

Autonomous control of a device using a technical motivation value

The present invention relates to a method for the autonomous control of a device, wherein the device generally moves in a physical real world and does not need to be further specified in terms of its physical configuration. The method offers the advantage that the device performs autonomous learning and continuously improves the knowledge or behaviour it has learnt. In general, it overcomes the disadvantage of the prior art, namely that training data must first be created, as is the case with conventional methods of artificial intelligence. In general, the method is universally applicable and the device learns autonomously, constantly correcting its own knowledge base in the process. Furthermore, a device is proposed which is configured to carry out the method, as well as a system arrangement comprising several of the proposed devices. In addition, a computer program product and a computer-readable storage medium are proposed, which carry out the method steps or cause a computer to carry out the method.

Various methods from the field of artificial intelligence (AI) are known from the prior art, which involve providing training data and having algorithms recognise patterns during a learning phase. In this context, it is possible for specific algorithms to recognise patterns within the database and subsequently extract them. Thus, implicit knowledge is extracted from large volumes of data. Once the training phase is complete, the correspondingly trained algorithms are applied to actual data and, in turn, generate implicit dependencies from this to solve real-world problems. A disadvantage here is that the training data must first be provided, and undesirable influence on the learning phase can already occur at this stage, which is not only error-prone but also time-consuming.

Furthermore, swarm intelligence is known from the prior art, whereby multiple devices are provided and these then solve a problem collaboratively in a distributed manner. Here too, the control and coordination of the individual participants is complex and, in some cases, prone to errors. This is particularly true when distributed learning is to take place, which in turn must be coordinated.

Artificial neural networks are also known from the prior art, which, based on graph theory, provide neurons and connections—i.e. nodes and edges—. It is known that these networks mimic functions of the human brain and are capable of learning in the process. Thus, edge weights can be varied, new edges can be added or old ones deleted, and existing nodes can be deactivated or new nodes added. In this way, a dynamically learning overall system is created.

However, using a KNN in the manner described above presents at least four problems:

1. For objects on which the KNN has not been trained, the KNN provides no information to call up the routines assigned to the object.
2. The programmed properties or behaviours of the recognised and identified objects may be incorrect, may have changed in the meantime, or may still be unknown.
3. One or more scientists, experts, etc. select the training data and specify the desired outcomes. Even in unsupervised learning, the training data and hyperparameters are still specified by humans, and this cannot rule out errors.
4. A KNN cannot correct errors. Each piece of information is distributed across all parameters of the KNN, similar to a hologram in which every data point contains information about the entire image. A KNN must therefore always be completely reset and retrained from scratch.

The device referred to here as 'Ego', on the other hand, does not fail when encountering unknown objects, unlike KNNs, as seen, for example, in the video where the self-driving car suddenly spots two boxes on the road. KNN generally stands for a Artificial Neural Network.

In the state of the art, there is a need to create an autonomous control system for devices in such a way that time-consuming preparatory work, such as the provision of selected and interpreted training data, and the training process itself can be prevented or minimised.

There is therefore a need for a self-learning system that is also capable of learning collectively or estimating what other participants are likely to do or are even capable of doing. Consequently, there is a need for a system comprising multiple participants, wherein each participant actually retains knowledge about the other participants and further develops this, which is referred to here as empathy .

Accordingly, it is an objective of the present invention to propose a method for the autonomous control of a device, which acts independently solely on the basis of provided target information and thereby builds up action knowledge. The proposed method is intended to be capable of recognising the scope of action of effectors of any device and to create an improved method or at least an alternative method for the autonomous control of the device or devices. Furthermore, it is an object of the present invention to propose a device for carrying out the method, as well as a system arrangement comprising a plurality of the proposed devices. Furthermore, it is an object to propose a computer program product and a computer-readable storage medium containing instructions that carry out the method.

The objective is achieved by the features of patent claim 1. Further advantageous embodiments are set out in the sub-claims.

- Accordingly, a method is proposed for the autonomous control of a device, comprising initialisation by means of randomised actuation of effectors and reading out a plurality of
- 5 internal sensor units configured to measure internal device properties and a readout from a plurality of external sensor units configured to measure external environmental properties; storing interrelationships existing between the actuation of the effectors and the read-out device properties and environmental properties as action instructions, such that an action instruction transforms an initial triplet of actuation, device properties and environmental
- 10 properties into a target triplet; and controlling the device in accordance with at least one predefined target device property and/or at least one predefined target environmental property using at least one stored action instruction, which specifies how the at least one predefined target device property and/or the at least one predefined target environmental property is set on the basis of the output triplet and the target triplet.
- 15 The proposed method serves the purpose of autonomous control, i.e. the creation of action instructions, of a device which is generally of a physical nature. This may be a motor vehicle, or alternatively a robot or a production plant. With regard to mobility, no restrictions are envisaged here, so that the device can move on land, in the air, on water or even underwater. In this context, 'autonomous' means that the method ensures the device learns
- 20 independently and also independently recognises which courses of action are available. These can be learnt, and the sensors are used to determine which action leads from which starting point to which result. Generally, it is possible for the device to set its own objectives, or for the objectives to be transmitted externally. In the case of self-generated objectives, this may, for example, involve maintaining operational capability. An internal objective might,
- 25 for instance, be to seek out a charging station when the battery level is low. An external objective can be communicated by specifying which task or activity the device is to perform.

- In a preparatory step, initialisation takes place by randomly activating effectors. An effector is generally a physical unit that has an impact on the real world. This may, for example, affect the device itself, such as steering, accelerating, braking, etc., and/or involve the targeted
- 30 manipulation of objects by, for instance, robotic arms, either directly or using tools, or by instruments with remote effects, be they projectiles or beams, or, for example, fire-fighting agents such as water or fire-extinguishing foams. In the case of a car, this could be the brakes, the accelerator pedal, etc., but also the headlights, indicators, etc. The effectors are

actuated and then, using internal and external sensors, the effects of these actions are measured. This is recorded and can then be used in the further course of the process. In this way, actions are learnt and used effectively in subsequent process steps.

Since the proposed device or method can be operated or controlled entirely without any prior
 5 knowledge, the preparatory step involves randomised operation, i.e. arbitrary actuation, as it must first be learnt which actions are even possible and what effects they can have, without them already being performed or trained with a specific goal in mind, so that what is learnt is not tied to specific external goals and might not be applicable to other external goals later on. In further iterations of the process, the learned actions are then performed in a goal-
 10 oriented manner. During randomised actuation, the system parameters of the effectors are tested and for example, a robotic arm is moved into all possible positions. This is recorded in each instance, and it is determined which action of the effector has what effect on the real world. In the example of a headlamp, it is possible that it only provides the on and off parameters. The headlamp is then switched on and off, and external sensors measure the
 15 headlamp's impact on the environment, whilst internal sensors measure the headlamp's operating parameters, such as temperature changes. In the case of more advanced LED spotlights, it is also possible to adjust both the colour and the intensity. This is tested by means of randomised activation, and the internal and external results are then logged.

According to the invention, internal sensor units and external sensor units are proposed. In
 20 this context, 'internal' and 'external' do not refer to the arrangement of the sensor units, but rather, in accordance with the present invention, denote that the sensor units measure external conditions relative to the device or, analogously, internal conditions. Internal conditions are all system parameters of the device itself. External conditions are environmental variables, i.e. those relating to objects surrounding the device. Thus, device
 25 properties are measured by means of the internal sensor units and environmental properties by means of the external sensor units. If, for example, the device has a certain battery state, this is an internal parameter. If an object is transferred from A to B by means of a robotic arm, this is an external state, whereas the state of the robotic arm is, in turn, an internal state. Typically, therefore, internal and external parameters interact, and actions, for
 30 example, always result in a reduced battery state, whilst these actions in turn affect external properties. In the opposite direction, there is an interaction whereby effectors can be used to guide the device to a charging point, whereupon a charging process is initiated, which in turn influences the internal battery state.

The recorded interrelationships are then stored, thereby logging how the actuation of the effectors influences the internal and external parameters. The device properties are thus stored together with the environmental properties, thereby defining action instructions. The actuation of the effectors is therefore an action that influences both the device itself and the environment. For example, if the device travels from a first geographical point to a second geographical point, this changes the device's environment—which is an external parameter—and this also alters the device's internal state, for instance through a change in temperature or a change in the battery's condition. Thus, a triple can be stored that specifies what the internal state was, what the external state was, and what action is then carried out. This is converted into a new internal state and a new external state. Thus, there is an initial triplet comprising the actuation of the effectors, device properties and environmental properties, and this is converted into a target triplet. In the target triplet, the new internal state, the new external state and possible actions can be specified.

The concept of triples is intended merely to illustrate the general transition of system states. In general, it is also possible as an alternative for internal states and external states to be converted into a new internal state and a new external state by means of a function. The action thus consists of a mapping function from a first tuple to a second tuple. The skilled person will recognise that the cardinality of both descriptions is equivalent. Either an initial tuple is transformed into a target tuple by means of a function, or an input triple is formed, specifying internal states, external states and an action, whereby this action then leads to a target triple which has a new internal state, a new external state and a further field. The additional field may remain empty; however, it is preferable that at least one action be performed here that is now possible in this state. The final field may also be populated such that, for example, a vector containing identifiers for further actions is introduced here.

To illustrate this with an example, the device can actuate the effector known as the motor drive and move forwards. What is now recorded is an output triplet comprising an internal state, namely a battery charge level, an external state, namely an image signal of a physical environment, and an action instruction, namely 'Drive'. This output vector or output triplet is now converted into a target triplet, which comprises a new battery charge level, a new image signal of the environment, and actions that would now be possible, such as moving forwards or backwards. During the action of moving forwards, it is also possible to specify that, for example, braking or activating the headlights would be possible.

In an alternative example, the initial triplet of the battery state and the image signal of the

environment is converted by means of the 'Move' function into a target triplet, which describes the new battery state and a new image signal of the environment. Thus, a record is made of what occurs internally and externally during a specific action. This knowledge can be reused in further action steps towards a predefined goal.

- 5 In further procedural steps, target device properties or predefined target environment properties may be provided. The device can thus be controlled in such a way that the at least one predefined target device property and/or the at least one predefined target environment property is established. The action instructions are known from the preceding procedural steps, and it is therefore known how an internal state and an external state can
10 be transformed into a further internal state and a further external state. The target specifications, i.e. the target device property and/or the target environment property, serve this purpose. These target properties thus determine what is to be achieved, and this is accomplished by executing the action instructions. The system therefore reads out the current state of the device and then selects stored actions which, starting from the actual
15 state, achieve the target state. In a preferred case, this is achieved by executing a single action instruction that immediately meets the target specifications. In a typical case, however, the initial state will be successively transformed into the target state by means of several action properties. Several action instructions are therefore linked in such a way that the target specifications are ultimately achieved.
- 20 To illustrate this with an example, the method may have determined that the vehicle can move from A to B to C to D. Consequently, the instructions are stored to enable the car to travel from A to B, from B to C, and from C to D. It is also implicitly stored that the car can travel from A to C and from A to D. Furthermore, it is stored that the car can drive from C to D. These possibilities are now utilised, and if a target specification states that the car is to
25 drive from A to C, the device has two options to choose from: namely, to drive from A to B to C, or to drive from A to C. In general, it would also be possible to drive from A to D and then from D to C. Which option is selected can also be defined in the target specifications. For example, a maximum journey time can be defined in terms of ' '. Furthermore, a maximum energy consumption can be defined. In preparatory process steps, actions are therefore
30 executed randomly and these are then available at runtime. In general, the method may provide for randomised actuation of the effectors, but it may also additionally provide for the reuse of previously learnt action instructions. Thus, a learning phase of randomised actuation can be carried out, as well as an execution phase that draws on previously stored actions. These can also be alternated in any order. Furthermore, during the execution

phase, the stored action knowledge is refined using internal and external sensors, thereby generating new action instructions that may be preferred in relation to a different target.

Furthermore, even whilst an action is being performed, the objectives may change in such a way that, for example, a battery status becomes critical; in this respect, the objective of

5 maintaining operational capability is prioritised to such an extent that the system recognises it is now necessary to proceed to a charging station. This results, at least temporarily, in an overriding objective, namely to proceed to a charging station, even if this does not achieve the objective of proceeding to the original target coordinates. As soon as a certain charge level is reached, the objective of the vehicle's own operability is downgraded again and the
10 objective of proceeding to the destination is increased to such an extent that the journey is now continued.

According to one aspect of the present invention, the method is iterated in a learning manner, such that new action instructions are constantly identified, each of which converts an initial triplet into a target triplet, and these action instructions are stored for controlling the

15 device. This has the advantage that the method is established, either wholly or at least partially, in such a way that new situations are constantly recognised and action instructions are created in such a way that effectors are actuated. Thus, the method can be varied in such a way that the actuation of the effectors no longer takes place randomly, but rather existing action instructions can be combined, or it is also possible for parts of existing action
20 instructions to be randomised, so that new possible action instructions are created for which it is also clear which interactions they trigger. Furthermore, the method can also be established in such a way that it is established from the storage of interrelationships right through to the control of the device. Thus, even when executing known action instructions, interrelationships are stored and new parameters can be identified. For example, it can be
25 recognised that executing the same action twice results in different target parameters. This may be the case, for instance, when a crosswind arises during a car journey. This is stored, and the external sensor then detects that a wind must have occurred. A new output triplet is thus created, and the effectors can be used to randomly test how to countersteer in such a situation. New interrelationships have therefore been identified, and if such an output triplet
30 is identified again, it is now clear which action must be performed to achieve the desired target triplet.

According to a further aspect of the present invention, the reading of internal and/or external effectors is carried out in such a way that individual reference values are stored and intermediate values are interpolated and/or extrapolated. This has the advantage that the

reference values can be selected according to their availability and that they can also be estimated in such a way that a measurement does not have to be available for every possible value. Rather, existing measurements can be used to estimate which values would result under normal circumstances. Thus, further reference values can be calculated

5 mathematically.

According to a further aspect of the present invention, the target device property and/or the target environment property is influenced, at least temporarily, by a device property and/or an environment property. This has the advantage that the target specifications can also be changed or prioritised. If, for example, a journey is to be made from a first geographical point

10 to a second geographical point, stopping at a charging station may be prioritised in the meantime, as otherwise the final destination would not be reached. This thus has the advantage of always guaranteeing that the target specifications are met and creating a fail-safe procedure.

According to a further aspect of the present invention, several action instructions are

15 combined into a choreography that transforms a source triplet into a target triplet. This has the advantage that even complex, i.e. composite, actions can be executed, and as the process progresses or is established, new action instructions are constantly created, which are combined into increasingly efficient choreographies. The proposed method is thus improved iteratively.

20 According to one aspect of the present invention, the device is a manufacturing robot with two robotic arms, each having an effector (gripper), which are designed in accordance with the device, i.e. they 'know' how to grasp, move, manipulate, etc. objects.

The device is now shown just once how the new workpieces, Workpiece-1 and Workpiece-2, are to be grasped, turned and then joined together. The result is then marked as the target

25 for the device. The device can now repeat this at any time: show it once, rather than training extensively.

If the two workpieces now appear on a conveyor belt, two alter egos are created for each, and based on the situation, the initial triplet, and the associated target, the motivation value drops into negative territory, causing the device to recognise a call to action and join the two

30 together, which leads to a motivation value that brings the device to a standstill or resting position. This is an example of how the device's ability to learn from observation can be utilised.

According to a further aspect of the present invention, the method is executed on a first device and, using external sensor units of the first device, a second device is detected which also executes the method, and the interactions generated by the second device are stored on the first device or the interactions are stored remotely and made available via an interface
5 at least to the first device and/or the second device. This has the advantage that the method is executed on multiple instances and the first device observes the capabilities of the second device. The behaviour of the second device is thus analysed and stored on the first device or centrally, accessible to the first device. The results of executing the method on the second device are thus read out and/or measured by the first device and stored centrally as potential
10 actions for the first device.

According to a further aspect of the present invention, the first device is controlled on the basis of the detected interactions of the second device. This has the advantage that the first device can be controlled in accordance with at least one predefined target device property and/or at least one predefined target environment property, using at least the stored action
15 instruction of the second device.

According to a further aspect of the present invention, interrelationships of the detected second device between its parameters and its actions are used to generate action instructions for the first device. This has the advantage that a network of devices is created which, by means of randomised actuation of effectors and reading out data from a plurality of
20 internal sensor units configured to measure internal device properties and from a plurality of external sensor units configured to measure external environmental properties, generate or interactions and then share these amongst themselves for respective or joint use.
Furthermore, the conversions of the output triples into target triples can be carried out collectively. Thus, corresponding choreographies are not limited to a single device executing
25 the method, but rather the choreographies can be carried out by several devices in a division of labour.

According to a further aspect of the present invention, an object is detected using external sensors and its detected parameters are compared with device characteristics and/or operational instructions of the devices by mapping them onto a copy of the central template,
30 whether these have been trained by the device itself or adopted and implemented from a comparable device. This has the advantage that the device can, in effect, clone itself or infer the properties of the object from its own properties. This can also be described as empathy.
If, for example, external sensors identify a device that exhibits characteristics similar to those

of the executing device, it is assumed that this newly detected device possesses capabilities similar to those of the executing device. To illustrate this with an example, we consider a vehicle that is carrying out the proposed invention. This first vehicle therefore carries out the proposed method and, using an external sensor—in this case an imaging unit—detects another, i.e. a second, device. The first device, or the first vehicle, has learnt that at a certain speed it can only brake to a limited extent and can only steer sideways to a limited extent. The first vehicle now detects that the second vehicle has similar dimensions and is travelling at a similar speed to the first vehicle. The instructions are then conceptually projected into the second vehicle, and the first vehicle recognises that the second vehicle can only brake to a limited extent and can only change direction to a very limited extent. If an overtaking manoeuvre is now initiated on a motorway, the first vehicle recognises that the second vehicle could brake or could also change lanes. It is thus concluded from its own behaviour, or rather from the possible initial triplets and the possible target t triplets, that this object has similar characteristics. Furthermore, the behaviour of the detected second vehicle can be monitored using the external sensors, and the vehicle's own data memory, which stores interrelationships, can then be updated. Thus, source triplets and target triplets can also be detected, and inferences can in turn be drawn from the second vehicle to the first vehicle.

According to a further aspect of the present invention, the detected parameters are used to predict the behaviour of the detected object. This has the advantage that it is recognised that the device itself executes a specific action instruction in a given situation, so that the other detected object would very likely also execute the same action in the same situation. The device's own actions or executed action instructions are therefore also logged, and it is then assumed that the detected object could behave in the same way or at least similarly.

According to a further aspect of the present invention, interrelationships of the detected object between its parameters and its actions are used to generate action instructions for the device. This has the advantage that external recognised interrelationships relating to the proposed device or the device controlled by the method can also be utilised. Thus, the device does not merely create interrelationships that are evaluated, but rather the external world can also be observed, and conclusions regarding the mapping of source triples to target triples can then be drawn.

According to a further aspect of the present invention, an effector is provided in the form of a processor, a memory, a motor, a gripper arm, a locomotion unit, a drive, a limb, a windscreen wiper, a direction indicator, a light, a steering mechanism and/or an actuator.

This has the advantage that, using the proposed device or the device to be controlled according to the method, all possible hardware units may be present. The implementation is merely illustrative and not exhaustive, so that any physical unit may be controlled autonomously in accordance with the proposed method.

- 5 According to a further aspect of the present invention, internal sensor units detect an effector state, an effector setting, an effector position, a size, a weight, a dimension, a speed, an acceleration, deceleration, a battery level, a range, a position, a processor load, a memory load and/or a system parameter of the device. This has the advantage that the internal sensor units can measure all parameters and states of the device that performs the method
10 or that is performed or controlled by the method.

- According to a further aspect of the present invention, external sensor units detect a distance, an object, a lidar signal, an image signal, a position, a size, a dimension, an acoustic signal and/or an external parameter. This has the advantage that the entire environment of the device can be analysed and detected. For this purpose, any possible
15 sensors that detect the environment in any way may be used individually or in combination.

- Since trained neural networks are proposed in accordance with one aspect of the present invention, it is advantageous that the central template of the invention assumes the functions of trained neural networks, but far exceeds those of the NN in terms of its capabilities and flexibility.
20

- The central template can either be trained itself or copied from a comparable device. In the case of an artificial intelligence system for a smart home, a central template can never learn to move on its own, for example to monitor the actions of an elderly person as part of health monitoring (keywords: empathy, fall detection, etc.); in the state of the art, it must be copied
25 from another source and implemented.

- The initial training or development of the central template is comparable to children at play and is not determined by the achievement of externally defined goals. It serves solely to learn how to act: what am I capable of doing? Later, by combining the possible actions learnt
30 in this way, the external goals can be achieved, whereby the path to them is or must be devised by the system itself. Otherwise, the actions would be fixed on specific goals from the outset (opening a door), only to fail when other goals (vacuuming) needed to be achieved.

In accordance with one aspect of the present invention, this involves testing one's own capabilities, such as accelerating, steering, etc. – similar to 'babbling'. However, the device simultaneously recognises and stores (and this goes beyond 'babbling') the changes (distances, speeds, acceleration, etc.) that are thereby brought about, so that it ceases to set its own actions (steering) as a target triplet, but rather the changes it can achieve in this way, and this goes beyond 'babbling'. Just as a child first learns (must learn) to crawl and then begins to crawl purposefully here and there. Thus, for children as well as for the device, the target triples become increasingly 'abstract' in comparison to their own, 'original' possibilities for action, which thus become mere means for ever more complex target triples. NN, on the other hand, are and remain limited to the trained initial and target triples.

The device, in accordance with one aspect of the present invention, sets its own goals or goal triples during the juvenile, or training, phase, just as small children do: I want to go there, I want to bump into that. This random, independent goal-setting negatively influences the internal motivational value; in other words, it is 'deflected'. Now, through its own actions, it attempts to change this value in the opposite direction, i.e. positively. A child would probably perceive this as (positive) excitement or tension (i.e. not as a negative feeling such as fear, for example). Technically speaking, each set of three goals can and is thus mapped onto an internal value through which the device can control its own behaviour. With NN, the user must specify the achievement concretely during the training phase; otherwise, the backpropagation process would not function.

During the learning phase, the device attempts to reach its destination by utilising its available actions—such as 'accelerating (moving forwards), steering and braking'—in accordance with one aspect of the present invention. Through its sensors, it detects that, for example, 'accelerating' whilst steering in the wrong direction does not affect the distance value in such a way as to positively alter the motivation value. Each action therefore immediately leads to a change in the motivation value, from which the device recognises whether it is getting closer to the target or not, so that it continues the action (in the case of a positive change) or alters it (in the case of no change or a negative change). Using the matrix of reference values and interpolation, the device can also determine whether a small change in the action (accelerating slightly more or slightly less) alters the motivation value positively or negatively, enabling it to independently optimise the actions required to reach the target. In this way, it can identify and select the faster or shorter route. If, for example,

the target lies diagonally to the right of the device (detectable via the sensors and the internal polar coordinate system) and it wishes to reach it, it can recognise that steering to the left (away from the target) negatively affects the change in the motivation value, whereas steering to the right has a positive effect.

5

Using the stored reference values, the device can, as part of a simulation calculation prior to deciding on a specific action (steering left or right), calculate the respective effect on the motivation value and then immediately select the optimal action (steering right in the above scenario) itself, without this having to be trained beforehand, as would be necessary with a
10 neural network. Of course, this is only possible if a sufficiently large database is available.

In accordance with one aspect of the present invention, restrictive constraints can also be taken into account, for example. Example: A road does not lead directly to the destination, which lies behind a freshly ploughed field, but only at a 45° angle. Now the simulation of the
15 straight path (maximum change in motivation) leads across the field with the risk of getting stuck (massive negative change in motivation) or further along the road (small but still positive change in motivation)

Thus, the device or method according to one aspect of the present invention learns, or is
20 subsequently capable of:

1. Actions alter the motivation value, whereby the inherent aim is to bring this value to a maximum positive value (whether this is 0, the largest or smallest possible value in a particular implementation is irrelevant to the principle).
2. Some actions influence the change in the motivation value more strongly, others less
25 so, so that the action(s) with the strongest influence are technically identifiable and selectable.
3. From the stored reference values and a new situation (initial triplet), the device can, in the sense of a simulation calculation, select or 'decide upon' the actions that are optimal for achieving the target in each case, without having been specifically trained
30 on the initial and target triplets.
4. The ability to simulate not only enables a direct action that transfers the initial triplet directly into the target triplet, but also allows the achievement of the target triplet to be broken down into several individual steps, for example, to avoid violating certain

constraints.

5. If the device can calculate the consequences of its own actions through a technical simulation (when will I be where and at what speed), the consequences of the actions of other, new, unknown objects can also be technically calculated via the simulation. The device copies its own object instance (the ego) with all data, methods and functions, and updates the j e temporary copy (the alter ego) with the current sensor readings.

The internal action instruction is, for example, the device's independent operation of the steering to align the device, for example, with the current target – it results from the simulation and the resulting optimal trajectory for reaching the target triple.

According to one aspect of the present invention, an external action instruction is the external setting of a (new) target triplet.

- 15 An initial triplet A is, according to one aspect of the present invention, a current situation consisting of the device, the other dynamic objects, the constraints of the scenario ..., which 'prompts' the device, through simulation – i.e. essentially the 'theoretical' testing of various developments – to find an optimal path (trajectory) to the target triplet.

- 20 According to one aspect of the present invention, the device's properties comprise its own physical characteristics (mass, size, 'horsepower', etc.), the ability to act independently (brake, steer, accelerate, ...), the ability to alter the environment (bump, nudge, push, etc.)

- 25 According to one aspect of the present invention, see, environmental properties are also the initial triplet: a current situation consisting of the device, the other dynamic objects, the constraints of the scenario, etc.

The target triplet can be described as follows according to one aspect of the present

- 30 invention: in the juvenile training phase, these would initially be purely random direct actions, the effects of which (distances, collisions, acceleration, ..) are stored, such that these then become the effects of direct positive (to be achieved) and negative (collisions: to be avoided at all costs) target triplets. Once a sufficiently large database exists, the device can set itself targets that are no longer directly recognisable, but can only be achieved, for example, after navigating around an obstacle: again, something that children learn after a certain time:

overcoming the 'out of sight, out of mind'

According to one aspect of the present invention, there is therefore no failure in the presence of unknowns. This is a technical solution to a technical problem.

- 5 According to one aspect of the present invention, the control of the device in accordance with at least one predefined target device property and at least one predefined target environment property, using at least one stored action instruction, is carried out as follows: The at least one action instruction specifies how the at least one predefined target device property and the at least one predefined target environment property are set based on the
10 initial triplet and the target triplet

The at least one action instruction converts the initial triplet into the target triplet. There is an initial configuration, and this is converted into a target configuration on the basis of one or more action instructions; in the latter case, this involves a chain of action instructions that convert individual initial triplets into target triplets.

- 15 At this stage, the method or device has not yet learnt how to achieve the predefined target device property and the at least one predefined target environment property on the basis of the learnt mappings or action instructions. The device must therefore select, from the machine-learned body of experience and the development forecasts of a dynamic situation, the action instructions that achieve the target configuration via the best route (optimal
20 trajectory). If the device approaches the target configuration, this action instruction can be used; the motivation value changes positively. If, instead, the device moves away from the target configuration, this action instruction cannot be used by it; the motivation value changes negatively. This can be evaluated or simulated on the basis of the stored action instructions. This results in the possibility, according to the invention, of independently
25 selecting the action instruction(s) that work best, i.e. those that are expected to provide the best path (optimal trajectory) to the target configuration .

- For example, when the proposed device is approaching, a distance can be determined sensorically using the proposed method. It has been learned that a starting triplet is converted into a target triplet. The starting triplet in the present example is 'Actuation' (switch
30 on the means of transport), 'Device properties' (the device can drive, steer, ...) and 'Environmental properties' (the device is located at geographical point A). The target triplet is Activation (the device turns round), Device Properties (the device cannot move due to an obstacle) and Environmental Properties (the device is located at geographical point B, but in front of an obstacle). The device now detects that the motivation value has changed

negatively, as an obstacle has been encountered. The action instruction(s) that maximally improve the motivation value are used, and the motivation value indicates how closely the transition from the initial triplet to the target triplet achieves the predefined target device property and the at least one predefined target environmental property.

- 5 As these triples are stored, this movement can therefore be simulated, and the action instruction that best achieves the target configuration is used. This results in a planned and proactive selection/use of the action instruction(s).

The motivation value may also take into account technical configurations of the device. For instance, high energy consumption may also reduce the motivation value. If the device is
 10 damaged, this may also negatively affect the motivation value. Further examples include excessive time requirements—in which case the shortest time requirement is used—or the occurrence of undesirable target configurations such as damage to external objects. A metric can be provided empirically that correlates the factors influencing the motivation value, and an optimum is determined empirically, which is then implemented in accordance
 15 with the action instruction. The effectors are controlled accordingly.

The task is also solved by a device configured to perform a method according to one of the preceding claims, comprising an initialisation unit configured to initialise by means of randomised actuation of effectors and reading out data from a plurality of internal sensor units configured to measure internal device properties, and a readout unit configured to read
 20 out data from a plurality of external sensor units configured to measure external environmental properties; a storage unit configured to store interrelationships existing between the actuation of the effectors and the read-out device properties and environmental properties as action instructions, such that an action instruction transforms an initial triplet of actuation, device properties and environmental properties into a target triplet; and a control
 25 unit configured to control the device in accordance with at least one predefined target device property and/or at least one predefined target environmental property using at least one stored action instruction which specifies how the at least one predefined target device property and/ or the at least one predefined target environmental property is to be set on the basis of the output triplet and the target triplet.

- 30 The problem is also solved by an arrangement comprising a plurality of devices which are communicatively coupled and exchange action instructions.

The problem is also solved by a computer program product comprising control commands which implement the proposed method or operate the proposed device.

According to the invention, it is particularly advantageous that the method can be used to operate the proposed devices and units. Furthermore, the proposed devices and units are suitable for carrying out the method according to the invention. Thus, each device implements structural features suitable for carrying out the corresponding method. However, the structural features may also be configured as method steps. The proposed method also provides steps for implementing the function of the structural features. Furthermore, physical components may equally be provided virtually or in a virtualised form.

Further advantages, features and details of the invention are apparent from the following description, in which aspects of the invention are described in detail with reference to the drawings. In this context, the features mentioned in the claims and in the description may each be essential to the invention either individually or in any combination. Likewise, the features mentioned above and those described further here may be used individually or in any combination. Components or parts with similar or identical functions are partly designated by the same reference numerals. The terms 'left', 'right', 'top' and 'bottom' used in the description of the embodiments refer to the drawings in an orientation with normally readable figure designations or normally readable reference numerals. The embodiments shown and described are not to be understood as exhaustive, but are intended as examples to illustrate the invention. The detailed description is intended to inform those skilled in the art; therefore, known circuits, structures and methods are not shown or explained in detail in the description, so as not to complicate the understanding of the present description. The figures show:

Figure 1: a schematic flowchart of a method for autonomously controlling a device according to an aspect of the present invention;

Figure 2: a diagram by means of which internal and/or external characteristics can be estimated. In this way, internal and/or external parameters can be measured, and interpolation can be performed between the sample points or further sample points can be extrapolated;

Figure 3: a representation of a real-world situation in which a child wishes to cross a road and the proposed device recognises the situation and acts accordingly;

Figure 4: a real-world road traffic situation in which the proposed device or method is applied in accordance with an aspect of the present invention; and

Figure 5: a further real-world situation in road traffic, in which the device practises cornering in accordance with a further aspect of the present invention.

Figure 6A: a schematic flowchart of a method for autonomously controlling a device in accordance with an aspect of the present invention during the ' ' juvenile phase, with the primary aim of refining its own actions; and

Figure 6B: a schematic flowchart of a method for autonomously controlling a device according to an aspect of the present invention in the adult, application phase for achieving its own objectives, taking into account and predicting the behaviour of other objects relevant to a situation.

Figure 1 shows, in a schematic flowchart, a method for the autonomous control of a device, comprising an initialisation step 100 involving the randomised actuation 101 of effectors and the reading 102 of data from a plurality of internal sensor units configured to measure internal device properties, and the reading 103 of data from a plurality of external sensor units configured to measure external environmental properties; a storing 104 of interrelationships existing between the actuation 101 of the effectors and the read-out device properties and environmental properties as action instructions, such that an action instruction transforms an output triplet of actuation 101, device properties and environmental properties into a target triplet; and controlling 105 the device in accordance with at least one predefined target device property and/or at least one predefined target environmental property using at least one stored 104 action instruction, which specifies how the at least one predefined target device property and/ or the at least one predefined target environmental property is to be set on the basis of the initial triplet and the target triplet.

Functionally, it replaces NNs trained with pre-selected and interpreted data, meaning it is not predetermined to focus on specific external targets or predefined external objects, which leads to much greater freedom in recognition and action. This approach allows short-term behavioural predictions to be made for the observed objects, thereby enabling the development of situations to be assessed and incorporated into the design of one's own actions. It also enables learning through observation and empathy. Aspects of the invention are:

- A central template as the core of the AI for capturing all objects in a situation;
- Data storage that enables the recognition of patterns;
- Behavioural predictions for all observed objects in a situation;
- Empathy capability based on the central template;

- The ability to learn from observations;
 - Recognition of factors that cannot be directly observed;
 - Recognition of natural laws and causal relationships;
 - Computer-compatible forms of theories that can be created, stored, reused and
- 5 continuously improved; and/or
- Self-assessment and the evaluation of self-generated theories as the basis for insight, learning and the selection of the best theory for achieving one's own goals .

As an example, a car that behaves according to these ideas is described. How does it assess a child or a box, how does it handle overtaking on the motorway, or how does it

10 recognise the law of gravity and crosswinds (as an example of recognising causalities and non-manifestable objects). It is further described how learning from observation is possible through empathy. All of this demonstrates that the capabilities of this AI extend beyond driving cars. The essential basis, however, is a physically existing robot in a physically existing environment that observes other objects directly or indirectly.

- 15 The following sections propose several aspects of the present invention which enable an exemplary implementation of the method, the device and/or the system configuration. The following aspects are to be understood merely as examples and may be applied individually or in combination.

Possible hardware for the robot or the proposed device:

- 20 This section describes the technical equipment that the robot or vehicle may possess in accordance with an aspect of the present invention.

Effectors:

- In simple cases, such as a car, the effectors would be – apart from windscreen wipers, indicators, lights, etc. – braking, acceleration and steering, with which the robot interacts with
- 25 its environment; even if it is merely a change in its own position, this also constitutes a change in the situation.

Internal sensors:

- The internal sensors for measuring and providing feedback on the system's own
- 30 characteristics, such as size, weight, speed, acceleration, etc., based on the actions of the

effectors. They are necessary for assessing the system's own actions and as a prerequisite for learning. This enables it to recognise its own actions – for example, the strength of braking and its effect (the extent of deceleration) – to establish connections and store them in memory, thereby learning independently and, consequently, continuously improving its own actions.

As a physical entity in a physical environment, the robot is subject to the laws of physics. A measured acceleration, i.e. a change in velocity, depends not only on mass but also, amongst other things, on how far the accelerator pedal is pressed or how hard the brake is applied. If these laws are not to be hard-coded, but the robot is to discover, store and apply them itself, it requires internal sensors to experience and store the consequences of using its effectors. The internal sensors, for example in a car, therefore record data such as size, speed, acceleration, etc.

External sensors:

In accordance with one aspect of the present invention, the external sensors record the environment (distances, objects, etc.), for example via lidar, image processing, etc.

Naturally, the robot must be made aware of the environment in a suitable manner. What exactly is required is described in more detail below. However, the most suitable systems should not be predetermined, as the information requirements are lower due to the use of the central template. It is not necessary to capture and process all information that is theoretically available.

Possible robot software:

This section describes the components through which the AI described here achieves its intelligence in accordance with one aspect of the present invention.

Data storage:

The data storage is a high-availability data store for data tuples comprising data from the internal sensors, the effectors, the actions and their motivation values (see below). The storage system links actions performed in an appropriate manner with the associated actual, experienced and observed results. It therefore stores only internal data in tables and, for example, no pixels from one or more images. Consequently, there is only data whose internal meaning is known.

Another part of this module, in accordance with one aspect of the present invention, is the repeated checking for internal consistency of the data. This means that the data must not contain any contradictions, structures leading to loops or circularity, etc. The aim is that only one decision may ever result from the data. The routine that checks this is also intended to
 5 remove stored redundant data, not only to keep the database as lean as possible, but also to prevent over-fitting. What is considered redundant is determined by the nature of the theory formation.

Figure 2 illustrates theory formation in accordance with one aspect of the present invention:

10 The central template can, using the simple means described here, replicate any functional interrelationship, store it and use it for forecasting.

Braking or acceleration depend functionally on the mass being accelerated and the force applied. However, this function is initially unknown, as the mathematical function of the change in velocity dV depends on the initial conditions V_n (current velocity, weight, etc.) and the use of the effectors E_m (force on the accelerator pedal, steering, etc.), i.e.:

$$15 \qquad dV = f(V_n, E_m) \qquad (1)$$

This function should be learnt by the robot itself rather than being pre-programmed, as this allows what has been learnt to be extended, corrected and improved in the same way later on. The effective acceleration during braking (or accelerating) is linked via the internal sensors measuring acceleration a to the (mathematically) independent variable of force F
 20 (how hard the respective pedal is pressed) and the mass m according to the law $F = m * a$. The principle by which the robot itself derives this equation is simple. To illustrate, let us take a (different, arbitrary) quadratic function $y = -(x-4)^2+9$, which is to be modelled.

Suppose that initially only the two measured values at points 11 and 12 exist. Now suppose that a prediction p is required at the point $x=3$ (point 2), which, via the straight line a between
 25 11 and 12, leads to a value of $y=3.8$. However, as the error relative to the ex post measured value $y = 8.0$ is far too large, this point 2R at $x = 3.0$ and $y = 8.0$ is incorporated into the memory. Consequently, for a new forecast, either the line b_1 between [11, 2R] or b_2 between [2R, 12] would be used. Over time, many further points (31, 32, 33, ...) emerge in the above example, through which the functional relationship between x and y can be
 30 modelled with arbitrary accuracy by the respective straight lines along the data tuples (c_1 , c_2 , c_3 , ...). The accuracy is limited only by experience not (yet) gained and the size of the

memory. Of course, provided that computing and storage capacities allow, this can be extended to any number of variables; instead of straight lines, hyperplanes $a_1x_1 + \dots + a_nx_n = \text{const}$ are used for forecasting by means of piecewise linear interpolation or extrapolation.

However, to keep the number of stored corner points or data tuples as small as possible, these can also be removed: if an ex post analysis reveals that a data tuple lies so close to, or even on, the line/hyperplane between two neighbouring tuples, then this tuple can be deleted as redundant. This not only reduces storage requirements but also computational effort, and the behaviour adapts increasingly well over time to the actual, albeit still unknown, mathematical function, which then leads to increasingly efficient behaviour over time.

- 10 This approach also prevents overfitting, which must be taken into account when selecting and determining the scope of the training data for a KNN. Redundant data points are thus not stored repeatedly, which would otherwise completely overshadow the rare events—the so-called ‘black swans’—that may have severe consequences.

- 15 Storing functional relationships between variables has a further advantage, as this also ensures that every possible inverse function is stored, since the stored data tuples do not distinguish between which values were ‘dependent’ and which were ‘independent’ when they occurred. In the function above, x is the independent variable and the corresponding y -value can be determined using the function $y = -((x-4)^2)+9$. In the database, one would have to search for the x -value and, if the exact x -value is not explicitly stored, the y -value can be approximated using the two nearest x -values. However, one can simply search the database for the y -value and determine an x -value using piecewise linear interpolation or extrapolation – this is much simpler than attempting to determine the inverse function of, for example, the equation above.

- 25 In this very simple way, any functional relationship can be reproduced with any degree of accuracy; the only prerequisite is sufficient practice, as with a complex sport. Of course, the accepted error ε can also be changed and adjusted over time, whether it needs to be reduced to increase the required precision, or whether it can be increased to reduce the storage and computational effort, as this in turn affects the number of stored key and data points; it is a continuous optimisation within the ongoing process.

- 30 Within the scope of this invention, and in accordance with one aspect of the present invention, a theory regarding the rules and laws of a current situation is therefore defined as a self-contained, consistent dataset from which the robot can calculate the predictions of this

theory by means of piecewise linear interpolation or extrapolation. The data set consists of the robot's own measurement results and, where applicable, the parameters of created alter egos. In this way, the robot can determine, store, reuse and improve each of its theories itself. As part of self-assessment (see below), it is also able to select and apply the best specific theory (data set) for a given situation, the one with the highest competence (see below).

Actions:

According to one aspect of the present invention, an action represents an ordered sequence of effector operations to achieve a goal. Actions can be designed using the functional relationships between effectors, the internal parameters and the known effects, which are known to the robot. Initially, during the juvenile learning phase (see below), simple actions are performed. The AI learns to accelerate and brake, then accelerate, wait and bump into a boundary (externally braked with a 'risk of injury'), accelerate, steer, brake and so on. These simple actions are then combined into increasingly complex 'choreographies', which are further optimised; for example, braking becomes smoother as pedal pressure decreases with falling speed, and cornering becomes more fluid. This ' ' optimisation is achieved through a motivation value (see below), which represents a kind of internal evaluation of the executed action. Thus, the 'choreography' (or part thereof) with the highest motivation value is selected from those stored. In this way, the AI can continuously improve the use of its effectors independently. Furthermore, by relating to the present moment and the experiences gained from actions, one's own situation at future points in time can be predicted, which forms the basis for one's own action planning.

Motivational values:

Motivational values reflect a kind of reward. On the one hand, they are determined internally after a self-set goal has been achieved (successful); on the other hand, a motivational value can also be communicated externally (well done). For example, if a route is covered with sharp steering movements, low speed and yet high lateral acceleration, this sequence of actions receives a lower motivational value than the result of the 'training', after which the same route is traversed faster and yet with lower lateral acceleration. This must, of course, also be stored in memory.

The motivation values thus represent a form of undefined, intrinsic goals. Long-term goals such as the end point of a journey can be linked, or short-term intermediate goals such as

visiting a petrol station when the tank is empty: as the fuel level drops, the motivation value for 'refuelling' slowly rises until it exceeds that of the long-term goal. The pursuit of the long-term goal is interrupted in favour of refuelling and then resumed afterwards.

Goal system:

- 5 According to one aspect of the present invention, the robot is always 'switched on', although it may, of course, have a power switch. It is therefore always performing an action. However, this also includes doing nothing (charging) or, whilst in a state of rest, optimising its own long-term memory of possible actions (sleeping). The action selected is always the one that currently has the highest motivation value. If an action not currently being performed
- 10 achieves a higher motivation value than the one currently being performed, the latter is cancelled, terminated or interrupted so that the action with the highest motivation can be carried out. If, on the journey from Munich to Hamburg, the motivation of such a car to refuel or recharge its batteries exceeds that of reaching Hamburg, it will interrupt the journey at the next opportunity and head for a petrol station.
- 15 In the juvenile phase (see below), i.e. the learning phase, of such a robot or car, the motivations for driving, braking, accelerating and cornering over short distances will be assigned high motivation values, simply because the internal long-term memory signals that learning should take place or that errors in action should be reduced in order to refine its own behaviour. Later, in the adult phase (see below), when such behaviour causes no or
- 20 only insignificant changes in the data storage, and error reduction approaches zero, such behaviour becomes 'boring'; it receives only low motivation values, and then, for example, energy saving is prioritised.

Thus, in accordance with one aspect of the present invention, this goal system avoids committing to reality, which would, after all, entail an interpretation of it that might prove to be

25 incorrect in the future.

The central template, the Ego:

The central template represents the core and basis of this AI in accordance with one aspect of the present invention. It is called the ego because it places itself at the centre of its data processing and initially views and evaluates everything from its own perspective. Therefore,

30 the ego does not need to be provided with or trained on data and its meaning (for example, image recognition); it creates everything itself. According to the invention, the central

template or the Ego in the device may have arisen independently through the autonomous learning of its own capabilities, or it may have been copied and implemented from a comparable device.

Subjective polar coordinate system:

- 5 All data from the observations and experiences of this AI are subjective or, in accordance with one aspect of the present invention, are made from the perspective of the ego. The spatial concept also follows this idea. No Cartesian coordinate system is designed with an assumed origin of the observed environment, in which the respective coordinates are assigned to each recognised object. Instead, a polar coordinate system may be used. This
- 10 not only automatically establishes the mathematical origin within the ego, but also the ego's subjective orientation with front-back, up-down, right-left, including the motion vector. The surrounding (empty) space is taken as a given. The observed objects are recorded solely in terms of their distance from one's own vantage point—that is, the origin of the polar coordinate system—and the angle. Empty space in the physical sense is therefore not
- 15 assumed to be a separate entity or dimension. It is there and is simply an option to move there, unless another object is there, standing or moving (towards it); in that case, the 'space' would be occupied or the space would not be empty. A further advantage of subjectivist polar coordinates arises in the following for the concept of alter egos (see below).

20 Juvenile phase:

- Within unconditional constraints, such as the integrity of the observed objects and the robot itself, the maintenance of operational readiness (battery charge level), objectives such as gentle acceleration and braking and optimised cornering, and others, the robot should/can learn the use of its effectors and the resulting consequences itself, primarily in order to
- 25 coordinate its effectors with its internal and external sensor systems and, through the intensive training 100–105, intermittently interrupted by rest periods, for example to recharge the batteries and maintain the internal database (redundancy, consistency, etc.), until the identified discrepancy between planning and result has been sufficiently resolved. In the case of a car, this would primarily involve the use of the accelerator, steering, braking,
- 30 reaching a destination (success) or missing it (error), etc. Initially, the ego sets itself simple goals for simple actions, which are carried out and stored along with their results; these are then later combined into increasingly complex 'choreographies'. This training phase comes to an end when the actions relating to self- or externally-set goals ('drive over there') can no

longer be optimised, that is, when the mathematical module can no longer reduce the error, or can only do so marginally, and/or the database can hardly be optimised any further (number and location of points in memory). As the rate of error reduction slows down, the juvenile phase draws ever closer to its end. A KNN, on the other hand, requires a large

5 volume of specifically selected and interpreted training data, whereas the ego needs only a real-world environment – one might say a playroom – where it cannot cause any damage, in order to test and develop its own abilities. It follows intrinsic goals rather than externally prescribed ones like a KNN, and thus does not interpret the environment or the world.

Adult phase:

- 10 In the juvenile phase, the device has learnt the ordered sequence of its effector actions for achieving the set goals. What value do the insights and experiences gained in the juvenile phase hold? All data about the world that the ego has come to know in the world in which it moves are its own data, its own observations and its own experiences. In a broader scientific-theoretical or philosophical sense, it must be noted that this data is true from the
- 15 ego's perspective, in the sense of 'that is how it is' or 'that is how it was'. This thus constitutes a natural and firm foundation for knowledge. Naturally, the ego must maintain its internal data. In addition to the optimisation already hinted at above, this also entails ensuring that it is free from contradictions and errors. Otherwise, as the manufacturer or operator of the robot, one could not assume that the robot will achieve the intended goals
- 20 with its programmed actions.

- The changes in the environment that are experienced and caused by the robot itself, such as a change of location resulting from accelerating, steering and braking, are experienced and stored by linking internal and external data. This is sufficient, and there is no need for any deeper understanding of the underlying causes beyond the robot's own effectors – in the
- 25 sense of questions such as: why does pressing the accelerator lead to acceleration, steering to cornering, and braking to deceleration? For a car's 'ego', the fact that 'there are three options' is enough to change location in a targeted manner. At this level, the 'That's how it is' is what matters, not the 'Why'. A pigeon on the road certainly has no idea how a car or a cyclist works. But if such a 'road user' were to pass it at a sufficient distance, it would remain
- 30 put; yet if the direction of movement were to change towards it, it would fly away. It may not know why, but it is aware of the possible behaviours and therefore pays close attention.

Alter Ego:

When the (adult) ego recognises an object in the observable environment via its external sensory systems, the ego simply creates a copy of itself and adjusts the parameters of the copy in accordance with the observed characteristics. The ego is thus the central template for all observable and non-observable objects and phenomena. Hence the names 'ego' and, as a consequence, 'alter ego' for the copies.

By using the ego as the central template, there are no longer any unknown objects . There are only unknown parameters, which can, however, be measured via external sensors, estimated from one's own experience (keyword: prejudice) and whose possible range can be narrowed down through further observations.

Just as the ego can calculate its behaviour using its methods, stored data and parameters and predict its own situation, the behaviour of the alter egos can also be calculated using the same methods, data and adapted parameters. The assumptions thus attributed to the alter egos are limited to the same physical laws and to the fact that short-term goals can be derived from the observed orientation and direction of movement. Whether what is observed is a box on the street, a kangaroo in Australia or an elderly lady with a walking frame.

In terms of programming, the ego should be available as an object. Then, for a newly encountered object, a copy of the ego can very simply be created, after which the parameters of the copy of the new object can be adjusted using external sensors and the system's own experience. There could be a list of the alter egos for the current situation, into which the new copy is inserted or created:

```
Class CEgo { ... };
```

```
CEgo Ego( ... );
```

```
// Start of training & learning, juvenile phase
```

```
...
```

```
// Start of the adult phase
```

```
CEgo AlterEgos[];
```

```
...
```

```
// new object detected
```

```
AlterEgos[i] = Ego.clone(); // Clone the ego for the object
```

```
AlterEgos[i].adjust(...); // Adjusting the parameters of the new object
```

```
...
```

The ego has thus internally created an alter ego as a predictable representation of an
 5 observed object in the environment and, as shown in Figure 6B, can form the inner loop for
 the continuous updating of the observed situation, consisting of object recognition 311,
 updating of the observable object parameters 312, forecasting the behaviour of the objects
 313, adjusting the sequence of effector deployments 314, and a second, outer loop following
 the attainment of the target and the entry into a rest and (also) charging phase 315, until it
 10 continues with the initialisation 310 for a new target.

It makes no difference whether the observed object is known or entirely new. If it is new, the
 range of possible feature values is greater, which requires increased attention (sampling
 frequency). This has the following advantages:

- Every observed object is fundamentally known!
- 15 - Behavioural prediction of all external objects using our own methods and data!
- Humans no longer need to intervene to programme in missing information.
- The ego continuously learns and improves its behaviour.
- The ego is capable of empathy and is aware of the current situation (see below).
- The ego can learn from observing the alter egos (see below).
- 20 - Alter egos can embody unknown causalities and rules (see below).

Figure 3 shows Scenario 1, 'a child', in accordance with one aspect of the present invention.

When the ego, for example the AI of a simple car, encounters another road user – in this
 example, a child – it creates a copy of itself and adjusts the parameters (distance d from the
 ego, orientation, speed v_K , ...):

25 In this way, the alter ego can use the ego's methods to predict where it will be and when.

And so, by comparing this with its own calculation of where it will be and when, the ego can determine whether a potentially dangerous situation might arise, in order to react accordingly, i.e. to brake. In this way, the child's short-term behaviour can be predicted using the system's own methods and experience. NOTHING needs to be specially programmed or
 5 trained for the child that would not already be available when this AI encounters a child for the first time.

Figure 4 shows Scenario 2 'Motorway' in accordance with one aspect of the present invention.

Another example that more clearly illustrates how powerful this seemingly simple idea of
 10 alter egos is, is a situation on a motorway involving a lorry with two cars behind it and a car in the overtaking lane, the second of which is our own car's alter ego. Based on its own knowledge of Newton's second law ($F=m*a$) regarding braking force and mass, and using the parameters of the observed cars and the lorry – namely, estimated mass, measured speed and estimated braking effect – the ego can, using its own methods alone, predict the
 15 braking distances of all road users in a potential accident scenario, thereby determining its own safe following distance. But let us first consider the 'thoughts' or assumptions of the ego regarding the other road users in order to predict their actions:

- The LKW0 is travelling at a certain speed, and the ego (PKWEgo) has learnt (without needing specific training) that road users of this size very rarely exceed it. The LKW's
 20 alter ego will therefore not predict any change in speed.
- The PKWEgo, the focus of the analysis, would be capable of driving faster and would overtake Truck0 whilst taking other road users into account. Overtaking would get the PKWEgo to its destination faster and thus currently has a higher motivational value; it
 25 prepares for the overtaking manoeuvre.
- Car1, another alter ego, thinks the same as the CarEgo by design, because the ego would act in the same way in its place. The ego therefore assumes that Car1 also wants to overtake Truck0. Naturally, the alter ego of Car1 'sees' CarEgo and will
 30 carry out the overtaking manoeuvre (as assumed by the ego) whilst taking into account the presence of CarEgo – for example, by indicating and steering with particular attention to CarEgo and any other road users in the overtaking lane. That would be what the ego of CarEgo 'thinks' and how it would predict the behaviour of Car1.

- Another potential road user, Car2, is approaching at high speed in the overtaking lane. For this too, the ego creates an alter ego and formulates a prediction of its actions based on the clear road ahead.

5

- As shown above, the programme, the Ego, can predict the entire situation involving all relevant road users and plan its own sequence of actuator inputs accordingly, so that no dangerous situation arises.

Alter Ego, consciousness:

- 10 The ego thus possesses an internal, complete picture of the observable environment, including all recognised objects, and the ability to predict their behaviour in the short term. This would be a possible definition of consciousness—one that is genuine, not feigned or imitated—and one that could be implemented in programming.

- 15 The short-term behavioural prediction, based on the ability to put oneself in the place of the observed objects, means here viewing the situation with all data converted for the alter egos, and calculating the development of the situation based on the calculated probable behaviour of the observed objects.

The ego in the motorway situation described above can, for all observed objects and without the need to program extra routines:

- 20
- predict their behaviour, assuming that all road users wish to proceed as quickly as possible and without an accident,
 - continuously adjust the predictions based on the observed actual behaviour (braking, accelerating, steering, indicating, flashing headlights, etc.) and
- 25
- overtake the lorry itself at a suitable moment.

The possible behaviour of all parties involved in this situation can be assessed by the ego without any difficulty, by applying its own methods and experiences (data) to them with the appropriate parameters. Doesn't a human do exactly the same?

- 30 The examples make it clear that the basis, the foundation and the starting point of this AI is always the ego, the central template. The more sophisticated its capabilities in terms of internal and external sensors, and the better its training during the juvenile phase, the more

intelligent its behaviour and the greater its potential to predict the behaviour of observed objects. If the ego were also given a parameter for its own vulnerability – in the case of a car, one would probably speak of deformability – the ego could also assess the risks of a strong or weak contact. BUT, the ego needs to be built, programmed and trained ONLY (once),

5 although the latter part can, of course, also be transferred from a trained ego. Observed objects need neither to be identified (KNNs and their training data), nor does the behaviour of the identified objects need to be programmed. One could, of course, object that the ego is not sufficient to predict the behaviour of other road users. But one need only consider how

10 humans attempt to assess the behaviour of other road users. Humans, too, do not know the long-term destinations of road users on a motorway; they, too, can only surmise the short-term goals of other road users (to get ahead as quickly as possible without an accident) and prepare for them.

The robot's capabilities:

Empathy:

15 It was shown above how the ego, with the help of its alter egos, can put itself in the situations of the observed objects in its environment in order to calculate their behaviour from their perspective. One could call this empathy, provided the term is not reserved exclusively for humans.

Learning from observation:

20 Learning from observation means observing behaviour that may be better than one's own and adopting it where appropriate and where possible. The basis is the comparison between the behaviour of others and one's own, and this comparison is made possible by the empathy defined above.

25 Empathy of the ego is realised here via alter egos, through which the ego places itself in the situation of other observed objects in order to predict their behaviour from their perspective of the environment. Suppose there is now a difference between predicted and observed behaviour. And if the alter ego now modifies the sequence of effector actions in such a way that the observed behaviour is replicated, the ego has the opportunity to copy the observed behaviour, provided it offers an advantage.

30 Figure 5 shows Scenario 3: different 'cornering behaviour'.

Let us imagine that the ego has learnt to drive through a bend always at the same distance from the right-hand edge of the road, the line LE. Now it observes a car ahead that cuts the bend by turning in earlier but less sharply, thereby also reducing lateral acceleration and thus driving round the bend faster, on the dotted curve LB.

- 5 Seen in this light, the empathy described here is a prerequisite for learning through observation and imitation of others' behaviour. (What is to prevent us from assuming something similar applies to humans, i.e. learning through imitation based on empathy?)

If the data from the external sensors, converted to the situation and position of the alter ego, is linked in this way to the stored, simple and more complex sequences of actions of the ego—which have, after all, been copied into the alter ego—the ego is able to learn from observation: It can recognise the difference between the self-planned sequence of actions of its own effectors (always maintaining the same distance from the edge of the road) and the observed sequence (braking and turning in earlier, smaller steering movements and higher cornering speed, and accelerating earlier), and determine that it could drive through the bends faster in this way. This form of learning is certainly faster than the usual trial and error.

Recognising causal laws:

In the section on theory formation, we explained how any observed functional relationship can be modelled. Let us use this ability to recognise laws of nature, for which alter egos may also be used where necessary and appropriate. Of course, these do not lie or drive on a road, but they cause changes that can be detected by external and/or internal sensors.

Let us imagine an experiment in which our AI ego observes a falling apple. It creates an alter ego and notices the accelerated movement towards the Earth. Now, the ego's first assumption will be that the apple itself caused the acceleration; like the car ego, it has its own 'accelerator pedal' to speed up. But then the ego will observe that no apple on the ground moves of its own accord and that other objects also fall to the ground and do not move any further once there. Furthermore, the ego is likely to have learnt about the law of gravity itself. On a sloping road, without pressing the accelerator, it would have noticed an acceleration and learnt that ' ', but only 'downhill', whilst 'uphill' it must press the accelerator harder. The steeper the angle α , the greater the force according to this formula we are familiar with (where g = acceleration due to gravity, 9.81 m/sec^2):

$$F = m * g * \sin(\alpha) \quad (2)$$

We recall that the ego does not, of course, explicitly know this equation or the equation for free fall, i.e. the law of gravity ($\sin(\alpha=90^\circ) = 1$), but through the data points and piecewise linear interpolation or extrapolation, it has an implicit understanding of this relationship between the relevant quantities.

- 5 The ego therefore behaves as if it knew that, as a physicist would say, as a body with significant mass, it is subject to the above law of gravitational attraction. When it now sees another object and creates its alter ego, that alter ego is also subject to this gravitational attraction. In this way, this insight effectively acquires the status of a law of nature that acts upon all bodies without needing to be specifically programmed. If the behaviour of such an
- 10 ego were observed from the outside, one could not say whether it actually knows the law of gravity, as a physicist does, or whether it is merely feigning this insight. Even if the ego sees a pigeon sitting on the street that flies away as soon as the ego approaches, it can recognise that the flapping of wings and upward acceleration correspond.

- Another example would be a strong crosswind. Let the ego be a large, empty van. When
- 15 driving straight ahead normally, the lateral acceleration is zero. But then, whilst driving straight ahead, the van is suddenly pushed sideways and the lateral acceleration sensor is triggered. In this case, the ego can simply create a new alter ego for the unexpected force and the unknown cause, and link it to this lateral acceleration. If the ego can then also capture data from the surroundings – such as whether it is in a forest or on a plain – or the
- 20 movement of branches, and add this to the data store, the ego has not only recognised a previously unknown variable (crosswind), it has also identified a possible causal link. If the ego could also see and assess the extent of the branches' movements, it could even roughly estimate the strength of the force acting sideways and thus the probable influence on driving behaviour. In this way, the ego has itself formulated a theory incorporating a new variable.
- 25 Much like the physicists who introduced dark matter and dark energy, even though they are neither visible nor observable, or like Nobel laureate Peter Higgs, who in the 1960s conceived of a particle that was then first detected in 2012 at CERN . The alter ego, that is, the copy of the self, can of course never recognise the true nature of a phenomenon using this approach in the sense of a search for truth, but it is sufficient for a competent grasp of
- 30 the situation.

Self-assessment:

The first form of knowledge acquisition has already been described. In the section on 'recognising causal laws', the example of a crosswind was discussed. A force (initially)

unknown to the ego suddenly generates lateral acceleration on a straight stretch, a measurable effect on the robot or the ego. To account for this, it creates an alter ego. However, further information is initially lacking to signal to the ego when the phenomenon occurs and with what intensity. Only by observing a temporal connection between the movements of the branches – sometimes stronger, sometimes weaker – and its own lateral acceleration – sometimes absent, sometimes present – can the ego link the two and infer the same cause for both. What is happening here? The ego cannot detect the crosswind itself, but it can detect its own reaction to it (lateral acceleration) and the reaction of the branches. In doing so, it generates an element that sometimes has an effect on the ego and sometimes does not, but which affects not only the ego but also other observable objects.

The second form of knowledge acquisition is based on comparing one's own behaviour with that of other objects. A prerequisite is the assumption that the behaviour of the observed objects is 'rational' and not randomised, although this could also be recognised as such, in which case the comparison would no longer be a criterion for potential knowledge gain. Two different behavioural comparisons are carried out, enabling the ego to determine whether its own competence (see below) in a situation is inferior, equivalent or superior to that of the observed objects. This means that the observed objects can perceive more, the same amount or less of the situation. An observed van suddenly slows down on a straight road. It appears to see something unknown to the observing ego; the ego perceives less than the object. If an inferior competence has been identified, this can be understood as a reason to investigate such situations specifically, in order to raise one's own competence to at least the level of the others. The comparison is not about whether the behaviour is right or wrong – that would only lead to the familiar problems of theories of truth – but simply whether and when behaviour changes and whether patterns of behaviour recur. The first behavioural comparison concerns different situations: do the observed objects display different behaviours in situations that are different for the ego? The second comparison concerns identical situations: does the behaviour of the observed objects repeat itself in situations that are the same for the ego, or does it vary? From these observations of multiple situations over a longer period, the ego can conclude

- that its own competence is inferior if, repeatedly, observed objects display different behaviours in situations that are the same for the ego and also vary their behaviour in situations that are different for the ego,
- that its own competence is superior if, repeatedly, observed objects repeat their

behaviour in situations that are the same for the ego but also do not vary their behaviour in situations that are different for the ego,

- that one's own competence is equivalent if objects observed repeatedly in situations that are the same for the ego repeat their behaviour but vary their behaviour in situations that are different for the ego.

Through these simple comparisons, the ego can recognise the extent of its own competence in certain situations compared to other objects. It should be noted that this is not an attempt to compare with the 'truth of the real world'.

- 10 From these comparisons, the ego can assess itself and, for example, deduce whether it should behave more cautiously in certain situations and attempt to identify what is not (yet) recognisable. It can be understood as an internal mandate to conduct targeted research as part of 'constant learning and the continuous expansion of insight and knowledge' (see above).

15 Competence:

However, the results of these comparisons can also be used externally to assess the competence of the robot's capabilities in order to decide on its potential applications. In contrast to the truth of knowledge about the real world that the ego has acquired, there are no linguistic problems when it comes to comparing competences and determining greater or lesser competence.

- 20 However, the concept of truth in relation to theories has another important aspect, which competence must also fulfil. Theories enable predictions. It is therefore important to know the quality of a theory and, consequently, of its predictions. Generally, attributing truth to a theory serves as a criterion for being able to apply it and being permitted to do so, with the advantage that the criterion of truth automatically excludes all competing theories, thus
- 25 avoiding the problem of deciding which theory to use. The concept of competence must achieve something comparable.

- 30 Competence is defined and used here as the grasping of the relevant influencing factors of a situation or similar situations. It is thus, first and foremost, a limited criterion, the degree of which can, however, be determined in a comparative procedure, in stark contrast to the truth of a theory, which presupposes validity that is unlimited in time and space. This, however,

can neither be proven nor does the criterion allow different theories to be compared with one another.

Competence is therefore better suited to describing a robot's capabilities than the concept of truth, and, unlike truth, it can be determined by the robot itself through a formal procedure for self-assessment.

Figure 6A shows, in a schematic flowchart, a procedure for the autonomous control of a device during the juvenile or training phase. The aim and end goal of the phase is the sufficiently precise use and deployment of the effectors, with the smallest possible error. Following initialisation 200, there is randomised actuation 201 of effectors and the reading out 202 of data from a multitude of internal sensor units configured to measure internal device properties, and a reading out 203 of data from a multitude of external sensor units configured to measure external environmental properties; a storage 204 of interrelationships existing between the actuation 201 of the effectors and the read-out device properties and environmental properties as action instructions, such that an action instruction converts an output triplet of actuation 201, device properties and environmental properties into a target triplet; and controlling 205 the device in accordance with at least one predefined target device property and/or at least one predefined target environmental property, using at least one stored 204 action instruction which specifies how the at least one predefined target device property and/ or the at least one predefined target environmental property is to be set on the basis of the initial triplet and the target triplet.

Figure 6A shows the Juvenile Phase:

- 200 Initialisation in preparation for learning
- 201 Randomised actuation of the effectors
- 202 Reading out 202 a large number of internal sensor units
- 203 internal device characteristics and a readout 203 from a plurality of external sensor units
- 204 Storing 204 of interrelationships
- 205 Detection of the motion precision error, immediately if too large, continue at 201

Optional: Rest phase for recharging and ex post for data processing:

Errors, redundancy, consistency,...

Figure 6B shows the adult phase:

- 310 Initialisation as preparation for achieving a longer-term goal
- 5 311 Recognition of the objects 311 in the current situation
- 312 Creating or updating the alter egos or their parameters 312
- 313 Predicting the behaviour 313 of the observed objects using the current alter egos
- 10 314 Calculation of the optimal use of effectors 314 for further autonomous target tracking
- 315 Rest phase 315 for recharging and ex-post data processing:
- Errors, redundancy, consistency, etc.

In this context, the terms 'robot', 'device', 'system' and 'ego' are used synonymously.

Claims

1. A method for the autonomous control of a device, comprising:

- an initialisation (100) by means of randomised actuation (101) of effectors and reading out (102) of a plurality of internal sensor units arranged to measure internal device properties,
5 and a reading out (103) from a plurality of external sensor units arranged to measure external environmental properties ;

- storing (104) interrelationships existing between the actuation (101) of the effectors and the read-out device properties and environmental properties as action instructions, such that an action instruction transforms an input triplet of actuation (101), device properties and
10 environmental properties into a target triplet; and

- a controlling (105) the device in accordance with at least one predefined target device property and at least one predefined target environmental property using at least one stored (104) action instruction , which specifies how the at least one predefined target device property and the at least one predefined target environmental property are set on the basis
15 of the initial triplet and the target triplet, wherein the at least one action instruction is used which maximally increases a motivation value, and the motivation value indicates how closely the conversion of the initial triplet into the target triplet achieves the at least one predefined target device property and the at least one predefined target environment property.

20 2. The method according to claim 1, characterised in that the method is iterated in a learning, randomised manner such that new action instructions are constantly identified, each of which converts a source triplet into a target triplet, and these action instructions are stored for controlling (105) the device.

25 3. The method according to claim 1 or 2, characterised in that the reading out (102, 103) of internal and/or external effectors is carried out in such a way that individual sample values are stored and intermediate values are interpolated and/or extrapolated.

4. The method according to one of the preceding claims, characterised in that the target device property and/or the target environment property is influenced, at least temporarily, by a device property and/or an environment property .

30 5. The method according to one of the preceding claims, characterised in that a plurality of

action instructions are combined into a choreography which transforms an initial triplet into a target triplet.

6. The method according to any one of the preceding claims, characterised in that the method is executed on a first device and, by means of external sensor units of the first device, a second device is detected which also executes the method and whose generated interactions are stored on the first device.
7. The method according to claim 6, characterised in that the first device is controlled (105) on the basis of the detected interactions of the second device.
8. The method according to one of claims 6 or 7, characterised in that correlations between the parameters and actions of the detected second device are used to generate action instructions for the first device.
9. The method according to one of the preceding claims, characterised in that action instructions are simulated prior to their implementation and the maximum motivation value is determined therefrom.
10. The method according to one of the preceding claims, characterised in that internal sensor units detect a motivation value, an actuator state, an actuator setting, an actuator position, a size, a weight, a dimension, a speed, an acceleration, a deceleration, a battery level, a range, a position, a processor utilisation, a memory utilisation and/or a system parameter of the device.
11. The method according to one of the preceding claims, characterised in that technical factors influencing the motivation value are empirically determined and these are correlated with a stored metric for determining the maximum motivation value.
12. A device configured to perform a method according to any of the preceding claims, comprising:
 - an initialisation unit configured to initialise (100) by means of randomised actuation (101) of effectors and reading out (102) from a plurality of internal sensor units configured to measure internal device properties, and reading out (103) from a plurality of external sensor units configured to measure external environmental properties;
 - a storage unit configured to store (104) interrelationships existing between the actuation

(101) of the effectors and the read-out device properties and environmental properties as action instructions, such that an action instruction converts an output triplet of actuation (101), device properties and environmental properties is converted into a target triplet; and

- a control unit configured to control (105) the device in accordance with at least one predefined target device property and/or at least one predefined target environmental property using at least one stored (104) action instruction, which specifies how the at least one predefined target device property and/or the at least one predefined target environmental property is set on the basis of the initial triplet and the target triplet, wherein the at least one action instruction is used which maximally increases a motivation value, and the motivation value indicates how closely the conversion of the initial triplet into the target triplet achieves the at least one predefined target device property and the at least one predefined target environment property.

13. A system arrangement comprising a plurality of devices according to claim 12 which are communicatively coupled and exchange action instructions.

- 14. A computer program product comprising instructions which, when the program is executed by at least one computer, cause the computer to perform the steps of the method according to any one of claims 1 to 11.

- 15. A computer-readable storage medium comprising instructions which, when executed by at least one computer, cause it to perform the steps of the method according to any one of claims 1 to 11.