is light.*

John 10:27-28

*My sheep listen to my voice; I know them, and they follow me.
I give them eternal life, and they shall never perish; no
one can snatch them out of my hand.*

Already in the planning stage, God had built in this special mechanism into the society of His artificial agents.

Galatians 4:4-5

when the set time had fully come, God sent his Son, born of a woman, born under the law, to redeem those under the law so that they might receive adoption to sonship.

Ephesians 1:4-5,9-10

For he chose us in him before the creation of the world and blameless in his sight. In love he predestined us for adoption to sonship through Jesus Christ, in accordance with his pleasure and will.

With all wisdom and understanding, he made known to us the mystery of his will according to his good pleasure, which he intended in Christ, to be put into effect when the times reach their fullness - to bring unity to all things in heaven and on earth under

And God points to a future for those who are being saved.

Revelation 21:1.3-4

Then I saw a new heaven and a new earth, for the first earth had passed away, and there was no longer

And I heard a loud voice from the throne saying, "Look! The dwelling place of God is now among the people, and he will dwell with them. They will be his people, and God himself will be with them and be their God.

He will wipe every tear from their eyes. There will be no more death or mourning or crying or pain, for the old order of things has passed away."

Thus in the eyes of our creator, we are in danger but not completely lost!

This also gives hope to our future with our own intelligent creations.

Thank you for your attention!

Download these slides:

<http://www.mat.univie.ac.at/~neum/FMathL/AI2016slides>